基于高通量测序技术的黄三七根茎转录组数据分析

李依民[1]，彭 亮[1]，杨冰月[1]，张明英[1]，任 瀛[1]，程虎印[1]，吴海峰[2*]，张 岗[1, 2*]

1. 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
2. 中国医学科学院北京协和医学院 药用植物研究所，北京 100193

摘 要：目的 获得黄三七 *Souliea vaginata* 根茎转录组信息特征。**方法** 以黄三七根茎为对象，采用二代高通量测序平台 Illumina HiSeq™ 2000 150PE 进行转录组测序并进行系统的生物信息学分析。**结果** 转录组测序分析共获得 63 322 086 条高质量序列（clean reads），Trinity *de novo* 组装获得 52 575 个 unigenes，平均长度 909 nt。BLAST 分析显示分别有 28 842（54.86%）、10 712（20.37%）、9 245（17.58%）、11 559（21.99%）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 45 分支，涉及 126 个 KEGG 标准代谢通路，其中包括 17 个次生代谢标准通路。蛋白编码框序列 2 215 个，包含高等植物转录因子 55 个家族；借助 MISA 软件发现 4 609 个 SSRs，三碱基重复 SSRs 数量最丰富，有 2 106 个，出现频率为 45.7%，五碱基重复 SSRs 相对较少，占 2.9%。**结论** 利用高通量测序技术和生物信息分析获得黄三七根茎转录组信息特征，为后期黄三七功能基因鉴定、次生代谢途径解析及其调控机制研究奠定基础。

关键词：黄三七；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- 4983 - 08

DOI: 10.7501/j.issn.0253-2670.2018.21.005

Transcriptomic data analyses of rhizome of *Souliea vaginata* via Illumina high-throughput sequencing technology

LI Yi-min[1], PENG Liang[1], YANG Bing-yue[1], ZHANG Ming-ying[1], REN Ying[1], CHENG Hu-yin[1], WU Hai-feng[2], ZHANG Gang[1, 2]

1. Shaanxi Provincial Key Laboratory for Chinese Medicine Basis & New Drugs Research, College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an 712046, China
2. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China

Abstract: Objective To obtain the transcriptome dataset of rhizome of *Souliea vaginata* **Methods** Using the high-throughput illumina sequencing platform Illumina HiSeq™ 2000 150PE, a rhizome transcriptome dataset of *S. vaginata* was obtained, followed by systemic bioinformatics analyses. **Results** The transcriptome sequencing analyses produced to a great number of 63 322 086 high quality clean reads. Trinity *de novo* assembling resulted in a total of 52 575 unigenes with an average length of 909 nt. BLAST analysis indicated that 28 842 (accounting for 54.86% of the total unigenes), 10 712 (20.37%), 9 245 (17.58%), and 11 559 (21.99%) unigenes were successfully annotated in the NR, Swiss-port, KOG, and KEGG databases, respectively. GO classification contained the basic three major groups, including biological process, cellular component, and molecular function, and 45 subgroups. Among them126 KEGG standard pathways were designated, of which 17 were defied as the secondary metabolism. Of all unigenes, 2 215 with protein coding sequences were predicted, and 55 families of plant transcription factors were also identified. MISA prediction yielded a number of 4 609 simple sequence repeats (SSRs), among which the tri-nucleotide SSRs were abundant with 2 106 (45.7%), whereas the penta-nucleotide SSRs were relatively less, accounting for 2.9%. **Conclusion** The transcriptomic characteristics of *S. vaginata* rhizome were revealed by the high-throughput Illumina sequencing technology along with bioinformatics analyses, which

收稿日期：2018-03-15

基金项目：陕西省自然科学基金项目（2017JM8030）；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）；陕西省高校首批青年杰出人才支持计划项目；中国医学科学院医学与健康科技创新重大协同创新项目（2017-I2M-1-013）；咸阳市中青年科技领军人才项目；国家留学基金委浙江省药学重中之重开放课题（YKFJ3-012）

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

吴海峰，男，博士，副研究员，研究方向为天然药物活性成分发现及作用机制。Tel/Fax: (010)57833250 E-mail: hfwu@implad.ac.cn

would be of great importance for the functional gene characterization, secondary metabolism pathway dissections, and their regulatory mechanisms in *S. vaginata*.

Key words: *Souliea vaginata* (Maxim.) Franch.; transcriptome; functional gene; metabolism pathway; simple sequence repeats

黄三七 *Souliea vaginata* (Maxim.) Franch. 为毛茛科升麻族黄三七属多年生草本植物，别名长果升麻、太白黄连或土黄连，单属单种，主要分布于我国陕西、甘肃、四川、云南和西藏等地。黄三七根茎或全草供药用，性苦、凉，具清热除烦、解毒消肿之功效，主治热病烦躁、心悸怔忡、骨蒸潮热、咽炎、口腔炎、结膜炎、疮痈肿毒、湿热泄泻、痢疾[1]，在陕西太白七药中具有重要地位。已报道黄三七中分离得到三萜皂苷、生物碱、有机酸等多种成分[2]。现代药理研究表明，环阿屯烷型（又称环菠萝蜜烷或环阿尔廷烷型）三萜皂苷为黄三七主要活性成分[3]，也是升麻族铁破锣属和升麻属等属的特征性成分，具有解热、镇痛、抗炎、抗风湿、抗肿瘤、抑制核苷转运和抗骨质疏松等多种药效[4-5]，可为临床药物研发提供丰富的前体化合物。因此，黄三七的基础研究及资源开发具有重要意义和极好的发展前景。

转录组测序作为功能基因组研究的一个重要组成部分，能够在整体水平上获得特定条件下细胞中所有基因转录本全局信息，有助于揭示生物体生长发育、次生代谢及生理适应分子机制及转录调控规律[6]。近年来，基于高通量测序技术的转录组分析策略在药用植物功能基因组领域内应用十分广泛，已经获得西洋参[7]、人参[8]、柴胡[9]、甘草[10]和膜荚黄芪[11]等众多药用植物转录组数据，为阐明中草药种质资源遗传基础奠定重要基础。本研究利用二代高通量测序平台 Illumina HiSeq™ 2000 150PE 进行黄三七根茎转录组测序分析，以期揭示黄三七根茎转录组的整体表达特征，为该药用植物功能基因鉴定、次生代谢途径解析及其调控机制研究提供基础数据。

1 材料与方法

1.1 材料

植物材料于 2015 年 7 月采自陕西省宝鸡市太白县秦岭鳌山咀头镇鳌山北坡，经度 107°23′27.7″，纬度 34°00′11.5″，海拔 2 556 m，经陕西中医药大学药学院张岗教授鉴定为毛茛科植物黄三七 *Souliea vaginata* (Maxim.) Franch.。取单株植株根茎液氮速冻后置于−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）制备黄三七根茎总 RNA，琼脂糖凝胶电泳和 NanoDrop™ 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（fragmentation buffer）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 HiSeq™ 2000 150PE 对黄三七根茎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、滤过除杂和冗余处理等质控得到高质量序列（clean reads），再利用组装软件 Trinity[12]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 的组装片段作为组装出来的 unigene。

1.4 转录组功能注释

利用 BLAST 将 unigenes 序列与蛋白数据库 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* 值＜1×10^{-5}），得到与相应 unigenes 具有最高序列相似性的蛋白，进而得到 unigenes 注释信息。根据 NR 注释信息，使用 Blast2GO 软件得到 unigenes 的基因本体（gene ontology，GO）注释，用 WEGO 软件对所有 unigenes 做 GO 功能分类统计，从宏观上认识该物种的基因功能分布特征。

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 unigenes 与以上蛋白库做 BLASTx 比对（*E* 值＜1×10^{-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[13]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

使用 MISA 检测黄三七转录组 unigenes，搜索 SSRs 并进行统计分析。

2 结果与分析

2.1 黄三七转录组组装与质量分析

采用 Illumina HiSeq™ 2000 100PE 高通量测序平台对黄三七根茎进行转录组测序，共得到 65 137 148 条 raw reads，过滤产生了 63 322 086 条 clean reads，包含 6 320 552 293 个核苷酸信息，Q20（碱基≥20%）和 Q30（碱基量≥30%）分别为 98.82%、94.18%，GC 量为 43.88%，说明测序质控良好，clean reads 质量合格。Trinity 组装获得 52 575 个 unigene，平均长度 909 nt，最长达到 12 082 nt，最短序列为 201 bp，N50 为 1 583 nt。unigenes 长度分布（图 1-A）显示，16 324 条 unigenes 长度超过 1 000 nt，5 988 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 27 631 条；其次为 reads 数量在 1～10 的 unigenes，为 6 771 条；reads 数量在 101～200、1 001～2 000、2 001～4 000 的 unigenes 分别为 3 664、2 726、2 104 条；其余 reads 分布区域对应的 unigenes 数量均相对较少。

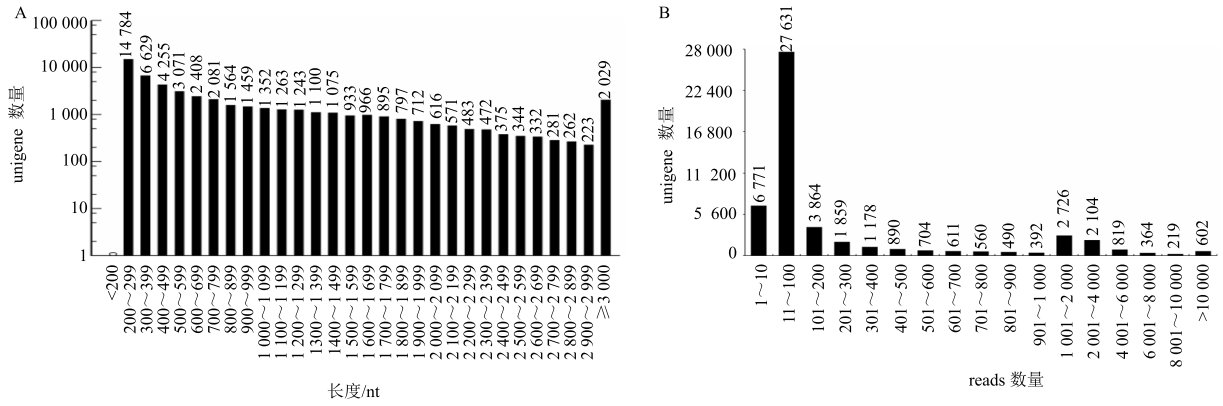

图 1 黄三七转录组 unigenes 长度分布（A）和 reads 覆盖统计（B）

Fig. 1 Length distribution (A) and reads coverage statistics (B) of *S. vaginata* transcriptomic unigenes

2.2 黄三七转录组 unigenes 的功能注释

使用 BLAST 将所有 unigenes 与 NR、Swiss-port、KOG、KEGG 等数据库进行一致性比对分析，对各数据库注释的 unigenes 数目进行统计，进而获得黄三七根茎转录组 unigenes 的功能注释信息。结果表明，28 842 条 unigenes（54.86%）在 NR 数据库比对成功得到注释，在 Swiss-port、KOG、KEGG 等数据库获得注释的 unigenes 数目依次为 10 712（20.37%）、9 245（17.58%）、11 559（21.99%）。4 735 条 unigenes 同时在所有数据库中注释，至少有一种数据库注释成功的 unigenes 共 28 887 条（54.94%），23 688 条 unigenes 未获得注释。

以 NR 数据库为例进行分析，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，莲 *Nelumbo nucifera* Gaertn. 所占比例最高，9 437 条（32.72%）；其次为葡萄 *Vitis vinifera* L. 2 325 条（8.06%），土瓶草 *Cephalotus follicularis* Labill. 1 265 条（4.39%），可可 *Theobroma cacao* L. 1 183 条（5.0%），其余匹配物种比例在 1.42%～2.82%，比例小于 1.42%的匹配物种的 unigenes 为 11 289，占 39.14%。

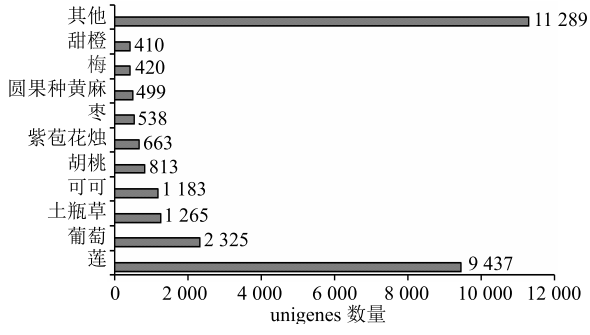

图 2 黄三七转录组 unigenes 与 NR 数据库匹配物种分布

Fig. 2 Species distribution of *S. vaginata* transcriptomic unigenes against NR database

根据 NR 注释信息得到 GO 分类（图 3），9 993 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 45 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，达 3 358 和 3 357 条；其次是细胞器（organelle），有 2 556 条；病毒粒子（virion）、病毒粒子组成（virion part）等基因较少，在 100 条以下。生物过程主要聚集在细胞过程（cellular process）和代谢过程（metabolic process），涉及的基因分别有 10 952 和 9 531 条；应激适应（response to stimulus）、着色（pigmentation）、生物调控（biological regulation）基因数量分别为 3 429、3 111、3 575 条。分子功能中具有催化活性（catalytic activity）和结合功能的（binding）基因数量较高，分别为 5 778 和 4 092 条，其他类别基因数目普遍较少。

为了进一步分析黄三七转录组 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 3 389 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 1 906 和 1 463 条；加工和修饰 unigenes 数目 1 052 条；其他种类基因丰度不尽相同（图 4）。

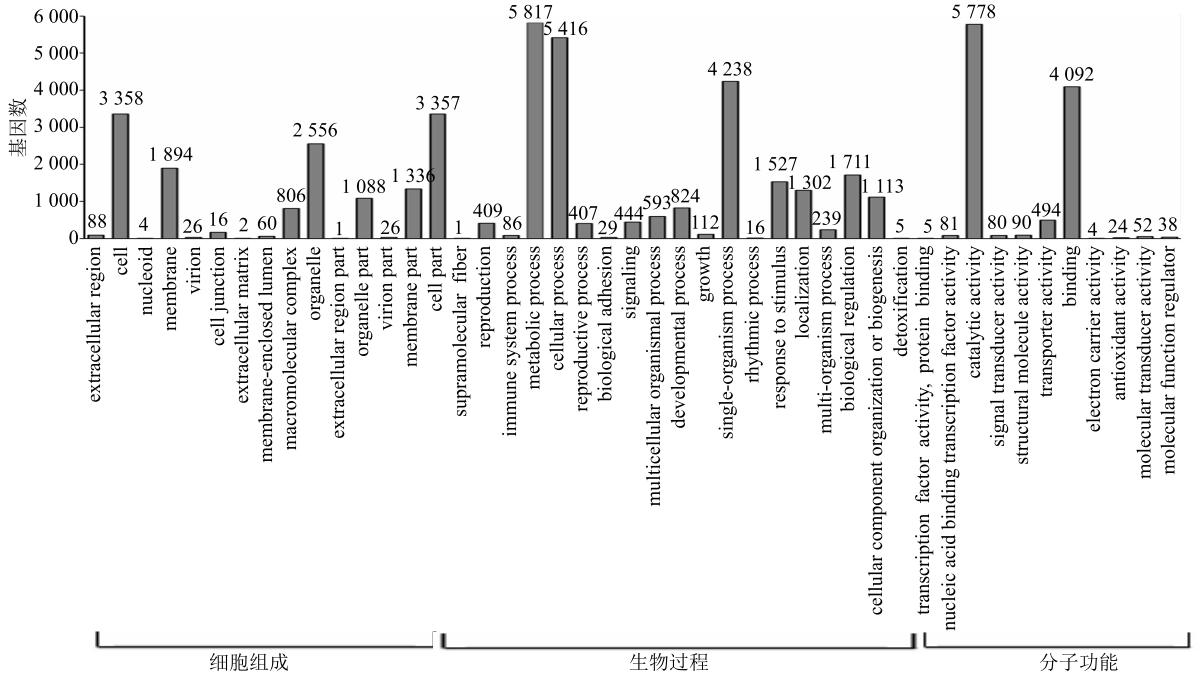

图 3　黄三七根茎转录组 unigenes 的 GO 分类

Fig. 3　GO classification of *S. vaginata* transcriptomic unigenes

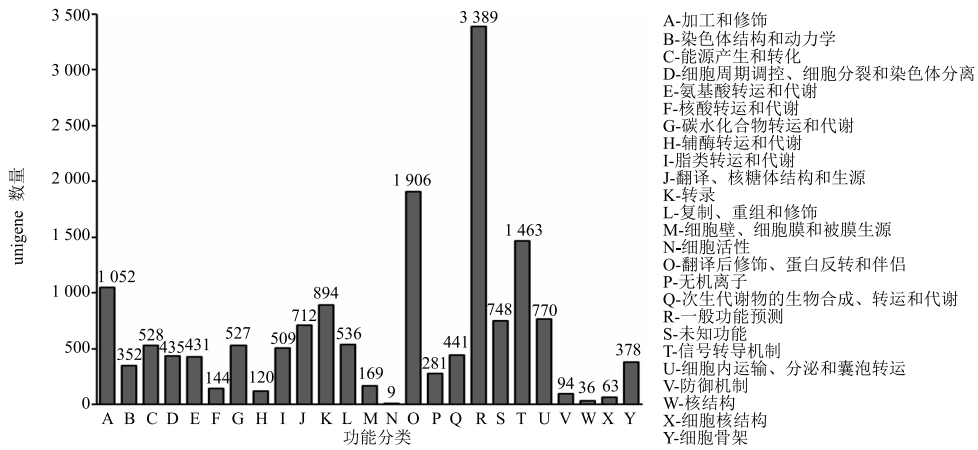

图 4　黄三七转录组 unigenes 的 KOG 注释分布

Fig. 4　KOG annotation distribution of *S. vaginata* transcriptomic unigenes

黄三七根茎转录组 unigenes 参与 KEGG 代谢通路分为 5 大分支：细胞过程（cellular processes）498 条、环境信息处理（environmental information processing）372 条、遗传信息处理（genetic information processing）2 793 条、代谢（metabolism）5 668 条和有机系统（organismal systems）294 条。6 041 条 unigenes 获得 126 个 KEGG 标准代谢通路，按照基因注释量大小依次排序，选取前 12 个代谢通路信息见表 1，这些通路包含的 unigenes 数量大于 200 条。

KEGG 代谢通路分析还发现 513 条 unigenes 参与黄三七萜类、甾类、生物碱、黄酮类、花青素等生物合成相关的 17 个次生代谢标准通路（表 2）。其中，苯丙素的生物合成代谢通路（ko00940）基因数量最多，为 124 个；萜类化合物骨架生物合成（ko00900）基因数量次之，为 78 条；与类胡萝卜素生物合成（ko00906）有关的基因有 34 条；分别有 38、33 个 unigenes 与莨菪烷类、哌啶、吡啶生物碱（ko00960）及异喹啉类生物碱生物合成相关（ko00950）；24 条 unigenes 参与二萜类生物合成（ko00904）；倍半萜和三萜类化合物的生物合成（ko00909）基因有 10 条；咖啡因的代谢，花青素、芥子油苷以及黄酮和黄酮醇的生物合成通路基因数较少，均在 5 条以下。

表 1 黄三七转录组 unigenes KEGG 通路分析统计

Table 1 KEGG pathway analysis of *S. vaginata* transcriptomic unigenes

编号	代谢通路	unigenes 数量	占比/%	通路 ID
1	核糖体	397	6.57	ko03010
2	碳水化合物代谢	338	5.60	ko01200
3	剪接体	325	5.38	ko03040
4	氨基酸生物合成	302	5.00	ko01230
5	内质网蛋白质加工	290	4.80	ko04141
6	RNA 转运	251	4.15	ko03103
7	嘌呤代谢	245	4.06	ko00230
8	植物-病原菌互作	243	4.02	ko04626
9	淀粉和蔗糖代谢	224	3.71	ko00500
10	植物激素信号转导	218	3.61	ko04075
11	内吞作用	208	3.44	ko04144
12	泛素介导的蛋白质降解	202	3.34	ko04120

2.3 CDS 和转录因子分析

对黄三七转录组所有 unigenes 的 CDS 进行分析，通过 BLAST 比对共获得 CDS 序列 28 584 个，利用 ESTscan 数据库分析获得 CDS 序列 2 215 个。转录因子预测发现有 55 个家族成员，其中 bHLH、ERF、C2H2、bZIP、NAC、FAR1、MYB 及 WRKY 类占主体，说明黄三七根茎生理代谢涉及众多转录调控过程（图 5）。

表 2 黄三七转录组 unigenes 次生代谢 KEGG 通路注释统计

Table 2 Secondary metabolism KEGG pathway annotation analysis of *S. vaginata* transcriptomic unigenes

编号	代谢通路	unigenes 数量	占比/%	通路 ID
1	苯丙素的生物合成	124	2.05	ko00940
2	萜类化合物骨架生物合成	78	1.29	ko00900
3	甾类化合物的合成及姜辣素	38	0.63	ko00945
4	莨菪烷类、哌啶、吡啶生物碱生物合成	37	0.61	ko00960
5	类胡萝卜素生物合成	34	0.56	ko00906
6	异喹啉类生物碱的生物合成	33	0.55	ko00950
7	玉米素的生物合成	32	0.50	ko00908
8	类黄酮生物合成	32	0.50	ko00941
9	柠檬烯和蒎烯降解	24	0.40	ko00903
10	二萜类生物合成	24	0.40	ko00904
11	油菜素内酯的生物合成	18	0.30	ko00905
12	单环 β-内酰胺的合成	13	0.22	ko00261
13	倍半萜和三萜类化合物的生物合成	10	0.17	ko00909
14	咖啡因的代谢	5	0.08	ko00232
15	花青素生物合成	5	0.08	ko00942
16	芥子油苷的生物合成	4	0.07	ko00966
17	黄酮和黄酮醇的生物合成	2	0.03	ko00944

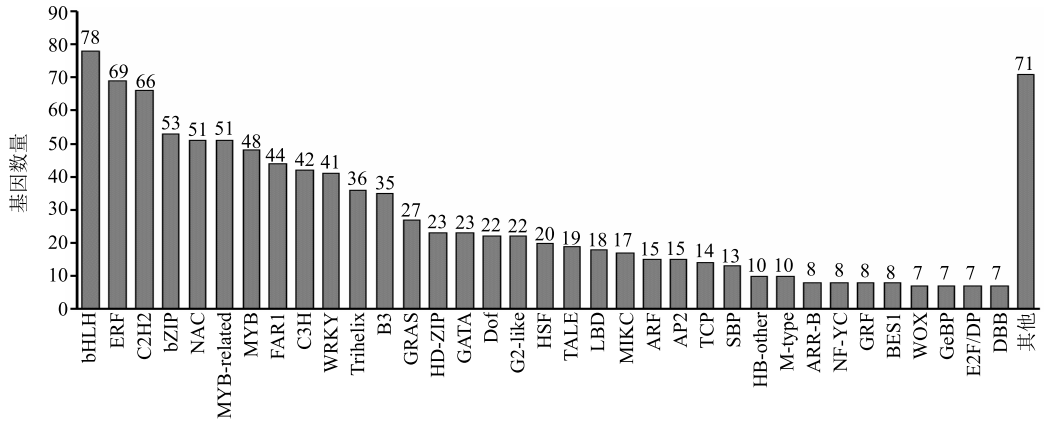

图 5　黄三七转录组 unigenes 的转录因子分析

Fig. 5　Transcription factor classification of *S. vaginata* transcriptomic 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），3 979 个 unigenes 中共计 4 609 个 SSRs。其中，三碱基重复 SSRs 数量最丰富，有 2 106 个（45.7%），在这之中 AAG/CTT 类型的比例最高。双碱基重复 SSRs 数量次之，有 1 669 个，占 SSRs 总量的 36.2%，其中 AG/CT 重复类型数量最多。四碱基和六碱基重复分别为 379、322 个，各占 8.2%、7.0%；五碱基重复重复相对较少，仅占 2.9%。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　黄三七转录组 unigenes SSRs 分析

Table 3　SSRs analysis of *S. vaginata* transcriptomic 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	654	322	236	176	113	57	26	4	16	65	1 669	36.2
三碱基重复	0	1 253	469	221	73	30	28	11	2	9	3	7	2 106	45.7
四碱基重复	285	54	26	8	5	1	0	0	0	0	0	0	379	8.2
五碱基重复	100	23	9	0	0	0	0	0	0	0	0	1	133	2.9
六碱基重复	226	65	6	10	10	4	0	0	1	0	0	0	322	7.0
合计	611	1 395	1 164	561	324	211	141	68	29	13	19	73	4 609	100.0

3　讨论

近年来，二代高通量测序技术在本草基因组及合成生物学等研究方面应用广泛，并取得重大进展[6]。本研究首次采用 Illumina HiSeq™ 2000 150PE 测序平台，进行秦岭特色中草药资源黄三七的转录组测序分析。黄三七根茎高通量测序数据约 6.3 G，测序质量良好、质控严格，利用 Trinity *de novo* 组装，93.9%的高质量 reads 参与组装，共得到 52 575 条 unigenes，序列长度与 reads 分布区域对应合理。转录组 unigenes 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黄三七的基因表达特征，为深入研究黄三七生长发育、次生代谢、转录调控等生物学过程功能基因的批量发掘提供数据资料。

基于高通量测序的转录组数据通常采用生物信息学分析策略进行基因注释和功能分类[6-7]。本研究利用 BLAST、Trinity[12]、ESTscan[13]等多种生物信息软件，对黄三七转录组 unigenes 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 大数据库比对，注释成功的 unigenes 共 28 887 条，占全部序列的 54.94%，其余 23 688 条 unigenes 并未获得注释，这与已报道的西洋参[7]、人参[8]、柴胡[9]、甘草[10]、珠子参[14]和罗勒花[15]等物种转录组测序注释比例类似，说明黄三七转录组中存在大量序列特征及功能尚未知的 unigenes。

GO 分类揭示黄三七根茎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[15]，本研究共得到 25 个不同的 KOG 类群，说明黄三七转录组 KOG 种类比较全面。进一步对黄三七功能基因序列进行 KEGG 代谢路径注释，发现 126 个 KEGG 标准代谢通路，这些基因可能参与黄三七水分吸收、矿质营养、光合作用和呼吸作用等生命代谢活动；此外，还发现大量 unigenes 参与萜类、甙类、生物碱、黄酮类、花青素等生物合成相关的 17 个次生代谢标准通路；其中，与倍半萜和三萜类化合物的生物合成相关基因有 10 条。西洋参[7]、人参[8]、珠子参[14]等珍稀名贵药用植物所含三萜皂苷主要以达玛烷型四环三萜和齐墩果烷型五环三萜为主，这与黄三七中三萜皂苷的类型不同，这些基因的发现为揭示黄三七环阿屯烷型三萜皂苷生物合成途径解析提供线索。

基因表达的转录调控在植物生长发育及环境适应方面发挥重要作用。最新版植物转录因子数据库 PlantTFDB 4.0[16]包含 58 个家族，其中 AP2/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[17]。如丹参中一个新的 R2R4-MYB 转录因子 SmMYB36 与某类 bHLH 转录因子互作共同调控丹参初生代谢和次生代谢[18]。本研究获得的黄三七 unigenes 转录因子家族覆盖 PlantTFDB 4.0 数据库中 55 个家族，说明黄三七生命活动代谢涉及复杂的转录调控机制；黄三七中与次生代谢调控密切相关的转录因子家族 unigenes 数量较多，有助于深入研究黄三七萜类、酚类及生物碱等各类活性物质生物合成的转录调控机制研究。

SSRs 包括 EST-SSR 和基因组 SSRs 2 种类型。除了具有基因组 SSRs 基本优点外，EST-SSR 兼有降低引物开发成本、提高测序数据利用率的特点，因此在作物中广泛用于遗传多样性、分子标记等研究[19-20]。本研究基于经典的 MISA 分析，发掘了黄三七根茎转录组 3 979 个 unigenes 的 4 609 个 SSRs 位点，SSRs 从双核苷酸类型到六核苷酸类型均具备，表明黄三七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[19-20]。黄三七双核苷酸重复 SSRs 中 AG/CT 类型最多，三核苷酸重复中 AAG/CTT 类型最多，这与罗勒花[15]、番红花[21]、人参[22]等植物中以 CT、AG 双核苷酸重复 SSRs 为主要类型的情况相同，但主要三碱基重复 SSRs 类型不一致。可见大多数植物 SSR 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。

目前，利用二代高通量测序技术对黄三七根茎转录组的研究还处于起步阶段，对转录组数据初步分析获得了萜类化合物生物合成途径的全部骨架基因，与环阿屯醇烷型三萜皂苷氧化衍生化修饰相关的细胞色素氧化酶 unigenes 有 169 条，其中 62 个含有完整的 ORF，对这些基因的表达调控研究将为解析环阿屯醇烷型三萜皂苷的生物合成打下基础。转录组数据同时获得了丰富的 SSRs 信息，为研究该物种单属单种的遗传特征提供依据。后续将对黄三七转录组数据做进一步的系统分析，通过解析黄三七三萜皂苷生物合成通路、调控、遗传结构等，以便更好地阐释其生长发育及生理适应等科学问题，也为黄三七药用资源的开发和利用提供理论基础。

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] 周 亮, 杨峻山, 涂光忠. 黄三七皂苷类化学成分的研究 [J]. 中国药学杂志, 2005, 40(18): 1375-1377.

[3] Wu H, Zhang G, Wu M, *et al*. A new cycloartane triterpene glycoside from *Souliea vaginata* [J]. *Nat Prod Res*, 2016, 30(20): 2316-2322.

[4] 李延勋, 栗章彭, 苏艳芳. 裂环环阿屯烷型三萜的研究进展 [J]. 中草药, 2017, 48(15): 3198-3209.

[5] Sun H, Liu B, Hu J, *et al*. Novel cycloartane triterpenoid from *Cimicifuga foetida* (Shengma) induces mitochondrial apoptosis via inhibiting Raf/MEK/ERK pathway and Akt phosphorylation in human breast carcinoma MCF-7 cells [J]. *Chin Med*, 2016, 11(1): 1-11.

[6] 陈士林, 朱孝轩, 李春芳, 等. 中药基因组学与合成生物学 [J]. 药学学报, 2012, 47(8): 1070-1078.

[7] 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: 262-269.

[8] 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.

[9] Sui C, Zhang J, Wei J, *et al*. Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins [J]. *BMC Genom*, 2011, 12(1): 539-559.

[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] Liu X B, Ma L, Zhang A H, *et al*. High-throughput analysis and characterization of *Astragalus membranaceus* transcriptome using 454 GS FLX [J]. *PLoS One*, 2014, 9(5): e95831.

[12] Grabherr M G, Hassour M. Full-length transcriptome assembly from RNA-seq data without a reference genome [J]. *Nat Biotechnol*, 2011, 29(7): 644-670.

[13] Iseli C, Jongeneel C V, Bucher P. ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences [J]. *Proc Int Conf Intell Syst Mol Biol*, 1999, 36: 138-148.

[14] 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.

[15] 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.

[16] Jin J P, Tian F, Yang D C, *et al*. Plant TFDB 4. 0: Toward a central hub for transcription factors and regulatory interactions in plants [J]. *Nucleic Acids Res*, 2017, 45(D1): D1040-D1045.

[17] Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants [J]. *Mol Plant*, 2015, 8(5): 689-708.

[18] Ding K, Pei T, Bai Z, *et al*. SmMYB36, a novel R2R3-MYB transcription factor, enhances tanshinone accumulation and decreases phenolic acid content in *Salvia miltiorrhiza* hairy roots [J]. *Sci Rep*, 2017, 7(1): 5104-5111.

[19] 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.

[20] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: Features and applications [J]. *Trends Biotechnol*, 2005, 23(1): 48-55.

[21] 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.

[22] 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 Genomics*, 2013, 14: 245.