

普通油茶开花与授粉期转录组测序及相关基因差异表达分析

龙雪燕¹, 古龙², 谭玉珊², 李万春¹, 胡玉玲²

(1. 浙江农林大学 林业与生物技术学院, 省部共建亚热带森林培育国家重点实验室, 浙江省森林芳香植物康养功能研究重点实验室, 浙江 杭州 311300; 2. 湖南应用技术学院 农林科技学院, 湖南 常德 415000)

摘要: 为了解普通油茶开花与授粉期转录组特征, 并寻找普通油茶该时期生殖过程相关差异表达基因, 为普通油茶生殖分子机制提供研究基础。以普通油茶湘林 210# 4 个时期 (花未开、授粉 2-3 d, 授粉 7 d 和 20 d) 的花为材料, 采用高通量转录组测序。结果表明: 除去低质量数据后共得到 21.57Gb 的 Clean reads, 在阈值为 \log_2 (倍数变化) >1 或 \log_2 (倍变化) <-1 时, 共检测到 85 134 个 Unigene (46 872 个上调和 38 262 个下调) 差异基因表达。基于 GO、KEGG、Pfam、SwissProt、eggNOG 和 NR 等 6 个数据库进行相似性比对, 依次分别有 33290、26896、29515、26821、38129 和 40433 个基因被注释。差异基因 GO 功能分化分析表明: 基因大量富集在生物学过程, 有 71367 种不同的 GO 功能, 主要有生物过程、DNA 转录调控、DNA 转录、蛋白质磷酸化、氧化还原过程、防御反应; 在 KEGG 分析中, 植物激素信号转导、淀粉和蔗糖代谢富集的差异基因数最多, 分别有 404 个和 313 个。差异基因在油茶柱头 FR1 时表达量最高, 两条通路差异注释基因分别为 SPL8、AHP1、SERK2、SAUR67、CXE16、SAUR21、EIN3、TGAL4、ARR6 和 BGLU40、SBEI、TRE1、Os10g0521000、exgA。

关键词: 普通油茶; 花期; 转录组; 差异基因

Differential expression analysis of genes in the reproductive process of ordinary *Camellia oleifera*

Long Xueyan¹, Gu Long², Tan Yushan², Li Wanchun¹, Hu Yuling²

(1. State Key Laboratory of Subtropical Silviculture, Zhejiang Provincial Key Laboratory of Forest Aromatic Plants-based Healthcare Function, School of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou 311300, China; 2. College of Agriculture and Forestry Science and Technology, Hunan Applied Technology University, Changde Hunan 415000, China)

Abstract: In order to understand the Transcriptome characteristics of common *Camellia oleifera* during flowering and pollination, and to find the differentially expressed genes related to the reproductive process of common *Camellia oleifera* during this period, it provides a research basis for the molecular mechanism of common *Camellia oleifera* reproduction. High throughput Transcriptome sequencing was used to sequence the flowers of the common *Camellia oleifera* Xianglin 210 # 4 periods (flowers not blooming, 2-3 days of pollination, 7 days and 20 days of pollination). The results showed that 21.57Gb Clean reads were obtained after removing

¹ 基金项目: 湖南省教育厅重点项目(22A0714); 湖南省“十四五”应用特色学科 (林学) (湘教通[2022]351)。

作者简介: 龙雪燕 (1998-), 女, 贵州铜仁人, 硕士研究生, 主要研究方向为经济林木发育与栽培生理。E-mail: 2363078950@qq.com。

**通讯作者 Corresponding author: 胡玉玲 (1977-), 男, 浙江温州人, 博士, 教授, 主要从事油茶育种与栽培研究, E-mail: huyulin@126.com。

low-quality data. When the threshold value was $\log_2(\text{multiple change}) > 1$ or $\log_2(\text{multiple change}) < -1$, 85134 Unigenes (46872 up-regulated and 38262 down-regulated) were detected. Based on six databases including GO, KEGG, Pfam, SwissProt, eggNOG, and NR, similarity comparisons were conducted, and 33290, 26896, 29515, 26821, 38129, and 40433 genes were annotated, respectively. The differentiation analysis of GO function of different genes shows that genes are largely enriched in biological processes, with 71367 different GO functions, mainly including biological processes, DNA Transcriptional regulation, DNA transcription, protein phosphorylation, protein phosphorylation, redox process, and defense response; In KEGG analysis, there were 404 and 313 differential genes in Plant hormone signal transduction, starch and sucrose metabolism enrichment, respectively. The expression level of differentially expressed genes is highest in *Camellia oleifera* stigma FR1, and the differentially annotated genes in the two pathways are SPL8, AHP1, SERK2, SAUR67, CXE16, SAUR21, EIN3, TGAL4, ARR6, BGLU40, SBEI, TRE1, Os10g0521000, and exgA, respectively.

Keywords: Ordinary *Camellia oleifera*; Flowering period; Transcriptome; Differential gene

油茶(*Camellia oleifera* Abel.)是山茶科(*Theaceae*)山茶属(*Camellia* L.),常绿灌木或小乔木,是我国特有的木本油料树种^[1],因其富含多种不饱和脂肪酸,保健价值极高,故被誉为“东方液体黄金”^[2]。生殖过程是生物延续的重要生命现象,与遗传、进化、乃至生命起源密切相关^[3]。普通油茶一般生殖过程包括五个阶段:花芽分化,花的发育,开花传粉,受精,果实及种子发育^[4],传粉和受精是油茶座果的重要环节。油茶是雌雄同花植物,花期从10月初到翌年2月份,盛花期在11月份左右^[5]。油茶亦是“抱籽怀胎”植物,具有严重的落花落果现象^[6],影响油茶产量,其原因有油茶授粉亲和力不高、自交不亲和性以及品种花期不遇等^[7],因此有必要对油茶生殖过程及其内在分子机制进行研究,以便制定相应的花期调控措施、授粉树配置、良种选育,对油茶增产具有重要意义。目前,油茶花发育相关研究主要集中在花发育过程花解剖结构变化、花期、花粉亲和性、自交可育性等方面,但开花授粉分子层面的研究报道较少。

转录组是指细胞中所有参与蛋白质翻译的信使RNA(mRNA)的总和^[8],转录组分析具有低成本、高通量等特点,从整体层面上进行分子调控网络研究,尤其适用于无参考基因组的物种,目前已在动物、植物和微生物中得到广泛应用^[9]。本研究利用转录组测序技术,分别对油茶授粉的花4个关键时期进行转录组测序,利用现有数据库与测序结果进行比对,分析油茶授粉前后差异基因变化情况,分别对所得差异表达基因进行GO和KEGG注释,进一步探讨油茶授粉发育过程中差异表达基因的功能分类及其所参与的相关通路,为今后挖掘油茶授粉柱头化学成分的调控基因提供借鉴,油茶的生殖过程分子机制奠定基础及利用基因工程培育新品种提供方向。

1 材料与方法

1.1 实验材料

研究材料采自贵州省松桃县盘兴镇正大油茶基地的4株普通油茶湘林#,树龄为8a,生长状况良好。分别采集2020年10月20日(花未开)、10月25日(授粉2-3d)、10月31日(授粉7d)、11月22日(授粉20d)花,去除萼片后用锡纸包好置于液氮中保存。进行三个生物学重复采样,分别记为FL₁-1, FL₁-2,

FL₁₋₃, FR₁₋₁, FR₁₋₂, FR₁₋₃, FR₂₋₁, FR₂₋₂, FR₂₋₃, FR₃₋₁, FR₃₋₂, FR₃₋₃。

1.2 普通油茶总 RNA 提取、文库制备及测序

采用 Trizol 试剂盒(Invitrogen, CA, USA)提取油茶 4 个时期的柱头总 RNA, 总 RNA 的数量和纯度通过 RIN 数>7.0 的 Bioanalyzer 2100 和 RNA 1000 Nano LabChip 试剂盒(安捷伦, 加利福尼亚州, 美国)进行分析。

按照 Illumina 公司的文库制备实验流程制备文库, 样本提取的总 RNA 经过质检合格后, 使用连接有 Oligo(dT)的磁珠富集真核生物 mRNA。经抽提的 mRNA 被片段化试剂(Fragmentation Buffer)随机打断成短片段, 以片段化的 mRNA 为模板, 用六碱基随机引物(Random hexamers)合成一链 cDNA, 随后加入缓冲液、dNTPs、RNaseH 和 DNA Polymerase I 进行二链 cDNA 合成。AMPure XP beads 纯化双链产物, 利用 T4 DNA 聚合酶和 Klenow DNA 聚合酶活性将 DNA 的粘性末端修复为平末端, 3' 末端加碱基 A 并加接头, AMPureXP beads 进行片段选择, 之后用 USER 酶降解含有 U 的 cDNA 第二链进行 PCR 扩增获得最终测序文库。文库质检合格后采用 Illumina Novaseq™ 6000 进行测序, 测序读长为双端 2*150bp(PE150)。

1.3 从头组装、单基因注释和功能分类

首先, 使用内部的 Cutadapt^[10]和 perl 脚本来删除包含适配器污染、低质量碱基和未确定碱基的读取。然后使用 FastQC 验证序列质量(<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)。包括清洁数据的 Q20、Q30 和 GC 内容。所有下游分析均基于高质量的干净数据。使用 Trinity 2.4.0 进行转录组的从头组装^[11]。Trinity 根据共享序列内容将转录物分组为簇。这样的转录物簇被松散地称为“基因”。该簇中最长的转录物被选为“基因”序列(又名单基因)。

将所有组装的 Unigenes 与非冗余(Nr)蛋白质数据库进行比对(<http://www.ncbi.nlm.nih.gov/>), 基因本体论(GO)(<http://www.geneontology.org/>), SwissProt(<http://www.expasy.ch/sprot/>), 京都基因和基因组百科全书(KEGG)(<http://www.genome.jp/kegg/>)和 eggNOG(<http://eggnogdb.embl.de/>)使用阈值为 Evalue<0.00001 的 DIAMOND^[12]的数据库。

1.4 差异表达单基因分析

Salmon^[13]用于通过计算 TPM^[14]来执行 Unigenes 的表达水平。通过 R package edgeR^[15]选择 log₂(倍数变化)>1 或 log₂(倍数变化)<-1 且具有统计学意义(p 值<0.05)的差异表达的 Unigenes。

2 结果与分析

2.1 转录组数据测序及序列组装

采用 Illumina 测序平台技术对普通油茶样本进行转录组测序, 共得到 25.72 Gb 的原始数据(raw data); 过滤掉不合格的序列, 共得到 21.57 Gb 有效数据, 占原始数据的 83.86%, 组装后 GC 含量与长度分布如

图 1~图 2 所示, 每组数据质量见表 1。各样本得到过滤后数据 clean reads 占原始数据的 89.87%、93.98%、85.94%、90.05%，其中 QC20 和 QC30 均超过 90%，说明转录组测序数据质量较高，可进行后续生物信息学分析。

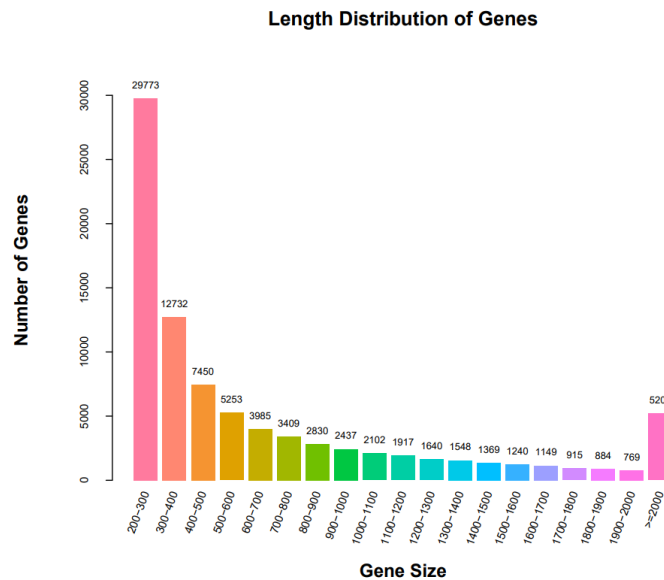


图 1 普通油茶授粉柱头 Unigene 组装长度统计

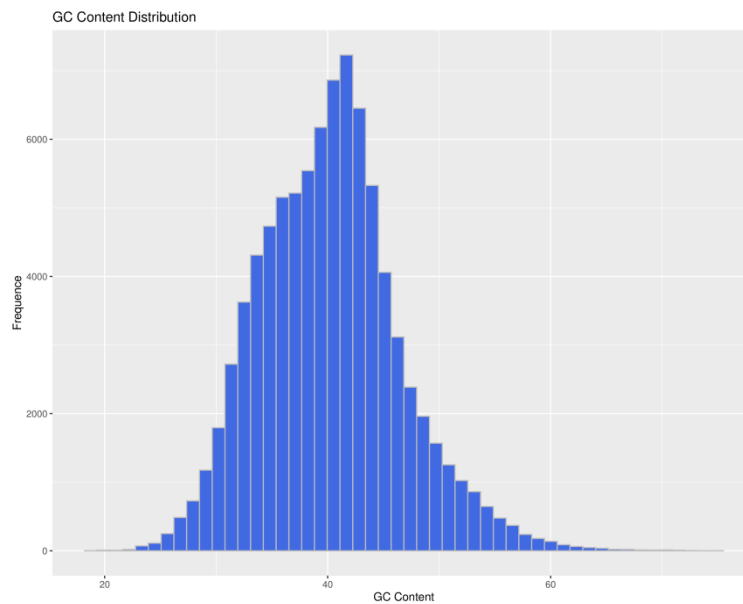


图 2 普通油茶授粉柱头 Unigene 的 GC 含量统计

表 1 过滤前后的 Reads 统计

Sample	Raw_Reads	Raw_Bases (原)	Valid_(clean)	Valid_Bases	Valid% (有	Q20%	Q3	GC%
--------	-----------	---------------	---------------	-------------	-----------	------	----	-----

(样品)	(原始数据)	始数据大小(G)	Reads (有效数据)	(有效数据大小(G)	效比例(%)	0%			
FL1	43 794 206	6.57G	39 285 087	5.49G	89.87	98.30	94.51	45.62	
FR1	40 486 891	6.07G	38 048 351	5.32G	93.98	98.34	94.62	45.75	
FR2	46 891 663	7.03G	40 391 519	5.65G	85.94	98.36	94.65	45.48	
FR3	40 540 441	6.05G	36 539 157	5.11G	90.05	98.38	94.68	45.48	

注: Q20: 质量值 ≥ 20 的碱基所占比例; Q30: 质量值 ≥ 30 的碱基所占比例; GC%: GC含量所占比例。

2.2 Unigene 功能注释

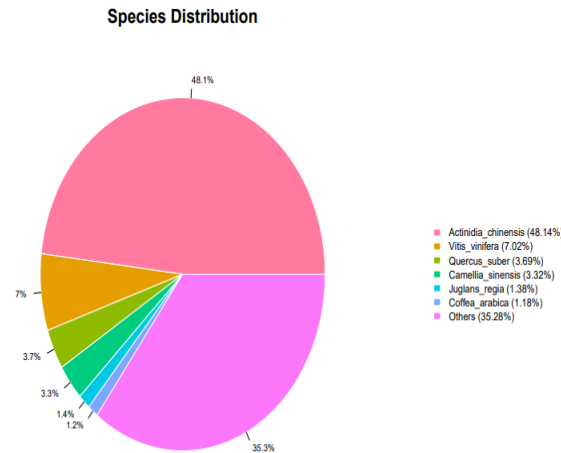
将 Unigene 序列与 NCBI 中的 GO 数据库、KEGG 数据库、KOG (euKaryotic Ortholog Groups) 数据库、Pfam 蛋白质数据库、Swiss-Prot 蛋白质数据库、eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) 数据库和 NR 非冗余蛋白质数据库进行比对, 最终获得注释信息的 Unigene 有 86602 条 (表 2)。共有 40433 条 Unigene 注释到 NR 数据库, 其中, 48.14% 的 Unigene 被注释到中华猕猴桃 (*Actinidia chinensis*) 中 (图 3: A)。利用 eggNOG 数据库进行功能注释, 被注释到的 Unigene 共有 38129 条, 分别注释到 23 个类群中。其中翻译后修饰、蛋白折叠、分子伴侣, 信号转导机制, 转录, 复制、重组和修复, 翻译、核糖体结构和生物合成是最大的 5 个类群; 核结构是最小类群 (图 3: B)。对普通油茶的 Unigenes 进行 GO 功能富集, 共注释到 191767 种不同的 GO 功能, 共有生物学过程、细胞组分和分子功能 3 大类 50 个功能组 (图 3: C)。其中生物学过程 (Biological Process) 占比最多, 有 71367 种不同的 GO 功能, 主要有生物过程、DNA 转录调控、DNA 转录、蛋白质磷酸化蛋白质磷酸化、氧化还原过程、防御反应; 细胞组分有 63232 种不同的 GO 功能, 主要是在细胞核, 其次是细胞质、质膜; 分子功能有 57158 种不同的 GO 功能, 主要有蛋白质结合和分子功能。

表 2 普通油茶授粉柱头 Unigene 的功能注释在各数据库分布情况 (Unigene 功能注释统计)

DB 注释数据库	Num 注释总数量	Ratio(%)
All	86602	100.00
GO	33290	38.44
KEGG	26896	31.06
Pfam	29515	34.08
swissprot	26821	30.97
eggNOG	38129	44.03
NR	40433	46.69

注: DB: 数据库名称; Num: Unigene 在数据库中被注释的数目; Ratio: 注释到的 Unigene 数和所有 Unigene 所占比例。

(A)



(B)

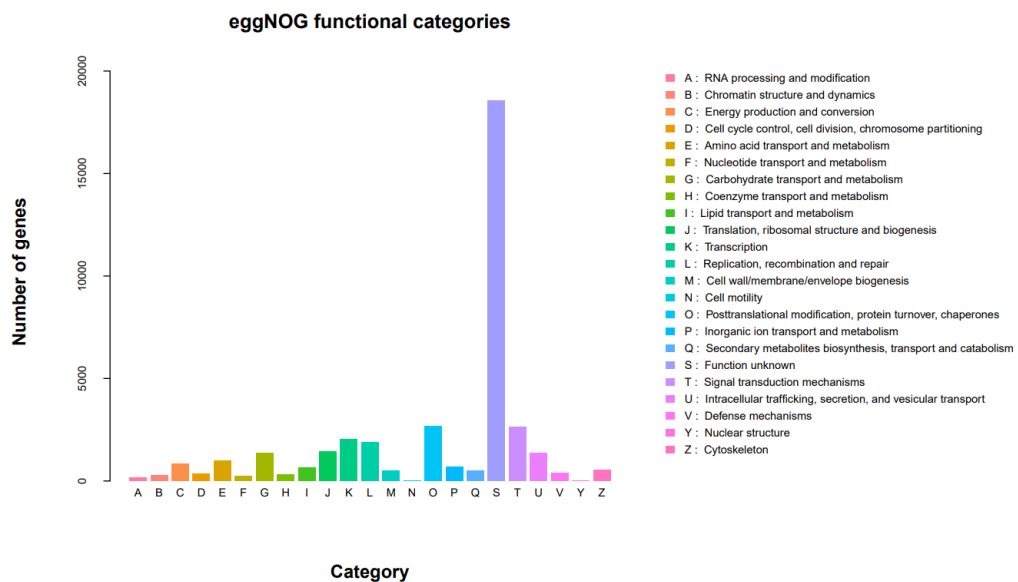
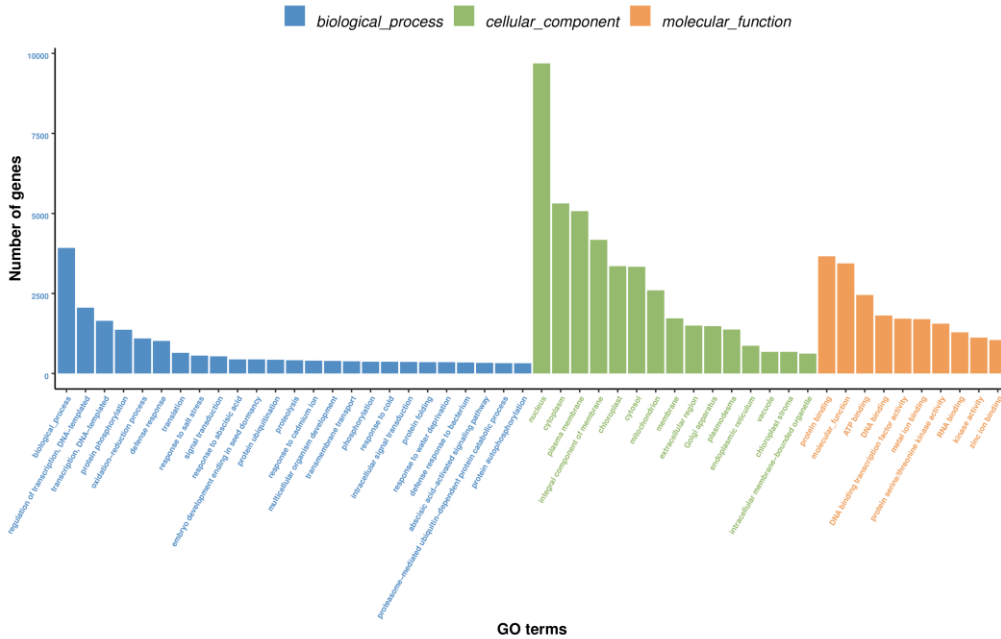


图 B 中横坐标的 A-Z 分别表示，A: RNA 加工与修饰；B: 染色体的结构域动态；C: 能量产生与转换；D: 细胞周期控制、细胞分裂、染色体他分离；E: 氨基酸的运输和代谢；F: 核苷酸的运输和代谢；G: 碳水化合物的运输与代谢；H: 辅酶的运输和代谢；I: 脂质的运输和代谢；J: 翻译、核糖体结构和生物合成；K: 转录；L: 复制、重组和修复；M: 细胞壁/膜/胞外被膜；N: 细胞运动；O: 翻译后修饰、蛋白折叠、分子伴侣；P: 无机盐的运输和代谢；Q: 次生代谢物的生物合成；S: 功能未知；T: 信号转导机制；U: 细胞内运输、分泌和囊泡运输；V: 防御机制；Y: 核结构；Z: 细胞骨架。

(C)



注：1~25：生物过程；26~40：细胞组分;41~50：分子功能；1.生物过程；2.DNA 转录调控；3.DNA 转录；4. 蛋白质磷酸化蛋白质磷酸化；5.氧化还原过程；6.防御反应；7.转化；8.对盐胁迫的反应；9. 信号转导；10. 对脱落酸的反应；11. 胚发育结束于种子休眠；12. 蛋白质泛素化； 13. 蛋白质水解；14. 对镉离子的反应；15.多细胞生物发育；16. 跨膜运输跨膜运输；17. 磷酸化； 18. 对冷胁迫的反应；19. 细胞内信号转导；20.蛋白质折叠；21. 对水分缺乏的反应；22. 对细菌的防御反应；23. 脱落酸激活的信号通路；24. 蛋白酶体介导的泛素依赖性蛋白质分解代谢过程；25.蛋白质自体磷酸化；26. 细胞核；27. 细胞质；28. 质膜；29. 膜的组成成分；30. 叶绿体；31. 细胞质；32. 线粒体；33.膜；34. 胞外区；35. 高尔基体；36. 胞间连丝；37. 内质网；38. 液泡；39. 叶绿体基质；40. 膜的组成成分；41. 蛋白质结合；42. 分子功能；43. A T P 结合；44. D N A 结合；45. D N A 结合转录因子活性；46. 金属离子结合；47. 蛋白质丝氨酸 / 苏氨酸激酶活性；48. R N A 结合 49. 激酶活性；50.锌离子结合。

图 3 普通油茶授粉柱头转录组功能注释 (A: NR 数据库注释物种分布; B: eggNOG 数据库分布统计; C: GO 功能注释)

2.3 差异基因表达分析

通过将花未开 (FL1)、授粉后 2-3 d (FR1)、授粉后 7 d (FR2)、授粉后 20 d (FR3) 四个时期花的文库基因表达相互对比发现, FR3_VS_FL1 上调基因数为 10369 条, 下调基因数为 9087 条; FR2_VS_FR1 上调基因数为 11748 条, 下调基因数为 8825 条; FR3_VS_FR1 上调基因数为 13937 条, 下调基因数为 10496 条; FR3_VS_FR2 上调基因数为 222 条, 下调基因数为 419 条 (图 5)。从上述数据分析可得, 随着油茶花

芽的发育，上调基因数量逐渐增多，说明随着花的发育，越来越多的基因被启动，但是到后期，差异表达基因明显减少，说明花芽发育已经趋于平稳。

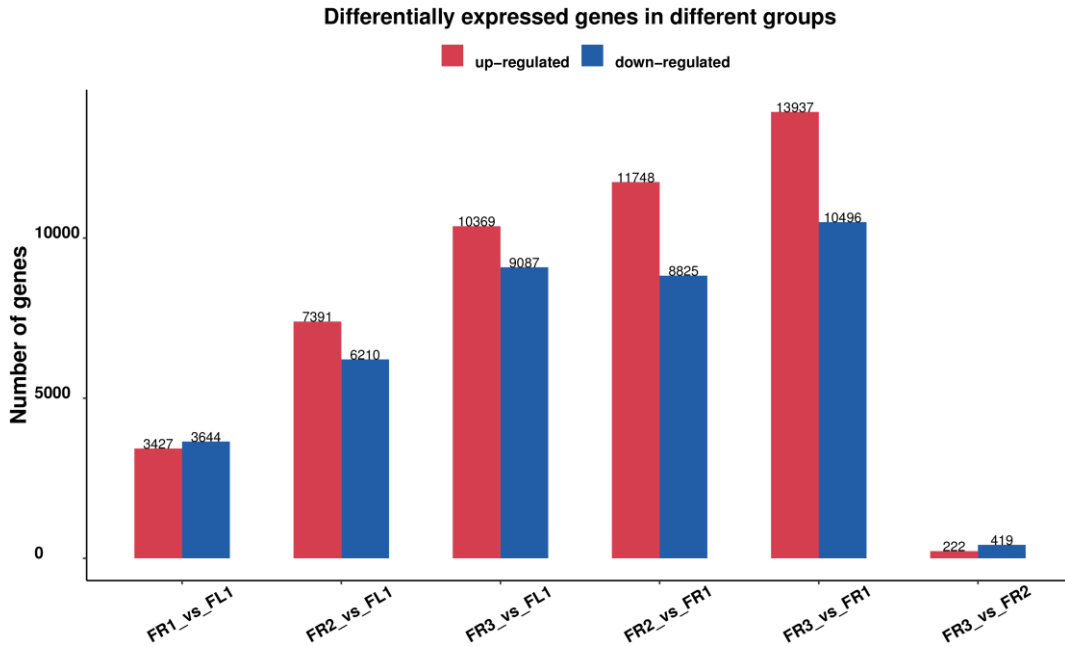


图 4 普通

油茶授粉花 4 个发育时期差异表达基因数

FR1_VS_FL1: 授粉 2-3 d/花未开; FR2_VS_FL1: 授粉后 7 d/花未开; FR3_VS_FL1: 授粉后 20 d/花未开; FR2_VS_FR1: 授粉后 7 d/授粉 2-3 d; FR3_VS_FR1: 授粉后 20 d/授粉 2-3 d; FR3_VS_FR2: 授粉后 20 d/授粉后 7 d。

2.4 KEGG Pathway 功能注释及分类

对普通油茶授粉前和授粉后的 Unigene 序列与 KEGG 数据库进行比对 (图 6), 得出差异表达的 20 条代谢通路, 分别是泛醌和其他萜类醌的生物合成、酪氨酸代谢、萜类骨架生物合成、淀粉和蔗糖代谢、植物激素信号转导、苯丙素的生物合成、光合作用-天线蛋白、

戊糖和葡萄糖醛酸相互转化、非同源末端连接、单萜类生物合成、错配修复、MAPK 信号通路-植物、吲哚生物碱生物合成、糖酵解/新生葡萄糖、DNA 复制、二萜类生物合成、氨基氨基酸代谢, 角质、木栓素和蜡的生物合成、碱基切除修复、糖尿病并发症中的 AGE-RAGE 信号通路。其中植物激素信号转导、淀粉和蔗糖代谢富集的 Unigene 序列最多。

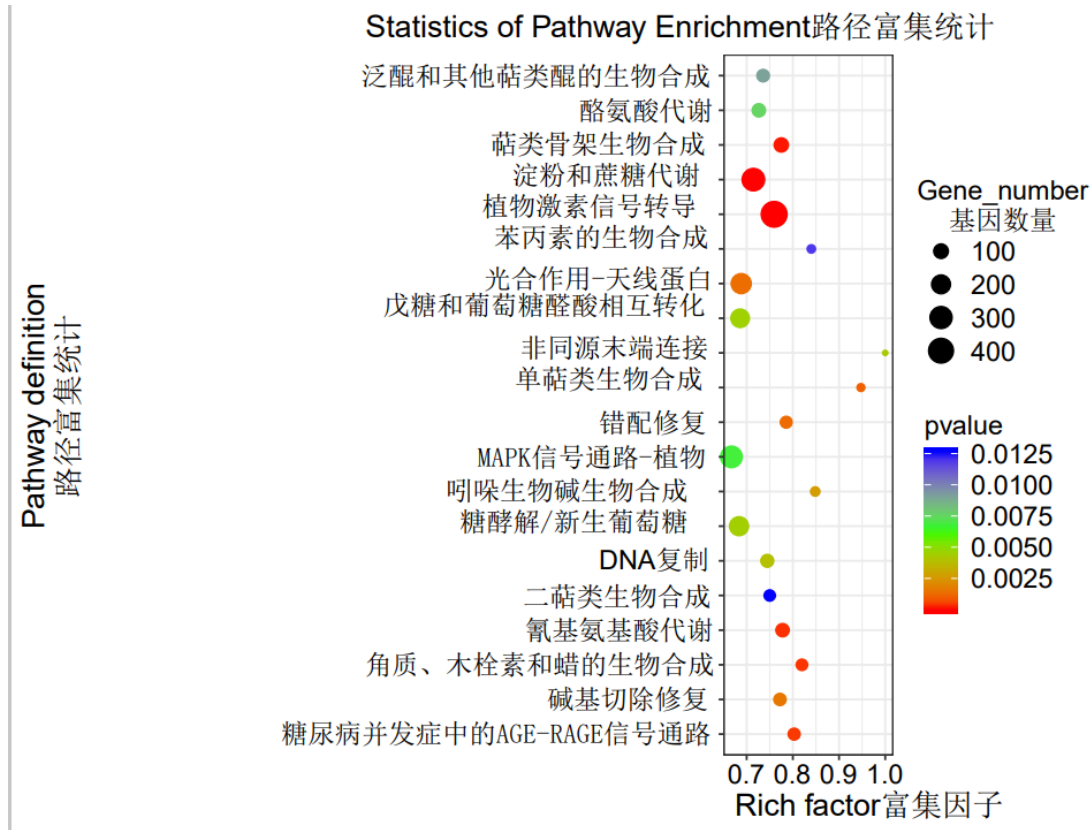


图 6 代谢通路 KEGG 富集分析

2.6 普通油茶授粉花差异基因表达筛选

在 KEGG 的分析基础上，进一步分析植物激素信号转导以及淀粉和蔗糖代谢途径，并以其为筛查对象。在植物激素信号转导途径中发现有 404 个差异基因富集，并进一步根据基因的表达倍数和表达量分别筛选在授粉前和授粉后各组中前 10 个表达量较高的差异基因(表 3)。根据统计结果显示，差异基因 SPL8、AHP1、SERK2、SAUR67、CXE16、SAUR21、EIN3、TGAL4、ARR6 在授粉 2-3 d (FR1) 时，表达量最高，之后表达量逐渐下降，差异基因 TGA2.2 在花未开 (FL1) 时，表达量最高，随后也是逐渐下降。推测在油茶花授粉过程中，这些基因相互作用并调节植物激素并发生极其活跃和复杂的变化。

在淀粉和蔗糖代谢途径中有 313 个差异基因富集，根据基因的表达倍数和表达量筛选在授粉前和授粉后各组中前 10 个表达量较高的差异基因 (表 4)。根据统计结果显示，差异基因 BGLU40、SBEI、TRE1、Os10g0521000、exgA 在授粉 2-3 d (FR1) 时，表达量均显著升高，授粉 7 d (FR2) 后表达量急剧下降，授粉 20 d (FR3) 表达量升高；差异基因 GLC1 和 HXK2 在授粉 7 d (FR2) 时表达量最高，在授粉 20 d (FR3) 后表达量急剧下降；差异基因 CEL3 和 BG5 在花未开 (FL1) 时表达较高，之后下降，到授粉 20 d (FR3) 时表达量最高。差异基因 UGP1 在授粉 2-3 d (FR1) 时表达量较高，之后下降，到授粉 20 d (FR3) 时表达量最高。

表 3 普通油茶授粉柱头显著差异表达基因（植物激素信号转导途径）

基因 ID	基因名称	功能描述	表达量 (TPM)			
			FL1	FR1	FR2	FR3
TRINITY_DN20136_c0_g3	SPL8	角鲨启动子结合样蛋白	9892.86	21790.9	909.63	1241.66
TRINITY_DN24743_c0_g2	AHP1	含组氨酸的磷酸转移蛋白	3580.48	5920.87	691.46	724.34
TRINITY_DN20402_c0_g1	SERK2	类脑脊髓炎不敏感 1 相关受体激酶 1 亚型 X2[小粒咖啡]	2955.88	3107.36	592.4	567.34
TRINITY_DN27434_c1_g3	SAUR67	生长素早期反应蛋白 SAUR68	1720.29	2762.93	581.41	541.78
TRINITY_DN21192_c0_g1	CXE16	羧酸酯酶	1381.46	1599.93	453.82	443.12
TRINITY_DN23658_c0_g6	SAUR21	生长素反应蛋白	1354.88	1465.43	444.24	443.12
TRINITY_DN24791_c0_g2	EIN3	乙烯不敏感的 3 样 1 蛋白[葡萄]	1202.89	1299.08	435.31	426.89
TRINITY_DN20569_c0_g6	TGAL4	转录因子 TGA9 样异构体 X3[小粒咖啡]	1062.48	1276.05	426.28	417.38
TRINITY_DN20647_c0_g2	ARR6	双组分反应调节剂, 如[中华猕猴桃]	1031.03	1192.01	426.28	396.58
TRINITY_DN28221_c0_g1	TGA2.2	转录因子样[中华猕猴桃]	1031.03	963.42	424.01	396.58

表 4 普通油茶授粉柱头显著差异表达基因（淀粉和蔗糖代谢途径）

基因 ID	基因名称	功能描述	表达量			
			FL1	FR1	FR2	FR3
TRINITY_DN15737_c0_g1	GLC1	葡聚糖内-1,3-β-葡糖苷酶	313.01	371.31	1214.44	595.11
TRINITY_DN17068_c0_g2	HXK2	己糖激酶样	224.31	267.79	583.98	321.91
TRINITY_DN17744_c0_g1	UGP1	UTP—葡萄糖-1-磷酸尿苷基转移酶	197.34	219.61	198.62	227.82
TRINITY_DN18130_c0_g1	BGLU40	β-葡萄糖苷酶 6 GH1 家族	162.10	214.76	178.20	186.19
TRINITY_DN18298_c0_g1	SBEI	1,4-α-葡聚糖分支酶 1/淀粉样[中华猕猴桃]	151.02	204.11	161.79	171.42
TRINITY_DN18336_c0_g1	TRE1	海藻糖酶	141.65	189.68	159.29	170.89
TRINITY_DN18354_c0_g1	Osl0g0521000	海藻糖酶	136.36	176.49	128.08	155.20
TRINITY_DN18484_c0_g1	exgA	葡聚糖 1,3-β-葡糖苷酶 A 前体, 部分[中华猕猴桃]	134.11	170.89	124.15	153.79
TRINITY_DN18557_c0_g1	CEL3	内切葡聚糖酶 9-样	132.84	112.20	112.15	146.72
TRINITY_DN18585_c0_g2	BG5	葡聚糖内-1,3-β-葡糖苷酶 BG4 前体[中华猕猴桃]	124.95	111.09	109.24	146.70

3 结论与讨论

利用 Illumina 测序技术对样本进行转录组测序及相关生物学信息分析, 共获得 154264114 条 clean reads, 组装成 86602 个 Unigene, 并通过普通油茶 Unigenes 序列与多个数据库进行比对, 对普通油茶 Unigenes 进行功能注释, 然而还存在一些 Unigenes 序列没被注释到相关信息, 这可能与所获得的 Unigenes 片段断裂, 相关测序数据注释信息欠缺以及从普通油茶转录组获得新基因等因素有关联^[16]。

通过 GO 数据库与普通油茶 Unigenes 序列进行功能注释以及基因生物学特征功能分类, 多达 191767 个 Unigenes 序列注释到生物学过程、分子功能、细胞组分等三大类相关信息。其中生物学过程 (Biological Process) 占比最多主要有生物过程、DNA 转录调控、DNA 转录、蛋白质磷酸化、氧化还原过程、防御反应, 本研究筛选的差异表达基因的 GO 功能分类结果与前人油菜授粉后种子研究较为相似^[17], 说明这些差异基因的富集在油茶花授粉过程中起了重要作用。油茶花授粉过程中, 花粉管萌发, 穿入柱头通向子房^[18], 大量代谢过程开始启动, 参与碳水化合物代谢的相关表达基因逐渐递增, 提供能量^[19]。本研究通过对差异表达基因分析可知, 大部分差异表达基因主要富集在植物激素信号转导, 其次是淀粉和蔗糖代谢。

油茶花授粉时, 需要一些基因来启动, 本研究结果表明植物激素信号转导的差异表达基因 SPL8、AHP1、SERK2、SAUR67、CXE16、SAUR21、EIN3、TGAL4、ARR6 在授粉 2-3 d 时表达量上升, 在授粉 7 d 时表达量急剧下降, 说明这些基因参与响应花粉刺激的分子过程。有研究表明油茶开花前后 2 d 是最佳授粉期^[20], 开花后 6 d 柱头开始枯萎^[21], 这和本研究结果一致。SPL8、SERK2 在花药发育中起着非常重要的作用已在杨柳^[22]、梁婉^[23]等人研究结果中得到证实, 此外, 还有研究表明 SPL8 基因在花药内的特定位置正确启动微孢子囊的形成, 促进孢子母细胞有规律地进入减数分裂^[24]。在拟南芥中, 细胞分裂素 (CK) 信号主要由组氨酸激酶 (AHKs), 组氨酸磷酸转移蛋白 (AHPs) 和应答调节因子 (ARRs) 介导^[25]。细胞分裂素参与配子发生、胚胎发生和种子发育的调节^[26], 因此, 推测本研究含组氨酸的磷酸转移蛋白基因 AHP1 在油茶授粉生殖过程中具有重要作用。SERK2 作为类固醇激素油菜素类固醇 (BRS) 信号组分是水稻中 BR 操作的潜在有用候选者^[27], 其被声称参与植物细胞胚胎发生能力的获得以及调节营养和生殖发育^[28]。李兴国^[29]等人研究结果表明细胞分裂素途径中的关键基因 ARR6 能够调节花发育。还有研究表明由 CK11, AHP 和 B 型 ARR 组成的遗传途径, 用于调节雌配子体发育和营养生长^[30]。SAUR67 是编码促进下胚轴细胞扩增的蛋白质的生长素响应基因^[31], 小生长素向上 RNA (SAUR21) 在茶叶插条的不定根 (AR) 形成中发挥作用^[32], EIN3 能够调控拟南芥种子休眠^[33], 这些基因以及 CXE16 和 TGAL4 在油茶生殖过程中的功能还未见报道。

油茶属于异花授粉植物, 需要通过大量分泌花蜜来吸引昆虫为其授粉。因此, 在油茶授粉时, 柱头外

面的表皮细胞会分泌含多种酶的糖质柱头液,有利于接受和选择花粉^[34]。本研究在淀粉和蔗糖代谢途径中发现差异基因 BGLU40、SBEI、TRE1、Os10g0521000、exgA 在授粉 2-3 天 (FR1) 时表达量最高,暗示着这些基因主要参与花蜜合成与分泌过程中的一些糖类以及相应酶催化过程,使得开花 2 天时花蜜糖浓度最高,柱头可授性也最高^[35]。这和李红姣等人^[36]对油茶花蜜规律的研究结果一致。张睿^[37]研究认为基因 BGLU40 参与酚类代谢调控,和本研究结果不同。海藻糖基因 TRE1、Os10g0521000 是拟南芥中的一种内源性物质,在调节植物中的碳水化合物分配中发挥作用^[38],海藻糖在含有花粉的花药中含量特别高,但它并不局限于这种组织^[39]。

本研究利用高通量技术分析普通油茶开花及四个授粉时期的转录组特征,分析授粉过程涉及的相关通路及关键表达基因,植物激素信号转导以及粉和蔗糖代谢通路可能是油茶授粉的关键调节通路,这些基因和通路在油茶异花授粉过程中有重要调控作用。

参考文献:

- [1] 庄瑞林. 中国油茶[M]. 北京: 中国林业出版社, 2008.
- [2] 陈永忠,邓绍宏,陈隆升,等. 油茶产业发展新论[J]. 南京林业大学学报(自然科学版), 2020, 44(1): 1-10.
- [3] 田惠桥, 朱学艺. 被子植物生殖生物学[M]. 北京:科学出版社, 2012.
- [4] 何汉杏, 康文星, 何秀春. 普通油茶及其优树生殖生态研究[J]. 经济林研究, 2002(4): 10-13.
- [5] 严学成. 油茶的花芽分化[J]. 华南农业大学学报(自然科学版), 1980.
- [6] 查小华, 江国强, 戴义, 等. 油茶结果枝与成花相关性研究[J]. 江西林业科技, 2012(2): 12-14.
- [7] 国家林业局国有林场和林木种苗工作总站. 中国油茶品种志[M]. 北京: 中国林业出版社, 2016.
- [8] 王跃, 毛开云, 王恒哲, 等. 转录组学测序技术应用与市场分析[J]. 生物产业技术, 2017(5): 11-17.
- [9] 王云生, 聂飞, 林顺权. 果树高通量测序的最新研究进展[J]. 基因组学与应用生物学, 2015, 34(9): 2034-2043.
- [10] Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet Journal* 17.
- [11] Grabherr, M. G., et al. (2011). "Full-length transcriptome assembly from RNA-seq data without a reference genome." ("A Bayesian approach for accurate de novo transcriptome assembly - Nature") *Nature Biotechnology*, 29(7), 644.
- [12] Buchfink, B., et al. (2015). Fast and sensitive protein alignment using diamond. *Nature Methods*, 12(1), 59.
- [13] Patro, R., et al. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417.
- [14] Mortazavi, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods*, 5(7), 621.
- [15] Robinson, M. D., et al. (2010). EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139.
- [16] 贾新平, 孙晓波, 邓衍明, 等. 鸟巢蕨转录组高通量测序及分析[J]. 园艺学报, 2014, 41(11): 2329-2341.
- [17] 王晓丹, 肖钢, 常涛, 等. 高油酸油菜脂肪酸代谢的蛋白质组学与转录组学关联分析[J]. 华北农学报, 2017, 32(6): 31-36.
- [18] 邹锋, 谭晓风, 袁德义, 等. 我国油茶生殖生物学研究进展[J]. 长江大学学报(自然科学版)农学卷, 2009, 6(1): 18-20+26.
- [19] 冷容, 胡彦婷, 周艳丽, 等. 人参种胚不同后熟发育阶段比较转录组学分析[J]. 分子植物育种, 2020, 18(17): 5638-5649.
- [20] 张琰. 六个油茶品种的授粉配置研究[D]. 武汉:华中农业大学, 2020.

- [21] 于钊妍, 刘艳菊, 徐玉芬, 等. 油茶生殖生物学及杂交育种研究进展[J]. 热带农业科学, 2022, 42(11): 44-49.
- [22] 杨柳. 拟南芥 SPL8 同源基因的克隆和功能分析[D]. 山西大学, 2018.
- [23] 梁婉. 玉米 LRR 家族受体激酶的鉴定与功能分析[D]. 兰州大学, 2017.
- [24] Unte, Ulrike S et al. "SPL8, an SBP-box gene that affects pollen sac development in Arabidopsis." *The Plant cell* vol. 15, 4 (2003): 1009-19.
- [25] Abdelrahman, Mostafa et al. "Defective cytokinin signaling reprograms lipid and flavonoid gene-to-metabolite networks to mitigate high salinity in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* vol. 118, 48 (2021): e2105021118.
- [26] Werner, T. & Schmuelling, T. Cytokinin action in plant development. *Curr. Opin. Plant Biol.* 12, 527–538 (2009).
- [27] Dong, Nana et al. "Regulation of Brassinosteroid Signaling and Salt Resistance by SERK2 and Dec. 2020.
- [28] Albrecht, Catherine et al. "The Arabidopsis thaliana SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis." *The Plant cell* vol. 17, 12 (2005): 3337-49.
- [29] 李兴国. 细胞分裂素调节拟南芥花发育的研究[D]. 山东农业大学, 2008.
- [30] Deng, Yan et al. "Arabidopsis histidine kinase CK11 acts upstream of histidine phosphotransfer proteins to regulate female gametophyte development and vegetative growth." *The Plant cell* vol. 22, 4 (2010): 1232-48.
- [31] Lorrai, Riccardo et al. "Genome-wide RNA-seq analysis indicates that the DAG1 transcription factor promotes hypocotyl elongation acting on ABA, ethylene and auxin signaling." *Scientific reports* vol. 8, 1 15895. 26 Oct. 2018.
- [32] Wang, Yongxin et al. "Integrated transcriptome and hormonal analysis of naphthalene acetic acid-induced adventitious root formation of tea cuttings (*Camellia sinensis*)." *BMC plant biology* vol. 22, 1 319. 4 Jul. 2022.
- [33] 忽钰哲. EIN3 通过 DOG1 调控拟南芥种子休眠的分子机制研究[D]. 华中农业大学, 2022.
- [34] 何汉杏, 康文星, 何秀春. 普通油茶及其优树生殖生态研究[J]. 经济林研究, 2002(4): 10-13.
- [35] 李红娇, 董坤, 韩明跃, 等. 油茶泌蜜规律及花粉活力和柱头可授性的变化[J]. 广西林业科学, 2012, 41(4): 327-330+355.
- [36] 李红娇, 董坤, 韩明跃, 等. 油茶泌蜜规律及花粉活力和柱头可授性的变化[J]. 广西林业科学, 2012, 41(4): 327-330+355.
- [37] 张睿. 铁核桃果实发育过程中酚类代谢的转录组分析[D]. 贵州大学, 2019.
- [38] Müller, J et al. "Trehalose and trehalase in Arabidopsis." *Plant physiology* vol. 125, 2 (2001): 1086-93.
- [39] Müller J, Boller T, Wiemken A (1995a) Trehalose and trehalase in plants: recent developments. *Plant Sci* 112:1-9.