

· 基础研究 ·

药用植物冬凌草高通量转录组测序与分析[△]

陈延清, 胡志刚, 刘大会, 黄必胜, 刘迪*

(湖北中医药大学 药学院, 湖北 武汉 430065)

[摘要] **目的:** 得到冬凌草高通量转录组数据, 为挖掘冬凌草二萜类生物合成途径关键基因, 研究冬凌草 EST-SSR 分子标记提供数据来源。**方法:** 利用 Illumina HiSeq 4000 高通量测序技术, 采用 Trinity 的方法从头组装、序列拼接和去冗余处理; 基于 BLAST 完成 Unigene 编码蛋白 NR 功能注释、GO 分类、KEGG 代谢通路注释和分类、功能基因挖掘, SSR 分子标记挖掘。**结果:** 获得冬凌草叶与茎两种不同组织共 12 GB 转录组数据, 得到 3 7961 个 Unigene 基因, 平均长度 1063 bp; 得到冬凌草叶和茎转录组有 60 条 unigenes 参与萜类化合物骨架合成、6 条 unigene 参与各种萜类化合物合成, 有 26 条 unigene 参与二萜化合物合成途径, 叶与茎两种组织中有差异表达的基因 4565 个, 其中在茎中表达较高的为 1668 个, 在叶中表达较高的有 2697 个, 与二萜类化合物合成相关的差异表达基因有 15 个。**结论:** 本研究为冬凌草二萜化合物生物合成途径中关键基因的挖掘和分子育种提供了数据信息, 为深入研究冬凌草中冬凌草甲素等有效成分的生物合成途径及其调控机制提供基础。

[关键词] 冬凌草; 二萜合成; 转录组分析; RNA-seq; 差异基因表达分析

High-throughput Transcriptome Sequencing and Analysis of *Isodon rubescens* (Hemsl.) H. Hara

CHEN Yan-qing, HU Zhi-gang, LIU Da-hui, HUANG Bi-sheng, LIU Di*

(College of Pharmacy, Hubei University of Chinese Medicine, Wuhan 430065, China;)

[Abstract] **Objective:** To obtain high-throughput transcriptome data of *Isodon rubescens* (Hemsl.) H. Hara, and provide data sources for the study of the key genes of diterpenoid biosynthetic pathway and EST-SSR molecular markers in *I. rubescens*. **Methods:** Illumina HiSeq 4000 high-throughput sequencing technology was used to obtain transcriptome data from the leaves and stems of adult *I. rubescens* by Trinity method for de novo assembly, sequence splicing, and de-redundant processing. Unigene-encoded proteins were based on BLAST, NR functional annotation, GO classification, KEGG metabolic pathway annotation. Moreover, EST-SSR molecular was preliminarily developed. **Results:** A total of 12 GB transcriptome data were obtained from the annual leaf and stem tissues, yielding 37, 961 Unigene genes with average length of 1063 bp. There were 60 unigenes involved in terpenoid skeleton synthesis, 6 unigenes involved in various terpenoids synthesis, and 26 unigenes involved in diterpenoid compound pathways stem from the leaf and stem. There were 4565 differentially expressed genes in the tissues, of which 1668 were highly expressed in stems and 2697 were highly expressed in leaves. There were 15 differentially expressed genes related to the synthesis of diterpenoids. **Conclusion:** This study provides data for the key genes in diterpenoid biosynthetic pathways and molecular breeding, providing a basis for in-depth study of biosynthetic pathways and regulatory mechanisms of oridonin and other active ingredients in *I. rubescens*.

[Keywords] *Isodon rubescens*; diterpene synthesis; transcriptome analysis; RNA-seq; differential gene expression analysis

doi:10.13313/j.issn.1673-4890.20180618001

大多数唇形科药用植物味甘苦, 性微寒, 中国民间习用于治疗失眠、消化不良、风湿病、肿瘤

等^[1-2]。现代化学和药理研究表明, 唇形科植物主要的生物活性成分为二萜化合物, 由于其复杂多样的

[△] [基金项目] 湖北省教育厅科学技术研究项目 (B2016077)

* [通信作者] 刘迪, 助理研究员, 研究方向: 中药资源开发利用与品质评价研究; Tel: (027) 68890106, E-mail: liuditem@126.com

碳骨架和具有多种抗肿瘤药理活性成分,二萜类化合物已经成为抗癌药物候选药物^[3-4]。冬凌草 *Isodon rubescens* (Hemsl.) H. Hara 属于唇形科中药用价值和经济价值较为重要的药用植物,其中全草中含量较高二萜类化合物冬凌草甲素、冬凌草乙素具有良好杀菌消炎、抗肿瘤、抗氧化等作用。目前,冬凌草中的二萜类化合物的研究与市场开发成为了热点^[5-6]。随着对冬凌草资源与品质的研究,其道地性和种质资源的遗传规律探索也逐渐深入^[7]。采用现代化学分析技术,对冬凌草最佳采收期^[8],以及对不同产地来源主要二萜活性成分冬凌草甲素和冬凌草乙素的含量差异进行了研究^[9],并建立了不同产地冬凌草药材的 HPLC-PDA 指纹图谱特征图谱^[10],阐述冬凌草药材的最佳采收时期和产地差异。利用随机引物标记 RAPD 和 ISSR^[11],以及特异标记简单序列重复 SSR (simple sequence repeat, SSR)^[12] 等分子标记技术,探索了冬凌草遗传多样性规律,为筛选优质种质资源提供了丰富的遗传基础,探讨药材道地性的形成机制。

高通量测序(next generation sequencing, NGS)技术是对传统测序的革命性改变,转录组学研究方法转录组测序(RNA-seq)揭示药用植物遗传育种和基因进化规律,其具有通量高、成本低、不受基因组序列信息限制等优点^[13]。对于非模式生物而言,例如冬凌草,转录组测序更偏重基因编码区域,与庞大的基因组信息相比,重复元件和富含 GC 的区域比较少,拼接相对容易^[14],转录组测序针对物种转录本直接进行测序,为基因组序列信息有限的非模式生物提供可能。本研究利用 Illumina HiSeq 4000 对药用植物冬凌草药用部位叶和茎进行了高通量转录组测序,建立冬凌草转录组数据库,为深入研究冬凌草中冬凌草甲素等有效成分的生物合成途径及其调控机制提供基础,为药用活性成分的合成代谢以及药材道地性品质的形成提供生物学依据。高效地发现非模式物种的新基因和 SSR 位点的信息分析,也为开展分子标记辅助育种、基因工程技术选育创制新的药用冬凌草优良品种奠定基础,对优质中药材的品种选育和规范化栽培生产具有重要的指导意义。

1 材料与方法

1.1 材料

本研究的冬凌草样品来自于道地产区河南济

源^[15],采于药材最佳采收期^[8],在同一生境下,选取冬凌草地上部分叶和茎为研究材料,迅速用液氮研磨,用于 RNA 样品提取。

1.2 总 RNA 的提取文库、构建和转录组测序

用 TRIzol[®] Reagent、试剂盒 Plant RNA Purification Reagent 试剂盒 (Invitrogen, 美国) 提取、纯化叶与茎总 RNA^[16]。分别用 Nanodrop 2000 与 Agilent 2100 Bioanalyzer 进行总 RNA 质量和纯度检测。为了获得尽可能丰富的冬凌草采收期转录组信息,特别地将同株的叶和茎的 RNA 混合。富集 mRNA 使用 Dynabeads Oligo(dT)25 的磁珠。在 Thermomixer 中加入 fragmentation buffer 将 mRNA 打断为短片段,添加 $0.1 \text{ mol} \cdot \text{L}^{-1} \text{ Mg}^{2+}$ 缓冲液到所得的 mRNA 中,高温下使其片段化后,然后将 mRNA 为模板反转合成 cDNA、然后通过磁珠纯化、末端修复、3' 末端加碱基 A、加测序接头等步骤后选择片段大小,进行 PCR 扩增, adaptor 构建冬凌草转录组测序文库,使用 Illumina HiSeq 4000 进行测序及分析。

1.3 测序数据质量处理及组装

使用 SeqPrep (<https://github.com/jstjohn/SeqPrep>)、Sickle (<https://github.com/najoshi/sickle>) 软件进行过滤原始数据,从而去除含有 Primer/Adaptor 的序列、含 N 比率超过 10% 的 reads 和含接头的低质量 reads,并且使 Q30 (碱基测序错误率低于 0.1%) 达到 80% 以上。从而获得去掉接头后的测序序列(cleanreads)进行后续分析。

由于缺乏冬凌草基因组的数据,因此利用 Trinity 软件对冬凌草叶与茎转录组重头组装(de novo assembly),测序序列之间的重叠(overlap)信息组装得到重叠群(contigs),依据两端信息(paired-end)和重叠群的相似性,然后使用 TGICL 软件进行聚类整合和延长得到非冗余的转录本(transcripts)。并从中选取最长的转录本作为非重复序列基因(unigenes)。Trinity 参数设置为 K-mer = 25,当序列延伸成重叠群时重复为 K-mer-1。拼接完毕后,将没有匹配结果的短读序的转录本删除。处理与注释冬凌草叶和茎的差异表达基因使用相同参考基因,得到同一时期冬凌草叶和茎所有 Unigenes 数据库使用 TGICL 的聚类组装策略。

1.4 功能基因注释

通过 Transdecoder version v 2.0.1 对 unigenes 的

CDS进行了预测。利用 Trinotate 软件包(v 2.0.0, <https://trinotate.github.io/>)的标准流程对 All-unigenes 进行了全面的注释,用 Swissprot^[17]、Pfam^[18]、eggNOG^[19]、Gene Ontology^[20]、SignalP^[21]、Rnammer^[22]和 KEGG Ontology 数据库注释。

我们使用 Trinotate 软件(版本:20140708)对预测出的编码序列进行功能注释。Trinotate 软件整合了多个数据库,可以从不同角度对转录本进行功能注释,使用软件或数据库如下:使用 diamond(版本:0.5.3)blastx, SwissProt、TrEMBL 和 kegg 作为数据库,对转录本进行注释;同时,还使用 blastp(版本:2.2.28+), SwissProt 作为数据库,对转录本进行注释;使用 hmmscan(HMMER)软件(版本:3.1), Pfam(Bateman et al. 2002)作为数据库,对预测的蛋白序列的结构域进行注释;用 SignalP(Petersen et al. 2011)(版本:V4)软件,预测信号肽;使用 tm-HMM(版本:2.0c)预测跨膜结构;使用 RNAmmer(版本:1.2)预测 rRNA 基因。

1.5 基因表达水平及表达差异量分析

基于 Trinity 拼接得到的转录本,我们进行表达丰度计算。最先利用 bowtie(版本:1.0.1;参数:mismatch=2)将每个样品的 reads 分别比对到组装出来的转录本上,得到 bam 格式的比对文件,然后使用 RSEM(版本:v1.2.17)(<http://deweylab.biostat.wisc.edu/rsem/>)计算比对到转录本上的 reads 的表达量。为了便于不同基因之间进行表达量的比较,我们使用 FPKM 对表达量进行标准化。FPKM(expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced)^[23]是每百万 fragments 中来自某一基因每千个碱基长度的 fragments 数目,其同时考虑了测序深度和基因长度对 fragments 计数的影响。

差异表达基因(differentially expressed genes, DEGs)是基于基因表达水平分析中得到的 read count 数据,根据筛选阈值筛选出来的表达差异的基因。参考 edgeR 进行差异基因分析,筛选阈值选择为 False Discovery Rate(FDR)≤0.05, |log₂FoldChange| ≥1。其中 FC(Fold Change)表示叶与茎间表达量的比值,其绝对值越大表明差异倍数越小, FDR 值越大则为表达差异越不显著^[24]。使用软件 GOseq(版本:3.0)对样品间差异表达基因进行 GO 库注释的分类统计和功能富集分析,使用 KOBAS([ko-bas.cbi.pku.edu.cn/home.do\)进行 KEGG pathway 富集分析,使用 Fisher 精确检验法对差异表达基因 KEGG 通路富集的显著性进行评价。使用超几何检验来找到差异表达基因中显著富集的途径。并经过多次检测和校正后,定义 P < 0.05 的通路途径为显著富集差异表达基因的途径。](http://</p></div><div data-bbox=)

1.6 SSR 分析

以冬凌草转录组为研究对象,使用 MISA(Microsatellite)软件进行 SSR 搜索。对 Unigenes 进行 SSR 检测,参数设置为“1~12、2~6、3~5、4~5、5~4、6~4”,1~12 代表单碱基重复至少 12 次才算 SSR,双碱基 6 次,三碱基 5 次,以此类推,重复单元最多有 6 个碱基。

2 结论与分析

2.1 总 RNA 质量检测 and 组装结果分析

经检测,样品总 RNA 的 A260/A280 均大于 1.8, 28 S/18 S 在 1.5~2.3 之间,反应 RNA 完整性的 RIN 值在 8.5~9.5 之间为质量满足建立和基因功能注释、分类及样本间异同性分析。

通过组装,叶与茎冬凌草样品共获得 3 7961 个 Unigene 序列。All-Unigene 总长为 40366712 bp,平均长度为 1063 bp,反应组装效果的 N50 值为 1810。其中 N50 是评估组装质量最常用的指标之一,它反映了组装结果的连续性和完整性。所得 Unigene 的数目随长度的增加而减少,其 Unigene 大多集中在 200~3000 bp 之间,平均 Unigene 为 1063 bp。总的来说,Unigene 的整体长度分布均匀(图 1)。

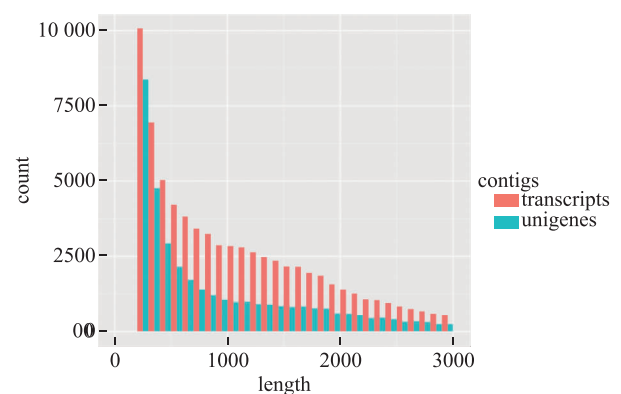


图 1 All-Unigene 长度分布统计

2.2 Unigene 蛋白数据库比对结果

基于 BLAST 算法对基因相似性分析,在 37 961 个

Unigene 中, PFam 库中 17842Unigene 同源匹配得到信息, 在 SignalP 数据库中有 2394 个, 在 eggNOG 数据库中有 8769 个, 在 GO 数据库中有 18 776 个, KEGG 数据库中有 23 936 个(表 1)。通常情况下测序的难度与基因组、转录组复杂程度有关。此外, 由于缺乏参考基因组信息, 当前研究非模式生物的转录组仍具有一定挑战。即使目前国内对药用植物的从头转录组研究已有报道, 但测序深度和基因检出数目仍有待完善和提高^[25-26]。

表 1 All-Unigene 蛋白数据库统计结果

注释/工具	基因数量	比例数
All	37 961	100. 00%
Annotated	27 382	72. 13%
Blast_hit	24 234	63. 84%
Pfam	17 841	47. 00%
SignalP	2394	6. 31%
Gene_Ontology(GO)	18 776	49. 64%
eggNOG	8769	23. 10%
TmHMM	6510	17. 15%
RNAMMER	6	0. 02%
KEGG Ontology(KO)	23 936	63. 05%

2.3 GO 注释和分类

GO 注释和分类是转录组数据研究中的一种国际标准化的基因功能分类手段, 它提供了一套动态更新的标准词汇表, 用来全方面描述基因和基因产物的特性。GO 功能分析由细胞组分、分子功能以及生物学过程三大部分组成。通过 GO 分析, 37961Unigene 中有 18 775 个获得了 GO 注释, 被注释为“细胞组分”有 15 639 个; 16 455 个被注释为“分子功能”, 1541515639 被注释为“生物学过程”。我们对其中的 level2 的 GO 注释进行了统计(图 2)。

2.4 KEGG 注释和分类

KEGG 是全面分析生物体基因产物的代谢合成途径以及产物功能的参考数据库, 根据生化途径, 提供基因的功能注释和分类信息, 对比分析表达基因在生物学上的系统行为。将 6456 个 Unigene 映射到 KEGG Pathway 数据库中, 注释为代谢行为相关的序列共有 4235 个。其中涉及代谢(metabolism)、生物系统(organismal systems)环境信息处理(environmental

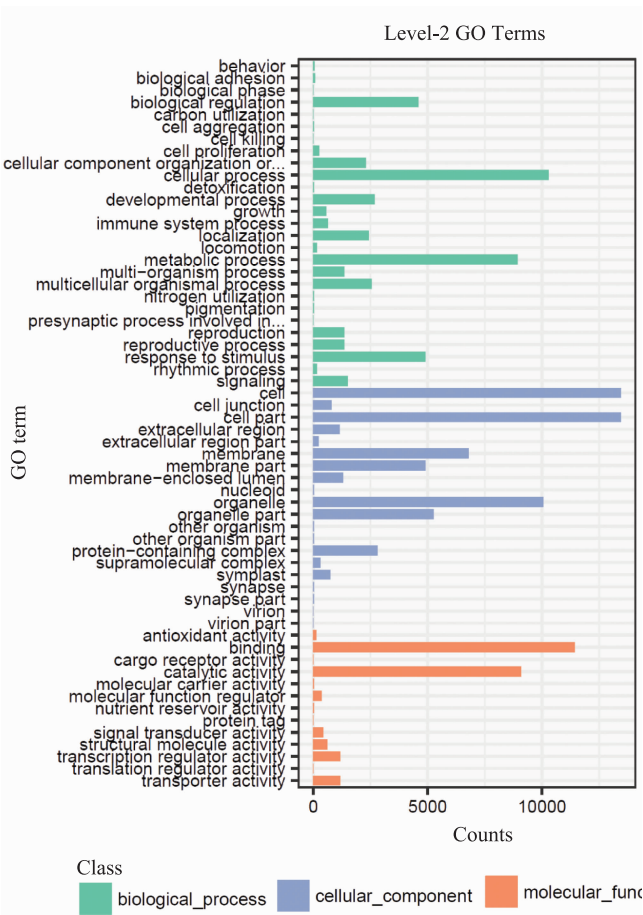


图 2 Level-2 的 GO 注释结果

information processing)、细胞进程 (cellular processes) 及遗传信息处理 (genetic information processing) 5 个大类及 102 条代谢通路, 包含 Unigene 最多的是 Ribosome(核糖体)共有 389 个, 占糖酵解/糖异生 (glycolysis/gluconeogenesis) 共有 226 个, 占 6%; 其次是 Oxidative phosphorylation(氧化磷酸化)共有 244 个, 占 3.8%, 这些转录水平上的数据变化充分说明了冬凌草药用部位在特定发育时期的新陈代谢变化(图 3)。进一步研究还发现, 在代谢途径中, 与中药有效活性成分生物合成相关 Unigenes 序列分别是苯丙素类合成 (phenylpropanoid biosynthesis) 101 条、萜类化合物骨架合成 (terpenoid backbone biosynthesis) 60 条、各种萜类化合物合成 (sesquiterpenoid, triterpenoid, monoterpene and diterpenoid biosynthesis) 6 条、(莨菪类、哌啶类、吡啶类)生物碱合成 (tropane, piperidine and pyridine alkaloid biosynthesis) 15 条、黄酮类成分合成 (flavonoid, flavone, flavonol and isoflavonoid biosynthesis) 30 条(图 4)。

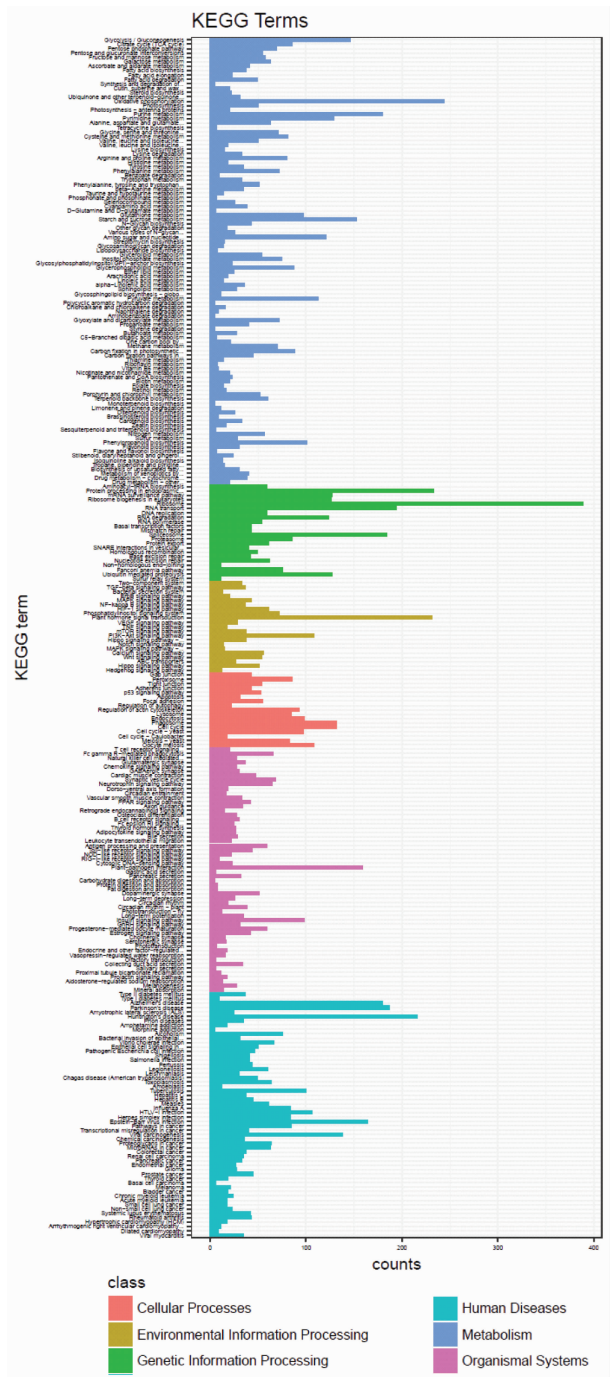


图 3 KEGG 注释和分类结果

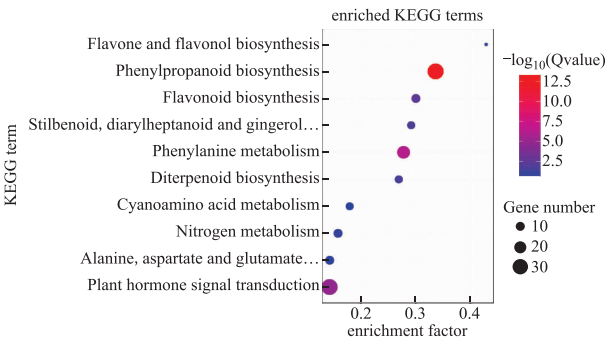


图 4 KEGG 富集分析图

2.5 二萜化合物合成途径相关基因

植物二萜化合物主要合成途径已经明确，分为三步：最先，活性异戊烯焦磷酸(IPP)及其异构体二甲基烯丙基焦磷酸(DMAPP)生成；之后为二萜化合物通用前体香叶基香叶基焦磷酸(GGPP)碳骨架的合成；最后在相关酶的作用下发生环化、羟基化等从而形成完整的二萜化合物分子^[27]。利用本地化 Blast 的检索方式在冬凌草中确定了本数据中与二萜化合物生物合成相关的候选基因。发现二萜合成途径上的相关基因共有 26 个，其中与二萜骨架合成有关的基因有 9 个，其中与焦磷酸合成酶(copalyl diphosphate synthase)有关的 Unigene 有 6 个，与贝壳杉烯合酶合成酶(KSL, kaurene synthase-like synthase)有关的 unigene 有 3 个，具有细胞色素 P450(CYP450, cytochrome)功能的 Unigene 共有 5 个。

2.6 差异表达基因的分析

通过冬凌草叶与茎两个不同组织样品的转录本/基因的表达量概率(FPKM scores)密度分布(图 5)，不同样品间差异基因在 KEGG 通路富集具有特别显著性差异($P < 0.01$)的通路描述，可以鉴定出一些在叶和茎两种组织在次生物质合成与代谢途径中的差异表达基因。对冬凌草叶与茎两种组织进行差异表达分析，得到差异表达基因 4565 个，其中在茎中表达较高的为 1668 个，在叶中表达较高的有 2697 个，与二萜类化合物合成相关的差异基因有 15 个。冬凌草叶与茎两种组织差异基因表达图如图 6。

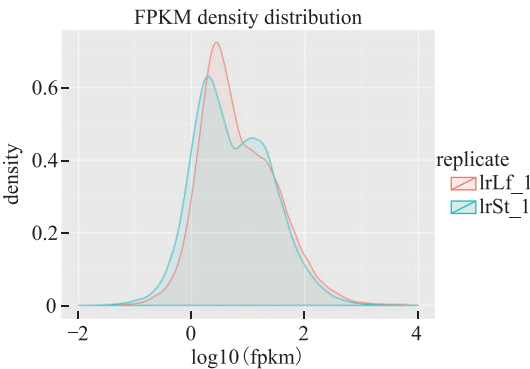


图 5 冬凌草叶和茎基因表达量密度分布结果

2.7 SSR 分析结果

简单重复序列又称微卫星序列(Simple Sequence Repeat, SSR)由于具有高度多态性，克服一般分子标记不确定、非特异的缺点，已在分子辅助育种、指纹数据库的构建、品种纯度鉴定及遗传多样性分

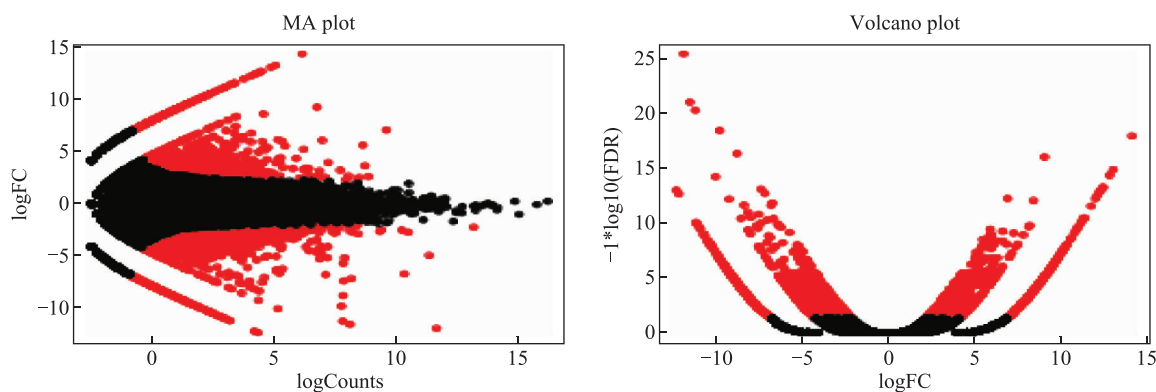


图 6 冬凌草叶与茎差异表达基因的 MA 图和火山图

析等方面得到广泛应用^[28-29]。在本次所测得的冬凌草转录组中,共检测到 8787 条 SSR 序列,分布于 7333 条 unigene 中。SSR 中二核苷酸重复的 SSR 标记有 4016 条,占比 45.70%,是主要类型;其次是单核苷酸重复基序(2480 条,占 28.22%),接着是三核苷酸重复基序(1707 条,占比 19.43%)。

3 讨论

本文通过 Illumina Hiseq4000 平台,利用 RNA-seq 技术对道地产地冬凌草的叶与茎进行了转录组测序与功能分析,初步建立了它们的基因表达图谱,并通过比较药用部位和非药用部分合成途径差异酶基因的发掘初步阐述了不同组织冬凌草成分差异的分子机制。并分析了转录组中的 SSR,揭示了冬凌草中遗传规律,也为药用冬凌草的遗传多样性研究和分子标记开发奠定了分子基础。

冬凌草叶和茎两个组织样品的高通量测序分析共获得 3 7961 个 Unigene 序列。大量 Unigene 在多个数据库中不能匹配到已知基因,可能由于数据库中由于缺少冬凌草基因组信息;通过对转录组数据的差异表达分析发现与冬凌草合成途径有关的差异基因主要注释在苯丙素类化合物、萜类化合物骨架合成、生物碱类化合物以及黄酮类成分生物合成途径等通路中。冬凌草中的活性成分为二萜类物质^[6],研究其生物合成途径下游关键基因尤为重要。基于本文转录组数据,陈^[30]等人成功克隆并初步验证了下游 *IrCYP71* 功能基因,利用转录组数据挖掘药用植物的功能基因已经成为重要手段^[31]。对药用植物的基因功能分化进行深入分析,并与活性成分种类、数量和积累变化相结合,进而从分子生药学角度揭示由于不同生

长时间和生长环境基因表达的差异与药用部位中次生代谢成分差异的机制,才能为药材质量控制与提高提供理论依据^[32]。高通量转录组测序技术的应用同时也加速了 SSR 位点的挖掘,本研究中冬凌草的 All-Unigenes 转录本中 SSR 出现的频率特点为二核苷酸的重复单元类型最为丰富,这为后续冬凌草的种质资源和遗传多样性研究提供了丰富的分子标记数据。通过构建药用植物生长发育的不同时期、不同器官和不同产地的转录组文库,可以全方位地研究药用植物中活性成份生物合成的关键酶基因的表达水平,从而有效地揭示次生代谢产物生物合成的途径及其调控机制。

参考文献

- [1] 陈川川,刘越,郭凤仙,等.冬凌草生物学研究进展[J].湖北农业科学,2016,55(19):4901-4906.
- [2] 国家药典委员会.中华人民共和国药典:一部[M].北京:中国医药科技出版社,2015:115.
- [3] 冯卫生,李红伟,郑晓珂,等.冬凌草化学成分研究进展[J].中国新药杂志,2008,17(23):2003-2007.
- [4] 郭萍,李玉山,郭远强.冬凌草化学成分和药理活性研究进展[J].药物评价研究,2010,33(2):144-147.
- [5] 尹磊.冬凌草二萜类物质合成途径中相关功能基因分析[D].郑州:河南中医药大学,2016.
- [6] 刘淑云,谭英姿,范明松,等.冬凌草甲素研究进展[J].中国当代医药,2012,19(13):17-18.
- [7] 崔璨.冬凌草种质资源研究[D].郑州:河南中医学院,2010.
- [8] 李洋,董诚明,徐鹏,等.冬凌草适宜采收期的研究[J].中国现代中药,2014,16(10):824-828.
- [9] 陈随清,尹磊,宋君,等.不同产地冬凌草种质资源分子生物学分析[J].亚太传统医药,2016,12(16):5-9.
- [10] 靳怡然,杜英峰,田婷婷,等.HPLC-PDA 指纹图谱结合

- 主成分分析评价不同产地冬凌草药材的质量[J]. 中草药, 2015, 46(15): 2291-2295.
- [11] 李国民, 李忠超, 许凯扬, 等. 抗癌植物冬凌草种质资源遗传多样性分析[J]. 热带亚热带植物学报, 2008, 16(2): 116-122.
- [12] 黎晓英, 刘胜贵, 王丹, 等. 鱼腥草转录组 SSR 位点信息分析及其多态性研究[J]. 中草药, 2016, 47(10): 1762-1767.
- [13] 赵振宇, 王仕玉, 郭凤根, 等. 转录组测序及其在药用植物上的应用[J]. 基因组学与应用生物学, 2017(2): 820-825.
- [14] Rasool S, Mohamed R. Plant cytochrome P450s: nomenclature and involvement in natural product biosynthesis[J]. Protoplasma, 2016, 253(5): 1197-1209.
- [15] 郭琳, 程永现, 白明, 等. 冬凌草现代研究分析[J]. 中医学报, 2015, 30(3): 412-414.
- [16] Wei M M, Song M Z, Fan S L, et al. Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene expression profiling. BMC Genomics, 2013, 14(1): 97-113.
- [17] Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. [J]. 2003, 31(1): 365-370.
- [18] Finn R D, Bateman A, Clements J, et al. Pfam: the protein families database [J]. Nucleic Acids Research, 2014, 42(Database issue): 222-230.
- [19] Spang A, Saw J H, Jørgensen S L, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes[J]. Nature, 2015, 521(7551): 173-179.
- [20] Fawcett T. An introduction to ROC analysis [J]. Pattern Recognition Letters, 2005, 27(8): 861-874.
- [21] Petersen T N, Brunak S, Von H G, et al. SignalP 4. 0: discriminating signal peptides from transmembrane regions[J]. Nature Methods, 2011, 8(10): 785-786.
- [22] Lagesen K, Hallin P, Rødland E A, et al. RNA mmer: consistent and rapid annotation of ribosomal RNA genes [J]. Nucleic Acids Research, 2007, 35(9): 3100-3108.
- [23] Li B, Dewey C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [J]. BMC Bioinformatics, 2011, 12(1): 93.
- [24] Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res, 1997, 7(10): 986-995.
- [25] Cusido R M, Onrubia M, Sabater-Jara A B, et al. A rational approach to improving the biotechnological production of taxanes in plant cell cultures of *Taxus* spp [J]. Biotechnology Advances, 2014, 32(6): 1157-1167.
- [26] 刘红亮, 郑丽明, 刘青青, 等. 非模式生物转录组研究[J]. 遗传, 2013, 35(8): 955-970.
- [27] 高原, 王翌霞, 张亮, 等. 转录组测序深度对表达基因检出及表达量估计的影响[J]. 云南农业大学学报(自然科学), 2014, 29(5): 642-647.
- [28] 罗兵, 孙海燕, 徐港明, 等. SSR 分子标记研究进展[J]. 安徽农业科学, 2013, 41(12): 5210-5212.
- [29] Yue J, Zhu Z C. Optimization of SSR reaction system for identifying the authenticity of maize hybrids. [J]. Journal of Anhui Agricultural Sciences, 2011, 50(4): 291-296.
- [30] 陈延清, 胡志刚, 黄必胜, 等. 冬凌草 IrCYP71 基因的克隆和功能[J]. 中国实验方剂学杂志, 2018, 24(14): 29-35.
- [31] 宋经元, 徐志超, 陈士林. 本草基因组学专辑简介[J]. 中国科学生命科学, 2018, 48(4): 349-351.
- [32] Xu H, Song J, Luo H, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. [J]. Molecular Plant, 2016, 9(6): 949-952.

(收稿日期 2018-06-18)

(上接第 1475 页)

- [11] Phillips S J, Schapire R E. A maximum entropy approach to species distribution modeling [C]//International Conference on Machine Learning. ACM, 2004: 83.
- [12] 乔慧捷, 胡军华, 黄继红. 生态位模型的理论基础、发展方向与挑战[J]. 中国科学: 生命科学, 2013, 43(11): 915-927.
- [13] 孙洪兵, 孙辉, 蒋舜媛, 等. 基于 3S 技术的羌活区划研究 I. 基于 MaxEnt 和 ArcGIS 的羌活生长适宜性分析及评价[J]. 中国中药杂志, 2015, 40(5): 853-862.
- [14] 卢有媛, 杨燕梅, 马晓辉, 等. 中药秦艽生态适宜性区划研究[J]. 中国中药杂志, 2016, 41(17): 3176-3180.
- [15] 刘蒙蒙, 邢咏梅, 郭顺星, 等. 基于 Maxent 生态位模型预测药用真菌猪苓在中国潜在适生区[J]. 中国中药杂志, 2015, 40(14): 2792-2795.
- [16] 王运生, 谢丙炎, 万方浩, 等. ROC 曲线分析在评价入侵物种分布模型中的应用[J]. 生物多样性, 2007, 15(4): 365-372.
- [17] 王茹琳, 高晓清, 王闰利, 等. 基于 MaxEnt 的非洲橘硬药马在全球及中国的潜在分布区预测[J]. 中国农学通报, 2014, 30(28): 315-320.

(收稿日期 2018-04-28)