银杏细胞转录组高通量测序及分析

张 楠[1] 孙桂玲[2] 戴均贵[3] 杨艳芳[1] 刘洪伟[1] 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 北京 100091)

(2 中国科学院昆明植物研究所 资源植物与生物技术重点实验室 昆明 650201)

(3 中国医学科学院/北京协和医学院 药物研究所天然药物活性物质与功能国家重点实验室 北京 100050)

摘要 利用 Illumina 的 Genome Analyzer IIx 对银杏(*Ginkgo Biloba*)细胞转录组进行高通量测序,挖掘银杏内酯和紫杉醇生物合成基因,特别是新的羟基化酶基因,为今后最终完善红豆杉细胞紫杉醇生物合成途径中未知的羟基化步骤作准备。通过测序,获得了银杏细胞 69 286 个 contig,56 387个 scaffold,32 032 个 unigene。Unigene 平均长度 636bp。另外从 gap 分布、GC 含量、基因组 coverage 等方面对 unigene 进行评估,数据显示测序质量好,可信度高。通过分析 unigene 的表达和功能注释等信息,发现 66 个属于 CYP450 基因家族,726 个参与次生代谢物合成,其中 59 个与萜类合成有关,17 个与二萜类合成相关。利用生物信息学方法从 Michigan State University 银杏成熟叶、侧根、成熟果实、无菌苗以及次生茎的转录组数据中找到了与银杏细胞 CYP450 高度同源的紫杉烷羟基化酶候选基因 15 个,为后续研究奠定了基础。

关键词 银杏细胞 Genome Analyzer IIx 转录组

中图分类号 Q949

"转录组"最先由 Velculescu 等[1]人提出,指细胞在特定状态下全部表达的 RNA 的总和,反映相同基因在不同条件下表达水平的差异,并能揭示不同基因的相互作用及各自功能。与起初的基因组学研究相比,转录组学更具动态性,能进一步表现细胞内的生命活动,为诠释细胞的功能提供了更具价值的参考信息[2]。基于转录组学在功能基因组学研究中的重要价值,我们利用 Illumina 的 Genome Analyzer IIx 高通量测序平台研究了银杏细胞的转录组并对其进行了分析,以期挖掘新的功能基因。因为 Illumina 的 Genome Analyzer IIx 系统能迅速、准确地确定碱基序列,比传统基因挖掘方法更具优势。

银杏是一种重要的药用植物,其细胞转录组的研究有可能发现一些与萜类活性成分(如银杏内酯)生物合成有关的候选基因。红豆杉(*Taxus*)含有具抗癌作用的紫杉醇已广为所知,但由于天然植物红豆杉的紫杉醇产量有限,要大规模生产相关药物就必须开发新途径,寻找紫杉醇生物合成途径中的关键酶基因就是值得探索的方向[3-4]。综合国内外研究,在红豆杉细胞紫杉醇母核的生物合成途径中,除紫杉烷 9α-羟基化酶(taxoid 9 alpha-hydroxylase)和紫杉烷 1β-羟基化酶(taxoid 1 beta-hydroxylase)基因外,其余的羟基化酶基因均得到克隆[5]。参与紫杉醇生物合成的羟基化酶都属于细胞色素 P450(cytochrome P450,简称 CYP450)。由于红豆杉细胞中的 *CYP450* 基因数量太多,美国华盛顿州立大学 Rodney Croteau 院士实验室研究了多年都没有从红豆杉中克隆到紫杉烷 9α-羟基化酶基因[7],这提示我们从红豆杉中直接克隆到此基因的难度很大。2001 年戴均贵等研究发现培养的银杏(*Ginkgo Biloba L.*)悬浮细胞能生物转化 2α,5α,10β,14β-四乙酰氧基-紫杉-4(20),11-二烯(sinenxan A,简称 SIA)(图 1),即银杏细胞可以对 SIA 进行特异性的 9α 羟基化且产率接近 70%[8-12],此结果表明银杏细胞含有紫杉烷 9α-羟基化酶的活性,对紫杉醇衍生物具有相似的底物催化功能。这一结果启发我们,通过研究银杏细胞的转录组寻找紫杉烷 9α-羟基化酶(taxoid 9alpha-

收稿日期: 2013-01-15 **修回日期:** 2013-02-28

* **通讯作者,电子信箱:** qiudy@caf.ac.cn

hydroxylase)编码基因,再 blast 红豆杉基因序列,就有可能发现红豆杉细胞中与之同源的基因,并最终鉴定出红豆杉中紫杉烷9α-羟基化酶,以完善红豆杉紫杉醇生物合成这一重要代谢步骤。现将银杏细胞有关转录组的测序方法及结果报道如下。

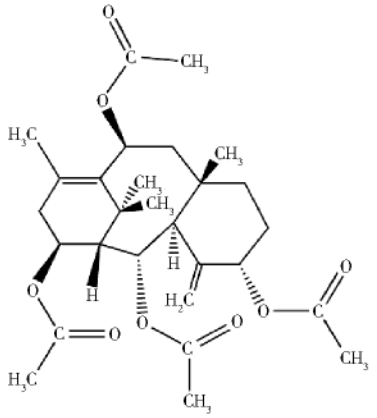

图1　Sinenxan A 的结构(分子式 $C_{28}H_{40}O_8$)

Fig. 1　The structure of sinenxan A (Formula $C_{28}H_{40}O_8$)

1　材料与方法

1.1　材　料

银杏愈伤组织在添加 0.2mg/L 6-BA,1.5mg/L 2,4-D 和 2mg/L IAA 的 MS 固体培养基上培养,取培养后 10d 新鲜的银杏继代愈伤组织 1g,迅速用液氮冷冻,存于 -80℃备用。

1.2　研究方法

根据 Chang 等人的 CTAB 法[13] 提取细胞的总 RNA,用带有 Oligo(dT)的磁珠富集 mRNA,加入破碎缓冲液(fragmentation buffer)打断 mRNA 为 200~700 的短片段,以此为模板,用六碱基随机引物(random hexamers)合成第一条 cDNA 链;再加入缓冲液、dNTPs、RNase H 和 DNA polymerase I 合成第二条 cDNA 链;cDNA 用 QIAquick PCR Purification Kit(Qiagen 试剂公司)纯化,经 EB 缓冲液洗脱后进行末端修复、加 polyA 及连接测序接头;采用琼脂糖凝胶电泳筛选不同大小的片段,之后 PCR 扩增,建好的文库最后用 Illumina GA IIx 进行双末端测序。原始的测序结果去除制备文库时产生的接头序列、两端低质量序列、低复杂度序列、长度小于 75bp 的序列,再进行转录组序列的 SOAPdenovo[14] 从头组装,最后对各个序列作注释与分类。为了统计所有匹配到 repeats,rRNA,tRNA,snoRNA 等非编码且保守 RNA 的原始 Reads 数目,我们首先利用 RepeatMasker(http://www.repeatmasker.org)和 INFERNAL 软件(http://rfam.sanger.ac.uk/)[15],找出所有的含有 repeats 和 ncRNAs 的 unigenes,之后用 RSEM 软件[16] 找出所有匹配以上 unigenes 的所有原始 reads。除了我们自己测定的银杏细胞转录组,我们还从 GenBank SRA database(http://www.ncbi.nlm.nih.gov/sra)中下载了 Michigan State University 银杏成熟叶、侧根、成熟果实、无菌苗以及次生茎这 5 种组织的转录组数据(SRA ID: SRR325161、SRR325162、SRR325164、SRR325165、SRR325166),并用 ABySS 软件[17] 进行组装,通过 CAP3 软件[18] 处理以去除冗余。通过生物信息学方法,利用我们银杏细胞转录组测序得到的 CYP450 unigene 从 Michigan State University 银杏转录组数据中 blast 出可能的 P450 羟基化酶基因。

2　结　果

2.1　银杏细胞转录组从头组装与序列拼接

2.1.1　银杏细胞 contig 序列　读碱基程序(base calling)将测序的原始图像还原为序列数据,即原始读序(raw reads),去除杂质序列后得到 clean reads。测序产量统计原始读序 30 740 036 个,总核苷酸数 2 305 502 700bp。短读序组装软件 SOAPdenovo 能从头组装转录组:首先将具有一定长度重叠(overlap)的 reads 连成更长的片段,生成不含未知碱基的片段称为 contig,再将 reads 与 contig 比对,通过 paired-end reads 确定来自同一转录本的不同 contig 以及它们之间的距离,SOAPdenovo 将其连在一起,中间的未知序列用 N 表示,得到 scaffold。进一步利用 paired-end reads 对 scaffold 做补洞处理,最后得到含 N 最少,两端不能再延长的序列即是 unigene。测序结果显示:银杏细胞 Contig 平均长度 344bp,总条数 69 286,总碱基数 23 840 388。Contig N50 是 484bp。银杏细胞 contig 组装质量统计见表1,长度分布见图2。

表1　银杏细胞 contig 的组装质量统计

Table 1　*Ginkgo biloba L.* cells contig quality

长度 Length	数量 Number	百分比 Percent
100-200nt	32 584	47.03%
200-300nt	13 643	19.69%
300-400nt	6 860	9.90%
400-500nt	4 028	5.81%
>=500nt	12 171	17.57%

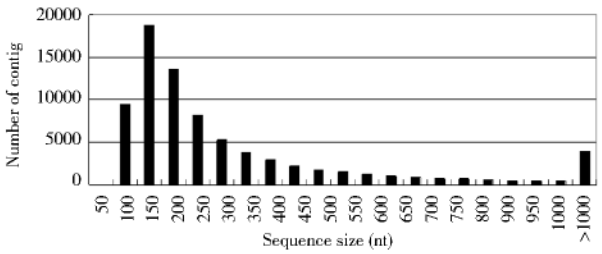

图 2 银杏细胞 contig 长度分布

Fig. 2 *Ginkgo biloba L.* cells contig sequence size

2.1.2 **银杏细胞** unigene **序列** **银杏细胞** scaffold **平均长** 424bp，**总条数** 56 387，N50 **为** 713bp。**由** scaffold **组装出** unigene 32 032 **个，平均长** 636bp，N50 **为** 904 bp，**所有** unigene **中能确定方向的序列有** 23 194 **个，不能确定方向的有** 8 838 **个，序列联配时** 93.7% **未见** gap **空位**（30 020 **个**），**表明测序质量很高**（**图** 3）。

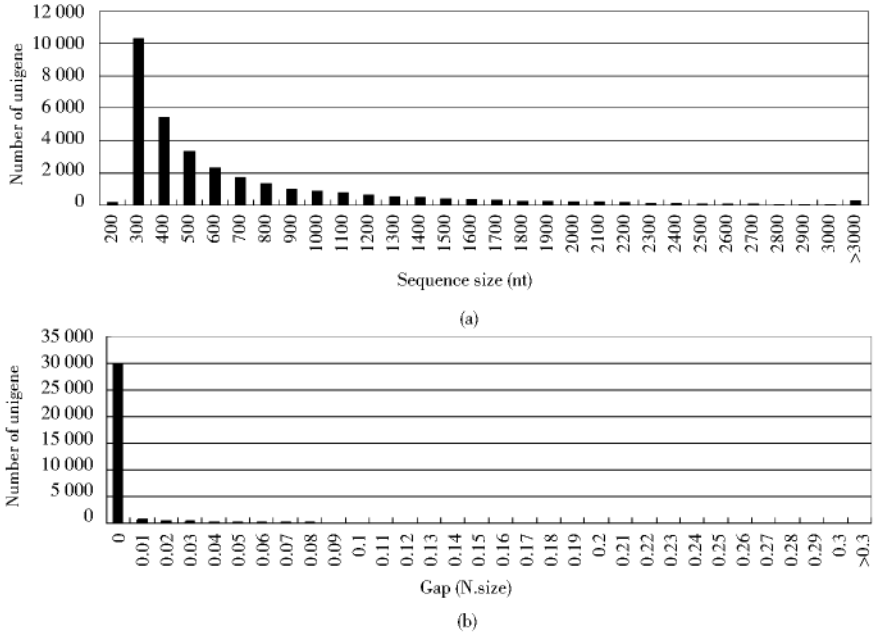

图 3 银杏细胞 unigene 的长度分布和 gap 分布

Fig. 3 Length and gap distribution of *Ginkgo biloba L.* cells unigenes

a: Length distribution of *Ginkgo biloba L.* cells unigenes b: Gap distribution of *Ginkgo biloba L.* cells unigenes

GC **含量是基因组碱基序列的重要特征之一，能反映基因结构、功能与进化信息，**GC **分布不均匀导致基因组不同** GC **含量的片段其性质和功能也有差异。银杏细胞** Unigene **中** GC% **为** 40% ~50% **的序列**（22 994 **个**）**占总序列的** 72%，GC% **为** 30% ~40% **的序列**（7 881 **个**）**占总序列** 25%，GC% **为** 50% ~60% **的序列**（1 085 **个**）**占总序列** 3%，**另外** GC% **为** 20% ~30% **的序列有** 53 **个，**GC% **为** 60% ~70% **的序列有** 16 **个，而百分含量过高**（**大于** 70%）**或过低**（**小于** 20%）**的** Unigene **不存在，表明** GC **含量基本呈正态分布。**

为了进一步挖掘银杏转录组的特征，我们使用 RepeatMasker **和** INTERNA **软件对所有的** Unigenes **进行了预测，结果表明** 2 265 **条** unigenes **被预测为** repeats、**及** rRNA、tRNA、snoRNA **等非编码** RNA **基因。经** RSEM **软件分析，发现** 701 727 **条原始** reads（**占所有原始** Reads **的** 2.28%）**匹配到了** repeats，rRNA，tRNA，snoRNA **等非编码且保守** RNA **基因。**

2.1.3 **银杏细胞测序深度与覆盖率** **分析基因组遗传变异时，基因组的测序深度**（depth）**指借助测序平台从基因组中读出的碱基总量**（bp）**与该基因组样本平均容量的比值，为评价测序质量的指标之一。基因组覆盖率**（coverage）**可增加检测基因组上每个碱基的可靠性，能更好地识别基因组中变异的发生频率。深度与覆盖率之间呈正相关，测序的错误率或假阳性会随着深度的提升而下降。重测序的个体，当测序深度在** 10 ~15X **以上时，覆盖率和测序错误率能得以控制。银杏**

细胞的 Unigene 覆盖率在 95% 以上的序列共 22 624 个,占所有 Unigene 的 70.6%,测序深度在 10X 以上的序列有 15 364 个,占所有 Unigene 的 48.0%。

2.2 基因表达与功能注释分析

将 blast 比对结果中 rank 最高的蛋白确定为该 unigene 的编码区序列(CDS),根据标准遗传密码表将编码区序列翻译成氨基酸序列(序列方向 5′→3′),再用软件 ESTScan[19] 预测未能与任一数据库比对上的 Unigene,得到预测的编码区核酸序列(序列方向 5′→3′)和氨基酸序列。在银杏细胞 Unigene 中,用 Blast 搜索分别得到 31 982 个碱基序列和 CDSs,ESTScan 预测了 1 851 个核苷酸序列和 CDSs。按照 nr、SwissProt、KEGG[20] 和 COG 的优先级顺序将 All-Unigene 序列与蛋白库做 blastx 比对(evalue < 0.000 01),若某 unigene 序列比对上高优先级数据库中的蛋白,则不进入下一轮比对,否则自动跟下一个库进行比对,如此循环直到与所有蛋白库比对完为止。注释分析给出 All-Unigene 的表达量和功能注释等信息,其中功能注释信息包括 Swissprot 注释、Pathway 注释、COG 功能注释以及 Gene Ontology(GO)功能注释。All-Unigene 的表达量是根据多少 reads map 到 unigene 上来统计的。本测序中属于 CYP450 的 unigene 共 66 个,通过生物信息学方法,利用我们银杏细胞 CYP450 unigene 序列从 Michigan State University 银杏转录组序列中 blast 出可能的羟基化酶基因 1 412 个,其中紫杉烷 10β 羟基化酶同源基因 281 个,紫杉烷 13α 羟基化酶同源基因 287 个,紫杉烷 2α 羟基化酶同源基因 282 个,紫杉烷 7β 羟基化酶同源基因 282 个,紫杉烷 5α 羟基化酶同源基因 280 个;经比对,挑选与红豆杉羟基化酶序列一致性(identity)高、score 分值大于 100 且在银杏各组织(成熟叶片、侧根、成熟果实、无菌苗、次生茎)中显著表达的候选基因 15 个,目前正在克隆这 15 个基因的全长,并将进行有关蛋白质的表达与功能鉴定工作。

2.2.1 Unigene KEGG 注释 KEGG 是分析基因产物在细胞中的代谢途径以及这些基因产物的功能的数据库,利用它可以进一步研究基因在生物学上的复杂行为。根据 KEGG 的注释信息能进一步得到 unigene 的 pathway 注释[21-22]。银杏细胞 unigene 中共有 726 个与次生代谢物生物合成有关,这些合成途径主要有:花青素合成途径(anthocyanin biosynthesis [PATH: ko00942])、甜菜红碱合成途径(betalain biosynthesis [PATH: ko00965])、油菜素内酯合成途径(brassinosteroid biosynthesis [PATH: ko00905])、咖啡因合成途径(caffeine metabolism [PATH: ko00232])、类胡萝卜素合成途径(carotenoid biosynthesis [PATH: ko00906])、二萜合成途径(diterpenoid biosynthesis [PATH: ko00904])、黄酮和黄酮醇合成途径(flavone and flavonol biosynthesis [PATH: ko00944])、类黄酮合成途径(flavonoid biosynthesis [PATH: ko00941])、芥子甙合成途径(glucosinolate biosynthesis [PATH: ko00966])、吲哚生物碱合成途径(indole alkaloid biosynthesis [PATH: ko00901])、异喹啉生物碱合成途径(isoquinoline alkaloid biosynthesis [PATH: ko00950])、柠檬烯和松萜合成途径(limonene and pinene degradation [PATH: ko00903])、类单萜合成途径(monoterpenoid biosynthesis [PATH: ko00902])、苯丙烷合成途径(phenylpropanoid biosynthesis [PATH: ko00940])、芪类化合物,二芳基庚烷和姜醇合成途径(stilbenoid, diarylheptanoid and gingerol biosynthesis [PATH: ko00945])、类萜骨架合成途径(terpenoid backbone biosynthesis [PATH: ko00900])、萜类生物碱,哌啶生物碱和嘧啶生物碱合成途径(tropane, piperidine and pyridine alkaloid biosynthesis [PATH: ko00960])、玉米素合成途径(zeatin biosynthesis [PATH: ko00908])。其中参与二萜合成途径(diterpenoid biosynthesis [PATH: ko00904])和类萜骨架合成途径(terpenoid backbone biosynthesis [PATH: ko00900])的 unigens 分别有 17 和 59 个。

2.2.2 Unigene GO 注释 通过与拟南芥的蛋白质组序列比对,获得银杏细胞 unigene 的 GO 分类信息。Gene Ontology(简称 GO)是一个国际标准化的基因功能分类体系,它提供了一套标准词汇表(controlled vocabulary)来全面描述生物体中基因和基因产物的属性。GO 有三个本体(ontology),分别描述基因的分子功能(molecular function)、所处的细胞组分(cellular component)、参与的生物过程(biological process)。根据 nr 注释信息,使用 Blast2GO 软件[23] 得到 unigene 的 GO 注释信息。Blast2GO 是被广泛认可的 GO 注释软件,之后再用 WEGO 软件[24] 对所有 unigene 进行 GO 功能分类统计,从宏观上认识银杏细胞的基因功能分布特征。银杏细胞 unigene 中有 3 876 条归入分子功能,5 111 条归入细胞组分,4 856 条归入生物过程(图 4)。

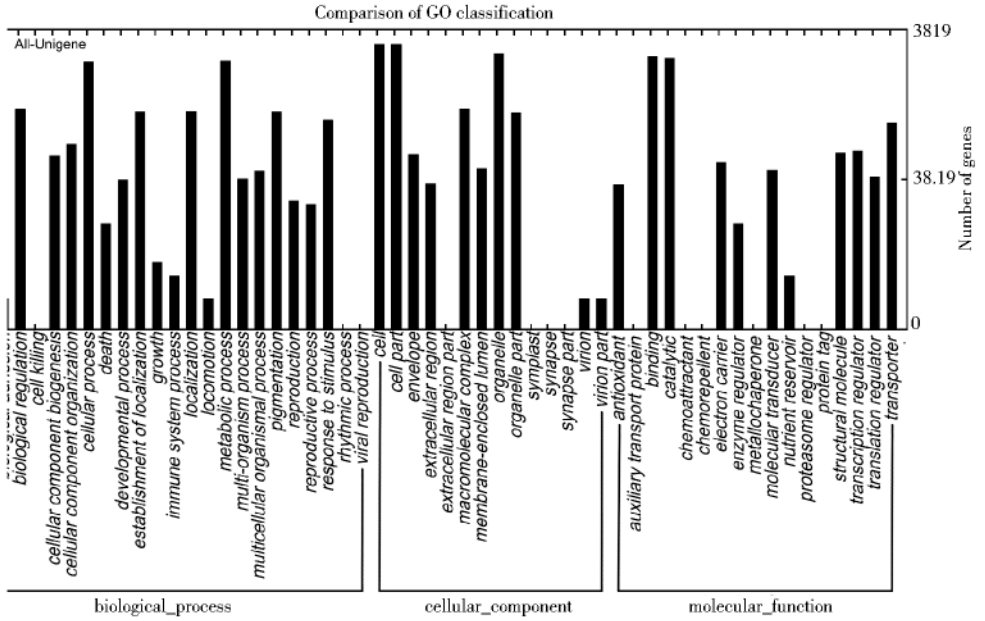

图 4 银杏细胞 GO 分类

Fig. 4 Comparison of *Ginkgo biloba L.* cells GO classification

2.2.3 Unigene COG 注释 COG 是对基因产物进行直系同源分类的数据库，每个 COG 蛋白都被假定来自祖先蛋白，COG 数据库是根据细菌、藻类、真核生物的完整基因组的编码蛋白、系统进化关系而构建的。将银杏细胞 unigene 和 COG 数据库进行比对，预测 unigene 可能的功能并对其做功能分类统计，结果表明有相当多的 unigene(262 个) 参与次生代谢物的合成、运输与降解(图 5)。

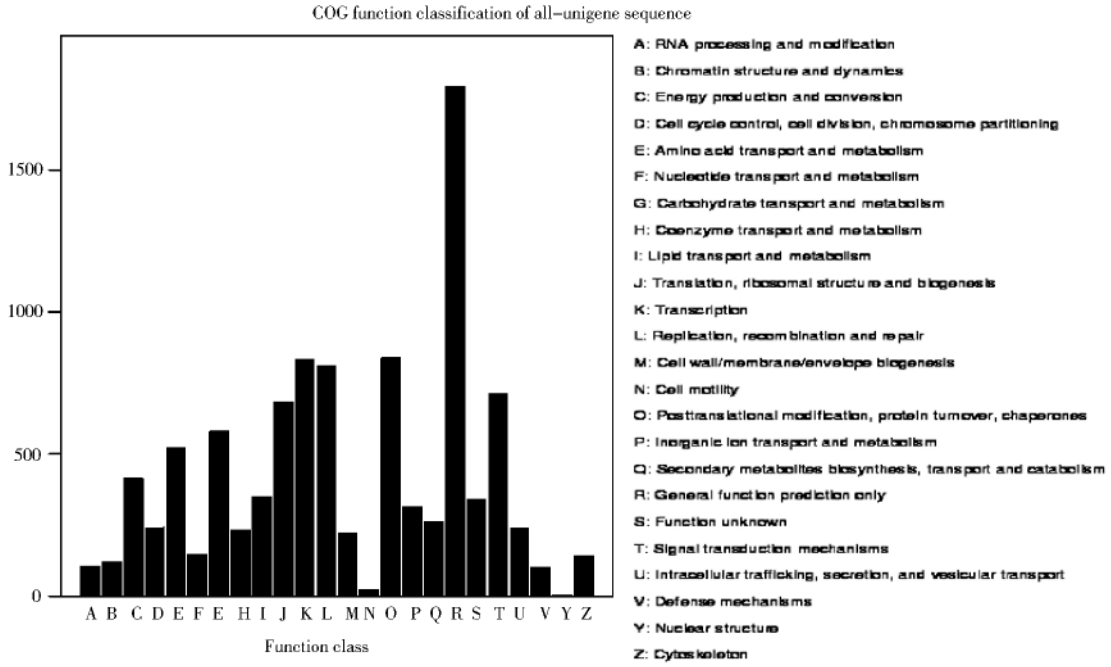

图 5 银杏细胞 Unigene COG 功能分类

Fig. 5 COG function classification of *Ginkgo biloba L.* cells unigenes

3 讨 论

银杏内酯和紫杉醇的生物合成主要以萜类和二萜类的次生代谢为主,本实验测得的 Unigene 中参与萜类代谢的有 59 个,二萜类的 17 个。生成 Taxol 的主要反应涉及两大部分:紫杉烷母核巴卡亭Ⅲ和苯基异丝氨酸侧链。两部分分别合成再连接形成紫杉醇。在母核形成过程中,所有羟基化反应均依赖细胞色素(Cytochrome)P450 单加氧酶催化[25-30],具体有紫杉烷 13α-羟基化酶、紫杉烷 9α-羟基化酶、紫杉烷 7β-羟基化酶、紫杉烷 2α-羟基化酶、紫杉烷 1β-羟基化酶等,最终形成巴卡亭Ⅲ[31-41],除紫杉烷 9α-羟基化酶和紫杉烷 1β-羟基化酶外,其余羟基化酶的编码基因均已被成功克隆。本测序中属于 CYP450 的 Unigene 共 66 个,但由于测序数据为 2G,片段大多不是全长。利用生物信息学方法从 Michigan State University 5 种银杏组织转录组测序得到的序列中我们又 blast 出可能的 P450 羟基化酶基因 1412 个,其中有 10β 羟基化酶、13α 羟基化酶、2α 羟基化酶、7β 羟基化酶、5α 羟基化酶的候选基因,由于 Michigan State University 5 种银杏组织转录组数据比我们的银杏组织转录组数据要大很多,这样就可以使有关羟基化酶候选基因的长度变得更长,从而为今后克隆有关基因的全长并进行蛋白功能鉴定打下了很好的基础。我们的结果也说明自己少量的转录组测序数据结合网上大量的同一物种的转录组测序数据,可获得序列长度更长甚至全长的有关候选基因。如果能从中发现银杏的 9α 羟基化酶基因同时证实其功能,并在红豆杉组织中克隆到序列相似、功能相同的基因,则将完善红豆杉中紫杉醇合成途径的 P450 羟基化反应,有关研究正在进行中,结果将另文报道;另一方面,由 sinenxan A 经银杏细胞转化生成的 NPB-014 产物是一个具有全新结构、理想的肿瘤 MDR 逆转剂的候选药物[8-12,42],该酶的成功克隆将有助于肿瘤 MDR 逆转剂 NPB-014 的大量合成,为 NPB-014 的商业化生产起到一定推动作用。

参考文献

[1] Velculescu V E, Zhang L, Zhou W, et al. Characterization of the yeast transcriptome. Cell, 1997, 88: 243-251.

[2] 吴琼,孙超,陈士林,等. 转录组学在药用植物研究中的应用. 世界科学技术(中医药现代化),2010,12(3):457-461.
Wu Q, Sun CH, Chen SH L, et al. Application of transcriptomics in the studies of medicinal plants. World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica, 2010, 12(3): 457-465.

[3] 冯磊. 中国红豆杉紫杉烷 13α-羟基化酶基因的克隆及表达. 曲阜:曲阜师范大学, 2006.
Feng L. Cloning and expression of taxane 13 alpha hydroxylase gene from Taxus chinensis. Qufu: Qufu Normal University, 2006.

[4] 胡国武,温廷益,元英进. 紫杉醇合成代谢途径中紫杉烯合成酶 cDNA 的克隆. 生物工程学报,2000,16(2):158-160.
Hu G B, Wen T Y, Yuan Y J. Cloning of taxadiene synthase cDNA from the cell Line of Taxus cuspidate. Chinese Journal of Biotechnology, 2000, 16(2): 158-160.

[5] 胡国斌,丛日山,梅兴国,等. 中国红豆杉细胞紫杉醇合成期特异表达新基因 TS1-FL 的部分 cDNA 克隆. 武汉大学学报(理学版),2004,50(4):472-476.
Hu G B, Cong R SH, Mei X G, et al. Partial cDNA Cloning of a novel gene TS1-FL specifically expressed in the taxol synthesis phase of Taxus chinensis cells. J. Wuhan Univ. (Nat. Sci. Ed.), 2004, 50(4): 472-476.

[6] 胡国斌,梅兴国,龚伟,等. 中国红豆杉细胞紫杉醇合成期与非紫杉醇合成期基因表达差 异初步分析. 生物工程学报,2002,18(4):512-515.
Hu G B, Mei X G, Gong W, et al. Differences in gene expression between Taxus chinensis cells during taxol-synthesis phase and those during non-taxol-syn thesis phase. Chinese Journal of Biotechnology, 2002, 18(4): 512-515.

[7] Croteau R, Ketchum R E, Long R M, et al. Taxol biosynthesis and molecular genetics. Phytochem Review, 2006, 5(1): 75-97.

[8] Dai J G, Yang L, Sakai J I, et al. A Taxatetraene from microbial transformation of Sinenxan A. Chinese Chemical, 2005, 16(6): 738-742.

[9] Dai J G, Guo H Z, Lu D D, et al. Biotransformation of 2a, 5a, 10β, 14 β, tetra-acetoxy-4 (20), 11-taxadiene by Ginkgo cell suspension cultures. Tetrahedron, 2001, 2: 4677-4679.

[10] Dai J G, Zhang M, Ye M, et al. Biotransformation of 14-Deacetoxy-13-oxo sinenxan A by Ginkgo Cell Cultures. Chinese Chemical, 2003, 14(8): 804-806.

[11] Dai J G, Ye M, Guo H Z, et al. Substrate specificity for the hydroxylation of polyoxygenated 4(20), 11-taxadienes by Ginkgo cell suspension cultures. Bioorganic Chemistry, 2003, 31: 345-356.

[12] Dai J, Cui Y, Zhu W, et al. Biotransformation of 2alpha, 5alpha, 10beta, 14beta-tetraacetoxy-4 (20), 11-taxadiene by cell suspension cultures of Catharanthus roseus. Planta Med, 2002, 68: 1113-1117.

[13] Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from Pine Tree. Plant Mol Biol Rep, 1993, 11(2):

113-116.

[14] Li R Q, Zhu H M, Ruan J. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res, 2010, 20: 265-272.

[15] Griffiths-Jones S, Bateman A, Marshall M, et al. Rfam: an RNA family database. Nucleic Acids Res, 2003, 31(1): 439-441.

[16] Li B, Dewey C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 2011, 12: 323.

[17] Simpson JT, Wong K, Jackman SD, et al. ABySS: a parallel assembler for short read sequence data. Genome Res, 2009, 19: 1117-1123.

[18] Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res, 1999, 9: 868-877.

[19] Iseli C, Jongeneel C V, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. ISMB, 1999, 138-148.

[20] Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res, 2008, 36: D480-D484.

[21] 林萍,曹永庆,姚小华,等. 普通油茶种子4个发育时期的转录组分析. 分子植物育种, 2011, 9(4): 498-505.
Lin P, Cao Y Q, Yao X H, et al. Transcriptome analysis of Camellia oleifera Abel seed in four development stages. Molecular Plant Breeding, 2011, 9(4): 498-505.

[22] 李滢,孙超,罗红梅,等. 基于高通量测序454 GS FLX的丹参转录组学研究. 药学学报, 2010, 45(4): 524-529.
Li Y, Sun CH, Lou H M, et al. Transcriptome characterization for Salvia miltiorrhiza using 454 GS FLX. Acta Pharmaceutica Sinica, 2010, 45(4): 524-529.

[23] Conesa A, Gotz S, Garcia-Gomez J M, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 2008, 21(18): 3674-3676.

[24] Ye J, Fang L, Zheng H K, et al. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res, 2006, 34: W293-W297.

[25] 仇燕,王刚. 紫杉醇合成代谢途径中苯丙基转移酶cDNA的克隆. 应用与环境生物学报, 2004, 10(1): 043-045.
Qiu Y, Wang G. CLoning of phenylpropanoyltransferase functioning in taxol biosynthesis. Chin J Appl Environ Biol, 2004, 10(1): 043-045.

[26] 阮仁余,孔建强,郑晓东,等. 中国红豆杉细胞色素P450还原酶的基因克隆、表达与活性分析. 遗传, 2010, 32(11): 1187-1194.
Ruan R Y, Kong J Q, Zheng X D, et al. cDNA cloning, heterologous overexpression and activity analysis of cytochrome P450 reductase of Taxus Chinensis. Hereditas (Beijing), 2010, 32(11): 1187-1194.

[27] 王宾会. 紫杉烷类内生真菌的分离及紫杉醇合成相关基因的克隆. 上海: 上海交通大学, 2006.
Wang B H, Isolation of taxane-producing endophytic fungi and molecular cloning of genes involved in taxol biosynthesis. Shanghai: Shanghai Communication Universty, 2006.

[28] 王伟. 一. 中国红豆杉紫杉醇生物合成基因的研究; 二. 银杏二萜环化酶克隆的初步研究. 北京: 中国协和医科大学, 2002.
Wang W. Part 1 Studies on the taxol biosynthetic genes in Taxus chinesis, Part 2. Preliminary studies on the cloning of diterpene cylase enzyme in Ginkgo biloba. Beijing: Peking Union Medical College, 2002.

[29] 肖颖,赵冬,王刚. 紫杉醇合成途径中紫杉烯合成酶cDNA的克隆. 中国农业科学, 2006, 39(10): 2138-2146.
Xiao Y, Zhao D, Wang G. cDNA cloning of taxadiene synthase functioning in taxol biosynthesis. Scientia Agriculra Sinica, 2006, 39(10): 2138-2146.

[30] 肖颖,赵冬,王刚. 紫杉醇生物合成途径及其相关酶的研究进展. 河北师范大学学报, 2006, 30(2): 222-228.
Xiao Y, Zhao D, Wang G. Progress in studies of taxol biosynthesis and taxol biosynthetic enzymes. Journal of Hebei Normal Universit y (Natural Science Edition), 2006, 30(2): 222-228.

[31] Schoendorf A, Rithner C D, Williams R M, et al. Molecular cloning of a cytochrome P450 taxane 10b-hydroxylase cDNA from Taxus and functional expression in yeast. PNAS, 2000, 98: 1501-1506.

[32] Jennewein S, Rithner C D, Williams R M, et al. Taxol biosynthesis: Taxane 13alpha-hydroxylase is a cytochrome P450-dependent monooxygenase. PNAS, 2001, 98: 13595-13600.

[33] Chau M D, Croteau R B. Molecular cloning and characterization of a cytochrome P450 taxoid 2a-hydroxylase involved in Taxol biosynthesis. ABB, 2004, 427: 48-57.

[34] Jennewein S, Long R M, Williams R M, et al. Cytochrome P450 taxadiene 5a-Hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of Taxol biosynthesis. Chemistry & Biology, 2004, 11: 379-387.

[35] Long R M, Croteau R. Preliminary assessment of the C13-side chain 2′-hydroxylase involved in Taxol biosynthesis. BBRC, 2005, 338: 410-417.

[36] Lia H, Horiguchi T, Croteauc R, et al. Studies on taxol biosynthesis: preparation of taxadiene-diol and triol-derivatives by deoxygenation of taxusin. Tetrahedron, 2008, 64(27): 6561-6567.

[37] Funk C, Croteau R. Induction and characterization of a cytochrome P450-Dependent camphor hydroxylase in tissue cultures of common sage (Salvia officinalis). Plant Physiol, 1993, 101: 1231-1237.

[38] Cahoon E B, Ripp K G, Hall S E, et al. Transgenic production of

epoxy fatty acids by expression of a cytochrome P450 enzyme from Euphorbia lagascae seed. Plant Physiology,2002,128:615 - 624.

[39] Kaspera R,Croteau R. Cytochrome P450 oxygenases of taxol biosynthesis. Phytochem,2006,5:433-444.

[40] Chau M D,Jennewein S,Walker K,et al. Taxol biosynthesis: molecular cloning and characterization of a cytochrome P450 taxoid 7β-Hydroxylase. Chemistry & Biology,2004,11:663-672.

[41] Hefner J,Rubenstein S M,Ketchum R E B,et al. Cytochrome P450-catalyzed hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5a-ol the first oxygenation step in taxol biosynthesis. Chemistry & Biology,1996,3: 479-489.

[42] 张猛,尹大力,刘红岩,等. 紫杉烷类多药耐药逆转剂的合成及其逆转耐药活性. 药学学报,2003,38(6):424-429.
Zhang M,Yin D L,Liu H Y et al. Synthesis and drug resistant reversal activities of taxane-like multi-drug resistant reversal agents. Acta Pharmaceutica Sinica,2003,38(6):424-429.

Sequencing and Analysis of the Transcriptome of *Ginkgo biloba* L. Cells

ZHANG Nan[1] SUN Gui-ling[2] DAI Jun-gui[3] YANG Yan-fang[1] LIU Hong-wei[1] QIU De-you[1]

(1 State Key Laboratory of Tree Genetics and Breeding, The Research Institute of Forestry ,Chinese Academy of Forestry, Beijing 100091,China)

(2 Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China)

(3 State Key Laboratory of Bioactive Substances and Functions of Natural Medicines,Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College,Beijing 100050,China)

Abstract To sequence the transcriptome of *Ginkgo biloba* L. cells, Illumina Genome Analyzer IIx was used. One purpose is to discover candidate genes involved in ginkgolide biosynthesis and new hydroxylases in Ginkgo biloba cells such as taxoid 9-alpha hydroxylase, which will complete the unknown hydroxylation steps in taxol biosynthesis pathway in Taxus sp. A total of 69 286 contigs, 56 387 scaffolds and 32 032 unigenes with average length of 636bp were generated. Unigene qualities from several aspects like gap distribution, GC content, gene coverage were assessed. The results indicate that the sequencing data is good with high quality and reliability. Analyzed the information of unigene expression and functional annotation, we found that 66 unigenes belong to CYP450 gene family, 726 relate to secondary metabolism among which 59 involved in terpenoid metabolism and 17 involved in diterpenoid biosynthesis. At last, 15 hydroxylase candidates were selected by bioinformatics analysis our transcriptome data of *Ginkgo biloba* L. cells and the Michigan State University transcriptome data of *Ginkgo biloba* L. tissues. These candidate genes selection work set foundation for the further research.

Key words *Ginkgo biloba* cells Genome Analyzer IIx Transcriptome