

白鲜根转录组高通量测序与数据分析

李依民, 张化为, 陈莹, 张明英, 刘阿萍, 宋小妹, 杨新杰*, 张岗*

陕西中医药大学药学院 陕西省中药基础与新药研究重点实验室, 陕西 西安 712046

摘要: 目的 获得白鲜 *Dictamnus dasycarpus* 根转录组信息特征。方法 以白鲜根为研究对象, 采用 Illumina HiSeq™ 2000 150PE 进行高通量转录组测序并进行数据分析。结果 转录组测序共获得 69 643 286 条高质量序列 (clean reads), Trinity *de novo* 组装获得 49 050 条 unigenes, 平均长度 841 nt。BLAST 分析显示分别有 31 636 (64.49%)、22 367 (45.60%)、19 246 (39.23%)、12 595 (25.68%) 条 unigenes 在 NR、Swiss-port、KOG、KEGG 数据库得到注释信息, 可归为 GO 分类的生物过程、细胞组分和分子功能 3 大类 42 分支, 涉及 132 个 KEGG 标准代谢通路, 其中包括 18 个次生代谢标准通路。进一步分析获得 90 个基因参与多种生物碱生物合成通路。蛋白编码框序列 1 908 个, 高等植物转录因子 55 个家族; 借助 MISA 软件发现 4 579 个 SSRs, 三碱基重复最丰富, 有 2 021 个, 出现频率为 44.1%; 五碱基重复 SSRs 仅占 3.5%。结论 利用高通量测序技术获得白鲜根转录组信息特征, 为后期白鲜基因功能鉴定、次生代谢途径解析及其调控机制研究奠定基础。

关键词: 白鲜; 转录组; 功能基因; 代谢通路; 简单重复序列

中图分类号: R282.12

文献标志码: A

文章编号: 0253-2670(2018)21-4975-08

DOI: 10.7501/j.issn.0253-2670.2018.21.004

High-throughput transcriptome sequencing of roots of *Dictamnus dasycarpus* and data analyses

LI Yi-min, ZHANG Hua-wei, CHEN Ying, ZHANG Ming-ying, LIU A-ping, SONG Xiao-mei, YANG Xin-jie, ZHANG Gang

Shaanxi Key Laboratory of Basis & New Drugs Research of Chinese Medicine, College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an 712046, China

Abstract: **Objective** To obtain the transcriptome dataset of roots of *Dictamnus dasycarpus*. **Methods** The root transcriptome dataset of *D. dasycarpus* was obtained using the high-throughput sequencing platform Illumina HiSeq™ 2000 150PE, followed by systemic bioinformatics analyses. **Results** A great number of 69 643 286 high quality clean reads were obtained by the transcriptome sequencing analyses. Using Trinity *de novo* assembling, a total of 49 050 unigenes were finally obtained, with an average length of 841 nt. BLAST analysis indicated that 31 636 (accounting 64.49% of the total unigenes), 22 367 (45.60%), 19 246 (39.23%), 12 595 (25.68%) unigenes were successfully annotated in the NR, Swiss-port, KOG, and KEGG databases, respectively. And GO classification contained the basic three major groups, including biological process, cellular component, and molecular function with 42 subgroups. A total of 132 KEGG standard metabolic pathways were designated, 18 of which were defined as the secondary metabolism. Further analysis revealed that a total of 90 unigenes were involved in the biosynthesis of various alkaloids. Of all unigenes, 1 908 were predicted to have CDS, and 55 families of plant transcription factors were also identified. Using MISA prediction, 4 579 simple sequence repeats (SSRs) were obtained, among which the tri-nucleotide SSRs were abundant with 2 021 (44.1%), whereas the penta-nucleotide SSRs accounted for 3.5%. **Conclusion** The root transcriptome of *D. dasycarpus* revealed by the high-throughput sequencing technology will be important for gene functional characterization, secondary metabolism pathway exploration, and regulation mechanism research in this species.

Key words: *Dictamnus dasycarpus* Turcz.; transcriptome; functional unigenes; metabolism pathway; simple sequence repeats

收稿日期: 2018-04-05

基金项目: 陕西省教育厅专项 (18JK0216); 陕西中医药大学新进博士科研启动经费 (104080001); 陕西省教育厅科学研究项目 (17JK0215); 陕西省高校青年杰出人才支持计划项目; 陕西省科技厅社发攻关项目 (2013K14-03-08); 咸阳市中青年科技领军人才项目

作者简介: 李依民, 女, 博士, 讲师, 研究方向为中药资源与分子生药学。E-mail: 2051058@sntcm.edu.cn

***通信作者** 张岗, 男, 博士, 教授, 研究方向为中药资源与分子生药学。Tel/Fax: (029)38185165 E-mail: jay_gumling2003@aliyun.com
杨新杰, 男, 博士, 副教授, 研究方向为中药品种、品质及资源开发研究。E-mail: yangxj211@126.com

白鲜 *Dictamnus dasycarpus* Turcz. 为芸香科白鲜属多年生草本植物, 干燥根皮作中药白鲜皮药用, 又名八股牛、野花椒根皮、山牡丹、臭根皮等, 被历代本草及《中国药典》2015 年版收载, 性味苦、咸, 寒, 具有祛风燥湿、祛风除湿、解毒之功效, 主治风热湿毒所致的风疹湿疹、疥癣、黄疸及风湿热痹等^[1]。现代中药化学研究揭示白鲜皮包含挥发油、生物碱、柠檬苦素、多糖、香豆素、黄酮及甾体类等多种有效成分^[2]。其中, 白鲜碱、黄柏酮等具有良好的抗炎^[3]、神经保护^[4]、抗癌^[5]等药理活性。生药学方面开展了白鲜营养器官发育与成分积累关系^[6]及药材真伪鉴别^[7]等研究, 为白鲜皮的质量控制研究及临床应用提供保障。

基于二代高通量测序技术的转录组学, 是本草基因组学研究的一个重要组成部分, 能够在整体水平上揭示细胞所有基因转录本信息特征, 在中草药基因挖掘、次生代谢调控、遗传进化以及道地性研究等方面发挥重要作用^[8]。因此, 该策略在中草药功能基因组研究领域得到快速应用, 现已经获得人参^[9]和甘草^[10]等众多物种的转录组数据, 为阐明中草药种质资源遗传基础奠定基础。白鲜皮临床功效明确、活性成分多样, 具有极高的研究和开发价值。然而, 白鲜资源以野生为主, 随着临床需求量的不断增加及连年的无序采挖, 导致其野生种质资源严重枯竭, 白鲜资源可持续利用研究是亟待关注的热点领域。本研究利用二代高通量测序平台 Illumina HiSeqTM 2000 150PE 进行白鲜根转录组测序分析, 以期揭示白鲜根转录组的整体表达特征, 为该药用植物提供丰富基因资源, 有助于后期基因鉴定、次生代谢途径及调控机制解析。

1 材料与方法

1.1 材料

植物材料于 2015 年 6 月采自陕西省宝鸡市太白县咀头镇鳌山, 经度 107°27'22.5", 纬度 33°53'26.2", 海拔 1 725 m, 经陕西中医药大学张岗教授鉴定为芸香科植物白鲜 *Dictamnus dasycarpus* Turcz.。取单株植株的根于液氮中速冻后, 置于-80℃冰箱备用。

1.2 RNA 提取与文库构建

采用 EASYspin 植物 RNA 快速提取试剂盒 (Aidlab, 中国) 制备白鲜根总 RNA, 琼脂糖凝胶电泳和 NanoDropTM 2000 分光光度计 (Thermo Fisher, 美国) 检测完整性。用带有 Oligo (dT) 的磁珠富集 mRNA, 加入碎片化缓冲液将 mRNA 打断成短片段,

用六碱基随机引物 (random hexamers) 合成 cDNA 第 1 链; 然后加入缓冲液、dNTPs、RNase H 和 DNA polymerase I 合成 cDNA 第 2 链; 再经过 QiaQuick PCR 试剂盒 (QIAGEN, 德国) 纯化并加 EB 缓冲液洗脱之后做末端修复, 加 poly (A) 并连接测序接头, 然后用琼脂糖凝胶电泳进行片段大小选择, 最后进行 PCR 扩增构建测序文库。

1.3 转录组测序与组装

利用 Illumina HiSeqTM 2000 150PE 对白鲜根转录组文库进行高通量测序。测序原始图像数据经碱基识别 (base calling) 转化为序列数据原始序列 (raw reads), 经数据评估、过滤除杂和冗余处理等质控得到高质量序列 (clean reads), 再利用短 reads 组装软件 Trinity 做转录组 *de novo* 组装分析。Trinity 首先将具有一定长度重叠 (overlap) 的 reads 连成更长的片段, 这些通过 reads overlap 得到的不含 N 的组装片段即为 unigene。

1.4 转录组功能注释

利用 BLAST 将 unigenes 与蛋白数据库 NR、Swiss-port、蛋白相邻类的聚簇 (KOG) 和 KEGG (京都基因与基因组百科全书) 进行比对 ($E < 1 \times 10^{-5}$), 得到与相应 unigenes 具有最高序列相似性的蛋白, 得到 unigenes 注释信息。根据 NR 注释信息, 使用 Blast2GO 得到 unigenes 基因本体 (gene ontology, GO) 注释, 用 WEGO 对所有 unigenes 做 GO 功能分类, 从宏观上认识物种基因功能分布特征。

1.5 蛋白编码框 (CDS) 和转录因子预测

按 NR、Swiss-Prot、KOG 和 KEGG 的优先级顺序将 unigene 序列与以上蛋白库做 BLASTx 比对 ($E < 1 \times 10^{-5}$) 并确定该 unigenes 编码区的核酸序列 (序列方向 5'→3') 和氨基酸序列。利用 ESTScan 预测与以上数据库比对不上的 unigenes 的编码区及序列方向。将所预测的 unigenes 编码蛋白序列与植物转录因子数据库 plant TFDB 进行 hmmscan 比对, 搜索转录因子家族及其成员。

1.6 简单重复序列 (simple sequence repeats, SSRs) 特征检测

用 MISA (MicroSAteellite identification tool) 检测白鲜转录组 unigenes 搜索 SSRs 并进行统计分析。

2 结果与分析

2.1 白鲜转录组组装与质量分析

采用 Illumina HiSeqTM 2000 150PE 高通量测序平台对白鲜根进行转录组测序, 共得到 69 643 286

条 raw reads, 过滤产生了 67 538 832 条 clean reads, 包含 6 735 850 678 个核苷酸信息, Q20 (碱基量 $\geq 20\%$) 和 Q30 (碱基量 $\geq 30\%$) 分别为 98.81%、94.21%, GC 量为 43.23%, 说明测序质控良好, clean reads 质量合格。Trinity 组装获得 49 050 个 unigenes, 平均长度 841 nt, 最长达到 13 135 nt, 最短序列为 201 bp, N50 为 1 498 nt。

2.2 白鲜转录组 unigenes 的功能注释

用 BLAST 将所有 unigenes 与 NR、Swiss-port、KOG、KEGG 等数据库进行比对, 对各数据库注释的 unigenes 数目进行统计, 获得白鲜根转录组 unigenes 的功能注释信息。结果表明, 31 636 条 unigenes (64.49%) 在 NR 数据库中得到注释, 在 Swiss-port、KOG、KEGG 等数据库获得注释的 unigenes 数目依次为 22 367 (45.60%)、19 246 (39.23%)、12 595 (25.68%)。注释成功的 unigenes 共 31 955 条 (65.15%), 17 095 条 (34.85%) 未获得注释, 在 4 个数据库同时得到注释的 unigene 有 9 530 条。

unigenes 长度分布 (图 1-A) 显示, 12 009 条 unigenes 长度超过 1 000 nt, 4 930 条序列大于 2 000 nt。reads 在 unigene 上的覆盖情况统计 (图 1-B) 显示, 所含 reads 数量在 11~100 的 unigenes 数量最多, 为 26 692 条; reads 数量在 101~200 的 unigenes 次之, 为 3 998 条; reads 数量在 1 001~2 000、1~10、2 001~4 000 的 unigenes 分别为 3 122、2 195、2 392 条; 其余 reads 分布区域对应的 unigenes 数量均相对较少。以 NR 注释为例分析, unigenes 注释同源基因的物种分布如图 2 所示, 在相似序列匹配度较高的物种中, 甜橙 *Citrus sinensis* (L.) Osbeck 所占比例最高, 19 630 条 (68.06%); 其次为可可 *Theobroma cacao* L. 1 333 条 (4.62%), 土瓶草 *Cephalotus follicularis* Labill. 1 274 条 (4.42%), 克莱门柚 *Citrus clementina* Hort. ex Tan. 677 条 (2.35%), 紫苞花烛 *Anthurium amnicola* Dressler 451 条 (1.56%), 葡萄 *Vitis vinifera* L. 427 条 (1.48%); 其余匹配物种比例在 1.17%~1.23%, 比例小于 1.17% 的匹配物种的 unigenes 为

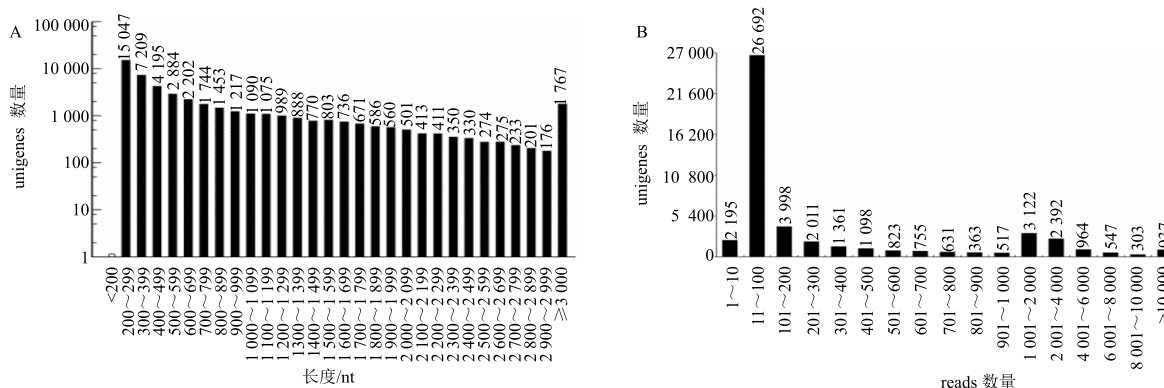


图 1 白鲜转录组 unigenes 长度分布 (A) 和 reads 覆盖统计 (B)

Fig. 1 Length distribution of *D. dasycarpus* transcriptome unigenes (A) and reads coverage (B)

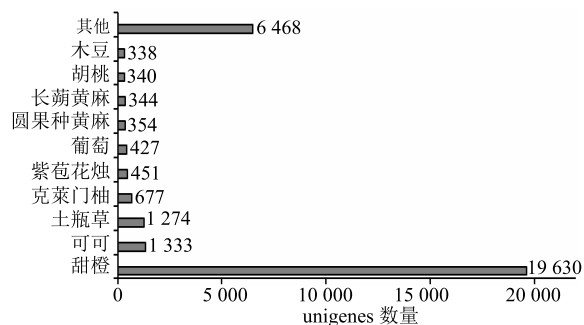


图 2 白鲜转录组 unigenes 与 NR 数据库匹配物种分布

Fig. 2 Species distribution of *D. dasycarpus* transcriptome unigenes against NR database

6 468 条, 占 12.74%。根据 NR 注释信息得到 GO 功能分类 (图 3), 27 208 条 unigenes 被注释到生物过程、细胞组分和分子功能 3 个 GO 类别的 42 个小组。细胞组分中细胞 (cell) 和细胞部分 (cell part) 相关基因丰度最高, 均为 1 693 条; 其次是细胞器 (organelle), 有 1 181 条; 病毒粒子 (virion)、病毒粒子组成 (virion part) 等基因较少, 均为 25 条。生物过程主要聚集在代谢过程 (metabolic process)、细胞过程 (cellular process) 和单一生物体过程 (single organism process), 涉及的基因分别有 3 915、3 229 和 2 369 条; 定位 (localization)、应激适应 (response to stimulus)、生物调控 (biological regulation) 基因

数量分别为 655、608、501 条。分子功能中具有催化活性 (catalytic activity) 和结合功能 (binding) 的基因和数量较高, 分别为 4 005 和 2 638 条, 转运体活性 (transporter activity) 基因数目 295 条, 其他类别基因数目普遍较少。

为了进一步分析白鲜转录组 unigenes 的功能, 进行 KOG 功能分类分析, 共得到 25 个不同的 KOG 功能类群、种类比较全面, 包括大多数的生命活动; 一般功能预测的基因数量最多, 有 6 728 条; 翻译后修饰, 蛋白反转、伴侣和信号转导机制类次之,

分别为 3 515 和 3 049 条; 转录和 RNA 加工与修饰 unigenes 数目为 1 759 和 1 773 条; 其他种类基因丰度不尽相同 (图 4)。

白鲜根转录组 unigenes 参与 KEGG 代谢通路分为 5 大分支: 细胞过程 (cellular processes) 593 条、环境信息处理 (environmental information processing) 450 条、遗传信息处理 (genetic information processing) 2 785、代谢 (metabolism) 12 420 条和有机系统 (organismal systems) 282 条。6 609 条 unigene 获得 132 个 KEGG 标准代谢通路,

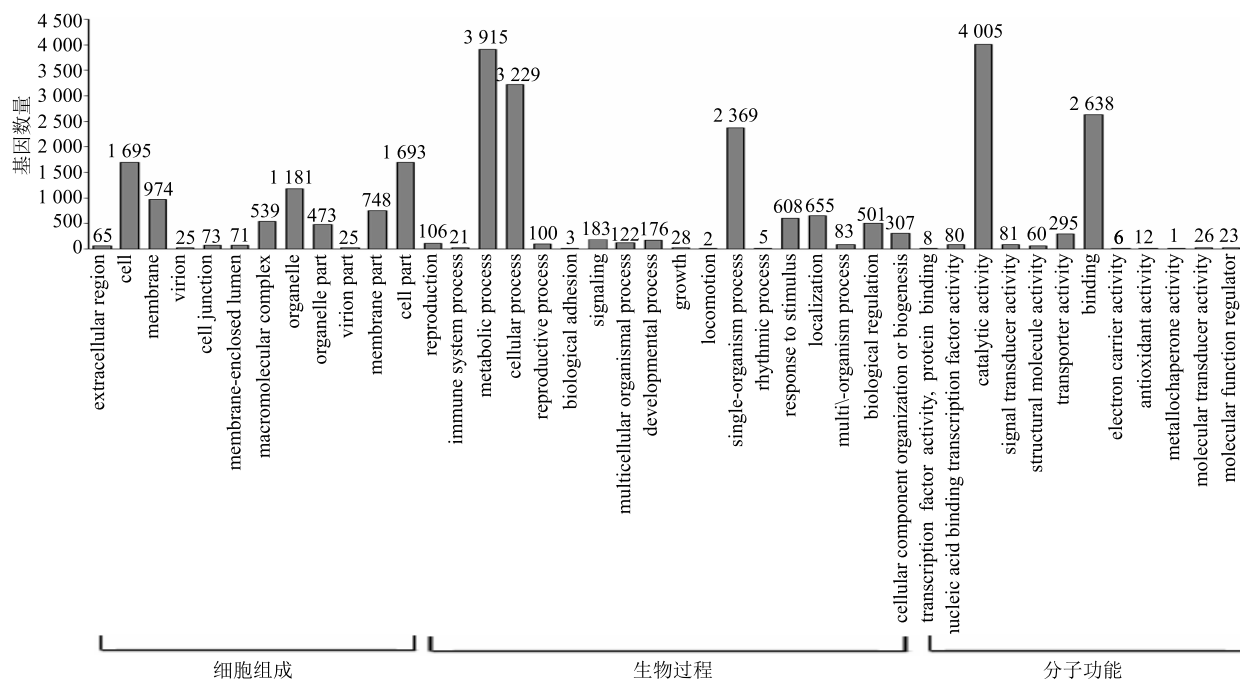


图 3 白鲜转录组 unigenes 的 GO 分类

Fig. 3 GO classification of *D. dasycarpus* transcriptome unigenes

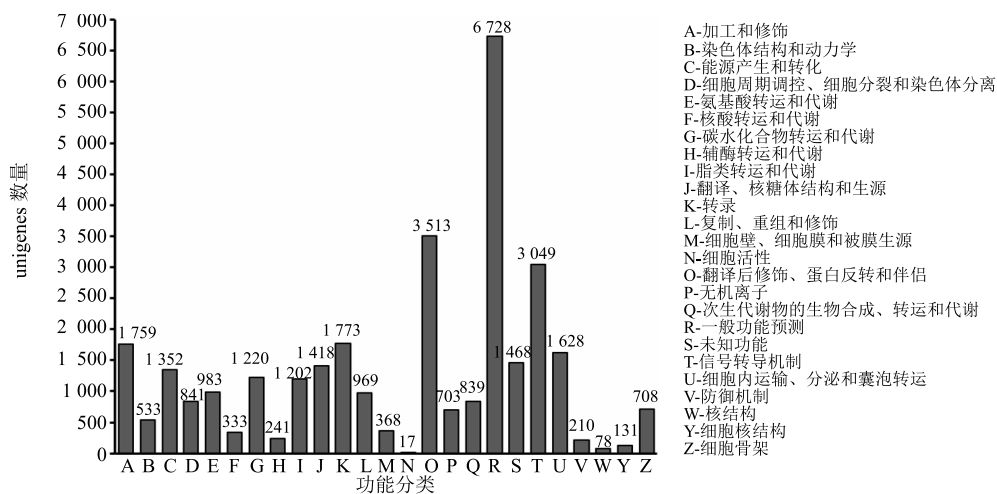


图 4 白鲜转录组 unigenes 的 KOG 注释分布

Fig. 4 KOG annotation distribution of *D. dasycarpus* transcriptome unigenes

按照基因注释量大小依次排序,选取前 18 个代谢通路信息见表 1, 这些通路包含的 unigenes 数量均大于 200 条。

KEGG 代谢通路分析还发现 501 条 unigenes 参与

表 1 白鲜转录组 unigenes KEGG 通路分析统计

Table 1 KEGG pathway analysis of *D. dasycarpus* transcriptome unigenes

编号	代谢通路	unigenes 数量	占比/%	通路 ID
1	代谢通路	2 688	40.67	ko01100
2	次生代谢物生物合成	1 469	22.23	ko01110
3	抗生素生物合成	723	10.94	ko01130
4	多种环境下微生物代谢	590	8.93	ko01120
5	碳代谢	425	6.43	ko03010
6	核糖体	417	6.31	ko04626
7	氨基酸生物合成	366	5.54	ko01200
8	内质网蛋白质加工	312	4.72	ko04141
9	剪接体	281	4.25	ko03040
10	淀粉和蔗糖代谢	276	4.18	ko00500
11	植物激素信号转导	276	4.18	ko04075
12	RNA 转运	265	4.01	ko03013
13	胞吞作用	249	3.77	ko04144
14	泛素介导的蛋白质降解	240	3.07	ko04120
15	植物-病原菌互作	235	3.56	ko04626
16	嘌呤代谢	230	3.48	ko00230
17	氨基糖和核苷酸糖代谢	207	3.13	ko00520
18	糖酵解途径	203	3.07	ko00010

表 2 白鲜转录组 unigenes 次生代谢 KEGG 通路注释统计

Table 2 Secondary metabolism KEGG pathway annotation analysis of *D. dasycarpus* transcriptome unigenes

编号	代谢通路	unigenes 数量	占比/%	通路 ID
1	苯丙素生物合成	186	2.810	ko00940
2	萜类化合物骨架生物合成	75	1.130	ko00900
3	萜类、生物碱、黄酮类等生物合成相关的 16 个次生代谢标准通路 (表 2)。	48	0.730	ko00960
4	苯丙素的生物合成代谢通路 (ko00940) 基因数量最多, 为 186 条。	45	0.680	ko00950
5	萜类化合物骨架生物合成 (ko00900) 基因数量次之, 为 75 条; 其中, 与类胡萝卜素生物合成 (ko00906) 有关的基因有 41 条; 26 条 unigenes 参与柠檬烯和蒎烯降解 (ko00903); 倍半萜和三萜类化合物的生物合成 (ko00909) 基因有 14 条。	41	0.620	ko00906
6	咖啡因代谢、花青素、黄酮和黄酮醇生物合成的通路基因数均在 5 条以下。	26	0.390	ko00903
7	生物碱为白鲜的主要成分之一, 种类多种多样, 其中白鲜碱、黄柏酮等呋喃喹啉类生物碱为重要代表。在白鲜根转录组数据中找到了莨菪烷、哌啶、吡啶、异喹啉、吡啶酮类生物碱生物合成 KEGG 标准通路 (表 2)。	24	0.360	ko00904
8	48 个 unigenes 参与莨菪烷、哌啶、吡啶类生物碱生物合成 (ko00960); 45 个 unigenes 参与异喹啉类生物碱生物合成 (ko00950); 吡啶酮生物碱生物合成 (ko01058) 的通路基因数仅有 1 条, 未发现呋喃喹啉类生物碱代谢标准 KEGG 通路。	16	0.240	ko00905
9	2.3 CDS 和转录因子分析	14	0.210	ko00740
10	对白鲜转录组所有 unigenes 的 CDS 进行分析, 通过 BLAST 比对共获得 CDS 序列 31 259 个, 利用 ESTscan 数据库分析获得 CDS 序列 1 908 个。转录因子预测发现有 55 个家族成员, 其中 ERF、C2H2、bHLH、WRKY、MYB-related、NAC、MYB、bZIP、及 C3H 类占主体, 说明白鲜根生理代谢涉及众多转录调控过程 (图 5)。	12	0.180	ko00908
11		10	0.150	ko00966
12		8	0.120	ko00902
13		4	0.060	ko00232
14		4	0.060	ko00944
15		1	0.015	ko00942
16		1	0.015	ko01058

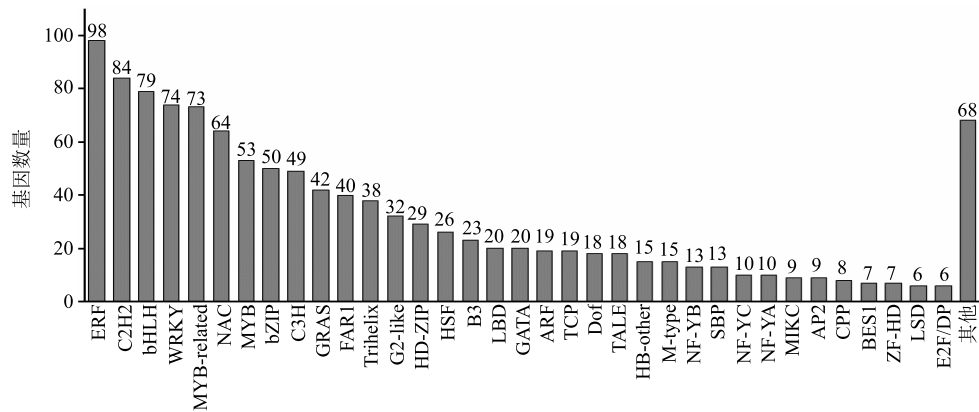


图 5 白鲜转录组 unigenes 的转录因子分析

Fig. 5 Transcription factor classification of *D. dasycarpus* transcriptome unigenes

2.4 SSRs 特征分析

利用 MISA 软件对转录组 unigenes 进行 SSRs 分析 (表 3), 3 836 个 unigenes 中共计 4 579 个 SSRs。三碱基重复 SSRs 数量最丰富, 有 2 021 个 (44.1%), 其中, AAG/CTT 类型的比例最高。双碱基重复 SSRs 数量次之, 有 1 578 个, 占 SSRs

总量的 34.5%, 其中 AG/CT 重复类型数量最多。四碱基和六碱基重复分别为 422、400 个, 各占 9.2%、8.7%; 五碱基重复重复相对较少, 仅占 3.5%。此外, 还发现 SSRs 重复单元数量也存在一定变化, 其中重复 5、6 次的比例最高, 重复 4、7 次的次之。

表 3 白鲜转录组 unigenes SSRs 分析

Table 3 SSRs analyses of *D. dasycarpus* transcriptome unigenes

重复	重复单元数量												合计	占比/%
	4	5	6	7	8	9	10	11	12	13	14	≥15		
二碱基重复	0	0	612	365	215	149	97	59	16	7	15	43	1 578	34.5
三碱基重复	0	1 103	488	232	64	48	48	15	5	7	5	6	2 021	44.1
四碱基重复	311	79	23	7	2	0	0	0	0	0	0	0	422	9.2
五碱基重复	115	37	5	0	1	0	0	0	0	0	0	1	158	3.5
六碱基重复	272	91	13	16	4	2	1	1	0	0	0	0	400	8.7
合计	698	1 310	1 141	620	286	199	146	75	21	14	20	49	4 579	100.0

3 讨论

基于高通量测序技术的转录组研究, 通过对样本的总 RNA 进行测序, 可以获得转录本丰度、可变剪接等全局信息, 结果准确、分析可靠且重复性较高, 因此有助于从整体水平上揭示生物体生长发育、次生代谢及生理适应及转录调控规律, 在本草基因组学研究方面应用非常广泛也并取得重大进展^[8]。本研究首次采用 Illumina HiSeq™ 2000 150PE 测序平台, 进行常用中药原植物白鲜的转录组测序分析, 测序质量良好、质控严格, *de novo* 组装获得 67 538 832 条高质量 reads 参与组装, 共得到 49 050 条 unigenes, 序列长度与 reads 覆盖区域对应合理。unigene 序列信息量庞大, 数据基本涵盖全转录组信

息, 能够反映秦岭特殊环境条件下白鲜的基因表达特征, 为下一步研究白鲜生长发育、次生代谢表达调控等生物学过程功能基因的发掘及遗传机制研究提供基础资料。

高通量转录组测序策略通常偶联生物信息学分析进行系统 reads 滤过、质控、组装, 基因注释和功能分类等。本研究利用 BLAST、Trinity、ESTscan 等多种软件, 对白鲜转录组 unigene 序列进行注释和功能分类。基于 BLAST 分析, 将所有 unigenes 与 NR、Swiss-port、KOG、KEGG 等 4 大数据库进行比对, 成功注释 31 955 条, 占全部序列的 65.15%, 17 095 条 unigenes 未获得注释, 这与已报道的人参^[9]、甘草^[10]和罗勒花^[11]等物种转录组测

序注释比例类似,说明基于高通量测序技术的药用植物转录组有大量序列需深入挖掘,白鲜转录组尚存在大量序列特征及功能未知的 *unigenes*。

GO 分类揭示白鲜根的转录组特性与生物过程、细胞组分和分子功能相关;KOG 功能分析从基因组水平寻找直系同源体,预测未知 ORF 的生物学功能,可大大提高基因功能注释的准确性^[11],本研究共得到 25 个不同的 KOG 类群,说明白鲜转录组 KOG 种类比较全面。进一步对白鲜功能基因序列进行 KEGG 代谢路径注释,发现 131 个 KEGG 标准代谢通路,这些基因可能参与白鲜水分吸收、矿质营养、光合作用和呼吸作用等生理代谢过程;还发现大量 *unigenes* 参与萜类、生物碱、黄酮类、花青素等生物合成相关的 16 个次生代谢标准通路。目前,罂粟科、茄科、小檗科及菊科等多种药用植物中关于莨菪烷、哌啶、吡啶、异喹啉或吡啶酮类生物碱生物合成通路及其调控机制有较大进展^[12]。本研究转录组数据中也发掘这些类型生物碱生物合成通路中的关键基因,印证了白鲜生物碱的多样性。然而,呋喃喹啉类生物碱 KEGG 标准通路未能得到注释,可能由于该类生物碱研究数据相对匮乏。尽管如此,这些基因信息为进一步白鲜生物碱生物合成通路解析及其调控研究提供重要线索。

转录因子的表达调控在植物细胞代谢过程中起决定作用。ERF、bHLH、MYB 和 WRKY 等家族在植物细胞甲羟戊酸、苯丙烷类代谢途径调控中起关键作用^[13]。本研究获得的白鲜 *unigenes* 转录因子家族覆盖最新版 PlantTFDB 4.0 数据库^[14]中的 55 个家族,与萜类、生物碱、黄酮类次级代谢物合成有关的转录因子大量表达,说明白鲜生命活动代谢涉及复杂的转录调控机制,有助于进一步深入研究。

转录组测序数据除了能够用于生物样本的基因挖掘及其表达调控研究,还可开发大量 EST-SSR,因此在遗传多样性、分子标记等方面应用广泛。本研究基于经典的 MISA 分析,发掘了白鲜根转录组 3 836 个 *unigenes* 中共计 4 579 个 SSRs 位点,SSRs 从双核苷酸类型到六核苷酸类型均具备,表明白鲜基因组内具有较高丰度的 SSRs。重复类型以三核苷酸为主,双核苷酸所占比例次之。这与以三核苷酸重复类型为主的主要作物水稻、大麦或棉花等的研究结果一致^[15-16]。白鲜转录组中 AG/CT 类型 SSRs 最多,三核苷酸重复中 AAG/CTT 类型最多,这与番红花^[17]、人参^[18]等药用植物中双核苷酸重复

SSRs 为主要类型的情况一致。可见大多数植物 SSRs 重复主要以双核苷酸和三核苷酸为主,但不同物种的重复序列存在差别。

基于高通量测序的白鲜转录组学研究初步获得大量重要数据信息,后续通过系统分析,以期全面解析以白鲜碱、黄柏酮等为主要活性成分的生物合成通路及其调控机制,揭示该药用植物的群体遗传结构特征,为白鲜药用资源的可持续开发提供理论基础。

参考文献

- [1] 南京中医药大学. 中药大辞典 (上、下册) [M]. 上海: 上海科学技术出版社, 2006.
- [2] Lv M, Xu P, Tian Y, *et al.* Medicinal uses, phytochemistry and pharmacology of the genus *Dictamnus* (Rutaceae) [J]. *J Ethnopharmacol*, 2015, 171: 247-263.
- [3] Kim H, Kim M, Kim H, *et al.* Anti-inflammatory activities of *Dictamnus dasycarpus* Turcz. root bark on allergic contact dermatitis induced by dinitrofluorobenzene in mice [J]. *J Ethnopharmacol*, 2013, 149(2): 471-477.
- [4] Gilsaeng J, Byun E, Li B, *et al.* Neuroprotective effects of constituents of the root bark of *Dictamnus dasycarpus* in mouse hippocampal cells [J]. *Arch Pharmacol Res*, 2014, 37(3): 421-421.
- [5] Soo P H, Hong N R, Seok A T, *et al.* Apoptosis of AGS human gastric adenocarcinoma cells by methanolic extract of *Dictamnus* [J]. *Pharmacog Mag*, 2015, 11(Suppl 2): S329-S336.
- [6] 毛少利, 周亚福, 王宇超, 等. 白鲜营养器官解剖结构及其与白鲜碱的积累关系 [J]. 西北植物学报, 2015, 35(6): 1135-1141.
- [7] 赵 静, 庞其昌, 马 骥, 等. 光谱成像技术快速鉴别白鲜皮及其伪品八角枫 [J]. 中国中药杂志, 2014, 32(8): 1835-1837.
- [8] 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.
- [9] Chen S, Luo H, Li Y, *et al.* 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng* [J]. *Plant Cell Rep*, 2011, 30(9): 1593-1601.
- [10] Ramilowski J A, Sawai S, Seki H, *et al.* *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals [J]. *Plant Cell Physiol*, 2013, 54(5): 697-710.
- [11] 刘 雷, 赵 欢, 冉茂中, 等. 罗勒花和叶的转录组数据组装及基因功能注释 [J]. 中草药, 2017, 48(17): 3612-3618.

- [12] He S M, Song W L, Cong K, *et al.* Identification of candidate genes involved in isoquinoline alkaloids biosynthesis in *Dactylicapnos scandens* by transcriptome analysis [J]. *Sci Rep*, 2017, 7(1): 9119.
- [13] Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants [J]. *Mol Plant*, 2015, 8(5): 689-708.
- [14] Jin J P, Tian F, Yang D C, *et al.* PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants [J]. *Nucleic Acids Res*, 2017, 45(D1): D1040-D1045.
- [15] Cardle L, Ramsay L, Milbourne D, *et al.* Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. *Genetics*, 2000, 156(2): 847-854.
- [16] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: Features and applications [J]. *Trends Biotechnol*, 2005, 23(1): 48-55.
- [17] 陈国庆. 番红花 EST 资源的 SSR 信息分析 [J]. 广西植物, 2011, 31(1): 43-46.
- [18] Li C, Zhu Y, Guo X, *et al.* Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer [J]. *BMC Genom*, 2013, 14: 245.