

Molecular and biochemical analyses of avocado (*Persea americana*) reveal differences in the oil accumulation pattern between the mesocarp and seed during the fruit developmental period

Yu Ge (✉ geyu@catas.cn)

Chinese Academy of Tropical Agricultural Sciences <https://orcid.org/0000-0001-7153-5035>

Xiangshu Dong

Yunnan University

Yuanzheng Liu

Chinese Academy of Tropical Agricultural Sciences

Ying Yang

Chinese Academy of Tropical Agricultural Science

Rulin Zhan

Chinese Academy of Tropical Agricultural Sciences

Research

Keywords: Avocado, Industrial oils, Oil biosynthesis, Tissue-specific, RNA-seq, Transcription factor, Long non-coding RNA

Posted Date: April 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20746/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientia Horticulturae on January 1st, 2021. See the published version at <https://doi.org/10.1016/j.scienta.2020.109717>.

Abstract

Background: The avocado (*Persea americana*) mesocarp and seed contain high-value oil with broad industrial applications. The oil contents in these two tissues vary considerably at maturity. Additionally, the molecular mechanism underlying the tissue-specific oil accumulation in the developing avocado mesocarp and seed remains unclear, which has hampered the exploration of the utility of avocado for oil production.

Results: To clarify the mechanisms mediating the differences in oil contents and fatty acid compositions, the transcriptomes and oil bodies were compared between the oil-storing tissues during the fruit developmental period. The results revealed the increasing and fluctuating trends in the oil accumulation in the developing avocado mesocarp and seed, respectively. Additionally, striking differences in the lipid droplets between the mature mesocarp and seed were revealed in confocal microscopy images. Subsequently, the gene transcription profiles of the developing mesocarp and seed were characterized via a comprehensive transcriptome analysis involving second-generation sequencing and single-molecule real-time sequencing techniques. The tissue-specific transcription of lipid-related genes contributing to fatty acid synthesis, triacylglycerol assembly, and triacylglycerol storage was examined, with most of the lipid-related genes expressed at higher levels in the developing mesocarp than in the developing seed. A weighted gene co-expression network analysis uncovered 291 transcription factors that were commonly or uniquely correlated with the oil contents in the avocado mesocarp and seed. Moreover, 11 trans-acting and 79 cis-acting long non-coding RNAs were identified as common or unique to the developing avocado mesocarp and seed. These long non-coding RNAs may regulate the expression of 43 lipid-related genes. Finally, a network of genes associated with oil accumulation in the developing avocado mesocarp and seed was established.

Conclusions: The results of this study further elucidate the tissue-specific oil biosynthesis and related regulatory network in the avocado mesocarp and seed. Furthermore, tissue-specific lipid-related genes, putative transcription factors, and putative long non-coding RNAs affecting oil accumulation were identified. Our data may also be useful for characterizing tissue-specific oil accumulation at the transcriptomic level, thereby identifying candidate genes for improving the oil production of related plant species.

Background

Vegetable oils have traditionally been used primarily in the food industry, but they are increasingly being used for other applications, including those related to pharmaceuticals and cosmetics [1]. There has recently been growing interest in the utility of vegetable oils as phytochemical alternatives to the synthetic substances commonly used in foods, pharmaceuticals, and cosmetics. This effort is supported by consumer concerns over the safety of products containing synthetic chemicals, which are believed to cause or increase the severity of health problems and pollute the environment [2, 3]. Avocado (*Persea americana* Mill.) is native to Meso-America, and is considered to be an important and economically

valuable tropical and subtropical woody oil-bearing crop [4]. Regarding oil content, avocado fruit is exceeded only by the fruits of palm and olive trees [5]. The relatively high proportion of unsaturated fatty acids in avocado fruit oil [6, 7] is associated with superior skin permeability and sunscreen performance, leading to the use of avocado oil in the cosmetics industry worldwide as a natural ingredient replacing artificially synthesized chemicals [8, 9]. Furthermore, the oil extracted from avocado seeds may be used in the biofuel industry as an alternative biodiesel source instead of the conventional petroleum-based diesel fuel [5, 10].

The lipid biosynthesis process in oil-rich plant tissues mainly involves the following two stages: fatty acid (FA) biosynthesis in plastids and triacylglycerol (TAG) assembly in the endoplasmic reticulum (ER) [11, 12]. The lipid biosynthesis pathway has been relatively well elucidated, and the associated genes and enzymes have been identified and in many cases, functionally validated. However, the characterized molecular mechanism is mostly limited to oil-producing seeds rather than other oil-producing tissues [2, 13–15]. In nature, there are diverse plant species that accumulate oil in non-seed tissues, such as the mesocarp layer of oil palm, olive, avocado, and Chinese tallow (*Triadica sebifera*) [16–18], the outer surface layer of bayberry fruit (*Myrica rubra*) [19], the stolon tuber of yellow nutsedge (*Cyperus esculentus*) [20], the stem tissue of oil firewood (*Tetraena mongolica*) [21], and the caruncle of castor bean [3]. A few recent investigations examining the oil palm mesocarp, olive mesocarp, castor caruncle, and nutsedge tuber tissues revealed key differences in the transcriptional control of lipid biosynthesis between nonseed and seed tissues [3, 18, 20, 22]. Avocado, like oil palm and olive, is one of the few plant species in which the mesocarp, a nonseed tissue, and the seed accumulate substantial amounts of oil [7]. Thus, avocado may be a model plant species for studying the differences in the tissue-specific molecular mechanisms underlying the synthesis of oil in the mesocarp and seed. However, the differences in the expression patterns of genes associated with oil accumulation between the developing avocado mesocarp and seed are relatively unknown.

The fact that most genes governing FA biosynthesis share the same temporal transcription pattern strongly suggests that they are coregulated and probably share common cis- and trans-regulatory elements [23]. Previous studies confirmed that WRINKLED1 (*WRI1*) is essential for the transcriptional regulation of FA biosynthesis because it controls the expression of key genes in the fatty acid biosynthetic pathways in oil-rich seed and non-seed tissues [7, 11, 18, 24–28]. Additionally, LEAFY COTYLEDON 1 (*LEC1*), LEAFY COTYLEDON 2 (*LEC2*), FUSCA 3 (*FUS3*), and ABSCISIC ACID INSENSITIVE 3 (*ABI3*) are transcriptional master regulators that mediate storage oil biosynthesis in seed oil-producing crops, with *LEC1* and *LEC2* directly regulating *WRI1* expression in *Arabidopsis thaliana* (*Arabidopsis*) [18, 29, 30]. However, a set of new transcription factors, EgNF-YA3, EgNF-YC2, and EgABI5, reportedly modulate the transcription of oil biosynthetic pathway genes in the oil palm mesocarp [18]. Avocado accumulates considerably more oil in its mesocarp than in its seed [31], suggesting the existence of a tissue-specific fatty acid metabolism transcriptional network.

Transcriptomes generated by second-generation sequencing (SGS) technology platforms have broadly been applied to illuminate or reconstruct several important biochemical pathways [32–35]. However, the

short transcript sequences generated by SGS techniques decrease the accuracy of sequence assembly and hinder sequence analyses of genes and enzymes involved in downstream metabolic pathways [36, 37]. Single-molecule real-time (SMRT) technology, which is a third-generation sequencing technology [38] that can produce much longer reads than SGS platforms, has been applied to effectively capture full-length transcript sequences [39]. Combining Illumina and SMRT technologies has recently been confirmed as an effective strategy for sequencing and analyzing the transcriptomes of several crops to clarify the transcriptomic complexity or the genes contributing to important biochemical pathways [40–44]. Moreover, integrated analyses of the de novo assembly involving short-read and full-length transcriptome sequences and metabolomics data may enable the identification of the genes underlying a particular phenotypic trait and provide researchers with information regarding gene-to-metabolite networks for species of interest [2, 28, 45].

Long non-coding RNAs (lncRNAs) are potentially important gene regulators in eukaryotic cells, especially in some key biological processes [46]. There remains some ambiguity surrounding their numbers and characteristics in genomes [47]. Predicting and functionally annotating lncRNAs is generally challenging, but useful, because they are not orthologous and there is a lack of homologous sequences between closely related species [48]. However, the functions of very few lncRNAs have been elucidated [49, 50]. Hence, the lncRNA information for one species is not necessarily suitable for predicting lncRNAs in other species. To date, the functions of lncRNAs in avocado mesocarp and seed developmental processes, especially in terms of oil accumulation, have been largely unexplored.

In this study, we first applied Illumina HiSeq and PacBio sequencing technologies to analyze the lipid biosynthesis in the avocado mesocarp and seed. A comparison of the transcriptomes, FA profiles, and lipid droplet histological characteristics of the two oil-storing tissues during the fruit developmental period provided insights into the tissue-specific molecular mechanisms controlling the variations in oil content and composition. Second, we conducted a weighted gene co-expression network analysis (WGCNA) to identify key transcription factors (TFs) associated with oil contents based on the transcriptional data for the developing avocado mesocarp and seed. Third, lncRNAs in the two oil-storing tissues were identified by examining genome-guided PacBio sequences, after which lncRNAs associated with genes in the lipid biosynthetic pathways were compared between the two oil-storing tissues. The results of this study provide insights into tissue-specific oil biosynthetic networks in the two oil-storing avocado tissues, and may broaden our understanding of the synthesis of oils in plants.

Results

Oil accumulation in the developing avocado mesocarp and seed

Avocado fruit is rich in oil, which is mainly stored in the mesocarp and seed. The fruit developmental period has been divided into five stages based on the days after full bloom (DAFB) [51], with the fruit considered to be physiologically mature at 215 DAFB. The oil content in the mesocarp continued to

accumulate throughout the fruit developmental period (Fig. 1a). From 75 to 110 DAFB, the mesocarp oil content more than quadrupled from 0.64 to 2.52 g/100 g fresh weight (FW), after which it accumulated slowly from 110 to 180 DAFB. Finally, the mesocarp oil content increased by almost 3-fold from 180 to 215 DAFB, peaking at 15.20 g/100 g FW at 215 DAFB. In contrast, the seed oil content was much lower than the mesocarp oil content at each of the five fruit developmental stages, and it fluctuated slightly throughout the analyzed period (Fig. 1a).

The fatty acid composition also varied between the mesocarp and seed during the fruit developmental period (Fig. 1b; Additional file 1: Table S1). Among the major fatty acids, palmitic (C16:0) or linoleic acid (C18:2) were the most abundant in different mesocarp developmental stages, whereas C18:2 was consistently the predominant fatty acid in the seed (Fig. 1b). The C16:0 and C18:2 compositions in the mesocarp changed substantially during the mid-to-late developmental stages, with a steady increase in the palmitoleic acid (C16:1) level and a decrease in the linolenic acid (C18:3) level (Fig. 1b). In the seed, there were almost no variations in the composition during the developmental period, and unlike the mesocarp, the seed contained a higher proportion of C18:2 and a lower proportion of C16:1, C18:0, and C18:3 (Fig. 1b). The striking difference in the fatty acid profiles between the mesocarp and seed might suggest the existence of tissue-specific fatty acid metabolism transcriptional networks. The complete ion chromatograms of fatty acid methyl esters of the avocado mesocarp and seed at 215 DAFB are provided in Additional file 2: Figure S1.

Histological analyses of lipid droplets in the mature avocado mesocarp and seed

The lipid droplets in the mature avocado mesocarp and seed at 215 DAFB underwent a histological analysis. A mass of large lipid droplets (approximately 10–25 μm in diameter; Fig. 2a) formed in the mature mesocarp, occupying most of the cell volume when the total lipid content peaked at 215 DAFB (Fig. 1a). In contrast, a few small lipid droplets (approximately 2 μm in diameter; Fig. 2b) were detected at the edge of the cell wall in the avocado seed.

Transcriptomic analysis via SMRT sequencing and de novo transcript assembly based on short reads in the two analyzed oil-producing tissues

To comprehensively characterize the gene expression dynamics, the transcriptomes of the developing avocado mesocarp and seed were analyzed based on the avocado genome sequence [4]. The analysis included 30 de novo transcript assemblies from paired-end Illumina RNA-seq reads and full-length transcripts for the two tissues derived from PacBio SMRT sequencing. After filtering for quality, 1,428 million clean paired-end reads were generated, with an average of 84.36% reads mapped to the avocado reference genome (Additional file 3: Table S2). The SMRT sequencing resulted in 512,163 and 421,641 circular consensus reads for the avocado mesocarp and seed, with a mean length of 2,088 and

2015 bp, respectively. These reads included 448,992 and 364,137 full-length non-chimeric (FLNC) reads for the avocado mesocarp and seed, respectively. Additionally, 130,999 and 100,927 consensus isoforms were acquired for the avocado mesocarp and seed, with a mean length of 2,016, and 1,954 bp, respectively, based on the clustering of FLNC sequences. After polishing consensus sequences, 125,335 and 92,951 high-quality consensus sequences were obtained for the avocado mesocarp and seed, respectively.

Because the sequencing depth and accuracy are greater for the Illumina platform than for the SMRT platform, the Illumina RNA-seq reads were used to improve the full-length transcript quality and to determine the gene expression levels at the five fruit developmental stages. After eliminating redundancies with CD-HIT, 60,374 and 47,907 nonredundant transcripts were retained. The SMRT sequencing raw data are available in the GenBank database (avocado mesocarp: accession number PRJNA551932; avocado seed: accession number PRJNA559779). The Illumina RNA-Seq raw data are also available in the GenBank database (accession number PRJNA541745).

Tissue-specific transcription of lipid-related genes involved in FA synthesis and TAG assembly

Thirty-two unigenes affecting plastid FA biosynthesis and 24 unigenes influencing TAG assembly in the ER were identified (Additional file 4: Table S3). Most of the known lipid-related enzymes were detectable in the avocado mesocarp and seed. A hierarchical cluster analysis was performed based on the normalized fragments per kilobase of transcript per million mapped reads (FPKM) values of the transcripts in the mesocarp and seed at five fruit developmental stages. The results indicated that all differentially expressed lipid-related genes contributing to FA synthesis and TAG assembly in the avocado mesocarp and seed belonged to two clusters (Fig. 3). Regarding the avocado mesocarp, Cluster I comprised 40 unigenes whose expression levels tended to slowly increase (Fig. 3a). These included 27 and 13 unigenes related to the FA biosynthesis pathway and TAG assembly, respectively. Cluster II consisted of 5 and 11 unigenes involved in the FA biosynthesis pathway and TAG assembly, respectively, and expression levels tended to slowly decrease (Fig. 3a). In terms of the avocado seed, Cluster I included 12 and 6 unigenes related to TAG assembly and FA synthesis, respectively, with expression levels that increased slightly (Fig. 3b). Cluster II comprised 26 and 12 unigenes affecting FA synthesis and TAG assembly, respectively, with decreasing expression levels (Fig. 3b).

Most of the lipid-related genes contributing to FA synthesis and TAG assembly were more highly transcribed in the mesocarp than in the seed at more than four fruit developmental stages (Fig. 4; Additional file 4: Table S3). We observed that 32 unigenes related to FA synthesis were expressed in both the avocado mesocarp and seed, with average transcript levels of 173.31 and 64.23 FPKM/stage, respectively, whereas 24 unigenes affecting TAG assembly were expressed in the mesocarp and seed, with average transcript levels of 38.05 and 16.05 FPKM/stage, respectively. However, 14 unigenes were more highly transcribed in the developing seed than in the developing mesocarp, although the average

transcript levels of most of these 14 unigenes were less than 30 FPKM/stage (Fig. 4; Additional file 4: Table S3).

Transcriptional divergence of TAG storage genes between two oil-storing avocado tissues

One oil-body oleosin (OBO), one steroleosin (STERO), and two caleosin (CALO) unigenes were expressed in the avocado mesocarp and seed (Fig. 4). These four unigenes were more highly transcribed in the seed than in the developing mesocarp. Specifically, OBO expression was almost undetectable in the developing mesocarp, whereas two CALO paralogs had average transcript levels exceeding 120 FPKM/stage in the developing mesocarp (Additional file 5 Table S4). Additionally, two lipid droplet-associated protein 2 (LDAP2) paralogs were abundantly expressed in the developing mesocarp and seed, with LDAP2-1 more highly transcribed than the other three types of lipid droplet-associated genes in the developing mesocarp (Fig. 4; Additional file 5: Table S4).

Tissue-specific transcriptional regulation of transcription factors

Of the 8,113 putative TFs differentially expressed in the developing mesocarp and seed, 350 and 464 TFs were unique to the mesocarp and seed, respectively (Additional file 6: Figure S2). Moreover, these 8,113 putative TFs were distributed in 213 families, but mainly in the bHLH (268), RLK-Pelle_DLSV (237), C3H (235), C3H (203), and NAC (203) families (Additional file 7: Table S5).

To determine which TFs play pivotal roles in seed oil accumulation in the developing avocado mesocarp and seed, we conducted a WGCNA examining differentially expressed TFs and the oil contents of the developing avocado mesocarp and seed. A total of 291 TFs were commonly or uniquely significantly correlated with the oil contents in the avocado mesocarp and seed (Fig. 5a–c). These TFs belong to 69 families each in the mesocarp and seed, with the most abundant TFs identified as AP2/ERF-ERF (9), WRKY (7), and bHLH (6) in the mesocarp as well as bHLH (7) and MYB (6) in the seed (Additional file 8: Table S6). Additionally, some well-known TF genes that regulate lipid biosynthesis were also identified by the co-expression analysis, including PaWRI1 in the mesocarp and PaABI3 in the seed (Fig. 5a, b; Additional file 8: Table S6). The results of the gene ontology (GO) enrichment analysis indicated that the TFs correlated with the oil content were associated with transcription factor activity, sequence-specific DNA binding, and protein serine/threonine kinase activity in the mesocarp (Fig. 5d) as well as protein serine/threonine kinase activity and regulation of transcription (DNA-templated) in the seed (Fig. 5f). Furthermore, plant hormone signal transduction pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database were enriched among the lipid-related TFs in the avocado mesocarp and seed (Fig. 5e, g). Some hub TFs were identified in the gene co-expression networks for the mesocarp and seed (Fig. 5a, b; Additional file 8: Table S6). These hub TFs were most highly co-expressed with other TFs associated with lipids, implying they may contribute to oil biosynthesis.

Variations in lncRNAs between two oil-rich avocado tissues

Of the 5,299 lncRNAs differentially expressed between the developing avocado mesocarp and seed, 198 and 271 were unique to the mesocarp and seed, respectively (Additional file 9: Table S7; Additional file 10: Figure S3a). Additionally, on the basis of lncRNA–mRNA pairs 1,608 trans-acting lncRNAs were identified, of which 120 and 232 were unique to the developing avocado mesocarp and seed, respectively (Additional file 10: Figure S3b). Moreover, 4,657 cis-acting lncRNAs were revealed, including 174 and 215 lncRNAs that were unique to the developing mesocarp and seed, respectively (Additional file 10: Figure S3c). Finally, the results suggested that 11 trans-acting lncRNAs and 79 cis-acting lncRNAs help regulate the expression levels of 43 lipid-related genes involved in FA synthesis, TAG assembly, and TAG storage in the developing avocado mesocarp and seed (Additional file 11: Table S8).

Validation of candidate differentially expressed genes involved in lipid metabolism

The relative expression levels and temporal transcription patterns were calculated for the key genes identified based on the oil-related transcriptome sequencing data. Specifically, 10 genes, including six lipid-related genes, two TF genes, and two lncRNA genes, were analyzed. The $2^{-\Delta\Delta Ct}$ values of the selected genes were generally consistent with the RNA sequencing results (Additional file 12: Figure S3). Therefore, the quantitative real-time (qRT)-PCR data confirmed the validity of the transcriptomic data.

Discussion

In this study, the total lipid accumulation patterns in the avocado mesocarp and seed tended to increase and fluctuate slightly, respectively, during the fruit developmental period (Fig. 1a). These findings were consistent with the results of previous avocado studies [6, 7]. However, in contrast to these earlier investigations, several studies involving oil palm have confirmed that the total lipid contents in three oil-storing tissues (mesocarp, embryo, and endosperm) exhibit similar increasing trends in developing fruits [11, 12, 25]. Our histological results suggest that the total lipid content in avocado is generally directly proportional to the volume of the lipid droplets in the mesocarp and seed (Fig. 2). Several previous studies proved that the lipid droplets of the oil palm mesocarp and endosperm are relatively large (approximately 5–20 μm in diameter) and occupy most of the cellular volume at harvest, whereas those of the oil palm embryo are smaller (approximately 2 μm in diameter) and are located at the cell periphery at harvest [11, 12, 25]. Similarly, the total lipid contents of three oil-rich tissues of oil palm are consistent with lipid droplet volumes [11, 12, 25].

In this study, the average expression levels of 32 unigenes associated with FA synthesis were 4.56- and 4.00-fold higher than those of 24 unigenes contributing to TAG assembly in the developing mesocarp and seed, respectively (Additional file 4 Table S3). These results suggest that the genes related to FA synthesis may be more important than those related to TAG biosynthesis for the substantial accumulation of oil in avocado fruit tissues. Similar results were reported in previous studies that

examined oil-rich tissues (e.g., mesocarp, seed, and tuber) in diverse oil-bearing crops [2, 7, 11, 20, 52]. Our data further confirm that the temporal transcriptional pattern of lipid-related genes is conserved in various oil-rich tissues of oil-bearing crops.

The conversion of pyruvate to fatty acids in plastids involves at least 13 enzymes and/or protein complexes (Fig. 4). Regarding the four subunits of the pyruvate dehydrogenase complex, each one is encoded by one unigene, of which PaPDH-E1 α represented an average of 40% of the transcripts of these four unigenes in the developing mesocarp (Additional file 4: Table S3). This was similar to the findings of an earlier avocado mesocarp study [7]. In contrast, an average of 43% of the transcripts of these four unigenes corresponded to PDH-E1 β in the developing seed (Additional file 4: Table S3). However, PDH-E1 β is the predominant transcript in three oil-rich tissues of oil palm [25]. Two unigenes encoding three subunits of acetyl-CoA carboxylase were expressed in the avocado mesocarp and seed (Fig. 4), with PaACC-BC expression undetectable in the avocado mesocarp [7]. A comparison of the expression of these six acetyl-CoA carboxylase unigenes revealed that the average PaACC-BC-2 transcript levels, which were 259.92 and 55.74 FPKM/stage in the developing avocado mesocarp and seed, respectively, were greater than the transcript levels of the other five unigenes (Additional file 4: Table S3). In contrast to avocado, ACC-Ct β is more highly transcribed in oil palm than three other unigenes in the mesocarp, embryo, and endosperm [25]. However, in oilseed crops, BCCP2 is predominantly expressed in *Ricinus communis*, *Brassica napus*, and *Tropaeolum majus* seeds [52].

Stearoyl-acyl carrier protein desaturase (SAD) is a precursor for the biosynthesis of polyunsaturated fatty acids and catalyzes the conversion of stearoyl-acyl carrier protein (ACP) to oleoyl-ACP [53]. Seven homologous AtSAD genes were identified in the Arabidopsis genome (AtFAB2, AtDES1, AtDES2, AtDES3, AtDES4, AtDES5, and AtDES6) [54]. In the current study, two homologous PaSAD unigenes, PaFAB2 and PaDES6, were expressed in the developing avocado mesocarp and seed, with PaFAB2 more abundantly transcribed (Fig. 4). The orthologs of FAB2 have often been detected in some oleic acid-rich tissues from nonseed and seed oil crops, in which they are generally more highly expressed than other SAD homologs, implying FAB2 is important for oleic acid synthesis [7, 20, 25, 53]. In contrast to PaFAB2, PaDES6 was mainly transcribed in the developing seed rather than in the developing mesocarp (Fig. 4). Similar studies revealed that SAD homologs usually exhibit tissue-specific expression patterns in oil palm and olive [25, 55]. The EgFAB2 paralogs are expressed in three oil-storing tissues (mesocarp, embryo, and endosperm), whereas EgDES5 is only transcribed in the embryo and endosperm [25]. Seven OeuSAD homologs in the wild olive genome are also differentially expressed among four tissues [55].

The PaSAD and PaACP4 transcript levels were greater than the transcript levels of the other unigenes contributing to the FA synthesis in the plastids of the developing avocado mesocarp and seed, and were generally consistent with the observed oil accumulation, indicative of the key roles for these genes related to avocado oil composition (Fig. 1a, Fig. 4). Similar results were obtained for various oil-storing tissues of other crops [2, 7, 20].

Acyl-ACP thioesterase catalyzes the hydrolysis of the acyl-ACP intermediates during the final step of FA biosynthesis to release free FA [56]. This enzyme has been divided into two families according to distinguishing substrate specificities, namely FatA and FatB, which determine, to some extent, the FA composition of storage lipids in plants [57]. Specifically, FatA exhibits high substrate specificity toward monounsaturated oleoyl-ACP, whereas FatB subfamily members exhibit specificity for palmitoyl- and stearoyl-ACPs [58]. In this study, we detected transcripts for one PaFatA and two PaFatB paralogs in the transcriptomes of the developing mesocarp and seed (Fig. 4); however, all PaFatA and PaFatB genes were expressed in the avocado mesocarp in an earlier investigation [7]. The PaFatA expression level was more than 4-fold (75 DAFB) and 45-fold (215 DAFB) higher in the developing mesocarp than in the developing seed (Additional file 4: Table S3). Similarly, C18:1 contents were more than 6- to 40-times greater in the developing mesocarp than in the developing seed from 110 to 215 DAFB (Additional file 1: Table S1). These results suggest that PaFatA expression is generally positively correlated with C18:1 contents, which is consistent with the results of previous studies [25, 58]. In contrast, the expression levels of two PaFatB paralogs were unrelated to the C16:0 and C18:0 contents in the avocado mesocarp and seed in the present study. The two PaFatB paralogs were more highly expressed in the avocado seed than in the mesocarp during most of the five fruit developmental stages, whereas C16:0 and C18:0 were less abundant in the developing avocado seed than in the developing mesocarp throughout the examined period (Fig. 4; Additional file 1: Table S1). This observation may be explained by the lower substrate specificity of FatB than of FatA [59]. Additionally, the enzymes encoded by two PaFatB paralogs were unable to efficiently hydrolyze 16:0-ACP and 18:0-ACP in the avocado mesocarp and seed, but may have complementary roles that enhance the PaFatA-mediated hydrolysis of 18:1-ACP. The EgFatB in the oil palm embryo and endosperm catalyzes the hydrolysis of 18:1-ACP very efficiently because of a lack of EgFatA [25].

Long-chain acyl-CoA synthetase (LACS) is an enzyme that exports newly synthesized free FAs in the plastid to the ER and converts them to acyl-coenzyme A (acyl-CoA) [60]. In Arabidopsis, LACS is encoded by a small gene family consisting of nine members [61]. In the current study, four homologous LACS genes were expressed in the avocado mesocarp and seed (PaLACS1, PaLACS4, PaLACS8, and PaLACS9) (Fig. 4). Both PaLACS8 and PaLACS9 were relatively abundantly transcribed in the developing avocado mesocarp, with average transcript levels of 119.99 and 87.47 FPKM/stage, respectively, whereas PaLACS4 might mainly contribute to the acyl activation in the avocado seed, with an average transcript level of 105.22 FPKM/stage (Additional file 4: Table S3). Accordingly, ER-associated LACS (PaLACS8) and plastidial LACS (PaLACS9) appear to codetermine the acyl activation from the plastid to the ER in the avocado mesocarp, but ER-associated PaLACS4 is important for the export of FAs between the plastid and ER in the avocado seed. However, a study by Kilaru et al. [7] indicated that LACS9 is most abundantly expressed in the avocado mesocarp. Regarding oil palm, EgLACS9 and two EgLACS4 paralogs are expressed in the oil palm mesocarp, embryo, and endosperm, but the expression of the LACS8 ortholog is undetectable in the oil palm [25]. These three LACS unigenes are expressed in the developing oil palm embryo and endosperm, but EgLACS9 and an EgLACS4 paralog are reportedly also expressed in the developing oil palm mesocarp [25]. Another study determined that CeLACS4 and CeLACS8 account for

approximately 70% of the transcripts of seven LACS unigenes in the oil-rich tuber tissue of *C. esculentus* [20]. In oilseed crops and *Arabidopsis*, LACS9 orthologs are the most abundantly expressed unigenes in the seed [52, 62]. Mutational studies in *Arabidopsis* suggested the enzymes encoded by LACS4 and LACS9 exhibit overlapping functions [62]. Both LACS8 and LACS9 may be similarly associated with plastidial fatty acid export in sunflower seeds [58].

The TAG assembly in the ER requires a series of acylations of glycerol-3-phosphate with acyl-CoA and a subsequent dephosphorylation. In this study, we uncovered at least 12 enzymes involved in TAG assembly. For example, GPAT catalyzes the acylation of glycerol-3-phosphate to yield lysophosphatidic acid during the first step of membrane and storage glycerolipid assembly [63]. To date, nine GPAT gene family members have been identified in *Arabidopsis*. Three GPAT unigenes (one GPAT1 and two GPAT9 paralogs) were expressed in the avocado seed, whereas only two GPAT9 paralogs were transcribed in the avocado mesocarp (Fig. 4; Additional file 4: Table S3). Kilaru et al. [7] also suggested that only GPAT9 is expressed in the avocado mesocarp. Similarly, GPAT9 is reportedly more highly transcribed than the other family members, with GPAT1–8 either unexpressed or expressed at low levels in diverse oil crops [2, 20, 25, 52]. Unlike in other oil crops, the GPAT1 expression level was high in the developing avocado seed, and higher than the expression levels of two GPAT9 paralogs at 215 DAFB (Additional file 4: Table S3).

Among the various unigenes encoding 1-acylglycerol-3-phosphate acyltransferase (LPAAT), LPAAT2 was more highly expressed than LPAAT1 and two LPAAT5 paralogs in the avocado mesocarp and seed (Fig. 4). Similarly, LPAAT2 was identified as the predominantly expressed gene in the avocado mesocarp in a previous study [7]. The LPAAT2 expression levels were similar between the avocado mesocarp and seed; however, LPAAT1 and two LPAAT5 paralogs were transcribed at higher levels in the developing mesocarp than in the developing seed (Additional file 4: Table S3). Two kinds of PP, namely phosphatidate phosphohydrolase (PAH) and lipid phosphate phosphatase (LPP), have functions related to de novo DAG synthesis [64]. In the current study, PaPAH1, two PaPAH2 paralogs, two PaLPP2 paralogs, and two PaLPP3 paralogs were expressed in the avocado mesocarp and seed (Fig. 4). Four of these unigenes, PaPAH1, PaPAH2-1, PaLPP2-1, and PaLPP2-2, were more highly transcribed in the developing mesocarp than in the developing seed, whereas the other three unigenes exhibited the opposite expression pattern (Additional file 4: Table S3). Previous studies implied that LPP rather than PAH is the more probable candidate enzyme responsible for generating DAG [65, 66]. In our study, the average expression levels of four LPP unigenes were 29.40 and 25.23 FPKM/stage in the developing mesocarp and seed, respectively. These expression levels were higher than the average transcript levels of three other LPP unigenes, which were 11.61 and 8.97 FPKM/stage in the developing mesocarp and seed, respectively (Additional file 4: Table S3).

Earlier research confirmed that DGAT, which transfers acyl-CoA to the sn-3 position of diacylglycerol (DAG), is a key rate-limiting enzyme for TAG assembly in the ER [11]. The PaDGAT1 and PaDGAT2 transcript levels were on average 15.71- and 2.32-fold higher in the developing mesocarp than in the developing seed, respectively, which was consistent with the oil content differences observed between the two oil-rich tissues in this study (Fig. 1a; Additional file 4: Table S3). Additionally, PaDGAT2 was

transcribed more abundantly than PaDGAT1 by an average of 1.64- and 6.79-fold in the developing mesocarp and seed, respectively (Additional file 4: Table S3), suggesting DGAT2 is the predominant enzyme synthesizing TAG in the avocado mesocarp and seed. Similarly, the DGAT2 transcript level is more than 3-times greater than that of DGAT1 in the *C. esculentus* tuber [20], but DGAT1 is more highly transcribed in oil seeds and fruits [7, 11, 25].

In addition to the de novo DAG synthesis via the Kennedy pathway, DAG precursors may be transferred from PC to DAG by PDAT to form lyso-PC and TAG [11]. Our data revealed two PaPDAT paralogs and two LPCAT paralogs were expressed in the avocado mesocarp and seed. The two PaPDAT paralogs were more highly transcribed in the avocado seed than in the mesocarp, whereas there were no differences in the transcription of the PaLPCAT paralogs between the avocado mesocarp and seed (Additional file 4: Table S3). Another possible route involves the PC-derived production of DAG as the substrate for TAG synthesis via the reversible activities of two enzymes, phosphatidylcholine:DAG cholinephosphotransferase (PDCT) and/or cytidine-5'-diphosphocholine:DAG cholinephosphotransferase (CPT), with the reversibility of the catalyzed reactions enabling the enrichment of DAG in PUFAs [7]. In our study, the PaCPT transcript level in the developing mesocarp was on average 57.61-fold higher than that in the developing seed. In contrast, the PaPDCT expression level in the developing seed was on average 259-fold higher than that in the developing mesocarp (Additional file 4: Table S3). Similarly, in an earlier investigation, the PaCPT expression levels were on average 6-fold higher than the PaPDCT expression levels in the avocado mesocarp [7]. Additionally, CPT orthologs are more highly expressed than PDCT orthologs in most oil-rich seeds and fruits [20, 25, 52]. Notably, PaCPT and PaPDCT exhibited tissue-specific transcriptional levels, yet PaCPT and PaPDCT do not contribute to major changes in the channeling of FAs from the PC pool for TAG synthesis in the avocado mesocarp and seed because of their relatively low expression levels (< 5 FPKM/stage in the mesocarp and seed) (Additional file 4: Table S3). Collectively, these results suggest that the flux through PC may contribute to the considerable accumulation of TAG in the avocado mesocarp and seed.

The enzyme encoded by FAD2 catalyzes the introduction of a second double bond in oleic acid, which generates polyunsaturated linoleic acid [67]. In our study, two PaFAD2 paralogs were expressed in the avocado mesocarp and seed, with higher transcription levels in the developing mesocarp than in the developing seed, consistent with the differences in the C18:2 contents between these two tissues (Fig. 4; Additional file 1: Table S1). Additionally, the PaFAD2-1 expression patterns in the developing mesocarp and seed were also in accordance with the changes in the C18:2 contents in the developing mesocarp and seed.

Lipid droplets accumulating TAG are encircled by a phospholipid monolayer and abundant amphipathic proteins, and they function as the hub for metabolic processes [68]. An earlier study proved that OBO, STERO, and CALO are TAG storage-related genes [20]. Recently, new lipid droplet-associated genes have been identified, namely LDAP1 and LDAP2 [69, 70]. In our study, PaLDAP2-1 and two PaCALO paralogs, with average transcript levels greater than 120 FPKM/stage, were the main genes involved in the formation of lipid droplets in the mesocarp, whereas four unigenes (PaLDAP2-1, PaOBO, and two PaCALO

paralogs), with average transcript levels greater than 120 FPKM/stage, were the primary genes contributing to the formation of lipid droplets in the seed (Fig. 4; Additional file 4: Table S3). In contrast to avocado, earlier transcriptomic studies revealed that the substantial expression of OBO, STERO, and CALO in the oil-bearing seed and tuber tissues of oil crops is likely important for stabilizing TAG in developing seeds and tubers [2, 20, 52].

In the current study, co-expression analyses identified potential TFs regulating the expression of lipid-related genes affecting oil accumulation. Previous investigations proved that WRI1, LEC1, LEC2, FUS3, and ABI3 encode key TFs regulating oil biosynthesis [18]. Both WRI1 and ABI3 were expressed in the avocado mesocarp and seed, respectively, in the current study, but they were not classified as hub genes in the co-expression network. A total of 143 and 141 TFs were uniquely correlated with the oil content in the avocado mesocarp and seed, respectively, implying the oil biosynthesis regulatory network varies between these two tissues. Similarly, different ABA-responsive TF genes (EgNF-YA3, EgNF-YC2, and EgABI5) were identified as regulators of oil accumulation in the oil palm mesocarp instead of the TF genes (WRI1, LEC1, LEC2, FUS3, and ABI3) identified in other oilseed crops, although all of these TF genes belong to the nuclear factor Y (NFY) and basic leucine zipper (bZIP) families [18, 23]. Our analysis of the avocado mesocarp and seed transcriptomes uncovered 10 hub TF genes in the avocado mesocarp and seed, of which PaPBS1-1 and PaRAP2-3 are highly expressed and encode enzymes that interact with other lipid-related TFs in the developing avocado mesocarp and seed, respectively (Fig. 5a, b, Additional file 8: Table S6). We identified PaPBS1-1 as a serine/threonine-protein kinase, and a GO term enrichment analysis indicated that lipid-related TFs are associated with protein serine/threonine kinase activities in the avocado mesocarp and seed. Thus, serine/threonine-protein kinases may affect oil accumulation. Serine/threonine-protein kinase participants in histone modifications, and histone modifications play essential roles in chromatin remodeling and gene expression regulation [71]. Recently, increasing evidence has demonstrated that histone modifications provide a key switch for oil accumulation in *A. thaliana* [71]. Additionally, PaRAP2-3 was identified as an ethylene-responsive TF, and the subsequent KEGG analysis implied that the lipid-related TFs are involved in plant hormone signal transduction pathways in the avocado mesocarp and seed. Similarly, previous studies suggested that hormone-responsive TFs (such as WRI1) regulate the expression of several fatty acid biosynthetic genes [11, 18, 25]. Moreover, many TF genes, such as PaRAP2-1, PaCOL4-1, PaGRP-2, PaMBF1B, PaEIN3-1, PaCDL1, and PaASIL2, were highly expressed in the developing avocado mesocarp and seed (Additional file 8: Table S6). Future analyses of these TFs with forward and backward genetic methods may clarify their relationships in avocado.

Recent studies confirmed the importance of lncRNAs for regulating gene expression in eukaryotic cells, especially during some key biological processes [46]. However, the number of lncRNAs encoded in genomes as well as their characteristics remain largely unknown [72]. In the current study, 5,299 lncRNAs were identified as expressed in the developing avocado mesocarp and seed, after which 11 trans-acting lncRNAs and 79 cis-acting lncRNAs corresponding to 43 lipid-related unigenes were detected in the developing avocado mesocarp and seed (Additional file 8: Table S7; Additional file 10: Table S8). Our data also revealed that nine cis-acting lncRNAs were specifically expressed in the avocado mesocarp or

seed (Additional file 10: Table S8). Further analyses indicated that lncRNAs PB.19359.1, PB.6205.1, PB.4443.3, PB.12340.3, and PB.19743.2 are more highly expressed (average transcript levels > 20 FPKM/stage) in the developing avocado mesocarp than in the developing seed (Additional file 8: Table S7). Previous studies also demonstrated that lncRNAs often exhibit tissue-specific expression patterns [73]. Furthermore, lncRNAs PB.19359.1 and PB.6205.1 were positively co-expressed with PDH (E3), whereas lncRNAs PB.4443.3, PB.12340.3, and PB.19743.2 expression levels were positively related with FAD2-2 expression (Additional file 4: Table S3; Additional file 8: Table S7). Therefore, these five lncRNAs may play substantial roles in accelerating oil accumulation, especially in the developing avocado mesocarp.

Conclusions

In summary, the regulated oil biosynthesis in the developing avocado mesocarp and seed was explored at the transcriptome, metabolic, and histological levels in this study. Bioinformatic analyses indicated that tissue-specific oil biosynthesis transcriptional networks in the developing avocado mesocarp and seed may involve multiple components, including various lipid-related genes, TFs, and lncRNAs. The data generated in the current study as well as in previous investigations were used to develop a regulatory network for the oil biosynthesis in the developing avocado mesocarp and seed that includes the vital lipid-related genes identified in this study (Fig. 4). Moreover, the expression profiles of known lipid-related genes in the developing avocado mesocarp and seed were summarized (Additional file 4: Table S3; Additional file 5: Table S4). Our results indicate that avocado is a promising model plant for characterizing tissue-specific oil biosynthesis and the associated regulatory network. Future studies should focus on functionally annotating the identified candidate genes, TFs, and lncRNAs to further expand our knowledge regarding oil biosynthesis in the avocado mesocarp and seed.

Methods

Plant materials

Avocado fruits (cultivar 'Hass') were collected from six 10-year-old trees consisting of Zutano clonal rootstock (two trees were classified as a unit for each biological replicate) from April 2018 to September 2018 at the Chinese Academy of Tropical Agricultural Sciences (Danzhou city, Hainan province, China; latitude: 19.52°N, longitude: 109.57°E, and altitude: 200 m above sea level). The fruits in each tree were marked during the main flowering season (i.e., February 2018), and collected at five time-points (75, 110, 145, 180, and 215 DAFB) until the fruits reached physiological maturity (defined as the ability to ripen after harvest). The 215 DAFB time-point was set as the harvest time. Three sets of fruits were randomly collected for each biological replicate in each developmental stage and quickly brought to the laboratory. The first set of nine fruits was morphologically characterized as previously described [51]. The second set was freeze-dried for an oil body analysis. The third set was immediately frozen at -80 °C for the subsequent transcript and FA composition analyses.

Analysis of the fatty acid composition by gas chromatography–mass spectrometry (GC-MS)

The FA composition was analyzed by GC-MS as described by Ge et al. [9]. Qualitative and quantitative analyses of the FA composition were performed with an Agilent 7890B-7000B GC-MS system (Santa Clara, CA, USA) equipped with a DB-5MS column (60 m × 0.25 mm i.d., 0.25-mm film thickness) using helium (1.3 mL/min) as the carrier gas. The fatty acid methyl esters (FAMES) were identified by comparing the retention times of the peaks with those of commercial standards and by comparing the respective ion chromatograms with those reported in the NIST 2011 library. Methyl nonadecanoate was added as an internal standard and the FAMES were quantified with the calibration curves for the standards ($R^2 \geq 0.995$). The FAME contents were measured for three biological replicates, each comprising three technical replicates.

Histological analyses

To visualize the lipid droplets in the mature avocado mesocarp and seed at 215 DAFB, sample sections that were 10–15 mm thick were prepared with the CV5030 vibratome (Leica Biosystems, Wetzlar, Germany) and then washed three times with 0.2 M phosphate-buffered saline. They were then stained with 0.5 mg/mL Nile red and examined with the LM510 confocal laser scanning microscope (Carl Zeiss AG, Oberkochen, Germany), with a laser excitation at 543 nm and a 40 × objective lens.

Illumina transcriptome library preparation and sequencing

The total RNA extraction and analysis (i.e., determination of concentration and integrity) were completed as described by Ge et al. [51]. Poly-T oligo-attached magnetic beads were applied to purify mRNA from the total RNA. Five mesocarp and seed samples underwent an RNA-seq analysis involving three biological replicates per sample. The fragmentation step was completed with divalent cations in the NEBNext First Strand Synthesis Reaction Buffer (5×) at an elevated temperature. First-strand cDNA was synthesized with a series of random hexamer primers and reverse transcriptase, after which the second cDNA strand was produced with DNA Polymerase I and RNase H. The cDNA libraries used for sequencing were constructed by ligating the cDNA fragments to sequencing adapters and amplifying the fragments by PCR. The libraries were then sequenced with the Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA).

PacBio Iso-seq library preparation and sequencing

The SMARTer PCR cDNA Synthesis Kit was used to synthesize cDNA from the same RNA samples used for Illumina sequencing. The cDNA was synthesized using mixed RNAs from five fruit developmental stages for each mesocarp and seed sample. After the PCR amplification, quality control analysis, and purification, full-length cDNA fragments were obtained with the BluePippin Size Selection System protocol for the construction of each mesocarp and seed cDNA library (1–6 kb). Selected full-length cDNA was ligated to the SMRT bell hairpin loop. A Qubit 2.0 fluorometer was then used to determine the cDNA library concentration. The quality of the cDNA library was assessed with the Agilent 2100

Bioanalyzer. Finally, each SMRT cell for the mesocarp and seed was sequenced with the PacBio Sequel \times instrument (Pacific Bioscience, Menlo Park, CA, USA).

Illumina data analysis

Raw data (raw reads) in the fastq format were first processed with in-house perl scripts. Clean data (clean reads) were obtained by removing reads containing adapters or poly-N sequences as well as low-quality reads from the raw data. At the same time, the Q20, Q30, GC-content, and sequence duplication level of the clean data were calculated. All downstream analyses were based on the high-quality clean data. These clean reads were then mapped to the reference genome sequence with HISAT2. Only reads with a perfect match or one mismatch were further analyzed and annotated based on the reference genome. Gene expression levels were calculated according to the expected number of fragments per kilobase of transcript per million mapped reads (FPKM). Genes differentially expressed between two samples were analyzed with the DESeq R package (1.10.0) [74]. The P-values were adjusted with the Benjamini & Hochberg method. Genes identified by DESeq with a false discovery rate ≤ 0.01 and a fold-change ≥ 2 were defined as differentially expressed.

PacBio data analysis

Raw reads were processed into error-corrected reads of insert (ROIs) using the Iso-seq pipeline with min Full Pass = 0.00 and min Predicted Accuracy = 0.80. Next, the FLNC transcripts were identified by searching for the poly-A tail signal and the 5' and 3' cDNA primers among the ROIs. Iterative Clustering for Error Correction (ICE) was used to obtain consensus isoforms, and full-length consensus sequences from ICE were polished with Quiver. The low-quality full-length transcript isoforms were corrected based on the Illumina short reads with the default settings of Proovread. High-quality and corrected low-quality transcript isoforms were confirmed as nonredundant with CD-HIT (version 1) (<http://weizhongli-lab.org/cd-hit/>). Full-length consensus sequences were mapped to the reference genome with GMAP [75]. Mapped reads were further collapsed with the pbtranscript-ToFU package, with min-coverage = 85% and min-identity = 90%. The 5' difference was not considered when collapsing redundant transcripts.

LncRNA identification based on PacBio sequences

Four computational approaches (CPC/CNCI/CPAT/Pfam) were combined to sort non-protein-coding RNA candidates from the putative protein-coding RNAs in the transcripts. Putative protein-coding RNAs were filtered out according to a minimum length and exon number threshold. Transcripts longer than 200 nt and with more than two exons were selected as lncRNA candidates to be screened further with CPC/CNCI/CPAT/Pfam, which can distinguish the protein-coding genes from the non-protein-coding genes.

To clarify the lncRNA biological functions, the target genes were predicted based on cis-acting and trans-acting modes [76]. The sliding window strategy was used to search for cis-acting target genes within 10/100 kb upstream and downstream of lncRNAs. To identify the trans-acting lncRNAs, the correlations

between the expression levels of lncRNAs and all protein-coding genes were calculated. The trans-acting lncRNAs were identified based on Pearson's correlation coefficient (r) > 0.9.

Analysis of transcription factors

Transcription factor gene families were identified according to categorically defined TF families and criteria from the KO, KOG, GO, Swiss-Prot, Pfam, Nr, and Nt databases with the default parameters of iTAK (version 1.2). The TFs were identified and classified as previously described [77].

Functional annotation of transcripts

Genes were functionally annotated following a BLASTX search (E-value threshold of 10^{-5}) of the following databases: Clusters of Orthologous Groups of proteins (KOG/COG), Non-supervised Orthologous Groups (eggNOG), Swiss-Prot, Pfam, and NCBI nonredundant protein sequence (Nr). Additionally, the KEGG Automatic Annotation Server [78] was used to assign KEGG metabolic pathways to the unigenes. Moreover, the unigenes were annotated with GO terms with the Blast2GO (version 2.5) program [79] based on the BLASTX matches in the Pfam and Nr databases (E-value threshold of 10^{-6}).

GO and KEGG enrichment analysis

The GO functional enrichment and KEGG pathway analyses of the differentially expressed genes were completed with Goseq R (version 1.20.0) [80] and KOBAS (version 3.0)[81].

Validation of transcript levels by quantitative real-time PCR

The transcription levels of 10 unigenes related to fatty acid biosynthesis expressed in the avocado mesocarp and seed were validated by qRT-PCR. Details regarding the qRT-PCR primers are presented in Additional file 13: Table S9. The qRT-PCR was conducted with a 96-well plate and the QuantStudio 7 Flex Real Time PCR System (Applied Biosystems, Foster City, CA, USA). The RNA extraction and qRT-PCR assay were completed as described by Ge et al. [51]. The PaActin7 gene was used as an endogenous control for normalizing the data and the $2^{-\Delta Ct}$ method was used for analyzing the PCR data. For each sample, the qRT-PCR analysis was conducted with three biological replicates, each comprising two technical replicates.

Co-expression analysis

A gene co-expression network was constructed according to the manual (labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/#manual Install) for the WGCNA R package. The data for the differentially expressed transcripts derived from the NGST sequencing data for the avocado mesocarp and seed were used to construct the co-expression network. The differentially expressed transcripts with FPKM value/stage ≥ 1 were used for the WGCNA. The GO and KEGG analyses were conducted with Blast2GO [79] and KOBAS [78], respectively. Co-expressed genes were counted with Metascape (metascape.org/gp/index.html#/main/step1) and their enrichment was statistically evaluated with Cytoscape 3.5 (<https://cytoscape.org/>).

List Of Abbreviations

FA: Fatty acid; TAG: Triacylglycerol; ER: Endoplasmic reticulum; WRI1: WRINKLED1; LEC1: LEAFY COTYLEDON 1; LEC2: LEAFY COTYLEDON 2; FUS3: FUSCA 3; ABI3: ABSCISIC ACID INSENSITIVE 3; SGS: Second-generation sequencing; SMRT: Single-molecule real-time; lncRNAs: Long non-coding RNAs; WGCNA: Weighted gene co-expression network analysis; TFs: Transcription factors; DAFB: Days after full bloom; FLNC: Full-length non-chimeric; FPKM: Fragments per kilobase of transcript per million mapped reads; OBO: Oil-body oleosin; STERO: Steroleosin; CALO: Caleosin; LDAP2: Lipid droplet-associated protein 2; KEGG: Kyoto Encyclopedia of Genes and Genomes; SAD: Stearoyl-acyl carrier protein desaturase; LACS: Long-chain acyl-CoA synthetase; LPAAT: 1-acylglycerol-3-phosphate acyltransferase; DAG: Diacylglycerol; PDCT: Phosphatidylcholine:DAG cholinephosphotransferase; CPT: Cytidine-5'-diphosphocholine:DAG cholinephosphotransferase; GC-MS: gas chromatography–mass spectrometry; FAMES: Fatty acid methyl esters; ROIs: Reads of insert; ICE: Iterative Clustering for Error Correction; KOG/COG: Clusters of Orthologous Groups of proteins; eggNOG: Non-supervised Orthologous Groups; Nr: NCBI nonredundant protein sequence;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data sets supporting the results of this article are included within the article.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the National Natural Science Foundation of China (grant number 31701883), Natural Science Foundation of Hainan Province of China (grant number 2019RC264), Earmarked fund for the belt and road tropical project (grant number BARTP-07), and Integrated demonstration of key techniques for the industrial development of featured crops in rocky desertification areas of Yunnan-Guangxi-Guizhou provinces (grant number SMH2019-2021).

Authors' contributions

Conceived and designed the experiments: YG and RZ. Performed the experiments: XD, YL and YY. Analyzed the data: YG and XD. Contributed materials: RZ. Wrote the paper: YG.

Acknowledgments

We thank Dr. Yajima for editing the English text of a draft of this manuscript.

References

1. Lourith N, Kanlayavattanakul M, Mongkonpaibool K, Butsaratrakool T, Chinmuang T. Rambutan seed as a new promising unconventional source of specialty fat for cosmetics. *Ind Crop Prod*. 2016;83:149-154.
2. Tian B, Lu TQ, Xu Y, Wang RL, Chen GQ. Identification of genes associated with ricinoleic acid accumulation in *Hiptage benghalensis* via transcriptome analysis. *Biotechnol Biofuels* 2019;12:16.
3. Wan X, Liu Q, Dong B, Pillai SV, Huang FH, Singh SP, et al. Molecular and biochemical analysis of the castor caruncle reveals a set of unique genes involved in oil accumulation in non-seed tissues. *Biotechnol Biofuels* 2019;12:158.
4. Rendón-Anaya M, Ibarra-Laclette E, Méndez-Bravo A, Lan TY, Zheng CF, Carretero-Paulet L, et al. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proc Natl Acad Sci USA* 2019;116:17081-17089.
5. Avocado and olive oil methyl esters. *Biomass Bioenerg*. 2013;58:143-148.
6. Villa-Rodríguez JA, Molina-Corral FJ, Ayala-Zavala JF, Olivas GI, Gonzalez-Aguilar GA. Effect of maturity stage on the content of fatty acids and antioxidant activity of 'Hass' avocado. *Food Res Int*. 2011;44:1231-1237.
7. Kilaru A, Cao X, Dabbs PB, Sung, HJ, Rahman MM, Thrower N, et al. Oil biosynthesis in a basal angiosperm: transcriptome analysis of *Persea americana* *BMC Plant Biol*. 2015;15:203.
8. Santo MAZ, Alicieo TVR, Pereira CMP, Ramis-Ramos G, Mendonca CRB. Profile of bioactive compounds in avocado pulp oil: Influence of the drying processes and extraction methods. *J Am Oil Chem Soc*. 2014;91:19-27.
9. Ge Y, Si XY, Wu B, Dong XS, Xu ZN, Ma WH. Oil content and fatty acid composition of the seeds of 16 avocado (*Persea americana*) accessions collected from southern China and their application in a soap bar. *J Agr Sci*. 2018;10:69-78.
10. Giraldo L, Moreno-Piraján JC. Lipase supported on mesoporous materials as a catalyst in the synthesis of biodiesel from *Persea americana* oil. *J Mol Catal B-Enzym*. 2012;77:32-38.
11. Tranbarger TJ, Dussert S, Joet T, Argout X, Summo M, Champion A, et al. Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiol*. 2011;156:564-584.
12. Guerin C, Joet T, Serret J, Lashermes P, Vaissayre V, Agbessi MDT, et al. Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *Plant J*. 2016;87:423-441.

13. Van Erp H, Kelly AA, Menard G, Eastmond PJ. Multigene engineering of triacylglycerol metabolism boosts seed oil content in Arabidopsis. *Plant Physiol.* 2014;165:30-36.
14. Zhang DJ, Zhao MX, Li S, Sun LJ, Wang WD, Cai CM, et al. Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean. *Plant J.* 2017;90:1120-1133.
15. Qi ZM, Zhang ZG, Wang ZY, Yu JY, Qin HT, Mao XR, et al. Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development. *Plant Cell Environ.* 2018;41:2109-2127.
16. Yang XQ, Pan H, Zeng T, Shupe TF, Hse CY. Extraction and characterization of seed oil from naturally grown Chinese tallow trees. *J Am Oil Chem Soc.* 2013;90:459-466.
17. Ibarra-Laclette E, Méndez-Bravo A, Pérez-Torres CA, Albert VA, Mockaitis K, Kilaru A, et al. Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genomics* 2015;16:599.
18. Yeap WC, Lee FC, Shan DKS, Musa H, Appleton DR, Kulaveerasingam H. WRI1-1, ABI5, NF-YA3 and NF-YC2 increase oil biosynthesis in coordination with hormonal signaling during fruit development in oil palm. *Plant J.* 2017;91:97-113.
19. Simpson JP, Ohlrogge JB. A novel pathway for triacylglycerol biosynthesis is responsible for the accumulation of massive quantities of glycerolipids in the surface wax of bayberry (*Myrica pensylvanica*) fruit. *Plant Cell* 2016;28:248-264.
20. Yang ZL, Ji HY, Liu DT. Oil biosynthesis in underground oil-rich storage vegetative tissue: comparison of *Cyperus esculentus* ruber with oil seeds and fruits. *Plant Cell Physiol.* 2016;57:2519-2540.
21. Wang G, Lin Q, Xu Y. *Tetraena mongolica* Maxim can accumulate large amounts of triacylglycerol in phloem cells and xylem parenchyma of stems. *Phytochemistry* 2007;68:2112-2117.
22. Shah ZH, Hamooh BT, Daur I, Rehman HM, Alghabari F. Transcriptomics and biochemical profiling: current dynamics in elucidating the potential attributes of olive. *Curr Issues Mol Biol.* 2017;21:73-98.
23. Kong Q, Ma W. WRINKLED1 transcription factor: How much do we know about its regulatory mechanism? *Plant Sci.* 2018;272:153-156.
24. Focks N, Benning C. *wrinkled1*: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* 1998;118:91-101.
25. Dussert S, Guerin C, Andersson M, Joet T, Tranbarger TJ, Pizot M, et al. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.* 2013;162:1337-1358.
26. Adhikari ND, Bates PD, Browse J. *WRINKLED1* rescues feedback inhibition of fatty acid synthesis in hydroxylase-expressing seeds. *Plant Physiol.* 2016;171:179-191.
27. Ji XJ, Mao X, Hao QT, Liu BL, Xue JA, Li RZ. Splice variants of the castor *WRI1* gene upregulate fatty acid and oil biosynthesis when expressed in tobacco leaves. *Int J Mol Sci.* 2018;19:146.

28. Deng Y, Zheng H, Yan ZC, Liao DY, Li CL, Zhou JY, et al. Full-length transcriptome survey and expression analysis of *Cassia obtusifolia* to discover putative genes related to aurantio-obtusin biosynthesis, seed formation and development, and stress response. *Int J Mol Sci.* 2019;19:2476.
29. Mu JY, Tan HL, Zheng Q, Fu FY, Liang Y, Zhang J, et al. *LEAFY COTYLEDON1* is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiol.* 2008;148:1042-1054.
30. Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. *Plant J.* 2007;50:825-838.
31. Ge Y, Si XY, Cao JQ, Zhou ZX, Wang WL, Ma WH. Morphological characteristics, nutritional quality, and bioactive constituents in fruits of two avocado (*Persea americana*) varieties from hainan province, China. *J Agr Sci.* 2017;9:8-17.
32. Kaur G, Arya SK, Singh B, Singh S, Dhar YV, Verman PC, et al. Transcriptome analysis of the palmarosa *Cymbopogon martinii* inflorescence with emphasis on genes involved in essential oil biosynthesis. *Ind Crop Prod.* 2019;140:111602.
33. Wang WJ, Wu Y, Xu HH, Shang Y, Chen YC, Yan ME, et al. Accumulation mechanism of indigo and indirubin in *Polygonum tinctorium* revealed by metabolite and transcriptome analysis. *Ind Crop Prod.* 2019;141:111783.
34. Ye P, Liang SC, Wang XM, Duan LX, Jiang-Yan FY, Yang JF, et al. Transcriptome analysis and targeted metabolic profiling for pathway elucidation and identification of a geraniol synthase involved in iridoid biosynthesis from *Gardenia jasminoides*. *Ind Crop Prod.* 2019;132:48-58.
35. Zhou SF, Chen J, Lai YS, Yin GH, Chen PL, Pennerman KK, et al. Integrative analysis of metabolome and transcriptome reveals anthocyanins biosynthesis regulation in grass species *Pennisetum purpureum*. *Ind Crop Prod.* 2019;138:111470.
36. Chao YH, Yuan JB, Li SF, Jia SQ, Han LB, Xu LX. Analysis of transcripts and splice isoforms in red clover (*Trifolium pretense*) by single-molecule long-read sequencing. *BMC Plant Biol.* 2018;18:300.
37. Tian JY, Feng SJ, Liu YL, Zhao LL, Tian L, Hu Y, et al. Single-molecule long-read sequencing of *Zanthoxylum bungeanum* transcriptome: identification of aroma-related genes. *Forests* 2018;9:765.
38. Chen J, Tang XH, Ren CX, Wei B, Wu YY, Wu QH, et al. Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. *BMC Genomics* 2018;19:548.
39. Luo D, Zhou Q, Wu YG, Chai XT, Liu WX, Wang YR, et al. Full-length transcript sequencing and comparative transcriptomic analysis to evaluate the contribution of osmotic and ionic stress components towards salinity tolerance in the roots of cultivated alfalfa (*Medicago sativa*). *BMC Plant Bio.* 2019;19:32.
40. Filichkin SA, Hamilton M, Dharmawardhana PD, Singh SK, Sullivan C, Ben-Hur A, et al. Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intro retention, and isoform ratio switching. *Front Plant Sci.* 2018;9:5.
41. Yang LF, Jin YH, Huang W, Sun Q, Liu F, Huang X.Z. Full-length transcriptome sequences of ephemeral plant *Arabidopsis pumila* provides insight into gene expression dynamics during

- continuous salt stress. *BMC Genomics* 2018;19:717.
42. Zuo CM, Blow M, Sreedasyam A, Kuo RC, Ramamoorthy GK, Torres-Jerez I, et al. Revealing the transcriptomic complexity of switchgrass by PacBio long-read sequencing. *Biotechnol Biofuels* 2018;11:170.
 43. Chao Q, Gao ZF, Zhang D, Zhao BG, Dong FQ, Fu CX, et al. The developmental dynamics of the *Populus* stem transcriptome. *Plant Biotech J.* 2019;17:206-219.
 44. Chao YH, Yuan JB, Guo T, Xu LX, Mu ZY, Han LB. Analysis of transcripts and splice isoforms in *Medicago sativa* by single-molecule long-read sequencing. *Plant Mol Bio.* 2019;99:219-235.
 45. Ye JB, Cheng SY, Zhou X, Chen ZX, Kim SU, Tan JP, et al. A global survey of full-length transcriptome of *Ginkgo biloba* reveals transcript variants involved in flavonoid biosynthesis. *Ind Crop Prod.* 2019;139:111547.
 46. Liu J, Wang H, Chua NH. Long noncoding RNA transcriptome of plants. *Plant Biotechnol J.* 2015;13:319-328.
 47. Randell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet.* 2012;13:329-342.
 48. Kim JA, Roy NS, Lee IH, Choi AY, Choi BS, Yu YS, et al. Genome-wide transcriptome profiling of the medicinal plant *Zanthoxylum planispinum* using a single-molecule direct RNA sequencing approach. *Genomics* 2019;111:973-979.
 49. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 2012;24:4333-4345.
 50. Ochogavía A, Galla G, Seijo JG, González AM, Bellucci M, Pupilli F, et al. Structure, target-specificity and expression of *PN_LNC_N13*, a long non-coding RNA differentially expressed in apomictic and sexual *Paspalum notatum*. *Plant Mol. Biol.* 2018;96:53-67.
 51. Ge Y, Cheng ZH, Si XY, Ma WH, Tan L, Zang XP, et al. Transcriptome profiling provides insight into the genes in carotenoid biosynthesis during the mesocarp and seed developmental stages of avocado (*Persea Americana*). *Int J Mol Sci.* 2019;20:4117.
 52. Troncoso-Ponce MA, Kilaru A, Cao X, Durrett TP, Fan JL, Jensen JK, et al. Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.* 2011;68:1014-1027.
 53. Du HW, Huang M, Li JS. Modification of the fatty acid composition in *Arabidopsis* and maize seeds using a stearyl-acyl carrier protein desaturase-1 (*ZmSAD1*) gene. *BMC Plant Biol.* 2016;16:137.
 54. Kachroo A, Shanklin J, Whittle E, Lapchyk L, Hildebrand D, Kachroo P. The *Arabidopsis* stearyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant Mol Biol.* 2007;63:257-271.
 55. Unver T, Wu ZY, Sterck L, Turktas M, Lohaus R, Li Z, et al. Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci USA* 2017;114:9413-9422.
 56. Jiang FY, Zhao L, Yandea-Nelson MD, Nikolau BJ. Two distinct domains contribute to the substrate acyl chain length selectivity of plant acyl-ACP thioesterase. *Nat Commun.* 2018;9:860.

57. Aznar-Moreno JA, Venegas-Calcrón M, Martínez-Force E, Garcés R, Salas JJ. Acyl carrier proteins from sunflower (*Helianthus annuus*) seeds and their influence on FatA and FatB acyl-ACP thioesterase activities. *Planta* 2016;244:479-490.
58. Aznar-Moreno JA, Venegas CM, Martínez-Force E, Garcés R, Mullen R, Gidda SK, et al. Sunflower (*Helianthus annuus*) long-chain acyl-coenzyme A synthetases expressed at high levels in developing seeds. *Physiol Plant*. 2014;150:363-373.
59. Salas JJ, Ohlrogge JB. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. *Arch Biochem Biophys*. 2002;403:25-34.
60. Klett EL, Chen SF, Yechoor A, Lih FB, Coleman RA. Long-chain acyl-CoA synthetase isoforms differ in preferences for eicosanoid species and long-chain fatty acids. *J Lipid Res*. 2017;58:884-894.
61. Shockey JM, Fulda MS, Browse JA. *Arabidopsis* contains nine longchain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. *Plant Physiol*. 2002;129:1710-1722.
62. Jessen D, Roth C, Wiermer M, Fulda M. Two activities of long-chain acyl-coenzyme A synthetase are involved in lipid trafficking between the endoplasmic eeticulum and the plastid in *Arabidopsis*. *Plant Physiol*. 2015;167:351-366.
63. Petit J, Bres C, Mauxion J-P, Tai FWJ, Martin LBB, Fich EA, et al. The Glycerol-3-Phosphate Acyltransferase GPAT6 from tomato plays a central role in fruit cutin biosynthesis. *Plant Physiol*. 2016;171:894-913.
64. Craddock CP, Adams N, Bryant FM, Kurup S, Eastmond PJ. PHOSPHATIDIC ACID PHOSPHOHYDROLASE regulates phosphatidylcholine biosynthesis in *Arabidopsis* by phosphatidic acid-mediated activation of CTP:PHOSPHOCHOLINE CYTIDYLYLTRANSFERASE activity. *Plant Cell* 2015;27:1251-1264.
65. Bates PD, Stymne S, Ohlrogge J. Biochemical pathways in seed oil synthesis. *Curr Opin Plant Biol*. 2013;16:358-364.
66. Bourgis F, Kilaru A, Cao X, Ngando-Ebongue GF, Drira N, Ohlrogge JB, et al. Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc Natl Acad Sci USA* 2011;108:12527-12532.
67. Feng T, Yang Y, Busta L, Cahoon EB, Wang HC, Lv SY. *FAD2* gene radiation and positive selection contributed to polyacetylene metabolism evolution in campanulids. *Plant Physiol*. 2019;181:714-728.
68. Walther TC, Chung JY, Farese Jr. RV. Lipid droplet biogenesis. *Annu Rev Cell Dev Bi*. 2017;33:491-510.
69. Gidda SK, Watt S, Collins-Silva J, Kilaru A, Arondel V, Yurchenko O, et al. Lipid droplet-associated proteins (LDAPs) are involved in the compartmentalization of lipophilic compounds in plant cells. *Plant Signal Behav*. 2013;8:e27141.
70. Horn PJ, James CN, Gidda SK, Kilaru A, Dyer JM, Mullen RT, et al. Identification of a new class of lipid droplet-associated proteins in plants. *Plant Physiol*. 2013;162:1926-1936.
71. Wang TY, Xing JW, Liu XY, Liu ZS, Yao YY, Hu ZR, et al. Histone acetyltransferase general control non-repressed protein 5 (GCN5) affects the fatty acid composition of *Arabidopsis thaliana* seeds by

- acetylating fatty acid desaturase3 (*FAD3*). *Plant J.* 2016;88:794-808.
72. Yandell M, Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet.* 2012;13:329-342.
73. Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 2014;24:444-453.
74. Anders S. Analysing RNA-Seq data with the DESeq package. *Mol Biol.* 2010;43:1-17.
75. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 2005;21:1859-1875.
76. Li A, Zhang J, Zhou Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinforma.* 2014;15:311.
77. Perez-Rodriguez P, Riano-Pachon DM, Correa LG, Rensing SA, Kersten B, Mueller-Roeber B. PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 2010;38:822-827.
78. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36:480-484.
79. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008;36:3420-3435.
80. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11:R14.
81. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39:316-322.

Additional Files

Additional file 1: Table S1. Fatty acid composition of total lipid in developing avocado mesocarp and seed during the fruit development.

Additional file 2: Figure S1. Total ion chromatograms of fatty acid methyl ester of avocado mesocarp and seed at 215 DAFB (Harvest).

Additional file 3: Table S2. Summary of 30 Illumina RNA-seq data and the alignment with the reference genome sequence.

Additional file 4: Table S3. Analysis of differential expression and annotation of unigenes participating in FA synthesis and TAG assembly.

Additional file 5: Table S4. Analysis of differential expression and annotation of unigenes participating in TAG storage.

Additional file 6: Figure S2. Identification of the putative expressed transcription factors between mesocarp and seed during avocado fruit development.

Additional file 7: Table S5. Overview of the putative expressed transcription factors in developing avocado mesocarp and seed.

Additional file 8: Table S6. Overview of lipid-related transcription factors in developing avocado mesocarp and seed based on WGCNA.

Additional file 9: Table S7. Overview of the predicted expressed lncRNAs in developing avocado mesocarp and seed.

Additional file 10: Figure S3. Identification of expressed lncRNAs between mesocarp and seed during avocado fruit development.

Additional file 11: Table S8. Overview of the predicted lncRNAs in cis/trans-regulation of the expressions of lipid-related genes participating in FA synthesis and TAG assembly.

Additional file 12: Figure S4. Validation of the transcriptomic data with quantitative RT-PCR. Ten unigenes and lncRNAs were used in this validation in avocado mesocarp (**a**) and seed (**b**). The comparative FPKM and $2^{-\Delta\Delta Ct}$ at 75 DAFB were used as the control for normalization. Results represent the mean of three biological replicates and two technical replicates (mean \pm SD, $n=6$). Refer to Additional file 4 Table S3 for the detail information on the selected unigenes (*PaPDH(E1 β)*, *PaACP4-2*, *PaSAD(FAB2)*, *PaLACS8*, *PaFAD2-2*) participating in FA synthesis and TAG assembly. Refer to Additional file 5 Table S4 for the detail information on the selected unigenes (*PaLDAP2-1*) participating in TAG storage. Refer to Additional file 8 Table S6 for the detail information on the selected lipid-related transcription factors (*PaPBS1-1* and *PaRAP2-3*). Refer to Additional file 9 Table S7 for the detail information on the selected lncRNAs (PB.19359.1 and PB.4443.3).

Additional file 13: Table S9. Primers used in qRT-PCR.

Figures

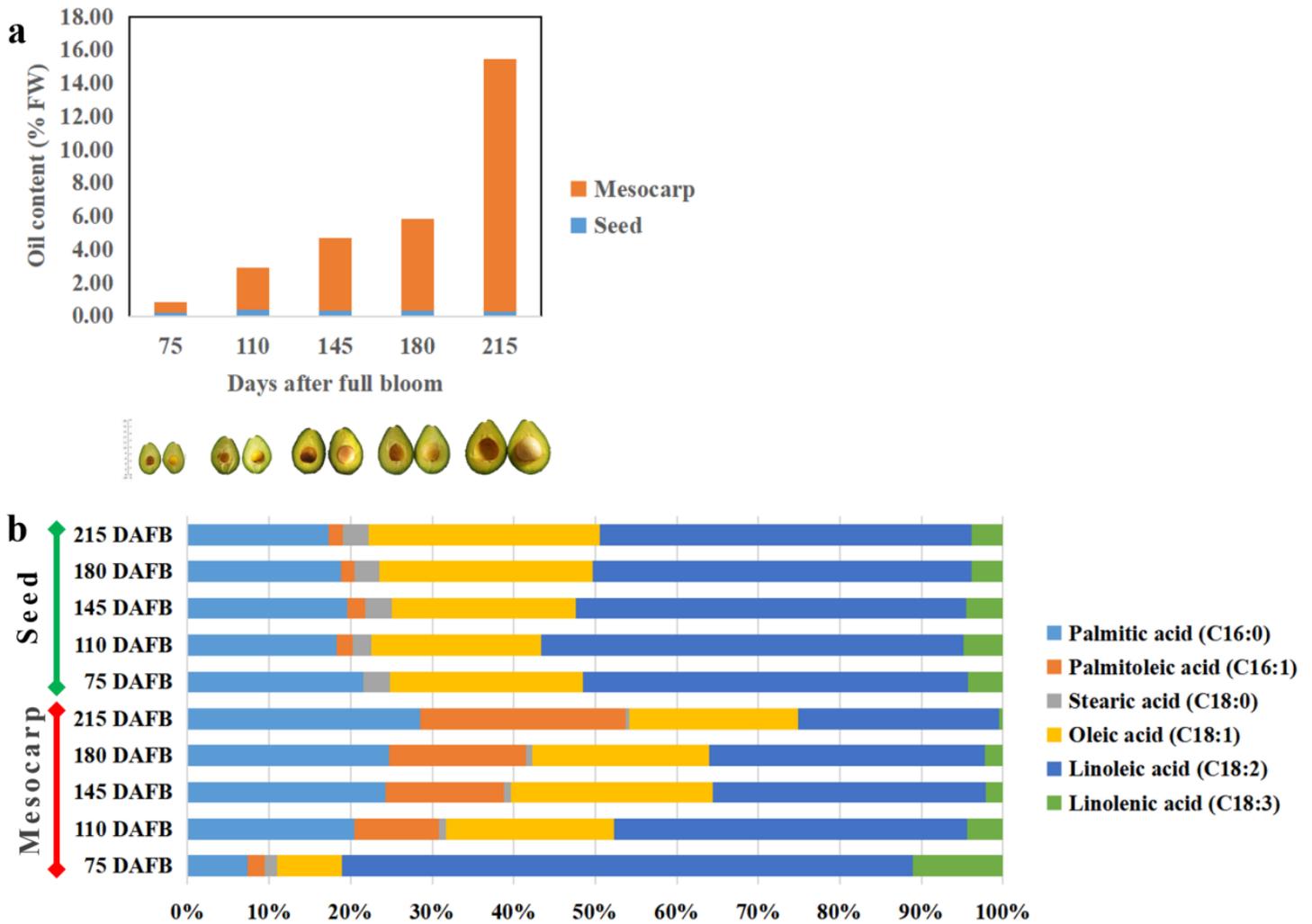


Figure 1

Total lipid content and composition of developing avocado fruits. a Oil content in the avocado mesocarp and seed during the fruit developmental period. The oil contents of individual mesocarps and seeds were measured by gas chromatography–mass spectrometry and are presented as a percentage of fresh weight (n = 9). b Fatty acid composition in the developing avocado mesocarp and seed during the fruit developmental period (n = 9)

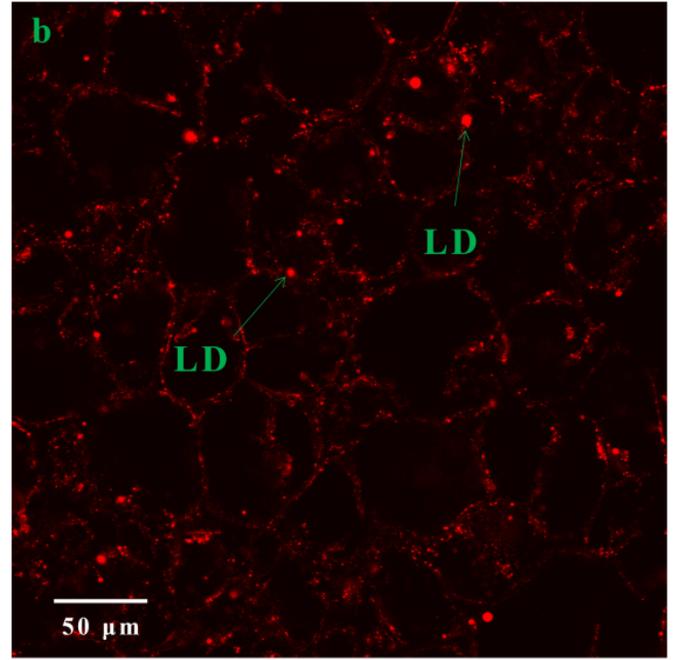
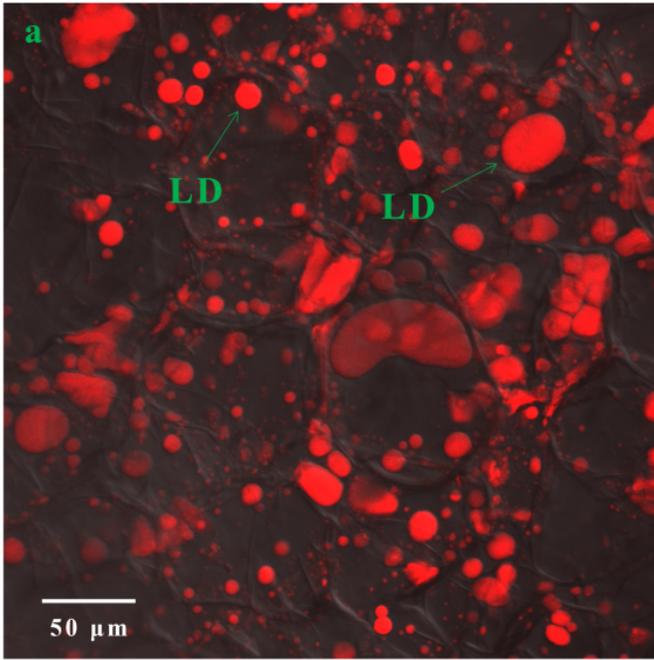


Figure 2

Confocal microscopy images of lipid droplets in the mature avocado mesocarp (a) and seed (b). Lipid droplets were visualized by Nile red staining at 215 DAFB. LD: lipid droplet

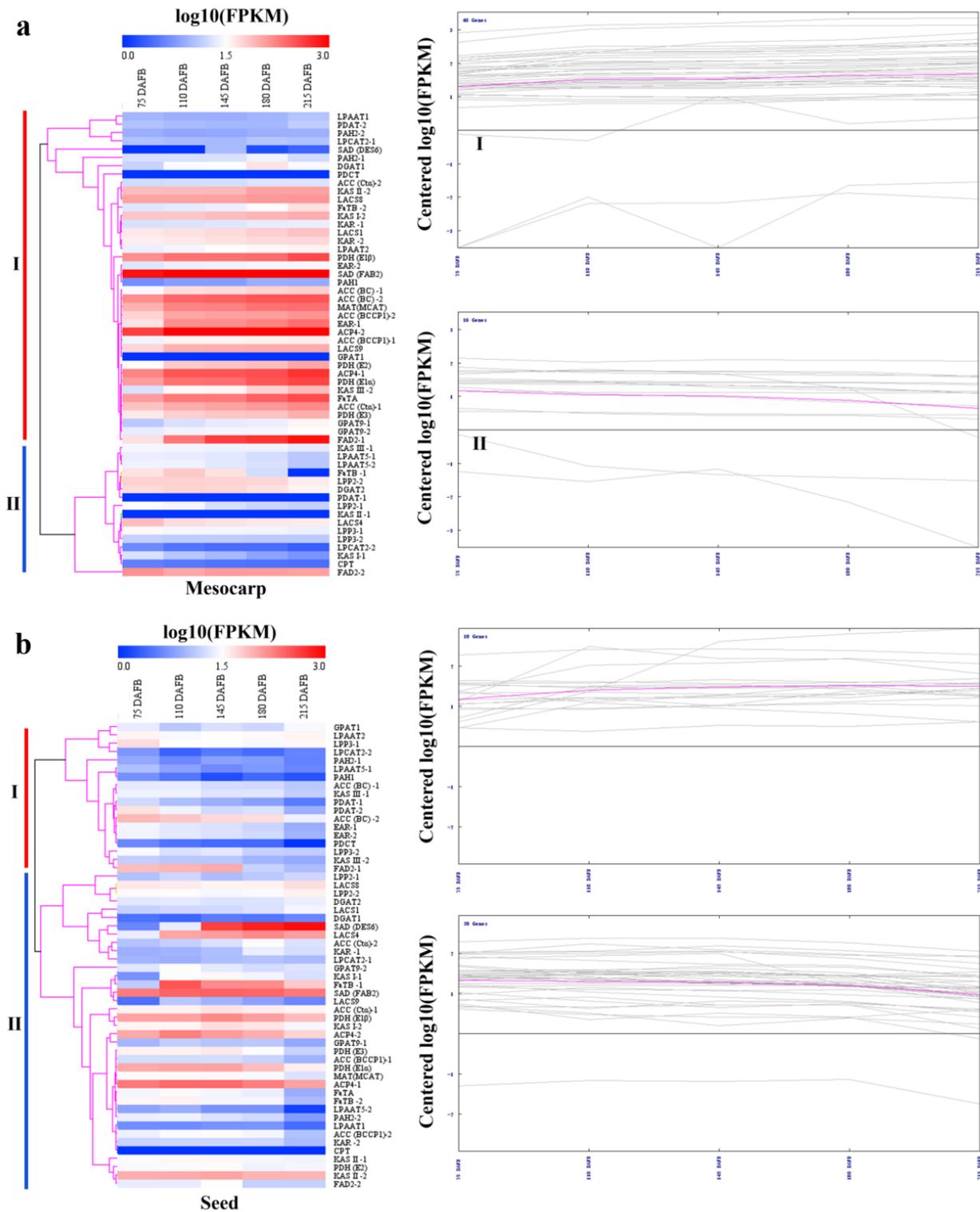


Figure 3

Cluster analysis of the avocado mesocarp and seed lipid-related genes associated with FA synthesis and TAG assembly. a Hierarchical clustering dendrogram of the lipid-related genes involved in FA synthesis and TAG assembly in the avocado mesocarp. Red and blue reflect high and low expression levels, respectively. Different gene expression patterns were detected in the two gene clusters. b Hierarchical clustering dendrogram of the lipid-related genes affecting FA synthesis and TAG assembly in the avocado

seed. Red and blue reflect high and low expression levels, respectively. Different gene expression patterns were detected in the two gene clusters

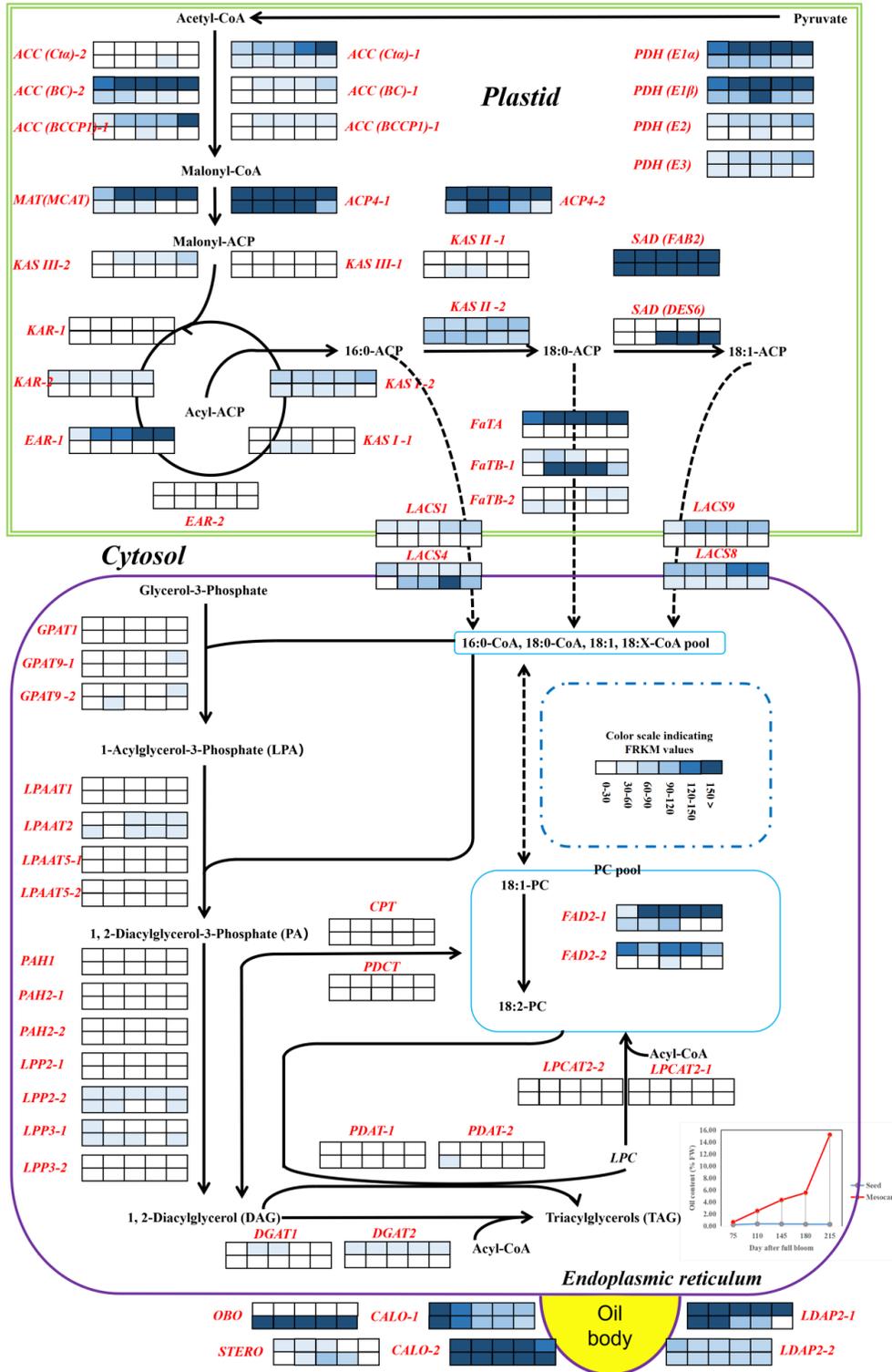


Figure 4

Transcript levels of the lipid-related genes associated with FA synthesis and TAG assembly in the avocado mesocarp and seed. Gene expression levels at 75, 110, 145, 180, and 215 days after full bloom are indicated with colored bars, with mesocarp and seed results presented in the top and bottom rows of

bars, respectively. FPKM: fragments per kilobase of transcript per million mapped reads; PDH (E1 α): pyruvate dehydrogenase E1 component subunit alpha; PDH (E1 β): pyruvate dehydrogenase E1 component subunit beta; PDH (E2): E2 component of pyruvate dehydrogenase complex; PDH (E3): E3 component of pyruvate dehydrogenase complex; ACC (Cta): carboxyltransferase α -subunit of heteromeric ACCase; ACC (BC): biotin carboxylase subunit of heteromeric acetyl-CoA carboxylase (ACCase); ACC (BCCP1): biotin carboxyl carrier protein of heteromeric ACCase; MAT (MCAT): malonyl-CoA:ACP malonyltransferase; ACP4: acyl carrier protein 4; KAS III: ketoacyl-ACP synthase III; KAR: ketoacyl-ACP reductase; EAR: enoyl-ACP reductase; KAS I: ketoacyl-ACP synthase I; KAS II: ketoacyl-ACP synthase II; SAD: stearoyl-ACP desaturase; FaTA: acyl-ACP thioesterase A; FaTB: acyl-ACP thioesterase B; LACS: long-chain acyl-CoA synthetase; GPAT: glycerol-3-phosphate acyltransferase; LPAAT: 1-acylglycerol-3-phosphate acyltransferase; PAH: phosphatidate phosphatase; LPP: phosphatidate phosphatase; DGAT: acyl-CoA:diacylglycerol acyltransferase; CPT: diacylglycerol cholinephosphotransferase; PDCT: phosphatidylcholine:diacylglycerol cholinephosphotransferase; PDAT: phospholipid:diacylglycerol acyltransferase; LPCAT: 1-acylglycerol-3-phosphocholine acyltransferase; LDAP: lipid droplet-associated protein; OBO: oil-body oleosin; CALO: caleosin; STERO: steroleosin

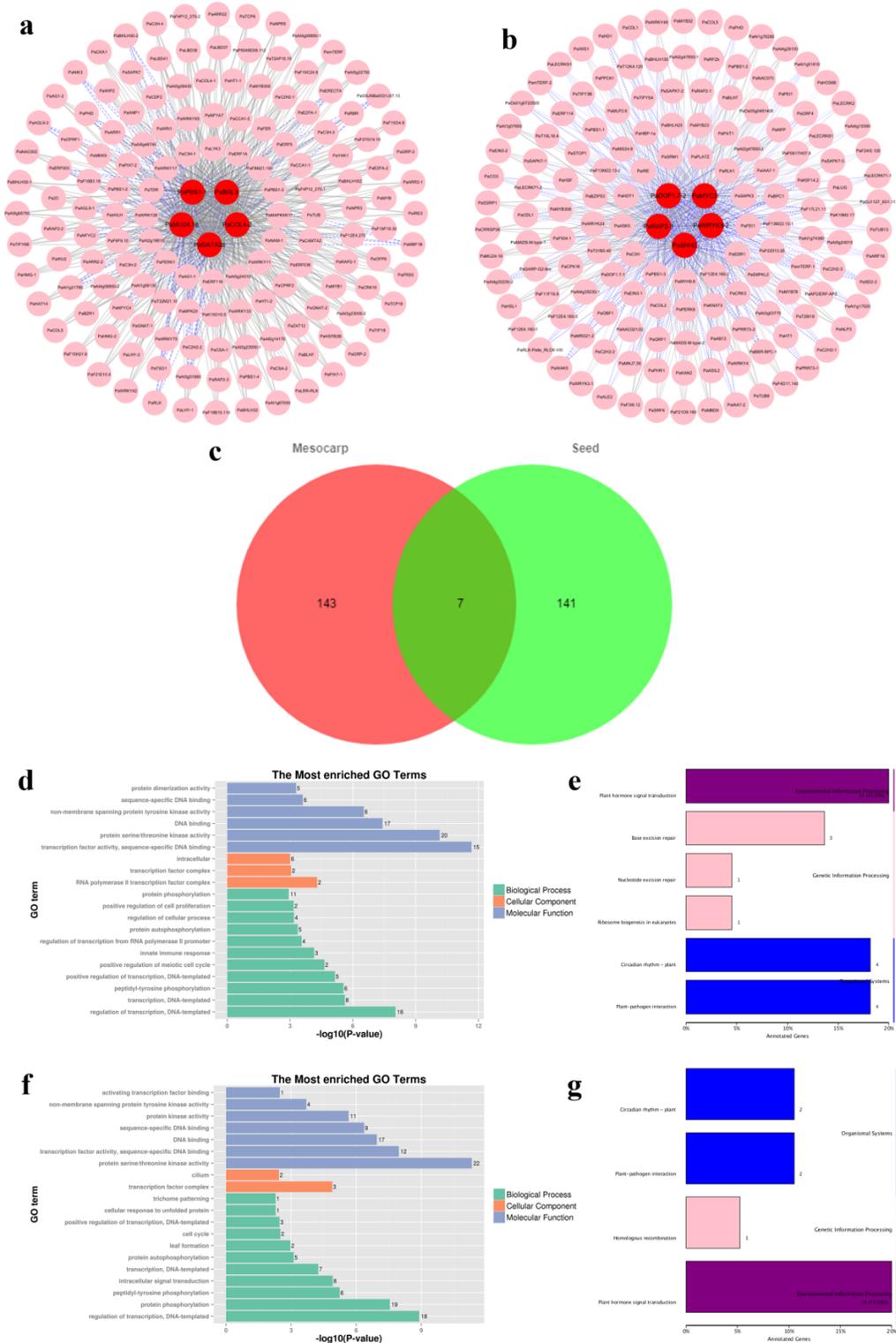


Figure 5

Functional clustering of lipid-correlated transcription factors. a Visualization of the lipid-correlated transcription factor enrichment analysis of the avocado mesocarp with Metascape. b Visualization of the lipid-correlated transcription factor enrichment analysis of the avocado seed with Metascape. Red circular nodes indicate the hub lipid-correlated transcription factors, whereas the pink nodes indicate the other transcription factors. Solid gray and dotted blue lines indicate positive and negative co-expression,

respectively. c Identification of differentially expressed lipid-correlated transcription factors between the mesocarp and seed in developing avocado fruit. d GO term enrichment of lipid-correlated transcription factors in the avocado mesocarp. e KEGG pathway analysis of lipid-correlated transcription factors in the avocado mesocarp. f GO term enrichment of lipid-correlated transcription factors in the avocado seed. g KEGG pathway analysis of lipid-correlated transcription factors in the avocado seed. Detailed results are listed in Additional file 8: Table S6.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile4TableS3.xlsx](#)
- [Additionalfile12FigureS4.tif](#)
- [Additionalfile7TableS5.xlsx](#)
- [Additionalfile13TableS9.xlsx](#)
- [Additionalfile8TableS6.xlsx](#)
- [Additionalfile1TableS1.xlsx](#)
- [Additionalfile9TableS7.xlsx](#)
- [Additionalfile2FigureS1.tif](#)
- [Additionalfile5TableS4.xlsx](#)
- [Additionalfile3TableS2.xlsx](#)
- [Additionalfile11TableS8.xlsx](#)
- [Additionalfile6FigureS2.tif](#)
- [Additionalfile10FigureS3.tif](#)