第 50 卷 第 9 期
2014 年 9 月

林 业 科 学
SCIENTIA SILVAE SINICAE

Vol. 50, No. 9
Sep., 2014

doi:10.11707/j.1001-7488.20140905

油茶花发育转录组测序及相关基因表达分析*

胡玉玲 姚小华 任华东 王开良 林 萍

(中国林业科学研究院亚热带林业研究所 富阳 311400)

摘 要: 通过对油茶成花过程的转录组测序及其成花相关基因的表达分析,结果表明: 转录组测序总共获得28 448 847个 reads,5 742 023 480 bp 数据量,GC 含量为 46.52%; 拼接成大于 200 bp 以上的 Unigenes 有 94 476 条,N50 长度为 806 bp,其中1 kbp以上的 Unigenes 共 12 643 条,占 Unigene 总数的 13.38%; Unigenes 在各数据库中功能注释数目,在 COG 中有 9 095 条,在 GO 中有 27 201 条,在 KEGG 中有 6 431 条,在 Swissprot 中有 24 534 条,在 TrEMBL 中有 36 393 条,在 Nr 中有 36 400 条,在 Nt 中有 30 858 条; 茎尖中 *FLC*,*FCA* 和 *FT* 基因表达量较 *AP1*,*AP2* 和 *PI* 基因低,*FT* 基因(ID: Unigene60063)是油茶成花的关键基因,*PI* 基因(ID: Unigene56059)与雄蕊发育关系紧密。

关键词: 油茶; 转录组; 高通量测序; 成花基因; 成花途径

中图分类号: S718.46 **文献标识码**: A **文章编号**: 1001-7488(2014)09-0036-08

Sequencing of Transcriptome Relevant to Flowering and Analysis of Floral-Related Genes Expression in *Camellia oleifera*

Hu Yulin Yao Xiaohua Ren Huadong Wang Kailiang Lin Ping

(*Research Institute of Subtropical Forestry, Chinese Academy of Forestry Fuyang* 311400)

Abstract: In this study, total RNA of stem apex and bud in *Camellia oleifera* was sequenced and its flowering relation genes were analysis. The results showed that transcriptome sequencing revealed a total of 28 448 847 reads, 5 742 023 480 bp data, and 46.52% GC content in whole genome. The Unigenes numbers with more than 200 bp were 94 476, and the Unigenes were very similar with grape' s. The N50 length was 806 bp, and 1 kbp or longer Unigenes were 12 643, accounting for 13.38% of total Unigenes database. The annotation number of Unigenes function in each database was : 9 095 Unigenes in COG, 27 201 Unigenes in GO, 6 431 Unigenes in KEGG, 24 534 Unigenes in Swissprot, 36 393 Unigenes in TrEMBL, 36 400 Unigenes in Nr, and 30 858 Unigenes in Nt. The expression profile showed that *FLC*, *FCA*, *FT* expression were lower than *AP1*, *AP2*, *PI* expression in the stem apex. The *FT* gene (ID: Unigene60063) may be a key gene for *Camellia* flower, and *PI* genes (ID: Unigene56059) are closely related with the stamen development.

Key words: *Camellia oleifera*; transcriptome; high-throughout sequencing; flowering genes; flowering way

山茶属(*Camellia*)油茶组(Sect. *Oleifera*)植物是我国重要的木本油料树种,茶油富含的不饱和脂肪酸对人体具有重要的保健功能(韩宁林等, 2009; 王文杰等, 2007),其中油茶(*Camellia oleifera*)是我国目前第一位的主栽物种,不管其栽培面积还是产量都居主导位置(庄瑞林, 2008)。油茶是典型的"抱子怀胎"植物,雌雄同花,花期一般在 10 月初到翌年 2 月,盛花期在 11 月中旬(查小华等, 2012)。春梢生长量非常大,一般春梢占三梢(春梢、夏梢和秋梢)98%以上(唐光旭, 1984)。油茶花开在当年生的春梢上,花量非常大,一支 7 张叶片的春梢上常常能开 30 朵以上,因此春梢是油茶的主要结果枝,与油茶产量密切相关(黎章矩等, 1992; 曾燕如等, 2009)。油茶是典型的常绿植物,春梢上的叶片一般有 15~16 个月的寿命,春梢一般 3 月中下旬开始萌发,5 月初开始花芽分化。油茶自交可育性低,如果花期遇到寒冷和阴雨天气会影响授粉受精,同时幼果树上过冬,不同品种及立地条件下的油茶产量

收稿日期: 2013-12-05; 修回日期: 2014-02-26。

基金项目: 国家科技支撑项目"油茶高产优质新种质创制"(2009BADB1B01); 油茶产业升级关键技术研究与示范(2009BADB1B00)。

* 姚小华为通讯作者。

差异非常大(林少韩等，1981；高本年，1981)。因此研究油茶花发育特点，并对花期及花器官形成进行有效调控，对解决油茶产量及提高油茶育种效率具有重要作用。目前对花发育调控方式主要有 2 种，即通过农事和转基因手段，转基因技术成功运用于油茶的花发育调控，其调控效果较农事手段明显。转基因成功的关键，除建立良好的遗传转化体系外，就是要清楚被调控生物的分子发育机制。

目前对油茶的成花机制研究鲜有报道，多数研究侧重于花芽形态和受精方面的研究(袁德义等，2011；王湘南等，2011)。全基因组测序是研究生物发育分子机制的重要内容和手段，但是油茶的染色体数为 $2n = 90$，其基因组非常庞大，基于目前的测序条件，其成本相当昂贵。从目前相关文献来看，侧重了解生物发育某一方面的基因表达，最有效的办法是通过新一代测序技术对油茶花发育过程的转录组进行高通量测序(Mortazavi *et al.*，2008)。目前油茶转录组测序虽然也取得一些进展(林萍等，2011；陈英等，2011)，但是侧重在花发育方面的研究却未见报道。本文对油茶花发育过程的 6 个阶段的转录本进行高通量测序，并探讨花发育相关基因 *FLC* (Flowering locus C)，*FCA* (Flowering time control)，*FT* (Flower locus T)，*AP1* (Apetala1)，*AP2* (Apetala2) 和 *PI* (Pistillata) 在花芽发育过程中的表达变化，以期为油茶成花机制研究提供基础数据。

1　材料与方法

1.1　试验材料

油茶长林 4 号(*Camellia oleifera* 'ChangLin 4')优良无性系。

1.2　采样方法

根据相关文献报道(庄瑞林，2008；袁德义等，2011；王湘南等，2011)，分别在花芽分化前(5 月 8 日)、萼片形成期(5 月 28 日)、花瓣形成期(6 月 15 日)、雌雄蕊形成期(7 月 3 日)、子房和花药形成期(7 月 25 日)及雌雄蕊成熟期(8 月 25 日)采集花芽用于总 RNA 分离及转录组测序；在春梢萌芽期(3 月 19 日)分别采集新芽和根尖，而后每隔 10 天取 1 次茎尖，最后 1 次为 5 月 28 日(花芽分化完成进入萼片形成期)。样品采集后置于液氮中带回，放在 −80 ℃的超低温冰箱中保存备用。

1.3　RNA 提取

RNA 提取采用艾德莱公司 EASYspin Plus 植物 RNA 快速提取试剂盒(目录号：RN38)，采用微量分光光度计 Nanodrop 2000 及琼脂糖凝胶电泳检测 RNA 质量，保证 RNA 样品浓度 400 $ng \cdot \mu L^{-1}$，28S∶18S大于 1.8。

1.4　转录组上机测序样品的制备

利用富含 ploy(T) 低吸附磁珠对 Total RNA 中的 mRNA 进行富集纯化处理，在高温条件下，利用二价阳离子打断 mRNA 以选取合适大小的目的片段，利用反转录酶和随机引物将打断后 mRNA 片段反转录形成 cDNA 第 1 条链，然后加入缓冲液、dNTPs，在 RNase H 和 DNA polymerase I 的作用下合成第 2 条 cDNA，对反转录合成的双链 cDNA 进行末端修复、3' 末端加 A、连接接头，通过琼脂糖凝胶电泳，筛选一定范围大小的片段。将前步处理好的样品进行精确定量(qubit)，在芯片(flow cell)表面进行桥式 PCR，使 DNA 片段扩增为单分子 DNA 簇(此过程在 Cluster Station 中进行)，单分子 DNA 簇形成后将 Flow cell 移入 Hi-Seq 中，进行 Illumina HiSeq™ 2000 测序。

1.5　转录组测序数据处理与分析

测序得到的原始数据去除只含有测序接头序列，或 N 含量过高以及过短的序列数据。利用 Trinity(http://trinityrnaseq.sourceforge.net) 软件通过序列之间的 Overlap 将序列延伸成 Contig，再根据序列的 Paired-end 信息，将 Contig 连接成转录本序列；对于 Trinity 组装结果，通过组装的 Component 从潜在的可变剪接转录本选取组装最长的转录本作为该样品的 Unigene 序列，最后将 Unigene 序列与 Nr，Nt，SwissProt，TrEMBL，GO，COG，KEGG 数据库比对，获得 Unigenes 的注释信息。如果不同库之间的比对结果有矛盾，则按 Nr，Swiss-Prot，KEGG 和 COG 的优先级确定 Unigene 的序列方向，比对不上的 Unigenes 则用软件 ESTScan 预测其编码区并确定序列的方向。

1.6　Real Time-PCR

根据相关文献及前期试验在转录组测序数据中，利用同源比对，筛选出 *FT*，*FCA*，*FLC*，*AP1*，*AP2* 和 *PI* 基因序列，利用 Primer 5 进行引物设计，引物序列见表 1。将 1.3 中提取的 RNA，通过 TAKARA 公司 PrimeScript® Reverse Transcriptase 试剂盒(商品编号：D2680A)方法(参考说明书)合成 cDNA 第 1 链，取其中 4 μL 合成产物稀释 10 倍，用于试验；采用 *CESA* 内参(Zhou *et al.*，2013)，根据 TAKARA 公司 SYBR® Premix Ex Taq™ (Tli RNaseh Plus)(商品编号为 DRR420S)说明书进行荧光定量 PCR 反应(ABI 7300)，每个反应重复 3 次。

表 1 引物序列

Tab. 1 Primer sequences

基因名称 Gene name	引物序列 Primer sequences	基因长度 Gene length/bp	退火温度 Annealing temperature(T_m)/℃
FT	FTF: 5′-GATGCTCCCAGCCCAAGTGAT-3′ FTR: 5′-CTCCGCAAAGTCCCGAGTGTT-3′	231	61.9
FCA	FCAF: 5′-TCATCAGAGGACCAAGACGACG-3′ FCAR: 5′-TGGTGTGGGCAGATGGATAGAG-3′	138	61.9
FLC	FLCF: 5′-GCAAATCTGGACTCTGCTTCAC-3′ FLCR: 5′-GAGGTGGATTGGTCAAGAAAGC-3′	205	60.1
AP1	AP1F: 5′-GGAGATGCGTGCTTGTTGAGGA-3′ AP1R: 5′-ACCCTAAGCCTCCCAACAGCAT-3′	247	61.9
AP2	AP2F: 5′-GCTGCTTCACCACTCTTACCTCT-3′ AP2R: 5′-CACTGAGTCTGAGGTGGAAGGTA-3′	202	62.0
PI	PIF: 5′-TGGAAAGATGCTGGAGGAGGATA-3′ PIR: 5′-AGGCTTCTCTCCCACACAAAACA-3′	255	60.2

2 结果与分析

2.1 转录组分析

2.1.1 转录组样品原始测序 reads 的数据结果与评估 将花芽发育的 6 个时期的 RNA 按照等量的方式混合成 1 个样，测序总片段数(total reads)为 28 448 847条，总碱基数为 5.74 Gbp，GC 含量为 46.52%。CycleQ 是测序质量评估的一个重要标准，其值越大测序误差越小，从图 1 可以看出本次测序 CycleQ20 percentage 为 100.00%，CycleQ30 percentage 绝大部分也接近 100.00%，说明测序结果符合要求，可以进行后续分析。

2.1.2 测序数据组装及结果统计 使用 Trinity 软件对样品数据进行组装，获得对应的 Transcripts(转录本)序列。Trinity 是一个专门对高通量 RNA-seq 数据进行全长转录本构建的软件，不同于其他的组装软件，最终构建出来的全长转录本不包含 Gap，是一个比较好的无基因组全长转录本的组装软件。N50 长度是衡量组装效果的一个重要指标，从表 2 中 Contig 和 Transcripts 的 N50 长度的数值大小可以看出，组装效果基本符合要求。

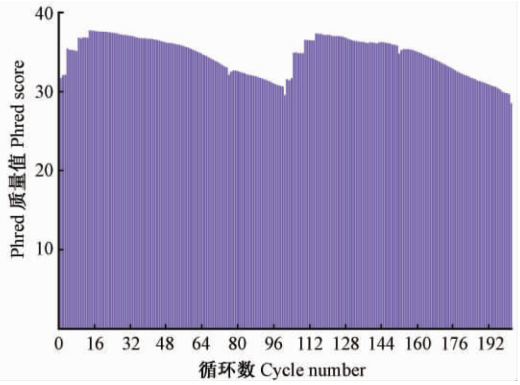

图 1 CycleQ 分布

Fig. 1 Distribution map of the CycleQ

表 2 测序数据组装和统计

Tab. 2 Sequencing data assembling and statistics

重叠群 Contigs			转录本 Transcripts		
重叠群长度 Contig length/bp	总数 Total number	占总数百分比 Percentage(%)	转录本长度 Transcript length/bp	总数 Total number	占总数百分比 Percentage(%)
0～300	5 150 211	98.71	200～300	48 993	30.96
300～500	35 277	0.68	300～500	37 781	23.87
500～1 000	18 361	0.35	500～1 000	30 411	19.22
1 000～2 000	9 602	0.18	1 000～2 000	27 357	17.29
2 000+	4 180	0.08	2 000+	13 721	8.67
总数 Total number	5 217 631			158 263	
总长度 Total length	262 993 383			126 640 737	
N50 长度 N50 length	49			1 381	
平均长度 Mean length	50.404 75			800.191 7	

2.1.3　Unigenes 数据库构建　由于后续的表达谱分析是建立在同一个参考基因的基础上,故采用相似度的方式构建 Unigenes 数据库,即当序列的比对长度达到较长序列的 80% 或者较短序列的 90% 时,认为 2 条序列同源; 取较长的序列作为代表,对样品先前聚类得到的基因序列作进一步的聚类,整合得到该项目物种 Unigenes 数据库。从表 3 看出共得到 Unigenes 94 476 条,其中长度在 1 kbp 以上的 Unigenes 有 12 643 条,占 Unigene 库总数的 13.38%,N50 长度为 806 bp,说明组装效果较好。

表 3　Unigenes 长度统计结果

Tab. 3　The statistical results of unigenes length

Unigenes 长度 Unigenes length/bp	总数 Total number	占百分比 Percentage(%)
200～300	40 923	43.32
300～500	27 155	28.74
500～1 000	13 755	14.56
1 000～2 000	8 386	8.88
2 000 +	4 257	4.51
总数 Total number	94 476	
总长度 Total length	53 644 108	
N50 长度 N50 length	806	
平均长度 Mean length	567.806 723 400 652	

2.1.4　Unigenes 与其他物种相似性比较　从图 2 可以看出,在目前的 Nr 数据库中,油茶转录组测序组装的 Unigenes 与葡萄(*Vitis vinifera*)的 Unigenes 相似数量最多,达到了 53.44%,其次是其他物种,毛果杨(*Populus trichocarpa*)为第三,而与油茶近亲的种即茶树(*Camellia sinensis*)相似的数量仅为 0.58%。

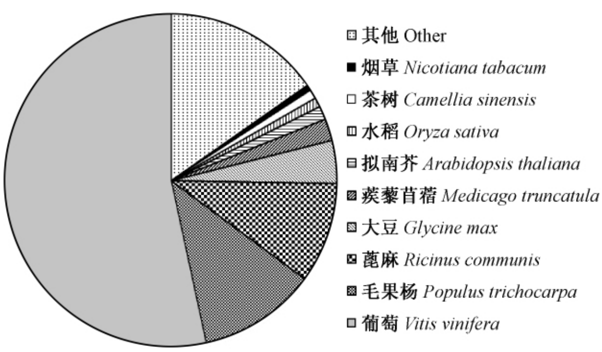

图 2　在 Nr 数据库中 Unigenes 与其他物种的比较

Fig.2　Unigenes similarity comparison with other species in Nr database

2.1.5　Unigenes 的功能注释　使用 BLAST 软件将 Unigenes 序列与 Nr, Nt, SwissProt, TrEMBL, GO, COG, KEGG 数据库比对,获得 Unigenes 的注释信息。从表 4 可以看出,最终获得注释信息的 Unigenes 有 41 240 条,在 COG 和 KEGG 数据库中注释的 Unigenes 数量都在 10 000 条以下,其他数据库则都超过 20 000 条,大部分集中在 30 000 条左右。

表 4　Unigenes 的功能注释在各数据库分布情况

Tab. 4　The database distribution of unigenes functional annotation

注释数据库名称 Database name	被注释数量 Annotated number	300 bp≤长度<1 000 bp 300 bp≤Length<1 000 bp	长度≥1 000 bp Length≥1 000 bp
COG	9 095	2 981	5 104
GO	27 201	11 406	10 515
KEGG	6 431	2 372	2 904
SwissProt	24 534	10 064	10 376
TrEMBL	36 393	16 403	12 113
Nr	36 400	16 412	12 117
Nt	30 858	13 165	11 460
被注释 Unigenes 总量 Annotated total number	41 240	18 743	12 220

2.1.6　Pathway 注释分析　Unigenes 通过与 KEGG 数据库比对,对 13 647 条 Unigenes 进行了 Pathway 注释,涉及的 Pathway 有 286 个。其中有 200 条 Unigenes 以上的 Pathway 有: 剪接体(346 条)占总数的 2.54%,分子伴侣和可折叠催化(247 条)占总数 1.81%,染色体(360 条)占总数 2.64%,内质网中的蛋白质加工(213 条)占总数的 1.56%,泛素系统(314 条)占总数 2.3%,植物激素信号转导(236 条)占总数 1.73%,转录因子(267 条)占总数 1.96%,糖基转移酶(224 条)占总数 1.64%,蛋白激酶(201 条)占总数 1.47%,DNA 复制和重组蛋白(281 条)占总数 2.06%,氧化磷酸化(237 条)占总数 1.74%,核糖体(294 条)占总数 2.15%,其他 Pathway 的 Unigenes 数量都在 200 条以下,大部分集中在 100 条以内。涉及与植物花发育相关的 Pathway 有: 植物激素信号转导,共有 236 条 Unigenes,占总数 1.73%; 植物昼夜调节,共 13 条 Unigenes,占总数 0.1%; 淀粉和蔗糖代谢,共有 168 条 Unigenes,占总数 1.23%。

2.1.7　GO 数据库注释分析　在 GO 数据库中(图 3),执行细胞组成功能(cellular component) Unigenes 有 42 106 条,其中达到 1 000 条以上 Unigenes,分别为: 核 5 102 条、质膜 4 259 条、线粒体 3 414 条、叶绿体 2 589 条、不可或缺的膜 2 116 条、胞质溶胶 1 999条、细胞膜 1 548 条、胞浆膜界囊泡 1 413 条、细胞质 1 330 条、质体 1 194 条、胞间连丝 1 140 条、胞外区 1 053 条、高尔基体 1 047 条,其他都在 700 条以下。执行分子功能(molecular function)有 47 192条 Unigenes,其中大于 1 000 条有: ATP 结合(3 069条)、锌离子结合(1 680 条)、核酸结合(1 658

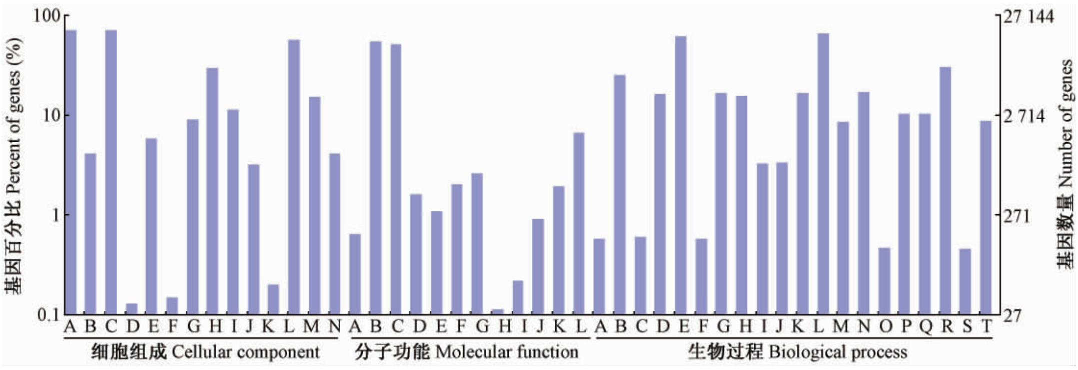

图 3 Unigenes 在 GO 中功能注释统计结果

Fig. 3 Unigenes functional annotation in the GO

细胞组成：A：细胞；B：细胞连接；C：细胞要素；D：细胞外基质；E：胞外区；F：胞外区要素；G：大分子复合物；H：膜；I：膜要素；J：膜封闭腔；K：核仁；L：细胞器；M：细胞器要素；N：共质体。分子功能：A：抗氧化活性；B：绑定；C：催化活性；D：电子载体活性；E：酶调节活性；F：分子转导活性；G：核酸结合的转录因子的活性；H：营养库活性；I：蛋白结合转录因子活性；J：受体活性；K：结构分子活性；L：转运活性。生物过程：A：生物附着；B：生物调控；C：细胞增殖；D：细胞成分和生物合成；E：细胞过程；F：凋亡；G：发育过程；H：定位系统建立；I：生长；J：免疫系统过程；K：定位；L：代谢过程；M：多机体过程；N：多细胞组织过程；O：色素沉着；P：再生；Q：再生过程；R：刺激应答；S：律动过程；T：信号传导。

Cellular component: A: Cell; B: Cell junction; C: Cell part; D: Extracellular matrix; E: Extracellular region; F: Extracellular region part; G: Macromoleular complex; H: Membrane; I: Membrane part; J: Membrane-enclosed lumen; K: Nucleoid; L: Organelle; M: Organelle part; N: Symplast. Molecular function: A: Antioxidant activity; B: Binding; C: Catalytic activity; D: Electron carrier activity; E: Enzyme regulator activity; F: Molecular transducer activity; G: Nucleic acid binding transcription factor activity; H: Nutrient reservoir activity; I: Protein binding transcription factor activity; J: Receptor activity; K: Structural molecule activity; L: Transporter activity. Biological process: A: Biological adhesion; B: Biological regulation; C: Cell proliferation; D: Cellular component organization or biogenesis; E: Cellular process; F: Death; G: Developmental process; H: Establishment of localization; I: Growth; J: Immune system process; K: Localization; L: Metabolic process; M: Multi-organism process; N: Multicellular organism process; O: Pigmentation; P: Reproduction; Q: Reproductive process; R: Response to stimulus; S: Rhythmic process; T: Signaling.

条）、核苷酸结合（1 470 条）、金属离子结合（1 442 条）、DNA 结合（1 353 条）、蛋白结合（1 229 条）、丝氨酸/苏氨酸激酶活性（1 037 条）、绑定（1 019 条）。涉及油茶生物过程功能（biological process）的 Unigenes 有 80 231 条，其中最多是氧化还原过程 1 833条，其次是蛋白磷酸化 1 161 条，其他全部都在 1 000条以下。

2.1.8 COG 数据库注释分析 图 4 可以看出，在 COG 数据库中 Unigenes 执行的功能分成：翻译，核糖体结构和生物合成，RNA 加工和修饰，转录、复制、重组和修复，染色质结构和动力学，细胞周期调控，细胞分裂，染色体分区，核结构，防御机制，信号转导机制，细胞壁、膜和包膜合成，细胞运动，骨架，胞外结构，细胞内运输，分泌和膜泡运输，翻译后修饰，蛋白质，伴侣，能源生产和转换，碳水化合物的运输和代谢，氨基酸转运和代谢，核苷酸转运和代谢，辅酶运输和代谢，脂质运输和代谢，无机离子转运和代谢，次生代谢产物的生物合成、运输和代谢等。其中只有一般功能的 Unigenes 数量最多，共有 2 489 条，其次是具复制、重组和修复功能的 Unigenes，数量共有 1 420 条，转录功能 Unigenes 数量有 1 214 条，信号转导机制 Unigenes 数量有 1 065 条，其他功能的 Unigenes 都在 1 000 条以下，其中较多的为翻译后的修饰功能，蛋白质周转和伴侣功能 Unigenes 有 837 条，糖类转运与代谢 Unigenes 有 690 条。

2.2 花发育相关基因在成花转变过程中的表达

从图 5 可以看出 *FLC*，*FT* 和 *FAC* 较 *AP1*，*AP2* 和 *PI* 在茎尖表达量相对较低。*FLC* 在根系中表达量最高，5 月 18 日表达量最低；*FT* 在 3 月 19 日表达量最高，4 月 18 日开始到 5 月 18 日表达量变动不大；*FAC* 在 3 月份的表达量较其他时期高，其与 *FLC* 基因表达量存在相互消长的关系。*PI* 在 5 月中旬表达量最高，*AP2* 则在根部和 5 月中上旬表达量较高，*AP1* 在 3 月份表达量都较高。

2.3 花发育相关基因在花器官形成过程中的表达

从图 6 可以看出，在花器官形成过程中 *AP2* 和 *PI* 基因相对表达水平较其他基因高，*AP2* 基因在雌雄蕊成熟期（8 月 25 日）相对表达量最高，其次是在雌雄蕊形成期（7 月 3 日），在萼片形成期（5 月 28 日）相对表达量最低。*PI* 基因在雌雄蕊形成期（7 月 3 日）相对表达量最高，其次为子房与花药形成期（7 月 25 日），再次为花瓣形成期（6 月 15 日）。

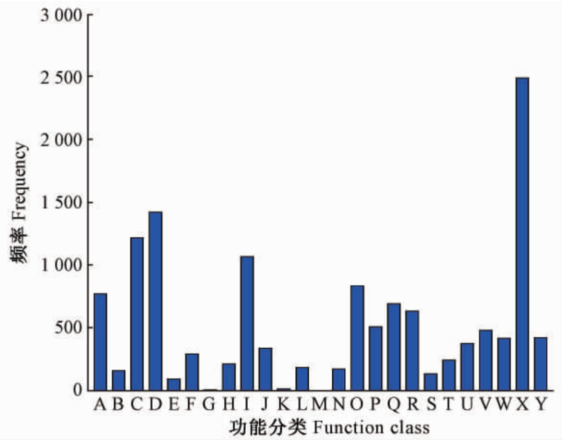

图 4　COG 数据库中功能注释的 unigenes 分类

Fig. 4　COG function classification of unigenes

A：翻译、核糖体结构和生物发生；B：RNA 加工和修饰；C：转录；D：复制、重组和修饰；E：染色体结构和动力学；F：细胞周期调控、细胞分裂、染色体分离；G：核结构；H：防御机制；I：信号传递机制；J：细胞壁/细胞膜生物发生；K：细胞活性；L：细胞骨架；M：胞外结构；N：细胞间运输、分泌物和囊泡运动；O：翻译后修饰、蛋白翻转、伴侣；P：能源产生与转化；Q：碳水化合物转运和代谢；R：氨基酸转运和代谢；S：核酸转运和代谢；T：辅酶转运和代谢；U：脂类转运和代谢；V：无机离子转运和代谢；W：次生代谢物生物合成、转运和代谢；X：只有一般功能预测；Y：未知功能。

A：Translation，ribosomal structure and biogenesis；B：RNA processing and modification；C：Transcription；D：Replication，recombination and repair；E：Chromatin structure and dynamics；F：Cell cycle control，cell devision，chromosome partitioning；G：Nuclear structure；H：Defense mechanisms；I：Signal transduction mechanisms；J：Cell wall/ membrane/ envelope biogenesis；K：Cell motility；L：Cytoskeleton；M：Extracellular structures；N：Intracellular trafficking，secretion，and vesicular transport；O：Posttranslational modification，protein turnover，chaperones；P：Energy production and conversion；Q：Carbohydrate transport and metabolism；R：Amino acid transport and metabolism；S：Nucleotide transport and metabolism；T：Coenzyme transport and metabolism；U：Lipid transport and metabolism；V：Inorganic ion transport and metabolism；W：Secondary metabolites biosynthesis，transport and catabolism；X：General function prediction only；Y：Function unknown.

AP1 基因在花发育的 5 个时期表达差异不明显，*FT* 和 *FAC* 在 6 月 15 日和 7 月 3 日表达量较其他时期高，但从总体上看在花器官发育过程中 *FCA* 和 *FLC* 基因的表达量较在成花转变期低。

3　结论与讨论

3.1　转录组测序分析

转录组学（transcriptomics）是功能基因组学研

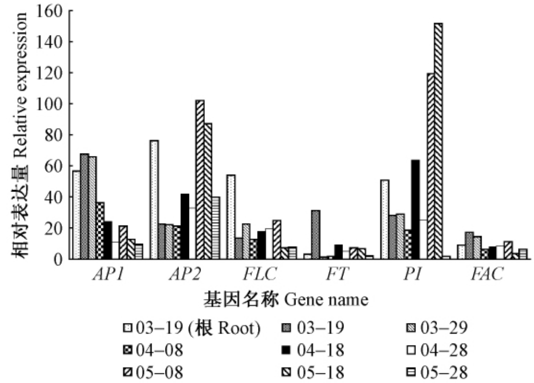

图 5　成花相关基因在茎尖的表达

Fig. 5　The expression of flowering gene in shoot tips

除 03 – 19 含根尖部位，其他时期只指春梢茎尖部位。In addition to 03 – 19 containing root tip，other times only refer to the spring shoot tip

AP1：Apetala1；*AP2*：Apetala2；*FLC*：Flowering locus C；*FT*：Flower locus T；*PI*：Pistillata；*FAC*：Flowering time control. 下同。The same below.

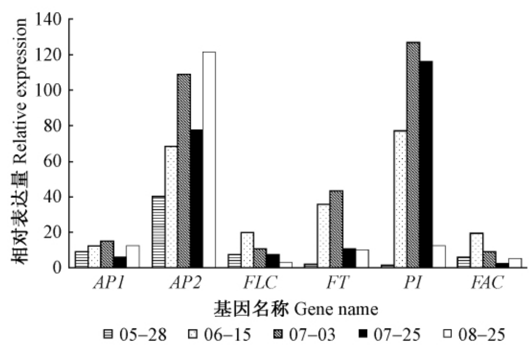

图 6　花器官形成过程中成花相关基因的表达

Fig. 6　The expression of floral organ gene during the process of floral organ formation

究的一个重要内容，它是从整体水平上研究细胞中基因转录的情况及其转录调控规律。转录组的概念由 Velculescu 等（1997）首先提出。转录组是指特定细胞在某一功能状态下全部表达的基因总和，一般指的是 mRNA 总和，基因表达具有特定的空间性和时间性特征（Lander *et al.*，2001），因此知道 mRNA 序列是转录组研究的重要基础；随着新一代测序技术的发展，其高通量性、低成本、快速、准确的特点可大大提高测序的效率，降低试验成本，为从 RNA 水平研究基因表达提供有效方法（Nagalakshmi *et al.*，2008；Rosenkranz *et al.*，2008；Hoen *et al.*，2008；Wilhelm *et al.*，2009；Jack，2004）。本研究从花发育的 6 个时期获得了近 6 Gbp的测序数据，测序数据的拼接基本符合后续生物信息学分析，并获得了近 12 643 条 1 kbp 以上 Unigenes，其中大部分具有完整的编码区，为后

续准确获得目的基因提供便利，同时为有效运用转基因技术奠定基础。

3.2 花发育相关基因表达分析

根据现有的研究文献表明：*CO*（Constant）（Cerdan *et al.*，2003）是植物光周期途径的关键基因；*LFY*（LEAFY）（Weigel *et al.*，1992）和 *FT*（Huang *et al.*，2005）是成花调控网络的关键基因。本文研究根据转录组测序的基础数据库，通过对 *CO* 和 *LFY* 基因的同源和相似性比对，结果只发现成花相关基因 *FT*。Zeevaart(2008)研究发现，在拟南芥(*Arabidopsis thaliana*)的花发育不同阶段 *FT* 基因都有一定的表达，其主要在植物叶片中表达，在其他生物体部分(根、茎和花芽)表达量相对较低，这与本文分析结果一致。在花温度诱导自主成花途径中(Blazquez，2001)，发现大量编码热激蛋白的基因 *HSP*(heat shock proteins)、促花基因 *FCA* 及抑花基因 *FLC*。本研究通过对测序数据中 *HSP* 基因的表达量进行分析发现，随着环境温度升高，油茶花芽中 *HSP* 基因表达量急剧增加，表达量极高，通常是其他基因表达量的 10^3 倍以上，同时对油茶采样株系的观察可以发现，随着春季环境气温回升，油茶的春梢生长在 40 天里可以生长 50 cm。从本文研究中还可以看出 *FCA* 基因在春梢发育初期及花芽分化期表达量明显增加，在花芽分化完成后的花器官形成期表达量相对较小，*FLC* 基因其表达量与 *FCA* 存在相互消长的关系。*AP1* 和 *AP2* 都属于 MADS-box 基因，是真核生物中一类重要的转录调控因子，属于开花 ABC 模型的 A 类基因(Mandel *et al.*，1992；Bowman *et al.*，1993)；从本文成花过程表达分析来看，A 类基因不仅与花的萼片形成有重要关系，而在前期的花芽转变过程也可能具有调控作用。*PI* 属于 B 类基因，对花的雄蕊和花瓣形成具有重要作用(Howell，1998)；从本文花芽分化过程各个时期的表达量来看，*PI* 基因还可能影响油茶的花芽分化。因此对 *FT* 及 *PI* 基因的改造及遗传改良可以获得早花和雄性不育油茶品种，为油茶高效栽培和育种打下良好基础。

3.3 油茶成花途径及成花关键基因分析

目前植物成花转变主要有光周期途径、春化途径、赤霉素途径、糖代谢途径和自主途径(潘瑞炽，2004；王忠，2002)。在本研究中没有发现与植物光周期关键基因 *CO* 的同源基因，也没有发现与春化相关的基因及控制花发端由营养生长转变成生殖生长的关键基因 *LFY*，而发现随着环境温度的升高(从 18 ℃ 到 37 ℃)合成热激蛋白的相关基因(*HSP*)表达量急剧攀升，春梢在良好水肥条件下快速生长，并在春梢生长停滞期完成了花芽分化，开始进入花器官形成阶段；鉴于前期试验发现，长林 4 号油茶经赤霉素处理对其花芽分化存在一定的抑制作用，而在室温(25 ℃)进行不同光周期和光强照射则不能完成花芽分化。结合上面分析可以初步判定，油茶成花途径是基于一定温度作用(持续高温)自身遗传决定的自主过程。首先，油茶春梢在持续升高的环境温度作用下编码热激蛋白的 *HSP* 基因表达量迅速升高，油茶在良好的水热及养分条件作用下春梢快速生长，春梢生长过程中促进成花基因 *FCA* 表达，并抑制 *FLC* 基因表达，同时上调成花网络关键基因 *FT*，*FT* 基因表达促进 *AP1* 和 *PI* 基因表达，而 *AP1* 和 *PI* 基因的表达引起花器官一系列相关基因的表达，从而完成了花芽分化。

参 考 文 献

陈 英，江香梅，张 露，等. 2011. 基于油茶组 59 万条 EST 序列的转录组学初步分析. 林业科学，47(2)：161－163.

高本年. 1981. 普通油茶适宜花期的初步探讨. 江苏林业科技，(4)：41－44.

韩宁林，赵学民. 2009. 油茶高产品种栽培. 北京：中国农业出版社.

黎章矩，施拱生，吴德晔. 1992. 油茶生育规律及其影响因子研究. 经济林研究，10(1)：21－26.

林 萍，曹永庆，姚小华，等. 2011. 普通油茶种子 4 个发育时期的转录组分析. 分子植物育种，9(4)：498－505.

林少韩，徐乃焕. 1981. 油茶花期生态及结实力研究. 林业科学，17(2)：113－122.

潘瑞炽. 2004. 植物生理学. 北京：高等教育出版社.

唐光旭. 1984. 油茶春梢生长发育与物质转化的研究. 经济林研究，2(2)：44－51.

王文杰，陈长庚，程 剑. 2007. 茶油在医药保健方面的积极作用. 中国食物与营养，9(4)：48－51.

王湘南，蒋丽娟，马 力，等. 2011. 油茶花芽分化的形态解剖学特征观测. 中南林业科技大学学报，31(8)：22－27.

王 忠. 2002. 植物生理学. 北京：中国农业出版社.

袁德义，邹 锋，范晓明，等. 2011. 油茶花芽分化及雌雄配子体发育的研究. 中南林业科技大学学报，31(3)：65－70.

曾燕如，黎章矩，戴文圣. 2009. 油茶开花习性的观察研究. 浙江林学院学报，26(6)：802 － 809.

查小华，江国强，雷小林，等. 2012. 油茶结果枝与成花相关性研究. 江西林业科技，(2)：12－14.

庄瑞林. 2008. 中国油茶. 北京：中国林业出版社.

Blazquez M A. 2001. Flower development pathways. Cell Science at a Glance，113:3547－3548.

Bowman J L，Alvarez J，Weigel D，*et al.* 1993. Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. Development，119：721－743.

Cerdan P D，Chory J. 2003. Regulation of flowering time by light quality. Nature，423：881 －885.

Hoen P A C，Ariyurek Y，Thygesen H H，*et al*. 2008. Deep sequencing-based expression analysis shows major advances in robustness，resolution and inter-lab portability over five microarray platforms. Nucleic Acids Research，36(21)：e141.

Howell S H. 1998. Molecular genetic of plant development. London：Cambridge University Press.

Huang T，Böhlenius H，Eriksson S，*et al*. 2005. The mRNA of the *Arabidopsis* gene FT moves from leaf to shoot apex and induces flowering. Science，309(5741)：1694 －1696.

Jack T. 2004. Molecular and genetic mechanisms of floral control. Plant Cell，16(suppl 1)：S1 －S17.

Lander E S，Linton L M，Birren B，*et al*. 2001. Initial sequencing and analysis of the human genome. Nature，409(6822)：860 －921.

Mandel M A，Gustafson-Brown C，Savidge B，*et al*. 1992. Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature，360：273 －277.

Mortazavi A，Williams B A，Wold B，*et al*. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods，5(7)：621 －628.

Nagalakshmi U，Wang Z，Waern K，*et al*. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science，320(5881)：1344 －1349.

Rosenkranz R，Borodina T，Lehrach H，*et al*. 2008. Characterizing the mouse ES cell transcriptome with Illumina sequencing. Genomics，92(4)：187 －194.

Velculescu V E，Zhang L，Zhou W，*et al*. 1997. Characterization of the yeast transcriptome. Cell，88(2)：243 －251.

Weigel D，Alvarez J，Smyth D R，*et al*. 1992. *Leafy* controls floral meristem identity in *Arabidopsis*. Cell，69(5)：843 －859.

Wilhelm B T，Landry J R. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods，48(3)：249 －257.

Zeevaart J A D. 2008. Leaf-produced floral signals. Curr Opin in Plant Biol，11(5)：541 －547.

Zhou C F，Lin P，Yao X H，*et al*. 2013. Selection of reference genes for quantitative real-time PCR in six oil-tea camellia based on RNA-seq. Molecular Biology，47(6)：836 －851.

（责任编辑　徐　红）