

基于转录组分析华细辛甲基丁香酚生物合成途径的相关基因

林懋怡, 牛 卉, 刘晋杰, 柳 威, 刘 忠*

上海交通大学药学院, 上海 200240

摘 要: 目的 获得华细辛 *Asarum sieboldii* 转录组数据库和差异表达基因, 识别华细辛甲基丁香酚生物合成相关基因。方法 以华细辛地下部分(根)和地上部分(叶片)2个样本作为供试材料, 采用 Illumina Hiseq 4000 进行转录组测序, 并从头拼接, 对 Unigene 进行功能注释和差异性分析。结果 共获得 12.25 Gb 数据, 拼接得到 129 003 个 Unigene, 可归类于 52 个 GO 分类, 涉及 363 条 KEGG 代谢通路。分析发现, 共有 439 个差异表达基因, 其中上调基因占 38.3%, 下调基因占 61.7%, 差异较大; 136 个 Unigene 与苯丙素类次生代谢途径相关, 其中 44 个 Unigene 参与甲基丁香酚的生物合成过程。结论 本研究得到大量华细辛转录本信息, 所发掘的与甲基丁香酚生物合成相关的 Unigene 为相关基因的分离、功能验证及其表达调控机制的阐明提供了重要依据。

关键词: 华细辛; 转录组; 苯丙素类生物合成途径; 甲基丁香酚; Unigene

中图分类号: R282.12 **文献标志码:** A **文章编号:** 0253-2670(2017)15-3160-08

DOI: 10.7501/j.issn.0253-2670.2017.15.023

Identification of genes related to biosynthesis of methyleugenol in *Asarum sieboldii* by transcriptome sequence analysis

LIN Mao-yi, NIU Hui, LIU Jin-jie, LIU Wei, LIU Zhong

School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: Objective To obtain the transcriptome sequence database and differentially expressed genes of *Asarum sieboldii*, and to identify the genes related to the biosynthesis of methyleugenol, the main chemical compound in this species. **Methods** Roots and leaves of the plants were chosen as experiment materials. The transcriptome sequence database was constructed by applying an Illumina Hiseq 4000 Sequencing Platform. Unigenes were assembled by BLAST similarity searches and annotated with GO and KEGG orthologs identifiers. Moreover, differentially expressed genes were analyzed. **Results** 12.25 Gb database was obtained, among which 129 003 unigenes were annotated to be involved in 52 GO-terms and 363 metabolic pathways. After analysis, 439 differentially expressed genes were observed, the up-regulated genes account for 38.3% and the down-regulated genes account for 61.7%. In addition, 136 unigenes involved in phenylpropanoid biosynthesis in *A. sieboldii*, and 44 unigenes that were associated with biosynthesis of methyleugenol were identified. **Conclusion** Unigenes explored in this study will significantly contribute to genome-wide research and analysis of this species.

Key words: *Asarum sieboldii* Miq.; transcriptome; phenylpropanoid biosynthesis pathway; methyleugenol; Unigene

华细辛 *Asarum sieboldii* Miq. 为马兜铃科细辛属多年生草本植物, 广泛分布于长江流域。华细辛是使用历史最悠久的细辛道地药材, 历版《中国药典》均将其收载为正品细辛的基原之一^[1-2]。华细辛主要活性成分有甲基丁香酚、黄樟醚、 α -细辛醚、细辛脂素等^[3], 它们均通过苯丙素类生物合成途径生成。

转录组是特定细胞或组织在某一功能状态下基因表达转录产物的总和。基于转录组的分析可全面

揭示有机体的不同组织在不同生长发育阶段或者在环境条件改变情况下基因表达的情况, 从而反映该有机体生命活动的状态^[4]。转录组测序可在缺乏遗传背景资料的情况下, 直接对任意物种的转录组进行分析^[5]。近年来, 由于转录组测序技术方法的快速发展, 并且具有高通量、速度快、价格低的优势, 已被广泛地应用于对非模式生物的研究中^[6]。根据文献报道, 已有几十种药用植物进行过转录组测序,

收稿日期: 2017-03-31

基金项目: 国家自然科学基金项目(81373962, 31570325); 上海交通大学医工交叉基金项目(YG2013MS72)

作者简介: 林懋怡(1993—), 女, 在读硕士研究生, 主要从事中药次生代谢研究。E-mail: lynmaoyi@foxmail.com

*通信作者 刘 忠(1967—), 男, 副教授, 主要从事中药次生代谢研究。E-mail: liuzhong@sjtu.edu.cn

如当归^[7]、红花^[8]、枸杞^[9]、半夏^[10]等,在此基础上,还进一步对多个基因进行了克隆与表达分析^[11-12],阐述了黄酮类^[13]、萜类^[14]等次生代谢产物生物合成的遗传基础。

目前,细辛类药用植物遗传基础的研究十分缺乏,仅限于 PAL^[15]、MADS-box^[16]中单基因的克隆,难以全面深入地理解细辛次生代谢产物生物合成的调控机制。本研究采用 Illumina Hiseq 4000 高通量测序平台,对华细辛不同部位(根和叶片)的转录组进行高通量测序,建立华细辛转录组数据库,分析基因表达水平的差异,识别甲基丁香酚生物合成途径相关基因,为华细辛品质改良次生代谢工程的开展提供依据。

1 材料与方法

1.1 材料

本研究样品采集于陕西安强县华细辛栽培种植基地,经上海交通大学药学院刘忠副教授鉴定为马兜铃科细辛属华细辛 *Asarum sieboldii* Miq.。收集 10 个盛花期代表性华细辛样品,分别取其根和叶片,用液氮研磨成粉,存于 -80 °C 备用。

1.2 转录组测序

采用植物总 RNA 提取试剂盒(Tiangen, 北京)分别从华细辛根和叶片中提取 RNA。所得总 RNA 经过 NanoDrop 2000 和 Agilent 2100 系统质检合格后备用。按照 NEBNext Ultra RNA Library Prep Kit 要求建立测序样本的文库,经过 Agilent 2100 和 Agilent High Sensitivity DNA Kit 检测合格之后,在 cBOT 自动成簇,最后在 Illumina Hiseq 4000 测序平台上进行双端测序(Paired End)。转录组测序委托苏州金唯智生物技术有限公司完成。

1.3 测序数据质量分析

获得测序原始数据之后,对测序数据质量进行评估。使用二代测序数据质量统计软件 Cutadapt (version 1.9.1) 对测序原始数据(Pass Filter Data)去除接头以及低质量序列。随后采用 FastQC^[17]对测序得到的 reads 进行检测,质量分数(Q)与测序错误率(P)关系为 $Q = -\lg P$ 。

1.4 转录组 De novo 组装

采用 Trinity (vr20140717)^[18]软件对样品数据进行从头组装,然后使用 Cd-hit (v4.5.4) 对得到的组装结果通过序列聚类做进一步的序列拼接和去冗余处理,最终得到非冗长的 Unigene 序列。并通过 bowtie2 (v2.1.0) 将 clean data 比对到 Unigene 上,

获得比对上 Unigene 的 reads 数量和比例。

1.5 Unigene 功能注释、分类和代谢途径分析

使用 BLAST 程序对 Unigene 与蛋白质数据库进行比对期望值(E 值) $< 1 \times 10^{-5}$, 蛋白质数据库包括 NCBI 非冗余蛋白数据库 Nr、GO、COG、SwissProt、KEGG。根据 GO 注释的基因序列,运用 WEGO 软件^[19]按功能进行统计。根据 COG 功能注释的结果,按功能对序列进行归纳。依据 KEGG 注释的信息,把序列按代谢通路进行一级分类和二级分类。对基因注释信息进行整理,分析华细辛中苯丙素类生物合成途径,搜索甲基丁香酚生物合成相关的基因。

1.6 基因表达水平分析

以拼接序列作为参考,经 RSEM (v1.2.4) 将 clean data 与功能基因 Unigene 进行比对,使用 FPKM (Fragment Per Kilo Bases per Million reads) 方法计算基因表达量,对测序深度和基因长度进行归一化,使不同长度的基因在不同测序深度下得到的基因表达水平估计值具有可比性。

1.7 差异表达基因筛选及分析

基于基因表达水平分析获得的 readcount 数进行差异表达基因筛选,使用 Bioconductor 软件包的 DESeq (v1.18.0) 进行分析,对检测的结果按照差异显著性标准[差异基因表达变化 2 倍以上且错误发现率(FDR)值 ≤ 0.05] 进行筛选,统计基因显著性差异表达情况,并对差异基因进行分类注释及 pathway 富集分析。

2 结果与分析

2.1 序列组装

总共得到约 12.25 Gb 数据量,reads 过滤后 Q30 (测序错误率小于 0.1%的碱基数量百分比) 为 92.48%,说明绝大部分 reads 质量较好,可用于后续的序列拼接。经从头组装后获得 20 341 742 个 Contig,平均长度为 54.98 nt, N50 长度为 48 nt; 去冗余处理后最终获得 129 003 个 Unigene,平均长度为 873.28 nt, N50 长度为 1 278 nt, GC 量为 42.50% (表 1)。从 Unigene 长度分布情况来看, 200~500 bp 所占比例最大,共 56 165 个 (43.54%), 大于 1 000 bp 共有 11 702 个 (9.07%) (图 1)。

将 clean data 与 Unigene 进行比对统计分析,根和叶片的 reads 覆盖率达到 77.21%和 76.83%,唯一比对到参考基因的 reads 数量分别为 55 647 450 和 55 866 228,可用于基因表达分析。

表 1 序列拼接的数据统计

Table 1 Statistic of sequence assembly result

序列种类	计数	总碱基数量	最短长度/nt	最长长度/nt	平均长度/nt	N50/nt	GC/%
Contig	20 341 742	1 118 293 483	25	16 273	54.98	48	45.68
Unigene	129 003	112 655 511	201	21 571	873.28	1 278	42.50

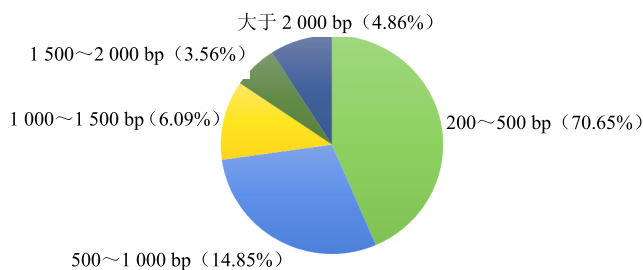


图 1 Unigene 不同长度区间分布图

Fig. 1 Length distribution of Unigenes

2.2 功能注释

基于 BLAST 算法, 将获得的 Unigene 与数据库中蛋白序列进行相似性比对。在 129 003 个 Unigene 中, 共有 67 875 个 (52.61%) 基因获得功能注释信息, 在 Nr 数据库中获得注释信息最多, 共 67 050 个 (98.78%), 具体 Unigene 各数据库注释个数结果见图 2。

Nr 功能注释结果中, 12.98% 的 Unigene 同源性高于 95%, 同源性在 60%~80% 的 Unigene 所占比例最高, 为 33.83%。Nr 注释的物种分布显示, 比对到莲 *Nelumbo nucifera* Gaertn. 的 Unigene 最多, 为 12 156 个, 其次为葡萄 *Vitis vinifera* Linn. (4 169 个)、拟南芥 *Arabidopsis thaliana* (Linn.) Heynh. (2 688 个) 和油棕 *Elaeis guineensis* Jacq. (2 448 个)。GO (gene ontology) 数据库是基因本体联合会 (Gene Ontology Consortium) 所建立的数据库, 旨在建立一个适用于各种物种的、对基因和蛋白质功能进行

限定和描述的、并能随着研究不断深入而更新的生物语言词汇标准。GO 定义了三级结构的标准语言 (ontologies), 即生物途径 (biological process)、分子功能 (molecular function) 和细胞组分 (cellular component)。通过 GO 注释分类, 发现共有 20 693 个 Unigene 获得 GO 功能注释, 其中 29.66% 被注释为细胞组分, 31.33% 被注释为分子功能, 39.01% 被注释为生物途径 (图 3)。

3 大功能又被细分为 52 个功能组, 其中“细胞组分”组中, 细胞部分 (cell part) 和细胞 (cell) 被注释的最多, 均为 53.3% (11 030 个); “分子功能”部分中, 结合功能 (binding) 和催化功能 (catalytic) 所占比例最大, 分别为 55.4% (11 467 个) 和 52.5% (10 861 个); 而“生物学途径”最主要的为代谢过程 (metabolic process) 和细胞过程 (cellular process), 依次为 61.5% (12 725 个) 和 59.5% (12 335 个)。

结合 COG 数据库对华细辛 Unigene 进行比对, 预测 Unigene 功能并分类统计, 从宏观上了解华细辛表达基因的功能分布特征。研究结果表明, 32 484 个 Unigene 得到注释信息, 并且分为 25 类。其中, 一般功能预测 (R) 最多, 共有 4 431 个, 然后依次为翻译后修饰、蛋白质转换、分子伴侣 (O), 信号转导机制 (T), 翻译-核糖体结构和生源 (J), 依次为 3 938、3 674 和 2 855 个。

为识别华细辛中代谢途径, 将 129 003 个 Unigene 进行 KEGG 映射分析, 共 19 112 个功能基因序列获得注释, 映射到 363 条代谢途径分支中。其中, 包含功能基因序列最多的是核糖体途径、碳代谢途径和氨基酸生物合成途径, 依次为 1 306、882、787 个功能基因。参照 KEGG 数据库, 获得功能注释的 Unigene 可分为 6 大类的代谢途径, 43 个小类, 比对到代谢作用的 Unigene 最多, 共 12 716 个, 其中涉及次生代谢产物生物合成途径的功能基因共 644 个。具体种类和数量见图 4。

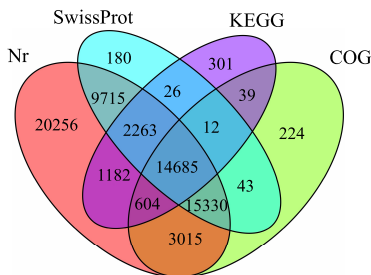


图 2 Unigene 各数据库注释结果维恩图

Fig. 2 Venn diagram of various annotations of Unigenes

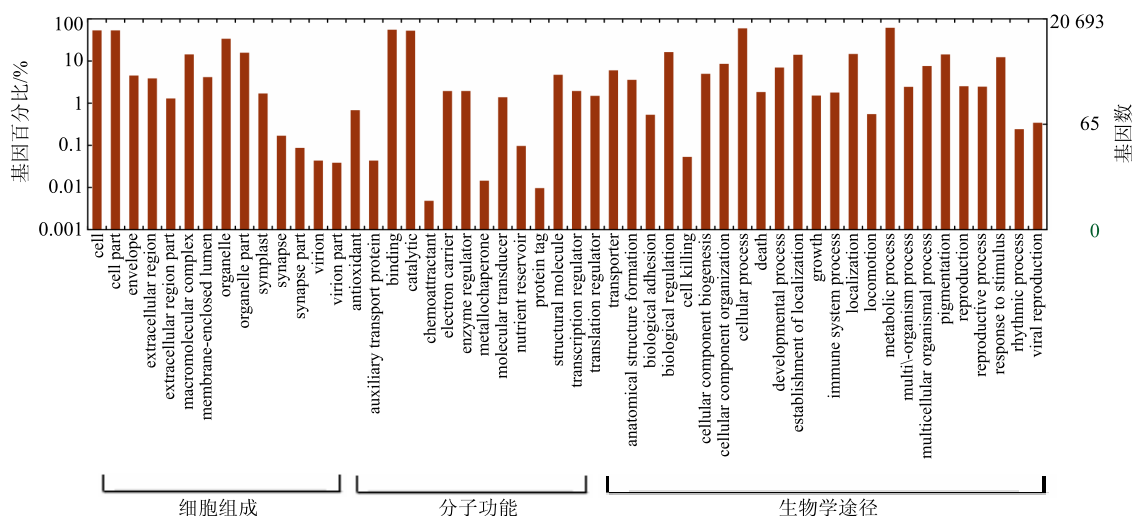


图 3 Unigene 的 GO 注释和分类

Fig. 3 GO Annotation and classification of Unigenes

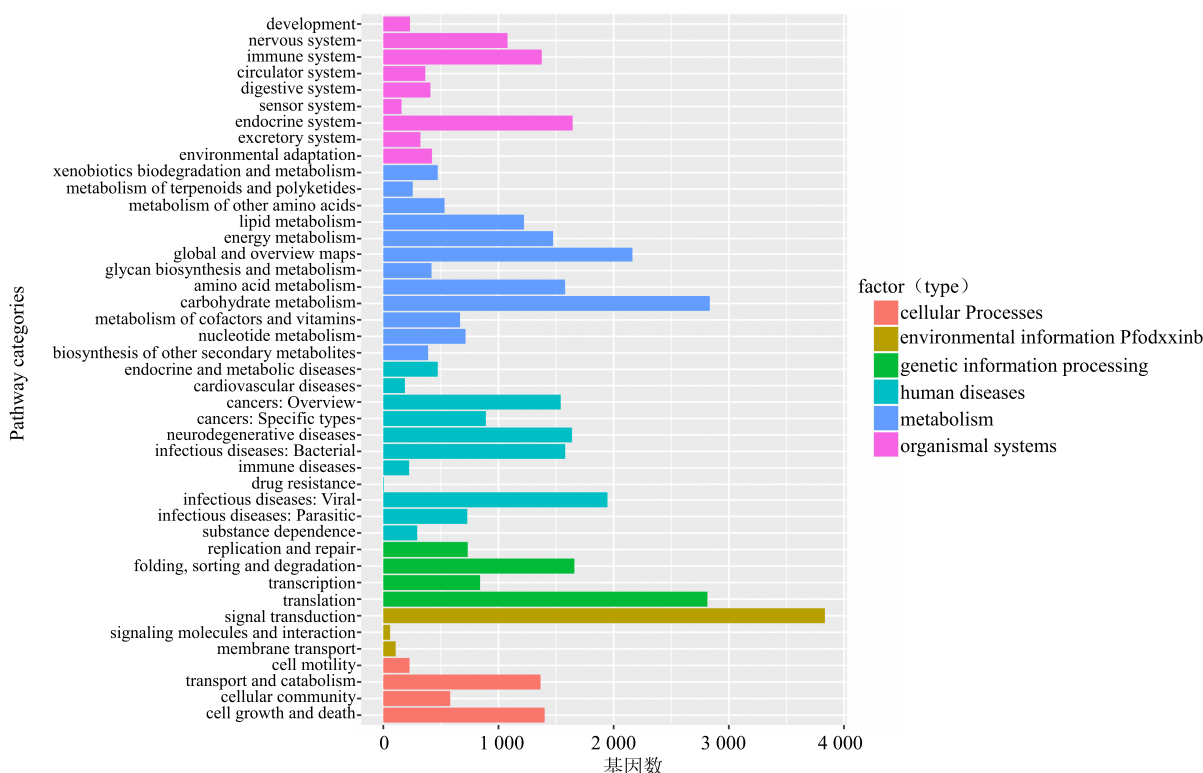


图 4 Unigene 的 KEGG 的注释和分类

Fig. 4 KEGG Annotation and classification of Unigenes

2.3 差异表达基因分析

一个基因表达水平的直接体现就是其转录本的丰度情况, 转录本丰度越高, 则基因表达水平越高。计算 FPKM 即得到基因表达水平相对值, 根和叶片分别有 82.35% 和 84.91% Unigene 的 FPKM 值在 0~1, 大于 60 的 Unigene 分别有 2 266 和 2 169 个, 绝大部分基因得到一定的表达^[20]。

通过差异表达基因的筛选, 获得 2 个样本间 493 条差异性表达基因。其中, 叶片相对于根上调基因共有 189 条 (38.3%), 下调基因共有 304 个 (61.7%)。将 493 条差异性表达基因与数据库 Nr、GO、SwissProt、COG、KEGG 进行比对, 结果显示共有 480 个差异表达基因获得注释。获得 GO 注释的差异表达基因共有 224 个, 有 95 个基因获得生物途径

注释, 细胞组分获得 36 个功能注释, 分子功能得到 93 个功能注释。

获得 COG 注释的差异表达基因共有 104 个基因, 其中大部分基因与翻译、核糖体结构、蛋白质转换等相关。归类于次生代谢生物合成、运输和分解代谢 (secondary metabolites biosynthesis) (transport and catabolism) 的基因具有 14 个。具体分类情况见图 5。

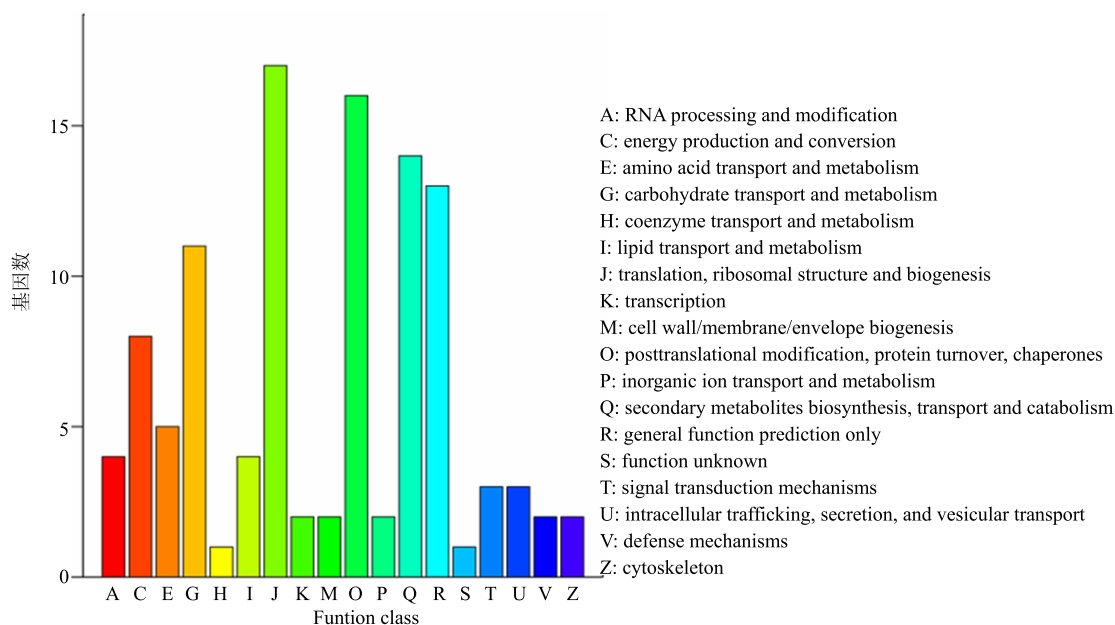


图 5 差异表达基因的 COG 分类

Fig. 5 COG function classification of differentially expressed unigenes

2.4 苯丙素类生物合成途径分析

苯丙素类是一类广泛存在的天然产物, 具有多方面的生物活性。在高等植物体内, 苯丙素类均是莽草酸途径而来, 莽草酸衍生的苯丙氨酸在苯丙氨酸解氨酶 (PAL) 催化下转化为反式肉桂酸, 从而将代谢流引向苯丙素类次生代谢途径。因此, 将华细辛转录组信息与 KEGG 数据库中已有的苯丙素类代谢研究结果进行比对分析, 发现共有 136 个功能基因序列映射到苯丙素类生物合成途径 (ko00940)。具体途径中注释到的 Unigene 见图 6。

甲基丁香酚生物合成途径已经完全明确^[21]。苯丙氨酸在 PAL、C4H、4CL、COMT、CCR、CAD 等酶的依次作用下生成松柏醇, 此后, 由松柏醇酯酰转移酶 (CFAT) 催化, 松柏醇生成松柏脂^[22], 随后经丁香酚合成酶 (EGS) 转化为丁香酚, 最后, 在丁香酚-O-甲基转移酶 (EOMT) 作用下进行甲基化反应, 生成甲基丁香酚。

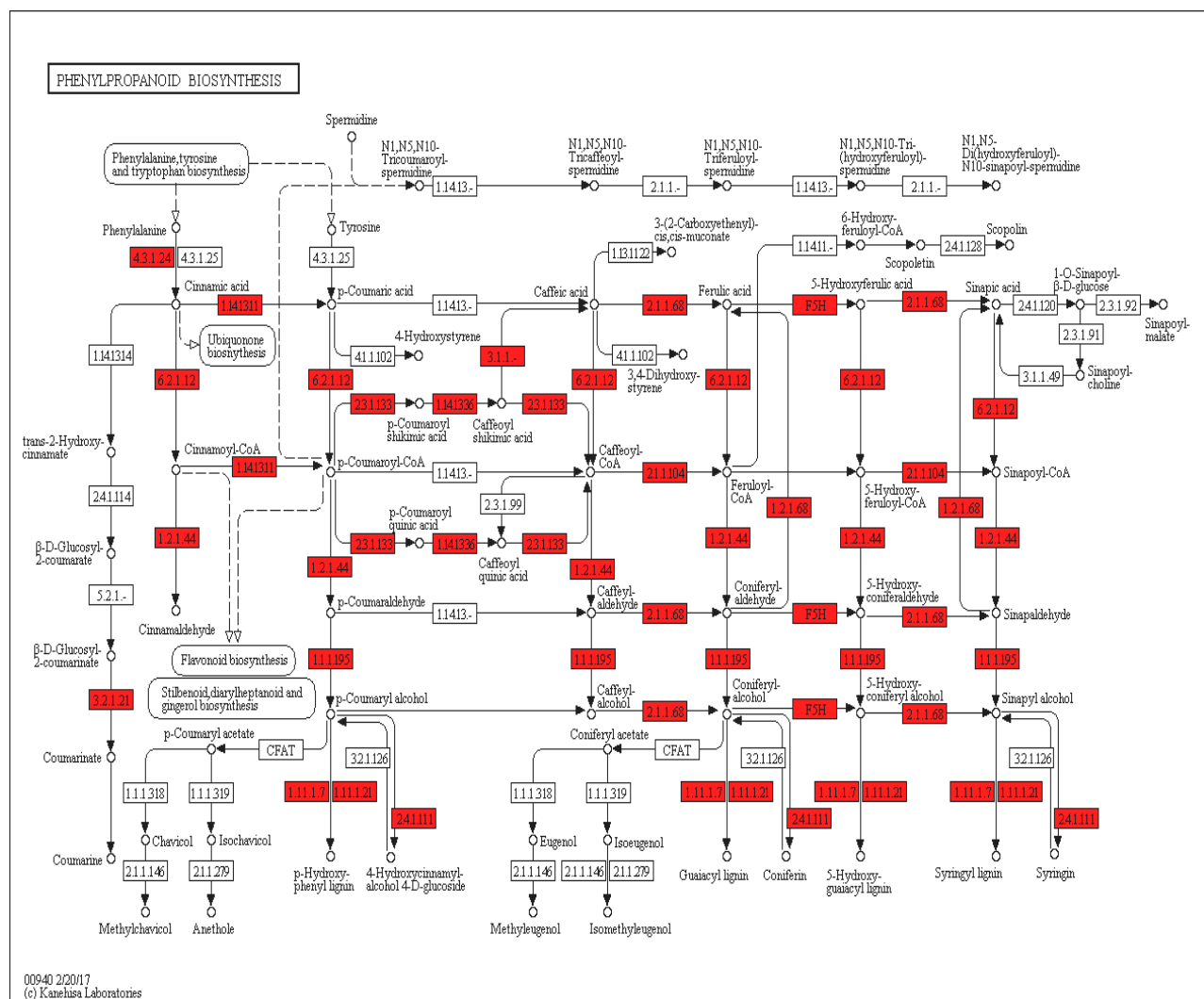
筛选与甲基丁香酚生物合成相关的 Unigene,

与 KEGG 数据库进行比对法发现, 有 229 条基因获得注释, 共涉及 128 条途径 (多个基因涉及 2 个及以上的途径), 显著富集的有 8 条途径, 最为丰富的是核糖体途径, 共有 63 个基因, 其次是碳代谢途径和光合作用途径, 分别有 22 和 16 个基因, 与次生代谢合成相关的途径有黄酮类生物合成途径和萜类生物合成途径, 共有 4 个差异表达基因。

发现华细辛转录组数据库中共有 44 个候选基因, 具体酶和涉及基因见表 2。与苯丙素类合成相关的表达差异较大的基因共有 3 个, 其中, 关键酶咖啡酸 O-甲基转移酶 (OMT, c229186_g1_i1) 催化底物进行甲基转移反应, 其根表达量为 29.18 (相对值), 叶片表达量为 0.03 (相对值), 为典型的下调基因。

3 讨论

本研究首次采用 Illumina HiSeq 4000 高通量测序平台对华细辛根和叶片进行转录组测序, 获得了 129 003 个 Unigene。FastQC 的 Q30 (92.48%) 和 N50 长度 (1 278 nt) 分析, 以及与 5 个蛋白质数据库比对 (共有 67 875 个 Unigene 获得注释) 的研究结果, 都表明样本测序质量和拼接结果良好, 说明 Illumina HiSeq 4000 高通量测序技术可获得大量高质量非模式物种的 Unigene, 在识别药用植物的功能基因和构建基因调控网络等方面具有十分广阔的应用前景^[23]。然而, 由于测序深度的影响和 Trinity 算法的局限性, 转录组测序数据分析中仍可能含有



深色表示注释到的 Unigene

Red parts indicate the Unigenes are annotated

图 6 华细辛苯丙素类生物合成途径

Fig. 6 Phenylpropanoid biosynthesis pathways of *A. sieboldii*

一些假阳性的转录体^[24]。因此，除加强测序深度和完善算法的准确性之外，需要对获得的 Unigene 进行功能验证，排除假阳性转录体，而得到可靠的有功能的基因序列，为后续调控次生代谢奠定良好的理论基础。另一方面，依然有大量 Unigene (47.39%) 未受到功能注释，可能的原因之一是序列过短 (< 500 bp)，难以获得同源信息，因此增强对未知功能基因的发掘能力也是非常重要的。

地下部分（根）和地上部分（叶片）2 个样本间共有 439 个差异表达基因，通过 GO、COG 注释和 Pathway 富集分析，证实华细辛根和叶片转录水平的差异较大。与甲基丁香酚生物合成相关的 Unigene 有 44 个，其中关键酶 COMT 在根中表达

量显著高于叶片中表达量，结合华细辛有效成分积累动态变化的研究结果，即药材采收期华细辛地下部分（根与根茎中）有效成分的量明显高于地上部分有效成分的量^[25-26]，因此，提示可将 COMT 开发为华细辛甲基丁香酚生物合成基因工程的重要调控靶点，通过上调 COMT 的表达，有望提高甲基丁香酚量。此外，甲基丁香酚生物合成途径的下游基因（CFAT、EGS、EOMT）尚未发现，这可能由于数据库资源的缺乏，使得信息不够全面，从而个别表达不够丰富的基因未得到注释。因此，在后续的研究工作中，需要加大挖掘深度，以获得这些尚未得到的下游途径基因序列，为其功能分析与验证乃至遗传操作提供基础。

表 2 甲基丁香酚生物合成相关的 Unigene
Table 2 Unigenes related to methyl eugenol biosynthesis

基因编号	酶及 KO 号	EC 编号	酶的功能
c96041_g1_i1 c125382_g5_i1 c127959_g1_i1 c147849_g1_i2 c147849_g1_i4 c124838_g1_i1 c142973_g1_i1 c147267_g1_i1 c229186_g1_i1	4-香豆酸辅酶 A 连接酶 (K01904)	EC: 6.2.1.12	催化底物连接 CoA 生成相应的 CoA 酯
c61990_g1_i1 c184504_g1_i1 c123395_g10_i1 c116951_g1_i1 c171718_g1_i1 c98285_g1_i1 c124761_g8_i1 c126212_g1_i1 c187899_g1_i1 c209904_g1_i1 c234622_g1_i1 c83117_g1_i1 c98146_g1_i2 c118623_g1_i4 c118623_g2_i1 c161778_g1_i1 c191858_g1_i1 c121549_g1_i1 c195093_g1_i1 c102482_g2_i1 c122608_g1_i1 c122608_g3_i2 c122608_g3_i4 c122608_g3_i6 c201863_g1_i1 c113752_g2_i1 c113752_g3_i1 c153713_g1_i1 c69547_g1_i1 c74310_g1_i1 c74310_g2_i1 c74310_g3_i1 c158054_g1_i1 c188831_g1_i1 c198944_g1_i1	咖啡酸辅酶 A- <i>O</i> -甲基转移酶 (K00588) 咖啡酰莽草酸酯酶 (K18368) 肉桂酸辅酶 A 还原酶 (K09753) 肉桂酸脱氢酶 (K00083) 松柏醛脱氢酶 (K12355) CYP98A3 (K09754) 苯丙氨酸解氨酶 (K10775) 莽草酸- <i>O</i> -羟基肉桂酰转移酶 (K13065) 肉桂酸-4-羟基化酶 (CYP73A, K00487)	EC: 2.1.1.104 EC: 3.1.1 EC: 1.2.1.44 EC: 1.1.1.195 EC: 1.2.1.68 EC: 1.14.13.36 EC: 4.3.1.24 EC: 2.3.1.133 EC: 1.14.13.11	以维生素 B ₁₂ 为辅酶, 催化底物间进行甲基转移反应, 合成阿魏酸辅酶 A 催化水解咖啡酰莽草酸酯生成咖啡酸 催化还原多种脱氢肉桂酸的 CoA 酯生成相应的肉桂醛 催化还原多种肉桂醛生成相应的肉桂醇 催化松柏醛脱氢生成阿魏酸 细胞色素家族成员, 催化羟基化反应 催化苯丙氨酸生成反式肉桂酸 催化 2 条以羟基肉桂酰 CoA 为起点的不同支路合成咖 啡酰辅酶 A 细胞色素家族成员, 催化羟基化反应, 催化反式肉桂 酸生成香豆酸

参考文献

- [1] 蔡少青, 王璇, 朱姝, 等. 中药细辛的本草考证 [J]. 北京医科大学学报, 1997, 29(3): 233-235.
- [2] 中国药典 [S]. 一部. 2015.
- [3] Ramalingam M, Kim S J. Phytochemical, toxicological and pharmacological studies of *Asiasari Radix et Rhizoma*: A review [J]. *Trop J Pharm Res*, 2015, 14(3): 545-554.
- [4] 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.
- [5] 王尧龙, 黄璐琦, 袁媛, 等. 药用植物转录组研究进展 [J]. 中国中药杂志, 2015, 40(11): 2055-2061.
- [6] 张召宝, 侯林, 潘晴, 等. 中药高通量转录组研究进展 [J]. 中国中药杂志, 2014, 39(9): 1553-1558.
- [7] 杨杰, 王金权, 丁维俊, 等. 基于 Illumina HiSeq 2000 测序技术对当归根的转录组特性研究 [J]. 中草药, 2015, 46(8): 1216-1222.
- [8] Liu X, Dong Y, Yao N, *et al.* De novo sequencing and analysis of the safflower transcriptome to discover putative genes associated with safflower yellow in *Carthamus tinctorius* L. [J]. *Inter J Mol Sci*, 2015, 16(10): 25657-25677.
- [9] Zhao S, Tuan P A, Li X, *et al.* Identification of phenylpropanoid biosynthetic genes and phenylpropanoid accumulation by transcriptome analysis of *Lycium chinense* [J]. *Bmc Genom*, 2013, 14: 802-809.
- [10] Zhang G, Jiang N, Song W, *et al.* De novo sequencing and transcriptome analysis of *Pinellia ternata* identify the candidate genes involved in the biosynthesis of benzoic acid and ephedrine [J]. *Front Plant Sci*, 2016, 7: 1209-1213.
- [11] 李建芳, 黄静, 程芳婷, 等. 地黄属核基因与叶绿体基因 DNA 条形码比较研究 [J]. 中草药, 2017, 48(1): 165-171.
- [12] 张玲, 王艳芳, 麦迪, 等. 红花二氢吡啶二羧酸合酶基因片段的分离及表达分析 [J]. 中草药, 2017, 48(2): 351-355.
- [13] 夏奇峰, 赵致, 刘红昌, 等. 艾纳香黄酮类物质生物合成途径分析 [J]. 中国中药杂志, 2016, 41(19): 3630-3636.
- [14] 赵灿, 郭丽娜, 彭玉帅, 等. 三七总皂苷生物合成的关键酶及其调控研究进展 [J]. 中草药, 2015, 46(19): 2954-2965.
- [15] 张劲松. 北细辛愈伤组织诱导及苯丙氨酸解氨酶基因片段的克隆 [D]. 哈尔滨: 东北农业大学, 2014.
- [16] 赵银河. 被子植物基部类群三白草和细辛开花相关基因的克隆, 功能与进化的研究 [D]. 北京: 中国农业大学, 2006.
- [17] Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments [J]. *Bioinformatics*, 2012, 28(16): 2184-2185.
- [18] Grabherr M G, Haas B J, Yassour M, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nat Biot*, 2011, 29(7): 644-652.
- [19] Ye J, Fang L, Zheng H, *et al.* WEGO: a web tool for plotting GO annotations [J]. *Nucleic Acids Res*, 2006, 35(S1): W293-W297.
- [20] Yoshida Y, Tomiyama T, Maruta T, *et al.* De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions [J]. *Bmc Genom*, 2016, 17: 182-191.
- [21] Koeduka T, Fridman E, Gang D R, *et al.* Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester [J]. *PNAS*, 2006, 103(26): 10128-10133.
- [22] 向林, 陈龙清. 花香的基因工程研究进展 [J]. 中国农业科学, 2009, 42(6): 2076-2084.
- [23] 林勇. 面向下一代测序技术的 de novo 序列拼接工具综述 [J]. 小型微型计算机系统, 2013, 34(3): 627-631.
- [24] 常征. 基于 RNA 测序技术的转录组从头拼接算法研究 [D]. 济南: 山东大学, 2014.
- [25] 徐玫, 李景华. 细辛不同部位的马兜铃酸及挥发油含量研究 [J]. 中国药业, 2007, 16(10): 29-31.
- [26] 王栋, 王喜军, 夏晓晖. 北细辛不同生长期甲基丁香酚和黄樟醚的气相色谱法定量分析 [J]. 色谱, 1997, 15(1): 87-88.