

• 药材与资源 •

基于转录组测序的越南安息香根、茎和叶基因表达分析

王小敏¹, 许姝姝², 詹若挺¹, 马新业^{1*}

1. 广州中医药大学 中药资源科学与工程研究中心, 岭南中药资源教育部重点实验室, 国家中成药工程技术研究中心 南药研发实验室, 广东 广州 510006
2. 广州中医药大学中药学院, 广东 广州 510006

摘要: 目的 对越南安息香 *Styrax tonkinensis* 进行转录组测序, 获得其根、茎和叶的转录组信息特征。方法 以越南安息香的根、茎和叶作为研究对象, 使用 Illumina HiSeq™ 2000 进行越南安息香根、茎和叶的转录组测序分析。结果 转录组测序根、茎和叶共获得 53 835 045 条高质量序列 (clean reads), Trinity de novo 组装获得 69 151 条 Unigenes, 平均长度 778.51 nt。BLAST 分析表明分别有 41 412 (59.89%)、31 189 (45.10%)、25 539 (36.93%)、16 749 (24.22%) 个 Unigenes 在 Nr、Swiss-port、KOG、KEGG 数据库中得到注释, 可归入 GO 分类的细胞组分、生物过程和分子功能 3 大类 46 分支, 涉及 129 条 KEGG 标准代谢通路, 其中有 31 个次生代谢标准通路。蛋白编码框序列 3 461 个, 涉及高等植物转录因子 54 个家族。使用 MISA 软件挖掘 10 974 个简单重复序列 (SSRs), 二碱基重复 SSRs 数量最为丰富, 有 6282 个 (57.24%), 五碱基重复 SSRs 最少, 占 2.45%。结论 利用高通量技术和生物信息分析获得了越南安息香根、茎和叶的转录组信息特征, 为后期越南安息香基因功能鉴定、次生代谢途径解析及调控机制的研究奠定基础。

关键词: 越南安息香; 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(2021)08-2392-08

DOI: 10.7501/j.issn.0253-2670.2021.08.023

Transcriptomic data analysis of roots, stems, and leaves of *Styrax tonkinensis*WANG Xiao-min¹, XU Hua-hua², ZHAN Ruo-ting¹, MA Xin-ye¹

1. Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education, Research Center of Chinese Herbal Resource Science and Engineering, South Medicine Research and Development Laboratory of National Engineering Research Center for Chinese Patent Medicine, Guangzhou University of Chinese Medicine, Guangzhou 510006, China
2. School of Chinese Materia Medica, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

Abstract: Objective To obtain the transcriptome information characteristics of roots, stems, and leaves by transcriptome sequencing of *Styrax tonkinensis*. **Methods** The roots, stems, and leaves of *S. tonkinensis* were selected as the research objects, and Illumina HiSeq™ 2000 was used to carry out the transcriptome sequencing analysis of these roots, stems, and leaves. **Results** A total of 53 835 045 high-quality sequences (clean reads) were obtained from the roots, stems and leaves by transcriptome sequencing, and 69 151 Unigenes were assembled by Trinity de novo, with an average length of 778 nt. BLAST analysis showed that 41 412 (59.89%), 31 189 (45.10%), 25 539 (36.93%), and 16 749 (24.22%) Unigenes were annotated in the Nr, Swiss-port, KOG, and KEGG databases, respectively, which could be classified into 46 branches of cell components, biological processes and molecular functions of the three major classes in GO classification, involving 129 KEGG standard metabolic pathways, of which 31 had secondary metabolic pathways. There were 3 461 protein coding frame sequences, involving 54 families of higher plant transcription factors. MISA software was used to mine 10 974 simple repeat sequences (SSRs), in which the two-base repeat SSRs was the most abundant, with 6 282 (57.24%), and the five-base repeat SSRs were the least, accounting for 2.45%. **Conclusion** The transcriptome information characteristics of *S. tonkinensis* roots, stems, and leaves were obtained by high-throughput technology and

收稿日期: 2020-08-06

基金项目: 国家自然科学基金青年基金项目 (81102764); 广东省教育厅重点提升平台建设项目—岭南中药资源教育部重点实验室 (2014KTSPT016)

作者简介: 王小敏 (1990—), 男, 硕士研究生, 研究方向为芳香药用植物资源学。Tel: 19860209295 E-mail: 807548046@qq.com

*通信作者: 马新业 (1976—), 男, 博士, 副研究员, 主要从事中药资源学研究工作。Tel: 15817036306 E-mail: usermxy@163.com

bioinformatics analysis, which laid a foundation for the study of functional identification, secondary metabolic pathway analysis and regulation mechanism of *S.tonkinensis*.

Key words: *Styrax tonkinensis* (Pierre) Crail ex Har-tw.; transcriptome; functional genes; metabolism pathway; simple sequence repeats

越南安息香 *Styrax tonkinensis* (Pierre) Crail ex Har-tw. 为安息香科 (Styracaceae) 安息香属落叶乔木, 又名越南安息香、白背安息香和滇桂野茉莉, 俗称“东京野茉莉”, 广泛分布于我国岭南地区, 老挝、缅甸等国也有分布; 其茎皮受到伤害后的分泌物干燥后作为安息香^[1], 始载于《新修本草》, 味辛、苦, 平, 无毒; 主心腹恶气鬼疰。安息香的现代药理学认为安息香具有抗细菌^[2]、抗真菌^[3]、抗补体^[4]、抗氧化^[5]、抗白血病^[6]、抗肿瘤^[7]等药理活性。现代中药学揭示安息香含有香脂酸类成分, 包括苯甲酸、松柏树乙醚、香草醛、苯甲酸苄酯、2-丙炔酮^[8], 此外, 安息香含有三萜类成分, 包含苏门答刺树脂酸、泰安息香树脂酸、齐墩果酸^[6], 可为药物研发提供丰富的化合物前体。因此, 安息香的资源开发和基础研究具有重要的价值和前景。

转录组测序对于没有基因组的物种来说, 可以有效地表征和鉴定植物中次生代谢物质的生物合成相关途径, 揭示植物体的生长发育、生理适应性以及探索其中的基因序列和表达水平^[9-11]。近年来, Illumina 测序技术已广泛应用于研究植物基因组, 并且也应用在番茄^[12]、人参^[13]、扁豆^[14]等物种中, 为其种植资源和基因遗传打下基础。本研究选择越南安息香幼苗的根、茎和叶作为基础材料, 使用 Illumina HiSeq-2000 采集 RNA-seq 读数, 以期挖掘安息香根、茎和叶的整体基因表达特征, 为该植物基因功能研究、次生代谢途径解析和调控探究提供数据基础。

1 材料与方法

1.1 RNA 提取与测序

植物材料于 2015 年 10 月采自广州中医药大学, 经广州中医药大学马新业副研究员鉴定为越南安息香 *S. tonkinensis* (Pierre) Crail ex Har-tw.。取单株植物根、茎、叶液氮冷冻后迅速存于 -80 °C 冰箱至使用。

越南安息香幼苗根、茎和叶的总 RNA 分离后, 用含有 Oligo(dT) 的磁珠富集 mRNA, 加入 fragmentation buffer 使 mRNA 成为短片段, 合成 cDNA 第 1 条链使用的是六碱基随机引物 (random hexamers), 使用添加 EB 缓冲液的 QiaQuick PCR 试剂盒纯化并洗脱, 经末端修复、加测序接头、加 poly(A), 大小片段的回收使用琼脂糖凝胶电泳, 最

后通过 PCR 扩增完成 cDNA 文库制备, 由广州基迪奥生物科技有限公司 (广州) 使用 Illumina HiSeq™ 2000 进行测序。

1.2 序列的从头组装和功能注释

采用 FastQC 软件 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) 评估原始数据的质量, NGS QC Toolkit (v2.3.3) 软件舍弃低质量读取 (<http://59.163.192.90:8080/ngsqctoolkit/>)。使用 Trinitysoftware^[11] 对 RNA-Seq 从头组装获得 Unigene, 其总体表达量使用 RPKM 法^[15]计算。

为了注释转录组, 利用 BLAST 将 Unigenes 比对到蛋白数据库 Nr、Swiss-prot、蛋白相邻类的聚簇 (KOG) 和京都基因与基因百科全书 (kyoto encyclopedia of genes and genomes, KEGG) (E 值 $< 1 \times 10^{-5}$), 获得与相应 Unigenes 具有最高相似性的蛋白, 从而得到相应的注释, 使用 Blast2GO 软件对 Unigenes 进行 GO (gene ontology) 分类, 用 WEGO 软件对获得的 Unigenes 进行 GO 分类, 宏观上认识该物种基因功能的分布情况。

1.3 蛋白编码框 (sequence coding for aminoacids in protein, CDS) 及转录因子的鉴定

按照 Nr、Swiss-Prot、KOG 和 KEGG 优先顺序把 Unigenes 序列和以上数据库进行 BLASTx 比对 ($E < 1 \times 10^{-5}$), 并确定该 Unigene 编码区的核苷酸序列 (序列方向 5' 至 3' 端) 及氨基酸序列。使用 ESTScan 预测与以上数据库对比不上的 Unigenes 的编码区和序列方向。将预测到的 Unigenes 编码蛋白序列和植物转录因子数据库 plantTFDB 进行 hmmscan 比对, 获得转录因子家族和相关成员。

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

MISA 软件被用到越南安息香转录组 Unigenes 的检测, 搜索 SSRs 同时进行统计分析。

2 结果与分析

2.1 转录组组装和质量评估

通过 Illumina HiSeq™ 2000 测序, 越南安息香的根、茎和叶分别获得 65 475 824、66 555 768、56 770 948 条 raw reads, 滤过后获得 64 267 196 (98.15%)、65 122 652 (97.85%)、55 714 504 (98.14%) 条 clean reads, 分别包含滤过后核苷酸

8 033 399 500、8 140 331 500 和 6 964 313 000 个。Q20 (碱基量 $\geq 20\%$) 均大于 98%和 Q30 (碱基量 $\geq 30\%$) 均 $>96\%$, 说明测序数据控制良好, clean reads 质量合格。Trinity 组装得到 69 151 条 Unigenes, 平均长度 778.51 nt, 最长达 10 486 nt, 最短为 201 nt, N50 等于 1333。

2.2 Unigenes 功能注释

用 BLAST 将所获得的 Unigenes 比对到 Nr、Swiss-port、KOG、KEGG 数据库, 并统计所注释到的 Unigenes 数目和功能信息。结果显示, 41 412 (59.89%) 条 Unigenes 在 Nr 数据库得到注释, 在 Swiss-prot、KOG 和 KEGG 数据库分别注释了 31 189

(74.3%)、25 539 (60.9%)、16 749 (39.9%) 条 Unigenes。成功注释的 Unigenes 共 41 951 条, 占 66.67%, 27 200 (39.33%) 条未得到注释。Unigenes 长度分布见图 1-A, 17 387 条 Unigenes 长度大于等于 1000 nt, 5674 条 Unigenes 超过 2000 nt。图 1-B 展示的是 reads 在 Unigene 上的覆盖度, reads 在 11~100 的 Unigenes 数量最多, 有 27 240 条; 其次是 reads 在 1~10 的 Unigenes, 有 5150 条; reads 在 101~200、2001~4000、1001~2000、 $>10\ 000$ 的数量分别是 4669、4226、3935 和 3186 条; 其他 reads 在 Unigenes 上的覆盖度均相对较小。

以注释到 Nr 数据库的 Unigenes 为例, 同源物

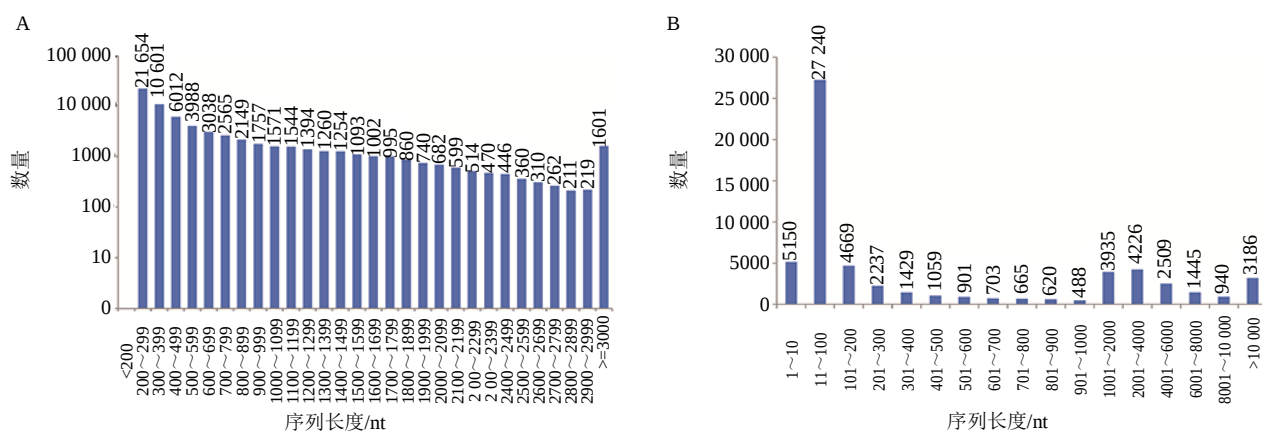


图 1 Unigenes 的长度分布 (A) 和 reads 的覆盖范围 (B)

Fig. 1 Length distribution of unigenes (A) and coverage of reads(B)

种如图 2, 在匹配度最高的物种中, 葡萄 *Vitis vinifera* L. 占的比率为最高, 达 4160 (10.05%) 条; 其次为欧洲油菜 *Brassica napus* L. 1889 (4.56%), 黄豆 *Glycine max* (Linn.) Merr. 1630 (3.94%), 毛果杨 *Populus trichocarpa* L. 1615 (3.90%), 蒺藜苜蓿 *Medicago truncatula* L. 1614 (3.90%), 拟南芥 *Arabidopsis thaliana* L. 1577 (3.81%), 粳稻 *Oryza sativa Japonica* Group 1563 (3.77%), 可可树 *Theobroma cacao* L. 1537 (3.71%), 番茄 *Solanum lycopersicum* L. 1400 (3.38%), 苹果 *Malus domestica* B. 1262 (3.05%); 匹配度在 3.04%~1.46% 的有 13 304 条, 小于 1.46% 匹配度的有 9 861 条, 占 23.81%。根据 Nr 进一步获得 GO 分类信息, 如图 3, 148 884 条 Unigenes 被注释到细胞组分、生物过程和分子功能 3 个 GO 类别的 46 个小组。生物过程中代谢过程 (metabolic process)、细胞过程 (cell process) 和多细胞生物过程 (multicellular organismal process) 丰度最高, 分别为 15 930、14 748 和 11 806 条; 生物调节 (biological regulation) 和应激适应

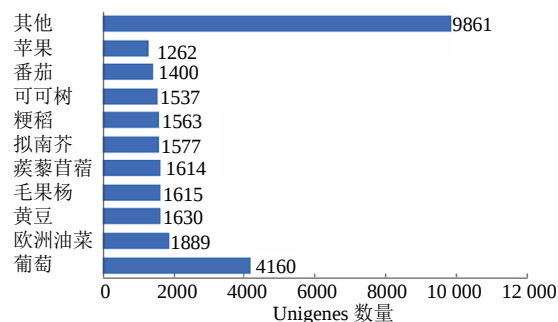


图 2 Unigenes 在 Nr 数据库中的物种匹配

Fig. 2 Species distribution of Unigenes against Nr database

(response to stimulus) 分别有 5087 和 4630 条。分子功能中催化活性 (catalytic activity) 和结合功能 (binding) 的基因最多, 分别达 15 160 和 14 029 条, 其他基因表达量都较少。细胞组分中细胞 (cell) 及细胞组分 (cell part) 较高, 分别达 10 915 和 10 914 条; 细胞器 (organelle)、细胞膜 (membrane) 分别有 8509 和 5938 条; 细胞膜组分 (membrane part)、器官组分 (organelle part) 和大分子复合物 (macromolecular complex) 分别有 4125、3713 和 2712 条, 其他基因均较小。

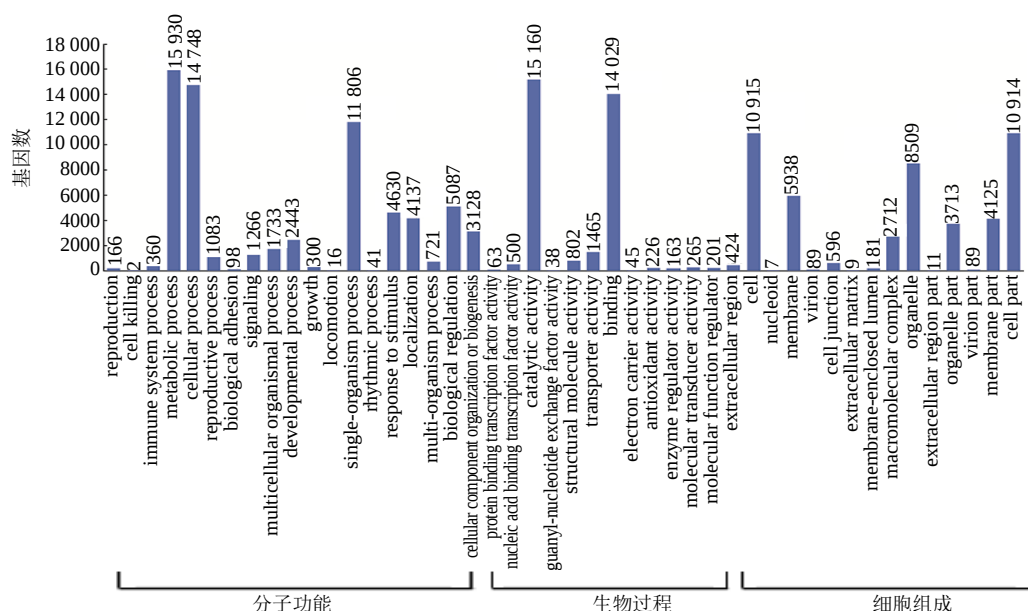


图 3 转录组 Unigenes 的 GO 分布

Fig. 3 GO classification of transcription Unigenes

为了进一步分析越南安息香的转录组 Unigenes 的功能,使用 KOG 功能分类分析(图 4),一共获得 25 种 KOG 分类,种类较全面,其中,一般功能预测的基因(general function prediction only)达 8 025 条,为最多;信号转导机制类(signal transduction mechanisms)、翻译后修饰、蛋白反转和伴侣(posttranslational modification, protein turnover, chaperones)次之,分别有 5082 和 4862 条;其他分类基因丰度参差不齐。

越南安息香根、茎和叶的转录组 Unigenes 参与 KEGG 代谢通路被分为 5 类:代谢(metabolism)8878 条、遗传信息处理(genetic information processing)3 936 条、细胞过程(cellular processes)753 条、有机系统(organismal systems)573 条和环境信息处理(environmental information processing)548 条。其中,7533 条 Unigenes 被归类于 94 个标准代谢通路,根据注释量选择前 9 个代谢通路展示见表 1,这些通路含有的 Unigenes 数量均不低于 200 条。

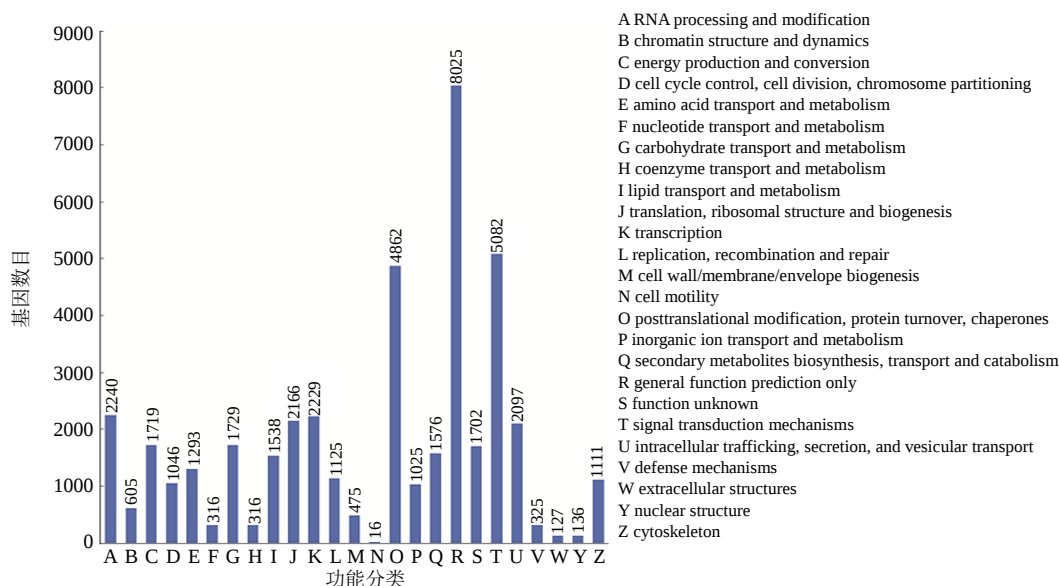


图 4 转录组 Unigenes 的 KOG 分布

Fig. 4 KOG classification of transcriptomic Unigenes

表 1 转录组 Unigenes 的 KEGG 通路统计分析

Table 1 KEGG pathway analysis of transcriptomic Unigenes

编号	代谢通路	Unigenes 数量	占比/%	通路 ID
1	oxidative phosphorylation	365	4.85	ko00190
2	starch and sucrose metabolism	310	4.12	ko00500
3	purine metabolism	285	3.78	ko00230
4	phenylpropanoid biosynthesis	276	3.66	ko00940
5	glycolysis / gluconeogenesis	242	3.21	ko00010
6	amino sugar and nucleotide sugar metabolism	225	2.99	ko00520
7	cysteine and methionine metabolism	221	2.93	ko00270
8	pyrimidine metabolism	214	2.84	ko00240
9	carbon fixation in photosynthetic organisms	203	2.69	ko00710

进一步分析 KEGG 发现, 940 条 Unigenes 参与苯丙烷类、类黄酮类、萜类等 20 个次生代谢标准生物合成通路 (表 2)。其中, 苯丙烷类生物合成 (ko00940) 注释到的基因数最多, 达 276 条; 其次是类黄酮的生物合成 (ko00941), 基因数为 100 条; 萜类化合物骨架生物合成 (ko00900) 有关基因 93 条; 80 条基因与芪类化合物、二芳基庚烷和姜辣素 (ko00945) 有关; 分别有 64、46、45 和 41 条基因与类胡萝卜素生物合成 (ko00906)、异喹啉生物碱生物合成 (ko00950)、二萜类生物合成 (ko00904) 和萜类生物合成 (ko00960) 有关; 37 条参与玉米素的生物合成 (ko00908); 油菜素类固醇的生物合成 (ko00905) 有 35 条; 34 条基因涉及倍半萜和三萜类的生物合成 (ko00909);

花青素、单环内酰胺类、柠檬烯和蒎烯的降解、单萜类生物合成、黄酮和黄酮醇类合成、异黄酮类合成、芥子油苷的生物合成、咖啡因的合成和甜菜红色素的生物合成相关的基因均较少, 在 30 条以下。这些数据为越南安息香次生代谢机制研究提供数据基础。

2.3 CDS 和转录因子分析

对越南安息香所有 Unigenes 进行 CDS 分析, 通过 BLAST 获得 41 351 个 CDS 序列, 利用 ESTscan 数据库得到 3461 条 CDS。转录因子通过预测, 被分到 54 个转录因子家族, 其中 C2H2、ERF、bHLH、MYB、GRAS、NAC、MYB-relate 和 WRKY 占主体, 说明众多转录调控过程 (图 5) 参与了越南安息香根、茎和叶的生理代谢。

表 2 转录组 Unigenes 次生代谢的 KEGG 通路注释统计

Table 2 Secondary metabolism KEGG pathway annotation analysis of transcriptomic Unigenes

编号	代谢通路	Unigenes 数量	占比/%	通路 ID
1	phenylpropanoid biosynthesis	276	29.36	ko00940
2	flavonoid biosynthesis	100	10.64	ko00941
3	terpenoid backbone biosynthesis	93	9.89	ko00900
4	stilbenoid, diarylheptanoid and gingerol biosynthesis	80	8.51	ko00945
5	carotenoid biosynthesis	64	6.81	ko00906
6	isoquinoline alkaloid biosynthesis	46	4.89	ko00950
7	diterpenoid biosynthesis	45	4.79	ko00904
8	tropene, piperidine and pyridine alkaloid biosynthesis	41	4.36	ko00960
9	zeatin biosynthesis	37	3.94	ko00908
10	brassinosteroid biosynthesis	35	3.72	ko00905
11	sesquiterpenoid and triterpenoid biosynthesis	34	3.62	ko00909
12	anthocyanin biosynthesis	26	2.77	ko00942
13	monobactam biosynthesis	15	1.60	ko00261
14	limonene and pinene degradation	14	1.49	ko00903
15	monoterpenoid biosynthesis	12	1.28	ko00902
16	flavone and flavonol biosynthesis	10	1.06	ko00944
17	isoflavonoid biosynthesis	6	0.64	ko00943
18	glucosinolate biosynthesis	3	0.32	ko00966
19	caffeine metabolism	2	0.21	ko00232
20	betalain biosynthesis	1	0.11	ko00965

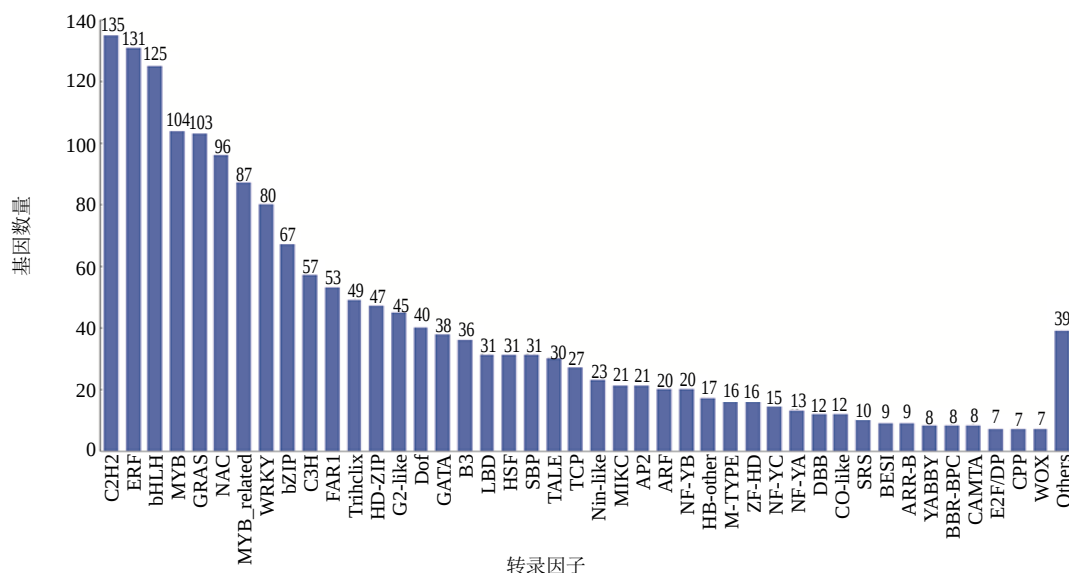


图 5 转录因子分析

Fig. 5 Transcription factor analysis

2.4 SSRs 特征分析

MISA 软件用于 Unigenes 的 SSRs 检索，8712 个 Unigenes 中有 10 974 个 SSRs (表 3)。二碱基重复 SSRs 丰度最高，为 6282 (57.24%)

个，其中，AG/CT 类型所占比例最高；其次是三碱基重复，3072 (27.99%) 个；四、五和六碱基分别有 777 (7.08%)、269 (2.45%) 和 574 (5.23%) 个。

表 3 转录组 Unigenes SSRs 分布

Table 3 SSRs analysis of 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	2057	1267	973	712	503	307	81	9	48	325	6282	57.24
三碱基重复	0	1760	734	319	91	70	64	6	10	7	2	9	3072	27.99
四碱基重复	549	164	42	17	5	0	0	0	0	0	0	0	777	7.08
五碱基重复	206	45	13	2	1	1	0	0	0	1	0	0	269	2.45
六碱基重复	405	145	11	10	0	3	0	0	0	0	0	0	574	5.23
合计	1160	2114	2857	1615	1070	786	567	313	91	17	50	334	10 974	100.00

3 讨论

高通量测序技术在药用植物的研究上已经展开了较为广泛的应用，并取得了重大的进展^[16]。该研究首次对越南安息香采用 Illumina HiSeq™ 2000 技术获得其转录组数据并进行相关分析，测序质量与质控结果均良好，53 835 045 条 clean reads 组装得到 69 151 条 Unigenes，其长度与 reads 覆盖度均合理。Unigenes 数据巨大，基本涵盖所有转录组信息，初步揭示越南安息香根、茎和叶的基因表达特征。

高通量测序获得的转录组数据的相关分析与生物信息学分析密切相关。本研究通过生物信息学分析对所获得的转录组数据进行 reads 滤过、组装，

以及基因注释和功能的分类等，涉及的软件有 BLAST、Trinity、ESTscan 等。

基于 BLAST 分析，69 151 条 Unigenes 通过 4 大数据库 (Nr、KEGG、KOG 和 Swiss-port) 比对有 41 951 (60.67%) 条被成功注释，该注释占比与已经报道的甘草^[17]、黄精^[18]、地黄^[19]等物种类似，反映了越南安息香中存在大量未知序列特征及功能的 Unigenes，有待进一步探究。

越南安息香转录组的 GO 分类揭示其特性与细胞组分、细胞过程和分子功能相关；通过 KOG 数据库从基因水平寻找直系同源物种，预测 ORF 的生物学功能，很大程度上提高了注释的准确度^[20]，该研究获得较为全面的 KOG 类群 (25 个)。进一步通

过 KEGG 数据库对越南安息香转录组分析,发现了 94 个标准代谢通路,这些通路可能与越南安息香的水分和矿物质吸收、光合和呼吸作用等生理过程相关。此外,还发现 20 条次生代谢标准通路,涉及了苯丙烷类、黄酮类及萜类等生物合成。其中,参与苯丙烷类生物合成、萜类生物合成的 Unigenes 分别有 276 个和 91 个。越南安息香的次生代谢产物以苯丙烷类和萜类为主^[21],本研究转录组数据为其生物合成通路及调控提供重要的依据。

转录因子在基因表达上起到重要的调控作用。C2H2、AP2/ERF、bHLH、MYB 等家族在植物细胞苯丙烷类、甲羟戊酸代谢途径中起到调控作用^[22]。越南安息香 Unigenes 覆盖了 plantTGdb 数据库 54 个转录因子家族,涉及到次生代谢调控的转录因子较多,说明了越南安息香植物体中复杂的转录调控机制。

转录组数据应用于构建遗传图谱和分析遗传多样性是其另外的功用^[23],本研究挖掘了越南安息香 8 712 个 Unigenes 的 10 974 个 SSRs 位点,重复类型以二核苷酸为主,三核苷酸次之,与以二核苷酸为主的当归^[23]的研究一致。双核苷酸重复以 AG/CT 类型为主,三核苷酸重复以 AAG/CTT 类型为主,与当归^[23]、甘草^[23]等植物相同。初步认为,植物中 SSRs 重复以双核苷酸和三核苷酸重复为主,同时,不同的物种又有所区别。

越南安息香的转录组数据研究初步获得了大量的基因信息,后续通过进一步系统分析,以期揭示越南安息香中苯丙素脂、三萜类等活性物质的生物合成、调控机制,以及该物种的遗传特征,为越南安息香资源的开发及利用提供数据基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Burger P, Casale A, Kerdudo A, et al. New insights in the chemical composition of benzoin balsams [J]. *Food Chem*, 2016, 210: 613-622.
- [2] Orchard A, Viljoen A, van Vuuren S. Antimicrobial essential oil combinations to combat foot odour [J]. *Planta Med*, 2018, 84(9/10): 662-673.
- [3] Shin S. Anti-*Aspergillus* activities of plant essential oils and their combination effects with ketoconazole or amphotericin B [J]. *Arch Pharm Res*, 2003, 26(5): 389-393.
- [4] Min B S, Oh S R, Ahn K S, et al. Anti-complement activity of norlignans and terpenes from the stem bark of *Styrax Japonica* [J]. *Planta Med*, 2004, 70(12): 1210-1215.
- [5] Min B S, Na M K, Oh S R, et al. New furofuran and butyrolactone lignans with antioxidant activity from the stem bark of *Styrax Japonica* [J]. *J Nat Prod*, 2004, 67(12): 1980-1984.
- [6] Wang F, Hua H M, Pei Y H, et al. Triterpenoids from the resin of *Styrax tonkinensis* and their antiproliferative and differentiation effects in human leukemia HL-60 cells [J]. *J Nat Prod*, 2006, 69(5): 807-810.
- [7] Sabbah D A, Al-Tarawneh F, Talib W H, et al. Benzoin schiff bases: Design, synthesis, and biological evaluation as potential antitumor agents [J]. *Med Chem*, 2018, 14(7): 695-708.
- [8] Castel C, Fernandez X, Lizzani-Cuvelier L, et al. Characterization of the chemical composition of a byproduct from *Siam benzoin* gum [J]. *J Agric Food Chem*, 2006, 54(23): 8848-8854.
- [9] Ha J, Kang Y G, Lee T, et al. Comprehensive RNA sequencing and co-expression network analysis to complete the biosynthetic pathway of coumestrol, a phytoestrogen [J]. *Sci Rep*, 2019, 9(1): 1934.
- [10] Xu Z C, 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]. *Plant J*, 2015, 82(6): 951-961.
- [11] Grabherr M G, Haas B J, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nat Biotechnol*, 2011, 29(7): 644-652.
- [12] Zhang Z H, Cao B L, Li N, et al. Comparative transcriptome analysis of the regulation of ABA signaling genes in different rootstock grafted tomato seedlings under drought stress [J]. *Environ Exp Bot*, 2019, 166: 103814.
- [13] Tang J R, Lu Y C, Gao Z J, et al. Comparative transcriptome analysis reveals a gene expression profile that contributes to rhizome swelling in *Panax japonicus* var. *major* [J]. *Plant Biosyst - Int J Deal Aspects Plant Biol*, 2020, 154(4): 515-523.
- [14] Cao Z, Li L, Kapoor K, et al. Using a transcriptome sequencing approach to explore candidate resistance genes against *Stemphylium blight* in the wild lentil species *Lens ervoides* [J]. *BMC Plant Biol*, 2019, 19(1): 399.
- [15] 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.

- [16] Haile Z M, Nagpala-De Guzman E G, Moretto M, *et al.* Transcriptome profiles of strawberry (*Fragaria vesca*) fruit interacting with *Botrytis cinerea* at different ripening stages [J]. *Front Plant Sci*, 2019, 10: 1131.
- [17] Liu Y L, Zhang P F, Song M L, *et al.* Transcriptome analysis and development of SSR molecular markers in *Glycyrrhiza uralensis* fisch [J]. *PLoS One*, 2015, 10(11): e0143017.
- [18] Wang S Q, Wang B, Hua W P, *et al.* De novo assembly and analysis of *Polygonatum sibiricum* transcriptome and identification of genes involved in polysaccharide biosynthesis [J]. *Int J Mol Sci*, 2017, 18(9): E1950.
- [19] Zhou Y Q, Wang X N, Wang W S, *et al.* De novo transcriptome sequencing-based discovery and expression analyses of verbascoside biosynthesis-associated genes in *Rehmannia glutinosa* tuberous roots [J]. *Mol Breed*, 2016, 36(10): 1-11.
- [20] Mudado M d e A, Ortega J M. A picture of gene sampling/expression in model organisms using ESTs and KOG proteins [J]. *Genet Mol Res*, 2006, 5(1): 242-253.
- [21] Hovaneissian M, Archier P, Mathe C, *et al.* Analytical investigation of Styrax and benzoin balsams by HPLC-PAD-fluorimetry and GC-MS [J]. *Phytochem Anal*, 2008, 19(4): 301-310.
- [22] Sun R Z, Cheng G, Li Q, *et al.* Light-induced variation in phenolic compounds in cabernet sauvignon grapes (*Vitis vinifera* L.) involves extensive transcriptome reprogramming of biosynthetic enzymes, transcription factors, and phytohormonal regulators [J]. *Front Plant Sci*, 2017, 8: 547.
- [23] Chen C, Chen Y J, Huang W J, *et al.* Mining of simple sequence repeats (SSRs) loci and development of novel transferability-across EST-SSR markers from de novo transcriptome assembly of *Angelica dahurica* [J]. *PLoS One*, 2019, 14(8): e0221040.

[责任编辑 时圣明]