

文章编号:1001-1498(2015)04-0597-07

## 红松 4 个组织的转录组数据分析与次生代谢产物的表达差异初探

张 振, 张含国\*, 周 宇, 张 磊, 于宏影, 张 莉

(东北林业大学 林木遗传育种国家重点实验室, 黑龙江 哈尔滨 150040)

关键词: 红松; 转录组; 基因注释; 次生代谢途径

中图分类号: S791.245 文献标识码: A

### Differential Gene Expression Analysis on Secondary Metabolites and Transcriptome Sequencing to Four Tissues of *Pinus koraiensis*

ZHANG Zhen, ZHANG Han-guo, ZHOU Yu, ZHANG Lei, YU Hong-ying, ZHANG Li

(State Key Laboratory of Forest Genetics and Tree Breeding, Northeast Forestry University, Harbin 150040, Heilongjiang, China)

**Abstract:** This research compared the transcriptomes of four different tissues of *Pinus koraiensis* by using RNA-seq technique and a lot of differentially expressed unigenes were obtained. A total of 21.3 GB of data was obtained and 41 476 expressed unigenes were received after assembling. Through GO and COG classification, the commented and transcribed sequences were classified into 24 or 56 functional categories, which referred to many biochemical processes such as material and energy metabolism, signal transduction, transcriptional regulation and defense reaction. Through pathway enrichment analysis, the pathways associated with grease synthetic fatty acid biosynthetic and terpenoids synthesis were found. Two unigenes of metabolic pathways in KEGG database were compared and 16 key enzyme genes of fatty acid synthesis in other species homologous and 34 key enzyme genes of terpenoid biosynthetic pathway were obtained. By high-throughput sequencing and bioinformatics research, a large number of *Pinus koraiensis* transcriptome unigenes was obtained. This research provides references for the research on the genes related to secondary metabolites synthesis in *Pinus koraiensis*.

**Key words:** *Pinus koraiensis*; transcriptome; gene annotation; secondary metabolites biosynthesis

红松(*Pinus Koraiensis*)为松科松属乔木,主要分布在中国东北东部,俄罗斯远东南部及朝鲜半岛,日本岛也有零星分布。红松是中国Ⅱ级重点保护野生植物,既是优良的用材树种,又是珍贵的食用干果和油料树种<sup>[1]</sup>。红松树干圆满通直,叶5针一束,球果两年成熟,花期5—6月,种子在次年9—10月成熟。红松木材可作建筑、造纸,树皮中含有松树脂,可作药用,松仁是一种营养价值很高的果品<sup>[2-4]</sup>,含有大

量的粗蛋白,油脂,多糖,粗纤维,除此之外,还富含维生素、矿物质以及钙、磷、锰、钴、铜、锌等微量元素<sup>[4-5]</sup>,与意大利石松(*P. pinea*),西伯利亚红松(*P. sibirica*),油松(*P. sinensis*),瑞士五针松(*P. cembra*),克罗拉多果松(*P. edulis*),单叶果松(*P. monophylla*)等松科球果为目前世界松仁商品主要种类。目前,在红松的遗传育种进程中,主要以可食用坚果林选育及品质改良为主,但针对红松的分子生

收稿日期: 2013-10-17

项目基金: 林业公益性行业科研专项“红松果材兼用林良种选育与定向培育技术研究”(201204320)

作者简介: 张 振(1986—),男,在读博士,主要从事林木遗传改良研究。E-mail: zhenzh19860516@163.com

\* 通讯作者: 教授,博士生导师,主要研究方向: 林木遗传育种与改良

物学研究进展较慢,红松的基因组及转录组数据还比较缺乏,对生长发育及代谢途径、遗传图谱构建等的研究较少,影响了红松遗传改良的研究进程。

EST 技术是开展转录组研究的有效方法<sup>[6]</sup>,已被广泛应用于动、植物和微生物的基因表达谱分析、功能基因预测等<sup>[7-9]</sup>。目前公共数据库中已有松属多个树种的 EST 序列,如火炬松、北美短叶松、海岸松<sup>[10-12]</sup>。截止 2014 年 5 月,从 GenBank 查找发现松属 EST 序列有 730 037 条,但红松的 EST 序列仅为 65 条。随着测序技术的发展,将高通量测序应用到红松的转录组研究中,不仅可降低测序的成本和时间,而且可获得的丰富的数据,有利于松属的转录组和生长发育相关研究。本研究采用 Illumina 公司开发的 HiSeq2000 测序技术进行松属树种转录组的研究,构建红松均一化 cDNA 文库、测序,对序列再做基因功能注释和分类等,以研究红松生长发育过程中重要基因的表达,获得包括脂肪酸类合成、萜类化合物合成等的合成相关的生物酶基因及参与植物次生代谢的基因,获得的 Unigenes 为红松及其相近物种的群体基因组学和功能基因组学提供了有价值的候选基因,这些基因将有助于开展选择育种和基因工程研究。

## 1 材料与方法

### 1.1 样本材料来源

样本采集于黑龙江省林口县青山林场红松种子园内,种子园为 1979 年嫁接营建,株行距 4 m × 6 m。于 2013 年 6 月分别采集无性系‘LK15’与无性系‘LK20’的上部针叶(T1)、嫩梢(T2)、雌花(T3)和球果(T4)4 个组织,液氮速冻后 -80℃ 保存。每个无性系的 4 个组织材料单独提取 RNA 后,再将提取的 RNA 分组织混合用于转录组测序。

### 1.2 总 RNA 的提取和 mRNA 的分离

T1 - T4 总 RNA 提取方法参照 Trizol Reagent 说明书进行。用 1.2% 琼脂糖凝胶电泳和 Nanodrop 纯度检测, Qubit2.0 浓度检测及 Agilent 2100 质量检测。确定总 RNA 完整性和纯度质量后,用 NEB Next Poly(A) mRNA Magnetic Isolation Module (NEB, E7490)富集 mRNA。

### 1.3 cDNA 片段合成及测序文库构建

以 mRNA 为模板,用 NEB Next mRNA Library Prep Master Mix Set for Illumina (NEB, E6110) 和 NEB Next Multiplex Oligos for Illumina (NEB, E7500)

构建上机文库。使用 Qubit2.0 进行初步定量,稀释文库至  $2 \text{ ng} \cdot \mu\text{L}^{-1}$ ,制备好的文库用 1.8% 琼脂糖凝胶电泳进行检测文库插入片段大小(insert size),然后用 Library Quantification Kit-Illumina GA Universal (Kapa, KK4824)进行 QPCR 定量(文库有效浓度  $> 2 \text{ nmol} \cdot \text{L}^{-1}$ )。检测合格的文库在 illumina cbot 上进行簇的生成,最后用 Illumina HiSeqTM 2000 进行测序。

### 1.4 序列分析

利用 Illumina 平台将测序所得的图像数据转化为相应的核苷酸序列数据,对所产生的原始序列文件进行质量评估和可信度分析,并去除测序过程中低质量的序列和不确定的序列( $Q < 20$ )。使用 Trinity<sup>[13]</sup>软件对样品数据进行混合组装,通过序列之间的 overlap 信息组装得到 contig,再根据序列的 paired-end 信息和 contig 之间的相似性对 contig 进行聚类,然后在局部进行组装得到转录本,最后从局部中挑选最主要的转录本作为 Unigene。

### 1.5 序列注释、功能分类和生物学通路分析

使用 BLAST<sup>[14]</sup>软件将 Unigenes 序列与 NR<sup>[15]</sup>、SwissProt<sup>[16]</sup>、GO<sup>[17]</sup>、COG<sup>[18]</sup>、KEGG<sup>[19]</sup>数据库比对,获得 Unigenes 的注释信息( $E\text{-value} < = 1e-05$ ),再将注释上的序列按照不同的功能划分成不同的类别。最后对 KEGG 注释的基因功能信息进行生物学通路的注释和预测。

### 1.6 差异表达 Unigene 的筛选

利用比对软件 Bowtie 将各样品测序得到的 reads 与 Unigene 库进行比对,然后根据比对信息并利用 RSEM<sup>[20]</sup>进行表达量水平估计,我们利用 RP-KM<sup>[21]</sup>值来反映对应 Unigenes 的表达丰度。选择 EBSseq<sup>[22]</sup>进行差异表达分析,利用 FDR 方法分析 Unigene 的表达,在差异基因筛选中,我们选取  $FDR < 0.01$  且  $\text{Fold Change} \geq 2$  作为标准定义为差异 Unigene。

### 1.7 差异 Unigene 的 GO 和 pathway 分析

GO 功能显著性富集分析能确定差异表达基因行使的主要生物学功能。通过 Pathway 显著性富集能确定差异表达基因参与的最主要生化代谢途径和信号转导途径。

## 2 结果与分析

### 2.1 高通量测序与功能注释

通过对红松的针叶、茎、雌花和球果的转录组进

行高通量的测序,总共获得 21.3 GB 的数据,经过序列拼接,获得 4 901 106 个 contig 片段,大于 300 bp 的有 35 331 个,大于 1 kb 的有 14 661 个。通过 Trinity 组装共得到 71 238 条转录本和 41 476 条 Unigene。

与 BLAST 序列比对,最终获得有注释信息的 Unigenes 有 26 849 个(占全部 Unigene 的 64.73%)。将组装得到的 41 476 条红松 Unigene 与 Nr 数据库比对,有 11 837 条与云杉(*Picea sitchensis*)、2 830 条与葡萄(*Vitis vinifera*)等植物的序列同源。

如图 1 所示,在 COG 功能分类体系中,获得的功能注释涉及 24 个 COG 功能类别,其中,General function prediction only 转录物的比例最大(2 216 条),Nuclear structure 最少(1 条),其他种类基因的表达丰度不尽相同。

图 2 所示,GO 富集性分析将注释的 Unigene 序列分成细胞组分(Cellular component)、分子功能(Mo-

lecular Function)、生物过程(Biological process)3 个大类 56 个小类,分别包含了 16、16、24 个功能亚类,其中,Cellular component 中,cell part、cell、organelle 基因最多;Molecular Function 中,catalytic activity、binding 基因最多;Biological Process 中,metabolic、cellular 基因最多。

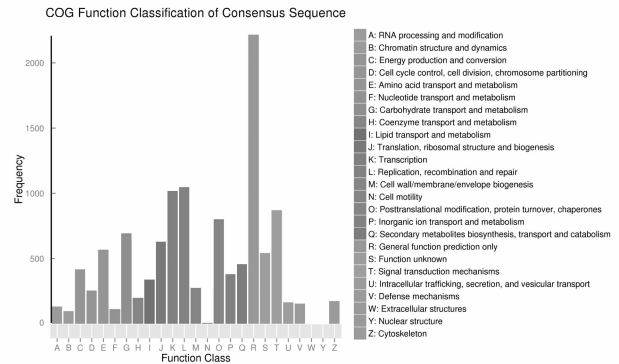


图 1 红松 unigene 的 COG 注释

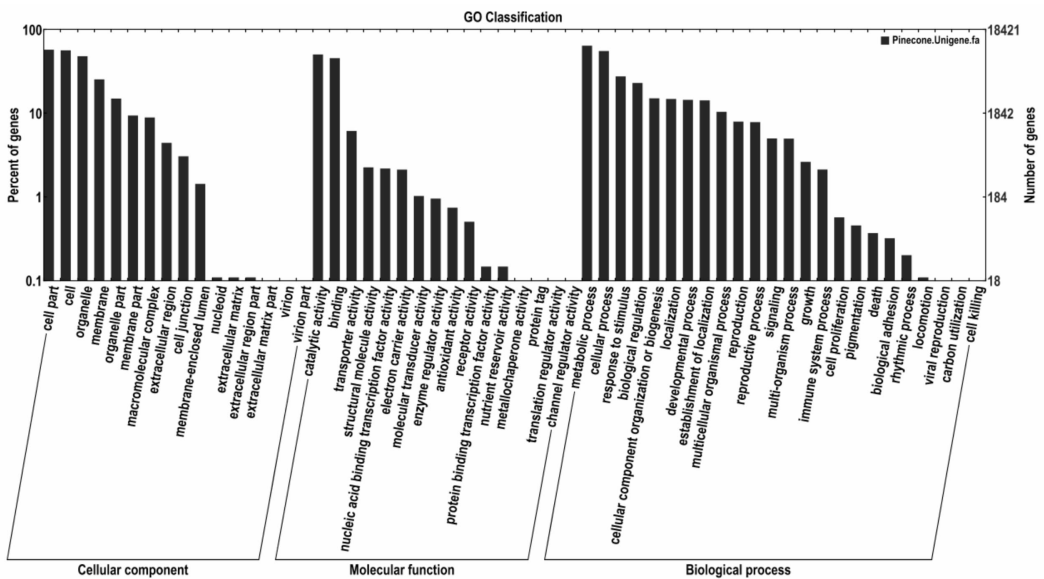


图 2 Unigene 的 GO 功能注释及分类统计结果

## 2.2 不同组织差异 Unigene 的 GO 分析

在转录组测序、组装和注释的基础上,将红松不同部位进行基因表达对比分析(见表 1)。发现嫩茎(T2)、花(T3)、球果(T4)分别与针叶(T1)相比,分别有 792 个、1 055 个、2 668 个 Unigene 上调表达,有 1 069 个、1 516 个、3 248 个 Unigene 下调表达;花(T3)、球果(T4)分别与嫩茎(T2)相比,分别有 834 个、2 841 个 Unigene 上调表达,有 802 个、3 016 个 Unigene 下调表达;球果(T4)与花(T3)相比,有 2 771 个 Unigene 上调表达,有 2 850 个 Unigene 下调

表达。研究发现球果分别与针叶、嫩茎、花的差异 Unigene 数量最多,说明球果在形成过程中的各途径中 Unigene 表达量丰富。通过 GO 分类,分别研究不同部位间差异 Unigene 的功能(见图 3)及不同部位间差异显著的 Unigene GO 富集。GO 功能分类表明,在所有组织部位间的差异基因,被分为细胞组分、分子功能、生物过程三大类。而细胞组分中 cell、organelle、cell part 三个亚类所占比例较高,分子功能中 catalytic activity、binding 两个亚类所占比例较高,生物过程中 metabolic proces、cellular process 所占比

例较高,同时,由图 3 可知,球果(T4)较针叶(T1)、茎(T2)、花(T3)四个组织部位的转录组差异 Unigene 在 GO 功能分类中所占的比例均较高。在此基础上,分别对不同部位间差异 Unigene 进行 GO 富集,我们以叶(T1)VS 果(T4)为例(见图 4),发现叶与果在 sequence-specific DNA binding transcription factor activity(122)、oxidoreductase activity(63)、metabolic process(183)、oxidation-reduction process(298)、carbohydrate metabolic process(83)、extracellular region(129)、plant-type cell wall(95)、cytoplasmic membrane-bounded vesicle(190)、apoplast(91)等方面差异显著。

方面差异显著。

表 1 注释的差异基因数量统计

Type	Total number	Up number	Down number
T1 vs T2	1 861	792	1 069
T1 vs T3	2 571	1 055	1 516
T1 vs T4	5 916	2 668	3 248
T2 vs T3	1 636	834	802
T2 vs T4	5 857	2 841	3 016
T3 vs T4	5 621	2 771	2 850

注:第一列表示样品组合,前一个样品为对照组,后一个样品为实验组。

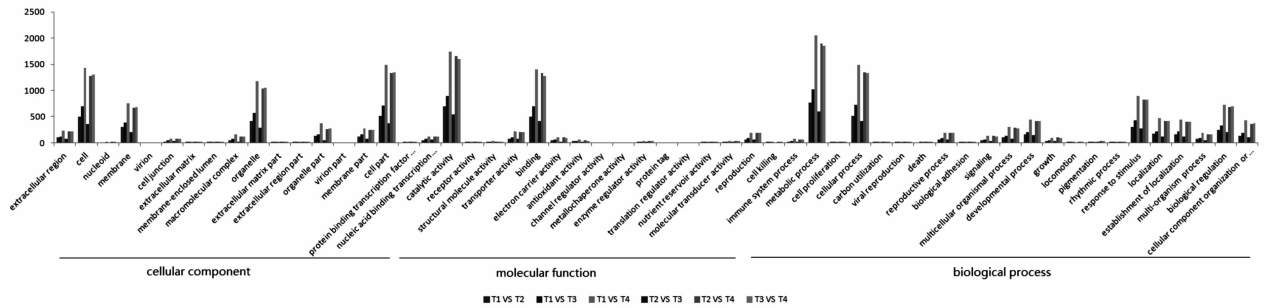


图 3 不同部位间差异 unigenes 的 GO 功能注释及分类统计

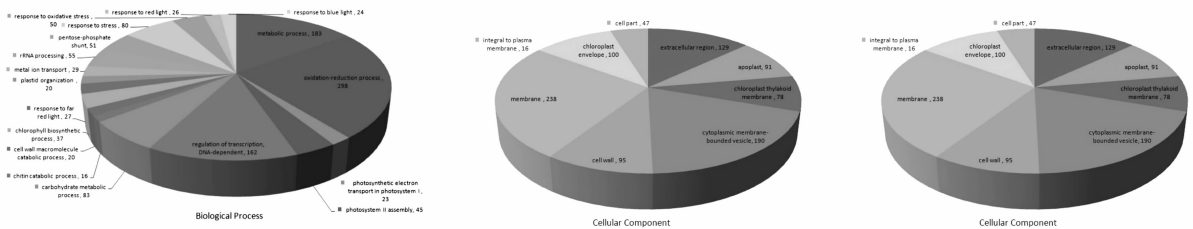


图 4 叶(T1)VS 果(T4)间的 GO 富集分析

## 2.3 不同组织差异 Unigene 的 pathway 注释分析

为了研究差异 Unigene 涉及的代谢途径,利用 KEGG 数据库分别对其进行了 pathway 富集分析(见图 5),同样以叶(T1)VS 果(T4)为例,发现叶与球果在 Phenylpropanoid biosynthesis(61)、Phenylalanine metabolism(43)、Photosynthesis(31)、Circadian rhythm(17)、Flavonoid biosynthesis(25) Diterpenoid biosynthesis(13)、Protein processing in endoplasmic reticulum(47)等途径差异显著。研究不同部位间差异 Unigene 的 pathway 富集分析的前 3 位差异 pathway 数据,表明,(T1)VS(T2)间与(T1)VS(T3)间的差异 pathway 一致,均为 Phenylpropanoid biosynthesis、Phenylalanine metabolism、Photosynthesis;(T1)VS(T4)间、(T2)VS(T4)间与(T3)VS(T4)间的差异

pathway 一致,均为 Phenylpropanoid biosynthesis、Phenylalanine metabolism、Protein processing in endoplasmic reticulum;(T2)VS(T3)间的差异 pathway 为 Phenylpropanoid biosynthesis、Phenylalanine metabolism、Plant hormone signal transduction。

## 2.4 红松次生代谢产物初探

2.4.1 脂肪酸合成途径相关基因 转录组数据库中注释到次生代谢产物代谢途径的 Unigene 基因有 669 个,代谢途径有 25 个不同的分支。对不同部位的转录组数据进行 Pathway 富集性分析发现共有 48 个 Unigene 可归类于脂肪酸生物合成途径,脂肪酸合成是油脂合成的基础,因此对脂肪酸合成途径基因的研究有助于更深入的认识油脂合成的机理。对每组两两比较获得的差异表达 Unigene 分别进行 Pathway

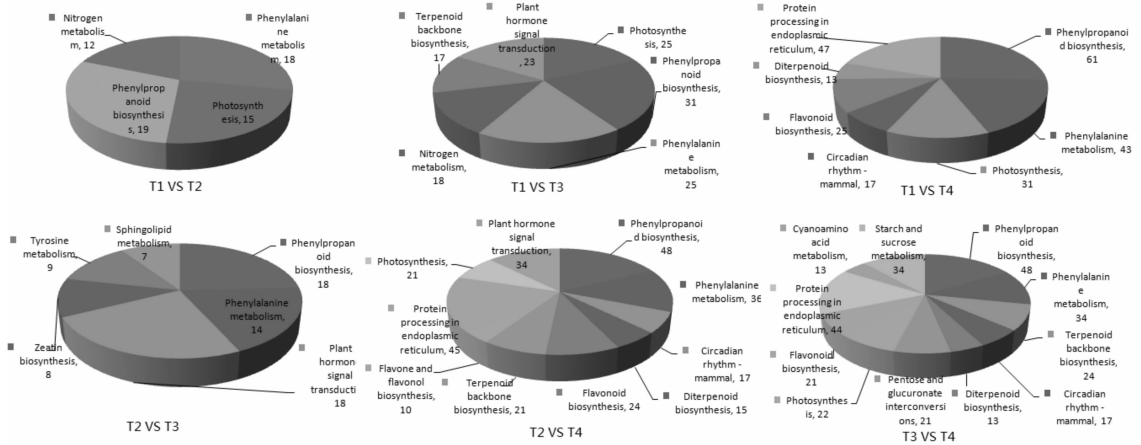


图5 不同部位差异 Unigene 的 KEGG 富集分析

富集性分析后发现,每组差异表达 Unigene 中都有 一些被归类于脂肪酸生物合成途径。将 48 个 Uni- gene 序列在 KEGG 数据库中比对后获得 16 个与 其他物种同源的脂肪酸合成相关的关键酶基因,片 段平均长度 1 397.65 bp,其中,KASI片段序列最长,为 2 542.5 bp;在表达水平上,以 FatB 和 SAD 最高,同时, 两个基因均在球果中表达量最高(见图6)。

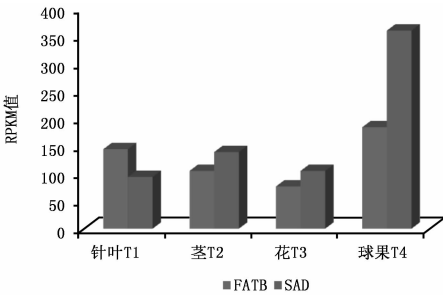


图6 FATB 与 SAD 基因在不同部位间的表达水平

2.4.2 挥发油、萜类化合物合成途径关键酶基因

萜类化合物是植物次生代谢物中的一类重要化 合物,被植物用于防御外病虫害,是针叶树的重要防 御物质。萜类化合物在针叶树中通常以复杂的混合物 形式存在,其数量和特性,会影响萜烯类化合物对潜 在的致病菌和植食性昆虫的防御作用<sup>[23]</sup>。目前,萜 类化合物已成为针叶树研究领域中的热点问题。萜 类合成酶(terpene synthases, TPS)是萜类化合物生 物合成过程中的关键酶之一。研究发现萜类化合物 生物合成有 2 条分别独立的途径,即甲羟戊酸途径 (MVA)和丙酮酸途径(DXP)<sup>[23]</sup>。这 2 条合成途径 分别在细胞溶质和细胞质体中形成 5C 单位的异戊 烯基二磷酸(IPP)和二甲丙烯焦磷酸(DMAPP),IPP 与 DMAPP 结合为牻牛儿基焦酸(GPP),GPP 去焦酸 后即生成单萜。IPP 与 GPP 头尾相连则产生法尼基

焦酸(FPP)释放出焦磷酸则可形成倍半萜。FPP 以 尾尾方式连接,即合成三萜类。IPP 和 FPP 相连会 产生牻牛儿基牻牛儿基焦酸(GGPP),去焦酸后形成 四萜。GGPP 可与多个 IPP 头尾连接合成多萜类产 物<sup>[24]</sup>。对不同部位的转录组数据进行 Pathway 富集 性分析发现共有 97 个 Unigene 可归类于萜类化合物 生物合成途径,共注释到 34 个萜类生物合成途径关 键酶基因,片段平均长度 880.6 bp,其中,GGPS 片 段最长,为 2 179.5 bp,在表达水平上,RPKM 值前 8 位 的基因在不同部位的基因表达量显示,在球果中的 萜类化合物基因表达量最少,除 HDR 外,在茎和花 中的表达量较高(见图7)。

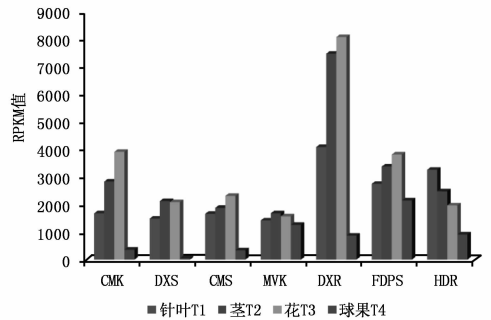


图7 RPKM 值前 8 位的基因在不同部位间的表达水平

3 结论与讨论

HiSeq2000 高通量测序数据量大,效率高,适合 于进行松属树种转录组的研究。Unigenes 及其功能 注释作为计算基因组学的产物,对育种来说就是一 笔宝贵的资源<sup>[25]</sup>。本研究采用 RNA-Seq 技术对红 松针叶、茎、花和球果 4 个部位间的转录组进行比 较研究,将红松不同部位间的差异表达的基因进行了 GO 和 Pathway 信号途径分类,初步明确了各个基因

编码的蛋白质的功能,为研究基因表达变化与油脂合成之间及与萜类化合物合成之间的关系奠定了研究基础,将有助于开展选择育种和基因工程研究。

本研究经过序列拼接,获得 4 901 106 个 contig 片段,通过 Trinity 组装得到 41 476 条 Unigene。通过对 Unigene 进行功能注释,最终获得有注释信息的 Unigenes 有 26 849 个,占全部 Unigene 的 64.73%。COG 功能分类将获得的功能注释涉及 24 个 COG 功能类别,GO 富集性分析将注释的 Unigene 序列分成 Cellular component、Molecular Function、Biological process 3 个大类 56 个小类。这一结果表明,采用高通量测序可以应用于红松不同部位间生长发育过程中表达的重要基因的研究。Li 等<sup>[26]</sup>采用高通量测序技术进行人参不同部位间的转录组分析,挖掘皂苷生物合成基因;李铁柱等<sup>[27]</sup>完成杜仲果实与叶片的转录组组装与功能注释。证实高通量测序应用于研究不同组织间的转录水平的可行性。Chen<sup>[28]</sup>等、Dawn E Hall 等<sup>[29]</sup>、Niu 等<sup>[30]</sup>、王晓峰等<sup>[31]</sup>分别采用高通量测序在针叶树中挪威云杉、美国黑松、油松、马尾松完成转录组测序,并进行基因资源的挖掘。

通过红松不同部位间的基因表达对比分析,发现,球果分别与针叶、嫩茎、花的差异 unigene 数量最多,说明生长发育期的球果存在着丰富的生化进程,处于基因高表达水平。由差异基因的 GO 分类中,在细胞组分中的 cell(细胞)、organelle(细胞器官)等亚类所占比例较高,分子功能中 catalytic activity(催化活性)、binding(拼接)两个亚类所占比例较高,生物过程中 metabolic proces(代谢过程)、cellular process(细胞过程)所占比例较高,而球果的高表达水平说明,差异基因的 GO 分类中主要体现为球果中含有大量快速生长相关的基因。不同部位差异 Unigene 的 pathway 注释分析与 GO 功能分析注释结果相识。

红松无论作为用材林,还是作为食品以及化工用品都具有较高的发展前景,目前对红松种仁中脂肪酸成分研究较多,表明,松仁脂肪酸成分中以不饱和脂肪酸为主,王振宇等<sup>[32]</sup>张振<sup>[33]</sup>等采用气质联用法检测了经超临界法提取的松仁油,不饱和脂肪酸的含量在 90% 以上,其主要成分为油酸、亚油酸、 $\alpha$ -亚麻酸和棕榈酸,其中亚油酸和  $\alpha$ -亚麻酸的含量最高。脂酰-酰基载体蛋白硫酯酶(acyl-acylcarrier protein thioesterase, Fat)是脂肪酸从头合成途径中

的关键酶,其主要功能是将脂酰基-ACP 水解成游离脂肪酸和 ACP,从而终止脂肪酸从头合成途径中脂肪酸链的延长<sup>[34]</sup>。由于底物不同,Fat 又被分为 FatA 和 FatB 两大家族,它们共同控制着植物体中饱和脂肪酸和不饱和脂肪酸的比例<sup>[35]</sup>。 $\Delta^9$ 硬脂酰 ACP 脱氢酶(stearoyl-ACP desaturase,SAD)在植物体中,以硬脂酰 ACP(stearoyl-ACP,C18-ACP)为底物,生成顺式不饱和双键的油酰基 ACP( $\Delta^9$ C18:1-ACP)接着在硫酯酶的作用下解离,转运至内质网进行三酰甘油的组装或再延长以及去饱和等过程,因此,SAD 是植物脂肪酸合成通路中一个重要的分支点,是不饱和脂肪酸合成中不可缺少的关键酶,直接决定饱和脂肪酸与不饱和脂肪酸的比例<sup>[36-37]</sup>,这对调节油料植物种子中脂肪酸组成具有重要意义。在表达水平上,以 FatB 和 SAD 最高,同时,两个基因均在球果中表达量最高。

植物体内许多萜烯成分的分布通常具有种属、器官、组织和生长发育阶段的特异性,表明了萜类合成酶(TPS)存在表达和调控的时空特异性。目前,在针叶树中已报道了 70 多种 TPS,已报道的针叶树 TPS 主要集中在松科(Pinaceae)植物中<sup>[23]</sup>。对萜类合成酶的深入研究,可以使我们通过 DNA 重组技术来改造萜类合成细胞中的代谢途径,以提高萜类最终产量或在不含萜类的生物中合成萜类<sup>[38]</sup>。红松转录组数据库的建立,获得了大量的转录本信息,可用于基因的克隆、表达谱分析、发现新基因、QTL 定位等多种用途,红松的研究多以定向培育、生长性状、化学成分分析等研究为主,其分子生物学研究进展较为缓慢,而红松转录组数据库的建立,为分子生物学的研究提供了重要的基础和依据;为红松次生代谢产物的生物合成与代谢调控提供了候选的基因资源,为红松活性成分含量指标和优良农艺性状改良奠定基础。

## 参考文献:

- [1] 马建路,庄丽文,陈 动,等.红松的地理分布[J].东北林业大学学报,1992,20(5):40-47.
- [2] Wolff R L, Pedrono F, Pasquier E, et al. General characteristics of *Pinus* spp. seed fatty acid compositions, and importance of Delta 5-olefinic acids in the taxonomy and phylogeny of the genus[J]. Lipids, 2000,35: 1-22.
- [3] Yoon T H, Im K J, Koh ET, et al. Fatty-Acid Compositions of *Pinus Koraiensis* Seed[J]. Nutr Res 1989,9:357-361.
- [4] Nergiz C, Donmez I. Chemical composition and nutritive value of *Pinus pinea* L. seed [J]. Food Chem, 2004,86: 365-368.

- [5] Zadernowski R, Naczek M, Czaplicki S. Chemical composition of *Pinus sibirica* nut oils[J]. Eur J Lipid Sci Tech, 2009,111; 698 – 704.
- [6] Li Y, Sun C, Luo H M, et al. Transcriptome characterization for *Salvia usung* 454 GS FLX[J]. Yao xue xue bao, 2010,45(4): 524 – 529.
- [7] Adams M. D, Kelley J M, Gocayne J D, et al. Complementary DNA sequencing: expressed sequence tags and human genome project [J]. Science, 1991,5013; 1651 – 1656.
- [8] Boguski M S, Tolstoshev C M, Bassett D E. Gene discovery in dbEST[J]. Science, 1994, 265(5181): 1993 – 1994.
- [9] Bouchez D, Höfte H. Functional genomics in plants [J]. Plant Physiol, 1998, 118(3): 725 – 732.
- [10] Allona I, Quinn M, Shoop E, et al. Analysis of xylem formation in pine by cDNA sequencing, Proc [J]. Natl. Acad. Sci., USA, 1998, 95(16): 9693 – 9698.
- [11] Chagn D, Chaumeil P, Ramboer A, et al. Cross-species transferability and mapping of genomic and cDNA SSRs in pines[J]. Theor. Appl. Genet, 2004, 109(6): 1204 – 1214.
- [12] Temesgen B, Brown G R, Harry D E, et al. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*pinus taeda* L.) [J]. Theor. Appl. Genet., 2001, 102(5): 664 – 675.
- [13] Manfred G G, Brian J H, Moran Y, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. Nature Biotechnology, 2011 (29): 644 – 652.
- [14] Stephen F A, Thomas L M, Alejandro A S, et al. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs[J]. Nucleic Acids Res, 1997 25(17): 3389 – 3345.
- [15] Yangyang DENG, Jianqi LI, Songfeng WU, et al. Integrated nr Database in Protein Annotation System and Its Localization [J]. Computer Engineering, 2006 32(5):71 – 74.
- [16] Apweiler R, Bairoch A, Wu C H, et al. UniProt: the Universal Protein Knowledgebase [EB/OL]. Nucleic Acids Res. <http://nar.oxfordjournals.org/cgi/content/full/32/suppl1/d115>, 2004,32(Database Issue):115.
- [17] Michael A, Catherine A B, Judith A B, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium[J]. Nat Genet, (25): 25 – 29.
- [18] Roman L T, Michael Y G, Darren A N, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution [J]. Nucleic Acids Res, 2000,128(1):33 – 36.
- [19] Minoru K, Susumu G, Shuichi K, et al. The KEGG resource for deciphering the genome[J]. Nucleic Acids Res, 2004, (32):277 – 280.
- [20] Bo Li, Colin N D. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome[J]. BMC Bioinformatics, 2011, 12:323 – 326.
- [21] Ali M, Brian A W, Kenneth M, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq[J]. Nature methods, 2008(5): 621 – 628.
- [22] Ning L, John A D, James A T, et al. EBSseq: an empirical Bayes hierarchical model for inference in RNA-seq experiments [J]. Bioinformatics, 2013 (29): 1035 – 1043.
- [23] 龚 治,李典谟,张 真,针叶树萜类合成酶基因研究进展[J]. 林业科学,2010,46(1):123 – 130.
- [24] 申培林,杨铁刹,张小全,等.烟草类萜类生物合成途径中关键酶基因的克隆与表达调控[J].江苏农业科学,2012,40(1): 37 – 40.
- [25] Pavy N, Laroche J, Bousque J, et al. Large-scale statistical analysis of secondary xylem ESTs in pine[J]. Plant Molecular Biology, 2005, 57:203 – 224.
- [26] Li C,Zhu Y,Guo X. et al. Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer[J]. BMC Genomics,2013, 14: 245 – 256.
- [27] 李铁柱,杜红岩,刘慧敏,等.杜仲果实和叶片转录组数据组装及基因功能注释[J].中南林业大学学报,2012,32(11):122 – 130.
- [28] Jun Chen, Severin U, Niclas G, et al. Sequencing of the needle transcriptome from Norway spruce (*Picea abies*Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms[J]. BMC Genomics2012, 13:589 – 604.
- [29] Dawn E H, Macaire M S Y, Sharon J, et al. Transcriptome resources and functional characterization of monoterpene synthases for two host species of the mountain pine beetle, lodgepole pine (*Pinus contorta*) and jack pine (*Pinus banksiana*) [J]. BMC Plant Biology,2013,13:80 – 94.
- [30] Shi-Hui Niu, Zhe-Xin Li, Hu-Wei Yuan, et al. Transcriptome characterisation of *Pinus tabulaeformis* and evolution of genes in the *Pinus* phylogeny[J]. BMC Genomics2013, 14:263 – 275.
- [31] 王晓锋,何卫龙,蔡卫佳,等.马尾松转录组测序和分析[J].分子植物育种,2013,11(3):385 – 392.
- [32] 王振宇,牛之瑞.红松种仁不饱和脂肪酸研究进展[J].油脂工程,2007,2:58 – 61.
- [33] 张 振,张 磊,张含国,等.铁力红松种子园无性系种子形态及营养成分变异研究[J].植物研究,2014,34(3): 356 – 363.
- [34] Jones A, Davies H M, Voelker T A. Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases [J]. Plant Cell, 1995, 7: 359 – 371.
- [35] Wu P Z, Li J, Wei Q, et al. Cloning and functional characterization of an acyl-acyl carrier protein thioesterase (JeFatB1) from *Jatropha curcas* [J]. Tree Physiol, 2009, 29: 1299 – 1305.
- [36] Luo T, Peng S M, Deng W Y, et al. Characterization of a new stearoyl-acyl carrier protein desaturase gene from *Jatropha curcas* [J]. Biotechnology Letters,2006,28: 657 – 662.
- [37] Slocombe S P, Cummins I, Jarvis R P, et al. Nucleotide sequence and temporal regulation of a seed-specific *Brassica napus* cDNA encoding a stearoyl-acyl carrier protein (ACP) desaturase [J]. Plant Mol Biol,1992,20: 151 – 155.
- [38] 黄 瑛,曾庆平.2006.萜类生物合成的基因操作[J].中国生物工程杂志, 26(1): 60 – 64.