

• 药材与资源 •

赶黄草的转录组测序及分析

袁 灿, 彭 芳[#], 钟文娟, 牟方生, 戢沛城, 龚一耘, 蒲德强, 叶 霄, 黄海燕, 杨 晓^{*}, 张 超^{*}
四川省农业科学院经济作物育种栽培研究所, 四川 成都 610300

摘要: 目的 获得赶黄草 *Penthorum chinense* 转录基因参考序列和表达量信息, 探索赶黄草药用活性成分生物合成的遗传基础, 为赶黄草的重要功能基因研究提供参考。方法 采用 Illumina HiSeq 2000 高通量测序技术, 对赶黄草进行转录组测序, 对获得的数据进行过滤组装, 对得到的 unigene 进行比对注释, 同时对本研究获得赶黄草活性成分合成相关代谢通路基因进行分析。结果 共获得了 40 005 442 条有效的短读序, 通过 *de novo* 拼接得到 42 306 个 unigene, 开放阅读框 (ORF) 分析共计得到 518 个 ORF, 其中 75 个 ORF 具有转录因子结构域。通过 KEGG 分析, 检测到 33、32、59、68 个 unigene 分别参与黄酮类生物合成、固醇类生物合成、萜类骨架生物合成和羧酸代谢。结论 本研究发现的这些基因对赶黄草遗传改良和药用活性物质的生物合成路径相关基因研究有着重要意义。

关键词: 赶黄草; 转录组; 槲皮素; 黄酮类; 固醇类; 萜类

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2017)21-4507-08

DOI: 10.7501/j.issn.0253-2670.2017.21.023

De novo assembly and transcriptome characterization of *Penthorum chinense*

YUAN Can, PENG Fang, ZHONG Wen-juan, MOU Fang-sheng, JI Pei-cheng, GONG Yi-yun, PU De-qiang,
YE Xiao, HUANG Hai-yan, YANG Xiao, ZHANG Chao

Industrial Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610300, China

Abstract: Objective In order to obtain the reference sequences and relative expression of transcription genes and study the genetic base of active ingredients in *Penthorum chinense*, which were useful for researching functional gene to *P. chinense*. **Methods** In this study, by performing Illumina Hiseq 2000 and *de novo* assembly, the transcriptome of whole plant was sequenced, the data were filtered and assembled, and the unigene was compared and annotated. Meanwhile, the genes related to the synthesis of metabolic pathway of active ingredients in *P. chinense* were analyzed. **Results** Totally, 40 005 442 valid short sequences were obtained, and 42 306 unigenes were spliced by *de novo*. Also, a total of 518 open reading frames (ORF) were obtained by ORF analysis, and 75 ORF of them had transcription factor domains. In addition, by performing KEGG pathway analysis, 33, 32, 59, and 68 unigenes were found to be involved in the pathway of flavonoid biosynthesis, steroid biosynthesis, terpenoid backbone biosynthesis, and 2-oxocarboxylic acid metabolism, respectively. **Conclusion** The datasets provided in this study will contribute significantly to genetic improvement and study on the genes related to biosynthesis pathway of pharmaceutical active substances from *P. chinense*.

Key words: *Penthorum chinense* Pursh.; transcriptome; quercetin; flavonoids; sterols; terpenoids

赶黄草又名水泽兰、水杨柳等, 学名扯根菜 *Penthorum chinense* Pursh., 是虎耳草科扯根菜属植物, 为苗族传统药物, 嫩叶、花和茎均可作药用。其性温、味甘、无毒, 具有清热解毒、活血散瘀、

利水消肿、退黄、平肝的功效^[1], 有“神仙草”之美誉。赶黄草始载于《救荒本草》, 随后在《中药大辞典》^[2]和《四川中药志》^[3]均有记载, 同时也被载入《四川省中药材标准》。其主要分布于我国云贵

收稿日期: 2017-04-19

基金项目: 四川省育种攻关项目 (2016NYZ0022); 四川省科技支撑计划项目 (2014SZ0134)

作者简介: 袁 灿, 硕士, 主要从事药用植物资源创新、遗传育种与栽培研究。Tel: (028)68907211 E-mail: 609669081@qq.com

*通信作者 张 超, 研究员。Tel: (028)68907207 E-mail: jychoazhang@163.com

杨 晓, 副研究员。Tel: (028)68907227 E-mail: 21sctcm@163.com

[#]并列第一作者

川一带,四川省古蔺县是栽培品的道地产区,以该地产赶黄草为原料制成的“古蔺肝舒”也是我国第一个获得世界原产地域产品保护的中成药。现代医学研究表明赶黄草具有很强的保护肝脏作用,此外还有利胆退黄、抗氧化、抑菌等作用。近年来许多活性成分从赶黄草中分离得到,主要为黄酮类、有机酸和甾醇等,其中以槲皮素为代表。目前对赶黄草的研究主要集中在化学成分和药理活性上,其药用活性成分分子遗传机制解析、相关代谢通路的基因发掘和育种利用却鲜有报道,仅林淑芳等^[4]利用 ISSR 标记对不同来源赶黄草进行了聚类分析。

转录组(RNA-seq)测序是研究生物在特定时期和特定组织基因表达的重要方法,转录组测序具有偏重基因编码区域,重复序列和 GC 富集区域较少,拼接相对容易的特点^[5-6]。近年来,转录组测序在药用植物研究上得到了较广泛的应用,并在中医药研究领域中发挥出重要作用^[7]。一批药用植物,如人参^[8]、三七^[9]、西洋参^[10]、川芎^[11]、丹参^[12]等转录组测序分析已完成,发现了大量与药用次生代谢产物合成调控相关的基因。但相对于模式植物和主要农作物,药用植物的遗传机制、基因组信息、功能基因等分子研究还比较薄弱。本研究利用 Illumina Hiseq 2000 对赶黄草全株进行了转录组测序,以期获得与其有效成分合成相关的基因信息,为赶黄草药用活性成分的生物合成调控、遗传育种提高赶黄草品质和产量等研究提供理论依据。

1 材料与方法

1.1 转录组测序

样品是四川省农业科学院经济作物育种栽培研究所选育,经四川省农作物品种审定委员会审定的赶黄草 2 号新品种,该品种经四川大学张浩教授鉴定为扯根菜 *Penthorum chinense* Pursh。选取赶黄草 2 号 3 个单株地上部分为研究材料,在现蕾前采样,等量混合成一个样品后液氮速冻保存于-80℃备用。采用 RNAqueous[®] Total RNA Isolation Kit[™] 试剂盒进行样品的总 RNA 抽提。用 RNAClean XP Kit 和 RNase-Free DNase Set 试剂盒进行纯化。纯化后的总 RNA 进行 mRNA 的富集,然后构建文库,文库用 Qubit[®] 2.0 Fluorometer 和 Agilent 2100 进行质量检测,文库质量合格后委托上海伯豪生物技术有限公司用 Illumina HiSeq 2000 进行测序,采用双端测序方式,测序长度 125 bp,最终获得数据量为 6.34 Gb。

1.2 测序数据的处理

用软件 FASTX-Toolkit (0.0.13) 过滤原始数据中低的短读序(reads):去除 reads 中质量(Q)大于 20 的碱基所占比例小于 50%的 reads;去除 3'端质量 Q 低于 20 的碱基,即碱基错误率小于 0.01,其中 $Q = -10 \log_{10} \text{error_ratio}$;去除接头序列;去除含有模糊碱基数大于 5%的 reads;去除长度小于 20 的测序片段 reads 和 ribosome RNA reads。

1.3 转录本拼接组装

测序数据用 CLC Genomics Workbench(6.0.4)^[13-15]的 scaffolding contig 算法进行 *de novo* 拼接,采用 word-size=45,最小重叠群长度≥300。应用 CAP3 EST 拼接软件对初始 unigene 进行第二次拼接,得到最终 unigene。

1.4 unigene 功能注释

拼接得到 unigene 序列利用 Blast+ (ncbi-blast-2.2.28+) 与 UniProt 数据库和 Nr 数据库进行 Blastx^[16]比对 ($E < 1 \times 10^{-5}$),应用 rpstblastn 程序将 unigene 与 CDD 数据库^[17]进行比对 ($E \leq 1 \times 10^{-5}$),进行 KOG 功能分类预测;利用 HMMER 3.0 软件的 hmmscan 程序与 Pfam 结构数据库进行比对 ($E = 0.01$);利用 EMBOSS (6.4.0) -getorf 对所有 unigene 进行开放阅读框(ORF)预测,选择最长且长度大于 300 bp 的 ORF 作为 unigene 的 ORF。利用 hmmsearch 程序对所有的 ORF 通过 HMM 模型与 PlantTFDB^[18]的转录因子结构域进行比对,对 ORF 进行转录因子预测。

1.5 GO 分析和 KEGG 代谢通路分析

将 unigene 序列与 GO (gene ontology) 数据库^[19]进行 blastx 比对 ($E < 1 \times 10^{-5}$),按分子功能(molecular function)、细胞组分(cellular component)、生物过程(biological process)进行分类。应用 KEGG KAAS 在线 pathway 比对分析工具对 unigene 进行 KEGG (kyoto encyclopedia of genes and genomes)映射分析。

1.6 基因表达定量分析

利用 RPKM (reads per kilo bases per million reads)标准^[20]对 reads 覆盖度进行基因表达量计算。

2 结果与分析

2.1 转录组测序及序列组装

将得到的 mRNA 经过反转录后制成 200 bp 插入文库,通过 Illumina Hiseq 2000 高通量测序,共获得 42 285 922 条读序,利用 FASTX-Toolkit 进行

滤过得到 40 005 442 条有效的短读序。平均读序长度为 94.74 bp，总核苷酸数 6.34 Gb，经过 *de novo* 拼接，得到的 contig 共计 50 625 个(≥200 bp)，总长度约 47.38 Mb，平均长度 936 bp，N50 为 1 342 bp；得到的初始 unigene 共计 43 143 条，总长度 47.53 Mb，平均长度 1 102 bp，N50 为 1 566 bp；应用 CAP3 拼接软件对初始 unigene 进行拼接，获得最终 unigene 为 42 306

个(≥300 bp)，总长度约 47.34 Mb，平均长度 1 119 bp，N50 为 1 601 bp，GC 比例为 39.58% (表 1)。

在总 unigene 中，所占比例最多的 unigene 大小为 400~600 bp，共 16 461 条 (38.91%)，大于 1 000 bp 的 unigene 有 8 689 条，占 20.54%；大于 2 000 bp 的有 6 084 条，占 14.38% (图 1)。unigene 整体长度分布较均匀。

表 1 序列拼接结果

Table 1 Statistics of sequence assembly result

序列种类	计数	总长度/bp	N25/bp	N50/bp	N75/bp	平均长度/bp	最长长度/bp	GC/%
contig	50 625 个	47 375 784	2 241	1 342	691	936	18 210	39.60
初始 unigene	43 143 条	47 532 326	2 507	1 566	804	1 102	18 210	39.60
最终 unigene	42 306 条	47 339 926	2 550	1 601	824	1 119	18 210	39.58

N25 将转录本从长到短排序，依次累加碱基数，当累计碱基数达到转录本总碱基数的 25% 时的最短转录本长度；N50、N75 为类似的方法统计 N25 length means the size of the contig, which selection criteria was the added length of the contig and contigs longer than it equal to 25% of total contigs length; N50 and N75 are similarly defined

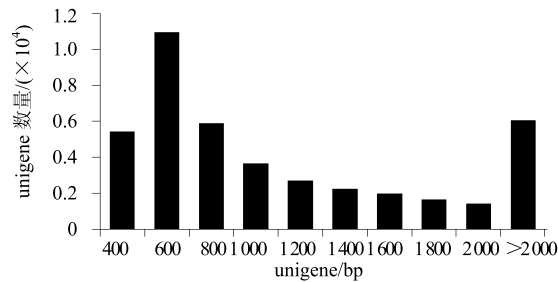


图 1 unigene 长度频数分布图

Fig. 1 Frequency distribution of sequence length

2.2 unigene 的功能注释

将拼接得到的 unigene 序列与 Uniprot、Nr、Pfam 数据库进行比对 (表 2)。42 306 个 unigene 比对结果显示，在 Nr 数据库中有 24 935 个 unigene 获得同源匹配信息，注释比例 58.9%，相似度在 60%~80% 的最多，约占 46%；与 Nr 数据结果类似，在 Uniprot 数据库中有 25 046 个 unigene 获得同源匹配信息，注释比例 59.2%，能比对上的序列较 Nr 数据库略多，在 Uniprot 数据库中仍然以 60%~80% 的相似度最多；2 个数据比对结果中，分别有 107 和 113 完全一致的序列，这些基因可能各个物种间保守性较高。Pfam 数据库是一个蛋白家族的大集合，通过与 Pfam 数据库比对，对 unigene 进行蛋白家族的注释，在 Pfam 数据库中有 9 544 个 unigene 获得同源匹配信息，注释比例 22.6%。

2.3 KOG 注释与分类

应用 rpstblastn 系统将 unigene 与 CDD 数据库

表 2 unigene 的功能注释

Table 2 Function annotation of unigenes

数据库	总序列数目	注释上数目	注释比例/%
Nr	42 306	24 935	58.9
Uniprot	42 306	25 046	59.2
Pfam	42 306	9 544	22.6

进行比对 ($E \leq 1 \times e^{-5}$)，进行 KOG 功能分类预测，14 138 个 unigene 被注释到 25 种 KOG 分类中 (图 2)。其中信号传导所占的 unigene 最多，为 16.0%，其次是翻译后修饰、蛋白质转换、伴侣蛋白，占 11.2%，再次是一般功能预测，占 10.4%，细胞运动类 unigene 最少，仅占 0.02%。有 2 043 个 unigene 为未知功能，占 4.4%。

2.4 ORF 预测和转录因子预测

利用 EMBOSS (6.4.0)-getorf 对所有的 unigene 进行 ORF 预测，共计得到 518 个 ORF。通过 hmmsearch 程序对所有的 ORF 与 PlantTFDB 的转录因子结构域进行比对，结果显示共有 75 个 ORF 含有转录因子结构域 (表 3)。共检测到 20 种不同类型转录因子，其中最多的为 NAC 类转录因子共 29 个，占有所有转录因子的 38.7%，其次是 WRKY 类转录因子共 19 个，占有所有转录因子的 25.3%，其余的各类型转录因子数量均较少，只有 1~5 个。

2.5 GO 注释和分类

将 unigene 序列与 GO 数据库进行 Blastx 比对分析，结果 12 960 个 unigene 按照功能被分为 3 大类：

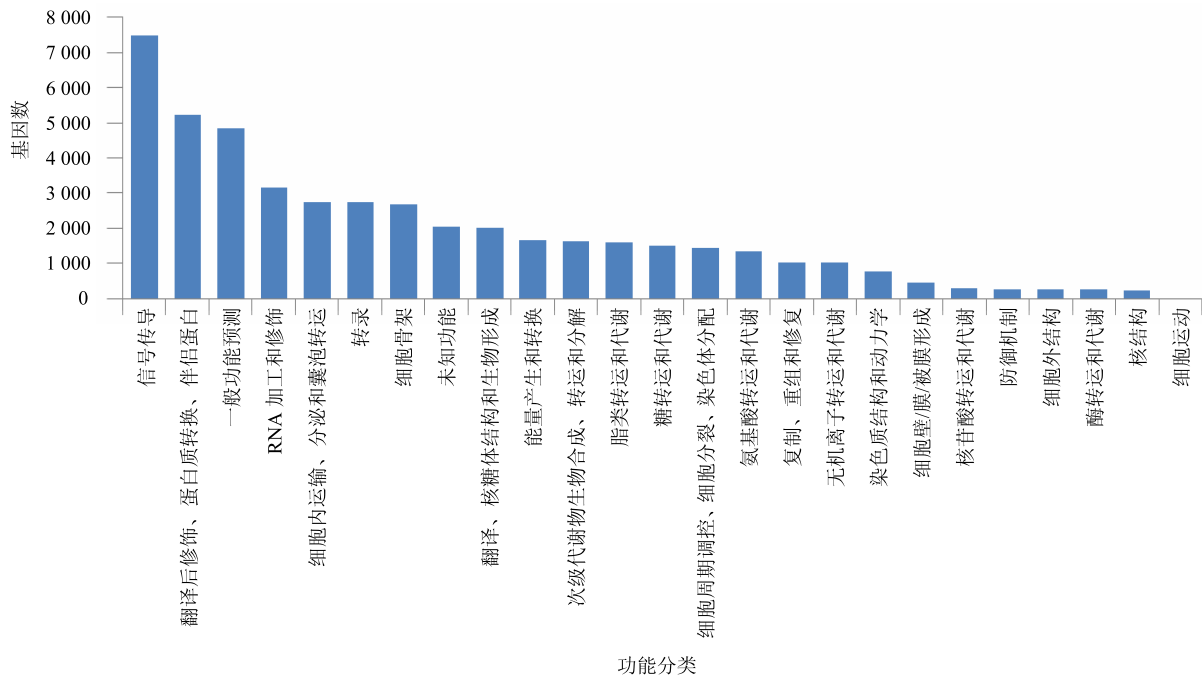


图 2 KOG 功能基因分类

Fig. 2 KOG functional categories

表 3 转录因子预测

Table 3 Prediction of transcription factors

转录因子类型	数量	转录因子类型	数量
AP2	1	GATA	1
ARF	1	HSF	1
ARR-B	1	LBD	1
bHLH	2	MYC	1
bZIP	1	MYB	2
C2H2	3	MYB_related	2
C3H	5	NAC	29
CAMTA	1	NF-X1	1
ERF	1	Trihelix	1
G2-like	1	WRKY	19

细胞组分、分子功能和生物学过程(图3),其中细胞组分 unigene 占 31.0%,分子功能 unigene 占 23.3%,生物学过程 unigene 占 45.7%。在这3大功能分类基础上,进一步细分为55个功能组。其中匹配最多的功能组分别是代谢过程有 8 561 个 unigene,结合功能有 7 208 个 unigene,细胞生理过程有 6 948 个 unigene,催化活性功能有 6 456 个 unigene。匹配最少的功能组分别是细胞运动、细胞外基质组分、蛋白标签,均只有 1 个 unigene 被匹配(图3)。

2.6 KEGG 注释和分类

将 42 306 个 unigene 序列进行 KEGG 映射分析,

共映射到 246 条代谢路径。包含功能基因序列较多的是代谢通路(2 382 个)、次生代谢合成通路(1 426 个)、抗生素物质生物合成通路(633 个)、不同环境下微生物代谢(586 个)。赶黄草含有黄酮、有机酸、萜类、甾体类等有效成分,其中以黄酮类化合物居多^[21]。本研究对赶黄草这4类重要药效物质相关的次生代谢产物生物合成路径进行统计分析,分析结果显示,分别有 33、68、59、32 个功能基因序列映射到黄酮、有机酸、萜类、甾体类合成4种代谢路径(表4)。

近年来许多研究表明赶黄草具有很强的肝保护作用,所以其活效成分研究主要集中在与其功能属性密切相关的黄酮类及槲皮素上^[22-23],据此本研究对赶黄草转录组测序获得的黄酮类生物合成的 33 个基因的表达量进行统计分析(表5),结果显示,33 个基因中 RPKM 最高的 5 个基因均为查耳酮合成酶(chalcone synthase, CHS)基因,最高 RPKM 达 1 460,较高的有类黄酮单加氧酶(flavonoid 3'-monooxygenase, F3'H)基因、柚皮素双加酶(naringenin 3-dioxygenase)基因、查耳酮异构酶(chalcone isomerase, CHI)基因,它们的 RPKM 分别为 184、174、137。

通过 KEGG 代谢通路分析,槲皮素生物合成主要有 3 个路径:一是通过咖啡酰辅酶 A 直接经过一

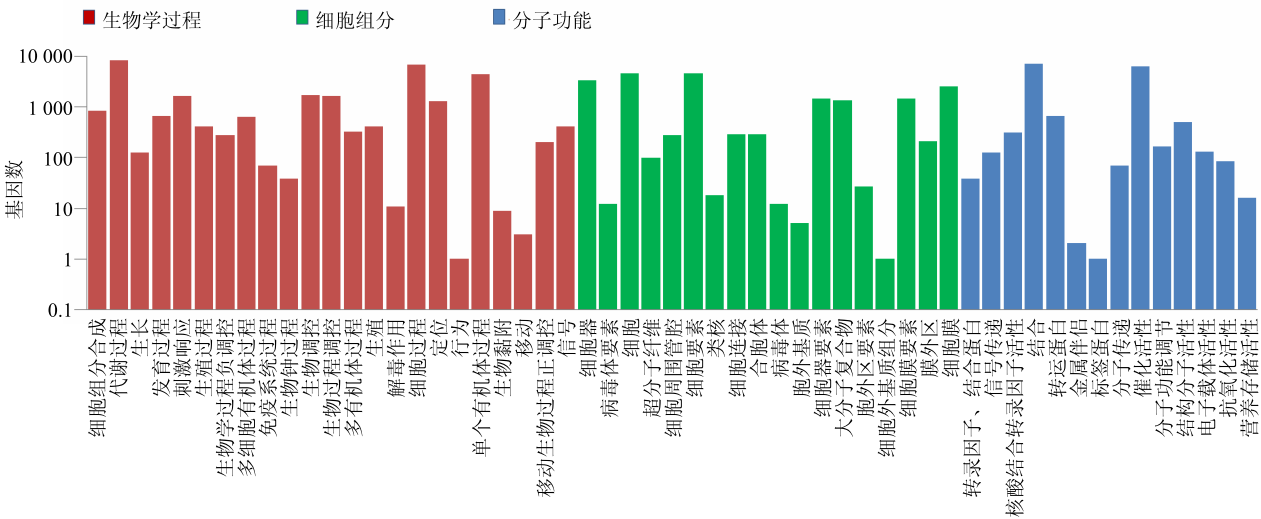


图 3 unigene 的 GO 注释与分类
Fig. 3 GO annotation and classification of unigenes

表 4 赶黄草重要药效物质相关的次生代谢产物生物合成途径

Table 4 Biosynthetic pathway of secondary metabolites involved in major active substances of <i>P. chinense</i>				
路径 ID	路径	覆盖到该通路的基因数	该通路中的基因数	通路中覆盖的功能基因序列数目
ko00941	黄酮类生物合成	12	20	33
ko00100	固醇类生物合成	19	31	32
ko00900	萜类骨架生物合成	29	54	59
ko01210	羧酸代谢	31	76	68
ko01110	次生代谢合成	410	956	1 246

系列酶催化后获得槲皮素；二是通过 P-香豆酰辅酶 A 经过一系列酶催化后生成山奈酚，山奈酚在经过 F3'H 催化后生成槲皮素；三是 P-香豆酰辅酶 A 在经过 CHS 催化后生成柚配基查耳酮，柚配基查耳酮在经过 F3'H 催化生成圣草酚查耳酮，然后通过第一个途径生成槲皮素，或者利用第 2 个途径对柚配基查耳酮进一步修饰后，转入第 1 个途径生成槲皮素。这 3 个生物合成途径中，参与的主要基因有查耳酮合成酶、查耳酮异构酶、黄烷酮-3-羟化酶、类黄酮羟化酶和黄酮醇合成酶，通过表达量分析发现这些基因的表达量均较高。

最新研究也提出将赶黄草酮 B（木脂素类）作为质量标志物（Q-Marker）^[24]。随后，本研究也对木脂素类相关的类苯基丙烷代谢途径相关基因进行了分析。共有 116 个 unigene 映射到该代谢通路中的 19 个关键酶上，这些基因的 RPKM 值在 10 以上占 50%左右，最高达 2 120。

3 讨论

3.1 Illumina HiSeq 2000 高通量测序技术在非模式物种功能注释基因的应用

随着二代测序成本的降低，测序技术不断的更

新，测序技术被广泛的应用到植物基因组研究中^[25]，近年在药用植物功能基因发掘的文献也越来越多^[26-27]。王晓玥等^[28]通过对人参、丹参、东北红豆杉、三七、银杏等进行转录组测序，发掘出 200 多个合成萜类、黄酮、酚酸类等次生代谢相关的基因以及修饰基因 2 000 多个，同时还提出用转录组测序技术来开展药材道地性的分子机制研究。然而至今没有识别赶黄草转录组特性的基因序列研究报道。本研究采用 Illumina HiSeq 2000 高通量测序平台对赶黄草转录组进行测序，获得 40 005 442 条有效的短读序，经过 *de novo* 拼接，得到的 contig 重叠群共 50 625 个，赶黄草功能基因序列 42 306 个。在 Nr、Uniprot、Pfam 数据库中获得同源匹配信息的功能基因数分别达 58.9%、59.2%和 22.6%。这些结果将有助于进一步研究赶黄草的生长发育和次生代谢产物的生物合成过程，为研究赶黄草相关基因的功能等提供参考。

3.2 赶黄草中转录因子分析

转录因子作为重要的调控基因，通过转录水平调控到植物生长发育、抗性、活性物质合成，近年

表 5 赶黄草黄酮类合成基因表达定量

Table 5 Gene expression analysis of flavonoids biosynthesis in *P. chinense*

unigene	表达量	通路编号	酶的描述
YC-GHC_no_rRNA.1_(paired)_contig_523	1 460.43	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_710	306.64	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_2256	290.10	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_2335	275.29	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_3004	273.53	K00660	chalcone synthase
First_Contig740	185.20	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_960	184.69	K05280	flavonoid 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_3245	174.20	K00475	naringenin 3-dioxygenase
YC-GHC_no_rRNA.1_(paired)_contig_6996	159.91	K05277	leucoanthocyanidin dioxygenase
YC-GHC_no_rRNA.1_(paired)_contig_711	156.24	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_517	137.03	K01859	chalcone isomerase
YC-GHC_no_rRNA.1_(paired)_contig_4824	130.48	K00487	trans-cinnamate 4-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_8564	84.76	K05278	flavonol synthase
YC-GHC_no_rRNA.1_(paired)_contig_4417	69.13	K08695	anthocyanidin reductase
YC-GHC_no_rRNA.1_(paired)_contig_709	62.07	K00660	chalcone synthase
YC-GHC_no_rRNA.1_(paired)_contig_3049	45.50	K05278	flavonol synthase
YC-GHC_no_rRNA.1_(paired)_contig_9093	39.40	K09754	coumaroyl quinate 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_15808	25.98	K00588	caffeoyl-CoA-O-methyltransferase
YC-GHC_no_rRNA.1_(paired)_contig_1486	25.64	K13082	flavanone 4-reductase
YC-GHC_no_rRNA.1_(paired)_contig_9639	16.33	K00588	caffeoyl-CoA-O-methyltransferase
YC-GHC_no_rRNA.1_(paired)_contig_15056	14.33	K13065	shikimate-O-hydroxycinnamoyltransferase
YC-GHC_no_rRNA.1_(paired)_contig_17459	14.33	K13065	shikimate-O-hydroxycinnamoyltransferase
YC-GHC_no_rRNA.1_(paired)_contig_11507	12.87	K05280	flavonoid 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_20811	7.48	K09754	coumaroyl quinate 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_15940	4.70	K05280	flavonoid 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_22770	4.09	K05280	flavonoid 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_2942	3.49	K13082	flavanone 4-reductase
YC-GHC_no_rRNA.1_(paired)_contig_25166	3.05	K01859	chalcone isomerase
YC-GHC_no_rRNA.1_(paired)_contig_27486	2.04	K05280	flavonoid 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_30655	2.03	K05280	flavonoid 3'-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_25877	1.74	K00487	trans-cinnamate-4-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_22109	0.74	K00487	trans-cinnamate-4-monooxygenase
YC-GHC_no_rRNA.1_(paired)_contig_40527	0.45	K00487	trans-cinnamate-4-monooxygenase

来在药用植物研究中被广泛关注^[29]。本研究通过对含有 ORF 结构的序列进行转录因子预测, 共有 20 类转录因子被检测到, 其中 NAC 类最多。NAC 类转录因子只存在于植物中^[30], 其数量众多, 一直是研究的热点^[31], 拟南芥和水稻中分别有 117 个和 151 个 NAC^[32]基因。NAC 被报道^[33]参与植物细胞壁形成、衰老、分枝、叶形等, 本研究所获得 NAC 类转录因子, 将有助于赶黄草高产、抗性基因的转录调控研究, 为赶黄草叶片、分枝等发育研究和育种应用奠定基础。在赶黄草中槲皮素等黄酮醇类化

合物被广泛研究, 黄酮醇类化合物也是许多药用植物的主要药性成分, 目前报道参与黄酮类代谢相关的转录因子有 MYB、MYC、WRKY、bHLH、WD40 等^[34-35], 在本研究得到的 WRKY 类转录因子数量仅次于 NAC 类转录因子, 这些转录因子的序列和表达量的信息对赶黄草活性物质合成代谢调控具有极为重要作用。

3.3 次生代谢合成

在《四川省中药材标准》2010 年版中, 槲皮素是赶黄草的唯一指标成分, 因此本研究对赶黄草槲

皮素合成通路的基因进行了分析。通过分析 KEGG 注释的代谢通路发现,在槲皮素合成的 3 个途径中,其合成酶能在赶黄草中被完全映射,且这些基因的表达水平都较高,这也是首次从基因表达水平上验证了赶黄草中槲皮素量较高。此外,通过分析还发现一些与槲皮素衍生物相关的基因的存在,这些基因同样高量表达,如异槲皮素合成过程中催化活性糖基转移的类黄酮-3-O-葡萄糖基转移酶(UDP-glucose)基因,槲皮素苷合成中的黄酮-3-O-鼠李糖糖基转移酶(UGT78D1)基因等,这与文献报道中在赶黄草中检测到槲皮素衍生物结果一致^[36]。黄酮醇合酶基因(FLS)在槲皮素合成中催化二氢黄酮醇到黄酮醇,是黄酮醇代谢途径关键酶,同时也是黄酮类合成途径与儿茶素合成途径的桥梁。在茶叶中黄酮醇类物质发生糖苷形成黄酮苷是绿茶汤色的重要组分,目前赶黄草也被大量开发做保健茶,本研究共得到 2 条 FLS 序列,其序列信息和表达量信息将有该基因的深入研究和赶黄草品质的改良。熊亮等^[24]认为赶黄草木脂素是比较特异的成分且具有修复损伤小鼠肝细胞的能力,与赶黄草功能相关且含量较高,适合作为 Q-Marker。本研究在对木脂素相关的苯丙烷类代谢途径进行分析,共计得到 116 个 unigene 参与该代谢途径。这一代谢途径源于莽草酸途径,其起始底物为苯丙氨酸和酪氨酸,再经过苯丙氨酸解氨酶(PAL)、4-香豆酸辅酶 A 连接酶(4CL)、反式肉桂酸 4 单氧酶(C4H)3 个酶的催化后生成黄酮和木质素代谢途径底物 P-香豆酰辅酶 A。PAL 为该代谢通路第 1 个催化酶,也是苯丙烷类代谢途径关键基因^[37],该基因在许多植物中存在多拷贝现象,其表达量呈现时空特异性,响应不同的环境信号。本研究在赶黄草中共鉴定出 12 条 PAL 序列,其中主要表达的序列有 2 条, PAL 家族仅少量基因表达其余同源基因沉默的现象在番茄中也被发现^[38]。C4H 为苯丙素代谢通路第 2 个催化反应酶,该酶属于细胞色素 P450(CYP450)家族,在不同的植物其拷贝数不同,本研究共鉴定出 4 条 C4H 序列。苯丙素代谢途径的第 3 个酶为 4CL,该酶控制苯丙烷类代谢走向不同的代谢途径,本研究共鉴定出 7 条 4CL 序列。本研究将促进这些基因全长序列的获得,为阐明赶黄草的苯丙素代谢通路以及木脂素合成机制奠定了基础。

参考文献

[1] 全国中草药汇编编写组. 全国中草药汇编(下册)[M].

第 2 版. 北京: 人民卫生出版社, 2000.

- [2] 江苏新医学院. 中药大辞典(上册)[M]. 上海: 上海人民出版社, 1977.
- [3] 四川中药志协作编写组. 四川中药志(第一卷)[M]. 成都: 四川人民出版社, 1979.
- [4] 林淑芳, 袁媛, 邵爱娟, 等. 不同来源赶黄草的 ISSR 分析[J]. 现代中药研究与实践, 2012, 26(3): 18-21.
- [5] Alagna F, D'Agostino N, Torchia L, et al. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development[J]. BMC Genomics, 2009, 10(1): 399-406.
- [6] Barakat A, DiLoreto D S, Zhang Y, et al. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection[J]. BMC Plant Biol, 2009, 9(1): 51-56.
- [7] 陈士林, 朱孝轩, 李春芳, 等. 中药基因组学与合成生物学[J]. 药学报, 2012, 47(8): 1070-1078.
- [8] Chen S L, Luo H M, Li Y, et al. 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng* [J]. Plant Cell Rep, 2011, 30: 1593-1601.
- [9] Luo H M, Sun C, Song J Y, et al. Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers[J]. BMC Genomics, 2011, 12(5): 1-15.
- [10] 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 Genomics, 2010, 11(1): 262-269.
- [11] Song T, Liu Z B, Li J J, et al. Comparative transcriptome of rhizome and leaf in *Ligusticum Chuanxiong* [J]. Plant System Evol, 2015, 301(8): 2073-2085.
- [12] Xu Z, Peters R J, Weirather J, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis[J]. The Plant J, 2015, 82(6): 951-961.
- [13] Su C L, Chao Y T, Alex C Y C, et al. De novo assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome[J]. Plant Cell Physiol, 2011, 52(9): 1501-1514.
- [14] Garg R, Patel R K, Tyagi A K, et al. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification[J]. DNA Res, 2011, 18(1): 53-63.
- [15] Bräutigam A, Mullick T, Schliesky S, et al. Critical assessment of assembly strategies for non-model species

- mRNA-Seq data and application of next-generation sequencing to the comparison of C(3) and C(4) species [J]. *J Exp Bot*, 2011, 62(9): 3093-3102.
- [16] Altschul S F, Gish W, Miller W, *et al.* Basic local alignment search tool [J]. *J Mol Biol*, 1990, 215(3): 403-410.
- [17] Marchler-Bauer A, Derbyshire M K, Gonzales N R, *et al.* CDD: NCBI's conserved domain database [J]. *Nucleic Acids Res*, 2014, 43: D222-D226.
- [18] Jin J, Zhang H, Kong L, *et al.* PlantTFDB 3. 0: A portal for the functional and evolutionary study of plant transcription factors [J]. *Nucleic Acids Res*, 2014, 42(1): 1182-1187.
- [19] Ashburner M, Ball C A, Blake J A, *et al.* Gene ontology: Tool for the unification of biology. The gene ontology consortium [J]. *Nat Genet*, 2000, 25(1): 25-29.
- [20] Mortazavi A, Williams B A, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq [J]. *Nat Methods*, 2008, 5(7): 621-628.
- [21] 郭 艳, 曾凡骏, 刘 敏, 等. 赶黄草有效成分及生物活性研究进展 [J]. *职业卫生与病伤*, 2016, 31(1): 41-43.
- [22] 汪洪武, 任启生, 冯长根, 等. 赶黄草中黄酮提取方法的研究 [J]. *中国药学杂志*, 2002, 37(7): 551-551.
- [23] 冯 浩, 王智民, 董歌扬, 等. 赶黄草化学成分的研究 [J]. *中国中药杂志*, 2001, 26(4): 260-261.
- [24] 熊 亮, 彭 成. 基于中药质量标志物 (Q-Marker) 的基本条件研究益母草和赶黄草的 Q-Marker [J]. *中草药*, 2016, 47(13): 2212-2220.
- [25] Zhang L, Jia H, Yin Y, *et al.* Transcriptome analysis of leaf tissue of raphanus sativus by RNA sequencing [J]. *PLoS One*, 2013, 8(11): e80350.
- [26] 张绍鹏, 金 健, 胡炳雄, 等. 珍稀药用植物珠子参的转录组测序及分析 [J]. *中国中药杂志*, 2015, 40(11): 2084-2089.
- [27] 杨 杰, 王金权, 丁维俊, 等. 基于 Illumina HiSeq 2000 测序技术对当归根的转录组特性研究 [J]. *中草药*, 2015, 46(8): 1216-1222.
- [28] 王晓玥, 宋经元, 谢彩香, 等. RNA-Seq 与道地药材研究 [J]. *药学报*, 2014, 49(12): 1650-1657.
- [29] 季爱加, 罗红梅, 徐志超, 等. 药用植物转录因子 AP2/ERF 研究与展望 [J]. *科学通报*, 2015, 60(14): 1272-1284.
- [30] Ooka H, Satoh K, Doi K, *et al.* Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana* [J]. *DNA Res*, 2003, 10(6): 239-247.
- [31] 彭 辉, 于兴旺, 成慧颖, 等. 植物 NAC 转录因子家族研究概况 [J]. *植物学报*, 2010, 45(2): 236-248.
- [32] Puranik S, Sahu P P, Srivastava P S, *et al.* NAC Proteins: Regulation and role in stress tolerance [J]. *Trends Plant Sci*, 2012, 17(6): 369-381.
- [33] 杨晓娜, 田 云, 卢向阳. NAC 转录因子在植物生长发育中的调控作用 [J]. *化学与生物工程*, 2014, 31(1): 1-5.
- [34] 高 珂, 王 玲, 吴素瑞, 等. 调控药用植物药效成分合成的转录因子研究进展 [J]. *中草药*, 2015, 46(20): 3100-3108.
- [35] 李宗艳, 李名扬. 调控植物类黄酮生物合成的转录因子研究进展 [J]. *南京林业大学学报: 自然科学版*, 2011, 35(5): 129-134.
- [36] 王 萌, 吴 霞, 江 云, 等. 赶黄草的研究进展 [J]. *食品与药品*, 2013, 15(3): 202-205.
- [37] Thomas V. Phenylpropanoid biosynthesis [J]. *Mol Plant*, 2010, 3(1): 2-20.
- [38] 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*, 283(48): 33591-33601.