

杨晚铃,白芷懿,邹明珠,等:鹿茸草全长转录组测序与次生代谢产物生物合成相关基因的挖掘[J].江西师范大学学报(自然科学版) 2023 47(1):99-110.

YANG Wanling, BAI Zhiyi, ZOU Mingzhu, et al. The full-length transcriptome sequencing and identification of related genes involved in secondary metabolism biosynthesis in *Monochasma savatieri* [J]. Journal of Jiangxi Normal University( Natural Science) 2023 47(1): 99-110.

文章编号:1000-5862(2023)01-0099-12

## 鹿茸草全长转录组测序与次生代谢产物 生物合成相关基因的挖掘

杨晚铃,白芷懿,邹明珠,王欣茹,谢建坤,张帆涛\*

(江西师范大学生命科学学院,江西 南昌 330022)

**摘要:**为获得鹿茸草的全长转录组信息,挖掘鹿茸草次生代谢化合物生物合成途径相关酶的基因,该文基于单分子测序技术,利用 Pacbio 高通量测序平台,对鹿茸草进行全长转录组测序,共获得 48 005 条去冗余的高质量转录本,与 NR、Swiss-Prot、GO、KEGG 等 8 个数据库进行 BLAST 比对,共有 45 362 个转录本被成功注释,注释率为 94.50%。其中有 389 条转录本被注释到 KEGG 的 10 条标准次生代谢生物合成通路中。对转录组数据进一步分析发现:参与鹿茸草苯丙素类生物合成的转录本有 194 条,参与生物碱类生物合成的转录本有 115 条,参与类黄酮化合物生物合成的转录本有 23 条,参与其他次生代谢产物的转录本有 57 条,参与次生代谢后氧化与糖基化修饰的转录本有 204 条。鹿茸草全长转录组的获得极大地丰富了鹿茸草的遗传信息,初步揭示了参与鹿茸草次生代谢产物合成相关的基因通路,为深入研究鹿茸草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:**鹿茸草;全长转录组;次生代谢;生物合成;遗传资源

中图分类号:S 567 文献标志码:A DOI: 10.16357/j.cnki.issn1000-5862.2023.01.13

### 0 引言

鹿茸草(*Monochasma savatieri* Franch. ex Maxim.)为玄参科鹿茸草属多年生根半寄生植物,别称有沙氏鹿茸草、白花鹿茸草、千重塔等,常生于阳光充足的多沙山坡及灌丛中,主要分布在中国的东南地区,如江苏省、安徽省、浙江省、福建省、江西省、湖南省等,在日本九州也有分布<sup>[1-2]</sup>。鹿茸草的根、茎、叶皆可入药,其性质平和,味略微苦涩,主要功效是凉血止血、泻火解毒和祛风除湿止痛,可用于治疗咳嗽、发热、吐血便血和风湿关节疼痛等多种病症<sup>[3]</sup>。鹿茸草的化学成分分析显示:鹿茸草含有苯丙素类、黄酮类、生物碱类、酚酸类以及多糖类等<sup>[4-5]</sup>多种成

分,这使得鹿茸草具有十分广泛的药理活性。现代药理学研究发现鹿茸草具有抗氧化<sup>[6]</sup>、抑菌<sup>[7]</sup>、抗炎<sup>[8]</sup>和抗病毒<sup>[9]</sup>等生物学活性,是生产国家中药保护品种炎宁颗粒的主药材之一<sup>[10]</sup>。目前,关于鹿茸草的化学成分和药理活性的研究较多,但关于其基因组水平的研究(如基因功能的分析、次生代谢产物生物合成相关基因的发掘和重要活性成分合成机制的阐明等)鲜见报道。

鹿茸草至今还没有公布的基因组序列。近年来,高通量测序技术已十分成熟,可以在缺少基因组信息的情况下对其进行转录组测序,获得转录组信息<sup>[11]</sup>。第 2 代测序技术虽然为植物研究积累了丰富的基因组数据,但是也存在一些技术缺陷,如读长较短、容易产生 PCR 扩增错误等,较难得到准确、全面

收稿日期:2022-03-16

基金项目:国家自然科学基金(31960370)和江西省自然科学基金(20202ACB205002)资助项目。

通信作者:张帆涛(1984—),男,江西临川人,副教授,博士,主要从事植物生物技术与分子生物学研究。E-mail: zhang84004@

163.com

的基因组测序信息,对深入挖掘和分析转录组数据有一定影响<sup>[12-14]</sup>。因此,第3代测序技术应运而生,其特点是超长读长,能读取包含单条完整转录本的序列信息,不需要模板扩增,且后期分析不需要组装就可得到高质量的全长转录本,可为许多无参考基因组的物种提供更全面的基因序列信息<sup>[15-16]</sup>。目前,第3代测序技术在植物(如薏苡<sup>[17]</sup>、艾蒿<sup>[18]</sup>、厚朴<sup>[19]</sup>、蒺藜苜蓿<sup>[20]</sup>、毛白菜<sup>[21]</sup>和当归<sup>[22]</sup>等)基因组的研究中已得到广泛的应用。本文利用 PacBio 第3代转录组测序技术对鹿茸草进行全长转录组测序,得到了大量的高质量转录本序列,分析了鹿茸草的基因表达特征与功能信息,挖掘了鹿茸草次生代谢合成的相关基因,为进一步阐明及调控鹿茸草次生代谢合成途径奠定了基础。

## 1 材料与方法

### 1.1 材料

鹿茸草于2021年6月采自江西省宜春市袁州区渥江镇(27°44'N,114°44'E),经宜春市原农科所所长高海军研究员鉴定。取新鲜鹿茸草的叶片和茎,在清洗干净后,吸干水分,置于液氮中速冻后于-80℃冰箱中保存备用。

### 1.2 RNA 提取、转录组文库构建及测序

委托北京百迈客生物科技有限公司(Biomarker Technologies, Beijing)进行鹿茸草总RNA提取、全长转录组文库构建及测序。在总RNA提取完成后,要对RNA样品的浓度和完整性等进行检测,取检验合格的高质量RNA样品用于全长转录组文库的构建。以下为文库构建的简要流程:首先合成mRNA的全长cDNA<sup>[23]</sup>,将得到的全长cDNA进行PCR扩增使其放大;然后修复全长cDNA末端并连上SMRT接头,消化核酸外切酶;最后得到测序文库。取检验合格的文库,利用PacBio SMRT测序平台进行全长转录组测序。

### 1.3 全长转录组数据获取

获取全长转录组序列的分析过程主要有3个阶段,它们分别是circular consensus sequencing(CCS)序列的生成并校正、全长序列的识别和isoform水平聚类得到一致序列<sup>[24]</sup>。首先,将全部的原始序列对应其接头转换成CCS序列,同时对CCS序列进行校正,获得序列的质量信息;然后,通过检测在CCS序

列中是否包含正确的5'引物、3'引物及polyA尾,将序列分成全长序列(包含5'引物、3'引物及polyA尾)和非全长序列;最后,对在全长序列中相似的序列(即同一转录本的多个拷贝)聚成一簇(cluster),每簇得到1条一致序列(consensus isoform),去除冗余的高质量序列用于后续分析。

### 1.4 序列信息注释

为了分析转录本的不同功能,使用BLAST<sup>[25]</sup>软件将获得的高质量非冗余转录本序列与NR(NCBI非冗余蛋白质数据库)<sup>[26]</sup>、Swiss-prot(蛋白质序列数据库)<sup>[27]</sup>、GO(基因本体论)<sup>[28]</sup>、COG(蛋白质直系同源簇)<sup>[29]</sup>、KOG(真核生物蛋白相邻类的聚簇)<sup>[30]</sup>、eggNOG(直系同源蛋白分组比对)<sup>[31]</sup>、Pfam(蛋白质家族)<sup>[32]</sup>、KEGG(京都基因和基因组百科全书)<sup>[33]</sup>数据库比对,从而获得转录本的注释信息。次生代谢产物生物合成通路按照在KEGG数据库中标准的次生代谢通路进行筛选。

### 1.5 CDS 及转录因子预测

使用TransDecoder<sup>[34]</sup>软件基于开放阅读框(open reading frame, ORF)的长度、对数似然函数值(log-likelihood score)、氨基酸序列与在Pfam数据库中蛋白质结构域序列的比对等信息,从转录本序列中识别可靠的潜在编码区序列(coding sequence, CDS),从而得到鹿茸草的CDS序列。转录因子的预测是使用植物转录因子预测软件iTAK<sup>[35]</sup>对植物TF、TR和PKS进行识别和分类,获得转录因子的类型和数量。

### 1.6 LncRNA 预测及靶基因预测

LncRNA不编码蛋白可通过CPC(coding potential calculator)<sup>[36]</sup>分析、CNCI(coding-non-coding index)<sup>[37]</sup>分析、pfam蛋白结构域分析、CPAT(coding potential assessment tool)<sup>[38]</sup>分析对转录本进行编码潜能筛选,判断其是否具有编码潜能,过滤掉有编码潜能的转录本,取以上4种分析软件鉴定得到的非编码转录本的交集作为最终预测的LncRNA结果。LncRNA可与mRNA进行碱基互补配对,利用LncTar<sup>[39]</sup>工具对得到的LncRNA进行靶基因预测。

## 2 结果与分析

### 2.1 全长转录组测序及序列分类校正

根据full passes $\geq 3$ 且序列准确性大于0.9的条

件,从原始序列中共提取获得了 173 830 条 CCS 序列,总碱基数为 386 701 738 bp,CCS 序列平均长度为 2 224 bp,平均测序深度( passes) 为 57. 较高的测序深度为后续获取高质量转录本序列提供了保障. 去除在 CCS 序列中的 cDNA 引物序列及 polyA 序列共获得 141 331 条全长非嵌合( full-length non-chimeric,FLNC) 序列,占 CCS 序列的 81. 30%. 使用 SMRTLink 软件的 IsoSeq 模块将在 FLNC 序列中相似的序列进行聚类分析,得到一致序列( consensus isoform) 72 245 条,筛选准确度高于 99% 的一致序列,共得到 72 231 条( 99. 98%) 高质量转录本. 去除在高质量转录本中的冗余部分,得到非冗余转录本序列 48 005 条,将其用于后续的各项分析.

2.2 转录本功能注释

2.2.1 序列比对及功能注释 将总共 48 005 条高质量的非冗余转录本与 NR、Swiss-prot、GO、COG、eggNOG、KOG、Pfam 和 KEGG 等 8 个数据库进行序列比对和功能注释,成功注释了 45 362 条( 94. 50%) 转录本( 见表 1) .

表 1 转录本功能注释数量统计

数据库	注释数量	占总转录本比例/%
NR 数据库	45 158	94. 07
Swiss-prot 数据库	29 643	61. 75
GO 数据库	32 340	67. 37
COG 数据库	20 790	43. 31
eggNOG 数据库	44 065	91. 80
KOG 数据库	29 653	61. 77
Pfam 数据库	37 712	78. 56
KEGG 数据库	20 952	43. 65
成功注释的转录本	45 362	94. 50
全部转录本	48 005	100. 00

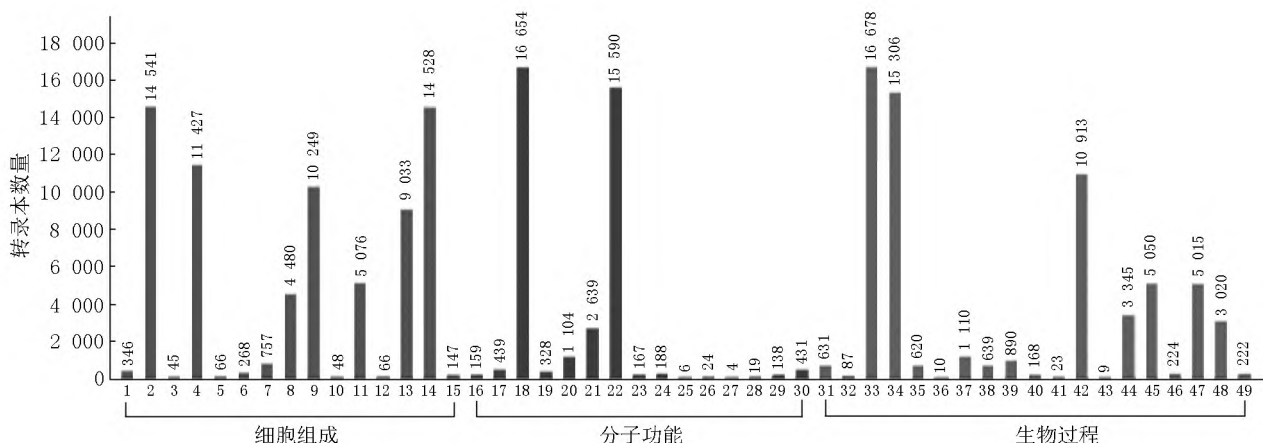
在 NR 数据库中得到注释的转录本最多,有 45 158 条( 占比 94. 07%) ; 其次是 eggNOG、Pfam、GO、KOG、Swiss-prot 和 KEGG 数据库分别注释了 44 065 条( 91. 80%) 、37 712 条( 78. 56%) 、32 340 条( 67. 37%) 、29 653 条( 61. 77%) 、29 643 条( 61. 75%) 和 20 952 条( 43. 65%) ; 而 COG 数据库注释到的转录本数量最少( 20 790 条,43. 31%) . 因为 NR 数据库成功注释的转录本数量最多,所以根据该数据库比对结果进行同源性分析. 结果表明: 在序列匹配度较高的 4 个物种中,芝麻( *Sesamum indicum* Linn.) 所

占的比例最高,为 16 641 条( 36. 85%) ; 其次是紫花风铃木( *Handroanthus impetiginosus* ( Mart. ex DC.) Mat-tos) 为 7 432 条( 16. 46%) 、斑点猴面花( *Erythranthe guttata* ( DC.) G. L. Nesom) 为 3 951 条( 8. 75%) 、木犀榄( *Olea europaea* Linn.) 为 690 条( 1. 53%) ; 其余比例小于 1. 53% 的匹配物种的序列为 16 444 条,占比 36. 41% .

2.2.2 转录本的 GO 分类 将测序获得的转录本在 GO 功能数据库中进行比对,发现有 32 340 条转录本能够比对到相似的基因,且大部分转录本能够对应多个不同基因. 结果表明: 成功比对到的 32 340 条转录本被分到细胞组分、分子功能和生物过程 3 个不同类别的 49 个小组中( 见图 1) .

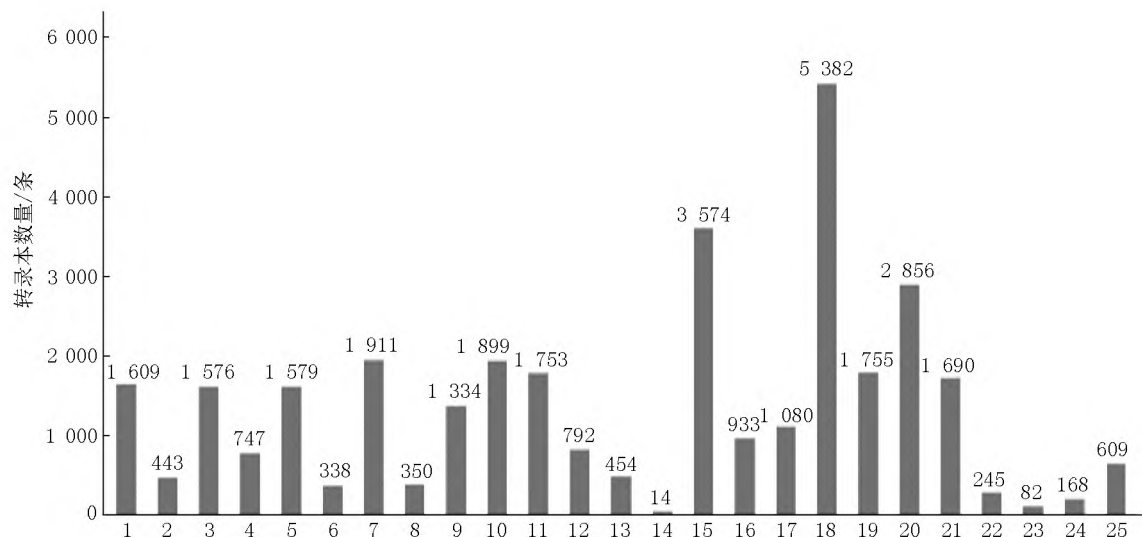
对不同类别的转录本进行统计分析发现: 在与细胞组成类别相关的转录本大部分被归类于细胞( cell) 子集( 14 541 条( 44. 96%) ) 和细胞部分( cell part) 子集( 14 528 条( 44. 92%) ) 中. 在归类到分子功能类别的转录本中,与催化活性( catalytic activity) 和结合( binding) 功能相关的转录本数量最多,分别有 16 654 条( 51. 50%) 和 15 590 条( 48. 21%) ,其他功能的转录本数量较少. 代谢过程( metabolic process) 是在生物过程类别中转录本数量最多的,共有 16 678 条( 51. 57%) . 此外,细胞过程( cell process) 和单一生物过程( single-organism process) 也比对到较多的转录本,分别为 15 306 条( 47. 33%) 和 10 913 条( 33. 74%) .

2.2.3 转录本的 KOG 分类 为了更好地分析鹿茸草转录组的功能,对在 KOG 数据库中被成功注释的 29 653 条转录本进行了分类. 结果表明: 这些转录本一共可分为 25 个不同的功能类群,且涉及的范围较广,涵盖了大部分的生命活动途径( 见图 2) . 其中,一般功能预测类的注释结果最多,有 5 382 条( 18. 15%) ; 在其他类群中,在转录后修饰、蛋白质转运与分子伴侣( 3 574 条,12. 05%) 、信号转导机制( 2 856 条,9. 63%) 、糖类转运与代谢( 1 911 条,6. 44%) 及翻译、核糖体结构和生物发生( 1 899 条,6. 40%) 类群注释的转录本数量也较多,而细胞外结构类群的转录本注释数量最少,仅有 14 条( 0. 05%) . 另外,在次生代谢产物生物合成、运输和分解代谢类群中注释到的转录本数量有 1 080 条,这对研究鹿茸草有效成分的生物合成和运输分解具有重要指导意义.



注: 图中立柱从左至右分别以 1~49 表示. 1 为胞外区; 2 为细胞; 3 为拟核; 4 为细胞膜; 5 为病毒粒子; 6 为细胞连接; 7 为膜封闭腔; 8 为大分子复合物; 9 为细胞器; 10 为胞外区组分; 11 为细胞器组分; 12 为病毒粒子组分; 13 为细胞膜组分; 14 为细胞组分; 15 为超分子复合物; 16 为转录因子活性, 蛋白质结合; 17 为核酸结合转录因子活性; 18 为催化活性; 19 为信号转导活性; 20 为结构分子活性; 21 为转运活性; 22 为结合; 23 为电子载体活性; 24 为抗氧化活性; 25 为金属伴侣活性; 26 为蛋白质标记物; 27 为翻译调节活性; 28 为营养库活性; 29 为分子转导活性; 30 为分子功能调节器; 31 为复制; 32 为免疫系统过程; 33 为代谢过程; 34 为细胞过程; 35 为生殖过程; 36 为生物附着; 37 为信号; 38 为多细胞生物过程; 39 为发育过程; 40 为发育; 41 为移动; 42 为单一生物体过程; 43 为节律过程; 44 为刺激响应; 45 为定位; 46 为多生物体过程; 47 为生物学调控; 48 为细胞组织的组成或合成; 49 为解毒.

图 1 转录本的 GO 分类



注: 图中立柱从左至右分别以 1~25 表示. 1 为 RNA 加工和修饰; 2 为染色质结构和动力学; 3 为能源生产和转换; 4 为细胞周期控制、细胞分裂、染色体分区; 5 为氨基酸转运和代谢; 6 为核苷酸转运和代谢; 7 为糖类转运与代谢; 8 为辅酶转运和代谢; 9 为脂质运输和新陈代谢; 10 为翻译、核糖体结构和生物发生; 11 为转录; 12 为复制、重组和修复; 13 为细胞壁/膜/包膜生物发生; 14 为细胞运动; 15 为转录后修饰、蛋白转运与分子伴侣; 16 为无机离子转运和代谢; 17 为次生代谢物合成、运输和代谢; 18 为一般功能预测; 19 为未知功能; 20 为信号转导机制; 21 为细胞内运输、分泌和囊泡运输; 22 为防御机制; 23 为细胞外结构; 24 为核结构; 25 为碳骨架.

图 2 转录本的 KOG 分类

2.2.4 转录本 KEGG 注释 总共有 20 952 条转录本被成功注释到 KEGG 数据库中,可归类到 128 个 KEGG 标准代谢通路. 其中,转录本注释数量大于 300 的通路共有 19 个,碳代谢通路注释到的转录本

数量最多,有 1 049 条 (5.00%); 其他注释到较多转录本的通路还有氨基酸的生物合成 (799 条, 3.81%)、核糖体 (758 条, 3.62%)、剪接体 (692 条, 3.30%) 和内质网蛋白质加工 (624 条, 2.98%) 等通

路(见表 2)。(0.93%);生物碱类的异喹啉生物碱生物合成与莨菪烷、哌啶和吡啶类生物碱生物合成通路的转录本注释数量次之,分别为 66 条(0.32%)和 49 条(0.24%);类黄酮生物合成的转录本注释数量为 23 条(0.11%),而芥子油苷生物合成代谢通路的转录本注释数量最少,仅有 1 条。

另外,进一步分析 KEGG 注释到的代谢通路,发现有 389 条转录本参与苯丙素类生物合成、异喹啉生物碱生物合成、莨菪烷、哌啶和吡啶类生物碱生物合成及类黄酮生物合成等 10 个次生代谢生物合成标准通路(见表 3)。在这些通路中,转录本注释数量最多的是苯丙素类生物合成代谢通路,有 194 条

表 2 转录本的 KEGG 通路统计分析

编号	通路	通路 ID	转录本数	占比/%
1	碳代谢	ko01200	1 049	5.00
2	氨基酸的生物合成	ko01230	799	3.81
3	核糖体	ko03010	758	3.62
4	剪接体	ko03040	692	3.30
5	内质网蛋白质加工	ko04141	624	2.98
6	RNA 转运	ko03013	571	2.73
7	糖酵解/糖异生	ko00010	476	2.27
8	淀粉和蔗糖代谢	ko00500	441	2.10
9	内吞作用	ko04144	439	2.10
10	丙酮酸代谢	ko00620	432	2.06
11	植物激素信号转导	ko04075	414	1.98
12	mRNA 监测通路	ko03015	400	1.91
13	嘌呤代谢	ko00230	391	1.87
14	氧化磷酸化	ko00190	377	1.80
15	泛素介导的蛋白水解	ko04120	342	1.63
16	过氧化物酶体	ko04146	338	1.61
17	氨基糖和核苷酸糖代谢	ko00520	321	1.53
18	半胱氨酸和蛋氨酸代谢	ko00270	310	1.48
19	乙醛酸和二羧酸代谢	ko00630	306	1.46

表 3 转录本次生代谢的 KEGG 通路注释统计

编号	通路	通路 ID	转录本数	占比/%
1	苯丙素类生物合成	ko00940	194	0.93
2	异喹啉生物碱生物合成	ko00950	66	0.32
3	莨菪烷、哌啶和吡啶类生物碱生物合成	ko00960	49	0.24
4	黄酮类生物合成	ko00941	23	0.11
5	单胺菌素生物合成	ko00261	19	0.09
6	芪类、二芳基庚烷类和姜酚的生物合成	ko00945	12	0.06
7	咖啡因代谢	ko00232	10	0.05
8	甜菜碱生物合成	ko00965	8	0.04
9	黄酮和黄酮醇的生物合成	ko00944	7	0.03
10	芥子油苷生物合成	ko00966	1	0.00

2.3 鹿茸草次生代谢生物合成途径相关酶的鉴定

2.3.1 苯丙素类生物合成相关基因挖掘 在植物次级产物代谢途径中,苯丙素类代谢是其中一条非常重要的途径,可产生多种重要的次生代谢产物<sup>[40]</sup>。在鹿茸草中与苯丙素类生物合成通路相关的

转录本有 194 个,注释到此合成途径中的 15 个酶(见图 3)。在鹿茸草的苯丙素类生物合成通路中有 69 个转录本注释到β-葡萄糖苷酶(β-glucosidase, BglB)、肉桂醇脱氢酶(cinnamyl-alcohol dehydrogenase, CAD)的转录本有 47 个、4-香豆酸辅酶 A 连接

酶( 4-coumarate-CoA ligase ,ACL) 有 25 个、过氧化物酶( peroxidase ,POD) 有 12 个、苯丙氨酸解氨酶( he-nylalanine ammonia-lyase ,PAL) 有 9 个. 此外, 还有松柏醛脱氢酶( coniferyl-aldehyde dehydrogenase , REF1) 、莽草酸 *O*-羟基肉桂酰基转移酶( shikimate *O*-hydroxycinnamoyl transferase ,HCT) 、肉桂酸 4-羟基化酶( cinnamate 4-hydroxylase ,C4H) 、肉桂酰辅酶 A 还原酶( cinnamoyl-CoA reductase ,CCR) 和咖啡酸氧甲基转移酶( caffeic acid *O*-methyltransferase ,COMT) 等多个酶( 见表 4) .

表 4 编码苯丙素类、生物碱类和类黄酮生物合成相关酶的转录本

合成途径	酶	转录本数	转录本编号
苯丙素类	BglB	69	LRC-yc _ transcript _ 35785/37249/58661/61622/8833/16647/34206/36387/37608/41972/53772/5606/58002/6014/61678/6339/6611/66134/6880/7441/18650/19897/20720/22972/23650/24406/26651/28806/2977/30549/33339/33776/34213/35098/35217/35262/37853/40519/40867/41077/4162/42154/42892/43501/43975/45068/49789/51075/51705/52408/61076/62108/62689/64339/64434/64664/6523/65838/65865/66365/6884/69168/69864/70080/7116/72067/9201/9207/9399
			LRC-yc _ transcript _ 11245/11527/11650/11862/11868/11882/12097/12170/12189/12554/16323/17633/17850/23601/23702/23827/25797/27323/27340/28075/32363/33320/41302/46740/49192/50155/52414/53263/55803/57837/58842/62065/63807/64453/64477/64945/65389/65522/67967/69329/69714/70293/71698/72016/8397/9305/9759
	CAD	47	LRC-yc _ transcript _ 21713/26724/32991/36203/38794/39400/41583/43709/50240/50773/51593/54764/56608/61115/62613/63233/63277/64316/64398/65025/65274/65382/67040/67463/7718
	4CL	25	LRC-yc _ transcript _ 11192/11915/12074/34290/47743/59763/60274/65237/66044/15128/24756/49633
	POD	12	LRC-yc_transcript_26055/4371/4564/48540/57805/5882/59436/62349/63793
	PAL	9	LRC-yc_transcript_21309/28365/35468/35534/48329/53101/63956
	REF1	7	LRC-yc_transcript_10343/28451/34759/36226/9615
	HCT	5	LRC-yc_transcript_27595/36266/59231/8738
	C4H	4	LRC-yc_transcript_10325/5622/572/64950
	CCR	4	LRC-yc_transcript_11559/20227/60103/71792
	COMT	4	LRC-yc_transcript_16824/68008/8006
	F5H	3	LRC-yc_transcript_52849/7810
	C3H	2	LRC-yc_transcript_8408
	UGT72E	1	LRC-yc_transcript_38451
生物碱类	CCoAOMT	1	LRC-yc_transcript_12962
	CSE	1	LRC-yc_transcript_70530/9998/17539/18647/25280/30784/39996/51797/53404/5453/54972/66338/68223/30126/9042/9967/10366/47306/60096/60503/63976
	GOT2	21	LRC-yc _ transcript _ 10448/18787/23392/24538/24987/3710/39783/46780/49002/53455/56653/66264/69586
	TAT	13	LRC-yc _ transcript _ 1279/3033/3207/34713/34795/37873/4266/4678/55421/65450/66112/67396
	AOC	12	LRC-yc_transcript_32369/34252/40120/4952/57332/60093/61984/62839
	TYR	8	LRC-yc_transcript_30789/52129/55114/60559/71600/8585
	TYDC	6	LRC-yc_transcript_25904/27453/31671/60165/7595
	PPO	5	LRC-yc_transcript_21126/43517
	TR1	2	LRC-yc_transcript_65706
	MAO	1	LRC-yc_transcript_9257
类黄酮	HisC	1	LRC-yc_transcript_21008/24139/34411/34446/47746/70719/7612
	F3'H	7	LRC-yc_transcript_43638/51082/70287
	CHS	3	LRC-yc_transcript_51887
	CHI	1	

2.3.2 生物碱类生物合成相关基因挖掘 生物碱类是一类含氮碱性的天然有机化合物,广泛存在于自然界中,是许多药用植物的重要活性成分<sup>[41]</sup>.在鹿茸草转录本的 KEGG 注释通路中,共有 115 个转录本注释到了 2 条关于生物碱类生物合成的通路,它们分别为异喹啉生物碱生物合成通路(66 个)与莨菪烷、哌啶和吡啶类生物碱生物合成通路(49 个).涉及生物碱生物合成通路的酶共有 9 种(见表 4),天冬氨酸转氨酶(aspartate aminotransferase ,GOT2)、酪氨酸转氨酶(tyrosine aminotransferase ,TAT)和伯

胺氧化酶(primary-amine oxidase ,AOC)是异喹啉生物碱生物合成通路与莨菪烷、哌啶和吡啶类生物碱生物合成通路共同注释到的酶.另外,异喹啉生物碱生物合成通路还注释到酪氨酸酶(tyrosinase ,TYR)、酪氨酸脱羧酶(tyrosine decarboxylase ,TYDC)、多酚氧化酶(polyphenol oxidase ,PPO)和单胺氧化酶(monoamine oxidase ,MAO).在异喹啉生物碱生物合成通路中注释到的 7 种酶大多为通路上游的关键酶,而通路下游的酶未被注释到(见图 3).

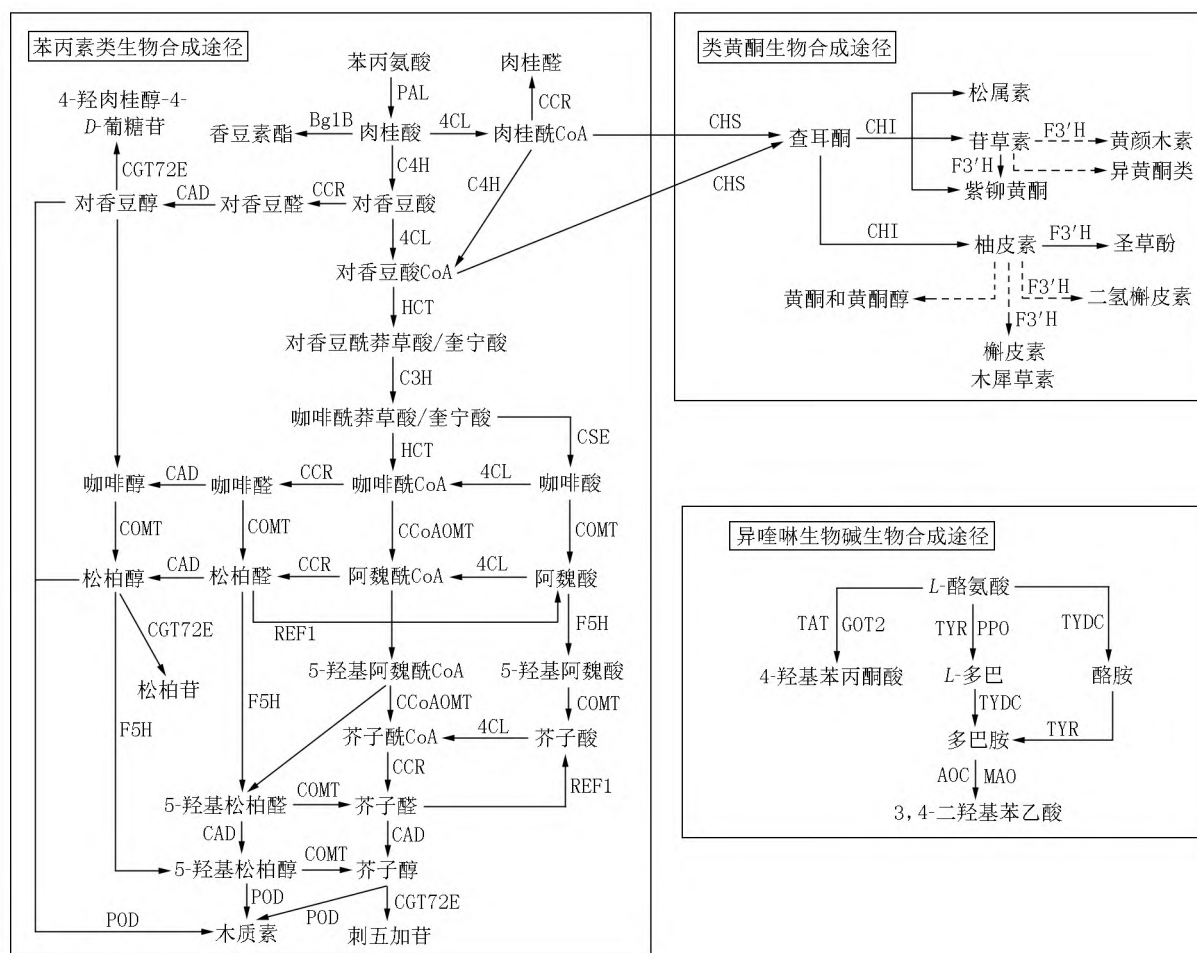


图 3 鹿茸草苯丙素类、黄酮类和异喹啉生物碱生物合成途径

2.3.3 类黄酮生物合成相关基因挖掘 在植物体内,类黄酮化合物是一类极为常见且重要的次生代谢产物,其结构与功能都十分多样<sup>[42]</sup>.类黄酮化合物在植物体内的生物合成是通过苯丙素生物合成途径来进行的,主要是从苯丙素生物合成途径中的对香豆酰 CoA 进入类黄酮的生物合成途径(见图 3).本文 KEGG 共注释到鹿茸草 23 条转录本参与类黄酮生物合成通路,该通路的专用酶有查耳酮合酶(chalcone synthase ,CHS)、查耳酮异构酶(chalcone

isomerase ,CHI) 和类黄酮-3'-羟化酶(flavonoid-3'-hydroxylase ,F3'H),它们分别注释到 3 条、1 条和 7 条转录本(见表 4).

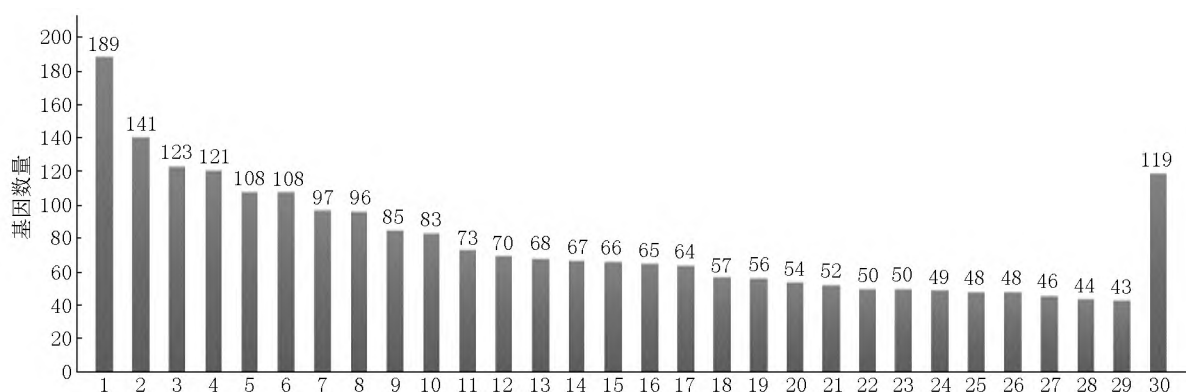
2.3.4 次生代谢后修饰酶相关基因挖掘 在植物次生代谢产物的后修饰中,氧化是最常见且极其重要的一种修饰反应,主要由细胞色素 P450(cytochrome P450 ,CYP450) 进行催化<sup>[43]</sup>.根据鹿茸草在 Swiss-Prot 数据库的比对结果中发现,共有 153 条转录本被注释到 23 个不同的 CYP450 家族中.CYP72、

CYP71 和 CYP83 家族的转录本较多,分别为 25 条、20 条和 13 条,而 CYP85、CYP89 和 CYP716 家族均只有 1 条转录本。糖基化也是存在于植物次生代谢过程中非常重要的一种修饰反应,主要由尿苷二磷酸-糖基转移酶(UDP-glycosyltransferases, UGT)催化<sup>[44]</sup>。在鹿茸草的 Swiss-Prot 数据库比对结果中,共搜寻到与 UGT 相关的转录本 51 条,分布在 12 个不同的 UGT 亚家族中。其中 UGT73 和 UGT86 亚家族的转录本最多,均为 10 条;其次是 UGT74 亚家族 9 条;最少的是 UGT91 和 UGT92 亚家族,各仅注释到 1 条转录本。

## 2.4 CDS 与转录因子分析

使用 TransDecoder 软件对转录本序列的编码区序列及其对应氨基酸序列进行预测,总共获得

46 061 条 ORF(open reading frame),其中完整的 ORF 有 32 529 条。完整 CDS 序列长度大致分布在 1~3 000 bp 之间,其中 99.21% 的 CDS 序列主要集中在 1~1 400 bp 之间。用植物转录因子预测软件 iTAK 预测在鹿茸草全长转录组序列中的转录因子,共预测到 54 个不同转录因子家族的 4 226 个转录因子(见图 4)。其中 C2H2、zn-clus、bZIP、C3H、bHLH、MYB-related、SNF2 和 CAMK\_CDPK 等家族的转录因子数目较多,分别为 189 个(4.47%)、141 个(3.34%)、123 个(2.91%)、121 个(2.86%)、119 个(2.82%)、108 个(2.56%)、108 个(2.56%)、97 个(2.30%)和 95 个(2.25%)。这说明在鹿茸草的各种生理代谢过程中,各种类型的转录因子发挥了重要的作用。



注:图中立柱从左至右分别以 1~30 表示不同的转录因子。1 为 C2H2; 2 为 zn-clus; 3 为 bZIP; 4 为 C3H; 5 为 bHLH; 6 为 MYB-related; 7 为 SNF2; 8 为 CAMK\_CDPK; 9 为 RLK-Pelle\_DLSV; 10 为 SET; 11 为 TRAF; 12 为 PHD; 13 为 WRKY; 14 为 TKL-PI-4; 15 为 B3-ARF; 16 为 NAC; 17 为 CMGC\_MAPK; 18 为 GRAS; 19 为 AUX/IAA; 20 为 Jumonji; 21 为 MYB; 22 为 CAMK\_CAMKL-CHK1; 23 为 GNAT; 24 为 FAR1; 25 为 GARP-G2-like; 26 为 HB-HD-ZIP; 27 为 AP2/ERF-ERF; 28 为 WNK\_NRBP; 29 为 RLK-Pelle\_RLCK-VIIa-2; 30 为其他类型。

图 4 转录因子分析结果

## 2.5 LncRNA 分析

利用 cnci 分析、cpc 分析、pfam 蛋白结构域分析和 cpat 分析方法对转录本进行 LncRNA 的预测;然后统计不同软件预测到的 LncRNAs 序列并绘制韦恩图进行分析,最终得到了 1 820 条 LncRNA 序列(见图 5)。其中,序列长度小于 300 nt 的 LncRNAs 的数量最多,有 804 条(44.18%);长度介于 300~600 nt 的 LncRNAs 有 455 条(25.00%);长度介于 600~900 nt 的 LncRNAs 有 140 条(7.69%);最长的 LncRNA 序列长度达 6 000 nt。将预测到的 1 820 条 LncRNAs 进行靶基因预测,其中 1 804 条(99.12%) LncRNA 成功预测到 36 140 条靶转录本,每个 LncRNA 预测的靶转录本数量范围是 1~30 条。

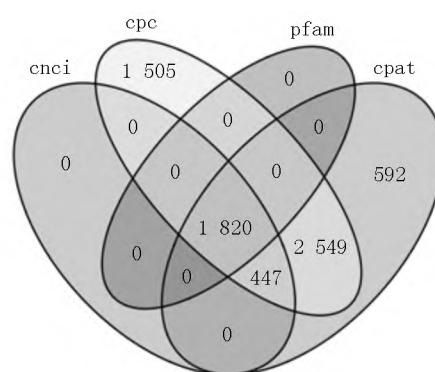


图 5 LncRNAs 韦恩图分析结果

## 3 讨论

鹿茸草全草可入药,具有抑菌、抗炎和抗病毒等



多种生物活性,是国家中药保护品种炎宁颗粒的主药材之一<sup>[45]</sup>。目前,有关鹿茸草化学成分、药理活性等方面的研究较为深入,但关于其分子生物学方面的研究很少,特别是次生代谢生物合成相关基因的发掘、克隆及功能鉴定等研究鲜见报道,严重阻碍了鹿茸草次生代谢产物的开发与利用。近年来,第3代高通量测序技术发展十分迅速,为鹿茸草等非模式植物功能基因的挖掘提供了极大便利<sup>[46]</sup>。本文采用PacBio第3代转录组测序技术对鹿茸草进行全长转录组测序,其平均测序深度为57X,在对FLNC序列聚类后,得到72 231条准确度大于99%的高质量转录本序列,去冗余后共获得了48 005条全长转录本。在无参考基因组的情况下,共有45 362条(94.50%)转录本被成功注释;在KEGG通路注释结果中,发现389条转录本参与了鹿茸草苯丙素类、生物碱类和类黄酮等10个次生代谢生物合成标准通路,这为鹿茸草次生代谢相关基因挖掘、克隆及功能鉴定等基因组学研究提供了丰富的数据基础。

苯丙素类化合物是在玄参科植物中十分常见且发挥重要功效的成分,研究发现苯丙素类化合物特殊的结构使其在抗氧化、抗肿瘤、阵痛降压和护心保肝等方面具有良好的生物活性<sup>[47-48]</sup>。其代谢途径以莽草酸途径的部分中间体为核心单位,形成苯丙氨酸进入代谢途径,从而产生苯丙烯、木质素、黄酮、香豆素和木脂素类等重要的次生代谢物<sup>[49]</sup>。该途径涉及PAL、C4H、4CL、HCT、CCR、POD、CAD和COMT等多个酶。其中,PAL、C4H和4CL催化该途径的前3步为该途径后续所有分支和所产生的代谢物提供基础,是在苯丙素类代谢途径中的3个关键酶。PAL处于苯丙素代谢途径的上游位置,能催化L-苯丙氨酸脱氨从而生成反式肉桂酸,可以调控初级与次级代谢之间的连接,是苯丙素代谢途径的限速酶<sup>[50]</sup>。C4H是苯丙素代谢途径的第2个关键酶,可以催化肉桂酸生成对香豆酸,作为该途径的第1个氧化反应,C4H在植物体内活性的大小会直接影响植物木质素和黄酮类化合物等生物合成代谢支路<sup>[51-52]</sup>。4CL作为苯丙素代谢途径的第3个关键酶,可以对不同的羟基肉桂酸进行催化生成对应的辅酶A酯,使苯丙素代谢途径进入不同的苯丙素类衍生物支路,对类黄酮、木质素和香豆素等化合物的合成具有重要调控作用<sup>[53]</sup>。本文鹿茸草全长转录组KEGG注释结果显示:发现与PAL、C4H和4CL这3个关键酶合成相关的转录本分别有9条、4条和25条。此

外,还有156条转录本可能参与在编码苯丙素类生物合成途径中的12种酶。这说明苯丙素生物合成途径在鹿茸草中较为活跃,相关酶基因的挖掘有利于后续对该途径的调控。

除苯丙素类化合物外,生物碱类和类黄酮化合物也是鹿茸草重要的次生代谢产物。生物碱类广泛存在于植物体内,具有抗病毒、抗肿瘤和消炎镇痛等多种功效<sup>[54-55]</sup>,其结构也复杂多样,可根据不同结构将其分为有机胺类、异喹啉类、莨菪烷类、吡啶类和吲哚类等多种类别<sup>[56]</sup>。本文KEGG注释结果显示:分别有66条和49条转录本注释到异喹啉生物碱生物合成通路与莨菪烷、哌啶和吡啶类生物碱生物合成通路中,涉及参与9种相关酶的合成。异喹啉生物碱的生物合成主要由酪氨酸经过天冬氨酸转氨酶(GOT2)、酶酪氨酸转氨酶(TAT)、4-羟基苯丙酮酸脱羧酶(4HPPDC)和酪氨酸脱羧酶(TYDC)等多种酶催化后生成4-羟基苯乙醛和多巴胺这2种重要的前体<sup>[57]</sup>。鹿茸草异喹啉生物碱的生物合成通路主要注释到上游前体合成中的相关酶,如GOT2、TAT和TYDC分别注释到了21条、13条和6条,但是催化生成4-羟基苯乙醛的4HPPDC却没有注释到相关的转录本。类黄酮化合物结构多样,可将其分为查尔酮、黄酮、黄酮醇、黄烷醇、花青素和原花青素或缩合单宁6个亚类<sup>[58]</sup>。不同的类黄酮化合物具有不同的生物学活性,目前发现其具有抑菌、抗氧化、抗衰老、调节血脂血糖和提高免疫力等多种功效<sup>[59-60]</sup>。通过对鹿茸草KEGG通路注释结果进行搜索,注释到了CHS、CHI和F3'H,这3个酶是类黄酮生物合成专用的关键酶。特别是CHS,它是类黄酮生物合成途径的第1个专用酶,用于合成查尔酮支架,然后全部的类黄酮化合物都是从查尔酮支架中衍生出来的。

次生代谢产物在合成后大多要对其骨架进行修饰,如氧化、糖基化、甲基化和乙酰化等修饰,在经过修饰后其结构、种类、稳定性和药用活性等都会发生变化,更有利于其在植物体内发挥作用<sup>[61-62]</sup>。CYP450在次生代谢产物氧化修饰中发挥了极大作用,它能将氧原子插入底物中,从而对苯丙素类、生物碱类和甾醇类等多种次生代谢产物进行修饰,提高其生物活性<sup>[63]</sup>。UGT负责次生代谢产物的糖基化修饰,将在糖基供体上的糖基转移至被修饰的次生代谢化合物上,从而使其亲水性、稳定性和生物活性等发生改变<sup>[64-65]</sup>。本文共有153条转录本被注释为CYP450基因,51条转录本被注释为UGT基因。在

鹿茸草转录组中多种 CYP450 和 UGT 的发现不仅有利于更具体地分析 CYP450 和 UGT 的催化功能,而且也有利于进一步阐明鹿茸草在次生代谢后在修饰中的氧化反应和糖基化反应机制。

## 4 结论

本文基于 Pacbio 第 3 代测序技术对鹿茸草进行全长转录组测序,获得了大量转录本数据信息,丰富了鹿茸草现有遗传信息,为后期进行鹿茸草遗传多样性分析、遗传图谱构建等分子生物学方面的研究提供了参考数据。同时,鹿茸草次生代谢产物生物合成相关基因的获得为进一步阐明次生代谢产物生物合成的分子机制奠定了基础。

## 5 参考文献

- [1] 中国科学院中国植物志编委会. 中国植物志: 第 68 卷 [M]. 北京: 科学出版社, 1963: 389-390.
- [2] ZHANG Muhan, CHEN Yulu, OUYANG Yao, et al. The biology and haustorial anatomy of semi-parasitic *Monochasma savatieri* Franch. ex Maxim [J]. Plant Growth Regulation 2015, 75(2): 473-481.
- [3] 国家中医药管理局《中华本草》编委会. 中华本草: 精选本(上册) [M]. 上海: 上海科技出版社, 1998: 338.
- [4] 杨燕妮, 王毅, 孙晓波, 等. 炎宁糖浆抗炎活性及其作用机制研究 [J]. 中国药理学通报 2018, 34(12): 1760-1766.
- [5] ZHENG Wei, TAN Xingqi, GUO Liangjun, et al. Chemical constituents from *Monochasma savatieri* [J]. Chinese Journal of Natural Medicines 2012, 10(2): 102-104.
- [6] SHI Mengfan, HE Wenjun, LIU Yanli, et al. Protective effect of total phenylethanoid glycosides from *Monochasma savatieri* Franch on myocardial ischemia injury [J]. Phytomedicine 2013, 20(14): 1251-1255.
- [7] 杨蓓芬, 崔敏燕. 绵毛鹿茸草的次生代谢产物含量与抑菌活性分析 [J]. 浙江中医药大学学报 2009, 33(2): 265-267.
- [8] 乡世健, 潘林燕, 王婉菱, 等. 鹿茸草抗炎、止咳作用的初步研究 [J]. 时珍国医国药 2012, 23(11): 2727-2728.
- [9] 傅建兴, 王利祥, 吴军. 炎宁颗粒治疗急性上呼吸道感染临床观察 [J]. 海峡药学 2010, 22(10): 171-173.
- [10] 马静, 许琼明, 吴文倩, 等. 炎宁颗粒有效部位化学成分研究 [J]. 中成药 2012, 34(10): 1946-1948.
- [11] 慧芳, 刘秀岩, 李宗谕, 等. 转录组测序技术在药用植物研究中的应用 [J]. 中草药 2019, 50(24): 6149-6155.
- [12] 罗纯, 张青林, 罗正荣. 第二代测序技术在植物遗传研究中的应用 [J]. 广东农业科学 2015, 42(3): 186-192.
- [13] 刘振, 徐建红. 高通量测序技术在转座子研究中的应用 [J]. 遗传 2015, 37(9): 885-898.
- [14] 张子敬, 刘燕蓉, 张顺进, 等. 第三代测序技术的方法原理及其在生物领域的应用 [J]. 中国畜牧杂志 2020, 56(6): 11-15.
- [15] 马丽娜, 杨进波, 丁逸菲, 等. 三代测序技术及其应用研究进展 [J]. 中国畜牧兽医 2019, 46(8): 2246-2256.
- [16] 俞晓玲, 姜文倩, 郑玲, 等. 单分子测序技术及应用研究进展 [J]. 生物化学与生物物理进展 2020, 47(1): 5-16.
- [17] 李文燕, 崔百元, 刘勤坚. 薏苡苗期叶片全长转录组测序、注释及 SSR 预测 [J]. 分子植物育种 2020, 18(15): 4855-4870.
- [18] CUI Yupeng, GAO Xinqiang, WANG Jianshe, et al. Full-length transcriptome analysis reveals candidate genes involved in terpenoid biosynthesis in *Artemisia argyi* [J]. Frontiers in Genetics 2021, 12: 659962.
- [19] 尹彦棚, 丁乔娇, 罗加伟, 等. 基于 Pacbio 第三代测序技术的厚朴基因组测序分析 [J]. 广西植物 2021, 41(8): 1251-1262.
- [20] 尚晓尧, 周玲芳, 尹芊芊, 等. 蒺藜苜蓿 (*Medicago truncatula*) 全长转录组测序及分析 [J]. 生物技术通报 2021, 37(8): 131-140.
- [21] CHEN Yeyu, YANG Huanchao, CHEN Yanling, et al. Full-length transcriptome sequencing and identification of immune-related genes in the critically endangered *Hucho bleekeri* [J]. Developmental & Comparative Immunology 2021, 116: 103934.
- [22] GAO Xue, GUO Fengxia, CHEN Yuan, et al. Full-length transcriptome analysis provides new insights into the early bolting occurrence in medicinal *Angelica sinensis* [J]. Scientific Reports 2021, 11(1): 13000.
- [23] CARTOLANO M, HUETTEL B, HARTWIG B, et al. cDNA library enrichment of full length transcripts for SMRT long read sequencing [J]. PLoS One 2016, 11(6): e0157779.
- [24] LI Weizhong, 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.
- [25] 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 Research 1997, 25(17): 3389-3402.

- [26] 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, 76.
- [27] BOUTET E, LIEBERHERR D, TOGNOLLI M, et al. UniProtKB/Swiss-Prot [M] // EDWARDS D. Plant bioinformatics: methods and protocols. Totowa: Humana Press, 2007: 89-112.
- [28] ASHBURNER M, BALL C A, BLAKE J A, et al. Gene Ontology: tool for the unification of biology [J]. Nature Genetics, 2000, 25(1): 25-29.
- [29] 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 Research, 2000, 28(1): 33-36.
- [30] KOONIN E V, FEDOROVA N D, JACKSON J D, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes [J]. Genome Biology, 2004, 5(2): R7.
- [31] JENSEN L J, JULIEN P, KUHN M, et al. eggNOG: automated construction and annotation of orthologous groups of genes [J]. Nucleic Acids Research, 2008, 36(S1): D250-D254.
- [32] FINN R D, BATEMAN A, CLEMENTS J, et al. Pfam: the protein families database [J]. Nucleic Acids Research, 2014, 42: D222-D230.
- [33] KANEHISA M, GOTO S, KAWASGHIMA S, et al. The KEGG resource for deciphering the genome [J]. Nucleic Acids Research, 2004, 32: D277-D280.
- [34] JO Yeonhwa, CHU Hyosub, CHO Jinkyong, et al. De novo transcriptome assembly of two different peach cultivars grown in Korea [J]. Genomics Data, 2015, 6: 260-261.
- [35] ZHENG Yi, JIAO Chen, SUN Honghe, et al. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases [J]. Molecular Plant, 2016, 9(12): 1667-1670.
- [36] KONG Lei, ZHANG Yong, YE Zhiqiang, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine [J]. Nucleic Acids Research, 2007, 35: W345-W349.
- [37] LUO Haitao, BU Dechao, SUN Liang, et al. De novo approach to classify protein-coding and noncoding transcripts based on sequence composition [M] // ALVAREZ M L, NOURBAKHSH M. RNA mapping: methods and protocols. Totowa: Humana Press, 2014: 203-207.
- [38] WANG Liguang, PARK H J, DASARI S, et al. CPAT: coding-potential assessment tool using an alignment-free logistic regression model [J]. Nucleic Acids Research, 2013, 41(6): e74.
- [39] LI Jianwei, MA Wei, ZENG Pan, et al. LncTar: a tool for predicting the RNA targets of long noncoding RNAs [J]. Briefings in Bioinformatics, 2015, 16(5): 806-812.
- [40] 曾兰亭, 杨子银. 茶树苯丙素类/苯环型挥发性物质的生物合成和胁迫响应的研究进展 [J]. 热带亚热带植物学报, 2019, 27(5): 591-600.
- [41] 梁晓莲, 刘纤纤, 李文莉, 等. 生物碱类化合物抗病毒研究进展 [J]. 辽宁中医药大学学报, 2021, 23(4): 51-57.
- [42] FERREYRA M L F, RIUS S P, CASATI P. Flavonoids: biosynthesis, biological functions, and biotechnological applications [J]. Frontiers in Plant Science, 2012, 3: 222.
- [43] 汪思远, 蒋世翠, 王康宇, 等. 植物细胞色素 P450 的研究进展 [J]. 吉林蔬菜, 2014(4): 41-45.
- [44] 郭淑, 罗红梅, 宋经元, 等. 糖基转移酶在植物次生代谢途径中的研究进展 [J]. 世界科学技术(中医药现代化), 2012, 14(6): 2126-2130.
- [45] 朱莹莹, 李敏, 陈重, 等. 炎宁颗粒的化学成分研究 [J]. 中成药, 2013, 35(5): 988-994.
- [46] XIAO Mei, ZHANG Ye, CHEN Xue, et al. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest [J]. Journal of Biotechnology, 2013, 166(3): 122-134.
- [47] 王博. 玄参属植物的化学成分和药理活性研究进展 [J]. 化工设计通讯, 2018, 44(6): 211.
- [48] 徐世林. 苯丙素苷类化合物的合成研究进展 [J]. 杨凌职业技术学院学报, 2008, 7(3): 31-32.
- [49] VOGT T. Phenylpropanoid biosynthesis [J]. Molecular Plant, 2010, 3(1): 2-20.
- [50] 程水源, 陈昆松, 刘卫红, 等. 植物苯丙氨酸解氨酶基因的表达调控与研究展望 [J]. 果树学报, 2003(5): 351-357.
- [51] 程俊, 程曦, 盛玲玲, 等. 杨山酥梨肉桂酸 4-羟化酶基因的克隆及表达分析 [J]. 农业生物技术学报, 2016, 24(11): 1698-1708.
- [52] 张华玲, 刘绪, 杨春贤, 等. 甘薯肉桂酸 4-羟化酶基因克隆及序列分析 [J]. 广西植物, 2018, 38(4): 501-508.
- [53] LAVHALE S G, KALUNKE R M, GIRI A P. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants [J]. Planta, 2018, 248(5): 1063-1078.
- [54] 张攀, 张宜凡, 陈群力, 等. 生物碱类天然产物抗乳腺癌肿瘤机制研究进展 [J]. 中国中药杂志, 2021, 46(2): 312-319.
- [55] 王玲燕, 夏桂阳, 夏欢, 等. 天然生物碱类镇痛成分的研究进展 [J]. 中国中药杂志, 2020, 45(24): 5829-5839.

- [56] MOHAMADI N ,SHARIFIFAR F ,POURNAMDARI M ,et al. A review on biosynthesis ,analytical techniques ,and pharmacological activities of trigonelline as a plant alkaloid [J]. Journal of Dietary Supplements ,2018 ,15( 2) : 207-222.
- [57] 刘金凤 ,黄鹏 ,卿志星 ,等. 苄基异喹啉类生物碱生物合成与代谢工程研究进展 [J]. 基因组学与应用生物学 , 2016 ,35( 8) : 2194-2200.
- [58] WINKEL-SHIRLEY B. Flavonoid biosynthesis: a colorful model for genetics ,biochemistry ,cell biology and biotechnology [J]. Plant Physiology 2001 ,126( 2) : 485-493.
- [59] PETRUSSA E ,BRAIDOT E ,ZANCANI M ,et al. Plant flavonoids: biosynthesis ,transport and involvement in stress responses [J]. International Journal of Molecular Sciences 2013 ,14( 7) : 14950-14973.
- [60] 王玲玲 ,边祥雨 ,高蔚娜 ,等. 植物类黄酮提取纯化技术研究进展 [J]. 营养学报 2019 ,41( 6) : 606-610.
- [61] 赵乐 ,朱昀昊 ,张莉 ,等. 基于转录组测序挖掘商陆皂苷甲生物合成相关基因 [J]. 药学报 ,2017 ,52( 9) : 1471-1480.
- [62] 李军玲 ,罗晓东 ,赵沛基 ,等. 植物萜类生物合成中的后修饰酶 [J]. 云南植物研究 2009 ,31( 5) : 461-468.
- [63] SCHULER M A ,WERCK-REICHHART D. Functional genomics of P450s [J]. Annual Review of Plant Biology , 2003 ,54: 629-667.
- [64] GACHON C M ,LANGLOIS-MEURINNE M ,SAINDRENAN P. Plant secondary metabolism glycosyltransferases: the emerging functional analysis [J]. Trends in Plant Science 2005 ,10( 11) : 542-549.
- [65] 陈欣 ,付锐锐 ,张鸿 ,等. 药用植物中 UGT 家族研究进展 [J]. 生物资源 2018 ,40( 5) : 443-449.

## The Full – Length Transcriptome Sequencing and Identification of Related Genes Involved in Secondary Metabolism Biosynthesis for *Monochasma savatieri*

YANG Wanling ,BAI Zhiyi ,ZOU Mingzhu ,WANG Xinru ,XIE Jiankun ,ZHANG Fantao \*

( College of Life Sciences ,Jiangxi Normal University ,Nanchang ,Jiangxi 330022 ,China)

**Abstract:** To obtain the full-length transcriptome sequence database and mine the genes of enzymes related to the secondary metabolic compounds' biosynthetic pathway in *Monochasma savatieri* ,based on Single Molecule Real Time ( SMRT) technology ,the full-length transcriptome is sequenced using Pacbio high-throughput sequencing platform. A total of 48 005 high-quality transcripts are obtained. 45 362 transcripts are successfully annotated by BLAST search against NR ,Swiss-Prot ,GO ,KEGG and other four public sequence databases ,with an annotation rate of 94. 50% . Among them 389 transcripts are annotated to the 10 standard secondary metabolic biosynthetic pathways of KEGG. Further analysis reveals that a total of 194 transcripts are involved in phenylpropanoid biosynthesis ,115 transcripts are related to alkaloid biosynthesis ,23 transcripts might participate in flavonoid biosynthesis ,and 57 transcripts are involved in other secondary metabolites. Additionally 204 transcripts relate to secondary metabolic oxidation and glycosylation post-modifications. The obtained data of this study greatly enriches the genetic information of *Monochasma savatieri* and preliminarily reveals the genes involved in the synthesis of secondary metabolites ,which lays a foundation for further studies on function and regulatory mechanisms of key genes involved in the synthesis of secondary metabolites in *Monochasma savatieri*.

**Key words:** *Monochasma savatieri* Franch. ex Maxim; full-length transcriptome; secondary metabolism; biosynthesis; genetic resource

( 责任编辑: 刘显亮)