

基于高通量测序的瑞香狼毒转录组数据分析

杨艳芳¹, 刘洪伟², 邱德有^{1*}

1. 中国林业科学研究院林业研究所, 林木遗传育种国家重点实验室 国家林业局林木培育重点实验室, 北京 100091

2. 河北省科学院生物研究所, 河北省主要农作物病害微生物控制工程技术研究中心, 河北 石家庄 050081

摘要: **目的** 获得瑞香狼毒 *Stellera chamaejasme* 转录组数据库代谢途径基因序列、SSR 以及转座子等信息。**方法** 以瑞香狼毒根作为受试材料, 采用二代测序方法中的 Illumina HiSeq 2000 进行转录组测序, 并进行系统的生物信息学分析。**结果** 共获得 26 785 872 个 Clean reads 片段, 拼接得到 47 053 条 Unigenes, 平均长度为 419 nt。将拼装所得到的 Unigene 序列利用 BLAST 工具分别与 Nr、Swiss-Prot、KEGG、COG 和 GO 数据库进行比对, 分别有 11 138 和 24 744 条 Unigene 在 Nr 和 Swiss-Prot 数据库中比对得到了注释信息, 可归于 36 个 GO 分类, 涉及 119 个 KEGG 标准代谢通路, 进一步分析发现 15 条萜类生物合成途径的关键酶基因。利用 MISA 软件发现 3 480 个 SSR, 数量最高的 SSR 类型为单碱基重复, 为 1 986 条, 出现频率为 57.07%, 最少的是六碱基重复 SSR, 只有 5 条, 出现频率仅为 0.14%。利用 RepeatMasker 在线工具针对瑞香狼毒转录组序列进行转座子预测分析, 结果共发现有 1 497 条转座子, 其中 E 值 $< 1 \times 10^{-5}$ 的序列有 827 条, 包含 22 种类型转座子, 数目最多的为 LINE/L1 类型 (405 条), 占比为 48.97%, 占比最少的为 DNA/Ginger、DNA/hAT、DNA/PIF-ISL2EU 和 LINE/Jockey 以及 LTR/Lenti 类型分别只有 1 条。**结论** 对瑞香狼毒进行高通量测序, 获得了大量基因序列信息以及 SSR 和转座子信息, 为今后分离克隆瑞香狼毒中佛波酯等有效成分生物合成的关键酶基因以及开展相关分子机制研究提供了数据资源和理论基础。

关键词: 瑞香狼毒; 转录组; 萜类; 佛波酯; 代谢通路

中图分类号: R282.12

文献标志码: A

文章编号: 0253-2670(2017)22-4740-08

DOI: 10.7501/j.issn.0253-2670.2017.22.024

Transcriptome characterization of *Stellera chamaejasme* with Illumina sequencing technology

YANG Yan-fang¹, LIU Hong-wei², QIU De-you¹

1. Research Institute of Forestry, Chinese Academy of Forestry, State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of State Forestry Administration, Beijing 100091, China

2. Biology Institute, Hebei Academy of Sciences, Hebei Engineering and Technology Center of Microbiological Control on Main Crop Disease, Shijiazhuang 050081, China

Abstract: Objective To obtain the transcriptome database and gene sequence, SSR as well as transposon information of *Stellera chamaejasme*. **Methods** Using the high-throughput sequencing platform (Illumina HiSeq 2000), a root transcriptome dataset of *S. chamaejasme* was obtained, and the sequencing results were analyzed with the bioinformatic way. **Results** With a total of 26 785 872 clean reads, 47 053 unigenes were assembled. All these unigenes were then blasted with Nr, Swiss-Prot, KEGG, COG, and GO databases. There were 11 138 and 24 744 unigenes were annotated with Nr and Swiss-Prot databases, respectively. The unigenes were involved in 36 GO-terms and 119 metabolic pathways. Further analysis showed that 15 unigenes were involved in terpenoids biosynthesis. Using MISA software, the results showed that there were 3 480 SSR from the 47 053 unigenes, and the most type of SSR was mononucleotide (1 986) with the frequency of 57.07%. Moreover, the hexanucleotide only had five repeat SSR and the frequency was only 0.14%. With RepeatMasker online tools to analyze the transposon of the transcriptome sequences, the results indicated that there were 1 497 transposons, and the number of transposons with $E < 1 \times 10^{-5}$ was 827. All the transposons were grouped into 22 types, and the LINE/L1 type (405) had the highest frequency (48.97%). The DNA/Ginger, DNA/hAT, DNA/PIF-ISL2EU, and LINE/Jockey as well as LTR/Lenti were the least type since each of them has only one transposon. **Conclusion** In this study, rich

收稿日期: 2017-07-18

基金项目: 中央级公益性科研院所基本科研业务费专项 (CAFYBB2014QB001); 国家自然科学基金项目 (31570675)

作者简介: 杨艳芳 (1978—), 女, 副研究员, 研究方向为林木次生代谢。Tel: (010)62889638 E-mail: echoyyf@caf.ac.cn

*通信作者 邱德有, 男, 研究员, 博士生导师, 主要从事植物次生代谢研究。Tel: (010)62889641 E-mail: qiudy@caf.ac.cn

sequence information of gene, SSR as well as transposon information of *Stellera chamaejasme* is helpful to carry out the research of the molecular mechanism of phorbol ester biosynthesis in *S. chamaejasme* in the future.

Key words: *Stellera chamaejasme* L.; transcriptome; terpenoid; phorbol ester; metabolic pathway

瑞香狼毒 *Stellera chamaejasme* L. 为瑞香科狼毒属多年生草本植物, 俗称断肠草、馒头花和红狼毒等。瑞香狼毒其根入药, 性味苦平, 有大毒, 《神农本草经》记载, 具有逐水祛痰、破积杀虫之功效。目前, 已有研究者对于瑞香狼毒的致毒原理展开了深入研究, 发现拥有较强的耐逆境能力, 且释放的多种化学成分能够对其他植物和微生物产生化感作用^[1-2]。这些化学成分主要有黄酮类、香豆素类、萜类、木脂素类和苯丙烯酸醇苷类等^[3-5], 它们还具有抑制肿瘤细胞生长、抗菌、抗病毒等作用, 尤其是二萜类物质具有良好的抗癌和抗艾滋病病毒活性^[6], 因此具有重要的药用价值和临床意义。

佛波醇及其酯类是一种四环二萜类化合物, 由于具有强大的抗癌活性而备受关注。例如, 从瑞香科植物平卧稻花 *Pimelea prostrata* Guill. 中分离得到的二萜类佛波醇类似物 prostratin^[7], 它不仅能促使体细胞减少分泌艾滋病病毒受体, 使艾滋病感染率下降; 而且它还可以逼出隐藏在免疫细胞内的艾滋病病毒^[8], 而且没有佛波醇及其酯类所具有的强烈致癌毒性, 极有希望开发成治疗艾滋病的新药^[9-11]。目前, 瑞香狼毒中有关药用成分分离鉴定以及药理作用方面的研究较为常见, 甚至已经有研究者利用化学合成方法仅仅用了 19 步就合成了佛波醇^[12], 而从巴豆油的佛波醇, 只需要经过 5 步就能成功地合成出 prostratin 及其类似物^[13]。但是, 有关 prostratin 等这些萜类、黄酮类等药用成分生物合成分子机制的研究较为少见。

转录组学 (transcriptomics) 是功能基因组学研究的重要组成部分, 能够在整体水平上研究特定条件下细胞中所有基因转录及转录调控规律。近几年, 高通量测序技术发展迅猛, 为在多个物种中利用转录组学技术探索获取基因序列、功能和转录表达模式等研究奠定了基础。本研究以瑞香狼毒为研究对象, 利用其根部组织为材料, 进行了转录组测序研究, 对获得的测序数据进行了组装、拼接和注释工作, 并初步发掘了黄酮类、二萜类等化合物代谢途径相关基因, 为今后瑞香狼毒相关药效成分基因的克隆及生物合成分子机制研究提供了基因资源和基础。

1 材料与方法

1.1 材料

植株种植于中国林业科学研究院温室中, 经中国林业科学研究院林业研究所邱德有研究员鉴定为瑞香科植物瑞香狼毒 *Stellera chamaejasme* L., 取单株植株的根为材料用于提取总 RNA, 构建转录组文库。植物总 RNA 提取试剂盒和 cDNA 反转录试剂盒购自北京天根生化科技 (北京) 有限公司。

1.2 转录组数据的组装与分析

1.2.1 文库构建 提取瑞香狼毒根的总 RNA, 用带有 Oligo(dT) 的磁珠富集 mRNA。加入 fragmentation buffer 将 mRNA 打断成短片段, 以 mRNA 为模板, 用六碱基随机引物 (random hexamers) 合成第一条 cDNA 链, 然后加入缓冲液、dNTPs、RNase H 和 DNA polymerase I 合成第 2 条 cDNA 链, 在经过 QiaQuick PCR 试剂盒纯化并加 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, 最终建好的测序文库用 Illumina HiSeqTM 2000 进行测序。

1.2.2 转录组数据组装 经除杂和冗余处理, 将测序所获得 clean reads 利用短 reads 组装软件 SOAPdenovo^[14] 进行从头组装, 得到 Contig, 然后再将 reads 比对回 Contig, 通过 paired-end reads 能确定来自同一转录本的不同 Contig 以及这些 Contig 之间的距离, 得到 Scaffold。进一步利用 paired-end reads 对 Scaffold 做补洞处理, 最后得到 Unigene。

1.2.3 转录组功能预测 利用 BlastX 将 Unigene 序列与蛋白数据库 Nr、Swiss-Prot、KEGG 和 COG 进行比对 (E 值 $< 1 \times 10^{-5}$), 得到与相应 Unigene 具有最高序列相似性的蛋白, 进而得到 Unigene 注释信息。

1.2.4 蛋白编码框 (CDS) 预测 按照 Nr、Swiss-Prot、KEGG 和 COG 的优先级与以上蛋白库进行 BlastX 比对 (E 值 $< 1 \times 10^{-5}$) 并确定该 Unigene 的编码区序列和方向, 然后根据标准密码子表将此编码区序列翻译成氨基酸序列, 由此得到该 Unigene 编码区的核酸序列 (序列方向 5'→3') 和氨基酸序列。利用 ESTScan^[15] 预测与以上数据库皆

比对不上的 Unigene 的编码区及序列方向。

1.2.5 SSR 特征检测 利用简单重复序列 (simple sequence repeat, SSR) 软件 MicroSATellite (MISA) 检测瑞香狼毒转录组组装得到的全部 Unigene, 进而找到所有的 SSR, 并对其进行统计分析。

1.2.6 转座子预测 将拼接后的瑞香狼毒 Unigene 序列提交到 Repeatmasker 网站, 利用在线工具 Protein-based Repeatmasking (<http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest>) 进行转座子预测。

2 结果与分析

2.1 瑞香狼毒转录组组装与质量分析

利用 Illumina 平台对瑞香狼毒的根进行转录组测序后, 共检测得到 26 785 872 个 Clean reads 片段。

在去除冗余序列后, 所获得的转录组测序结果共组装拼接出 101 723 个 Contigs, 平均长度为 230 nt, N50 为 289 nt, 最长达到了 7 649 nt。利用已有的 Contigs 数据进一步运用 paired-end joining and gap-filling 的方法, 将这些 Contig 组装成了 68 418 个 Scaffolds, 平均长度为 327 nt, 其中 3 112 个 Scaffolds 长度高于 1 000 nt。由这 68 418 个 Scaffolds 进而组装出 47 053 个 Unigenes, 平均长度为 419 nt, N50 为 499 nt。

对组装得到的 Contig 片段和 Unigene 片段的长度进行分析, 得到其序列长度分布图 (图 1)。结果显示, 序列共有 2, 534 条 Contig 序列长度超过了 1 000 nt (图 1-A), 而在组装出的 Unigene 中, 则有 3 874 条序列长度超过了 1 000 nt (图 1-B)。

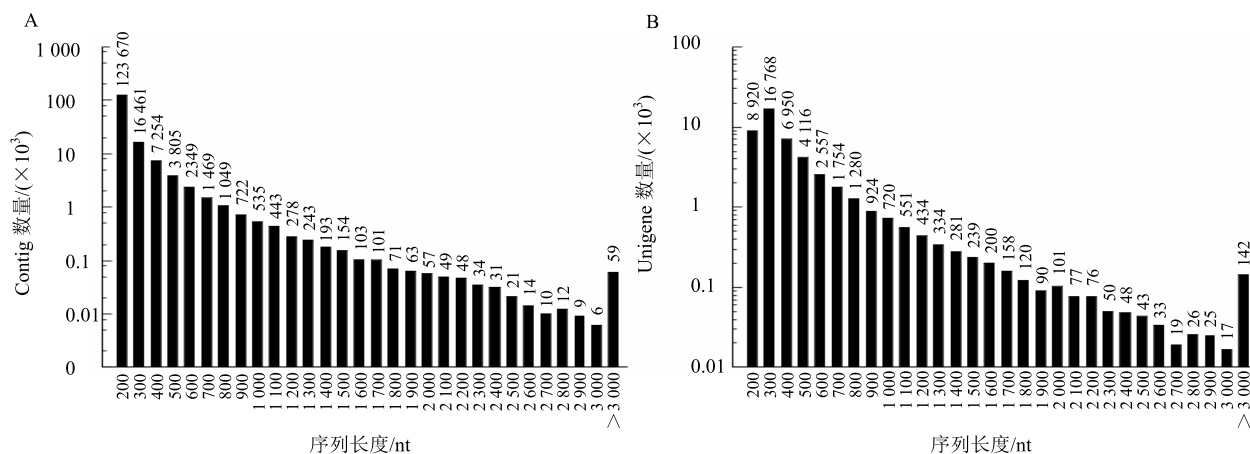


图 1 Contig (A) 和 Unigene (B) 序列长度分布

Fig. 1 Length distribution of Contig (A) and Unigene (B)

2.2 瑞香狼毒转录组 Unigene 的功能注释

将拼装所得到的 47 053 个 Unigene 序列分别与 Nr、Swiss-Prot、KEGG、COG 和 GO 数据库进行 Blast 比对, 并对每一个数据库中所注释得到的 Unigene 数目进行统计。通过注释分析得到瑞香狼毒的 Unigene 的功能注释信息。结果表明, 通过比对共有 11 138 条 Unigene (阈值 E 值 $< 1 \times 10^{-5}$) 在 Nr 数据库中比对得到了基因注释信息, 占总基因数的 23.7%, 在 Swiss-Prot 数据库中共有 24 744 条基因获得注释信息, 占基因总数的 52.6%。

将在 Nr 库中比对得到的 top hits 进行 E 值分布分析, 发现大约 43.3% 的 E 值位于 $1 \times 10^{-5} \sim 1 \times 10^{-30}$, 34.5% 的 Unigene 的 E 值在 $1 \times 10^{-5} \sim 1 \times 10^{-10}$; 而 E 值处于 $1 \times 10^{-30} \sim 1 \times 10^{-50}$ 和 $1 \times 10^{-50} \sim 1 \times 10^{-100}$ 的分别为 10.8% 和 8.4% (图 2-A)。进一

步进行相似度分析, 结果显示 53.5% 的 Unigene 的相似度在 30%~50%, 另外, 还有 15.7% 和 13.5% 的 Unigene 相似度分别位于 50%~60% 和 60%~80% (图 2-B)。此外, 从物种分布图可以看出, 在比对到 Nr 数据库中的 Unigene 中分别有 14.4%、12.1% 和 5.6% 与水稻 *Oryza sativa* Linn.、拟南芥 *Arabidopsis thaliana* (Linn.) Heynh. 和玉米 *Zea mays* Linn. 相匹配, 另外值得注意的是, 还与致病疫霉 *Phytophthora infestans* (Mont.) de Bary、长囊水云 *Ectocarpus siliculosus* (Dillw.) Lyngb 和莱茵衣藻 *Chlamydomonas reinhardtii* 分别具有 8.3%、7.3% 和 6% 的相似度 (图 2-C)。

Gene ontology (GO) 在生物信息学领域内广泛应用于对基因和蛋白质功能进行限定和描述, 主要包括 3 个分支: 生物过程 (biological process, BP)、

分子功能 (molecular function, MF) 和细胞组件 (cellular component, CC)。本研究获得的瑞香狼毒的 Unigene 被归纳成 36 个功能群组 (图 3), 且分别有 2 800、3 997 和 2 415 个 Unigenes 注释被分在生物过程、细胞组分和分子功能 3 大分支中。其

中, 有 824 条序列注释在“代谢进程”条目中, 与其他条目相比, 占比为 28.61% (图 3)。

为了进一步分析瑞香狼毒 Unigene 的功能, 研究组进行了 COG (cluster of orthologous groups) 功能分类分析, 共得到 25 个不同的功能类群 (图 4)。

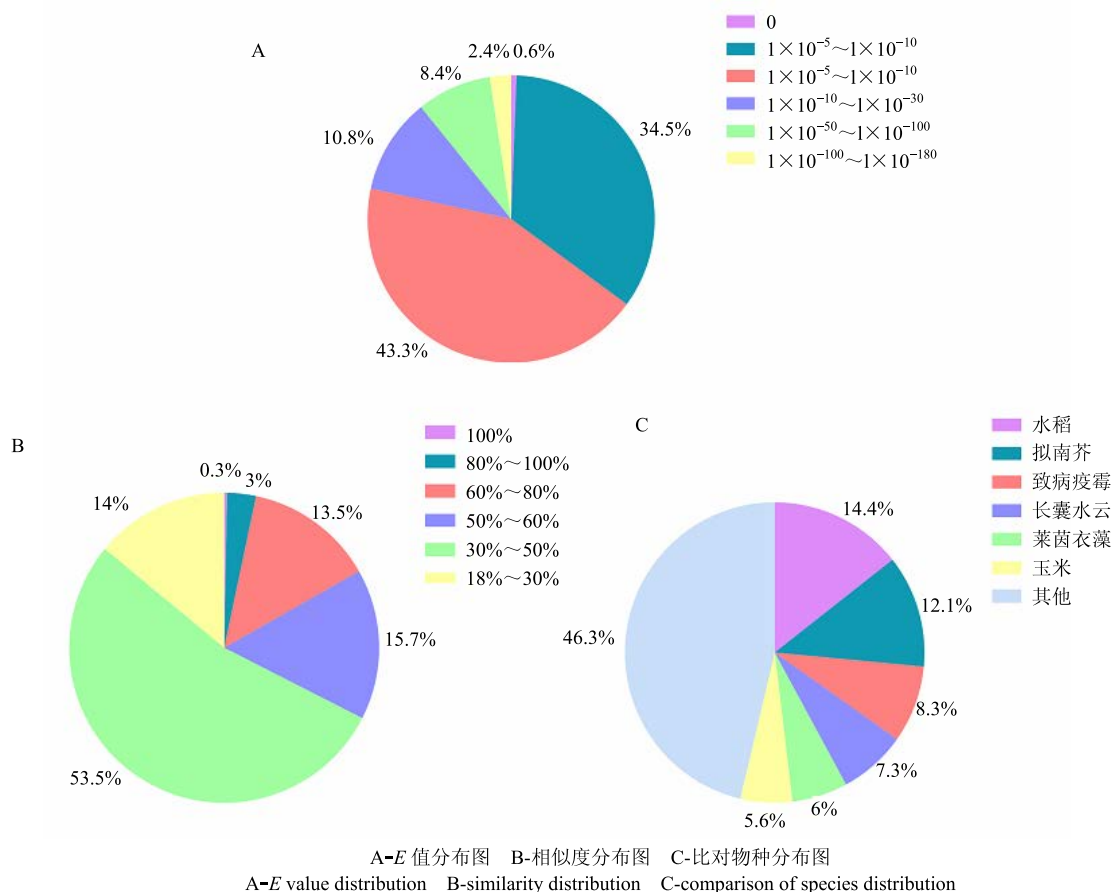


图 2 瑞香狼毒 Unigene 比对 Nr 数据库结果

Fig. 2 Statistics of homology search of transcripts against Nr peptide database

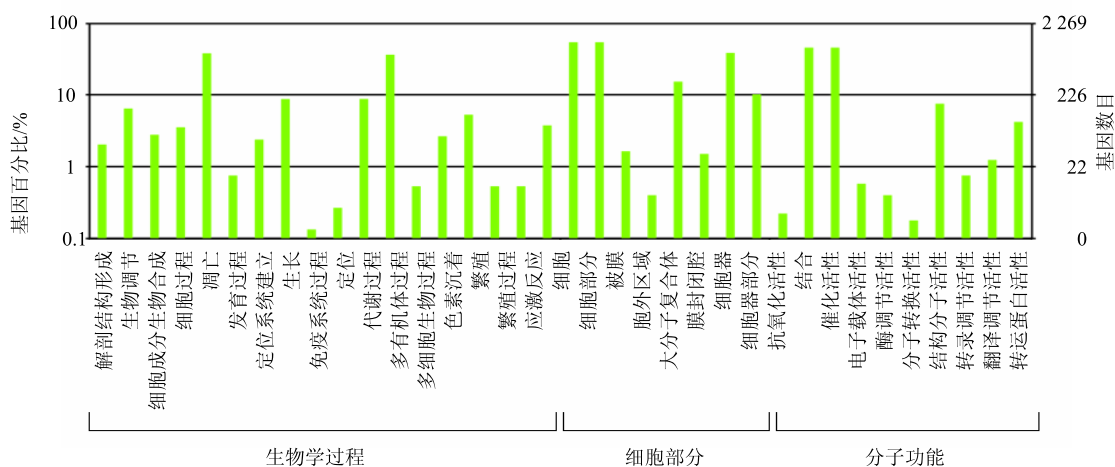


图 3 瑞香狼毒的 Unigene GO 功能分类图

Fig. 3 GO classification of unigenes

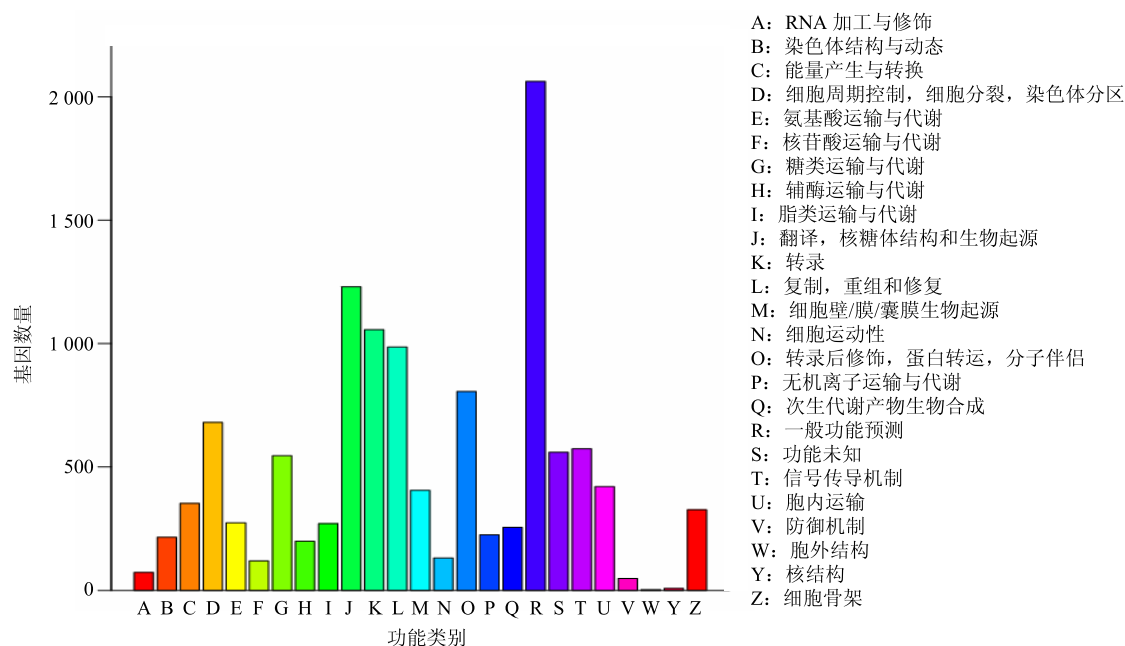


图 4 瑞香狼毒 Unigene 的 COG 功能分类

Fig. 4 COG function classification of *S. chamaejasme unigenes*

在这些类群中, 最大的类群为“一般功能预测类”(2 062 个, 17.43%), 最小的类群为“胞外结构”(3 个, 0.03%)。其他类群中, “翻译、核糖体结构和生物起源”(1 230 个, 10.40%)、“转录”(1 056 个, 8.93%)、“复制、重组和修复”(986 个, 8.34%), 所占比例也比较高。此外, “次生代谢产物生物合成、运输和分解代谢”占比为 2.16% (255 个), 对于研究瑞香狼毒中的有效成分的生物合成和运输分解具有指导意义。

KEGG 数据库比对分析发现, 共有 6 683 条序列被注释到 119 个 KEGG 通路中。其中, 共有 1 183 个成员与代谢途径 (ko01100) 相关, 有 484 个成员与生物碱生物合成有关 (ko01064、ko01065、ko01066、ko01063), 有 117 个成员与萜类和甾醇的生物合成有关 (ko01062)。此外, 有 15 个成员与萜类骨架合成 (ko00900) 有关, 有 12 个成员与苯丙氨酸生物合成有关 (ko00940)。还有少数成员与黄酮类合成有关 (4 个, ko00941; 1 个, ko00944; 1 个, ko00904)。

2.3 CDS 分析

对瑞香狼毒转录组所有 Unigene 的 CDS 进行预测, 通过 Blast 比对共获得 CDS 序列 60 661 个, 利用 ESTscan 数据库中进行分析, 可获得 CDS 序列 1 489 个。

2.4 萜类代谢基因分析

通过 KEGG 分析发现了 15 个与萜类骨架合成

有关的基因 (表 1), 其中, AACT 基因 (7 个 Unigene)、HMGS 基因 (2 个 Unigene)、HMGR 基因 (1 个 Unigene) 和 MVD 基因 (2 个 Unigene) 都属于 MVA 途径, 而在 MEP 途径中则没有注释到相关基因。此外, 还发现了下游途径中的 1 个 IPP1 基因 (Unigene44316_RX), 1 个 FDPS 基因 (Unigene3650_RX) 和 1 个 DHDDS 基因 (Unigene29072_RX)。同时, 通过比较这些 Unigene 的 RPKM 值可以看出, 表达值最高的是 AACT 基因, 其中有 2 个注释为 AACT 基因的 Unigene45755_RX 和 Unigene45792_RX 包含的 reads 数目超过了 1 000, 而 HMGR 和 MVD 基因表达量较低, 包含的 reads 数均只有 4 条。这些基因的发现, 都为进一步研究佛波酯类化合物生物合成分子机制奠定了研究基础。

2.5 转录组数据 SSR 特征分析

利用 MISA 软件对 47 053 条瑞香狼毒 Unigene 进行 SSR 特征分析, 结果发现检测的序列总长度为 19 701 894 bp, 共发现 3 480 个 SSR, 共有 3 035 条序列含有 SSR 重复单元, 占 Unigene 总数的 6.45%。此外, 含有 1 种以上的 SSR 位点的 Unigene 为 379 条, 而含有混合型 SSR 重复单元的 Unigene 为 254 条。此外, 瑞香狼毒 SSR 重复类型分析发现, 数量最高的为单碱基重复为 1 986 条, 出现频率为 57.07%, 其次为二碱基重复和三碱基重复 SSR 分别

有 833 条和 568 条，出现频率分别为 23.94%和 16.32%，最少的是六碱基重复 SSR 只有 5 条，出现频率仅为 0.14%（表 2）。

此外，在单碱基重复中重复基元频率最高最多

的是 A/T，占比为 90.53%，而二、三碱基重复最多的基元分别为 AC/GT 和 AGG/CCT，占比分别为 66.39%和 32.22%，出现频率最低的六碱基重复一共有 5 条，ACATAT/ATATGT 等不同类型各自 1 条。

表 1 瑞香狼毒中萜类骨架合成基因

Table 1 Genes related to tepenoids biosynthesis of *S. chamaejasme*

酶	EC 编号	Unigenes	长度/bp	reads 数目	RPKM
acetyl-CoA C-acetyltransferase (AACT)	EC 2.3.1.9	Unigene37907_RX	338	122	78.85
		Unigene40848_RX	428	135	68.90
		Unigene41585_RX	459	31	14.75
		Unigene43820_RX	613	416	148.24
		Unigene45755_RX	972	1 143	256.88
		Unigene45792_RX	985	1 172	259.92
		Unigene46404_RX	1 304	942	157.80
hydroxymethylglutaryl-CoA synthase (HMGS)	EC 2.3.3.10	Unigene11215_RX	596	98	35.92
		Unigene13662_RX	793	97	26.72
hydroxymethylglutaryl-CoA reductase (HMGR)	EC 1.1.1.34	Unigene24014_RX	198	4	4.41
diphosphomevalonate decarboxylase (MVD)	EC 4.1.1.33	Unigene20427_RX	178	4	4.91
		Unigene3231_RX	231	7	6.62
isopentenyl-diphosphate Delta-isomerase (IPPI)	EC 5.3.3.2	Unigene44316_RX	672	198	64.36
(2E,6E)-farnesyl diphosphate synthase (FDPS)	EC 2.5.1.1/EC 2.5.1.10	Unigene3650_RX	829	69	18.18
dehydrodolichyl diphosphate synthase (DHDDS)	EC 2.5.1	Unigene29072_RX	233	16	15.00

表 2 瑞香狼毒转录组中 SSR 重复类型分布

Table 2 SSR distribution in unigenes of *S. chamaejasme*

重复类型	总数	频率/%	最多重复基元	数目	比例/%
单碱基重复	1 986	57.07	A/T	1 798	90.53
二碱基重复	833	23.94	AC/GT	553	66.39
三碱基重复	568	16.32	AGG/CCT	183	32.22
四碱基重复	73	2.10	ATCC/ATGG	17	23.29
五碱基重复	15	0.43	AAAAC/GTTTT	10	66.67
六碱基重复	5	0.14	ACATAT/ATATGT	1	20.00
			ACCATC/ATGGTG		
			AGCAGG/CCTGCT		
			AGCTCC/AGCTGG		
			AGGCCC/CCTGGG		

2.6 转座子分析

利用 RepeatMasker 在线分析瑞香狼毒转录组序列的转座子进行预测，结果共发现有 1 497 条转座子，其中 827 条序列的 E 值<1×10⁻⁵，包含 22 种类型转座子，其中 LINE/L1 类型最多，共有 405 条，占比为 48.97%，其次分别为 LTR/ERV1 (91 条) 和 LTR/ERVK (73 条)，占比分别为 11.00%和 8.83%。

此外，DNA/Ginger、DNA/hAT、DNA/PIF-ISL2EU 和 LINE/Jockey 以及 LTR/Lenti 类型数目最少，分别只有 1 条，见表 3。

3 讨论

狼毒是我国传统中医用药，作为其药源植物瑞香狼毒的基因信息却非常匮乏。本研究采用 illumina 平台针对瑞香狼毒展开了转录组测序工作，共获得

表 3 瑞香狼毒转录组中转座子类型及数量

Table 3 Number and types of transposons in transcriptome of *S. chamaejasme*

转座子类型	数目	比例/%
DNA/Crypton	3	0.36
DNA/Ginger	1	0.12
DNA/hAT	1	0.12
DNA/hAT-Ac	49	5.93
DNA/hAT-Charlie	20	2.42
DNA/hAT-Tip100	7	0.85
DNA/P	5	0.60
DNA/PIF-ISL2EU	1	0.12
DNA/PiggyBac	9	1.09
DNA/TcMar-Marin	20	2.42
DNA/TcMar-Tc2	5	0.60
DNA/TcMar-Tigge	35	4.23
LINE/Jockey	1	0.12
LINE/L1	405	48.97
LINE/L2	4	0.48
LTR/ERV1	91	11.00
LTR/ERVK	73	8.83
LTR/ERVL	35	4.23
LTR/Gypsy	44	5.32
LTR/Lenti	1	0.12
LTR/Pao	15	1.81
RC/Helitron	2	0.24

了 47 053 个 Unigenes, 平均长度为 419 nt, 获得了大量瑞香狼毒中萜类、黄酮类等成分合成相关基因信息。研究发现, 经过 Blast 比对注释后, 瑞香狼毒的 Unigene 中除了分别与水稻、拟南芥和玉米相匹配度较高, 此外, 还与致病疫霉和藻类(长囊水云和莱茵衣藻)分别具有较高的的相似度(图 2-C)。推测原因可能是测序样品上有污染造成。此外, 本研究结果与之前本研究组测序分析的狼毒大戟的转录组数据也存在很大不同。狼毒大戟隶属于大戟科大戟属, 也含有佛波酯类化合物 prostratin。但是, 分析狼毒大戟转录组数据, 可以发现其 67.5% 的 Unigene 注释为同为大戟科的麻风树 *Ricinus communis* Linn., 其次有 17.8% 注释为毛果杨 *Populus trichocarpa* Torr. & Gray 和 8.4% 注释为葡萄 *Vitis vinifera* Linn. 等植物^[16]。此外, 瑞香狼毒并没有与任何其他能够产生佛波酯类化合物的作物, 如其他大戟科的植物具有较高相似性。分析原因, 有

可能是因为虽然瑞香狼毒和狼毒大戟等植物都含有佛波酯类化合物, 但是两者在进化关系上仍然较远, 且各自在合成佛波酯类的途径上或许也存在着较大差异。具体原因还需要今后进一步深入研究。

研究表明, 植物萜类的有 2 条生物合成途径: 甲羟戊酸(MVA)途径和 2-C-甲基-D-赤藓糖醇-4-磷酸(MEP)途径^[17]。本研究发现了 15 个二萜类化合物合成途径关键酶基因, 然而这些基因全部都来自 MVA 途径, 而合成萜类的另外一条途径 MEP 途径的关键酶基因却一个都没有被测到。可见, 在瑞香狼毒的萜类合成过程中, MVA 途径可能起到了主导作用。此外, 本课题组发现在这 15 个基因中, HMGR 和 MVD 基因表达量相对较低, 表明这两个基因极其可能是限制瑞香狼毒中 Prostratin 生物合成的限速步骤之一。这与前人研究证实的 HMGR 是 MVA 途径的限速步骤较为一致^[18]。

此外, 研究表明蓖麻烯合酶(casbene synthase, CS)基因可能是合成 prostratin 的母核巴豆烷的关键基因^[19]。然而, 在本研究中却没有找到测序序列被注释为 CS 的候选 Unigene 序列。利用本研究前期从狼毒大戟以及麻风树等植物中的 13 个 CS 基因序列为 query, 与本研究中的瑞香狼毒的 Unigene 进行 blast 比对, 也没有获得相应的候选 Unigene。本研究前期研究也发现利用 LC-MS 方法并未从其根部检测出 prostratin^[20]。因此推测, prostratin 及其前体可能不在瑞香狼毒根中合成或者量极低, 一般方法难以检测。另外一个原因, 推测是由于受本研究中所用的测序平台技术限制, 测序组装得到的 Unigene 序列太短, 以及数据库资源有限, 功能注释难度较大, 造成一批在根中表达丰度较低的基因, 如上述 MEP 途径中的关键酶基因和 CS 基因, 并未被得到准确组装和注释。因此, 在今后的研究工作中, 可以利用同源基因克隆的方法和更先进的测序平台针对瑞香狼毒中 prostratin 的生物合成途径分子机制进行进一步深入研究。

参考文献

- [1] 孙 庚, 罗 鹏, 吴 宁. 瑞香狼毒对青藏高原东部高寒草甸主要物种花粉萌发和种子结实的花粉化感效应[J]. 生态学报, 2010, 30(16): 4369-4375.
- [2] Jin H, Yang X Y, Yan Z Q, et al. Characterization of rhizosphere and endophytic bacterial communities from leaves, stems and roots of medicinal *Stellera chamaejasme* L. [J]. *System Appl Microb*, 2014, 37(5): 376-385.

- [3] 裴月湖, 冯宝民, 华会明, 等. 瑞香狼毒化学及药理研究进展 [J]. 中草药, 2001, 32(8): 764-766.
- [4] 赵明德, 王文颖. 瑞香狼毒生物学特性、生药鉴定及化学成分研究进展 [J]. 青海民族大学学报: 教育科学版, 2011(5): 61-64.
- [5] Yan Z Q, Guo H R, Yang J Y, *et al.* Phytotoxic flavonoids from roots of *Stellera chamaejasme* L. (Thymelaeaceae) [J]. *Phytochemistry*, 2014, 106(10): 61-68.
- [6] Asada Y, Sukemori A, Watanabe T, *et al.* Isolation, structure determination, and anti-HIV evaluation of tiglane-type diterpenes and biflavonoid from *Stellera chamaejasme* [J]. *J Nat Prod*, 2013, 76(5): 852-857.
- [7] Cashmore, A R, Seelye R N, Cain B F, *et al.* The structure of prostratin: a toxic tetracyclic diterpene ester from *Pimelea prostrata* [J]. *Tetrahedron Lett*, 1976, 17(20): 1737-1738.
- [8] Gustafson K R, Cardellina J H, McMahon J B, *et al.* A non-promoting phorbol from the Samoan medicinal plant, *Homalanthus nutans*, inhibits cell killing by HIV-1 [J]. *J Med Chem*, 1992, 35(11): 1978-1986.
- [9] Rullas J, Bermejo M, Garcia-Perez J, *et al.* Prostratin induced HIV activation and downregulates HIV receptor in peripheral blood lymphocytes [J]. *Antivir Ther*, 2004, 9(4): 545-554.
- [10] Hezareh M. Prostratin as a new therapeutic agent targeting HIV viral reservoirs [J]. *Drug News Perspect*, 2005, 18(8): 496-500.
- [11] Reuse S, Calao M, Kabeya K, *et al.* Synergistic activation of HIV-1 expression by deacetylase inhibitors and prostratin: implications for treatment of latent infection [J]. *PLoS One*, 2009, 4(6): 773-789.
- [12] Kawamura S, Chu H, Felding J, *et al.* Nineteen-step total synthesis of (+)-phorbol [J]. *Nature*, 2016 532(7597): 90-93.
- [13] Wender P A, Kee J M, Warrington J M. Practical synthesis of prostratin, DPP, and their analogs, adjuvant leads against latent HIV [J]. *Science*, 2008, 320(5876): 649-652.
- [14] Li R Q, Zhu H M, Ruan J, *et al.* De novo assembly of human genomes with massively parallel short read sequencing [J]. *Genome research*, 2010, 20(2): 265-272.
- [15] 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(3): 138-148.
- [16] Barrero R A, Chapman B, Yang Y F, *et al.* De novo assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes [J]. *BMC Genomics*, 2011, 12(1): 600.
- [17] 岳跃冲, 范燕萍. 植物萜类合成酶及其代谢调控的研究进展 [J]. 园艺学报, 2011, 38(2): 379-388.
- [18] 刘颖, 许巧仙, 席培宇, 等. 甘草 HMGR 基因 cDNA 的克隆及功能鉴定 (英文) [J]. 药学报, 2013, 48(5): 773-779.
- [19] Kirby J, Nishimoto M, Park J G, *et al.* Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae* [J]. *Phytochemistry*, 2010, 71(13): 1466-1473.
- [20] Tang Q, Su Z H, Han Z T, *et al.* LC-MS method for detecting prostratin in plant extracts and identification of a high-yielding population of *Euphorbia fischeriana* [J]. *Phytochem Lett*, 2012, 5(1): 214-218.