

孙诗瑶, 王晓丽, 曹子林, 等. 千针万线草根转录组测序及黄酮类化合物合成相关基因挖掘 [J]. 福建农业学报, 2022, 37 (8): 1008–1015.

SUN S Y, WANG X L, CAO Z L, et al. Transcriptome Sequencing and Identification of Genes Associated with Flavonoid Biosynthesis in *Stellaria yunnanensis* Roots [J]. *Fujian Journal of Agricultural Sciences*, 2022, 37 (8): 1008–1015.

千针万线草根转录组测序及黄酮类化合物合成相关基因挖掘

孙诗瑶¹, 王晓丽², 曹子林^{1*}, 张博林¹, 郭紫微¹, 赵思懿¹

(1. 西南林业大学生态与环境学院, 云南 昆明 650224; 2. 西南林业大学林学院, 云南 昆明 650224)

摘要: 【目的】通过高通量测序获取千针万线草根转录组信息, 在分子水平上进一步研究千针万线草根黄酮类的生物合成。【方法】采用高通量测序技术平台 Illumina Novaseq 6000 完成千针万线草根的转录组测序, 对 unigenes 进行功能注释和解析黄酮类化合物的生物合成相关基因。【结果】总共得到 34 137 条 Unigenes, 平均长度为 1 093.58 bp。注释到六大功能数据库 (NR、egg NOG、Pfam、Swiss-Prot、GO、KEGG) 中的 Unigenes 总数达到 22 369 条。千针万线草根 Unigenes 匹配至 NR 数据库的有 21 510 条, 与藜科的甜菜、藜麦和菠菜有高度同源性; 19 414 条 Unigenes 在 egg NOG 数据库得到 19 980 个注释并划分为 23 类; 19 942 条 Unigenes 在 GO 数据库中获得 69 356 个注释, 根据功能划分为细胞组分、分子功能及生物过程三大类, 分别对应 14、16、23 个亚类, 其中涉及生物过程较多; 6 505 条 Unigenes 富集在 KEGG 数据库的 131 条通路中, 代谢相关的通路占比最大, 其中筛选获得 80 个与黄酮类化合物代谢相关的基因, 共编码 16 个关键酶。同时, 有 724 个 Unigenes 被注释为转录因子。【结论】对千针万线草根进行转录组测序及基因功能注释, 并筛选到多个与黄酮类化合物合成相关的 Unigenes, 研究结果丰富了千针万线草根的遗传信息, 可以为进一步鉴定该物种药用成分合成的关键基因及其调控机制提供一定的参考依据。

关键词: 千针万线草; 根; 转录组; 测序; 基因功能注释

中图分类号: S 571.1

文献标志码: A

文章编号: 1008-0384 (2022) 08-1008-08

Transcriptome Sequencing and Identification of Genes Associated with Flavonoid Biosynthesis in *Stellaria yunnanensis* Roots

SUN Shiyao¹, WANG Xiaoli², CAO Zilin^{1*}, ZHANG Bolin¹, GUO Ziwei¹, ZHAO Siyi¹

(1. College of Ecology and Environment, Southwest Forestry University, Kunming, Yunnan 650224, China;

2. College of Forestry, Southwest Forestry University, Kunming, Yunnan 650224, China)

Abstract: 【Objective】Transcriptome of *Stellaria yunnanensis* root was sequenced and genes associated with the flavonoid biosynthesis identified. 【Method】Transcriptome of *S. yunnanensis* root was sequenced using the high-throughput Illumina Novaseq 6 000. Functional annotation of the unigenes were conducted and genes related to the biosynthesis identified.

【Result】There were 34 137 unigenes found with an average length of 1 093.58 bp. They annotated to 6 functional databases, including NR, egg NOG, Pfam, Swiss-Prot, GO, and KEGG, that numbered 22 369 in total. There were 21 510 unigenes annotated in NR database, which had high homology with *Beta vulgaris*, *Chenopodium quinoa*, and *Spinacia oleracea* of Chenopodiaceae, while 19 980 from 19 414 unigenes in egg NOG in 23 categories, and 69 356 from 19 942 unigenes in GO in 3 major functional categories of cellular components, molecular functions, and, especially, biological processes with 14, 16, and 23 subcategories, respectively. Of the unigenes, 6 505 were enriched in 131 pathways in KEGG with the largest proportion related to the metabolism pathways, 80 associated with the flavonoid biosynthesis encoding 16 key enzymes, and 724 annotated as transcription factors. 【Conclusion】The high-throughput transcriptome sequencing and gene function annotation on the *S. yunnanensis* roots were successfully performed. Several unigenes related to the flavonoid synthesis were identified. The

收稿日期: 2022-04-11 初稿; 2022-07-28 修改稿

作者简介: 孙诗瑶 (2000-), 女, 本科生, 研究方向: 生态学 (E-mail: 1435377729@qq.com)

* 通信作者: 曹子林 (1974-), 男, 博士, 副教授, 研究方向: 生态学 (E-mail: 1410214728@qq.com)

基金项目: 云南省高等学校大学生创新创业训练计划项目 (2020106770); 云南省高校优势特色重点学科 (生态学) 建设项目 (050005113111); 西南林业大学科研预研基金 (18200115)

information would facilitate further studies on the key genes and mechanisms in the formation of medicinal important functional ingredient in the plant.

Key words: *Stellaria yunnanensis*; root; transcriptome; sequencing; gene functional annotation

0 引言

【研究意义】千针万线草 (*Stellaria yunnanensis*) 为石竹科 (Caryophyllaceae) 繁缕属 (*Stellaria*) 的一种多年生草本植物, 其根具有补气健脾、养肝活血的药用价值^[1]。据研究, 繁缕属植物普遍含有大量的黄酮类化合物, 从中分离出的化合物主要有黄酮类、异黄酮类和黄酮醇类物质3类, 具广谱抗病毒药物的功效^[2]。因此, 研究千针万线草根的转录组和黄酮类代谢, 对其开发利用具有重要意义。【前人研究进展】谭瑞璞等^[3-4]优化了千针万线草中总黄酮的提取工艺, 并测定不同产地千针万线草中牡荆素的含量。在模式植物中, 与类黄酮合成相关的功能基因已被克隆并进行广泛的研究^[5]。目前, 黄酮类化合物的生物合成路径已较为明确^[6]。而就缺乏基因组信息的物种而言, 对于植物次生代谢中类黄酮合成途径的研究, 多停留在红花、中麻黄、半枝莲、金钱莲、铁皮石斛、藤茶等药用植物的基因挖掘和鉴定上^[7-12]。【本研究切入点】千针万线草基因组和转录组数据缺乏, 制约了该物种种质资源鉴定、遗传多样性分析和分子机制等研究的开展。因此, 对千针万线草根进行高通量转录组测序及基因功能注释, 可为后续研究奠定基础。【拟解决的关键问题】利用 Illumina Novaseq 6000 高通量测序平台对千针万线草根进行转录组测序, 应用公共数据库对其基因组进行功能注释及分类, 借助 KEGG 挖掘出该物种黄酮类化合物合成相关基因, 为开展黄酮类化合物合成的调控机制研究提供依据。

1 材料与方法

1.1 样品采集

千针万线草 (*Stellaria yunnanensis*) 根采自会泽县迤车镇箐口村山坡上, 埋于土中带回学校, 种于花盆放在苗圃, 每天正常浇水萌出根蘖苗后, 经西南林业大学曾觉民教授鉴定为千针万线草。于2021年7月取其根部以无菌水洗净, 滤纸吸去水分后装入冻存管, 液氮速冻、备用。

1.2 RNA 提取、文库的构建及测序

提取根的 total RNA 后, 检测 RNA 的浓度、纯度和完整性。参照文献[13]的方法, 将带有 Oligo (dT) 的磁珠和 ployA 配对 A-T 碱基。mRNA 从 total RNA 中分离后, 富集得到完整的 RNA 序列, 添加 fragmentation

buffer 随机断裂 mRNA, 利用磁珠筛选分离获得约 300 bp 小片段; 加入六碱基随机引物, 在逆转录酶存在下, 以 mRNA 作为模板, 反转录合成一链 cDNA, 再进行二链合成, 进而形成稳定的双链结构。连接 adaptor, 对目的条带大小合适的片段用 2% 的琼脂糖胶回收后, 再用 PCR 技术进行扩增; 最后, 对短序列片段用 Illumina Novaseq 6000 进行测序。

1.3 测序原始数据过滤及组装

测序得到原始数据后, 用软件 Seq Prep 将数据中带有大量接头序列、低质量序列末端、adapter 以及 N 的比例高于 10% 的碱基去除, 获得高质量的 clean reads。使用 Trinity 软件^[14]组装 denovo, 用 TransRate^[15]评估从头拼接结果中普遍存在的错误, 获取每一条 contig 的质量评分, 并将这些 contigs 评分整合后, 得到整个组装结果的分数, 筛选优化转录组获得的序列。利用 CD-HIT 软件^[16]对序列比对聚类, 剔除冗余、相似序列, 最终获得非冗余的序列。用 BUSCO 软件^[17]评价基因组或者转录组的组装完整性。

1.4 基因功能注释

利用 NCBI 将 Unigene 序列比对到 Nr、Pfam、SwissProt 和 egg NOG 数据库中 (evalue < 0.00001), 用 BLAST2 GO 软件对 Unigene 序列进行 GO 注释, 利用 KOBAST 软件对 Unigene 序列进行 KEGG 注释^[18], 并分类统计注释结果。用 TF 数据库对预测的蛋白序列进行 hmmscan 比对并得到转录因子家族。

2 结果与分析

2.1 千针万线草根转录组组装质量分析

对千针万线草根进行转录组测序后, 经过严格的质量控制及数据过滤得到 55400486 个高质量且干净的序列。用 Trinity 软件将这些高质量数据进行 de novo 组装, 共获得 37331463 条有效序列片段, N50 为 1799 bp, 平均长度为 1093.58 bp, GC 核苷酸的含量 (GC%) 为 41.02%, 千针万线草根 Unigene34137 条 (不含 N 的组装片段) 长度分布于 201~14683 bp (表1)。用长度分布柱状图分析组装后的基因如图1所示, 长度为 200~500 bp、501~1000 bp、1001~1500 bp、1501~2000 bp、2001~2500 bp、2501~3000 bp、3001~3500 bp、3501~4000 bp、4001~4500 bp、大于 4500 bp 的 unigene 分别占 40.2%、20.2%、13.3%、10.3%、6.7%、3.8%、2.2%、1.3%、0.8%、1.2%。

表 1 Unigene 基本信息表
Table 1 Basic information on unigenes

基因数量 Genes Num	GC核苷酸含量 GC Percentage/%	N50/bp	最大长度 Max Length/bp	最小长度 Min Length/bp	平均长度 Average Length/bp	组装的总碱基 Total assembled bases
34 137	41.02	1 799	14 683	201	1 093.58	37 331 463

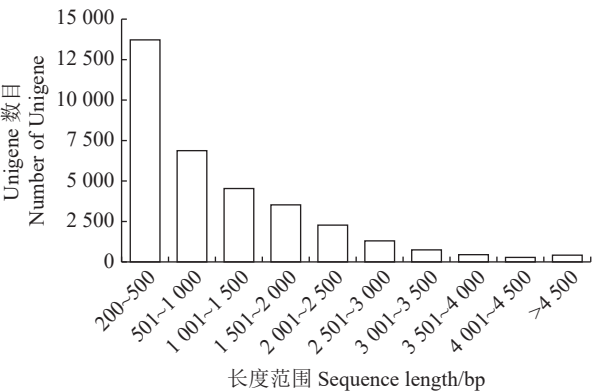


图 1 千针万线草根转录组 Unigenes 的长度分布

Fig. 1 Distribution of length of unigenes from *S. yunnanensis* roots

2.2 千针万线草根基因的总注释情况

将组装得到的基因在 Nr、Swiss-Prot、Pfam、egg NOG、GO、KEGG 数据库中逐一进行比对，所得到注释的基因条数分别为 21 510、17 208、17 702、19 414、17 191、10 285 条。至少有 22 369 条 unigenes 在一个数据库中进行了匹配和功能注释，有 11 768 条 unigenes 未比对到数据库，这表明功能未知的基因还较多。相比之下，注释到 Nr 数据库的基因数最多有 21 510 条，占比为 63.01%；注释到 KEGG 数据库的基因最少，共有 10 285 条，占比为 30.13%（表 2）。

2.3 千针万线草根基因的 NR 功能注释

共有 21 510 条 unigenes 比对到 NR 数据库，与其他物种比对后的结果如图 2 所示。注释序列物种中注释

到基因数量最多的是甜菜（*Beta vulgaris*），共有 6 473 条，占比为 30.09%；其余依次为藜麦（*Chenopodium quinoa*）、菠菜（*Spinacia oleracea*），分别有 6 124、3 954 条 unigenes，占比分别为 28.47%、18.38%，这 3 种植物占比高达 76.94%；其他 297 种物共有 4 959 条 unigenes，占比 23.06%。

2.4 egg NOG 功能注释

将千针万线草根基因比对到 egg NOG 数据库中，得到注释到的 Unigenes 有 19 980 条，占总 Unigenes 的 58.53%，被分为 23 类（图 3）。有 11 394 个未知的功能注释信息，其准确的生物学功能未确定，占所有功能注释信息的 57.03%；翻译后修饰、蛋白开关和分子伴侣注释到 1 264 条 unigenes，所占比例为 6.33%；转录注释到 1 144 条 unigenes，所占比例为 5.73%；信号传导机制注释到 895 条 unigenes，占 4.48%；胞内运输、分泌和囊泡运输注释到 702 条 unigenes，所占比例为 3.51%；碳水化合物运输和代谢注释到 632 条 unigenes，所占比例为 3.16%；翻译、核糖体结构与生物发生注释到 615 条 unigenes，所占比例为 3.08%；最少的是细胞运动和细胞核结构的 unigenes，分别仅有 6 条（0.03%）和 4 条（0.02%）。这些结果表明，千针万线草在翻译后修饰、蛋白开关和分子伴侣、信号转导机制、胞内运输、分泌和囊泡运输、碳水化合物运输和代谢、翻译、核糖体结构与生物发生等基因表达丰度较高。

2.5 GO 功能注释

注释到 GO 数据库中的 unigenes 一共有 17 191 条，获得 69 356 个功能注释，共分为生物过程、细胞组分和分子功能三类，依次得到 21 481、27 214 和 20 661 个功能注释，分别占比为 30.97%、39.24%、29.79%。3 个本体细分为 53 个功能亚类，分子功能类包括 16 个功能亚类，结合（8 993 个注释）所占比例最多，其次是催化活性（8 393 个注释）；生物学过程类占其中 23 个功能亚类，细胞进程（6 958 个注释）和代谢过程（5 972 个注释）占比较大，其次是生物调控（2 581 个注释）；细胞组分类包括 14 个亚类，“细胞”所占比例最多（8 164 个注释），其次是“膜”（5 897 个注释）和“细胞器”（4 792 个注释）。

表 2 Unigene 注释统计
Table 2 Annotation of unigenes

注释数据库 Annotation database	Unigene数目 Number of Unigene	比例 Percentage /%
Nr	21 510	63.01
Swiss-prot	17 208	50.41
Pfam	17 702	51.86
egg NOG	19 414	56.87
GO	17 191	50.36
KEGG	10 285	30.13
注释到的基因数量Annotation genes	22 369	65.53
未注释到的基因数量Without annotation gene	11 768	34.47

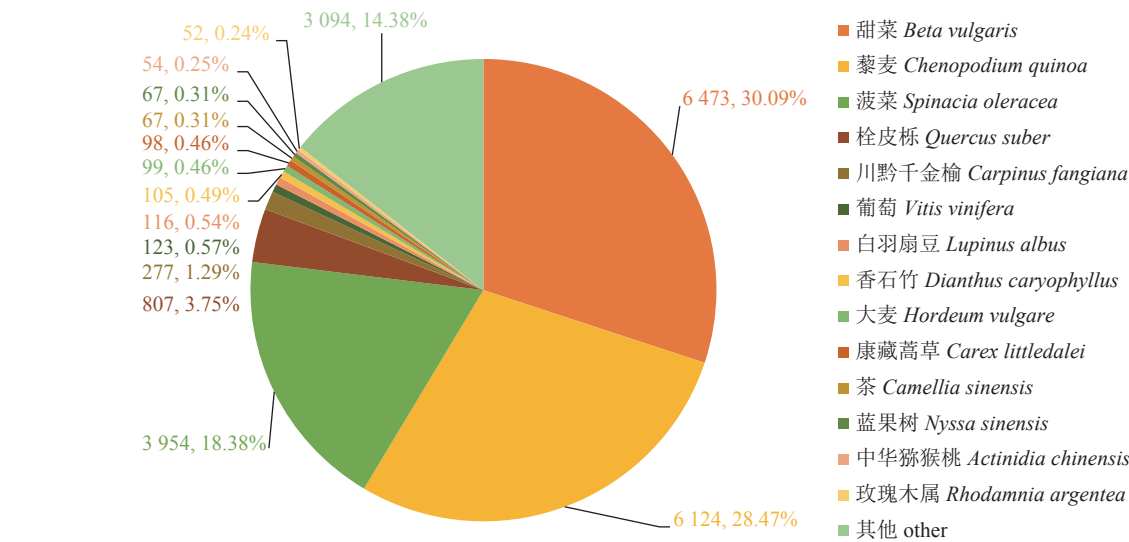


图 2 千针万线草根基因比对物种分类

Fig. 2 Species classification of unigenes



A-RNA 加工与修饰; B-染色质结构与动力; C-能量生产与转化; D-细胞周期调控、细胞分裂、染色体分裂; E-氨基酸转运与代谢; F-核苷酸转运与代谢; G-糖转运与代谢; H-辅酶转运与代谢; I-脂类转运与代谢; J-翻译、核糖体结构和生物合成; K-转录; L-复制、重组和修复; M-细胞壁、膜、包体生物合成; N-细胞运动性; O-蛋白质翻译后修饰与周转; P-无机离子转运与代谢; Q-次生代谢产物的生物合成; S-未知功能; T-信号传导机制; U-胞内运输、分泌和囊泡运输; V-防御机制; Y-核酸结构; Z-细胞骨架。

A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; Y: Nuclear structure; Z: Cytoskeleton.

图 3 千针万线草 unigene 的 egg NOG 分类

Fig. 3 Egg NOG function classification of *S. yunnanensis*

只有少数转录本被注释参与细胞杀伤、毒素活性、碳利用、介导分子活动、细胞增殖及发信号。这些

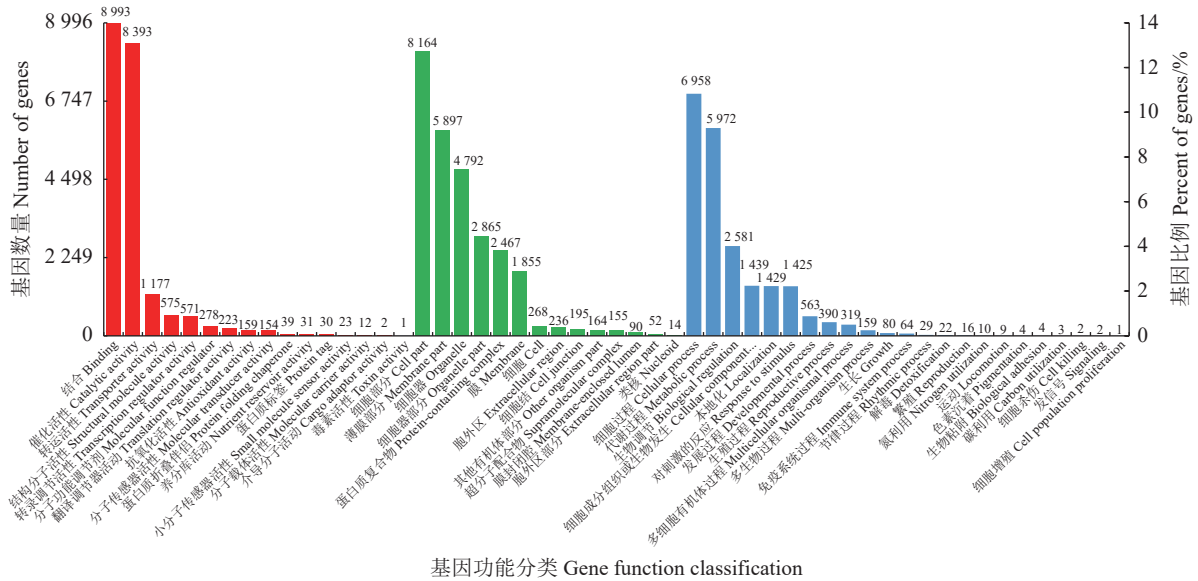
结果表明，以结合、催化活性、细胞、细胞过程、代谢过程相关的基因较多（图 4）。

2.6 KEGG 功能注释

在 KEGG 数据库共注释到 6505 个 unigenes，可归为 6 类，与代谢有关的通路共 100 条、注释到的 unigenes 有 3190 条，所占比例最高（49.04%）；与遗传信息处理有关的通路有 21 条，注释到的 unigenes 有 2115 条（占 32.51%）；环境信息处理相关的通路有 4 条，注释到的 unigenes 有 409 条（占 6.29%）；细胞过程相关的通路有 6 条，注释到的 unigenes 有 501 条（占 7.70%）；生物体系统相关的通路有 2 条，注释到的 unigenes 有 259 条（占 3.98%）；而人类疾病相关的通路有 2 条，注释到的 unigenes 有 31 条（占 0.48%）。代谢相关的通路细致分成 10 个亚类，其中排名前三的是碳水化合物代谢、氨基酸代谢、脂类物质代谢，分别占比为 11.79%、7.29%、6.41%，遗传信息加工分为 4 个亚类，环境信息处理和人类疾病各分为 2 个亚类，细胞过程和生物体系统的通路各分为 1 个亚类。其中“翻译”“折叠、分类和降解”“运输和代谢”“能量代谢”以及“信号转导”分别占 14.05%、9.99%、7.70%、5.70%、5.21%，其余 12 个亚类所占比例均小于 5%（图 5）。

2.7 千针万线草根中黄酮类物质生物合成相关基因挖掘

基于 KEGG 信号通路富集分析结果如表 3 所示，在千针万线草根的转录组数据中共筛选获得 80 条与黄酮类生物代谢相关的基因。其中 18 条 unigene 映射到了苯丙烷代谢通路（ko00940），45 条 unigene 映射到类黄酮生物合成代谢通路上（ko00941），10 条 unigene 映射到黄酮醇生物合成代谢通路



基因功能分类 Gene function classification

图 4 千针万线草根 GO 注释结果

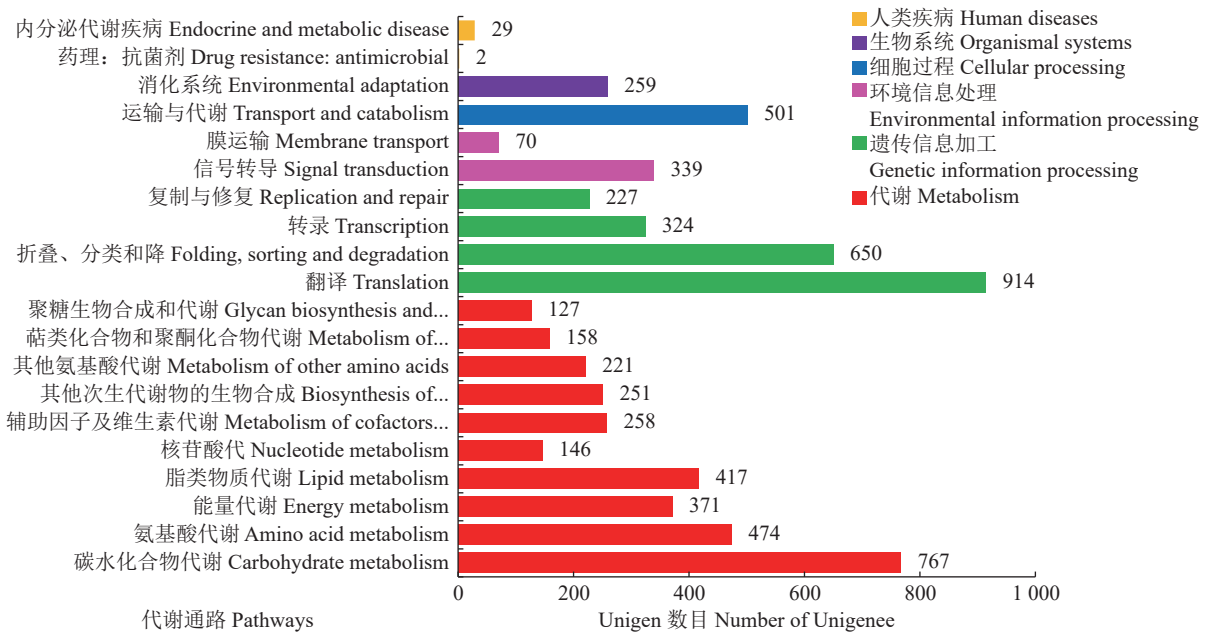
Fig. 4 Gene ontology (GO) categorization of *S. yunnanensis* roots

图 5 千针万线草根 unigenes 的 KEGG 注释结果及分类

Fig. 5 KEGG classification and metabolism pathways on annotated unigenes of *S. yunnanensis* roots

(ko00944), 7 条 unigene 映射到异黄酮生物合成代谢通路上 (ko00943)。

在苯丙烷代谢通路中, 18 条 unigene 分别编码苯丙烷途径中的 3 个关键酶: PAL、C4H、4CL。其中 5 条 unigene 编码该途径中第 1 个催化酶 PAL, 5 条 unigene 编码第 2 个催化反应酶 C4H, 该酶属于细胞色素 P450 (CY450) 家族。8 条 unigene 编码该途径的第 3 个酶 4CL, 其控制苯丙烷类代谢走向不同的代谢途径, 是一个重要分支点。在类黄酮生物合成代谢通路中, 45 条 unigene 共编码 10 种该通路下的关键酶, 包括查尔酮合成酶 (CHS)、查耳酮异构酶

(CHI)、黄烷酮-3-羟化酶 (F3H)、肉桂酸-4-单加氧酶 (CYP73A)、类黄酮-3'-单加氧酶 (CYP75B1)、根皮苷合酶 (PGT1)、花青素还原酶 (ANR)、莽草酸羟基肉桂转移酶 (HCT)、香豆酰脂 3'羟化酶 (C3'H)、咖啡酰辅酶 A-O 甲基转移酶 (CCOAMT)。其中, CHS 是催化黄酮类生物合成的第一个限速步骤, 是植物聚酮合成酶超家族的成员之一, 是植物类黄酮合成途径中的第一个关键结构。CHI 制约合成黄酮醇的含量。F3H 是黄烷酮分支点的一个核心酶。在黄酮醇生物合成代谢通路中, 分别编码该通路下的 2 种关键酶: 类黄酮-3'-单加氧酶 (CYP75B1) 和黄酮

表 3 千针万线草根转录组中黄酮类化合物合成相关基因
Table 3 Flavonoid biosynthesis-related gene in transcriptome of *S. yunnanensis* roots

代谢途径 Metabolic pathway	基因名称 Gene	KO编号 KO ID	EC编号 EC No.	Unigene数量 Number of unigene
苯丙烷代谢通路 Phenylpropane metabolic pathway	苯丙氨酸解氨酶基因 (<i>PAL</i>)	K10775	4.3.1.24	5
	肉桂酸-4-单加氧酶基因 (<i>CYP73A</i>)	K00487	1.14.14.91	5
	CoA连接酶基因 (<i>4CL</i>)	K01904	6.2.1.12	8
类黄酮代谢通路 Flavonoid metabolic pathway	花青素还原酶基因 (<i>ANR</i>)	K08695	1.3.1.77	1
	香豆酰脂3'-羟化酶基因 (<i>C3'H</i>)	K09754	1.11.13.36	2
	莽草酸羟基肉桂转移酶基因 (<i>HCT</i>)	K13065	2.3.1.133	12
	查耳酮异构酶基因 (<i>CHI</i>)	K01859	5.5.1.6	3
	肉桂酸-4-单加氧酶基因 (<i>CYP73A</i>)	K00487	1.14.14.91	5
	根皮苷合酶基因 (<i>PGT1</i>)	K22845	2.4.1.357	1
	查尔酮合成酶基因 (<i>CHS</i>)	K00660	2.3.1.74	12
	咖啡酰辅酶A-O甲基转移酶基因 (<i>CCOAMT</i>)	K00588	2.1.1.104	3
	类黄酮-3'-单加氧酶基因 (<i>CYP75B1</i>)	K05280	1.14.14.82	4
	黄烷醇-3-羟化酶基因 (<i>F3H</i>)	K00475	1.14.11.9	2
黄酮醇代谢通路 Flavonol metabolic pathway	类黄酮-3'-单加氧酶基因 (<i>CYP75B1</i>)	K05280	1.14.14.82	4
	黄烷醇-3-O-葡萄糖苷葡萄糖基转移酶基因 (<i>FG3</i>)	K22794	2.4.1.239	6
异黄酮代谢通路 Isoflavone metabolic pathway	异黄酮2'-羟化酶基因 (<i>CYP81E</i>)	K13260	1.14.14.90	7

醇-3-O-葡萄糖苷葡萄糖基转移酶 (FG3)。在异黄酮生物合成代谢通路中，7 条 unigene 均编码异黄酮 2'-羟化酶 (CYP81E)。

2.8 转录组 Unigene 的转录因子预测

千针万线草根转录组共预测到 724 个转录因子，可分为 33 个转录因子大家族。其中，MYB_superfamily (116 个, 16.02%) 属于最大家族，其次是ERF (59 个, 8.15%)，接着为bHLH (57 个, 7.87%)、C2C2 (53 个, 7.32%)、WRKY (52 个, 7.18%)、bZIP (47 个, 6.49%)、NAC (39 个, 5.39%)、C2H2 (35 个, 4.83%)、B3_superfamily (31 个, 4.28%)、GRAS (31 个, 4.28%)、LBD (30 个, 4.14%)、C3H (29 个, 4.01%)，其他 21 个家族共有基因 145 个，占 20.03% (图 6)。

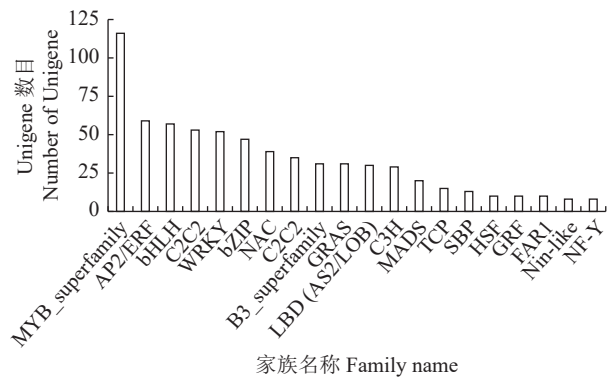


图 6 千针万线草根 Unigene 的转录因子预测
Fig.6 Predicted transcription factors of unigenes of *S. yunnanensis* roots

3 讨论

开展药用植物转录组研究是综合诠释植物基因组信息、发掘植物次生代谢生物合成途径中关键酶以及植物次生代谢成分生物合成途径与调控的新途径^[19]。本研究首次对药用植物千针万线草根进行转录组测序，测序产生 clean reads 的 Q20 值为 98.42%，Q30 值为95% 以上，说明测序质量合格。经denove 拼接组装，去除冗余后，共得到 34 137 条 Unigene，N50 为 1799 bp，平均长度为 1 093.58 bp。组装结果与杭白芷 (*Angelica dahurica*) 根 (N50 为 1 703 bp，平均长度 1 164 bp) 较为接近^[20]。N50 是评价组装序列完整性的重要指标，本研究中 N50>800 bp，表明组装片段完整，所得 Unigenes 数量基本涵盖全部转录信息，序列信息量庞大，可以满足转录组分析的要求。

甜菜 (*Betavulgaris*)、藜麦 (*Chenopodiumquinoa*)、菠菜 (*Spinacia oleracea*) 同属藜科，在 Nr 数据库注释中，注释到基因占比高达 76.94%，可能藜科与石竹科在亲缘关系上比较接近。而石竹科蝇子草属的滇白前注释到藜科这三种植物基因占比达 64.06%^[21]，千针万线草相比于滇白前，在与藜科亲缘关系上可能较近。在 GO 功能注释方面，可以分为 3 个大类和 53 个亚类，其中结合、催化活性、细胞过程及代谢过程的 Unigene 数目较多。通过 egg NOG 功能分类可知，参与翻译后修饰、蛋白开关和分子伴侣、转录、信号传导机制的 Unigene 较多，但有 11 394 条

未知功能基因, 占 57.03%。这是由于 egg NOG 更新比较快, 收录来源很广泛, 数据库收录的未知序列占比就高, 所以注释的结果未知序列占比较高。转录因子是能从 DNA 上读取并解释遗传“蓝图”的蛋白质组之一, 它和 DNA 相结合, 有助于启动调节基因转录的程序, 对植物应对非生物逆境胁迫具有重要影响^[22]。研究表明, MYB、AP2/ERF、bHLH、bZIP、NAC、WRKY 和 C2H2 转录因子对植物应对干旱、高盐和寒冷胁迫均有显著影响^[21]。对这七类转录因子进行统计, 千针万线草根中有 405 个 (55.94%), 可能与其抗逆适应相关。在野外观察发现, 千针万线草在向阳干旱贫瘠的陡坡也能正常生长, 这是其抗逆性强的一个方面, 是否能适应其他逆境, 还有待进一步研究。

通过 KEGG 数据库和通路分析发现, 涉及代谢途径的 Unigene (3190 个) 最多, 通路数量最多的也是涉及代谢途径的, 共 100 条。从本研究获得的基因注释结果来看, 有 5 个生物合成代谢途径 80 个 unigene 涉及到药用成分的合成, 其中包括黄酮、黄酮醇、异黄酮等物质。通过进一步筛选, 这些基因分别编码 16 种关键酶, 包括苯丙烷代谢途径所需的 3 种酶 (PAL、C4H、4CL), 类黄酮代谢所需的 10 种关键酶 (CHS、CHI、F3H、CYP73A、CYP75B1、PGT1、ANR、HCT、C3'H、CCOAMT) 和黄酮醇代谢途径所需的 2 种关键酶 (CYP75B1、FG3) 和异黄酮代谢途径关键酶 CYP81E。千针万线草根的药用成分主要是黄酮类物质, 通过对千针万线草根转录组测序、注释、分析以及黄酮类化合物合成相关基因的挖掘, 丰富了千针万线草根的遗传信息, 可以为进一步研究该物种药用成分合成的关键基因及调控机制、克隆关键基因及功能分析奠定基础, 也为千针万线草品质的形成提供理论依据。

参考文献:

- [1] 中国植物志编委会. 中国植物志(第26卷) [M]. 北京: 科学出版社, 1996: 133.
- [2] 马利华, 陈斌, 谢东浩, 等. 繁缕属植物抗病毒活性物质基础组分结构研究 [J]. 中草药, 2012, 43 (4): 799-805.
MA L H, CHEN B, XIE D H, et al. Study on antiviral material basis and component structure in plants of *Stellaria* L [J]. *Chinese Traditional and Herbal Drugs*, 2012, 43 (4): 799-805. (in Chinese)
- [3] 谭瑞璞, 刘海周, 邓义德, 等. 千针万线草中总黄酮提取工艺的研究 [J]. 云南中医学院学报, 2009, 32 (1): 9-11.
TAN R P, LIU H Z, DENG Y D, et al. Study on extraction process of total flavonoids in *Stellaria yunnanensis* franch [J]. *Journal of Yunnan University of Traditional Chinese Medicine*, 2009, 32 (1): 9-11. (in Chinese)
- [4] 谭瑞璞, 刘海周, 梁晓原. HPLC法测定千针万线草中牡荆素的含量 [J]. 云南中医学院学报, 2009, 32 (4): 28-29,32.
TAN R P, LIU H Z, LIANG X Y. Determination of vitexin in *Stellaria yunnanensis* by HPLC [J]. *Journal of Yunnan University of Traditional Chinese Medicine*, 2009, 32 (4): 28-29,32. (in Chinese)
- [5] 刘伟, 王俊磊, 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.
LIU W, WANG J Y, LI M, et al. Transcriptome sequencing analysis of gene expression of flavonoid biosynthesis in *Ginkgo biloba* [J]. *Chinese Traditional and Herbal Drugs*, 2018, 49 (23): 5633-5639. (in Chinese)
- [6] 邹丽秋, 王彩霞, 匡雪君, 等. 黄酮类化合物合成途径及合成生物学研究进展 [J]. 中国中药杂志, 2016, 41 (22): 4124-4128.
ZOU L Q, WANG C X, KUANG X J, et al. Advance in flavonoids biosynthetic pathway and synthetic biology [J]. *China Journal of Chinese Materia Medica*, 2016, 41 (22): 4124-4128. (in Chinese)
- [7] LI H Y, DONG Y Y, YANG J, et al. *De novo* transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids [J]. *PLoS One*, 2012, 7 (2): e30987.
- [8] 邓楠, 史胜青, 常二梅, 等. 基于中麻黄萌发种子转录组的黄酮类化合物合成途径基因的挖掘 [J]. 林业科学研究, 2014, 27 (6): 758-763.
DENG N, SHI S Q, CHANG E M, et al. Exploring flavonoid biosynthetic pathway genes based on transcriptome of *Ephedra intermedia* germinating seeds [J]. *Forest Research*, 2014, 27 (6): 758-763. (in Chinese)
- [9] 邹毅辉, 刘昌勇, 林泽燕, 等. 半枝莲转录组高通量测序及黄酮类相关基因解析 [J]. 福建农业学报, 2018, 33 (12): 1242-1250.
ZOU Y H, LIU C Y, LIN Z Y, et al. Transcriptome sequencing and flavonoid biosynthesis-related genes of *Scutellaria barbata* D. Don [J]. *Fujian Journal of Agricultural Sciences*, 2018, 33 (12): 1242-1250. (in Chinese)
- [10] 邹福贤, 许文, 黄泽豪, 等. 金线莲转录组测序及其黄酮类合成相关基因分析 [J]. 中国药科大学学报, 2019, 50 (1): 66-74.
ZOU F X, XU W, HUANG Z H, et al. Analysis of transcriptome sequencing and related genes of flavonoid biosynthesis from *Anoectochilus roxburghii* [J]. *Journal of China Pharmaceutical University*, 2019, 50 (1): 66-74. (in Chinese)
- [11] 林江波, 王伟英, 邹晖, 等. 基于转录组测序的铁皮石斛黄酮代谢途径及相关基因解析 [J]. 福建农业学报, 2019, 34 (9): 1019-1025.
LIN J B, WANG W Y, ZOU H, et al. Transcriptome analysis on pathway of and genes related to flavonoid synthesis in *Dendrobium officinale* [J]. *Fujian Journal of Agricultural Sciences*, 2019, 34 (9): 1019-1025. (in Chinese)
- [12] 许明, 杨志坚, 黄学敏, 等. 藤茶高通量转录组分析及黄酮类化合物合成相关基因挖掘 [J]. 南方农业学报, 2020, 51 (8): 1797-1805.
XU M, YANG Z J, HUANG X M, et al. Transcriptome analysis of *Ampelopsis grossedentata*(Hand. Mazz.)W. T. Wang and mining of

- putative genes involved in flavonoid biosynthesis [J]. *Journal of Southern Agriculture*, 2020, 51 (8) : 1797–1805. (in Chinese)
- [13] 袁茂, 江明锋, 徐亚欧, 等. 藏鸡不同发育阶段腿部肌肉组织转录组及microRNA联合分析 [J]. 畜牧兽医学报, 2019, 50 (12): 2400–2412.
- YUAN M, JIANG M F, XU Y O, et al. Analysis of transcriptome and microRNA in leg muscle of Tibetan chicken at different developmental stages [J]. *Acta Veterinaria et Zootechnica Sinica*, 2019, 50 (12) : 2400–2412. (in Chinese)
- [14] GRABHERR M G, HAAS B J, YASSOUR M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nature Biotechnology*, 2011, 29 (7) : 644–652.
- [15] SMITH-UNNA R, BOURSNELL C, PATRO R, et al. TransRate: Reference-free quality assessment of *de novo* transcriptome assemblies [J]. *Genome Research*, 2016, 26 (8) : 1134–1144.
- [16] LI W Z, GODZIK A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences [J]. *Bioinformatics*, 2006, 22 (13) : 1658–1659.
- [17] SIMÃO F A, WATERHOUSE R M, IOANNIDIS P, et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs [J]. *Bioinformatics*, 2015, 31 (19) : 3210–3212.
- [18] 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.
- [19] 朱昀昊, 董诚明, 郑晓珂, 等. 基于转录组测序的山茱萸次生代谢生物合成相关基因的挖掘 [J]. 中国中药杂志, 2017, 42 (2): 213–219.
- ZHU Y H, DONG C M, ZHENG X K, et al. Transcriptome analysis reveals genes involved in biosynthesis of secondary metabolism in *Cornus officinalis* [J]. *China Journal of Chinese Materia Medica*, 2017, 42 (2) : 213–219. (in Chinese)
- [20] 吴萍, 郭俊霞, 王晓宇, 等. 基于高通量测序技术的杭白芷(*Angelica dahurica*)根转录组组数据分析 [J]. 分子植物育种, 2020, 18 (10): 3207–3216.
- WU P, GUO J X, WANG X Y, et al. High-throughput transcriptome sequencing of roots of *Angelica dahurica* and data analyses [J]. *Molecular Plant Breeding*, 2020, 18 (10) : 3207–3216. (in Chinese)
- [21] 王建秋, 王晓丽, 曹子林. 滇白前种子转录组测序及功能注释 [J]. 种子, 2021, 40 (3): 52–59.
- WANG J Q, WANG X L, CAO Z L. Transcriptome sequencing and functional annotation of *Silene viscidula* seeds [J]. *Seed*, 2021, 40 (3) : 52–59. (in Chinese)
- [22] WANG H Y, WANG H L, SHAO H B, et al. Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology [J]. *Frontiers in Plant Science*, 2016, 7: 67.

(责任编辑：于洪杰)