

张雅玲, 王小安, 潘少霖, 等. 基于 RNA-Seq 的辣木叶转录组研究 [J]. 福建农业学报, 2017, 32 (10): 1111-1117.
ZHANG Y-L, WANG X-A, PAN S-L, et al. Study on Transcriptome of *Moringa oleifera* Lam. Leaves Based on RNA-seq [J]. *Fujian Journal of Agricultural Sciences*, 2017, 32 (10): 1111-1117.

基于 RNA-Seq 的辣木叶转录组研究

张雅玲, 王小安, 潘少霖, 叶新福*

(福建省农业科学院果树研究所, 福建 福州 350013)

摘要: 为探讨辣木叶营养价值形成的遗传基础, 采用 RNA-seq 对不同发育时期的辣木叶混合样品进行转录组测序, 获得 44 056 449 个 reads, 共组装得到 75 452 条 unigene, 平均长度为 795.69 nt, 其中有 15 824 条 unigene 长度大于 1 000 nt。采用公共数据库进行同源比对, 共有 30 847 条 unigene 得到注释。其中, 有 18 277 条 unigenes 被注释到 GO 数据库的 52 个功能组, 9 689 条 unigene 在 COG 数据库中得到注释。有 10 130 条 unigene 被富集到 129 条 KEGG 途径, 包括碳代谢、氨基酸生物合成和氧化磷酸化途径。

关键词: 辣木; 转录组; 功能注释

中图分类号: S 792.99

文献标识码: A

文章编号: 1008-0384 (2017) 10-1111-07

Study on Transcriptome of *Moringa oleifera* Lam. Leaves Based on RNA-seq

ZHANG Ya-ling, WANG Xiao-an, PAN Shao-lin, YE Xin-fu*

(Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian 350013, China)

Abstract: To explore the genetic basis for the nutritive value of *Moringa oleifera* leaves, the leaves at different developmental stages were mixed and used for transcriptome sequencing in the present study. With RNA-seq, 44 056 449 reads were generated, then assembled into 75 452 unigenes with an average length of 795.69 nt. 15 824 of these unigenes were longer than 1 kb. 30 847 unigenes were annotated by a similarity search against public databases. Of these, 18 277 unigenes were assigned into 52 GO function groups and 9 689 unigenes were annotated by COG database. 10 130 unigenes were enriched into 129 KEGG pathways, including carbon metabolism, biosynthesis of amino acids and oxidative phosphorylation.

Key words: *Moringa oleifera*; transcriptome; function annotation

辣木 *Moringa oleifera* Lam. 是一种原产亚洲的热带树种, 在亚洲、非洲和美洲的热带和亚热带地区均有栽培^[1]。辣木各个部分均可食用或作为药物用于治疗多种炎症引发的疾病^[2], 其中辣木叶因具有很高的营养价值和药用价值而备受关注^[3-4]。辣木叶富含蛋白质、矿物质、β-胡萝卜素和维生素多种营养物质^[5]。辣木叶的蛋白质含量高于牛奶^[6], 并含有所有必需氨基酸^[7]。研究表明辣木叶的维生素 A 含量高于胡萝卜, 维生素 C 含量高于柑橘, 钙含量高于牛奶, 铁含量高于菠菜, 钾含量高于香蕉^[8]。6 勺辣木叶粉即可满足孕妇和哺乳期妇女对铁和钙日需求量^[7]。辣木叶的提取物具有抗

菌、抗炎症和抗癌等活性^[9-11], 是许多国家用于治疗多种疾病的传统药物^[4]。此外, 辣木叶还是优质的蛋白质饲料^[12], 具有十分广阔的开发利用前景。

目前, 关于辣木的研究主要集中在营养、功能成分鉴定、食品开发利用等方面^[6, 9, 13-14]。仅有少量关于辣木营养和药用品质形成分子基础的报道。Tian 等^[15]完成了辣木基因组的测序, 并在此基础上对辣木转录因子进行了预测与分析^[16]。Kondo 等^[17]从辣木叶中分离得到了抗坏血酸合成酶基因, 并对其表达模式进行了分析, 发现该基因在叶片中表达量较高。Deng 等^[18]筛选出 2 个用于辣木实时

收稿日期: 2017-07-20 初稿; 2017-08-23 修改稿

作者简介: 张雅玲 (1987-), 女, 硕士, 研究实习员, 研究方向: 果树生物技术与遗传育种 (E-mail: 451476856@qq.com)

* 通讯作者: 叶新福 (1967-), 男, 教授, 研究员, 研究方向: 作物品质遗传育种研究 (E-mail: yexinfu@126.com)

基金项目: 福建省科技计划项目——省属公益类科研院所基本科研专项 (2016R1013-8)

荧光定量 PCR 分析的内参基因, 并进一步筛选了辣木黄酮类化合物、萜类化合物和脂肪酸合成相关基因^[19]。但辣木大量功能基因的注释亟待完善。本研究对不同发育阶段辣木叶片混合样品进行转录组测序和组装, 并对组装得到的 unigene 进行功能注释, 以期对辣木叶营养和药用品质相关功能基因挖掘及其分子机制解析奠定基础。

1 材料与方 法

1.1 试验材料

本试验所用材料为 3 年生印度辣木 PKM2 不同发育阶段的叶片 (展叶期后 0、10、20、30 d), 样品取自福建省农业科学院果树研究所辣木资源圃。于同一棵树上摘取生长良好无病虫害的辣木叶片, 采下后去除叶柄, 迅速用液氮速冻后保存于 -80℃ 冰箱。

1.2 RNA 提取、文库构建和 RNA-seq

采用 EZNA Plant RNA Kit (Omega Bio-tek) 进行不同阶段辣木叶片总 RNA 的提取。采用 Qubit 2.0 Fluorometer (Invitrogen, Life Technologies, CA, USA) 和 Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) 进行 RNA 浓度的测定和质量评价。将 4 个阶段辣木叶片的 RNA 等量混合后, 用于测序文库构建。辣木叶片 RNA 提取、文库构建和 RNA-seq 由北京百迈客生物科技有限公司完成。cDNA 文库的构建参照 Fang 等^[20]的方法。采用 Illumina HiSeq 4000 高通量测序平台对 cDNA 文库进行测序。测序原始数据已上传至 NCBI, 登录号为 SRR5765086。

1.3 转录组的组装与表达分析

对 Raw reads 进行数据过滤, 去除其中的接头序列及带有 N 或 10% 的碱基 Q 值 ≤ 20 的低质量 reads, 获得高质量的 Clean Data。获得高质量的测序数据之后, 采用 Trinity 法对其进行序列组装^[21]。所有参数均为默认参数。首先将测序 reads 打断为较短的片段 (K-mer), 然后将这些小片段延伸成较长的片段 (Contig), 并利用这些片段之间的重叠, 得到片段集合, 最后利用 De Bruijn 图的方法和测序 Read 信息, 在各个片段集合中分别识别转录本序列。采用 Bowtie 将各样品测序得到的 reads 与 unigene 库进行比对^[22], 根据比对结果, 结合 RSEM 进行表达量水平估计^[23]。利用 FPKM 值表示对应 unigene 的表达丰度。

1.4 功能注释和分类

采用 BLASTx 进行辣木叶片 unigene 序列的注释 (E 值 $< 10^{-5}$), 所用的数据库为 Nr、UniProt/Swiss-Prot、GO、COG、KEGG、KOG、Pfam 和 eggNOG。采用 Blast2GO^[24] 对 unigene 进行 GO 注释 (E 值 $\leq 10^{-5}$)。GO 功能分类结果采用 WEGO^[25] 进行作图。

2 结果与分析

2.1 辣木叶 RNA-Seq 测序与组装

为了获得尽可能丰富的辣木叶转录组信息, 选取展叶期及展叶期后 10、20 和 30 d 的辣木叶片, 分别提取不同时期辣木叶片的总 RNA, 并等量混合, 用于构建文库和 RNA-Seq 测序。HiSeq⁴⁰⁰⁰ 测序和数据过滤后, 共获得 44 056 449 个 reads, 总长度为 13 082 788 218 (13.08 Gb) 个碱基, GC 含量为 46.32%。碱基质量值 Q30 为 93.95%。

采用 Trinity 软件进行辣木叶转录组组装, 共得到 223 983 条转录本, 平均长度为 1 957.79 nt (表 1)。共得到 75 452 条 unigene, 总长度 60 036 488 nt, 平均长度为 795.69 nt, N50 长度为 1 543 nt, 组装完整性较高 (表 1)。其中, 长度为 200~500 nt 的 unigene 有 45 964 条 (60.92%), 长度为 500~1 000 nt 的 unigene 有 13 664 条 (18.11%), 长度大于 1 000 nt 的 unigene 有 15 824 条 (20.98%)。

表 1 辣木叶转录组组装结果

Table 1 Summary of sequencing and *de novo* assembly of *Moringa oleifera* Lam. leaves

长度区间/nt	转录本/条	unigenes/条
200~300	30299(13.53%)	25994(34.45%)
300~500	26316(11.75%)	19970(26.47%)
500~1000	28954(12.93%)	13664(18.11%)
1000~2000	46764(20.88%)	8266(10.96%)
2000+	91650(40.92%)	7558(10.02%)
总数量	223983	75452
总长度	438510672	60036488
N50 长度	3096	1543
平均长度	1957.79	795.69

注: 括号内数据为各长度区间对应的转录本、unigenes 组装比例。

2.2 基因功能注释

采用 Nr、UniProt/Swiss-Prot、GO、COG、

KEGG, KOG, Pfam 和 eggNOG 等 8 个数据库对辣木叶转录组的 75 452 条 unigene 进行注释(E 值 $< 10^{-5}$) (表 2)。结果表明, 有 40.88% 的 unigene (30 847 条) 得到注释, 剩余的 unigene (59.22%) 未能得到注释 (表 2)。长度大于 1 000 nt 的基因中仅有 14.24% (2 253 条) 未得到注释, 而小于 300 nt 的 unigene 中有 72.60% 未得到注释。Nr 数据库注释结果的 E 值分布结果表明, 得到注释的序列中有 54.56% (E 值 $< 10^{-60}$) 的序列与数据库中的序列具有很高的同源性, 有 33.94% ($10^{-60} < E$ 值 $< 10^{-15}$) 的序列与数据库中的序列具有较高的同源性。Nr 数据库注释结果的物种分布见图 1。辣木叶 unigene 与可可 (19.85%) 基因的匹配度最高, 其次为葡萄 (9.20%) 和甜橙 (5.42%)。

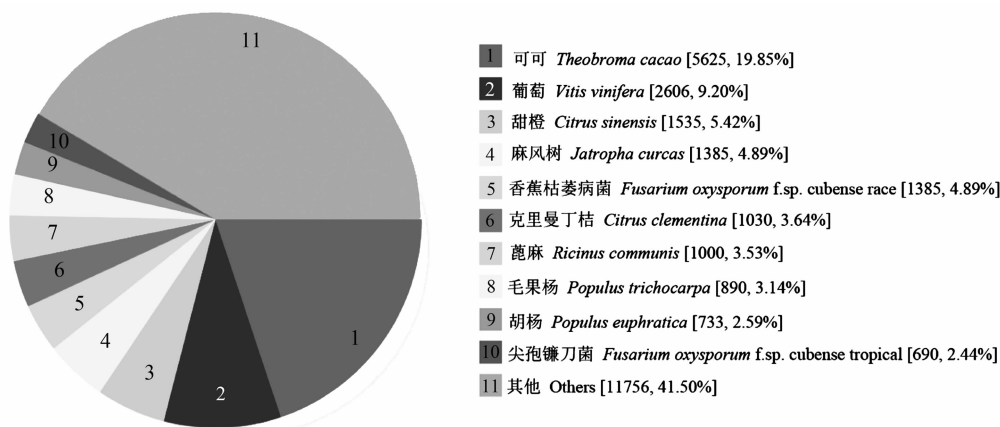


图 1 Nr 数据库 BLAST 比对结果的物种分布

Fig. 1 Species distribution of the BLAST search results in Nr database

将辣木叶转录组 unigene 与 GO 数据库进行比对, 有 18 277 条 unigenes 获得至少 1 条 GO 注释 (表 2)。这些 unigenes 被分为生物学过程、细胞组分和分子功能 3 大类和 52 个亚类 (表 3)。细胞组分分类中细胞、细胞部分和细胞器 3 类基因最多, 分别为 8 160、8 160 和 6 046 条。分子功能类中数量最多的为催化活性和结合分类, 分别为 9 878 和 9 483 条。生物学过程类中比例最高的为代谢过程和细胞过程, 分别为 12 587 和 10 821 条。

将辣木叶转录组 unigene 与 COG 数据库进行比对, 共有 9 689 条 unigene 得到注释, 占总 unigene 的 12.84% (表 2)。COG 数据库注释结果如图 2 所示, 根据功能共分为 25 类。数量最多的为“功能预测”, 共有 2 524 条, 占总数的

表 2 辣木叶 unigene 注释结果
Table 2 Summary of functional annotations for the assembled unigenes of *Moringa oleifera* Lam. leaves

数据库	注释基因数量/条	注释基因比例/%	300 nt ≤ 长度 < 1000 nt 注释基因数量/条	长度 ≥ 1000 nt 注释基因数量/条
Nr	28336	37.56	9441	13533
Swiss-Prot	16454	21.81	4690	9638
GO	18277	24.22	5936	9029
COG	9689	12.84	2573	5325
KEGG	10130	13.43	3146	5288
KOG	16850	22.33	5050	8305
Pfam	19541	25.90	5323	11339
eggNOG	28948	38.37	9255	13349
All	30847	40.88	10153	13571

26.05%。其次为“复制、重组和修复”、“转录”、“翻译、核糖体结构和生物合成”、“信号转导机制”和“翻译后修饰、蛋白折叠和伴侣蛋白”, 分别有 1 205、1 195、1 030、1 029 和 958 条。“核结构”的数量最少, 仅有 4 条。此外, 功能未知的基因有 286 条, 占总数的 2.95%。

为了更好地了解参与辣木叶生长发育的生物学途径, 采用 KEGG 数据库进行辣木叶转录组 unigene 注释。10 130 条 (13.14%) unigenes 与 KEGG 数据库中的序列匹配度高, 这些基因分别属于 129 条 KEGG 途径 (表 4)。其中, 3 027 条属于代谢途径。代谢途径中碳代谢、氨基酸生物合成和氧化磷酸化途径 unigene 数量最多, 分别为 458、399 和 285 条。参与次生代谢物合成的 unigene 共有 407 条。

表 3 辣木叶转录组 unigene 的 GO 功能分类
Table 3 GO classifications of unigenes for transcriptome of *Moringa oleifera* Lam. leaves

生物学过程		细胞组分		分子功能	
基因功能	基因数量/个	基因功能	基因数量/个	基因功能	基因数量/个
生殖	265	胞外区	512	蛋白质结合转录因子活性	40
免疫系统过程	254	细胞	8160	核酸结合转录因子活性	500
代谢过程	12587	类核	48	催化活性	9878
细胞过程	10821	膜	4251	受体活性	89
生殖过程	856	病毒体	5	鸟嘌呤核苷酸交换因子活性	46
生物附着	36	细胞连接	293	结构分子活性	650
信号传递	932	胞外基质	7	转运体活性	1296
多细胞组织进程	1495	膜腔	535	结合	9483
发育过程	1602	大分子复合物	2218	电子载体活性	364
生长	306	胞外区	6046	抗氧化活性	145
移动	10	胞外基质组分	1	金属伴侣蛋白活性	4
单组织过程	8955	胞外区域部分	8	酶调节活性	189
生物相	33	细胞器部分	2523	蛋白标签	1
韵律过程	37	病毒体部分	5	翻译调控活性	4
刺激应答	3287	膜部分	2187	营养库活性	20
定位	2855	细胞部分	8160	分子转换器活性	244
多生物过程	759	共质体	292		
生物条件	3263				
生物组分组织或生物合成	2133				

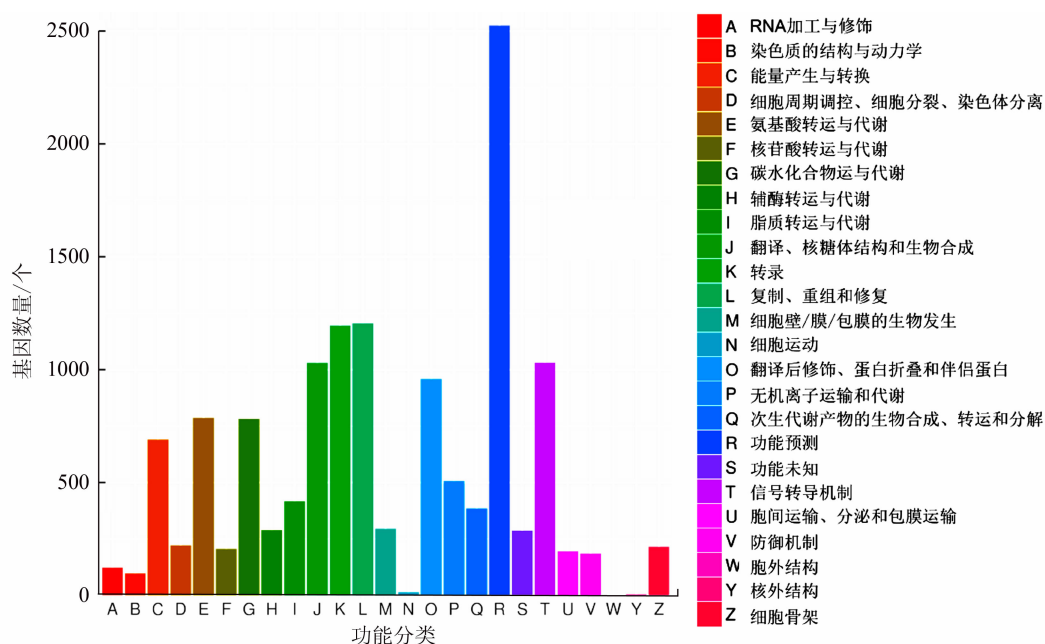


图 2 辣木叶转录组 unigene 的 COG 功能分类
Fig. 2 COG classifications of unigenes for transcriptome of *Moringa oleifera* Lam. leaves

表 4 辣木叶转录组 unigene 的 KEGG 功能分类

Table 4 KEGG classifications of unigenes for transcriptome of *Moringa oleifera* Lam. leaves

代谢通路	基因数量 /个	代谢通路	基因数量 /个	代谢通路	基因数量 /个
核糖体	517	磷酸肌醇代谢	77	缬氨酸,亮氨酸和异亮氨酸代谢	34
碳代谢	458	磷脂酰肌醇信号系统	77	叶酸-碳单位循环代谢通路	32
氨基酸生物合成	399	核苷酸切除修复	67	类黄酮生物合成	30
内质网蛋白加工	331	苯丙氨酸,酪氨酸和色氨酸生物合成	64	硒化合物代谢	29
氧化磷酸化	285	β -丙氨酸代谢	63	其他多糖降解	29
剪接体	266	抗坏血酸代谢	62	组氨酸代谢	28
淀粉和蔗糖代谢	255	脂肪酸降解	62	磷酸肌醇生物合成	25
RNA 转运	250	萜类骨架生物合成	62	高脚类,二烯类和姜素生物合成	24
糖酵解和糖异生	216	氰基氨基酸代谢	61	花生四烯酸代谢	23
嘌呤代谢	211	ABC 转运体	61	硫酸代谢	22
植物激素信号转导	206	RNA 聚合酶	61	核黄素代谢	21
内吞作用	197	酪氨酸代谢	60	二萜生物合成	21
mRNA 监控通路	195	同源重组	60	柠檬烯、蒎烯降解	20
氨基糖和核苷酸糖代谢	183	脂肪酸生物合成	57	牛磺酸和亚牛磺酸代谢	18
植物病原菌互作	183	丙酸代谢	57	烟酸和烟酰胺代谢	18
吞噬体	176	蛋白质输出	57	生物素代谢	18
泛素介导的蛋白水解	174	色氨酸代谢	55	叶酸生物合成	18
RNA 降解	162	DNA 复制	55	硫中继系统	18
丙酮酸代谢	160	卟啉和叶绿素代谢	53	维生素 B6 代谢	16
嘧啶代谢	153	囊泡运输中的陷阱相互作用	52	赖氨酸生物合成	15
苯丙素生物合成	153	不饱和脂肪酸生物合成	51	光合作用-天线蛋白	14
半胱氨酸和蛋氨酸代谢	152	基底转录因子	50	糖胺聚糖降解	14
柠檬酸循环	139	N-多糖生物合成	48	亚油酸代谢	14
光合生物中的碳固定	129	氮代谢	48	玉米素生物合成	14
苯丙氨酸代谢	125	赖氨酸降解	47	酮体合成和降解	13
真核生物核糖体合成	123	α -亚麻酸代谢	47	糖鞘脂类生物合成-globo 系列	12
二羧酸代谢	120	碱基切除修复	47	C5 分支的二元酸代谢	12
精氨酸和脯氨酸代谢	117	硫代谢	46	芳香化合物降解	11
2-氧代羧酸代谢	116	光合作用	44	倍半萜和三萜类化合物生物合成	10
脂肪酸代谢	115	丁酸代谢	44	芥子油苷生物合成	10
甘油磷脂代谢	107	错配修复	44	油菜素内酯生物合成	9
丙氨酸,阿斯帕特酸和谷氨酸酯代谢	104	脂肪酸延伸	42	非同源性末端接合	9
蛋白酶体	104	自噬调节	41	其他类型 α -聚糖生物合成	8
戊糖、葡萄糖醛酸转换	103	类固醇生物合成	40	单萜类生物合成	8
氨酰-tRNA 生物合成	103	泛醌和其他萜类生物合成	40	鞘糖脂生物合成	6
过氧化物酶体	103	醚脂类代谢	40	咖啡因代谢	5
谷胱甘肽代谢	96	嘧啶和吡啶生物碱生物合成	40	花色苷生物合成	4
甘氨酸,丝氨酸和三氨酸代谢	92	植物昼夜节律	40	硫辛酸代谢	3
磷酸戊糖途径	91	泛酸盐和辅酶 a 生物合成	38	异黄酮生物合成	3
果糖和甘露糖代谢	91	异喹啉生物碱生物合成	37	万古霉素抗药性	3
半乳糖代谢	89	角质,亚氨基酸和蜡生物合成	36	黄酮和黄酮醇生物合成	2
缬氨酸,亮氨酸和异亮氨酸降解	81	鞘脂类代谢	36	碳青霉烯生物合成	1
甘油酯代谢	77	类胡萝卜素生物合成	36	酮化合物糖单位生物合成	1

2.3 基因表达量分析

对 unigene 的表达量进行分析,发现 unigene 的 FPKM 值在 0~15 602.35,其中 FPKM 值大于 1 000 的 unigene 有 43 条,FPKM 值介于 100~1 000 的 unigene 有 779 条,FPKM 值介于 10~100 的 unigene 有 8 977 条,FPKM 值介于 1~10 的 unigene 有 10 762 条。表达量最高的 10 个基因如表 5 所示,其中有 7 个基因(二磷酸核酮糖羧化酶小链、叶绿体叶绿素 a/b 结合蛋白、二磷酸核酮糖羧化酶/加氧酶活化酶和铁氧还蛋白)与光合作用有关。根据 KEGG 注释结果,核糖体相关 unigene 的 FPKM 值在 0~647.44,其中有 214 个 unigene 的 FPKM 值大于 10,FPKM 值大于 100 的 unigene 有 118 条,FPKM 值最高的为 50S 核糖体蛋白 L12、40 核糖体蛋白 S4、60S 酸性核糖体蛋白 P0、50S 核糖体蛋白 L1 和 40S 核糖体蛋白 S14-3。参与氨基酸生物合成途径的 unigene 中有 167 条 FPKM 值大于 10,其中 FPKM 值大于 100 的 unigene 有 27 条,FPKM 值最高的依次为果糖-二磷酸醛缩酶、转酮醇酶和磷酸甘油酸激酶基因。参与内质网蛋白加工的 unigene 中有 141 条 FPKM 值大于 10。

表 5 辣木叶转录组中表达量最高的 10 条基因

Table 5 The 10 most abundant unigenes in transcriptome of *Moringa oleifera* Lam. leaves

编号	基因名称	FPKM 值
c18407.graph_c0	二磷酸核酮糖羧化酶小链	15602.35
c30652.graph_c0	叶绿体叶绿素 a/b 结合蛋白	12537.82
c35268.graph_c0	二磷酸核酮糖羧化酶小链	8535.69
c33703.graph_c0	类 Win 蛋白	5152.12
c8930.graph_c0	二磷酸核酮糖羧化酶/加氧酶活化酶	3702.24
c23488.graph_c0	泛素蛋白	2702.34
c30638.graph_c1	氨基环丙烷羧酸氧化酶	2458.66
c35274.graph_c0	叶绿体叶绿素 a/b 结合蛋白	2268.96
c8931.graph_c0	铁氧还蛋白	2057.79
c35272.graph_c0	叶绿体叶绿素 a/b 结合蛋白	2011.20

3 讨论与结论

RNA-seq 具有高通量和成本低等诸多优点,已广泛应用于植物功能基因组学研究^[26-31]。本研究采用 Illumina HiSeq 4000 对不同发育时期辣木叶的混合样品进行转录组测序,共组装得到 75 452

条 unigene,为辣木功能基因挖掘提供了大量的基础数据。邓丽婷^[19]对成熟辣木叶进行转录组测序仅组装得到 45 052 条 unigene,获得基因信息量明显低于本研究。这可能与本研究采用混合样品进行测序有关。王兴春等^[32]也认为采用混合样品进行测序有利于获得更丰富的基因信息。但不排除受测序平台等因素影响。

采用 Nr、GO、COG 和 KEGG 等数据库对辣木叶转录组的 unigene 进行注释,仅有 40.88% 的 unigene 获得注释信息。这可能与公共数据库中缺乏辣木及其近缘物种的基因组信息有关。此外,本研究中有 72.60% 小片段 unigene (长度小于 300 nt) 未得到注释,而长度在 1 000 nt 以上的 unigene 注释率达到 85.76%。可见,大量短片段序列的存在也是导致注释率较低的重要原因。KEGG 代谢途径分析结果表明,辣木叶中有大量核糖体、氨基酸生物合成和内质网蛋白加工等与蛋白质和氨基酸合成相关基因的表达。同时,COG 数据库注释结果也表明注释的基因中有大量与蛋白质合成相关的基因(“翻译、核糖体结构和生物合成”和“信号转导机制”和“翻译后修饰、蛋白折叠和伴侣蛋白”)。这些基因的表达为辣木叶高蛋白质和氨基酸含量奠定了基础。除了含有丰富的蛋白质,辣木叶中还含有大量具有生物活性的次生代谢物^[5]。次生代谢物的生物合成受大量基因的调控,合成与代谢途径十分复杂。克隆和鉴定次生代谢物合成和调控基因,解析次生代谢途径调控机制,是通过生物工程手段生产次生代谢物的前提。本研究采用公共数据库共注释了 30 847 条 unigene。其中,407 条基因参与次生代谢物的合成。这些次生代谢物合成相关基因为解析次生代谢途径及其调控机制提供了丰富的基础数据。

本研究建立了不同发育阶段辣木叶的转录组数据库,组装得到 75 452 条 unigene,并进行功能注释,共有 30 847 条 unigene 得到注释。其中,分别有 18 277、9 689 和 129 条在 GO 数据库、COG 数据库和 KEGG 数据库中得到注释。结果表明辣木叶中有大量氨基酸、蛋白质合成和次生代谢产物合成相关的基因表达,这些基因的表达为其高营养价值的形成奠定了物质基础。本研究的结果将为辣木叶营养与药用品质形成相关功能基因挖掘和分子标记开发提供参考,为辣木分子生物学研究与分子育种奠定基础。

参考文献:

[1] 党选民,曹振木,杨葵,等.热带珍稀植物辣木的特性及其开

- 发利用 [J]. 中国农学通报, 2004, 20 (20): 133—136.
- [2] ANWAR F, LATIF S, ASHRAF M, et al. *Moringa oleifera*: a food plant with multiple medicinal uses [J]. Phytotherapy Research, 2007, 21 (1): 17—25.
- [3] MISHRA S, SINGH P, SINGH S. Nutritional and medicinal value of *Moringa oleifera* leaves: potential and prospects [J]. Forestry Bulletin, 2011, 11 (1): 46—58.
- [4] SUJATHA B K, POONAM P. *Moringa Oleifera*-Nature's Gold [J]. Imperial Journal of Interdisciplinary Research, 2017, 3 (5): 1175—1179.
- [5] MOYO B, MASIKA P, HUGO A, et al. Nutritional characterization of *Moringa (Moringa oleifera Lam.)* leaves [J]. African Journal of Biotechnology, 2011, 10 (60): 12925—12933.
- [6] 刘凤霞, 王苗苗, 赵有为, 等. 辣木中功能性成分提取及产品开 展的研究进展 [J]. 食品科学, 2015, 36 (19): 282—286.
- [7] MISHRA S P, SINGH P, SINGH S. Processing of *Moringa oleifera* leaves for human consumption [J]. Bulletin of Environment, Pharmacology and Life Sciences, 2012, 2 (1): 28—31.
- [8] FAHEY J W. *Moringa oleifera*: a review of the medical evidence for its nutritional, therapeutic, and prophylactic properties. part 1 [J]. Trees for life Journal, 2005, (1): 5.
- [9] 彭磊, 田洋, 解静, 等. 世界辣木发展现状及市场前景分析 [J]. 世界农业, 2015, (9): 143—146.
- [10] WATERMAN C, CHENG D M, ROJAS-SILVA P, et al. Stable, water extractable isothiocyanates from *Moringa oleifera* leaves attenuate inflammation in vitro [J]. Phytochemistry, 2014, 103 (Supplement C): 114—122.
- [11] JUNG I L. Soluble extract from *Moringa oleifera* leaves with a new anticancer activity [J]. PLoS ONE, 2014, 9 (4): e95492.
- [12] RICHTER N, SIDDHURAJU P, BECKER K. Evaluation of nutritional quality of moringa (*Moringa oleifera Lam.*) leaves as an alternative protein source for Nile tilapia (*Oreochromis niloticus L.*) [J]. Aquaculture, 2003, 217 (1): 599—611.
- [13] 刘子记, 孙继华, 刘昭华, 等. 特色植物辣木的应用价值及发展前景分析 [J]. 热带作物学报, 2014, 35(9): 1871—1878.
- [14] 吴頔, 蔡志华, 魏焯昕, 等. 辣木作为新型植物性蛋白质饲料的研究进展 [J]. 动物营养学报, 2013, 25 (3): 503—511.
- [15] TIAN Y, ZENG Y, ZHANG J, et al. High quality reference genome of drumstick tree (*Moringa oleifera Lam.*), a potential perennial crop [J]. Science China Life Sciences, 2015, 58 (7): 627—638.
- [16] 田洋, 董扬, 盛军, 等. 辣木转录因子家族初步鉴定与分析 [J]. 云南农业大学学报: 自然科学版, 2016, 31(6): 450—461.
- [17] KONDO T, FUJIKAWA Y, UEDA A, et al. Cloning and gene expression analysis of ascorbic acid biosynthesis enzymes in *Moringa oleifera* [J]. African Journal of Agricultural Research, 2015, 10 (22): 2274—2285.
- [18] DENG L T, WU Y L, LI J C, et al. Screening reliable reference genes for RT-qPCR analysis of gene expression in *Moringa oleifera* [J]. PLoS ONE, 2016, 11 (8): e0159458.
- [19] 邓丽婷. 基于转录组测序的辣木内参基因筛选和三个代谢途径相关基因的表达分析 [D]. 广州: 华南农业大学, 2016: 48—49.
- [20] FANG Z Z, ZHOU D R, YE X F, et al. Identification of candidate anthocyanin-related genes by transcriptomic analysis of 'Furongli' plum (*Prunus salicina Lindl.*) during fruit ripening using RNA-Seq [J]. Frontiers in Plant Science, 2016, 7 (166): 1338.
- [21] GRABHERR M G, HAAS B J, YASSOUR M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. Nature Biotechnology, 2011, 29 (7): 644—652.
- [22] LANGMEAD B, TRAPNELL C, POP M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome [J]. Genome Biology, 2009, 10 (3): 1—10.
- [23] LI B, DEWEY C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [J]. BMC Bioinformatics, 2011, 12 (1): 323.
- [24] ONESA A, GÖTZ S, GARCÍA-GÓMEZ J M, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research [J]. Bioinformatics, 2005, 21 (18): 3674—3676.
- [25] YE J, FANG L, ZHENG H, et al. WEGO: a web tool for plotting GO annotations [J]. Nucleic Acids Res, 2006, 34 (S2): W293—W297.
- [26] 张少平, 邱珊莲, 郑云云, 等. 紫色黄秋葵转录组功能基因测序及分析 [J]. 核农学报, 2017, 31 (4): 643—653.
- [27] 向君亮, 黄殷. 葱蒜幼苗叶片转录组分析 [J]. 农业生物技术学报, 2017, 25 (3): 386—396.
- [28] 武红霞, 许文天, 罗纯, 等. 芒果果实转录组数据组装及基因功能注释简 [J]. 热带作物学报, 2016, 37 (11): 2191—2198.
- [29] 蔡年辉, 邓丽丽, 许玉兰, 等. 基于高通量测序的云南松转录组分析 [J]. 植物研究, 2016, 36 (1): 75—83.
- [30] 姜福星, 杨丽娟, 高素萍, 等. 碰碰香叶片的转录组测序分析 [J]. 分子植物育种, 2015, 13 (10): 2310—2319.
- [31] 吴超, 彭娟, 向林, 等. 基于高通量测序的铁皮石斛叶片转录组分析 [J]. 分子植物育种, 2012, 14 (12): 3334—3346.
- [32] 王兴春, 谭河林, 陈钊, 等. 基于 RNA-Seq 技术的连翘转录组组装与分析及 SSR 分子标记的开发 [J]. 中国科学: 生命科学, 2015, 45 (3): 301—310.

(责任编辑: 张 梅)