

山核桃种子脂肪代谢期 EST 序列的初步分析

黄银芝^{1,2}, 曾燕如¹, 周 秦³, 夏国华¹, 黄有军¹

(1. 浙江农林大学 亚热带森林培育国家重点实验室培育基地, 浙江 临安 311300; 2. 浙江省黄岩区头陀镇人民政府 农业办公室, 浙江 黄岩 318026; 3. 浙江省金华市农业科学研究院, 浙江 金华 321017)

摘要: 山核桃 *Carya cathayensis* 为重要的油料干果树种, 其非同一般的高含油率必定有其特殊的成油机制。为探索其成油机制, 在已构建的山核桃脂肪代谢相关 cDNA 文库基础上, 对 cDNA 进行随机克隆测序, 对序列进行拼接、功能注释, 基因本体论 (gene ontology, GO) 分类及基于拟南芥 *Arabidopsis thaliana* 的 FunCat 分类, 并对与脂肪代谢相关的全长 cDNA 序列进行蛋白质性质分析。结果共得到 1 010 条山核桃表达序列标签 (ESTs), 经拼接获得 188 个编码蛋白质的基因 (unigene), 其中 92 个 contigs 和 96 个 singlets。GO 分类将这些 unigene 分成细胞组分、分子功能和生物过程等 3 类, 并发现各序列编码的基因并不承担单一的功能, 有些序列在功能上有重叠; 同时参照拟南芥的基因功能分类, 将 unigene 序列在功能上分成 9 类, 得到了 143 个全长 cDNA 序列, 其中包括 14 个与脂肪代谢相关的 cDNA 序列。对 12 个与脂肪代谢相关的全长 cDNA 序列进行蛋白质性质分析, 发现它们均具有磷酸化位点; 除油质蛋白外都没有跨膜区域, 且不具信号肽结构, 为非分泌蛋白, 属于某个蛋白家族; 不同的油体蛋白具有不同的跨膜区, 且均具有信号肽结构, 属于 Oleosin super family 蛋白家族。表 4 参 35

关键词: 经济林学; 山核桃; 表达序列标签; 功能分类; 脂肪代谢; 蛋白分析

中图分类号: S722.3; Q753 **文献标志码:** A **文章编号:** 2095-0756(2015)02-0229-08

Analysis of ESTs associated with oil metabolism in seeds of *Carya cathayensis*

HUANG Yinzhi^{1,2}, ZENG Yanru¹, ZHOU Qin³, XIA Guohua¹, HUANG Youjun¹

(1. The Nurturing Station for the State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University, Lin'an 311300, Zhejiang, China; 2. Agricultural Office of Toutuo People's Government, Huangyan District, Huangyan 318026, Zhejiang, China; 3. Jinhua Institute of Agricultural Sciences, Jinhua 321017, Zhejiang, China)

Abstract: *Carya cathayensis* is an important oil species with dry nuts that are high oil content, for which there must be some special oil-synthesizing mechanism. In order to explore the mechanism, random cloning and sequencing of cDNAs from a fatty acid metabolism-associated cDNA library, sequence assembly and annotation, Gene Ontology (GO)- and *Arabidopsis thaliana*-associated MIPS Functional Catalogue (FunCat)-based classification of assembled sequences, and protein analysis of fatty acid metabolism-associated cDNA sequences were performed. A total of 1 010 expressed sequence tags (ESTs) was obtained, based on which 188 unigenes were assembled. There were 92 contigs and 96 singlets among the unigenes, which were classified into categories of cellular component, molecular function, and biological process based on classification of GO and found to overlap one another in functions. But against MIPS Functional Catalogue in *Arabidopsis thaliana*, these unigenes were classified into nine categories and 143 full-length cDNAs were obtained, from which 14 cDNA sequences were associated with metabolism of fatty acids. It had been found through proteomic analysis that 12 full-length

收稿日期: 2014-06-12; 修回日期: 2014-08-24

基金项目: 浙江省自然科学基金重点项目(Z13C160012); 浙江省科学技术创新团队项目(2011R50030); 国家高技术研究发展计划('863'计划)项目(2013AA102605)

作者简介: 黄银芝, 从事经济林培育与利用研究。E-mail: Huangyinzhi@126.com。通信作者: 曾燕如, 教授, 博士, 从事经济林培育与利用研究。E-mail: yzeng@zafu.edu.cn

fatty acid metabolism-related cDNAs had a phosphorylation site; they were all non-secretory proteins with no transmembrane domain and signal peptide except oleosins, belonging to certain protein family; and oleosins had different transmembrane domains and a signal peptide, belonging to a super oleosin family. [Ch, 4 tab. 35 ref.]

Key words: cash forestry; *Carya cathayensis*; expressed sequence tag; functional classification; fatty acid metabolism; protein analysis

山核桃 *Carya cathayensis* 是胡桃科 Juglandaceae 山核桃属 *Carya* 的木本油料树种, 其种仁的含油率平均达到 70%, 远远超过油菜新品系“超油 2 号” *Brassica napus* ‘Superoil No.2’ (52.8%)^[1], 为高含油量的树种, 是中国特有的名优干果。油料树种的成油途径及机制虽有共性, 但不同的物种又有各自的特异性。山核桃非同一般的高含油率必定有其特殊的成油机制。表达序列标签 (expressed sequence tags, EST) 是从已构建的 cDNA 文库中随机挑取克隆, 对 cDNA 进行测序得到的序列^[2-4]。由于 cDNA 由 mRNA 反转录而来, mRNA 是基因表达的产物, 因而 EST 代表生物体某种组织某一时期的基因表达, 也即具有时空性的特点。EST 分析有助于认识生物体生长发育、繁殖分化、遗传变异、衰老死亡等一系列生理生化过程^[5], 目前主要用于发现新基因、了解基因表达概况等, 是获得未知基因、发现新基因及开发表达序列标签-简单重复序列 (EST-SSR) 分子标记相对快速、简便的方法。在利用 EST 进行成油机制研究方面, 对薄荷 *Mentha × piperita* 腺毛的 EST 进行了功能分析, 以此来研究薄荷油的生物合成与分泌^[6]; 对 4 个时间点正在发育的蓖麻 *Richinus communis*, 欧洲油菜 *Brassica napus*, 卫矛 *Euonymus alatus*, 旱金莲 *Tropaeolum majus* 种子 EST 进行深度测序, 并对获得的序列进行比较分析发现, 4 物种在油脂储存的组织、光合能力及三酰甘油酯 (triacylglycerols, TAGs) 的结构与含量上彼此有差异, 但编码脂肪酸合成核心酶的基因 EST 是十分保守的, 在种子发育过程中与发育时间紧密相关, 参与甘油酯及其前体合成的基因表达具有明显的物种特异性^[7]; 对拟南芥 *Arabidopsis thaliana* 碳水化合物到种子油脂代谢途径的 EST 研究发现, 许多基因是在种子中特异表达的^[8]; 研究人员早就注意到, 催化脂肪酸代谢类似反应的酶, 其基因 EST 丰度水平在统计学上存在显著的差异^[9]; 对高油油棕 *Elaeis guineensis* 及低油耶枣 *Phoenix dactylifera* 的对比研究发现, 784 个转录因子中仅有 6 个转录因子 EST 水平 2 物种相差 15 倍, 其中 WRH1-like 的 EST 水平油棕比耶枣要高 57 倍, 它控制油脂合成与参与种子发育的上游因子无关, 参与了不同的调控网络, 该网络可能是棕榈果实所特有的^[10]。近年来, 随着实验技术的发展, 仪器设备的更新换代, 研究人员对数种油料作物开展了成油相关的转录组学研究, 如油茶 *Camellia oleifera*^[11], 油棕^[10]和麻风树 *Jatropha curcas*^[12]等, 研究结果相对于 cDNA 文库的测序结果序列短, 数据量大, 多集中在基因差异表达及 KEGG (kyoto encyclopedia of genes and genomes) 途径等。山核桃无类似的研究报道, 仅报道了山核桃成油相关 cDNA 文库的构建^[13]。本研究是在构建山核桃种子脂肪代谢相关 cDNA 文库的基础上, 对 cDNA 测序所得的 EST 序列进行分析的结果, 以期为后续深入研究山核桃成油机制提供线索。

1 材料和方法

1.1 材料

基于周秦等^[14]发现的山核桃油脂形成、转化、积累的关键时期, 黄银芝等^[13]构建了山核桃脂肪代谢相关的 cDNA 文库。该文库随机测序所得 EST 序列即为本文的原材料。

1.2 方法

1.2.1 cDNA 的克隆测序 参照 SmartTM cDNA library construction kit user manual (Clontech 公司) 的描述, 制作测序平板, 并随机挑取 1 个 λ 噬菌体克隆, 在重组噬菌体中经过活体剥离及完整质粒的环化, 将 λ 噬菌体 λ TriplEx2 克隆转化成质粒载体 pTriplEx2 的克隆, 并转化大肠埃希菌 *Escherichia coli*。转化子培养后, 分别用 pTriplEx2 测序引物 (5' sequencing primer TCCGAGATCTGGACGAGC/3' sequencing primer TAATACGACTCACTATAGGG) 直接对菌液进行 PCR^[10], 检测 cDNA 片段的大小, 并进行菌液保存。将经过 cDNA 片段大小检测过的菌液, 送上海生物工程有限公司, 用 T7 (5'-TAATACGACTCAC-TATAGGG-3') 和 pTriplEx2 测序引物 (5'-CTCCGAGATCTGGACGAGC-3') 进行测序。

1.2.2 数据处理与拼接 利用美国国家生物技术信息中心 (NCBI) 的在线软件 VecScreen (<http://www.ncbi>).

nlm.nih.gov/VecScreen/VecScreen.html)去除测序结果中的载体序列, 利用 DNASTAR 软件中的 SeqMan 模块去除冗余序列及小于 150 bp 的 EST 序列, 余下的高质量序列用 SeqMan 模块中的 Progress 方法进行聚类拼接, 生成编码蛋白质基因(unigenes), 包括 contigs 和 singlets。

1.2.3 功能注释 对得到的 unigenes, 先后在 NCBI 的非冗余蛋白数据库进行 Blastx 比对(序列对齐值大于 80 且序列同一性大于 35%)^[15-16], 在非冗余核酸数据库进行 Blastn 比对(序列对齐值大于 200 且序列同一性大于 75%)^[17]。同时, 用基因本体论(gene ontology, GO)分类, 并参照 Bevan^[18]对拟南芥基因组研究的方法, 对获得的功能已知的 cDNA 片段进行分类。

1.2.4 全长 cDNA 序列的获得 利用 Blastx 与其他生物已知的对应基因进行比对, 将包含经 ORF finder (open reading frame finder)(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>)预测且含有 ORF(开放阅读框)的 cDNA 视为全长 cDNA。

1.2.5 序列提交 GenBank 数据库 将序列转化为通用的 FASTA 格式, 再利用 Sequin 软件(<http://www.ncbi.nlm.nih.gov/Seqin/index.html>)将序列转化为 GenBank 提交格式, 并提交 GenBank。

1.2.6 山核桃脂肪代谢相关全长 cDNA 序列的蛋白分析 对获得的参与脂肪代谢的全长 cDNA 序列进行蛋白分析。利用 BioEdit 软件和瑞士生物信息学研究所(Swiss Institute of Bioinformatics)ExpASy 服务器(<http://www.expasy.org/>)上的 ProtParam 程序分析蛋白的分子量、等电点、氨基酸组成等。利用 ExpASy 服务器上的 ProtScale 程序对蛋白疏水性进行分析。利用丹麦科技大学(DTU)CBS 服务器(<http://www.cbs.dtu.dk/services>)上的 NetPhos 2.0 Server 程序、TMHMM Server 2.0 程序和 SignalP 2.0 程序分别对蛋白的磷酸化位点、跨膜区和信号肽位点进行分析。利用日本国家基础生物学研究所的 PSORT II 程序(<http://psort.nibb.ac.jp/form2.html>)对蛋白做亚细胞定位分析。用美国国家生物技术信息中心(NCBI)的 Conserved Domain Database(CDD)程序(<http://www.ncbi.nlm.nih.gov/cdd/>)预测该蛋白的保守结构域。

2 结果与分析

2.1 非冗余 EST 序列的获得

随机测序得到 1 010 条山核桃 EST, 去除低质量的序列和载体污染序列, 经 DNASTAR 软件拼接后获得 188 个编码蛋白质基因(unigenes), 其中 92 个 contigs 和 96 个 singlets, 序列总长度为 201.131 kb, 平均长度为 1 069.84 bp。

2.2 功能注释

将 unigenes 进行 Blastx 分析和 ORF 预测, 确认序列包含完整的 CDS 区, 获得全长 cDNA 序列。经分析, 188 条单一序列中共 143 个全长 cDNA 序列。将 unigenes 序列与 NCBI 的非冗余蛋白质数据库及非冗余核酸数据库比对, 提取有高度同源性的注释信息, 获得功能注释的序列 175 条。对这些有功能注释的序列再进行功能分类。

2.2.1 GO(gene ontology)分类 基因本体论 GO 从基因参与的生物过程、所处的细胞位置、发挥的分子功能等 3 方面描述基因及其产物的功能。175 条序列经分析分别归入细胞组分、分子功能和生物过程 3 类(表 1), 且发现各序列编码的基因并不承担单一的功能, 有些序列在功能上有重叠。细胞组分、分子功能和生物过程 3 类又可以分成不同的亚类, 发挥不同的功能(表 1)。

2.2.2 基于拟南芥基因功能的分类(MIPS functional catalogue) 根据所参与的细胞过程将 cDNA 序列分为 8 类(表 2), 分别是能量代谢(energy), 细胞结构(cell structure), 转录(transcription), 疾病防御(disease defense), 蛋白质合成(protein synthesis), 转运(transporters), 信号转导(signal transduction), 新陈代谢及次生代谢(metabolism and secondary metabolism)。从表 2 中可以看出, 与疾病防御相关的 cDNA 数量最多, 共有 44 条, 占功能注释 cDNA 序列的 25.1%, 其中尤以热激蛋白为多。在此类 cDNA 序列(表 3)中, 主要包括温度诱导脂蛋白(temperature-induced lipocalin, TIL), 过氧化物酶(peroxidase)、铜/锌过氧化物歧化酶(copper/zinc superoxide dismutase)和热激蛋白(heat-shock protein)等非生物胁迫相关蛋白, 也包括一些病程相关蛋白。细胞色素(cytochrome)是生物体内的一类重要的多功能血红素氧化还原酶类, 在生物防御外界不良环境影响方面起重要作用。另外还有谷胱甘肽转移酶(glutathione transferase), 金属硫蛋白(metallothionein protein)和泛素蛋白(ubiquitin-protein)等, 它们在植物抗病防御中也

表1 山核桃脂肪代谢期部分 cDNA 序列的基因本体论分类结果

Table 1 GO-based classification of unigenes from sequenced cDNAs of an oil metabolism-related library in *Carya cathayensis*

注释基因分类	unigene/个	亚类	百分比/%
细胞组分	98	细胞器部件(organelle part)	31.6
		细胞部件(cell part)	23.5
		细胞器(organelle)	15.3
		细胞(cell)	11.2
		大分子复合体(macromolecular complex)	10.2
		胞外区(extracellular region)	5.1
		无归类(unclassified)	3.1
分子功能	93	具有催化活性(catalytic activity)	46.2
		结合(binding)功能	25.8
		转运活性(transporter activity)	8.6
		结构分子活性(structural molecule activity)	4.3
		营养库活性(nutrient reservoir activity)	3.2
		抗氧化调节剂活性(antioxidant regulator activity)	2.1
		翻译调节活性(translation regulator activity)	1.1
		结合核酸转录因子活性(nucleic acid binding transcription factor activity)	1.1
		酶调节活性(enzyme regulator activity)	1.1
		无归类(unclassified)	6.5
生物过程	97	细胞过程(cellular process)	51.5
		刺激反应(response to stimulus)	15.5
		生物调节(biological regulation)	10.4
		代谢过程(metabolic process)	8.2
		细胞组分组织或生物发生(cellular component organization or biogenesis)	5.2
		多种生物过程(multi-organism process)	2.1
		发育过程(developmental process)	2.1
		生长(growth)	1.0
		生物黏附(biological adhesion)	1.0
		定位建立(establishment of localization)	1.0
		定位(localization)	1.0
运输(transport)	1.0		
合计	288		

起着重要作用。在与能量代谢蛋白相关的 cDNA 序列中发现了 14 个与脂肪酸代谢相关的基因序列(表 4), 主要有硬脂酰基 ACP 去饱和酶(stearoyl-ACP desaturase), 长链脂肪酸 CoA 连接酶(long-chain-fatty-acid CoA ligase), 乙酰-CoA 羧化酶羧基转移酶 β 亚基蛋白(acetyl-CoA carboxylase carboxyltransferase beta subunit protein), 苹果酸酯合成酶(malate synthase), 苹果酸酯脱氢酶(malate dehydrogenase)和 2-磷酸甘油酸酯脱水酶(2-phosphoglycerate dehydratase)等, 其中 12 个为全长 cDNA 序列; 另外还发现 5 条油体蛋白

表2 基于拟南芥基因功能的 cDNA 序列分类

Table 2 Classification of cDNA sequences based on gene functions in *Arabidopsis thaliana*

功能分类	所得 cDNA 数目	占功能注释 cDNA 百分比/%
能量代谢(energy)	25	14.3
细胞结构(cell structure)	9	5.1
转录(transcription)	2	1.1
疾病防御(disease defense)	44	25.1
蛋白质合成(protein synthesis)	31	17.7
转运(transporters)	5	2.9
信号转导(signal transduction)	12	6.9
新陈代谢及次生代谢(metabolism and secondary metabolism)	18	10.3
未知功能(unknown)	29	16.6
合计	175	100

表 3 与疾病防御相关的 cDNA 序列 (部分)

Table 3 Disease defense-related cDNA sequences in the nut of *Carya cathayensis* (part)

假定基因	来源	序列对齐值/bits	匹配率/%
peroxidase	<i>Canellia oleifera</i>	564	93
	<i>Cucumis melo</i>	560	94
ubiquitin-specific protease 23	<i>Arabidopsis lyrata</i>	106	42
ubiquitin-conjugating enzyme E2	<i>Zea mays</i>	317	95
E3 ubiquitin-protein ligase	<i>Arabidopsis thaliana</i>	215	67
ubiquitin carboxyl-terminal hydrolase isozyme L5	<i>Ricinus communis</i>	551	86
polyubiquitin	<i>Vitis vinifera</i>	873	99
cytochrome f	<i>Morus indica</i>	298	98
copper/zinc superoxide dismutase	<i>Melastoma malabathricum</i>	262	85
cysteine proteinase	<i>Glycine max</i>	540	70
proteasome subunit beta type	<i>Ricinus communis</i>	374	91
DHAR class glutathione transferase	<i>Populus trichocarpa</i>	357	80
glutaredoxin	<i>Populus trichocarpa</i>	164	61
lactoylglutathione lyase	<i>Ricinus communis</i>	313	78
pectin methylesterase	<i>Citrus bergamia</i>	380	87
chitinase	<i>Ricinus communis</i>	395	85
temperature-induced lipocalin	<i>Populus balsamifera</i>	337	84
EC protein III	<i>Ricinus communis</i>	134	66
metallothionein-like protein class II	<i>Fagus sylvatica</i>	91	90
metallothionein-like protein	<i>Quercus suber</i>	93	87
heat-shock protein	<i>Ricinus communis</i>	≥151	≥60
	<i>Glycine max</i>	229	69
heat shock protein 90	<i>Vitis pseudoreticulata</i>	402	95
heat-shock protein 70	<i>Hevea brasiliensis</i>	333	93
cytosolic class I small heat shock protein type 1	<i>Rhododendron simsii</i>	119	79
cytosolic class I small heat-shock protein HSP17.5	<i>Castanea sativa</i>	225	77
cytosolic class II small heat-shock protein HSP17.5	<i>Vitis vinifera</i>	190	67
chloroplast chaperonin 21	<i>Vitis vinifera</i>	206	84

(oleosin), 油体蛋白主要在植物种子中特异表达, 覆盖于油体表面, 在油体发生到分解消失过程中及对维持油体的稳定起着重要的生物学作用^[19-20]。

2.3 GenBank 数据库 EST 序列的递交

将有功能注释的 cDNA 序列递交 GenBank 数据库, 最终被接受的序列为 175 条, 登录号为 JN786116-JN786290, 占全部 unigenes 的 93.1%。

2.4 山核桃 cDNA 序列的蛋白分析

利用生物信息学的方法获得了 12 个与脂肪代谢相关的全长 cDNA 序列, 它们分别编码硬脂酰基载体去饱和酶蛋白(JN786139), 依赖 NADP 的甘油醛-3-磷酸脱氢酶(JN786163), β -酮-ACP 合成酶 II (JN786197), 长链脂肪酸 CoA 连接酶(JN786201), 线粒体丙酮酸脱氢酶激酶亚型 1(JN786205), 苹果酸合成酶(JN786209), 乙酰辅酶 A 羧化酶 β 亚基蛋白羧基(JN786226), 磷脂酰乙醇胺结合蛋白(JN786231), 脂质结合域(JN786242), 海藻糖-6-磷酸合成酶(JN786284), 油体蛋白 1(JN786177), 油体蛋白 2(JN786276)。将这些序列翻译成氨基酸序列, 并进行蛋白基本结构分析和保守功能域预测。结果表明, 这些蛋白均具有磷酸化位点, 可能具有磷酸化和自我磷酸化活性; 除油质蛋白外, 其他都没有跨膜区域, 且不具信号肽结构, 为非分泌蛋白, 均属于某个蛋白家族。油体蛋白中, 油体蛋白-1 有 3 个跨膜区, 油体蛋白-2 有 2 个跨膜区, 且均具有信号肽结构, 属于 Oleosin super family 蛋白家族。

表4 山核桃中与脂肪代谢相关的 cDNA 序列

Table 4 Fatty acid metabolism-related cDNA sequences in the nut of *Carya cathayensis*

假定基因	来源	序列对齐值/bits	匹配率/%
stearoyl-ACP desaturase	<i>Jatropha curcas</i>	725	89
nADP-dependent glycerol dehyde-3-phosphate dehydrogenase	<i>Pisum sativum</i>	920	91
beta-ketoacyl-ACP synthase II	<i>Jatropha curcas</i>	490	90
long-chain-fatty-acid CoA ligase	<i>Ricinus communis</i>	828	78
mitochondrial pyruvate dehydrogenase kinase isoform 1	<i>Glycine max</i>	464	90
malate synthase	<i>Ricinus communis</i>	1 012	86
acetyl-CoA carboxylase carboxyltransferase beta subunit protein	<i>Euonymus americanus</i>	520	94
phosphatidylethanolamine binding protein	<i>Ricinus communis</i>	275	76
patellin-5	<i>Ricinus communis</i>	207	68
trehalose-6-phosphate synthase	<i>Ricinus communis</i>	922	88
oleosin	<i>Corylus avellana</i>	139	80
2-phosphoglycerate dehydratase 1	<i>Hevea brasiliensis</i>	273	95
transaldolase	<i>Dimocarpus longan</i>	636	82
malate dehydrogenase	<i>Citrullus lanatus subsp.</i>	411	78

3 结论与讨论

参照基因本体论(gene ontology, GO)和拟南芥基因功能分类等生物信息学分析,有利于对蛋白质功能、代谢通路、相互作用等进行全面、快速、准确地分析。尽管近年来流行使用 RNA-seq 来分析基因的转录表达,虽然能分析相关的代谢途径及基因表达水平的差异等,但基于方法本身所得的序列都较短,而基于 cDNA 文库的测序,因方法上没有 RNA-seq 过程将 RNA 片段化的步骤,因而所得的 EST 序列往往更长,能够反映特定时空基因的表达情况,且较容易获得全长的 cDNA 序列。

从基于基因本体论分类的分子功能注释结果来看,建库时期的种子处于新陈代谢活跃的阶段,以参与细胞器部件、细胞部件、细胞过程的 cDNA 为多,且发现具有催化活性的序列所占比例较高,有酶或酶的结构域存在,说明各种酶活动活跃。事实上,文库构建所用样品的时间为 8 月 11-23 日^[13]。8 月 11 日前,山核桃种子通过细胞分裂,体积逐渐增大至正常大小;8 月 11 日后为种子内部物质转化期^[21-22],包括脂肪链的增长,油脂合成,饱和脂肪酸向不饱和脂肪酸转化等一系列新陈代谢过程。从基于拟南芥基因功能分类来看,细胞抗性与防御部分所占比例最大,其中尤以热激蛋白为多。热激蛋白又名胁迫蛋白^[20],高度保守^[23-24],存在于所有生物及所有生物的细胞中,胞外的热激蛋白可以刺激免疫系统中专门存在抗原的细胞,其功能之一是保护细胞防止其受到胁迫/凋亡^[23];主要的热激蛋白属于几个基因(蛋白)家族^[24]。热激的主要影响之一是使蛋白不能折叠,或不完全折叠,或不合适地折叠^[25],过多地发生依赖于胁迫强度及细胞体系的变化^[24]。山核桃果实成熟阶段主要在 8 月,在当地正是高温的炎炎夏日。热激蛋白的出现有利于种子抵抗高温胁迫而进行正常的发育。Hsp70 促进蛋白质的折叠过程,在环境胁迫的条件下,蛋白折叠对它的依赖性更强^[27]。Hsp90 对参与信号转导的近百个蛋白起调节作用,且与 Hsp70 等形成基于 Hsp70/Hsp90 的多蛋白陪伴机制,并通过该机制将参与信号转导的蛋白与 Hsp90 组合成复合体^[27]。过氧化物酶在植物抵抗病原菌感染的过程中起着防卫的作用^[28],同时诸如山葵过氧化物酶(horseradish peroxidase)可参与酚类氧化过程,使之形成多聚体及寡聚体,以减少酚类物质在植物体内的毒性。山核桃富含单宁,植物单宁(即植物多酚)为植物体内的复杂酚类次生代谢产物。温度诱导脂蛋白(TILs)是植物载脂蛋白中的一类,与非生物胁迫反应相关,在温度胁迫中起保护光合系统作用^[29],是叶黄素循环中的关键酶,在光氧化损害防卫中发挥功能^[30],同时作为耐热的必要组份,在严重热激诱导的脂类过氧化过程中起着防卫作用^[31]。铜/锌过氧化物歧化酶则催化过氧化氢(H₂O₂)产生游离的羟基^[32],而过氧化氢的累积是植物在逆境或衰老时体内活性氧代谢加强导致的。这也是为什么山核桃脂肪代谢期防御性 EST 数量多的原因,这和其果实生物发育期所处的高温干旱环境胁迫有关。

研究发现 13 条与山核桃脂肪代谢相关的序列,其中包括乙酰-CoA 羧化酶羧基转移酶 β 亚基蛋白。

乙酰-CoA 羧化酶是脂肪酸生物合成过程中的关键酶之一, 它在生物体内催化形成丙二酰 CoA (Malonyl-CoA), 而后者是脂肪酸生物合成和脂酰链延伸系统等重要代谢反应的底物, 被认为是生物体内一个基本的代谢底物和特定蛋白活性的调控代谢物^[33]。乙酰-CoA 羧化酶活性的过量产生可以增加大肠杆菌中脂肪酸生物合成的速率^[34]。这说明研究中构建的山核桃脂肪代谢相关 cDNA 文库及本文的研究结果对山核桃油脂脂肪代谢的研究是有帮助的。此外, 我们还构建了山核桃脂肪快速增长期(7 月 15 日至 8 月 11 日)的 cDNA 文库^[35], 可以补充本文库信息的不足。但脂肪快速增长期伴随着果实的生长, 也即果实逐渐增大, 内部胚由小变大, 液态胚乳逐渐增多, 并在胚的发育过程中为胚的生长提供营养, 而胚乳本身则逐渐为生长的胚所吸收而消失, 整个种壳内逐渐为胚所占据。因此, 此期大部分营养物质还是供应果实与种子本身的生长之用。基因表达产物蛋白质分子的各种生物功能与其复杂的结构密切相关。本研究利用生物信息学方法对脂肪代谢相关全长 cDNA 序列进行了详细的蛋白分析, 蛋白一些性质(如跨膜区结构域、磷酸化位点、二级保守结构域等)的预测会给我们了解该蛋白质的结构域或功能提供一些线索, 同时也可后续蛋白结构与功能、蛋白与核酸的互作、蛋白间的互作分析打下基础。

4 参考文献

- [1] 黄坚钦, 郑炳松, 黄有军, 等. 木本油料植物山核桃新品种的选育[J]. 生物质化学工程, 2006(B12): 178 - 181.
HUANG Jianqin, ZHENG Bingsong, HUANG Youjun, *et al.* Selection and breeding of new varieties of the oil-bearing tree hickory (*Carya cathayensis* Sarg.) [J]. *Biomass Chem Eng*, 2006(B12): 178 - 181.
- [2] HATEY F, TOSSER K G, CLOUSCARD-MARTINATO C, *et al.* Expressed sequenced tags for genes: a review [J]. *Gen Selection Evol*, 1998, **30**(5): 521 - 541.
- [3] 骆蒙, 贾继增. 植物基因组表达序列标签(EST)计划研究进展[J]. 生物化学与生物物理进展, 2001, **28**(4): 494 - 497.
LUO Meng, JIA Jizeng. Progress in expressed sequence tags (EST) project of plant genome [J]. *Prog Biochem Biophysics*, 2001, **28**(4): 494 - 497.
- [4] 陈全求, 詹先进, 蓝家祥, 等. EST 分子标记在基因组学中应用的研究进展[J]. 中国农学通报, 2010, **16**(3): 59 - 63.
CHEN Quanqiu, ZHAN Xianjin, LAN Jiexiang, *et al.* Study progress on application of EST (expressed sequence tags) in the functional genomics [J]. *Chin Agric Sci Bull*, 2010, **16**(3): 59 - 63.
- [5] HATEY F, TOSSER-KLOPP G, CLOUSEARD-MARTINATO C, *et al.* Expressed sequenced tags for genes: a review [J]. *Genet Sel Evol*, 1998, **30**(6): 521 - 541.
- [6] LANGE B M, WILDUNG M R, STAUBER E J, *et al.* Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes [J]. *PNAS*, 2000, **97**(6): 2934 - 2939.
- [7] TRONCOSO-PONCE M A, KILARU A, CAO Xia, *et al.* Comparative deep transcriptional profiling of four developing oil seeds [J]. *Plant J* 2011, **68**(6): 1014 - 1027.
- [8] WHITE J A, TODD J, NEWMAN T, *et al.* A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil [J]. *Plant Physiol*, 2000, **124**(4): 1582 - 1594.
- [9] MEKHEDOV S, de ILÁRDUYA O M, OHLROGGE J. Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis [J]. *Plant Physiol*, 2000, **122**(2): 389 - 401.
- [10] BOURGIS F, KILARUC A, CAO Xia, *et al.* Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning [J]. *PNAS*, 2011, **108**(30): 12527 - 12532.
- [11] 陈英, 江香梅, 张露, 等. 基于油茶组 59 万条 EST 序列的转录组学初步分析[J]. 林业科学, 2011, **47**(2): 161 - 163.
CHEN Ying, JIANG Xiangmei, ZHANG Lu, *et al.* Transcriptome characterization for *Camellia* Sect. *Oleifera* based on the 592499 ESTs [J]. *Sci Silv Sin*, 2011, **47**(2): 161 - 163.
- [12] NATARAJAN P, PARANI M. *De novo* assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing [J]. *BMC Genomics*, 2011, **12**(4): 191 - 202.
- [13] 黄银芝, 周秦, 黄有军, 等. 山核桃脂肪代谢相关 cDNA 文库的构建[J]. 浙江农林大学学报, 2011, **28**(1):

80 – 85.

HUANG Yinzi, ZHOU Qin, HUANG Youjun, *et al.* Construction of a fat metabolism related cDNA library in *Carya cathayensis* [J]. *J Zhejiang A & F Univ*, 2011, **28**(1): 80 – 85.

- [14] 周秦. 山核桃果实成熟期间基因表达的初步分析[D]. 临安: 浙江农林大学, 2010.
ZHOU Qin. *Fruit Ripening associated Gene Expression in Carya cathayensis* [D]. Lin'an: Zhejiang A & F University, 2010.
- [15] NEWMAN T, de BRUIJN F J, GREEN P, *et al.* Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones [J]. *Plant Physiol*, 1994, **106**(4): 1241 – 1255.
- [16] ANDERSON P A, LAWRENCE G J, MORRISH B C, *et al.* Inactivation of the flax rust resistance gene M associated with loss of a repeated unit within the leucine rich repeat coding region [J]. *Plant Cell*, 1997, **9**(4): 641 – 651.
- [17] HOFTE H, DESPREZ T, AMSENLEM J. An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNA from *Arabidopsis thaliana* [J]. *Plant J*, 1993, **4**(6): 1051 – 1061.
- [18] BEVAN M, BANEROFT I, BENT E, *et al.* Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana* [J]. *Nature*, 1998, **391**(6666): 453 – 455.
- [19] HUANG A H C. Oleosins and oil bodies in seeds and other organs [J]. *Plant Physiol*, 1996, **110**(4): 1055 – 1061.
- [20] FRANSDEN G I, MUNDY J, TZEN J T C. Oil bodies and their associated proteins, oleosin and caleosin [J]. *Physiol Plantarum*, 2001, **112**(3): 301 – 307.
- [21] 黎章矩, 高林, 王白坡, 等. 浙江省名特优经济树种栽培技术[M]. 北京: 中国林业出版社, 1995.
- [22] 解红恩, 黄有军, 薛霞铭, 等. 山核桃果实生长发育规律[J]. 浙江林学院学报, 2008, **25**(4): 527 – 531.
XIE Hongen, HUANG Youjun, XUE Xianning, *et al.* Growth and development of the *Carya cathayensis* nut[J]. *J Zhejiang For Coll*, **25**(4): 527 – 531.
- [23] LI Z, SRIVASTAVA P. *Current Protocols in Immunology* [M]. New York: John Wiley and Sons, Inc., 2004.
- [24] SCHLESINGE M J. Heat shock proteins [J]. *J Biol Chem*, 1990, **265**(21): 12111 – 12114.
- [25] HIGHTOWER L E. A brief perspective on the heat-shock response and stress proteins [J]. *Marine Environ Res* [J], 1993, **35**(1/2): IN1 – IN2, 79 – 83.
- [26] MAYER M P, BUKAU B. Hsp70 chaperones: cellular functions and molecular mechanism [J]. *Cell Molec Life Sci*, 2005, **62**(6): 670 – 684.
- [27] PRATT W B, TOFT D O. Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery [J]. *Exp Biol Med*, 2003, **228**(2): 111 – 133.
- [28] KARTHIKEYAN M, JAYAKUMAR V, RADHIKA K, *et al.* Induction of resistance in host against the infection of leaf blight pathogen (*Alternaria palandui*) in onion (*Allium cepa* var. *aggregatum*) [J]. *Indian J Biochem Biophys*, 2005, **42**(6): 371 – 377.
- [29] CHARRON J B F, OUELLET F, PELLETIER V, *et al.* Identification, expression, and evolutionary analyses of plant lipocalins [J]. *Plant Physiol*, 2005, **139**(4): 2017 – 2028.
- [30] CHARRON J B F, BRETON G, BADAWI M, *et al.* Molecular and structural analyses of a novel temperature stress-induced lipocalin from wheat and *Arabidopsis* [J]. *Fed Eur Biochem Soc Lett*, 2002, **517**(1/3): 129 – 132.
- [31] CHI W T, FUNG R W M, LIU H C, *et al.* Temperature-induced lipocalin is required for basal and acquired thermotolerance in *Arabidopsis* [J]. *Plant Cell Environ*, 2009, **32**(7): 917 – 927.
- [32] YIM M B, CHOCK B, STADTMAN E R. Copper, zinc superoxide dismutase catalyzes hydroxyl radical production from hydrogen peroxide [J]. *Proc Natl Acad Sci*, 1990, **87**(13): 5006 – 5010.
- [33] 石明旺. 油茶种子 EST 文库构建及油脂合成关键酶基因的分离鉴定[D]. 长沙: 中南林学院, 2004.
SHI Mingwang. *Establishment of ESTs Database from Camellia oleifera A. Seeds and Identification of Key Genes of Oil Synthetase* [D]. Changsha: Central South Forestry University, 2004.
- [34] DAVIS M S, SOLBIATI J, CRONAN J E Jr. Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli* [J]. *J Biol Chem*, 2000, **275**(37): 28593 – 28598.
- [35] 黄银芝. 山核桃脂肪代谢相关 cDNA 文库的构建及序列分析与利用[D]. 临安: 浙江农林大学, 2012.
HUANG Yinzi. *Construction of a Fat Metabolism-related cDNA Library and Analysis and Utilization of Sequences in Carya cathayensis* [D]. Lin'an: Zhejiang A & F University, 2012.