

‘黑油椿’香椿嫩芽高通量转录组测序及萜类代谢物质初步分析

赵 胡*, 唐开静, 范小莹, 吕伟健

(阜阳师范学院生物与食品工程学院, 抗衰老中草药安徽省工程技术研究中心, 安徽阜阳 236037)

摘 要: ‘黑油椿’香椿 (*Toona sinensis* ‘Heiyouchun’) 嫩芽富含萜类物质, 主要以倍半萜为主, 其中去氢香橙烯、9,10-脱氢异长叶烯和 β -石竹烯含量分别高达 4 440.71, 1 932.02 和 1 799.89 ng · g⁻¹。采用高通量 RNA-seq 测序技术 (Illumina HiSeq™4000) 对嫩芽进行转录组测序, 获得高质量的碱基数据量为 4.70 Gb, 共计 86 870 个转录本 (Transcript), 对这些转录本进一步组装, 得到 55 850 个单基因簇 (Unigene), 平均长度为 1 013 bp。序列同源性比对表明, 39 408 个单基因簇与数据库中其他物种基因存在不同程度的同源性, 功能注释匹配度最高的物种为甜橙 (*Citrus sinensis*)。将全部单基因簇进行 GO 数据库比对, 19 704 个单基因簇获得 GO 功能注释, 分为细胞组分、分子功能和生物学过程等 3 大类群和 54 个亚类, 其中与代谢过程、细胞组分、结合和催化活性及细胞过程相关的单基因簇占绝大多数。利用 COG 数据库进行比对, 14 186 个单基因簇分布于 25 个功能区域。KEGG 代谢通路数据库比对, 有 28 400 个单基因簇获得功能注释, 分为 6 大类, 21 个亚类和 135 个代谢途径分支。通过对转录组数据进一步分析发现, 参与 ‘黑油椿’ 嫩芽风味形成相关萜类合成的单基因簇共计 467 个, 其中参与萜类骨架生物合成、单萜生物合成、倍半萜和三萜类合成、二萜合成的单基因簇分别为 226、71、86、84 个, 这为进一步研究香椿萜类物质生物合成相关的基因功能及解析其风味物质形成的分子机制提供了基础数据。

关键词: 香椿; 转录组; 功能注释; 萜类代谢

中图分类号: S 644.4

文献标志码: A

文章编号: 0513-353X (2017) 11-2135-15

High-throughput Transcriptome Sequencing and Primary Analysis of Terpenoids Metabolism in *Toona sinensis* ‘Heiyouchun’ Sprout

ZHAO Hu*, TANG Kaijing, FAN Xiaoying, and LÜ Weijian

(School of Biology and Food Engineering, Fuyang Normal College, Engineering Technology Research Center of Anti-aging Chinese Herbal Medicine, Fuyang, Anhui 236037, China)

Abstract: *Toona sinensis* ‘Heiyouchun’ sprouts were rich in terpenoids that were mainly composed of sesquiterpenoids such as dehydroaromadendrene, 9,10-dehydroisolongifolene, and β -caryophyllene, their contents were as high as 4 440.71, 1 932.02 and 1 799.89 ng · g⁻¹, respectively. To understand terpenoids biosynthetic pathway, high-through RNA-seq technology was used to generate the transcriptome of *Toona sinensis* ‘Heiyouchun’ sprout and high-quality base data of 4.70 Gb and 86 870

收稿日期: 2017-06-22; 修回日期: 2017-10-23

基金项目: 安徽省教育厅自然科学基金重点项目 (KJ2016A552); 大学生创新创业计划项目 (201510371061, 201610371022)

* E-mail: zhaohu8196@sina.com

transcripts were acquired. Furthermore, a total of 55 850 unigenes with average length of 1 013 bp was obtained by de novo assembly. Sequence alignment analysis showed 39 408 unigenes in our transcriptomic data had sequence homology with those of other species at different degrees and the highest matching ratio of functional annotation to *Citrus sinensis*. Gene ontology analysis revealed that annotated 19 704 unigenes were grouped into 54 different categories in terms of cellular component, molecular function and biological process. Among them, the unigenes involved in metabolic process, cell composition, binding and catalytic activity, and cell processes were predominant. Based on the cluster of orthologous groups, 14 186 unigenes were further annotated and grouped into 25 functional categories. Moreover, 28 400 unigenes were annotated to 135 KEGG pathway and broadly divided into 6 categories of 21 branches. Our data indicated that 467 unigenes were mined and involved in terpenoids biosynthesis related to flavor formation of *Toona sinensis* 'Heiyouchun' sprout, including 226 for terpenoid backbone biosynthesis, 71 for monoterpenoids biosynthesis, 86 for sesquiterpenoids and triterpenoids biosynthesis, and 84 for diterpenoids biosynthesis, which laid a solid foundation for further study on the function of genes related to terpenoids biosynthesis and the molecular mechanism of flavor compounds formation.

Keywords: *Toona sinensis*; transcriptome; gene annotation; terpenoids metabolism

香椿 (*Toona sinensis* Roem) 嫩芽, 营养丰富, 香气浓郁, 风味鲜美, 深受人们的喜爱。安徽省是香椿的传统栽培区, 特别是太和县“太和香椿”(又称“贡椿”), 已成为安徽省的地方品牌产品之一。然而对香椿的分子遗传及转录组和基因组信息的研究十分缺乏。目前开展的香椿分子遗传研究主要集中在林业材用香椿遗传结构的分子标记分析, 缺乏菜用香椿营养风味品质相关的分子遗传学方面的研究, 其遗传背景和相关基因信息量极少, 给香椿营养风味品质形成的分子机制研究带来了严峻的挑战(刘军 等, 2008, 2009; Hsu et al., 2012; 裴宋雨 等, 2015)。

本研究以太和香椿优良品种‘黑油椿’的嫩芽为材料, 利用第 2 代高通量测序技术 Illumina HiSeq™ 4000 对其进行转录组测序, 以期分析香椿嫩芽转录组的特性及其基本的生物信息学特征, 为今后香椿重要性状关联的基因克隆、功能分析、次生代谢产物形成及代谢物含量调控相关基因的鉴定和挖掘提供基础资料, 为未来利用基因工程技术选育香椿新品种提供参考。

1 材料与方法

1.1 试验材料

试验香椿材料为 3 年生‘黑油椿’品种枝条上新生顶芽, 于 2016 年 4 月谷雨前后采自安徽省太和县香椿种植园。选取长约 10 cm 嫩芽, 液氮速冻, 并迅速转移至 - 80 °C 保存备用。

1.2 主要挥发性萜类物质的检测

参照 Arjan 等(2000)的方法, 取 1 g 新鲜嫩芽样品, 液氮研磨成粉后随即放入 20 mL 顶空瓶中, 用密封垫密封, 在 80 °C 水浴条件下用 SPME 萃取 1 h 后顶空抽气及 GC - MS (HS - GC - MS) (Agilent HS-7697A/ GC-7890A/MS-5975C, 气相色谱柱为 DB-5) 分析嫩芽全部萜类物质的含量。柱温控制程序: 50 °C 先保持 1 min; 然后以 10 °C · min⁻¹ 的速度升温到 100 °C 并保持 1 min; 此后以 4 °C · min⁻¹ 的速度升温到 200 °C 并保持 1 min; 此后再以 3 °C · min⁻¹ 的速度升温到 280 °C 并保持 7 min。

所得质谱图直接与该机 Nist 质谱库检出的标准图谱进行对比进行定性分析。以葵酸内酯为内标, 按面积归一化法计算各组分含量。试验设 3 个生物学重复, 每生物学重复有 3 个技术重复。

1.3 总 RNA 提取和检测

嫩芽总 RNA 提取采用 Trizol Reagent 方法, 并加入 DNaseI 除去残余的 DNA, 随后检测提取后总 RNA 质量及完整性、RNA 纯度 ($OD_{260/280}$ 比值), 最后用 Agilent 2100 Bioanalyzer 精确检测 RNA 完整性。每个样本取等量 RNA 混合组成 1 个 RNA 池, 测序待用。

1.4 转录组测序

样品检测合格后, 用带有 Oligo (dT) 的磁珠富集真核生物 mRNA, 加入试剂在恒温混匀仪中适温将 mRNA 断裂为短片段。以断裂后的 mRNA 为模板合成一链 cDNA, 然后配制第 2 链合成反应体系合成第 2 链 cDNA, 并使用试剂盒纯化回收、粘性末端修复、cDNA 的 3' 末端加上碱基 “A” 并连接接头, 然后进行片段大小选择, 最后进行 PCR 扩增。构建好的文库用 Agilent 2100 Bioanalyzer 和 ABI StepOnePlus Real-Time PCR System 质检, 合格后使用 Illumina HiSeq™ 4000 平台进行测序。

1.5 序列拼接和组装

测序的原始读序包含低质量、接头污染以及未知碱基 N 含量过高的读序, 去除测序接头污染、冗余重复等低质量的读序后, 获得干净读序数据, 统计干净读序的数量、总长度、Q20 和 GC 含量等。使用 Trinity 软件对干净读序 (去除 PCR 重复以提高组装效率) 进行从头组装获得转录本, 然后使用 Tgicl 软件将组装的转录本进行聚类去冗余, 得到样品的单基因簇, 分析转录本和单基因簇的长度及其分布。

1.6 序列注释和功能分类

将拼接后形成的高质量的单基因簇序列通过 BLASTX 与网站公布的公共蛋白数据库进行比对分析, 设置 $E\text{-value} < 10^{-5}$ 为阈值获取单基因簇功能注释。公共蛋白数据库包括 NCBI 非冗余核酸数据库 (Nr), Swiss-Prot 蛋白质序列数据库, GO (Gene Ontology, 基因本体论数据库), COG (Cluster of Orthologous Groups, 蛋白质直系同源数据库) 和 KEGG (Kyoto Encyclopedia of Genes and Genomes, 东京基因与基因组百科全书)。序列相似性比对最高的蛋白被用于 ‘黑油椿’ 嫩芽单基因簇功能注释。与 Nr 数据库比对后获得注释的单基因簇, 使用 Blast2GO 软件进一步进行 GO 注释, 并运用 WEGO 软件进行 GO 功能分类统计, 同时对单基因簇与 COG 数据库进行比对, 对其可能的功能进行分类统计。对单基因簇进行 KEGG 代谢通路分析, 进一步了解香椿嫩芽代谢途径、生物学功能及基因间存在的相互作用。将成功比对到各蛋白数据库的 CDS (Coding DNA Sequence) 的单基因簇和通过 ESTScan 软件预测 CDS 的 (数据库未比对上) 单基因簇合并进行序列长度分布统计。

2 结果与分析

2.1 ‘黑油椿’ 嫩芽挥发性香气成分及其含量

采用顶空固相萃取方法从新鲜采集的 ‘黑油椿’ 嫩芽中检测到的挥发性香气的成分主要有单萜类化合物 α -蒎烯和 D-柠檬烯; 倍半萜化合物主要有 β -石竹烯、 α -衣兰油烯、去氢香橙烯、9,10-脱氢异长叶烯、 α -愈创木烯、 α -芹子烯和其同分异构体、香橙烯及其氧化物等 (表 1)。从

检测的萜类化合物的成分构成上来看，倍半萜类化合物种类多而单萜化合物的种类较少（仅检测到两种）而且含量也相对较少， α -蒎烯和 *D*-柠檬烯的含量分别为 255.77 和 106.14 ng · g⁻¹；倍半萜类化合物中以去氢香橙烯的含量最高，为 4 440.71 ng · g⁻¹，9,10-脱氢异长叶烯和 β -石竹烯含量也较高分别为 1 932.02 和 1 799.89 ng · g⁻¹；衣兰烯和 α -愈创木烯含量为 1 295.52 和 1 286.76 ng · g⁻¹。具有相同分子式和分子量的同分异构体的化合物成分总量也很高，但同分异构体之间的相互转化而形成不同命名的各种倍半萜化合物的种类多样，含量分散。萜烯类氧化物的含量也较低。总体看来，嫩芽中挥发性香气物质以倍半萜为主，其中以去氢香橙烯的含量最高，9,10-脱氢异长叶烯和 β -石竹烯的含量的含量次之。这与文献报道（刘常金 等，2013）略有不同，可能与香椿的品种差异或测试的方法不同有关。

表 1 ‘黑油椿’香椿嫩芽萜类香气化合物及其含量
Table 1 Abundance of volatile terpenoid compounds in *Toona sinensis* ‘Heiyouchun’ sprouts

	序号 No.	化合物 Compound	分子式 Molecular formula	相对分子量 Relative molecular weight	含量/(ng · g ⁻¹) Content
单萜类 Monoterpenoids	1	α -蒎烯 α -Pinene	C ₁₀ H ₁₆	136.24	255.77 ± 27.80
	2	<i>D</i> -柠檬烯 <i>D</i> -Limonene	C ₁₀ H ₁₆	136.24	106.14 ± 6.50
倍半萜类 Sesquiterpenoids	3	α -芹子烯 α -selinene	C ₁₅ H ₂₄	204.35	402.68 ± 47.32
	4	β -芹子烯 β -selinene	C ₁₅ H ₂₄	204.35	404.22 ± 40.68
	5	α -可巴烯 α -Copaene	C ₁₅ H ₂₄	204.35	153.27 ± 13.10
	6	β -石竹烯 β -Caryophyllene	C ₁₅ H ₂₄	204.35	1 799.89 ± 83.74
	7	α -石竹烯 α -Caryophyllene	C ₁₅ H ₂₄	204.35	221.925 ± 38.60
	8	石竹烯氧化物 Caryophyllene oxide	C ₁₅ H ₂₄ O	220.39	259.42 ± 18.65
	9	顺式- β -法尼烯 Cis- β -Farnesene	C ₁₅ H ₂₄	204.35	54.84 ± 4.73
	10	(<i>Z,Z</i>)- α -法尼烯 (<i>Z,Z</i>)- α -Farnesene	C ₁₅ H ₂₄	204.35	233.95 ± 10.68
	11	衣兰烯 Ylangene	C ₁₅ H ₂₄	204.35	1 295.52 ± 66.60
	12	α -衣兰油烯 α -Murolene	C ₁₅ H ₂₄	204.35	104.02 ± 7.08
	13	反式-菖蒲烯 Trans-Calamenene	C ₁₅ H ₂₂	202.34	404.22 ± 37.22
	14	α -白菖考烯 α -Calacorene	C ₁₅ H ₂₄	204.35	365.64 ± 12.01
	15	(-)- β -榄香烯 (-)- β -Elemene	C ₁₅ H ₂₄	204.35	144.14 ± 12.36
	16	去氢香橙烯 Aromadendrene, dehydro-	C ₁₅ H ₂₂	202.34	4 440.71 ± 396.76
	17	9,10-脱氢异长叶烯 Isolongifolene,9,10-dehydro-	C ₁₅ H ₂₂	202.34	1 932.02 ± 194.25
	18	α -愈创木烯 α -Guaiene	C ₁₅ H ₂₄	204.35	1 286.76 ± 141.48
	19	γ -雪松烯 γ -Himachalene	C ₁₅ H ₂₄	204.35	122.05 ± 16.88
	20	异香橙烯环氧化物 Isoaromadendrene epoxide	C ₁₅ H ₂₄ O	220.39	383.28 ± 31.09
	21	香橙烯氧化物 Alloaromadendrene oxide-	C ₁₅ H ₂₄ O	220.39	120.82 ± 7.32

2.2 ‘黑油椿’嫩芽转录组测序及组装

‘黑油椿’嫩芽转录组测序获得原始数据量达 42.46 Mb，去除测序接头污染、冗余重复等低质量的读序后，得到干净读序数据量为 36.16 Mb，总高质量的碱基数据量为 4.70 Gb，碱基 Q20（测序错误率小于 1%）为 97.85%。测序数据结果表明，利用 Illumina HiSeq™ 4000 平台获得数据的数量及质量均较高，为下一步序列组装提供可靠原始数据。数据过滤后，使用 Trinity 软件对干净读序进行从头组装，共计获得 86 870 个转录本，总碱基数为 80 453 929，每个转录本的平均长度为 926 bp，N50（判断‘黑油椿’基因组拼接结果优劣的依据）为 1 649 bp，GC 含量占总碱基数 41.16%（表 2）。

转录本长度分布如表 3 所示，其中 200 ~ 1 000 bp 序列占 65.61%，1 000 ~ 2 000 bp 序列占 22.15%，2 000 ~ 3 000 bp 序列占 8.63%，大于 3 000 bp 序列占 3.61%。将这些转录本进一步组装，得到 55 850 个单基因簇，单基因簇分为两部分：一部分是基因簇，共计 20 037 个，同一个簇里面有若干个相似

度高（大于 70%）的读序（以 CL 开头，CL 后面接基因家族的编号）；其余的是不能组装的单一读序（以 Unigene 开头），代表单独的基因，共计为 25 813 个。全部单基因簇总碱基数为 56 626 762，每个单基因簇的平均长度为 1 013 bp，N50 为 1 680 bp，GC 含量占总碱基数 40.96%（表 2）。单基因簇长度分布如表 3 所示 300 ~ 1 000 bp 序列占 61.74%，1 000 ~ 2 000 bp 序列占 24.59%，2 000 ~ 3 000 bp 序列占 9.48%，大于 3 000 bp 序列占 4.19%。通过对比分析转录本和单基因簇长度分布发现，组装后的基因长度显著增加，组装效果较好，可进一步开展后续的生物信息学分析。

表 2 ‘黑油椿’香椿嫩芽转录组 Transcript 和 Unigene 数据组装质量指标
Table 2 Quality metrics of data assembly for transcript and Unigene in the transcriptome of *Toona sinensis* ‘Heiyouchun’ sprout

转录组 Transcriptome	数量 Number	碱基数 Bases number	平均长度 Mean length	N50	GC 含量/% GC content
转录本 Transcript	86 870	80 453 929	926	1 649	41.16
单基因簇 Unigene	55 850	56 626 762	1 013	1 680	40.96

表 3 转录本和单基因簇的序列大小
Table 3 Sequence size (bp) of Transcripts and Unigenes

长度范围/bp Length range	转录本 Transcript		单基因簇 Unigene	
	数量 Number	百分比/% Percentage	数量 Number	百分比/% Percentage
200 ~ 300	28 449	32.75	14 178	25.39
300 ~ 1 000	28 548	32.86	34 484	61.74
1 000 ~ 2 000	19 239	22.15	13 732	24.59
2 000 ~ 3 000	7 496	8.63	5 296	9.48
≥ 3 000	3 138	3.61	2 338	7.79

参照 Audic 和 Claverie（1997）的 FPKM 方法计算测序后组装得到的单基因簇的表达水平。统计分析表明，嫩芽单基因簇的 FPKM 平均值为 13.89，最大值为 5 056.6（CL6299.Contig2_All）。FPKM 值大于 500 的有 87 个单基因簇，这些基因广泛参与发育和代谢过程。FPKM 值低于 0.2 的有 224 个单基因簇，说明 Illumina HiSeq™4000 能检测到极低水平的基因表达，检测的灵敏度较高。

2.3 ‘黑油椿’嫩芽单基因簇的 Nr 功能注释

组装完毕后，对组装得到的单基因簇进行 Nr 和 SwissProt 数据库注释（E-value < 10⁻⁵）。注释结果表明，比对到 Nr 和 SwissProt 数据库的序列分别为 39 408 个和 24 356 个，分别占单基因簇总数的 70.56%和 43.61%。从 Nr 注释基因匹配的物种分布情况来看，嫩芽单基因簇与甜橙（*Citrus sinensis*）匹配度最高，为 54.32%，注释到克莱门柚（*Citrus clementina*）的序列为 23.81%，可可（*Theobroma cacao*）有 4.07%，葡萄（*Vitis vinifera*）有 1.86%，其余 15.94%序列分布于其它数个物种中。‘黑油椿’在 NCBI 中 Nr 数据库匹配度较高，可能与 Nr 数据库包括物种序列信息完整性有关（Xu et al., 2013），但由于香椿的遗传信息较少，部分序列在数据库中没有发现其同源序列。

2.4 ‘黑油椿’嫩芽单基因簇的 GO 功能分类

为了进一步从整体水平上了解‘黑油椿’嫩芽基因功能分布特征，表 4 对‘黑油椿’嫩芽单基因簇进行 GO 数据库比对分析发现，有 19 704 个单基因簇注释到 GO 数据库中，占全部单基因簇总数的 35.28%。按 GO 功能分类系统，在转录组全部单基因簇中，功能归类到生物学过程的单基因簇有 57 787 个，占 44.36%；归类到细胞构成的单基因簇有 44 588 个，占 34.22%；归类到分子功能的有 27 902 个，占 21.42%。以上 3 大功能分类又可进一步地分为 54 个亚类，每一功能模块包括的亚类数分别为 22、17 和 15。在 22 亚类构成的生物进程这一大类中，以细胞进程、代谢进程和单一有

表 4 ‘黑油椿’ 香椿嫩芽 Unigene GO 分类
Table 4 GO functional categories of *Toona sinensis* ‘Heiyouchun’ sprout unigenes

GO 分类 GO functional categories	数量 Number	百分比/% Percentage	亚类 Subgroups	数量 Number	百分比/% Percentage
生物学过程 Biological process	57 787	44.36	生物黏附 Biological adhesion	15	0.03
			生物相 Biological phase	12	0.02
			生物调节 Biological regulation	3 250	5.62
			细胞组分部分或生物合成 Cellular component organization or biogenesis	1 798	3.11
			细胞进程 Cellular process	12 145	21.02
			发育进程 Developmental process	1 632	2.82
			生长 Growth	305	0.53
			免疫系统进程 Immune system process	257	0.44
			定位 Localization	2 745	4.75
			细胞活动 Locomotion	16	0.03
			代谢进程 Metabolic process	13 352	23.11
			多细胞进程 Multi-cellular organism process	1 524	2.64
			有机体进程 Multi-organism process	714	1.24
			生物学进程负调节 Negative regulation of biological process	450	0.78
			生物学进程正调节 Positive regulation of biological process	521	0.90
			生物学进程调节 Regulation of biological process	3 089	5.35
			繁殖 Reproduction	246	0.43
			繁殖进程 Reproductive process	879	1.52
			应急反应 Response to stimulus	3 521	6.09
			节律进程 Rhythmic process	58	0.10
			信号传导 Signaling	1 274	2.20
			单一有机体进程 Single-organism process	9 984	17.28
细胞构成 Cellular component	44 588	34.22	细胞 Cell	10 587	23.74
			细胞连接 Cell junction	154	0.35
			细胞部分 Cell part	10 587	23.74
			胞外基质 Extracellular matrix	10	0.02
			胞外基质部分 Extracellular matrix part	3	0.0067
			胞外区域 Extracellular region	547	1.22
			胞外区域部分 Extracellular region part	5	0.01
			大分子复合体 Macromolecular complex	2 501	5.61
			膜结构 Membrane	5 741	12.88
			膜封闭内腔 Membrane-enclosed lumen	608	1.36
			膜部分 Membrane part	2 789	6.26
			拟核 Nucleoid	38	0.09
			细胞器 Organelle	7 852	17.61
			细胞器部分 Organelle part	2 987	6.70
			共质体 Symplast	163	0.37
			病毒体 Virion	8	0.018
			病毒体部分 Virion part	8	0.018
分子功能 Molecular function	27 902	21.42	抗氧化活性 Antioxidant activity	130	0.47
			结合活性 Binding	11 478	41.14
			催化活性 Catalytic activity	12 847	46.04
			电子载体活性 Electron carrier activity	258	0.92
			酶调节活性 Enzyme regulation activity	204	0.73
			鸟嘌呤核苷酸交换因子活性 Guanyl-nucleotide exchange factor activity	30	0.11
			金属伴侣蛋白活性 Metallochaperone activity	3	0.01
			分子转导活性 Molecular transducer activity	371	1.33
			核苷酸结合的转录因子活性 Nucleic acid binding transcription factor activity	409	1.47
			营养贮存活性 Nutrient reservoir activity	18	0.06
			蛋白质结合的转录因子活性 Protein binding transcription factor activity	57	0.20
			蛋白标签 Protein tag	7	0.03
			受体活性 Receptor activity	107	0.38
			结构分子活性 Structural molecule activity	594	2.13
			转运活性 Transporter activity	1 389	4.98

机体进程比例较大, 分别为 21.02%, 23.11%和 17.28%, 而以生物附着、生物相和细胞活动功能分类所属的单基因簇为少数, 仅分别有 15、12 和 16 个。细胞构成功能模块中以细胞和细胞部分占比例较大, 均为 23.74%, 其次是细胞器和膜结构占有一定的比例, 分别为 17.61%和 12.88%, 病毒和病毒体组分最少, 仅有 8 个单基因簇。在分子功能的模块中, 以结合活性和催化活性所占有的单基因簇比例为高, 分别为 41.14%和 46.04%, 而与金属伴侣蛋白活性和蛋白标签所占比例最少, 仅分别有 3 和 7 个单基因簇。从以上这些分析结果可以看出‘黑油椿’嫩芽基因表达总体水平上以与代谢过程相关的基因表达居多, 表明其在特定生长发育期内参与代谢相关功能基因数量与其较强的代谢能力相适应。

2.5 ‘黑油椿’嫩芽单基因簇的 COG 功能分类

表 5 对组装的全部单基因簇与 COG 数据库比对, 得到 14 186 个单基因簇注释到 COG 数据库中, 占全部单基因簇总数的 25.40%, 共获得 35 415 个单基因簇, 分布于 25 功能区域, 不同功能区域分布的单基因簇数各不相同, 其中以涉及作用范围更广泛的一般功能基因数最多, 占 18.12%, 转录 (9.50%)、信号转导机制 (7.10%)、复制重组和修复 (8.52%)、翻译、核糖体结构和生物发生 (6.31%)、翻译后修饰、蛋白折叠和分子伴侣 (7.08%) 包括的功能基因丰富, 碳水化合物运输和代谢所涉及的基因较多, 占 5.67%。另有 1 694 个单基因簇功能未知。值得关注的是由 957 个单基因簇与次生代谢物合成、运输及代谢相关, 这些单基因簇为后续进一步地研究‘黑油椿’嫩芽中与黄酮、萜类和皂苷成分等次生代谢物合成相关基因提供了基础数据。

表 5 ‘黑油椿’香椿嫩芽 Unigene COG 分类
Table 5 COG functional categories of *Toona sinensis* ‘Heiyouchun’ sprout unigenes

COG 功能分类 COG functional categories	数量 Number	百分比/% Percentage
翻译、核糖体结构和生物发生 Translation, ribosomal structure and biogenesis	2 236	6.31
转录 Transcription	3 365	9.50
信号转导机制 Signal transduction mechanism	2 513	7.10
次生代谢物合成、运输和代谢 Secondary metabolites biosynthesis, transport and catabolism	957	2.70
RNA 加工和修饰 RNA processing and modification	347	0.98
复制、重组和修复 Replication, recombination and repair	3 018	8.52
翻译后修饰、蛋白质折叠和分子伴侣 Posttranslational modification, protein turnover and chaperones	2 506	7.08
核苷酸运输和代谢 Nucleotide	336	0.95
核结构 Nuclear structure	8	0.02
脂质运输和代谢 Lipid transport and metabolism	795	2.24
细胞内转运、分泌和小泡运输 Intracellular trafficking, secretion and vesicular transport	875	2.47
无机离子运输和代谢 Inorganic ion transport and metabolism	865	2.44
一般功能预测 General function prediction only	6 417	18.12
未知功能 Function unknown	1 694	4.78
胞外结构 Extracellular structures	12	0.03
能量生成和转换 Energy production and conversion	954	2.69
防卫机制 Defense mechanisms	327	0.92
细胞骨架 Cytoskeleton	568	1.60
辅酶运输和代谢 Coenzyme transport and metabolism	673	1.90
染色质结构和活力 Chromatin structure and dynamics	458	1.29
细胞壁膜生物发生 Cell wall membrane biogenesis	1 417	4.00
细胞运动 Cell motility	354	1.00
细胞周期控制、细胞分裂、染色体分区 Cell cycle control, cell division, chromosome partitioning	1 482	4.18
碳水化合物运输和代谢 Carbohydrate transport and metabolism	2 007	5.67
氨基酸运输和代谢 Amino acid transport and metabolism	1 231	3.48

2.6 ‘黑油椿’嫩芽单基因簇的 KEGG 代谢通路分析

将‘黑油椿’嫩芽单基因簇与 KEGG 数据库比对，发现共有 28 400 个单基因簇获得注释，占全部单基因簇的 50.85%。对其进行统计分析，可将与 KEGG 代谢通路相关的单基因簇归为 6 大类、21 个亚类和 135 个代谢通路。由表 6 可以看出单基因簇在代谢相关通路中所占比例为最高，达 58.43 %；遗传信息处理相关的单基因簇数量次之，达 24.34%；与人类疾病相关的单基因簇数量，仅占 0.79%。鉴于代谢通路相关的单基因簇数量最多，进一步对这一大类所属的亚类进行分析，发现包括 11 个代谢路径，包括氨基酸代谢、其它次生代谢生物合成、碳水化合物代谢、能量代谢、全局代谢映射、糖生物合成和代谢、脂质代谢、辅助因子和维生素代谢、其它氨基酸代谢、萜类和聚酮化合物代谢、核苷酸代谢。

总体来看‘黑油椿’嫩芽中物质代谢活跃，既包含营养物质的形成，如氨基酸、维生素、脂类、糖类，一些次生代谢物，如黄酮、花青素、与‘黑油椿’香气形成相关的萜类物质，又包含嫩芽快速生长所需要的能量供应，细胞分裂所需要的核苷酸的合成，这与前期研究发现太和香椿嫩芽营养丰富，在这一时期具有很强的代谢活动能力相一致。

表 6 ‘黑油椿’香椿嫩芽 Unigene KEGG 分类
Table 6 KEGG categories of *Toona sinensis* (Heiyouchun) sprout unigenes

KEGG 功能分类 KEGG categories	数量 Number	百分比/% Percentage	亚类 Subgroups	数量 Number	百分比/% Percentage
生物系统 Organismal systems	1 721	5.35	环境适应 Environment adaptation	1 721	5.35
代谢 Metabolism	18 810	58.43	核苷酸代谢 Nucleotide metabolism	884	2.75
			萜类和聚酮化合物代谢 Metabolism of terpenoids and polyketides	780	2.42
			其它氨基酸代谢 Metabolism of other amino acid	628	1.95
			辅助因子和维生素代谢 Metabolism of cofactors and vitamins	907	2.82
			脂质代谢 Lipid metabolism	1 485	4.61
			糖生物合成和代谢 Glycan biosynthesis and metabolism	678	2.11
			全局路径 Global map	7 244	22.50
			能量代谢 Energy metabolism	1 082	3.36
			碳水化合物代谢 Carbohydrate metabolism	2 561	7.96
			其它次生代谢物的生物合成 Biosynthesis of other secondary metabolites	1 043	3.24
			氨基酸代谢 Amino acid metabolism	1 518	4.72
人类疾病 Human diseases	253	0.79	激素和代谢疾病 Endocrine and metabolic diseases	241	0.75
			药物抗性 Drug resistance	12	0.04
遗传信息处理 Genetic information processing	7 836	24.34	翻译 Translation	3 289	10.22
			转录 Transcription	1 554	4.83
			复制和修复 Replication and repair	667	2.07
			折叠、分类和降解 Folding, sorting and degradation	2 326	7.23
环境信息处理 Environmental information processing	1 617	5.02	信号转导 Signal transduction	1 372	4.26
			膜运输 Membrane transport	245	0.76
细胞过程 Cellular processes	1 953	6.07	运输和代谢 Transport and catabolism	1 953	6.07

‘黑油椿’嫩芽 135 个 KEGG 代谢通路，按每个代谢通路参与成功比对并功能注释的单基因簇

数量的高低依次排序如表 7 所示。由表 7 可以看出,代谢途径的单基因簇数量为第一,占全部 15.83%,其次是次生代谢合成和植物与病原互作,分别占 8.84%和 3.10%。这进一步表明‘黑油椿’嫩芽在其特定的生长发育阶段进行活跃的代谢反应,深入研究关键基因所参与这些代谢途径的反应,对揭示香椿嫩芽的风味品质形成机制具有重要意义。

表 7 ‘黑油椿’香椿嫩芽 unigene 的代谢途径分析
Table 7 Analysis of metabolic pathways of *Toona sinensis* (Heiyouchun) sprout unigenes

编号 No.	代谢途径 Pathway	基因数量 Number	百分比/% Percentage	编码 ID
1	代谢途径 Metabolic pathways	6 805	15.83	ko01100
2	次生代谢物生物合成 Biosynthesis of secondary metabolites	3 799	8.84	ko01110
3	植物病原菌相互作用 Plant-pathogen interaction	1 332	3.10	ko04626
4	RNA 运输 RNA transport	1 255	2.92	ko03013
5	剪接体 Spliceosome	1 185	2.76	ko03040
6	植物激素信号转导 Plant hormone signal transduction	1 140	2.65	ko04075
7	内吞作用 Endocytosis	1 129	2.63	ko04144
8	核糖体 Ribosome	902	2.10	ko03010
9	碳代谢 Carbon metabolism	872	2.03	ko01200
10	氨基酸生物合成 Biosynthesis of amino acids	866	2.01	ko01230
11	淀粉和蔗糖代谢 Starch and sucrose metabolism	848	1.97	ko00500
12	蛋白在内质网加工 Protein processing in endoplasmic reticulum	846	1.97	ko04141
13	mRNA 监视途径 mRNA surveillance pathway	833	1.94	ko03015
14	嘌呤代谢 Purine metabolism	675	1.57	ko00230
15	RNA 降解 RNA degradation	613	1.43	ko03018
16	嘧啶代谢 Pyrimidine metabolism	590	1.37	ko00240
17	泛素介导的蛋白降解 Ubiquitin mediated proteolysis	580	1.35	ko04120
18	糖酵解/糖异生 Glycolysis / Gluconeogenesis	477	1.11	ko00010
19	苯丙素生物合成 Phenylpropanoid biosynthesis	476	1.11	ko00940
20	氨基糖和核苷酸糖代谢 Amino sugar and nucleotide sugar metabolism	465	1.08	ko00520
21	核糖体合成 Ribosome biogenesis in eukaryotes	455	1.06	ko03008
22	甘油磷脂代谢 Glycerophospholipid metabolism	417	0.97	ko00564
23	植物昼夜节律 Circadian rhythm - plant	407	0.95	ko04712
24	半胱氨酸和蛋氨酸代谢 Cysteine and methionine metabolism	360	0.84	ko00270
25	氧化磷酸化 Oxidative phosphorylation	339	0.79	ko00190
26	甘油酯代谢 Glycerolipid metabolism	336	0.78	ko00561
27	丙酮酸代谢 Pyruvate metabolism	335	0.78	ko00620
28	戊糖和糖醛酸转换 Pentose and glucuronate interconversions	314	0.73	ko00040
29	过氧化物酶体 Peroxisome	303	0.70	ko04146
30	光合作用碳固定 Carbon fixation in photosynthetic organisms	299	0.70	ko00710
31	戊糖磷酸途径 Pentose phosphate pathway	286	0.67	ko00030
32	吞噬体 Phagosome	282	0.66	ko04145
33	吞噬调节 Regulation of autophagy	282	0.66	ko04140
34	半乳糖代谢 Galactose metabolism	275	0.64	ko00052
35	氨酰 - tRNA 合成酶 Aminoacyl-tRNA biosynthesis	272	0.63	ko00970
36	甘氨酸、丝氨酸和苏氨酸代谢 Glycine, serine and threonine metabolism	255	0.29	ko00260
37	核苷酸切除修复 Nucleotide excision repair	254	0.29	ko03420
38	脂肪酸代谢 Fatty acid metabolism	252	0.28	ko01212
39	ABC 转运蛋白 ABC transporters	250	0.27	ko02010
40	2 - 氧代羧酸代谢 2-Oxocarboxylic acid metabolism	248	0.27	ko01210
41	磷脂酰肌醇信号系统 Phosphatidylinositol signaling system	247	0.27	ko04070

续表 7

编号 No.	代谢途径 Pathway	基因数量 Number	百分比/% Percentage	编码 ID
42	胰岛素抗性 Insulin resistance	247	0.26	ko04931
43	N - 多糖生物合成 N-Glycan biosynthesis	238	0.26	ko00510
44	乙醛酸和二羧酸代谢 Glyoxylate and dicarboxylate metabolism	233	0.26	ko00630
45	磷酸肌醇代谢 Inositol phosphate metabolism	230	0.25	ko00562
46	卟啉和叶绿素代谢 Porphyrin and chlorophyll metabolism	229	0.25	ko00860
47	同源重组 Homologous recombination	227	0.25	ko03440
48	其它聚糖降解 Other glycan degradation	227	0.24	ko00511
49	萜类骨架生物合成 Terpenoid backbone biosynthesis	226	0.24	ko00900
50	果糖和甘露糖代谢 Fructose and mannose metabolism	226	0.24	ko00051
51	氰基乙酸代谢 Cyanoamino acid metabolism	219	0.23	ko00460
52	碱基切除修复 Base excision repair	218	0.22	ko03410
53	DNA 复制 DNA replication	218	0.20	ko03030
54	抗坏血酸代谢 Ascorbate and aldarate metabolism	217	0.20	ko00053
55	RNA 聚合酶 RNA polymerase	208	0.20	ko03020
56	谷胱甘肽代谢 Glutathione metabolism	205	0.20	ko00480
57	缬氨酸, 亮氨酸和异亮氨酸降解 Valine, leucine and isoleucine degradation	186	0.19	ko00280
58	基础转录因子 Basal transcription factors	183	0.18	ko03022
59	黄酮生物合成 Flavonoid biosynthesis	181	0.18	ko00941
60	芪类、二苯基庚酮和姜酚生物合成 Stilbenoid, diarylheptanoid and gingerol biosynthesis	180	0.17	ko00945
61	鞘脂生物合成 Sphingolipid metabolism	179	0.17	ko00600
62	光合作用 Photosynthesis	177	0.17	ko00195
63	酪氨酸代谢 Tyrosine metabolism	169	0.16	ko00350
64	三羧酸循环 Citrate cycle (TCA cycle)	169	0.15	ko00020
65	丙氨酸、天冬氨酸和谷氨酸代谢 Alanine, aspartate and glutamate metabolism	166	0.14	ko00250
66	精氨酸和脯氨酸代谢 Arginine and proline metabolism	165	0.14	ko00330
67	α - 亚麻酸代谢 α -Linolenic acid metabolism	162	0.13	ko00592
68	赖氨酸降解 Lysine degradation	161	0.13	ko00310
69	错配修复 Mismatch repair	156	0.13	ko03430
70	硫代谢 Sulfur metabolism	147	0.12	ko00920
71	苯丙氨酸、酪氨酸和色氨酸生物合成 Phenylalanine, tyrosine and tryptophan biosynthesis	144	0.12	ko00400
72	精氨酸生物合成 Arginine biosynthesis	143	0.12	ko00220
73	β - 丙氨酸代谢 β -Alanine metabolism	143	0.12	ko00410
74	苯丙氨酸代谢 Phenylalanine metabolism	143	0.11	ko00360
75	泛醌和其它萜醌生物合成 Ubiquinone and other terpenoid-quinone biosynthesis	142	0.11	ko00130
76	蛋白酶体 Proteasome	142	0.10	ko03050
77	泛酸酯和 CoA 生物合成 Pantothenate and CoA biosynthesis	135	0.10	ko00770
78	类胡萝卜素生物合成 Carotenoid biosynthesis	133	0.10	ko00906
79	蛋白质输出 Protein export	131	0.09	ko03060
80	类固醇生物合成 Steroid biosynthesis	124	0.09	ko00100
81	丙酸乙酯代谢 Propanoate metabolism	123	0.08	ko00640
82	糖基磷脂酰肌醇 (GPI) 锚生物合成 Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	121	0.08	ko00563
83	囊泡运输中的陷阱相互作用 SNARE interactions in vesicular transport	117	0.07	ko04130
84	柠檬烯和蒎烯 Limonene and pinene degradation	116	0.07	ko00903
85	脂肪酸延伸 Fatty acid elongation	114	0.07	ko00062
86	色氨酸代谢 Tryptophan metabolism	113	0.05	ko00380
87	醚脂质代谢 Ether lipid metabolism	112	0.05	ko00565
88	氮代谢 Nitrogen metabolism	111	0.03	ko00910
89	缬氨酸、亮氨酸和异亮氨酸生物合成 Valine, leucine and isoleucine biosynthesis	109	0.03	ko00290

续表 7

编号 No.	代谢途径 Pathway	基因数量 Number	百分比/% Percentage	编码 ID
90	脂肪酸生物合成 Fatty acid biosynthesis	106	0.03	ko00061
91	不饱和脂肪酸生物合成 Biosynthesis of unsaturated fatty acids	106	0.00	ko01040
92	脂肪酸降解 Fatty acid degradation	105	0.29	ko00071
93	角质、木栓质和蜡质生物合成 Cutin, suberine and wax biosynthesis	103	0.29	ko00073
94	莨菪烷、哌啶和吡啶生物碱生物合成 Tropane, piperidine and pyridine alkaloid biosynthesis	102	0.28	ko00960
95	多糖降解 Glycosaminoglycan degradation	97	0.27	ko00531
96	丁酸乙酯代谢 Butanoate metabolism	93	0.27	ko00650
97	异喹啉生物碱生物合成 Isoquinoline alkaloid biosynthesis	88	0.27	ko00950
98	倍半萜和三萜生物合成 Sesquiterpenoid and triterpenoid biosynthesis	86	0.26	ko00909
99	鞘糖脂生物合成-神经节系列 Glycosphingolipid biosynthesis - ganglio series	84	0.26	ko00604
100	二萜生物合成 Diterpenoid biosynthesis	84	0.26	ko00904
101	玉米素生物合成 Zeatin biosynthesis	83	0.25	ko00908
102	烟酸和烟酰胺代谢 Nicotinate and nicotinamide metabolism	79	0.25	ko00760
103	异黄酮生物合成 Isoflavonoid biosynthesis	77	0.25	ko00943
104	有机硒化合物代谢 Selenocompound metabolism	72	0.24	ko00450
105	单萜生物合成 Monoterpenoid biosynthesis	71	0.24	ko00902
106	叶酸碳库 One carbon pool by folate	71	0.24	ko00670
107	花生四烯酸代谢 Arachidonic acid metabolism	70	0.23	ko00590
108	组氨酸代谢 Histidine metabolism	65	0.22	ko00340
109	核黄素代谢 Riboflavin metabolism	61	0.20	ko00740
110	赖氨酸生物合成 Lysine biosynthesis	60	0.20	ko00300
111	生物素代谢 Biotin metabolism	57	0.20	ko00780
112	亚油酸代谢 Linoleic acid metabolism	56	0.20	ko00591
113	黄酮和黄酮醇的生物合成 Flavone and flavonol biosynthesis	56	0.19	ko00944
114	芥子油苷的生物合成 Glucosinolate biosynthesis	53	0.18	ko00966
115	C5 支链酸代谢 C5-Branched dibasic acid metabolism	51	0.18	ko00660
116	单环 β - 内酰胺的合成 Monobactam biosynthesis	50	0.17	ko00261
117	维生素 B6 代谢 Vitamin B6 metabolism	50	0.17	ko00750
118	其他类型的 O - 聚糖的生物合成 Other types of O-glycan biosynthesis	48	0.17	ko00514
119	叶酸生物合成 Folate biosynthesis	46	0.16	ko00790
120	光合作用 - 天线蛋白质类 Photosynthesis - antenna proteins	44	0.15	ko00196
121	硫中继系统 Sulfur relay system	41	0.14	ko04122
122	非同源末端连接 Non-homologous end-joining	41	0.14	ko03450
123	油菜素内酯生物合成 Brassinosteroid biosynthesis	39	0.13	ko00905
124	鞘糖脂生物合成 Glycosphingolipid biosynthesis - globo series	38	0.13	ko00603
125	硫胺素代谢 Thiamine metabolism	36	0.13	ko00730
126	芳烃降解 Degradation of aromatic compounds	33	0.12	ko01220
127	牛磺酸和牛磺酸代谢 Taurine and hypotaurine metabolism	32	0.12	ko00430
128	花青素生物合成 Anthocyanin biosynthesis	31	0.12	ko00942
129	吲哚生物碱的生物合成 Indole alkaloid biosynthesis	29	0.12	ko00901
130	苯并恶嗪酮生物合成 Benzoxazinoid biosynthesis	20	0.11	ko00402
131	酮体的合成与降解 Synthesis and degradation of ketone bodies	20	0.11	ko00072
132	硫辛酸的代谢 Lipoic acid metabolism	15	0.10	ko00785
133	咖啡碱代谢 Caffeine metabolism	12	0.10	ko00232
134	万古霉素抗性 Vancomycin resistance	12	0.10	ko01502
135	甜菜红素的生物合成 Betalain biosynthesis	1	0.01	ko00965

进一步挖掘与‘黑油椿’风味形成相关的萜类合成的单基因簇，发现涉及到萜类化合物骨架生物合成单基因簇 226 个 (Terpenoid backbone biosynthesis)，单萜生物合成单基因簇 71 个，倍半

萜和三萜合成单基因簇 86 个，二萜合成单基因簇 84 个。萜类化合物前体生物合成包括位于细胞质的甲羟戊酸途径和位于质体中的 5 - 磷酸脱氧木酮糖/2C - 甲基 4 - 磷酸 - 4D - 赤藓糖醇途径 (Vranová et al., 2013)。在甲羟戊酸代谢途径中发现与黑油椿嫩芽风味形成相关倍半萜合成酶类 (表 8)，包括乙酰辅酶 A 酰基转移酶、HMG-CoA 合成酶、HMG-CoA 还原酶、甲羟戊酸激酶、磷酸甲羟戊酸激酶、焦磷酸甲羟戊酸脱羧酶、异戊烯焦磷酸异构酶、异戊烯转移酶、倍半萜类合成酶。在 5 - 磷酸脱氧木酮糖/2C - 甲基 4 - 磷酸 - 4D - 赤藓糖醇途径中发现与‘黑油椿’嫩芽风

表 8 ‘黑油椿’ 香椿嫩芽转录组中参与萜类合成的酶
Table 8 The enzymes involving in terpenoids biosynthesis of transcriptome for *Toona sinensis* (Heiyouchun) sprout

EC 号 EC No.	酶名称 Enzyme definition	酶缩写 Enzyme abbreviatio	单基因簇数 Unigene number	编号 Entry
EC 2.3.1.9	乙酰辅酶 A 酰基转移酶 Hcetyl-CoA C-acetyltransferase	AACT	7	K00626
EC 2.3.3.10	羟甲基戊二酰 - CoA 合酶 Hydroxymethylglutaryl-CoA synthase	HMGS	8	K01597
EC 1.1.1.34	羟甲基戊二酰 - CoA 还原酶 Hydroxymethylglutaryl-CoA reductase	HMGR	17	K00938
EC 2.7.1.36	甲羟戊酸激酶 Mevalonate kinase	MK	2	K12742
EC 2.7.4.2	磷酸甲羟戊酸激酶 Phosphomevalonate kinase	PMK	4	K13273
EC 4.1.1.33	焦磷酸甲羟戊酸脱羧酶 Diphosphomevalonate decarboxylase	MPDC	4	K01597
EC 2.2.1.7	5 - 磷酸脱氧木酮糖合成酶 1-Deoxy-D-xylulose-5-phosphate synthase	DXS	14	K00021
EC 1.1.1.267	5 - 磷酸脱氧木酮糖还原异构酶 1-deoxy-D-xylulose-5-phosphate reductoisomerase	DXR	1	K00099
EC 2.7.7.60	4 - 磷酸 - 2C - 甲基赤藓糖醇 - 4 - 胞苷焦磷酸合成酶 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	MCT	4	K00991
EC 2.7.1.148	2C - 甲基赤藓糖醇 - 4 - 胞苷焦磷酸激酶 4-Diphosphocytidyl-2-C-methyl-D-erythritol kinase	CMK	3	K00919
EC 4.6.1.12	2C - 甲基赤藓糖醇 - 2,4 - 焦磷酸合成酶 2-C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase	MDS	1	K03526
EC 1.17.7.1	4 - 羟基 - 2 - 甲基 - 2 - E - 丁烯基 - 4 - 焦磷酸合酶 (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	HDS	4	K03526
EC 1.17.7.2	4 - 羟基 - 2 - 甲基 - 2 - E - 丁烯基 - 4 - 焦磷酸还原酶 (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate reductase	HDR	9	K03527
EC 5.3.3.2	异戊烯焦磷酸异构酶 Isopentenyl-diphosphate Delta-isomerase	IPPI	4	K01823
EC 4.2.3.27	异戊烯合成酶 Isoprene synthase	IPS	8	K12742
EC 2.5.1.12.5.1.10	法尼烯二磷酸合成酶 Farnesyl diphosphate synthase	FDPS/GGPPS	19	K00787
EC 4.2.3.25	(3S) - 芳樟醇合成酶 (3S)-linalool synthase	LIS	10	K15086
EC 4.2.3.15	月桂烯/罗勒烯合酶 Myrcene/Ocimene synthase	MYS/OCS	2	K12467
EC 4.2.3.20	(R) - 柠檬烯合成酶 (R)-Limonene synthase	LS	6	K15096
EC 4.2.3.108	1,8 - 桉叶油醇合成酶 1,8-Cineole synthase	CINS	1	K07385
EC 4.2.3.111	α - 松油醇合成酶 (-)-alpha-Terpineol synthase	TES	7	K18108
EC 4.2.3.46	α - 法尼烯合成酶 α -Farnesene synthase	AFS	3	K14173
EC 4.2.3.224.2.3.75	大根香叶烯 D 合成酶 (-)-Germacrene D synthase	GDS	20	K15803
EC 4.2.3.734.2.3.86	蛇床烯合成酶 Selinene synthase	SES	10	K14181
EC 4.2.3.57	β - 石竹烯合成酶 β -Caryophyllene synthase	QHS	1	K14184
EC 2.5.1.21	法尼酰二磷酸酯法尼酰基转移酶 Farnesyl-diphosphate farnesyltransferase	FDFT	3	K00801
EC 1.14.14.17	鲨烯单加氧酶 Squalene monooxygenase	SQLE	11	K00511
EC 5.5.1.13	ent - 柯巴基焦磷酸合酶 Ent-Copalyl diphosphate synthase	CPS	6	K04120
EC 4.2.3.19	贝壳杉烯合酶 Ent-Kaurene synthase	KS	16	K04121
EC 1.14.13.78	贝壳杉烯氧化酶 Ent-Kaurene oxidase	KO	1	K04123

味形成相关单萜合成酶类, 包括 5 - 磷酸脱氧木酮糖合成酶、5 - 磷酸脱氧木酮糖还原异构酶、4 - 磷酸 - 2C - 甲基赤藓糖醇 4 - 胞苷焦磷酸合成酶、2C - 甲基赤藓糖醇 4 - 胞苷焦磷酸激酶、2C - 甲基赤藓糖醇 - 2,4 - 焦磷酸合成酶、1 - 羟基 - 2 - 甲基 - 2 - 丁烯 - 4 - 焦磷酸合成酶、香叶基焦磷酸合成酶和单萜类合成酶等。

2.7 ‘黑油椿’嫩芽转录组单基因簇的 CDS 预测

将‘黑油椿’嫩芽全部 55 850 个 Unigene 与 Nr、Swiss-Prot、KEGG 和 COG 数据库按照优先级顺序进行 BLASTx (E-value < 10⁻⁵) 比对, 以确定单基因簇的 CDS 区序列, 并将 CDS 区序列翻译成氨基酸序列, 获得成功比对后的单基因簇的 CDS 区氨基酸序列和核苷酸序列。对于无法比对到蛋白质数据库的单基因簇用 ESTScan 软件预测后得到 CDS 区氨基酸序列和核苷酸序列。将所有 55 850 个单基因簇与蛋白质数据库比对后, 发现有 46 204 个单基因簇被成功比对注释, ESTScan 软件预测了 628 个单基因簇, 共预测了 46 832 个单基因簇, 其长度分布如表 9 所示。从表 9 中可以看出, 200 ~ 1 000 bp 的 CDS 序列占 68.01%, 1 000 ~ 2 000 bp 的占 22.44%, 2 000 ~ 3 000 bp 的占 6.69 %, 3 000 bp 以上的占 2.86%。

表 9 ‘黑油椿’香椿嫩芽转录组 Unigene 的 CDS 序列长度分布
Table 9 CDS length distribution of transcriptome for *Toona sinensis* ‘Heiyouchun’ sprout

长度范围/bp Length range	数量 Number	百分比/% Percentage
200 ~ 1000	31 849	68.01
1 000 ~ 2 000	10 508	22.44
2 000 ~ 3 000	3 130	6.69
≥ 3 000	1 341	2.86

3 讨论

有关香椿在食品、医药保健中的作用, 香椿种质繁育及资源保护等方面的研究已受到广泛关注 (陈刚 等, 2013; Wu et al., 2014; 肖华山 等, 2014), 对香椿风味形成相关酶基因的研究及人体必需氨基酸 (赖氨酸) 合成分子代谢途径而开展的转录组测序也有部分报道 (Hsu et al., 2012; Zhang et al., 2016)。但不同地域栽培的香椿品种营养成分及风味组成差异很大, 表明其种质的遗传背景具有高度的多样性 (李聚英 等, 2011; 刘常金 等, 2013)。^{‘黑油椿’}香椿风味独特、香味浓郁、营养价值高, 是“太和香椿”优良品种的典型代表。本试验中‘黑油椿’香椿嫩芽经 GC - MS 分析检测到了多种萜类化合物, 其中以倍半萜化合物成分最多, 含量最高的为去氢香橙烯, 达 4 440.71 ng · g⁻¹, 其次为 9,10 - 脱氢异长叶烯和 β - 石竹烯。这一结果表明‘黑油椿’的风味形成与这 3 种萜类物质的合成密切相关。

本研究中采用 Illumina HiSeqTM 4000 测序平台对‘黑油椿’香椿嫩芽转录组测序及分析, Q20 (测序错误率小于 1%) 和 N50 (判断‘黑油椿’香椿嫩芽基因组拼接结果优劣的依据) 分别为 97.85% 和 1 649 bp, 表明测序质量高, 能够满足后续分析的需要。对测序结果进行拼接组装后共获得 55 850 单基因簇, 组装后获得的单基因簇平均长度为 1 013 bp, N50 为 1 680 bp, 序列较长。将组装后全部的单基因簇与 NCBI 上的 Nr 数据库比对, 有 39 408 个获得功能注释。本研究获得的‘黑油椿’香椿嫩芽的转录组数据在序列数量、Q20、序列大小分布、N 值和与公共蛋白数据库的匹配率等方面比山东的香椿转录组数据高 (Zhang et al., 2016), 可能的原因是由于不同的测序平台及香椿的品种不同所造成的。将测序组装后获取的 55 850 个单基因簇与 NCBI 中 Nr 蛋白数据库进行比对分析发现有 39 408 个单基因簇获得注释, 占总数的 70.56%, 也有近 30% 的单基因簇未获得相关的注释信

息。这一现象在许多物种的转录组测序分析同样存在 (Shi et al., 2011; 贾新平等, 2014; Sun et al., 2015), 究其原因可能是多方面的, 如短片段单基因簇占比过高, 即使成功比对单数据库中匹配的序列本身也未能获得功能注释, 也不排除‘黑油椿’中存在新的功能基因的可能。KEGG 功能注释 28 400 个单基因簇参与 135 个代谢途径, 这些代谢途径不仅包括如碳水化合物代谢、氨基酸代谢、脂类等营养物质代谢等, 还涉及到次生物质如黄酮、萜类等与‘黑油椿’风味形成相关的物质代谢。太和香椿风味的形成与其萜类组成和含量密切相关, 风味物质主要构成为单萜和倍半萜类物质 (刘常金等, 2013)。为此本研究进一步挖掘‘黑油椿’单萜和倍半萜类物质的生物合成途径相关酶基因, 结果发现, 所有参与萜类合成酶基因在转录组数据中均有发现, 这为开展香椿风味物质形成的分子机理及相关基因表达调控的研究奠定了坚实的基础。结合黑油椿’嫩芽中检测的萜类物质的成分及含量分析发现, 基因转录本的数量与其含量存在一定的正相关性, 倍半萜合成酶单基因簇数明显高于单萜合成酶。与‘黑油椿’嫩芽风味成分直接相关的倍半萜合成酶有大根香叶烯 D 合成酶、蛇床烯合成酶、 α -法尼烯合成酶和 β -石竹烯合成酶, 单萜合成酶有 (3S)-芳樟醇合成酶、月桂烯/罗勒烯合成酶和 (R)-柠檬烯合成酶, 研究这些关键酶基因的表达调控对深入理解香椿风味形成的分子机制具有重要意义。

本研究首次利用高通量转录组测序技术建立了香椿优良品种‘黑油椿’嫩芽的转录组数据库, 获得了大量的转录本信息并进行了功能注释, 揭示了‘黑油椿’嫩芽转录组的整体表达特征, 为后续开展香椿基因组学、基因功能研究及分子标记辅助遗传育种的研究奠定了基础。

References

- Arjan G, Matias M I, Johan M, Robert H, Robert V. 2000. Molecular cloning and analysis of strictosidine β -D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Journal of Biological Chemistry*, 275 (5): 3051 – 3056.
- Audic S, Claverie J M. 1997. The significance of digital gene expression profiles. *Genome Research*, 7: 986 – 995.
- Hsu C, Huang P, Chen C, Mao C, Chaw S. 2012. Tangy scent in *Toona sinensis* (Meliaceae) leaflets: isolation, functional characterization, and regulation of *TsTPS1* and *TsTPS2*, two key terpene synthase genes in the biosynthesis of the scent compound. *Current Pharmaceutical Biotechnology*, 13: 1 – 12.
- Jia Xin-ping, Sun Xiao-bo, Deng Yan-ming, Liang Li-jian, Ye Xiao-qing. 2014. Sequencing and analysis of the transcriptome of *Asplenium nidus*. *Acta Horticulturae Sinica*, 41 (11): 2329 – 2341. (in Chinese)
- 贾新平, 孙晓波, 邓衍明, 梁丽建, 叶晓青. 2014. 鸟巢蕨转录组高通量测序及分析. *园艺学报*, 41 (11): 2329 – 2341.
- Li Ju-ying, Wang Jun, Dai Yun-qing, Su Chun-yuan, Chen Min. 2011. Composition of characteristic aroma of *Toona sinensis* and its change during storage. *Journal of Beijing Forestry University*, 33 (3): 127 – 131. (in Chinese)
- 李聚英, 王 军, 戴蕴青, 苏春元, 陈 敏. 2011. 香椿特征香气组成及其在贮藏中变化的研究. *北京林业大学学报*, 33 (3): 127 – 131.
- Liu Chang-jin, Zhang Jie, Zhou Zheng-yan, Xie Yan-hui, Wan Hong-ying, Liu Sheng-bin. 2013. Comparative analysis of volatile components in three cultivars of Chinese toon (*Toona sinensis*) by GC-MS. *Food Science*, 34 (20): 261 – 267. (in Chinese)
- 刘常金, 张 杰, 周争艳, 谢艳辉, 宛红颖, 刘胜斌. 2013. GC-MS 分析比较 3 个特产香椿品种的挥发性成分. *食品科学*, 34 (20): 261 – 267.
- Liu Jun, Chen Yi-tai, Jiang Jing-min, He Gui-ping, Yu Guo-min. 2009. Study on population genetic structure in *Toona ciliata* var. *pubescens* with SSR. *Forest Research*, 22 (1) : 37 – 41. (in Chinese)
- 刘 军, 陈益泰, 姜景民, 何贵平, 余国民. 2009. 毛红椿群体遗传结构的 SSR 分析. *林业科学研究*, 22 (1) : 37 – 41.
- Liu Jun, Chen Yi-tai, Sun Zong-xiu, Jiang Jing-min, He Gui-ping, Rao Long-bing, Wu Tian-lin. 2008. Spatial genetic structure of *Toona ciliata* var. *pubescens* populations in terms of spatial autocorrelation analysis. *Scientia Silvae Sinicae*, 44 (6): 45 – 52. (in Chinese)
- 刘 军, 陈益泰, 孙宗修, 姜景民, 何贵平, 饶龙兵, 吴天林. 2008. 基于空间自相关分析研究毛红椿天然居群的空间遗传结构. *林业*

- 科学, 44 (6) : 45 - 52.
- Pei Song-yu, Ren Zhi-yong, Jiao Zhen-bin, Cheng Yue-qin, Wang Hong-wei. 2015. Construction and characterization of enriched microsatellite library from *Toonka sinensis*. Seed, 34 (2) : 13 - 16. (in Chinese)
- 裴宋雨, 任志勇, 焦振彬, 程月琴, 王红卫. 2015. 香椿 (*Toona sinensis*) 微卫星富集文库构建及特性分析. 种子, 34 (2): 13 - 16.
- Shi Chengying, Yang Hua, Wei Chaoling, Yu Oliver, Zhang Zhengzhu, Jiang Changjun, Sun Jun, Li Yeyun, Chen Qi, Xia Tao, Wan Xiaochun. 2011. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics, 12: 131.
- Sun Hai-yue, Liu Yu-shan, Gai Yu-zhuo, Geng Jin-man, Chen Li, Liu Hong-di, Kang Li-min, Tian You-wen, Li Ya-dong. 2015. De novo sequencing and analysis of the cranberry fruit transcriptome to identify putative genes involved in flavonoid biosynthesis, transport and regulation. BMC Genomics, 16: 652.
- Vranová E, Coman D, Grussem W. 2013. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annual Review of Plant Biology, 64: 665 - 700.
- Wu Ji-you, Cheng Yong, Wu Qi-jun, Chen Ming-gao, Li Yan, Huang Ming-jun, Wang Xu-jun, Liu Qiu. 2014. Study on cutting propagation of *Toona ciliate* clones. Agricultural Science & Technology, 15 (11) : 1844 - 1846.
- Xiao Hua-shan, Xiao Xiang-xi, He Wen-guang, Lin Xing-chun, Zheng Yan-ling. 2014. Research advances on rapid propagation and germplasm in *Toona sinensis*. Subtropical Plant Science, 43 (4): 343 - 346. (in Chinese)
- 肖华山, 肖祥希, 何文广, 林兴春, 郑燕玲. 2014. 香椿种质资源与快速繁殖研究进展. 亚热带植物科学, 43 (4): 343 - 346.
- Xu Qiang, Chen Ling-ling, Ruan Xiao-an, Chen Di-jun, Zhu An-dan, Chen Chun-li, Bertrand D, Jiao Wen-biao, Hao Bao-hai, Lyon MP, Chen Jiong-jiong, Gao Song, Xing Feng, Lan Hong, Chang Ji-wei, Ge Xian-hong, Lei Yang, Hu Qun, Miao Yin, Wang Lun, Xiao Shi-xin, Biswas MK, Zeng Wen-fang, Guo Fei, Cao Hong-bo, Yang Xiao-ming, Xu Xi-wen, Cheng Yun-jiang, Xu Juan, Liu Ji-hong, Luo Jun-hong, Tang Zhong-hui, Guo Wen-wu, Kuang Han-hui, Zhang Hong-yu, Roose ML, Nagarajan N, Deng Xiu-xin, Ruan Yi-jun. 2013. The draft genome of sweet orange (*Citrus sinensis*). Nature Genetics, 45: 59 - 66.
- Zhang Xia, Song Zhen-qiao, Liu Tian, Guo Lin-lin, Li Xing-feng. 2016. De novo assembly and comparative transcriptome analysis provide insight into lysine biosynthesis in *Toona sinensis* Roem. International Journal of Genomics, 3: 1 - 9.

征 订

欢迎订阅 2018 年《保鲜与加工》

《保鲜与加工》为中文核心期刊, 中国科技核心期刊, RCCSE 中国核心学术期刊, 中国农业核心期刊, 中国北方优秀期刊, 中国学术期刊 (光盘版) 收录期刊, 美国《化学文摘》(CA) 收录期刊, 英国《国际农业与生物科学研究中心》(CABI) 收录期刊, 英国《食品科技文摘》(FSTA) 收录期刊。主管: 天津市农业科学院; 主办: 国家农产品保鲜工程技术研究中心 (天津)。

《保鲜与加工》杂志是我国农产品采后技术研究领域的中文核心期刊, 据中国知网的最新统计结果, 5 年复合影响因子为 1.340。本刊主要报道农产品保鲜与加工相关领域基础理论、新技术、新工艺、新设备、新材料的研究成果及国内外相关行业的动态与信息。主要设置专家论坛、保鲜研究、加工研究、检测分析、信息与物流、专题论述、食品安全、技术指南、行业资讯、科普沙龙、科技前沿、政策法规等栏目。适于科技人员、农业技术推广人员、相关企业管理和技术人员、大专院校师生及广大从事保鲜与加工技术研发领域的人士参阅。

欢迎在全国各地邮局 (所) 或本编辑部订阅, 欢迎广大读者踊跃投稿, 并诚邀刊登各类相关广告。

国际标准连续出版物号: ISSN 1009-6221, 国内统一连续出版物号: CN 12-1330/S

邮发代号: 6-146 双月刊, 逢单月 10 日出版, 单价 18 元, 全年 108 元。

通讯地址: 天津市西青区津静公路 17 公里处, 国家农产品保鲜工程技术研究中心 (天津) 《保鲜与加工》编辑部, 邮编: 300384

电话: 022-27948711, 联系邮箱: bxyjg@163.com, 投稿平台: www.bxyjg.com。