利用转录组数据挖掘东紫苏单萜生物合成相关基因

耿秀文，张爱丽，唐仁华，普春霞*

云南中医药大学 中药材优良种苗繁育工程中心，云南 昆明 650500

摘 要：目的 获得东紫苏 *Elsholtzia bodinieri* 转录组数据信息，了解东紫苏挥发油单萜生物合成途径。**方法** 采用 Illumina Hiseq 2000 高通量测序获得东紫苏的转录组数据，通过对数据进行组装、拼接及注释，挖掘其单萜类化合物代谢途径相关基因。**结果** 东紫苏 2 个样品材料测序后共获得 13.67 Gb 数据，91 169 356 条高质量序列，利用 Trinity 组装获得 93 327 条 unigenes，平均长度 1756 bp。将组装所得到的 unigenes 分别与 NCBI 官方的蛋白序列数据库（RefSeq non-redundant proteins，NR）、基因本体论数据库（gene ontology，GO）、真核生物蛋白质同源簇数据库（clusters of orthologous groups for eukaryotic complete genomes，KOG）、京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）等数据库进行 BLAST 比对分析。通过 KEGG 代谢通路分析，结果显示有 2 个单萜代谢相关途径，为萜类骨架生物合成（编号为 ko00900）和单萜类生物合成（编号为 ko00902），相关 unigenes 分别有 11 条和 30 条；进行实时荧光定量 PCR（reverse transcription PCR，RT-PCR）和测序，成功验证 6 个单萜合成相关候选基因全长 unigenes。**结论** 首次对东紫苏进行高通量转录组测序分析，获得了单萜生物合成的关键酶基因，为后续基因功能的研究奠定基础。

关键词：东紫苏；转录组；GPP 合酶；单萜生物合成；代谢途径

中图分类号：R282.12 **文献标志码：**A **文章编号：**0253 - 2670(2021)11 - 3373 - 10

DOI: 10.7501/j.issn.0253-2670.2021.11.025

High-throughput transcriptome sequencing of *Elsholtzia bodinieri* and excavation of genes related to monoterpene biosynthesis

GENG Xiu-wen, ZHANG Ai-li, TANG Ren-hua, PU Chun-xia

Engineering Research Center for Reproducing Fine Varieties of Chinese Medicinal Plants, Yunnan University of Chinese Medicine, Kunming 650500, China

Abstract: Objective To obtain the transcriptome data of *Elsholtzia bodinieri* and understand the pathway of monoterpene biosynthesis of essential oil from *E. bodinieri*. **Methods** In this study, Illumina Hiseq 2000 high-throughput sequencing was used to obtain the transcriptome data of *E. bodinieri*, and the genes related to the metabolism of monoterpenoids were mined by assembling, splicing and annotating the data. **Results** A total of 13.67 GB data and 91 169 356 clean reads were obtained after sequencing of the two sample materials of *E. bodinieri*, 93 327 unigenes were obtained by Trinity assembly with an average length of 1756 bp. The unigenes obtained from the assembly were blasted with NR, GO, KOG, KEGG and other databases. KEGG metabolic pathway analysis showed that there were two monoterpene metabolic pathways, terpenoid backbone biosynthesis (No. ko00900) and monoterpenoid biosynthesis (No. ko00902), with 11 and 30 related unigenes, respectively. Full-length unigenes of six monoterpene synthesis related candidate genes were successfully verified by RT-PCR and sequencing. **Conclusion** For the first time, high-throughput transcriptome sequencing analysis of *E. bodinieri* was carried out, and the key enzyme genes of monoterpene biosynthesis were obtained, which laid a foundation for the study of subsequent gene functions.

Key words: *Elsholtzia bodinieri* Vaniot; transcriptome; GPP synthase; monoterpene biosynthesis

东紫苏 *Elsholtzia bodinieri* Vaniot 为唇形科荆芥亚科（Nepetoideae）香薷族（Elsholtzieae）香薷属 *Elsholtzia* L.的植物，又名凤尾茶、牙刷草、小山茶、野山茶等，主产于云南滇中、滇南地区，生于海拔 1200～3000 m 的山坡草地、稀疏松林中或石山上。东紫苏全草入药，气清香，味辛、微苦，用于感冒、咽喉痛、头痛、扁桃腺炎、小儿口腔炎、牙痛、眼结膜炎、肝炎、消化不良[1-3]。东紫苏含有挥发油、黄酮类、三萜类、酚类等成分[4]。现代药理研究表明，东紫苏具有抑菌[5-6]、抗病毒[6-7]、抗氧化[8-9]、调血脂[10]的作用。

香薷属植物的挥发油被认为具有抗菌、抗氧化

收稿日期：2020-10-09

基金项目：国家自然科学基金资助项目（31660082）；云南省科技计划项目［2017FF117(-027)］

作者简介：耿秀文，女，硕士研究生，研究方向为中药资源开发与利用。E-mail: 871191768@qq.com

***通信作者：**普春霞，副教授。E-mail: yucca1977@163.com

等作用[11]。前期课题组对东紫苏 15 个居群叶的挥发油进行 GC-MS 分析[12]，初步确定 1,8-桉叶油醇（1, 8-cineole）、α-乙酸松油酯（α-terpinyl acetate）、β-蒎烯（β-pinene）、α-蒎烯（α-pinene）等为东紫苏叶挥发油的主成分，其中单萜为 72.469%～91.855%，倍半萜为3.727%～13.121%。付立卓等[13]、程伟贤等[14]的研究结果在主成分上与本课题组的研究结果基本一致[15]。根据文献报道，1, 8-桉叶油醇有抗炎和镇痛作用[16]，对气道炎症具有抑制作用[17]，能够减少鼻窦炎的恶化，预防慢性阻塞性肺病加重和改善哮喘[18-20]，并且还有解痉和抗菌活性[21]，以及一定的止泻作用[22]。

单萜化合物是由 2 个异戊二烯结构单元（C10）组成的链状或环状化合物，是一类结构多样的天然化合物，广泛存在于植物挥发油和树脂中[23]。植物单萜（monoterpenoid）由质体内的 4-磷酸-2-甲基赤藓糖（2-C-methyl-*D*-erythritol 4-phosphate，MEP）途径合成，香叶基二磷酸（geranyl diphosphate，GPP）是所有单萜共同的前体物质，GPP 在单萜合酶（monoterpene synthases，mono-TPS）的催化作用下生成单萜，单萜合酶是单萜生物合成的关键酶[24]。东紫苏叶挥发油的主成分 1,8-桉叶油醇应是由 GPP 在桉叶油醇合酶（cineole synthase）的作用下形成[25]。

转录组测序数据是挖掘基因以及参与各种代谢途径的酶的有效方法[26]，现已广泛应用于药用植物和芳香植物中次生代谢产物生物合成的关键基因的研究和鉴定[27]。本研究通过对东紫苏进行转录组测序，对测序数据进行组装、拼接及注释，并初步挖掘了东紫苏单萜类化合物代谢途径的相关基因，为东紫苏挥发油单萜合酶基因的克隆、次生代谢生物合成途径解析以及功能验证的研究提供了基因资源和理论基础。

1 材料

样本于 2015 年 11 月 3 日采自云南省昆明市呈贡区松茂水库（经度 102°55′18″，纬度 24°53′25″，海拔 2114 m），编号 SM；于 2015 年 11 月 7 日采自云南省昆明市嵩明县阿子营村（经度 102°47′43″，纬度 25°21′25″，海拔 2086 m），编号 AZY；原植物由云南中医药大学普春霞副教授鉴定为东紫苏 *E. bodinieri* Vaniot。样本材料采集后，移栽至云南中医药大学呈贡校区，于 2017 年 11 月 22 日采样后送转录组测序。

2 方法

2.1 东紫苏挥发油的提取及 GC-MS 分析

采用水蒸气蒸馏法提取 SM 和 AZY 2 个产地东紫苏叶片中的挥发油，利用 GC-MS 法分析检测其主要成分及含量。

2.1.1 气相条件 色谱柱：Agilent 19091J-115，HP-5（5% Phenyl Methyl Siloxan，50 m×0.32 mm×0.52 μm）；进样口温度 250 ℃，检测器（氢火焰离子检测器，FID）温度 250 ℃；柱箱升温程序：50 ℃升到 250 ℃，每分钟升 10 ℃；进样量 1 μL；分流比 20∶1；进样口压力 102.28 kPa；体积流量 2.1 mL/min；载气为高纯氦气。

2.1.2 质谱条件 电离方式：EI，电子轰击能量：70 eV，扫描范围 *m/z* 50～550；离子源温度 230 ℃；四极杆温度 150 ℃；定量方法：相对峰面积归一化法；谱库：NIST14。

2.2 转录组测序

采集东紫苏 2 个产地叶的样品材料，委托北京诺禾致源科技股份有限公司利用 Illumina Hiseq 2000 进行高通量转录组测序。

2.3 转录组数据组装

得到原始测序数据后，为保证信息分析质量，去除带接头的、N（N 表示无法确定碱基信息）的比例大于 10%的以及低质量 reads，对 raw reads 过滤，得到 clean reads，然后采用 Trinity 对 clean reads 进行拼接组装，最后得到 unigenes。

2.4 转录组功能注释

将组装得到的 unigenes 序列比对到 NCBI 官方的蛋白序列数据库（RefSeq non-redundant proteins，NR）、NCBI 官方的核酸序列数据库（nucleotide sequence database，NT）、京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）、Swiss-Prot 蛋白质数据库（A manually annotated and reviewed protein sequence database，Swiss-prot）、蛋白质家族数据库（Protein family，PFAM）、基因本体论数据库（gene ontology，GO）、真核生物蛋白质同源簇数据库（clusters of orthologous groups for eukaryotic complete genomes，KOG）这 7 个数据库进行基因功能注释。

2.5 东紫苏挥发油单萜合成相关候选基因筛选及克隆

从转录组数据中筛选出东紫苏挥发油单萜合成相关候选基因，获得基因序列和开放阅读框（ORF），筛选 ORF 长度大于 700 bp 的基因，分别设计 ORF 扩增引物。全部引物由生工生物工程（上海）股份有限公司合成（表 1）。

表 1 基因与引物序列

Table 1 Gene and primer sequences

基因 ID	编号	引物序列 (5’-3’)	ORF 长度/ bp
Cluster-11595.0	1	F1-1：ATGTGCAGCATTATAATTCAAGT	1848
		R1-1：CTCTTCAGGGGTGCCGTAGATATC	
		F1-2：TGGACAACTGGAGTGCTTCAACG	
		R1-2：TTATTTACTAGCTAGCTCATGCA	
Cluster-10292.10951	2	F2：ATGGCAGAAACAAAGAACATGTA	921
		R2：TTATTCGACAAAAGATGACACTTCC	
Cluster-10292.50707	3	F3：ATCTGCAGGCAGTTAGCCAGCA	1033
		R3：TTATAGATCATATAACATGCATG	
Cluster-10292.60008	4	F4：ATGAAGAACTCGTATGCAGTAGTTACA	834
		R4：CTAGAAAGGTGCAACCTCTATCTGATC	
Cluster-10292.9722	5	F5-1：ATGTCTAGCATCAGAATGCATGTGAC	1809
		R5-1：ATGATGAGGCTCAAACAGAGCAATTG	
		F5-2：ATTCACAAGCTAGAGGCAAGATGGTTC	
		R5-2：TCAAGAATATGGCCGGAACAACAAGC	
Cluster-10292.56890	6	F6：ATGACTGCTAGAGACGAGAAGAGAGG	819
		R6：TTATTCGACAAAAGATGACACTTCC	
Cluster-5267.0	7	F7：ATGGAGCTGAATCATCAACTAAC	912
		R7：TCAGAAGTCAGCCACATTTGTGCGG	
Cluster-1426.0	8	F8-1：ATGTCTACTACTGTAGTAACGC	1800
		R8-1：GAGTTCCTCCAAGGTACCATCG	
		F8-2：CTCGTGCTGGAGCTCGCCAAGC	
		R8-2：TCAATTGCAATGGAGGATCTGA	
Cluster-10292.51149	9	F9：ATGCATACATATATGCATATTA	882
		R9：CTAGAAAAGTGCAACCTCTATC	
Cluster-10292.60790	10	F10：ATGGTAGTTGCGGAGGTTCCCA	916
		R10：CTACTTTGTTCTGGTGATGACT	
Cluster-10292.60789	11	F11-1：ATGATGTATGCAAGAAGGCTTACC	1275
		R11-1：CCTAAGTTCAGGAAATTCCTCC	
		F11-2：ATTGCCCTGCTTGCTGGACAAAC	
		R11-2：CTACTTTGTTCTGGTGATGACTATG	

2.6 东紫苏单萜合成相关候选基因 RT-PCR 验证

取东紫苏叶片总 RNA 4 μL，按照 TaKaRa 公司的 PrimeScript™II 1st Strand cDNA Synthesis Kit 试剂盒说明书进行反转录反应，合成 cDNA。以 cDNA 为模板，反应体系（50 μL）：模板 2 μL，正反向引物各 1 μL，TaKaRa Taq DNA 聚合酶 25 μL，灭菌双蒸水 21 μL。PCR 反应程序为 98 ℃、30 s；98 ℃、10 s，58 ℃、20 s，72 ℃、2 min，30 个循环；72 ℃、8 min，12 ℃保温。PCR 产物经电泳分析后，送生工生物工程（上海）股份有限公司测序。

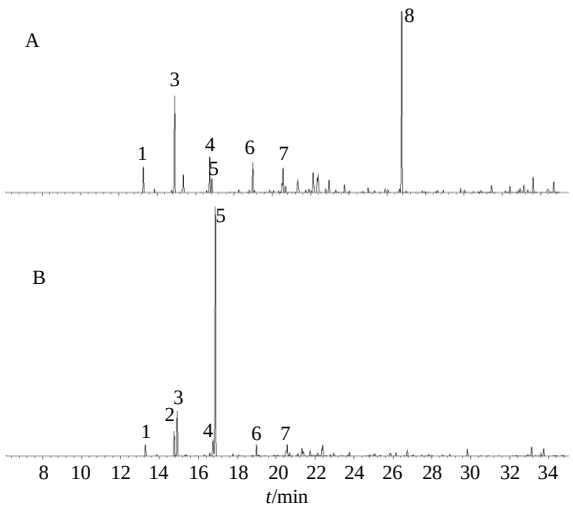
3 结果与分析

3.1 东紫苏挥发油 GC-MS 成分测定

按设定的 GC-MS 条件分析东紫苏挥发油的化学成分，经 NIST14 谱库检索、质谱分析等确定挥发油的化学成分，并用面积归一化法计算各组分的相对含量，其中 1,8-桉叶油醇和 α-乙酸松油酯均为含量较高的成分。从图 1 可以看出，产自松茂的东紫苏叶挥发油含量最高的成分为 α-乙酸松油酯，其次是 β-蒎烯；而产自阿子营的东紫苏叶挥发油含量最高的成分为 1,8-桉叶油醇，其次是 β-蒎烯。

3.2 东紫苏转录组数据组装与质量分析

东紫苏转录组测序总计得到 93 531 502 条 raw reads，过滤产生了 91 169 356 条高质量 clean reads，共计 13.67 Gb 的有效数据，Q20（Phred 数值大于 20 的碱基占总体碱基的百分比）SM 为 94.67%，AZY 为 94.62%；Q30（Phred 数值大于 30 的碱基占总体碱基的百分比）SM 为 86.81%，AZY 为 86.76%；GC



1-α-蒎烯 2-桧烯 3-β-蒎烯 4-*D*-柠檬烯 5-1,8-桉叶油醇 6-玫瑰呋喃 7-(−)-反式-松香芹醇 8-α-乙酸松油酯

1-α-pinene 2-sabinene 3-β-pinene 4-*D*-limonene 5-1,8-cineole 6-rosefuran 7-*trans*-(−)-pinocarveol 8-α-terpinyl acetate

图 1 SM (A) 和 AZY (B) 产东紫苏挥发油 GC-MS

Fig. 1 GC-MS charts of essential oil from *E. bodinieri* from SM (A) and AZY (B)

含量（clean reads 中 G 与 C 占 4 种碱基的百分比）SM 为 47.99%，AZY 为 47.17%。测序后得到的东紫苏转录组数据产出质量见表 2。运用 Trinity 软件将质控后得到的高质量序列进行组装，得到 153 657 个 transcripts，获得 93 327 个 unigenes，其中 unigenes 的总长度为 163 878 269 bp，平均长度为 1756 bp，N50 为 2333 bp。unigenes 长度分布图（图 2）显示，63 142 条 unigenes 长度超过 1000 bp，31 407 条 unigenes 长度超过 2000 bp。以上结果表明本研究转录组测序及组装完整性较高，能够进行后续的注释分析。转录组数据已上传至 NCBI 公共数据库，登录号为 SRR14567298、SRR14567299。

3.3 东紫苏转录组 unigenes 的功能注释

使用 BLAST 将拼装所得的 unigenes 分别与 NR、NT、KEGG、Swiss-prot、PFAM、GO、KOG 这 7 个数据库进行比对分析，对各数据库注释的 unigenes 数目进行统计，进而获得东紫苏转录组 unigenes 的功能注释信息，比对结果见表 3。结果表明，80 420 条 unigenes（86.17%）在 NR 数据库

表 2 东紫苏转录组数据产出质量

Table 2 Quality of output of *E. bodinieri* transcriptome data

样本	Raw reads/条	Clean reads/条	Clean bases/Gb	Q20/%	Q30/%	GC/%
SM	50 307 992	49 286 500	7.39	94.67	86.81	47.99
AZY	43 223 510	41 882 856	6.28	94.62	86.76	47.17

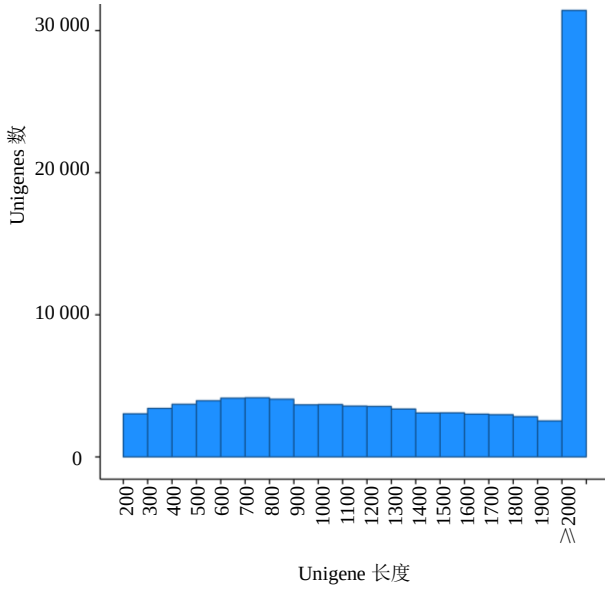


图 2 东紫苏转录组 unigenes 长度分布图

Fig. 2 Length distribution of unigenes in transcriptome of *E. bodinieri*

表 3 东紫苏转录组 unigenes 在各数据库注释情况

Table 3 *E. bodinieri* transcriptome unigenes annotated in each database

数据库	注释数目	占 unigene 百分比/%
NR	80 420	86.17
NT	65 009	69.65
KEGG	34 354	36.81
Swiss-Prot	64 873	69.51
PFAM	61 057	65.42
GO	61 194	65.56
KOG	25 391	27.20
Annotated in all databases	15 115	16.19
Annotated in at least one database	83 704	89.68
Total unigenes	93 327	100.00

比对成功得到注释，在 KEGG、Swiss-port、GO、KOG 等数据库获得注释的 unigenes 数目依次为 34 354 条（36.81%）、64 873 条（69.51%）、61 194 条（65.56%）、25 391 条（27.2%）。15 115 条 unigenes 同时在所有数据库中注释，至少在 1 个数据库注释成功的 unigenes 有 83 704 条（89.68%）。

通过与 NR 数据库进行比对注释，可以获取本物种基因序列与近缘物种基因序列的相似性，以及本物种基因的功能信息。从图 3 物种分布图可以看出与其他物种序列的相似性，在比对到 NR 数据库中的 unigenes 中，有 65.2%和芝麻 *Sesamum indicum* L.相匹配，其次是黄色猴面花 *Erythranthe guttata* (DC.) G. L. Nesom.，匹配度为 13.5%，丹参 *Salvia miltiorrhiza* Bunge、中粒咖啡 *Coffea canephora* Pierre ex Froehn.、葡萄 *Vitis vinifera* L.的匹配度分别为 1.6%、1.3%、1.2%，其他物种为 17.3%。

对基因进行 GO 注释之后，共有 61 194 条（65.56%）unigenes 注释到 GO 数据库，并根据 GO 功能分别将它们注释到生物过程（biological process）、细胞组分（cellular component）和分子功能（molecular function）3 大类中，共分为 56 个小组，注释结果见图 4。其中生物过程主要聚集在细胞过程（cellular process）、代谢过程（metabolic process），涉及的基因分别有 36 783、34 872 条；其次是单组织过程（single-organism process）和生物调节（biological regulation），基因数量分别为 27 213、12 627 条。细胞组分中细胞（cell）和细胞部分（cell part）相关基因数量较多，有 19 561 和 19 559 条，其次是细胞器（organelle），有 13 142 条。分子功能中具有结合功能（binding）和催化活性（catalytic activity）的基因数量较多，分别为 37 797 和 30 440 条，其他类别的基因数目普遍较少。

为了进一步分析东紫苏 unigenes 的功能，进行了 KOG 功能分类，共有 25 391 条（27.2%）unigenes 注释到 KOG 数据库，得到 25 个不同的 KOG 功能类群（图 5）。其中翻译后修饰、蛋白反转和伴侣（posttranslational modification, protein turnover, chaperones）的数量最多，有 3509 条；其次是一般功能预测（general function prediction only）的基因，数量为 3458 条；而数量最少的为细胞运动性（cell motility），仅有 23 条。

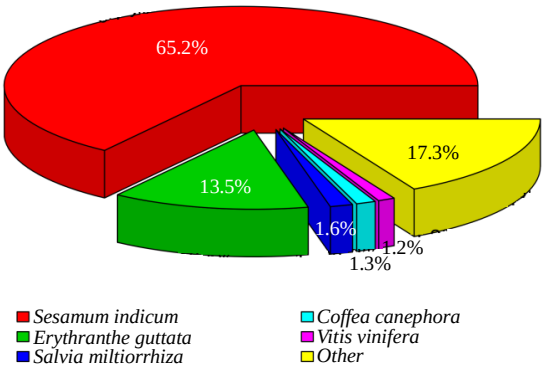


图 3　东紫苏转录组 unigenes 的物种分布图

Fig. 3　Species classification of unigenes in transcriptome of *E. bodinieri*

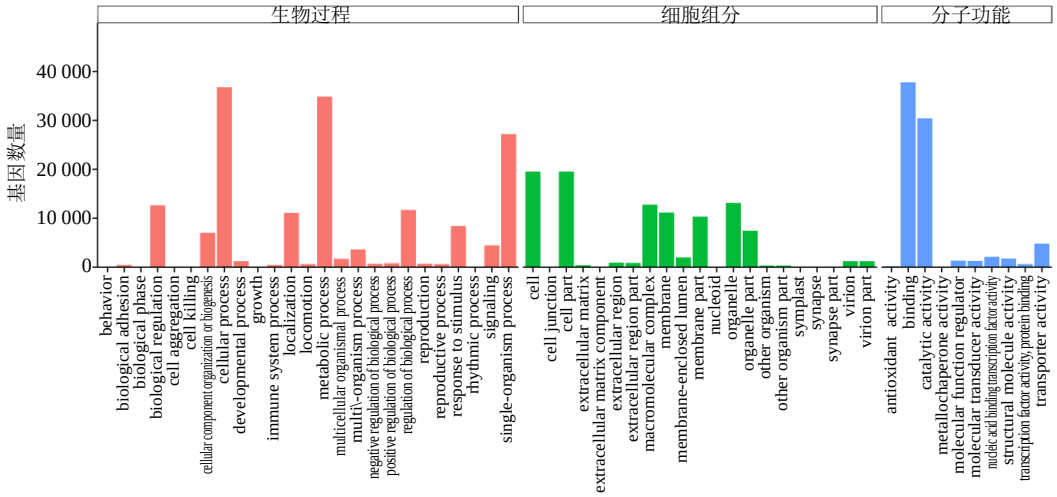


图 4　东紫苏转录组 unigenes 的 GO 功能分类注释

Fig. 4　GO functional classification annotation of unigenes of *E. bodinieri* transcriptome

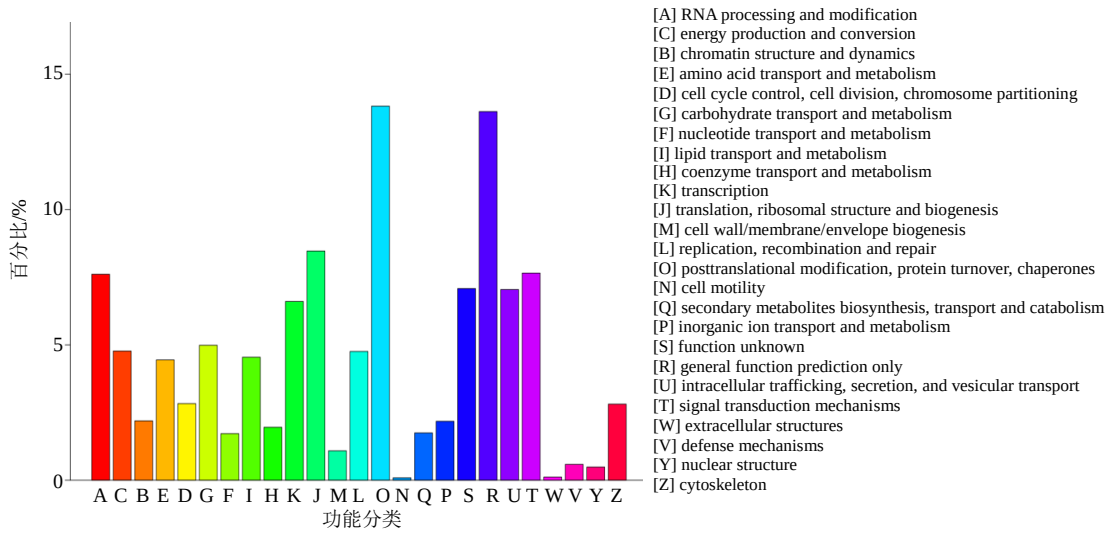


图 5　东紫苏转录组 unigenes 的 KOG 功能分类注释

Fig. 5　KOG functional classification annotation of unigenes of *E. bodinieri* transcriptome

将东紫苏 unigenes 与 KEGG 数据库比对，有 34 354 条（36.81%）unigenes 注释到该数据库中，分布于 129 条代谢通路中。KEGG 代谢通路（图 6）分为 5 大分支：细胞过程（cellular processes，A）1545 条、环境信息处理（environmental information processing，B）1299 条、遗传信息处理（genetic information processing，C）7448 条、代谢（metabolism，D）13 643 条和有机系统 E（organismal systems）1322 条。其中注释数量最多的是碳水化合物代谢（carbohydrate metabolism），为 2983 条，其次是折叠、分类和降解（folding，sorting and degardation）为 2687 条，数量最少的是膜运输（membrane transport），为 135 条。

3.4　东紫苏单萜代谢途径与相关基因

通过 KEGG 代谢通路分析，发现与东紫苏挥发油萜类生物合成相关的代谢通路有 4 条，分别是：萜类骨架生物合成，编号为 ko00900，相关基因有 218 条；单萜类生物合成，编号为 ko00902，相关基因有 30 条；柠檬烯和蒎烯降解，编号为 ko00903，相关基因有 38 条；倍半萜和三萜生物合成，编号

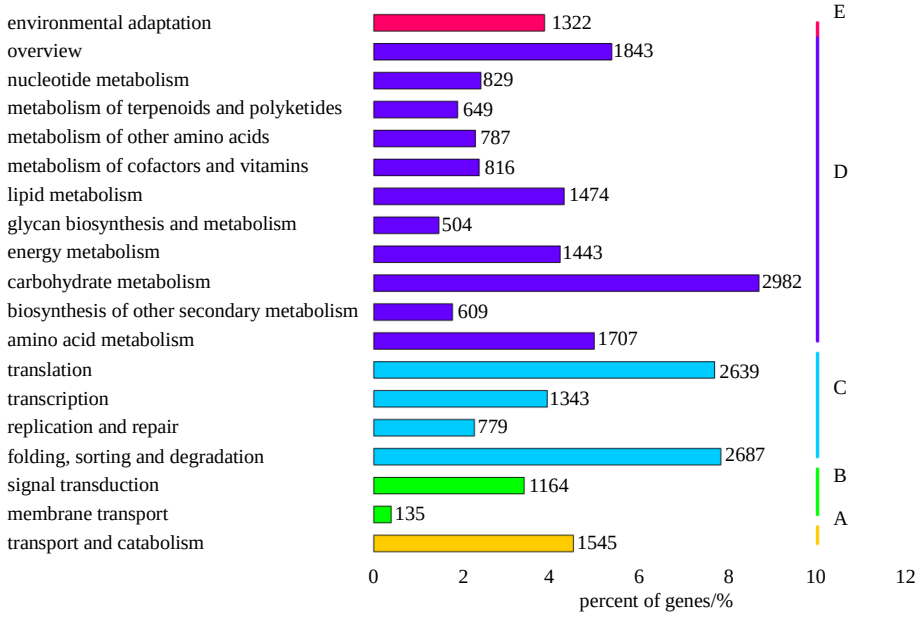


图 6　东紫苏转录组 unigenes 的 KEGG 分类注释

Fig. 6　KEGG classification annotation of unigenes of *E. bodinieri* transcriptome

为 ko00909，相关基因有 95 条。其中与东紫苏单萜合成相关的代谢通路是萜类骨架生物合成和单萜类生物合成。

与东紫苏单萜合成相关的酶分别有香叶基二磷酸合酶（geranyl diphosphate synthase，GPS），相关 unigenes 有 11 条；新薄荷醇脱氢酶[(+)-neomenthol dehydrogenase，NDS]，相关 unigenes 数量最多，有 22 条；*α*-松油醇合酶[(−)-alpha-terpineol synthase，ATS]，芳樟醇合酶[(3S)-linalool synthase，LIS]，月桂烯/罗勒烯合酶（myrcene/ocimene synthase，MS/OS），这 3 个酶相关 unigenes 数量较少，共有 8 条，具体见表 4。另外，通过比较这些 unigenes 的 FPKM 值可以看出，NDS 基因平均表达量是最高的，其中 SM 为 8.97，AZY 为 17.06；其次是 ATS 基因，平均表达量 SM 为 7.16，AZY 为 3.92；GPS 基因平均表达量 SM 为 2.29，AZY 为 1.64；LIS 基因和 MS/OS 基因平均表达量均偏低。

筛选得到的东紫苏 11 个单萜合成相关候选基因在 SM 和 AZY 2 个产地的 FPKM 值（图 7）。从图 7 中可以看到，基因 2（Cluster-10292.10951）注释为 NDS，在 2 个产地中的表达量均为最高，其中 SM 为 25.66，AZY 为 29.56；而基因 1（Cluster-11595.0）注释为 ATS，在 2 个产地中的表达量均较低，其中 SM 为 0.18，AZY 则为 0；基因 9（Cluster-10292.51149）注释为 NDS，基因 10（Cluster-10292.60790）和基因 11（Cluster-10292.60789）均注释为 GPS，这 3 个基因在 2 个产地中的表达量基本持平。

表 4 东紫苏挥发油单萜代谢途径的酶与相关基因

Table 4 Enzymes and related genes of monoterpene metabolic pathway of essential oil from *E. bodinieri*

代谢通路	通路编号	相关基因数量	关键酶	EC 编号	KO 编号	unigenes 数目	FPKM 平均值	
							SM	AZY
萜类骨架生物合成	ko00900	218	GPS	2.5.1.1	K14066	11	2.29	1.64
单萜类生物合成	ko00902	30	NDS	1.1.1.208	K15095	22	8.97	17.06
			ATS	4.2.3.111	K18108	3	7.16	3.92
			LIS	4.2.3.25	K15086	4	0.40	0.82
			MS/OS	4.2.3.15	K12467	1	0.19	0.12

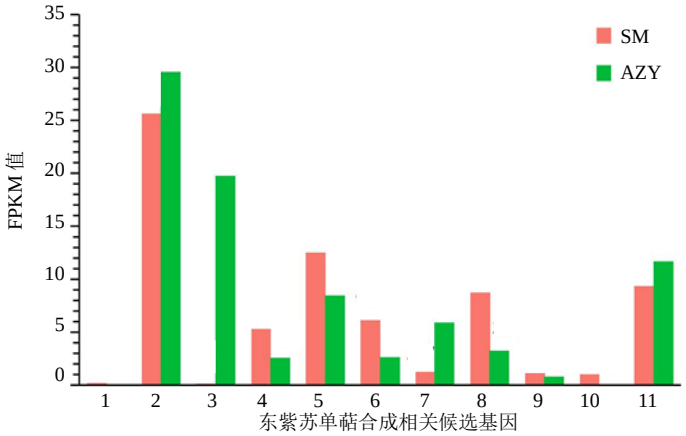
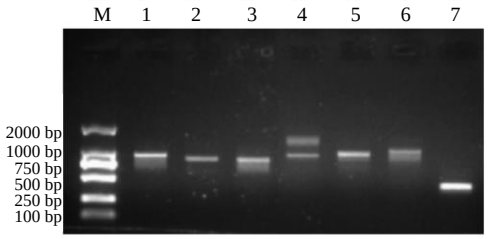


图 7 东紫苏单萜合成相关候选基因的表达量

Fig. 7 Expression of candidate genes related to monoterpene synthesis of *E. bodinieri*

3.5 东紫苏挥发油单萜合成相关候选基因全长 unigenes 的 RT-PCR 验证

利用表 1 中的 ORF 扩增引物进行 PCR 扩增，并进一步对 PCR 产物进行测序分析，得到以下结果：基因 2（Cluster-10292.10951）、基因 4（Cluster-10292. 60008）、基因 6（Cluster-10292.56890）、基因 7（Cluster-5267.0）、基因 10（Cluster-10292.60790）、基因 11（Cluster-10292.60789）这 6 个基因扩增后，经电泳分析，均产生明亮的目标条带（图 8），经测序分析后表明这 6 个基因与原 unigenes ORF 一致，这些候选 unigenes 为全长基因；另外基因 1（Cluster-11595.0）目前扩增出后半段，基因 5（Cluster-10292.9722）、8（Cluster-1426.0）这 2 个基因扩增出中间部分，下



M-Marker 1-2 号基因 2-4 号基因 3-6 号基因 4-7 号基因 5-10 号基因 6、7-11 号基因

M-Marker 1-2 gene 2-4 gene 3-6 gene 4-7 gene 5-10 gene 6,7-11 gene

图 8 东紫苏挥发油 6 个单萜合成相关候选基因 RT-PCR 凝胶电泳图

Fig. 8 RT-PCR gel electrophoresis of six candidate genes related to monoterpene synthesis in essential oil from *E. bodinieri*

一步将继续尝试全长扩增；基因 3（Cluster-10292.50707）、基因 9（Cluster-10292.51149）这 2 个基因扩增失败。

4 讨论

东紫苏作为云南特产物种，在云南主要用作民间药物及保健饮品[28]，其作为药用植物的基因资源部分一直是空白。本研究采用 Illumina Hiseq 2000 高通量测序平台，首次对东紫苏进行转录组测序分析，填补了东紫苏转录组数据信息的空白。通过测序共获得 13.67 Gb 数据，91 169 356 条高质量序列（clean reads），Trinity 组装获得 93 327 条 unigenes，平均长度为 1756 bp。

结合生物信息学分析方法对东紫苏转录组数据进行序列相似性、基因注释和功能分类，将所有 unigenes 分别与 NR、NT、KEGG、Swiss-prot、PFAM，GO、KOG 等数据库进行比对，经过 BLAST 比对分析后发现，有 80 420 条 unigenes 与其他近缘物种的已知基因具有相似性，与芝麻相似度最高，为 65.2%。在 KEGG 数据库比对中，有 34 354 条（36.81%）unigenes 注释到该数据库中，分布于 129 条代谢通路中，并发现 4 条萜类化合物代谢相关的途径。

RNA-Seq 技术的发展为转录组研究提供了机会，已成为发现与药用植物各种次生代谢途径生物合成相关基因的有力工具，即使是在没有参考基因组的物种中也是如此[29-30]。本研究是通过已注释基因的挖掘，在无参转录组中筛选目的基因，因与芝麻同源性匹配度最高，因此以芝麻作为参照来寻找目的基因。

前期课题组对东紫苏 15 个居群叶的挥发油进行 GC-MS 分析[12]，并根据主成分的差异，将其划分为不同的化学型，其中 SM 属于 T 型（1,8-桉叶油醇＜5%，5%＜α-乙酸松油酯＜20%），阿子营（AZY）属于 C 型（1,8-桉叶油醇>50%，α-乙酸松油酯＜5%），与之呼应的是 2 个个体在单萜合成相关候选基因表达量上的明显差异。已有研究表明与植物挥发油中高含量成分对应的单萜合酶基因其克隆成功率更高，主要原因可能在于基因表达量与其产物的积累量是一致的[31-34]，由于 2 个产地的东紫苏叶挥发油主成分差异明显，具有代表性，因此分析这 2 个产地的东紫苏叶在转录组数据中相关基因的表达量，有可能提高单萜合酶基因克隆的成功率。基于此，本研究在明确东紫苏挥发油主要成分的基础上，筛选出与单萜合成相关的代谢途径 2 条，单萜合成相关候选基因 41 个，通过 RT-PCR 验证其中的 6 个 unigenes 为全长基因，这为下一步研究东紫苏单萜生物合成及基因功能验证提供数据支持。

本研究筛选出的单萜合酶基因克隆成功率低，并且转录组注释的功能信息与东紫苏挥发油 GC-MS 数据分析结果不吻合，即注释的功能与挥发油的成分并不匹配，推测是受测序技术所限，以及数据库可供比对的基因资源有限，增加了功能注释的难度，致使一些东紫苏挥发油单萜合成相关的关键酶基因未能得到精准拼接、组装和注释。此外，由于相同物种来源的单萜合酶的相似性高于不同物种来源而具有相同功能的单萜合酶，序列相关性并不直接转化为产物相关性[23,33,35-38]，因此基于序列相似性仅可以判断其是否为单萜合酶，却不能预测其产物，所以已经克隆成功的 6 个单萜合成相关候选基因需要进行功能验证，确认其在东紫苏挥发油单萜生物合成过程中的催化功能，才能确定该酶属于哪一种单萜合酶，以便最终得到东紫苏 1,8-桉叶油醇合酶基因。

本研究首次采用高通量测序技术建立了东紫苏转录组数据库，获得了大量的转录本信息，并挖掘出单萜生物合成途径的关键酶基因，填补了东紫苏挥发油单萜合成途径分子研究的空白，为东紫苏挥发油单萜合酶基因的克隆、功能验证以及品种选育提供了基因资源及理论指导基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 中国科学院中国植物志编辑委员会. 中国植物志 (第 66 卷) [M]. 北京: 科学出版社, 1977: 336-338.

[2] 云南省食品药品监督管理局. 云南省中药材标准-第一册: 2005 年版 分辑号: 第一册 [M]. 昆明: 云南美术出版社, 2005: 82-83.

[3] 普春霞. 唇形科香薷属的系统演化与分类修订 [D]. 北京: 中国科学院大学, 2012.

[4] 毛媛, 刘克武, 刘克海, 等. 东紫苏化学成分研究概述 [J]. 中国中医药信息杂志, 2016, 23(3): 134-136.

[5] 胡浩斌, 王鑫, 刘建新, 等. 东紫苏根中抑菌活性成分的研究 [J]. 四川大学学报: 自然科学版, 2006, 43(4): 913-917.

[6] 段燕玲. 九种清热解毒类中草药抑菌、抗 HCV 活性研究 [D]. 昆明: 昆明理工大学, 2016.

[7] 钟金栋. 两种药用植物化学成分和生物活性研究 [D]. 昆明: 昆明理工大学, 2013.

[8] 张举成, 刘卫, 严和平, 等. 野山茶抗氧化活性研究及微量元素测定 [J]. 食品科技, 2007, 32(4): 206-207.

[9] 杨玉霞. 黑沙蒿和凤尾茶中二氢黄酮类化合物及其抗自由基活性研究 [D]. 开封: 河南大学, 2005.

[10] 吴永贵, 唐正平. 东紫苏的部分药效学实验研究 [J]. 现代中药研究与实践, 2009, 23(1): 51-53.

[11] Thanaseelungkoon N, Julsrigival J, Phannachet K, *et al*. Chemical compositions and biological activities of essential oils obtained from some *Apiaceous* and *Lamiaceous* plants collected in Thailand [J]. *Asian Pac J Trop Med*, 2018, 11(8): 486-494.

[12] 唐仁华, 杨礼攀, 张红霞, 等. 云南省 15 个居群的东紫苏叶挥发油化学型研究 [J]. 中药材, 2020, 43(4): 887-891.

[13] 付立卓, 李海舟, 李蓉涛. 2 种香薷属植物挥发油成分分析 [J]. 昆明理工大学学报: 理工版, 2010, 35(1): 88-92.

[14] 程伟贤, 古昆, 李聪, 等. 云南凤尾茶精油成分分析 [J]. 云南民族大学学报: 自然科学版, 2004, 13(2): 86-87.

[15] 罗燕, 王文静, 张红霞, 等. 东紫苏昆明居群叶特征与挥发油的相关性研究 [J]. 时珍国医国药, 2017, 28(5): 1216-1219.

[16] Santos F A, Rao V S. Antiinflammatory and antinociceptive effects of 1, 8-cineole a terpenoid oxide present in many plant essential oils [J]. *Phytother Res*, 2000, 14(4): 240-244.

[17] Bastos V P, Gomes A S, Lima F J, *et al*. Inhaled 1, 8-cineole reduces inflammatory parameters in airways of ovalbumin-challenged Guinea pigs [J]. *Basic Clin Pharmacol Toxicol*, 2011, 108(1): 34-39.

[18] Juergens U R, Engelen T, Racké K, *et al*. Inhibitory activity of 1,8-cineol (eucalyptol) on cytokine production in cultured human lymphocytes and monocytes [J]. *Pulm Pharmacol Ther*, 2004, 17(5): 281-287.

[19] Juergens U R. Anti-inflammatory properties of the monoterpene 1,8-cineole: Current evidence for co-medication in inflammatory airway diseases [J]. *Drug Res*, 2014, 64(12): 638-646.

[20] Worth H, Dethlefsen U. Patients with asthma benefit from concomitant therapy with cineole: A placebo-controlled, double-blind trial [J]. *J Asthma*, 2012, 49(8): 849-853.

[21] Hendry E R, Worthington T, Conway B R, *et al*. Antimicrobial efficacy of *Eucalyptus oil* and 1,8-cineole alone and in combination with chlorhexidine digluconate against microorganisms grown in planktonic and biofilm cultures [J]. *J Antimicrob Chemother*, 2009, 64(6): 1219-1225.

[22] Jalilzadeh-Amin G, Maham M. The application of 1,8-cineole, a terpenoid oxide present in medicinal plants, inhibits *Castor* oil-induced diarrhea in rats [J]. *Pharm Biol*, 2015, 53(4): 594-599.

[23] Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis [J]. *Proc Natl Acad Sci USA*, 1998, 95(8): 4126-4133.

[24] Tholl D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism [J]. *Curr Opin Plant Biol*, 2006, 9(3): 297-304.

[25] Gao Y, Honzatko R B, Peters R J. Terpenoid synthase structures: A so far incomplete view of complex catalysis [J]. *Nat Prod Rep*, 2012, 29(10): 1153-1175.

[26] Xiao M, Zhang Y, Chen X, *et al*. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest [J]. *J Biotechnol*, 2013, 166(3): 122-134.

[27] Chen C H, Zheng Y J, Zhong Y D, *et al*. Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora* [J]. *BMC Genomics*, 2018, 19(1): 550.

[28] 钟金栋, 冯峰, 李洪梅, 等. 东紫苏的化学成分研究 [J]. 昆明理工大学学报: 自然科学版, 2013, 38(5): 75-79.

[29] Ward J A, Ponnala L, Weber C A. Strategies for transcriptome analysis in nonmodel plants [J]. *Am J Bot*, 2012, 99(2): 267-276.

[30] Dillies M A, Rau A, Aubert J, *et al*. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis [J]. *Brief Bioinform*, 2013, 14(6): 671-683.

[31] Benabdelkader T, Guitton Y, Pasquier B, *et al*. Functional characterization of terpene synthases and chemotypic variation in three lavender species of section Stoechas [J]. *Physiol Plant*, 2015, 153(1): 43-57.

[32] Piechulla B, Bartelt R, Brosemann A, *et al*. The α-terpineol to 1,8-cineole cyclization reaction of tobacco terpene synthases [J]. *Plant Physiol*, 2016, 172(4): 2120-2131.

[33] Despinasse Y, Fiorucci S, Antonczak S, *et al*. Bornyl-diphosphate synthase from *Lavandula angustifolia*: A major monoterpene synthase involved in essential oil quality [J]. *Phytochemistry*, 2017, 137: 24-33.

[34] Radwan A, Kleinwächter M, Selmar D. Impact of drought stress on specialised metabolism: Biosynthesis and the expression of monoterpene synthases in sage (*Salvia officinalis*) [J]. *Phytochemistry*, 2017, 141: 20-26.

[35] Landmann C, Fink B, Festner M, *et al*. Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*) [J]. *Arch Biochem Biophys*, 2007, 465(2): 417-429.

[36] 徐应文, 吕季娟, 吴卫, 等. 植物单萜合酶研究进展 [J]. 生态学报, 2009, 29(6): 3188-3197.

[37] Demissie Z A, Cella M A, Sarker L S, *et al*. Cloning, functional characterization and genomic organization of 1,8-cineole synthases from *Lavandula* [J]. *Plant Mol Biol*, 2012, 79(4/5): 393-411.

[38] Borghi M, Xie D Y. Cloning and characterization of a monoterpene synthase gene from flowers of *Camelina sativa* [J]. *Planta*, 2018, 247(2): 443-457.

[责任编辑　时圣明]