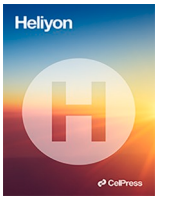
Heliyon 10 (2024) e38015

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

De novo assembly of *Idesia polycarpa* transcriptome and unsaturated fatty acid biosynthesis candidate genes Mining and functional Identification

Ruishen Fan [a,b,1], Boheng Wang [a,c,1], Hang Yu [a], Yiran Wang [a], Yanpeng Kui [a], Minmin Chen [b], Yibin Wang [a], Xiaoming Jia [a,*]

[a] College of Forestry, Northwest A&F University, Yangling, Shaanxi, China
[b] Rubber Research Institute, Chinese Academy of Tropical Agricultural Science, Haikou, Hainan, China
[c] East China Survey and Planning Institute of National Forest and Grassland Administration, Hangzhou, Zhejiang, China

ARTICLE INFO

Keywords:
Edible lipid
Unsaturated fatty acids
Idesia polycarpa
Transcriptome
UFA metabolism
Gene function

ABSTRACT

Unsaturated fatty acids (UFA) in lipids are the key to nutraceutical oil applications, with various potential applications in nutraceutical functional foods and pharmaceutical industries. In *Idesia polycarpa* (Salicaceae), more than 80 % of UFA have been found in the fruits; yet, the underlying genetic mechanism remains poorly understood. Due to the lack of theoretical research on the genes related to lipid biosynthesis and the complete genetic transformation system of *I. polycarpa* fruit, the selection and breeding of *I. polycarpa*, an excellent oil tree, has been severely restricted. In-depth understanding of the molecular mechanism and gene function of lipid biosynthesis of *I. polycarpa* fruit is therefore of great significance for the development of *I. polycarpa* resources. This is not only conducive to the genetic improvement of *I. polycarpa* by molecular breeding technologies but can also provide a reference for the study of the gene functions of other oil plants. In this study, the FA accumulation patterns of *I. polycarpa* fruits during 8 growth periods were analysed. Fruit from two developmental periods with different UFA levels were analysed for RNA sequencing by an Illumina NovaSeq 6000 HiSeq platform. *De novo* transcriptome assembly presented 115,350 unigenes and 4382 differentially expressed genes (DEGs). Functional annotation in the KEGG pathway and combined with DEG data revealed candidate genes potentially involved in UFA biosynthesis. Expression analysis of q-PCR of *IpDGAT2, IpGPAT, IpKASII, IpSAD, IpFAD2, IpFAD3* and *IpFAD8* suggested that these genes are highly involved in UFA biosynthesis. Full-length candidate genes were cloned and analysed by bioinformatic tools, and function analysis of *IpSAD* and *IpFAD3* showed that these genes regulated the products of linoleic acid metabolism. This study provides a foundation for UFA biosynthesis in *Idesia polycarpa*, facilitating its genetic breeding in the future.

1. Introduction

Agriculture is the foundation of the development of human society and the national economy and plays an important role in the

* Corresponding author.
E-mail address: jxm0601@163.com (X. Jia).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38015
Received 22 May 2023; Received in revised form 12 September 2024; Accepted 16 September 2024
Available online 19 September 2024

development of countries around the world. Edible oil, as a daily necessity, is an important part of agricultural production [1]. Since the new century, with the growth of the world population and the globalisation of trade, the trade of oil crops among countries has become increasingly frequent [2]. China is a large agricultural country, but it is also a country with a serious shortage of edible oil resources. Due to a variety of factors, China has been highly dependent on the import of edible vegetable oil in recent years and has become the world's largest importer of edible vegetable oil. According to the statistics of China's grain and oil production and sales in 2019, the self-sufficiency rate of edible oil in China is only 31 %, and nearly 70 % of the edible oil is dependent on imports, which is a major hidden danger to the safety of the edible oil supply in China. *Idesia polycarpa* Maxim (Salicaceae) is a dioecious tree largely distributed in China (Kim et al., 2006 [3]; [4]). The fruits of *I. polycarpa* are named oil grapes in China because they contain large amounts of unsaturated fatty acids (UFA) such as linoleic acid (LA, C18:2, 62.75–71.37 %) [5]. In addition, oil grapes also contain palmitoleic acid (POA, C16:1), oleic acid (OA, C18:1) and α-linolenic acid (ALA, C18:3). The oil content of seeds and pulp in *I. polycarpa* is 22.4%–25.9 % and 43.6 %, respectively [6]. Its fruits are harvested once per year, and their long-chain UFAs are promising feedstocks in the production of biodiesel or edible oil [7]. In addition, *I. polycarpa* oil also contains docosahexaenoic acid (DHA), vitamin E, β-sitosterol, and squalene, which are rare in vegetable oils and benefit human health [4,8,9].

In higher plants, the main chemical constituents of vegetable fats are acylglycerols, of which the main one is triacylglycerol (TAG) (Nelson and Cox, 2000 [10]). The pathway of TAG production has been extensively studied in many model plants ([11]; Li-Beisson et al., 2013). The biosynthesis of plant lipids can be roughly divided into three stages: first, fatty acids (FA) are synthesised in plastids, then, TAG is synthesised in the endoplasmic reticulum (ER), and finally, TAG is combined with oil-body proteins to form the microsomal structure of the oil body (Hills et al., 2004 [12]). Fatty acids are synthetic precursors of TAG and can be divided into saturated fatty acids (SFA), monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA). Acetyl-CoA, as a precursor of FA synthesis, is converted by pyruvate, which can be generated from sucrose by glycolysis [13]. In the plastid, acetyl-CoA is exported to the cytoplasm as acyl-CoAs by enzymes such as acetyl-CoA carboxylase (ACCase), fatty acid synthase complex (FAS), acyl-Acyl-carried proteins thioesterase and acyl-CoA synthetase. Both Acyl-CoAs and glycerol-3-phosphate are precursors in the TAG biosynthetic pathway. In the endoplasmic reticulum, acyl-CoAs are transferred to glycerol-3-phosphate by acyltransferases: glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAT), phosphatidic acid phosphatase (PAP) and diacylglycerol acyltransferase (DGAT) [14]. This acyl-CoA-dependent pathway is commonly called the Kennedy pathway. In addition to the Kennedy pathway, DGAT catalyses the synthesis of triacylglycerol from diacylglycerol (DAG). Two other pathways have been found in recent studies. In the first one, the acyl of sn-2 in phosphatidyl choline (PC) is transferred to DAG by phospholipid diacylglycerol acyltransferase (PDAT), forming TAG and haemolysis phosphatidyl choline. The other pathway was discovered in immature safflower seeds: an acyl of DAG is transferred to another DAG by trans acyltransferase (TA), thus forming TAG and monoacylglycerol, which is a reversible reaction [15]. Finally, the TAGs are packaged in lipid droplets, in specialised organelles, for seed germination and seedling growth.

The biosynthesis of SFA involves ACCase and FAS, wherein the FAS is a symmetrical dimer, with each monomer being a multifunctional complex with seven enzymes and one acyl carrier protein (ACP). The seven enzymes are malonyl-CoA-ACP transacylase (MCAT), acetyl-CoA-ACP transacylase (ACAT), ketoacyl-ACP synthase (KAS), ketoacyl-ACP reductase (KAR), hydroxyacyl-ACP dehydrase (HAS), enoyl-ACP reductase (EAR) and acyl-ACP thioesterase (including FATA and FATB). The ACCase catalyses the first step of fatty acid synthesis, which is the key factor affecting the synthesis of fatty acids. The synthesis of UFA consists of two steps: preliminary synthesis and further desaturation. Fatty acid desaturase (FAD) mainly includes acyl-ACP desaturase, acyl-CoA desaturase and acyl lipid desaturase.

Based on previous studies, KASII is a key enzyme in the extended 16:0-ACP to 18:0-ACP acyl protein synthesis, determining the 16-carbon FA/18 carbon FA ratio [16]. With RNA interference of *KASII* in *Arabidopsis thaliana* expression, the 16:0-ACP levels at late embryonic development are up to 53 %, seven times higher than in the wild type [17]. Stearoyl-ACP desaturase (SAD) can catalyse stearic acid (SA, C18:0) to OA. Quantitative real-time PCR (qRT-PCR) has shown that *GhSAD2* plays a role as dehydrogenase in fatty acid synthesis (Cai et al., 2016 [18]). By inhibiting the expression of *SAD2* in soybean, the content of SA is increased (Ping et al., 2008). Omega-6 fatty acid desaturase (FAD2) can catalyse OA to LA [19], determining oil quality and cold resistance in plants. Our previous study has confirmed that overexpression of *IpFAD2* in *A. thaliana* could catalyse LA synthesis; LA to ALA is catalysed by omega-3 FAD (FAD3). In fruit storage lipids, microsomal FAD2 and microsomal FAD3 are mainly responsible for the transformation from OA to ALA in the endoplasmic reticulum [20,21]. In the Kennedy pathway, GPAT is the first step in the catalytic biosynthesis of TAG. By changing the content of GPAT, it may regulate the metabolism of this pathway and thus control lipid accumulation [22]. Recently, GPAT has been cloned in many plants, such as *A. thaliana*, *Zea mays* and *Pisum sativum*. The DGAT catalyses the final step of TAG biosynthesis [23], making it a necessary enzyme for seed lipid storage [24–26]. In plants [27,28], mammals [29,30] and fungi (Lardizabal et al., 2004 [31]; [32]), DGAT1 and DGAT2 were identified as the key steps responsible for TAG synthesis. However, our understanding of the accumulation and biosynthesis of lipids in *I. polycarpa* has been relatively stagnant in recent years, and the production of oil grape in transgenic plants has not been reported.

In this study, the lipid and FA contents of oil grape from different developing periods were first analysed, and two periods with significant differences in lipid and UFA levels were selected. Because of no reference genome sequence to explore and elucidate the UFA biosynthesis and accumulation mechanism of lipids, *I. polycarpa* developing fruits from two periods were analysed by RNA-sequencing. By comprehensive transcriptome data analyses, qRT-PCR validation and bioinformatic tools, the candidate genes related to UFA biosynthesis were revealed. We propose a biosynthesis model to UFA in oil grape and verify the UFA biosynthesis function of some key genes by molecular biological methods for the first time. Both IpSAD and IpFAD3 can significantly increase the contents of OA and ALA in *A. thaliana* seeds, respectively. This approach could enable the further exploration of genetic engineering breeding of *I. polycarpa* and may also provide a clue for the further investigation of woody biodiesel plants.

2. Materials and methods

2.1. Plant materials and RNA Isolation

The material was a fine single plant of *I. polycarpa*, grown at a row spacing of 2 × 2 m at the experimental nursery of College of Forestry Northwest A&F University (108.06°E,34.25°N, Yangling, China); the sampling method followed Fan et al. [5]. We started harvesting the fruits at 70 days after pollination (DAP), every 10 days until the fruits were completely shed. Oil grapes of 70 DAP and 130 DAP were chosen for RNA transcriptome sequencing analysis, representing the fastest and slowest lipid accumulation stages of *I. polycarpa* fruits, respectively. The RNA was isolated by the MiniBEST Plant RNA Extraction kit (TaKaRa™, China), and the RNA purity and integrity of these samples were detected using an Agilent 2100 Bioanalyzer. The 70 DAP samples were labelled as unripe fruit (U1, U2, U3) in the control group, and the 130 DAP samples were labelled as ripe fruit (R1, R2, R3) in the experimental group, with different numbers representing different biological replicates. Based on this, 25 s/18 s of the RNA samples varied from 1.4 to 1.7, and the RNA integrity number of all samples varied from 7.8 to 9.1, indicating a high-quality RNA suitable for the construction of cDNA libraries.

2.2. Lipid content and fatty acid composition determination

The oil grapes were killed out at 120 °C for 16 h and pulverised. The lipid content was determined as described in Fan et al. [9], extracted by Soxhlet extraction. Total oil grape lipids were methyl-esterified according to the following method. Approximately 0.1 ml of oil grape was mixed with 2 ml of 5 % H_2SO_4 methanol (v/v) for transmethylation. As internal standard, we used 4 mg of glyceryl triheptadecanoate (C17:0) in each sample. Transmethylation was performed at 80 °C for 3 h. The mixture was kept at room temperature until cooled down, and 1 ml of n-hexane and 2 ml of 0.9 % NaCl solution were added to each sample to start the reaction. Subsequently, we centrifuged the mixture at 1500 g for 15 min; the upper layer contained the fatty acid methyl esters (FAMEs) and the n-hexane mixture. The FAMEs were determined by gas chromatography (Shimadzu GC-2010 PLUS, JAP), using the same parameters as described in Fan et al [5]. Each sample was analysed in triplicate.

2.3. cDNA library construction and transcriptome sequencing

The cDNA library construction and *de novo* transcriptome sequencing were performed by TSINGKE Co. Ltd (Beijing, China). In brief, mRNAs were enriched with magnetic beads containing Oligo (dT), and the mRNAs were fragmented into 350-bp sequences using fragmentation buffer and re-transcribed into first-strand cDNA using random hexamers. Subsequently, dNTPs, RNase H and DNA polymerase I were added to synthesise the second-strand cDNA, and the double-stranded (ds) cDNA was purified. The purified ds-cDNA was subjected to end repair, addition of poly-A tail, and ligation sequencing adapters. Finally, the ligated cDNA was amplified by PCR, and the established sequencing library was used for sequencing. After generating the library, an Agilent 2100 Bioanalyzer was used to detect the inserted fragment scope of the library, and the ABI StepOnePlus real-time PCR System was used to quantify the library concentration. After quality inspection, the constructed cDNA libraries were quantified on the Illumina NovaSeq 6000 HiSeq platform.

2.4. De novo assembly of the transcriptome

Raw data were first filtered by an internal software to remove the reads containing adapters, reads containing more than 5 % ambiguous poly-N and low-quality reads (reads with Phred score less than 10 accounted for more than 20 % of the total base as low-quality reads) to generate clean reads. Data volume and quality value of clean reads were calculated, and all clean reads were *de novo* assembled by the Trinity v. 2.0.6 software (Eisen et al., 2001 [33]), with Inchworm K = 25; the other parameters were min contig length 150, CPU 8, min kmer cov 3, min glue 3, bfly opts '-V 5-edge -thr = 0.1-stderr'. The assembled transcript was clustered, and redundancies were removed by Tgicl (Hoon et al., 2004) and assessed with BUSCO (http://busco.ezlab.org). The Tgicl software version v. 2.0.6 was used, and the parameters were I 40, c 10, v 25, O '-repeat stringency 0.95-minmatch 35-minscore 35'.

2.5. Unigene functional annotation

The reference genome is from Zuo [34].Assembled unigenes were annotated using BLAST (v 2.2.25, default parameters, http://blast.ncbi.nlm.nih.gov/Blast.cgi) [35] on the databases NT(NCBI nucleotide sequences), NR(NCBI non-redundant protein sequences), COG (Clusters of Orthologous Groups of proteins), KEGG (Kyoto Encyclopedia of Genes and Genomes) and SwissProt (a manually annotated and reviewed protein sequence database). The corresponding software and parameters were NCBI blast (v 2.2.25, *E-value* = 1E-05) for NT, NR, KEGG and SwissProt, NCBI blast (v 2.2.25, *E-value* = 1E-03) for COG.The GO annotation was performed using Blast2GO (v. 2.5.0, default parameters, *E-value* = 1E-06, https://www.blast2go.com) (Conesa et al., 2005) and NR annotation results.

2.6. Differential expression analysis

Clean reads were aligned to the assembled reference sequence by Bowtie 2 (v. 2.26, default parameters, http://bowtie-bio.

sourceforge.net/bowtie2/index.shtml) (Langmead et al., 2012). The Expression levels of unigenes and transcripts were calculated by RSEM (v. 1.2.12, default parameters, http://deweylab.biostat.wisc.edu/rsem) (Li et al., 2011 [36]) and measured by FPKM (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) (Florea et al., 2013). The DESeq2 was used for differential expression gene (DEG) analysis by the negative binomial distribution principle, according to Anders [37]. The *P*-values were adjusted by the Benjamini-Hochberg procedure to control the false discovery rate (FDR). Unigenes with |fold change| $\geq$ 2.00 and an adjusted *P*-values (*P*-adj) $\leq$ 0.05 were defined as DEGs. According to the results of DEG testing, DEGs were clustered by the pheatMap function in the R software for hierarchical clustering analysis. When multiple groups of differentially expressed genes are clustered at the same time, the differentially expressed genes in the intersection and those in the union are separately analysed by clustering. According to the KEGG annotation results and the official classification, we classified the differentially expressed genes in KEEG pathways and used the phyper function in the R software to conduct enrichment analysis. Generally, pathways with FDR $\leq$0.01 are considered to be significantly enriched.

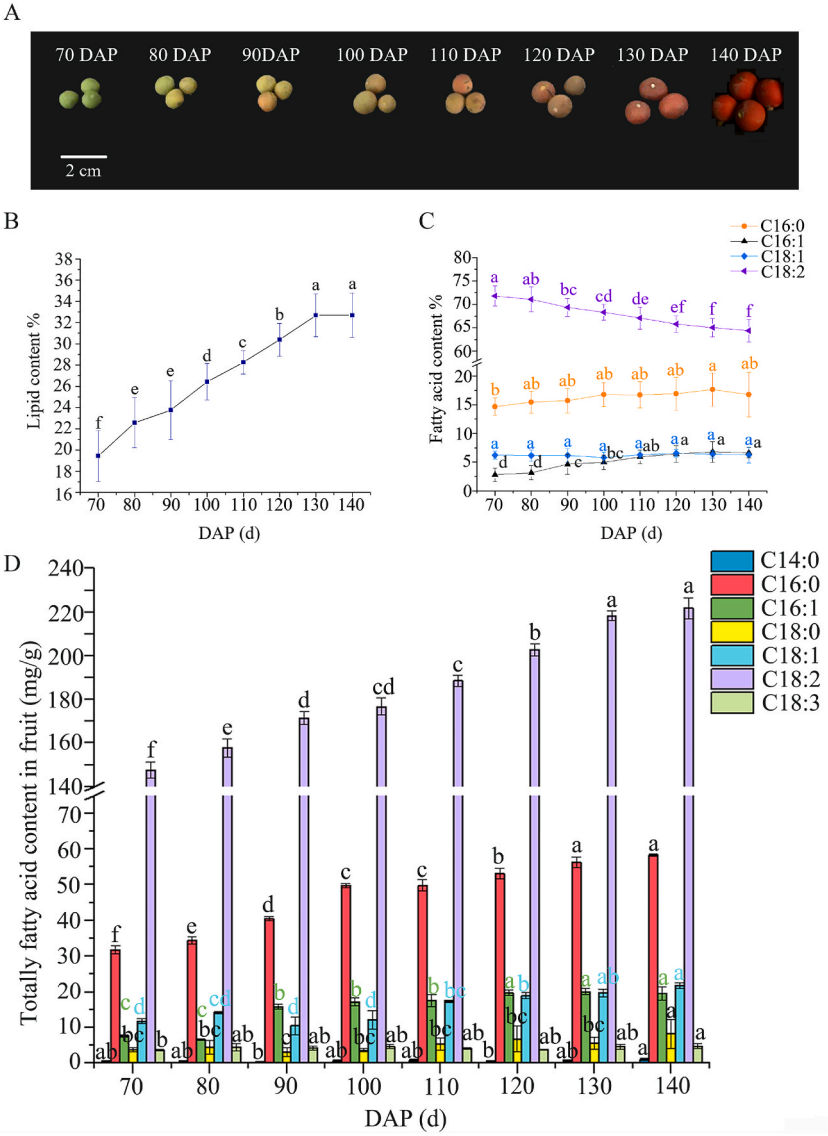


Fig. 1. Lipid and fatty acids of oil grape. (A) Developing oil grape in various periods. (B) Lipid content of oil grape in various periods. (C) Major fatty acid content of oil grape in various periods. (D) Major fatty acid accumulation of 100 g oil grape (WT) in various periods. All experiments were triplicates in biological replicates. Each column represents the mean of three biological replicates ±SD. Different letters above the columns indicate significant differences at $P \leq 0.05$.

2.7. Cloning and Bioinformatical analysis of full-length candidate genes

Full-length sequences of *GPAT, DGAT2, KASII, SAD, FAD2, FAD3* and *FAD8* were cloned by the RACE technology using the primers in Table S1; some of the sequences were uploaded to the NCBI. The orthologs from other species were determined via BLAST on the website NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The ORF of the full-length sequence was discovered by the ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The amino acid sequences of *I. polycarpa* and orthologs from other species sequence alignments were analysed using the software package DNAMAN v6, and corresponding phylogenetic trees were constructed with the neighbour-joining (NJ) method in the MEGA 6 software. The reliability of each tree was determined by bootstrap analysis with 100 replicates.

2.8. Validation of candidate genes by qRT-PCR

The validation of candidate genes was analysed by quantitative real-time PCR (qPCR). Seven primer pairs and control gene *UBA80* were designed by TSINGKE (Beijing, China), and the qPCR reactions were performed by the PrimeScript™ RT reagent kit (TaKaRa™, China). In brief, the final reaction volume was 10 μl, with the following standard amplification conditions: 98 °C for 0.5 min, 40 cycles of 5 s at 98 °C and 0.5 min at 58 °C, with a final step at 72 °C for 5 min. Each gene had a triplicate biological and technical duplication. The $2-\Delta\Delta Ct$ method was used to calculate the relative expression levels.

2.9. Function analysis of LA metabolism genes

The genes *IpSAD* and *IpFAD3*, which were responsible for the producing upstream and downstream metabolites of LA, were verified in a similar way as *IpFAD2* and *IpDGAT2* in our previous studies [5]. The plant expression vector pRI-101 (35s promoter) was recombinated with *IpSAD* and *IpFAD3* by *Xba*I and *Sal*I restriction sites, respectively. Recombinant plasmids were transformed into *A. thaliana*, and T_3 overexpression-positive seeds were selected for detection by GC [9].

3. Results

3.1. Lipid and fatty acid composition of I. polycarpa

In this part, we collected eight fruit samples at different periods and analysed the lipid content levels (Fig. 1A). During the growth periods, the lipid levels (19.43%–32.69 %) continuously increased until maturity, albeit with different increase rates (Fig. 1B). Except for 90 DAP and 130 DAP, the lipid contents of fruits increased significantly. During oil grape development, there were seven predominant fatty acids, namely myristic acid (C14:0), palmitic acid (PA, C16:0), POA (C16:1), stearic acid (SA, C18:0), OA (C18:1), LA (C18:2) and ALA (C18:3). In accordance with the accumulation pattern from the preparation to the mature stage, the relative content of LA decreased from 71.81 % to 64.38 %, and the proportions of C16:0 and C16:1 increased from 14.65 % to 17.62 % and from 2.78 % to 6.23 %, respectively (Fig. 1C). The contents of C14:0, SA and ALA were below 5 % in the growth periods, and the proportions of OA showed no obvious change. In addition, the composition of C16:0, C16:1, C18:2 and the lipid content in oil grape significantly differed at 70 DAP and 130 DAP.

To further dissect the relationship between fatty acid content and fruit quality, we predicted the contents of various fatty acids in 100 g of fruit, based on the previous results (Fig. 1D). Based on these findings, LA is the main component of fatty acids in oil grape, and its level directly affects fruit quality. This is consistent with the findings of a previous study [9]. In addition, we also should pay attention to the effect of the C16:0 and C16:1 contents on fruit quality. The fatty acid accumulation dynamic changes during oil grape growth periods provide useful information for identifying underlying enzyme reactions in UFA biosynthesis; these reactions are essential to genetic engineering breeding.

3.2. Illumina sequencing and de novo assembly of oil grape unigenes

In total, 18.73 Gb and 17.95 Gb clean reads were generated with 124.8874 M and 119.655 M clean base pairs at 70 DAP and 130 DAP, respectively. Among them, the Q20 and Q30 values of each sample were above 97.43 % and 92.83 %, and the GC content of each sample ranged between 43.19 % and 43.45 % (Table S2). Because there was no *I. polycarpa* genome sequence, all clean reads were *de novo* assembled using the Trinity software (Eisen et al., 2001), and redundancies were removed by Tgicl. A total of 115,350 unigenes were obtained, with a mean length of 1414 bp and an N50 size of 2179 bp (Table S3). The lengths of these unigenes varied from 200 to 15,660 bp and showed a consistent trend distribution (Fig. S1). The percentage of the complete BUSCO score was 91.0 %, and the length of N50 was larger than the mean length of unigenes (Fig. S2), indicating a high-integrity assembly.

3.3. Functional annotation of I. polycarpa unigenes

All the 115,350 unigenes were annotated against six public databases, including NR, NT, Swiss-Prot, KEGG, GO and COG. Of these, 92,643 (80.31 %) unigenes were annotated in known sequences of at least one database (Table S4); among of these, 84,231, 85,004, 67,091, 65,154, 41,482 and 53,380 unigenes aligned with the above databases, respectively. The NT possessed the highest number of annotated unigenes in these public databases.

The unigenes annotated from the NR database were aligned by BLAST (Fig. S3). There were 42,893 (50.92 %) and 31,154 (36.99 %) homologous unigenes that matched with *Populus euphratica* and *Populus trichocarpa*, respectively. In addition, the mapping of these unigenes to other species was less than 1 %. The results show that *I. polycarpa* has a high homology with the two Salicaceae species. For this reason, the latest phylogeny classification has placed *I. polycarpa* in the Salicaceae instead of the Galericaceae [38].

To evaluate the oil grape unigenes, functional categorisation was conducted in the COG database, where 41,482 unigenes were assigned to 25 functional classifications (Fig. S4). The largest class was the "general function prediction only" class with 12,176 unigenes, followed by the "Transcription" (5721). Only 32 and 4 unigenes were annotated to "Nuclear structure" and "Extracellular structures", respectively. It was noteworthy that 1781 unigenes were matched to "Lipid transport and metabolism", which may be closely related to the lipid metabolism of oil grape.

Regarding specific functions, GO analysis results showed that a total of 53,380 unigenes were divided into three main categories (Fig. S5). We could relate 26 subcategories to biological processes (BP) and 16 and 12 to cellular component (CC) and molecular function (MF), respectively. In the BP category, "metabolic process" and "cellular process" were the main subcategories, containing 29,974 and 28,333 unigenes. In the CC category, "cell" (23,374) and "cell part" (23,199) occupied a high percentage. For the MF category, the major subcategories were "binding" (26,226) and "catalytic activity" (26,137).

For KEGG annotation, 65,154 unigenes were divided into 6 KEGG categories and subdivided into 21 subcategories (Fig. 2). "Metabolism" was the largest category, with 14,396 unigenes in six main categories. In addition, 3048 unigenes were annotated in the "lipid metabolism" subcategory; these pathways might be crucial for fatty acid metabolism in oil grape.

3.4. Differentially expressed gene (DEG) analysis in developing oil grape

To estimate the gene expression levels of DEGs and identify candidate genes involved in fatty acid metabolism in oil grape, we introduced the FPKM values and collected the $\log_{10}$(FPKM+1) distribution of transcriptome. The integral FPKM values slightly increased from the preparation stage (70 DAP) to the mature stage (130 DAP). In the violin figure, the medians and other statistical magnitudes of two samples were almost the same (Fig. S6). Biological duplication is essential for any biological experiment, and high-throughput sequencing is no exception. The correlation indices in three biological replicates were all approximately 1 (Fig. S7). These results ensure the reliability of the subsequent differential gene analysis.

The DEGs were identified by DESeq, based on FPKM values. Comparing three biological replicates from 70 DAP and 130 DAP, we obtained 4382 (2182 up-regulated and 2202 down-regulated) DEGs with significantly increased or decreased FPKM values ($|\log_2$(Fold Change)$| > 1$ and q-value <0.001). There were 3455 DEGs mapped to 130 KEGG pathways. The 233 lipid metabolism DEGs (Fig. S8) were assigned to 15 pathways (Table S5), such as "Biosynthesis of unsaturated fatty acids"(23 DEGs), "Linoleic acid metabolism"(11 DEGs), "Fatty acid biosynthesis"(25 DEGs), "Fatty acid metabolism"(44 DEGs), "Fatty acid elongation"(20 DEGs) and "Glycerolipid metabolism"(47DEGs). These DEGs should be related to the fatty acid content difference between 70 DAP and 130 DAP; their specific functions need, however, more detailed investigations. All assembled Unigenes and DEG data are listed in Table S6.

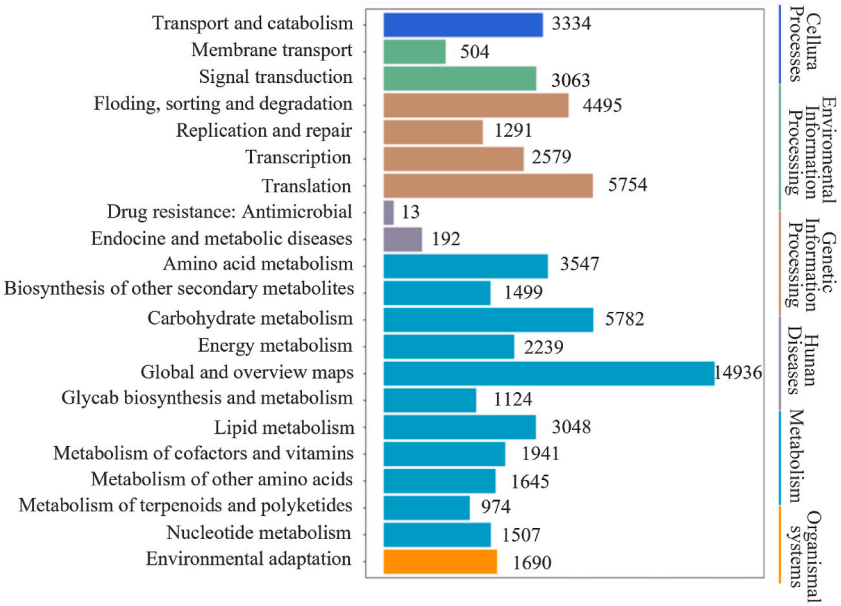


Fig. 2. KEGG annotation classification of *I. polycarpa.* The X-axis represents the number of unigenes, on the Y-axis, each colour represents a category, and each column is a subclass. The unigenes are distributed into six main categories: cellular processes, environmental information processing, genetic information processing, human diseases, metabolism and organismal systems.

3.5. Detailed analysis of candidate genes involved in lipid metabolism

In higher plants, the biosynthesis and accumulation of fatty acids is composed of multiple pathways [39]. Consequently, the DEGs are hierarchically clustered based on their expression patterns. A KEGG pathway enrichment analysis of DEGs demonstrated rich factors of the two pathways "Metabolic" and "Biosynthesis of secondary metabolites" (Fig. S9), indicating that the DEGs involved in these pathways were highly enriched and assigned to at least one pathway in Fig. S8. Candidate genes involved in lipid metabolism showed a higher expression level at 130 DAP (Fig. 3).

During oil grape development, the accumulation of lipid and fatty acids increased from 70 DAP to 130 DAP (Fig. 1B and D). Hence, DEGs assigned to lipid metabolism were screened by the FPKM value and the pathway-based analysis in Table S5. The 65 target genes are listed in Table S7; four transcription factors (*ABI3*, *FUS3*, *LEC2* and *WRI*) which related to lipid biosynthesis had a high expression level in fruits. The *ACC*, *MCAT*, *KAR*, *HAD*, *KAS I*, *KAS II*, *KAS III*, *SAD*, *FAD2*, *FAD3*, *FAD8*, *GPAT*, *LPAT*, *DGAT1* and *DGAT2* were considered to be important in TGA and UFA biosynthesis. Because there is no genome of *I. polycarpa*, these unigenes were compared to *Populus euphratica* and *Populus trichocarpa* through BLAST analysis in NCBI. These unigenes shared 74–100 % similarities, which might be due to differences among orthologous genes among different species. In this sense, these genes of *I. polycarpa* are orthologues similar to those of *P. euphratica* and *P. trichocarpa*. In addition, we cloned full-length sequences of *IpDGAT2*, *IpGPAT*, *IpKASII*, *IpSAD*, *IpFAD2-1*, *IpFAD3* and *IpFAD8* by the RACE technology. The sequences of these candidate DEGs and phylogenetic trees are listed in Supplementary Material 1.

Additionally, to validate the reliability of the FPKM values in the oil grape transcriptome, the expression patterns of target genes *IpDGAT2*, *IpGPAT*, *IpKASII*, *IpSAD*, *IpFAD2-1*, *IpFAD3* and *IpFAD8* which are highly important for *I. polycarpa* UFA biosynthesis were validated by q-PCR at 70 DAP and 130 DAP. The qPCR results show the veracity of the transcriptome data (Fig. 4); these candidate target genes were specifically expressed in different periods, which leads us to infer that they are important in the biosynthesis of UFA.

3.6. Candidate genes related to TAG assembly

Overall, we found thirteen acyltransferases; *IpDGAT2* had three transcripts, *IpGPAT9* and *IpDGAT1* had two transcripts, whereas *IpGPAT5*, *IpGPAT8*, *IpLPAT1-5*, *IpPDAT1-2* and *IpDGAT3* only had one transcript. The *IpLPAT4*, *IpLPAT5* and *IpPDAT2* were expressed

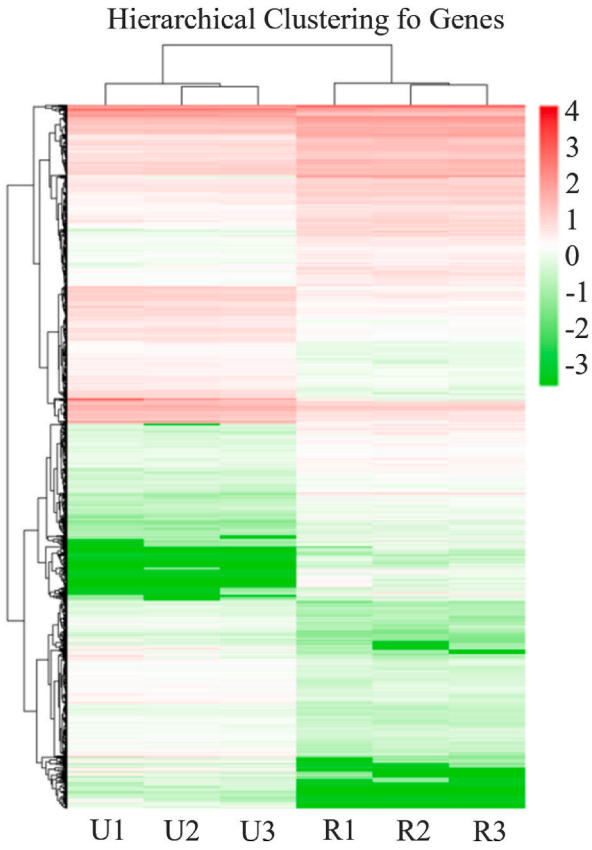


Fig. 3. Hierarchical clustering of oil grape in 70 DAP and 130 DAP. The horizontal axis represents the sample name and the clustering result of the sample, U and R represent samples at 70 DAP and 130 DAP. The vertical axis represents the candidate gene and the clustering result of the candidate gene. The different colors represent the gene expression level in the sample ($\log_2$(FPKM + 1)).

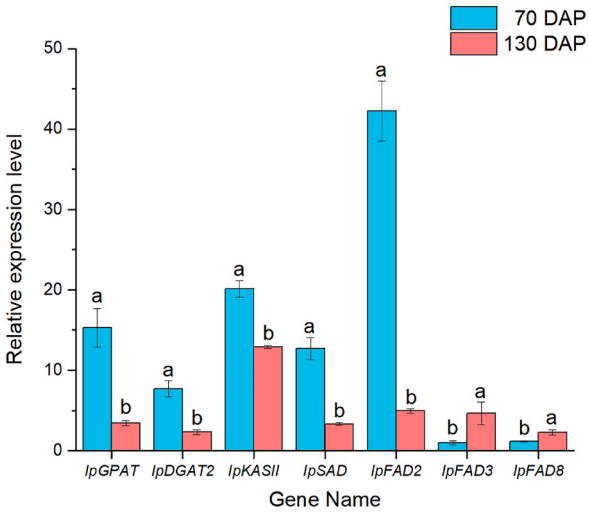


Fig. 4. Expression patterns of lipid biosynthesis DEGs validated by q-PCR. The internal standard gene is *UBA80*. Experiments were performed in triplicates in biological replicates.

at a relatively high level at 130 DAP, whereas other acyltransferases showed higher expression levels at 70 DAP (Table S7). We elucidated the Kennedy pathway for the production of TAGs. In this traditional pathway, the acyl-CoA of three acyl groups incorporate glycerol by acyltransferases, with GPAT, LPAT and DGAT being the three main acyltransferases [40]. The expression patterns of *IpDGAT2* and *IpGPAT* were similar, with three distinct peaks. In high plants, *GPAT* and *DGAT2* are mainly responsible for the first and final step of the Kennedy pathway, respectively. Therefore, the expression pattern of *IpGPAT* in *I. polycaipa* should be analysed in combination with that of *IpDGAT2*. The expression pattern of *IpDGAT2* was similar to that of *IpGPAT*, but there was a significant lag, that is, when the expression level of *IpGPAT* increases or decreases at a certain moment, the expression level of *IpDGAT2* in the next period also shows an upward or downward trend. This is consistent with the pattern of lipid accumulation in *I. polycaipa*, with the high expression level of *IpDGAT2* corresponding to the rapid increase in the lipid content (Fig. 1B and 5). After 130 DAP, the fruits were ripe, and the lipid concentration was highest. At the same time, the lipid content did not increase in this period, and the expression levels of *IpDGAT2* and *IpGPAT* were low. Expect for the Kennedy pathway, an acyl-CoA-independent pathway is another metabolic pathway for the biosynthesis of TAGs and involves *PDAT* to produce TAG. We found two *IpPDAT* expressed in oil grape, and their expression levels were significantly different. The *IpPDAT1* was expressed at a low level at 130 DAP, and *IpPDAT2* was expressed at a high level at 130 DAP. These genes might be an important factor in TAG biosynthesis.

The expression of *IpDGAT2* was much higher than that of other acyltransferases, which is a main factor in oil grape TAG assembly. In a previous study, silencing of *NtDGAT2* in *Nicotiana tabacum* caused the tobacco seed weight to reduce by 50 % [40]. In *A. thaliana* and *Nicotiana tabacum*, overexpression of *AtDGAT2* in leaves could significantly increase the TAG content and seed weight significantly [41]. In recent studies on *V. galamensis* and *R. communis*, *DGAT2* not only catalysed the biosynthesis of TAGs but also sequestrated unusual FAs into TAGs, such as aleuritic acid [42,43]. In our previous study, heterogeneous expression of *IpDGAT2* in *A. thaliana* caused increased seed lipid contents by 27.70 %, 34.34 %, and 39.09 % [5]. Combined with previous studies, increasing the *IpDGAT2* expression levels of *I. polycarpa* is a strategy to achieve high lipid contents of oil grape.

3.7. Candidate genes related to UFA assembly

The catalysis of the plastidial elongation of fatty acids by KASs has been demonstrated in the model plant *A. thaliana* [17]. In this study, we annotated *IpKAS I*, *IpKAS II* and *IpKAS III*. The RNA-Seq result illustrated that the expression of *IpKAS II_1* increased, whereas that of *IpKAS* decreased at 130 DAP. Combined with previous research and the fatty acid contents, the levels of C16:0 and C16:1 increased gradually during the oil grape growth period (Fig. 1C and D). We therefore assume that *IpKAS II_2* is a key enzyme in extending 16:0-ACP to 18:0-ACP acyl protein synthesis, which decided the C16 FA/C18 FA ratio in *I. polycarpa*. The q-PCR result showed that *IpKAS II_2* was expressed at a high level at 70 DAP, followed by a stable expression during the remaining development periods. The ratio of C18 was highest at 70 DAP and then decreased continuously (Fig. 5). This indicates that the synthesis of 18:0-ACP mainly occurred in the preparation stage of the oil grape development.

The UFA accumulation can be divided into two phases: preliminary synthesis and further desaturation. The SFA was desaturated by genes encoding for the plastid-localised FAD (*FAD4, FAD5, FAD6, FAD7* and *FAD8*), and the ER-localised FAD (*FAD2* and *FAD3*). In addition, long-chain acyl-CoA synthetase (*LACS*), *FATA* and *FATB* are also responsible for FA release (Bates et al., 2013; Li-Bession et al., 2013). We noted 12 FAD-related genes with differential expression levels (Table S7). There were three FADs annotated as SA desaturase; the *IpSAD_3* had a high expression level at 130 DAP, but *IpSAD_1* and *IpSAD_2* had high expression levels at 70 DAP. The SA (C18:1) content stably decreased during the developing periods of oil grape, which means that stearoyl-ACP-desaturase plays an

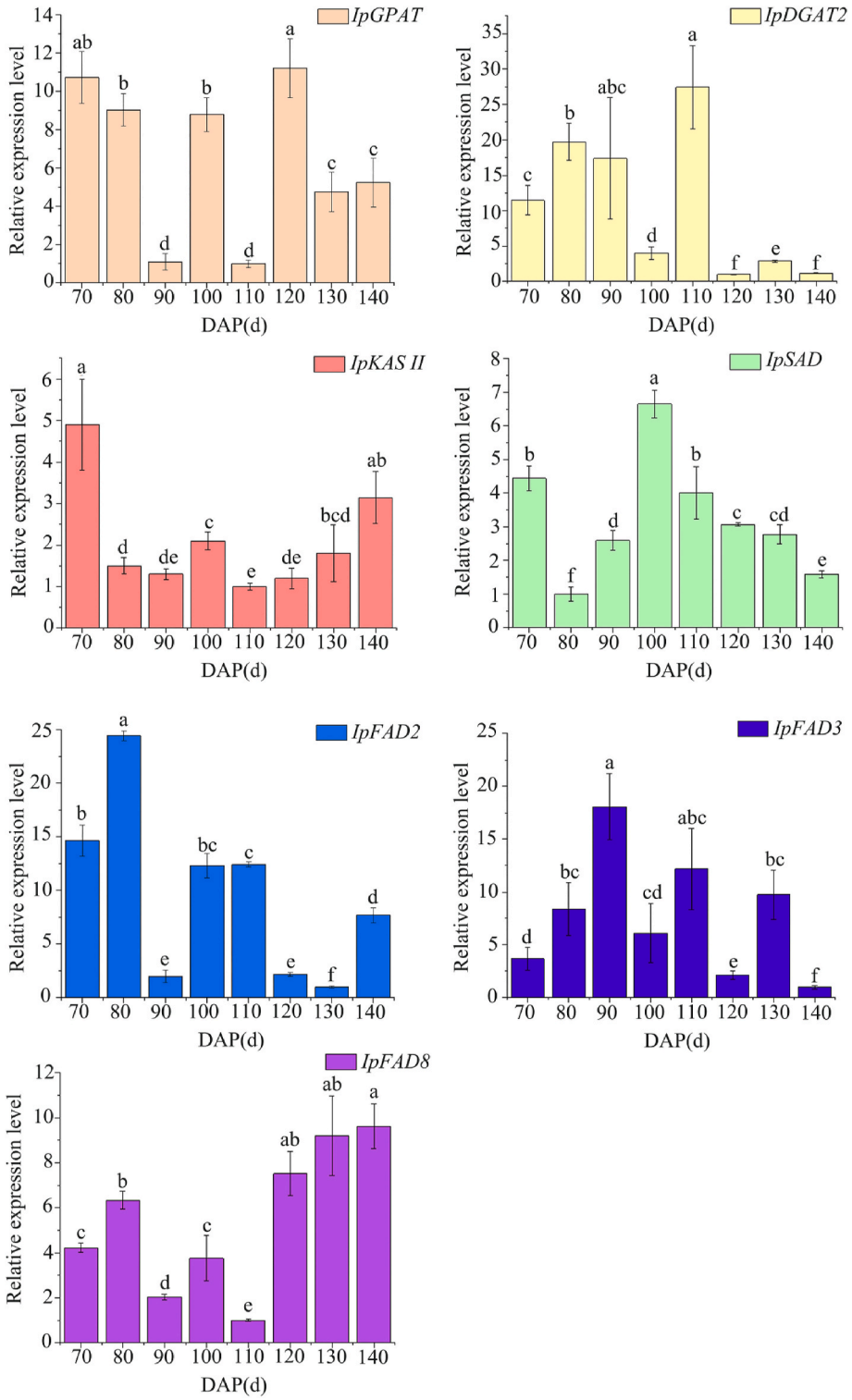


Fig. 5. Expression levels of seven target UFA-related DEGs of *I. polycarpa* in fruits at growth periods 70−140 DAP. The internal standard gene is *UBA80*. Experiments were performed in triplicate in biological replicates.

important role in OA biosynthesis. Compared with *IpSAD_2*, the expression level of *IpSAD_1* was more significant, and therefore, we selected *IpSAD_1* as candidate gene. The q-PCR analysis of *IpSAD* showed that it had a high expression level during 100 DAP to 110 DAP, and the content of OA was not significantly increased during this period (Fig. 1C and 5).

Since LA (C18:2) is the most abundant fatty acid in oil grape, *FAD2* is the major gene we should focus on (Hernandez et al., 2016 [44]). We found six omega-6 FAD genes by RNA-Seq (Table S7). Among them, five genes showed a significant difference between 70 DAP and 130 DAP and were localised in ER. The other gene, *IpFAD6*, localised in chloroplasts, had a relatively low expression in oil grape. The BLAST result of *IpFAD2-2* and *IpFAD2-3* genes showed consistency with *IpFAD2* (MF 693460), which we had cloned and analysed in a previous study [9]. Interestingly, the *IpFAD2-2* expression level was among the top 10 at 70 DAP and 130 DAP. The expression levels of four *IpFAD2* were high at 70 DAP, which is consistent with the LA content in oil grape. The BLAST result showed that *IpFAD2-1* showed 99 % similarity with MH 394210.1, which was 74.8 times higher in the seed than in the pericarp [8]. However, *IpFAD2-1* (MH 394210.1) and *IpFAD2-2* (MF 693460) exhibited the opposite expression patterns; the former was mainly expressed in the seed and the latter in the pericarp. Considering the expression levels and the proportions of seed and pericarp in fruit, we suggest that oil grape used distinct *FAD2* genes for LA biosynthesis, of our transcriptome data, MF 693460 was the main factor gene. The q-PCR result showed that the expression of *IpFAD2* increased at the early fruit developing stages and peaked at 80 DAP, followed by a gradual decline (Fig. 5). Another function of *FAD2* is to increase cold resistance in plants, especially seeds. A high PUFA level results in high cold resistance, which might be the reason for the increase in the expression of *IpFAD2* at 140 DAP.

Both *IpFAD3* and *IpFAD8* are two omega-3 FADs with high expression levels at 130 DAP (Table S7); they are responsible for the conversion of phosphatidylcholine-bound linoleate to linolenate within the endoplasmic reticulum, thus regulating the amount of C18:3 in seed oil [45]. The former is located in the ER and the latter in the chloroplast. Both *IpFAD3* and *IpFAD8* are responsible for the content of ALA in oil grape, and the expression levels of *IpFAD2*, *IpFAD3* and *IpFAD8* determine the ratios of LA and ALA as well as the PUFA content, thus affecting the quality of oil grape. Moreover, the activity of omega-3 FAD is also affected by temperature; the expressions of *IpFAD3* and *IpFAD8* in fruits were increased by low temperature but inhibited by high temperature, and the expression level of *IpFAD7* decreased under low-temperature stress. These results are consistent with previous findings on *Perilla frutescens* and *Descurainia sophia* (Tang et al., 2007).

3.8. Function analysis of LA metabolism in I. polycarpa

Three transformants (T_3), control stain and wild type (Col-0) were identified by PCR. Based on the results, pRI-101, *35S::IpSAD* and *35S::IpFAD3* were successfully transferred to *A. thaliana*. Cultivated under standard conditions, transgenic plants showed no obvious differences. The FA contents of seeds were determined and compared by GC. For the three transgenic lines of *IpSAD* overexpression in *A. thaliana* seeds (Fig. 6), the OA contents were 16.84 %, 17.38 % and 17.66 %, respectively, increasing by 10.43%–15.80 % compared to 15.25 % in Col-0. The eicosenoic acid (EOA, C20:1) contents were 16.63 %, 16.54 % and 16.46 %, respectively, increasing by 13.91%–15.09 % compared to the 14.45 % in Col-0. In the transformed seeds, the contents of OA and EOA increased significantly, whereas those of LA and ALA decreased significantly. In addition, the PA contents decreased significantly in transgenic line #7-6-18, resulting in total SFA contents in this line of only 16.04 %, which decreased by 4.41 % compared to the 16.78 % in Col-0. On the contrary, the total contents of SFA in other transgenic lines were 17.17 % and 18.46 %, increasing by 2.14 % and 10.01 % compared to Col-0. In addition, the eicosanoic acid (EA, C20:0) contents increased significantly in transgenic line #5-2-25. Other FA contents in transgenic lines showed no significant change compared with the control strain and Col-0.

For *IpFAD3* transgenic seeds (Fig. 7), the ALA contents were 13.13 %, 13.84 % and 13.49 %, respectively, increasing by 25.77%–29.21 % by compared to the 10.44 % in Col-0. The contents of ALA increased significantly, whereas those of LA decreased significantly. In three transgenic lines, LA contents were 32.62 %, 32.49 % and 32.34 %, decreasing by 16.66%–17.37 % by compared to the 39.14 %

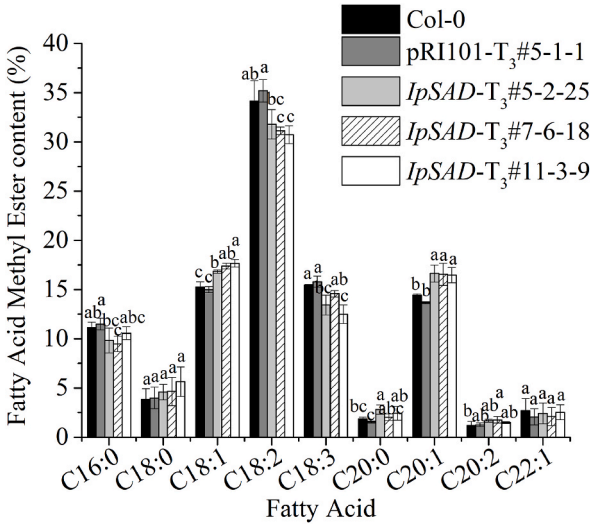


Fig. 6. Fatty acid methyl ester (FAME) contents in *IpSAD* overexpressed in *A. thaliana*. Different letters indicate significant differences at $P \leq 0.05$. Fatty acid proportions are expressed as mole percent. Experiments were performed in triplicate.

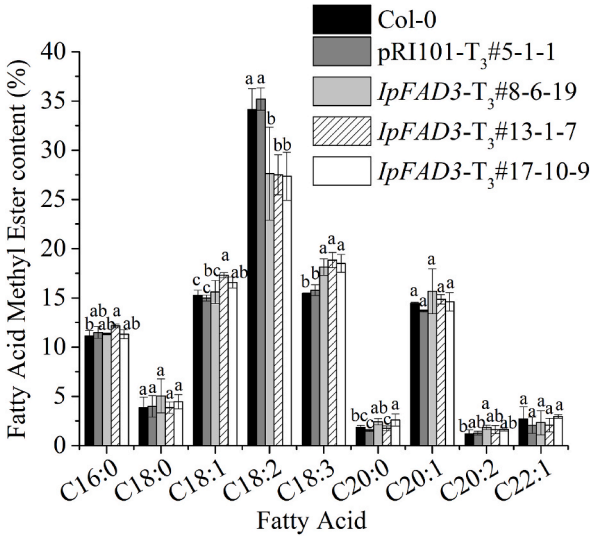


Fig. 7. Fatty acid methyl ester (FAME) contents in *IpFAD3* overexpressed in *A. thaliana*. Different letters indicate significant differences at $P \leq 0.05$. Fatty acid proportions are expressed as mole percent. Experiments were performed in triplicate.

in Col-0. Except for transgenic line #13-1-7, which showed a significant increase in C16:0 and OA contents, the other FA contents in transgenic lines showed no significant change. The total SFA contents were 18.46 %, 17.55 % and 18.13 %, respectively, increasing by 6.69%–12.22 % by compared to the 16.45 % in Col-0.

3.9. UFA biosynthesis pathway in I. Polycarpa

Using our cumulative data from transcriptomic comparison and biochemical analysis of 70 DAP and 130 DAP in oil grape, we propose a molecular basis for the UFA biosynthesis pathway for *I. polycarpa* (Fig. 8). In brief, at the beginning of FA biosynthesis, in the plastid, acetyl-CoA generates malonyl-CoA by ACC, which is then transferred to malonyl-ACP by MACT. Subsequently, cycled condensations of malony-ACP are catalysed by KAS, KAR, HAD, EAR. After six cycles, 16:0-ACP is generated, and after seven cycles, catalysed by KASII 18:0-ACP is generated. In addition, SAD may desaturate 16:0-ACP and 18:0-ACP to 16:1-ACP and 18:1-ACP in the plastid. Subsequently, acyl-ACP thioesterase (FATA and FATB) transport fatty acids from the plastid to the ER as acyl-CoAs. It should be noted that the reaction of 16:0-CoA synthesis to 18:0-CoA requires malonyl-CoA rather than acetyl-CoA, which was previously involved in the cyclic reaction; this step is irreversible. Under the action of FAD, 18:1-CoA is desaturated to 18:2-CoA and 18:3-CoA in the ER, and subsequently, 16:0-CoAs, 16:1-CoAs, 18:1-CoAs, 18:2-CoAs and 18:3-CoAs are assembled on the G3P by GPAT, LPAT and DGAT2. Finally, TAG is combined with oil body proteins to form the microsomal structure of the oil body in the cytoplasm. This pathway of UFA biosynthesis is consistent with that in most high plants, such as *Coriandrum sativum* [16].

4. Discussion

4.1. Correlation of lipid content and transcriptomes

As a woody plant species used for the production of biodiesel and edible oil, *I. polycaipa* provides oil with a high quantity and quality. It is an important new raw material species for food, health care products, medicine and the chemical industry, with a high development value and a huge market potential. In this study, we detected the accumulation patterns of lipid and FA contents at various growth periods of oil grape. Subsequently, *de novo* assembly transcriptomes at different time points, 70 DAP and 130 DAP, were functionally annotated in different public databases. The DEGs involved in lipid metabolism were screened by the DESeq2 method. The regulations and roles of some main UFA biosynthesis genes were selected and analysed, and finally, the expression patterns of some key genes were validated by q-PCR at different growth periods. Overall, the transcriptional patterns of lipid metabolism and the temporal accumulation patterns of lipid and FA contents were systematically analysed for the different time points to elucidate the metabolic and molecular mechanisms leading to differences in UFA biosynthesis between these two time points. Based on the results, we conclude the directed and regulated biosynthesis of UFA in oil grape. A better understanding of UFA accumulation in oil grape may facilitate ways to modify it, such as genetic engineering breeding of *I. polycarpa*.

Oil grape has high amounts of lipids (19.43%–32.69 %) and LA (64.38%–71.81 %) during fruit formation to maturation, whose contents differed greatly between 70 DAP and 130 DAP. In addition, the contents of C16:0 and C16:1 also significantly differed between 70 DAP and 130 DAP; this indicates that fruits from these two periods have different biochemical mechanisms for UFA biosynthesis. We assume that these differences are based on the mechanisms for lipid accumulation and the external environment. In

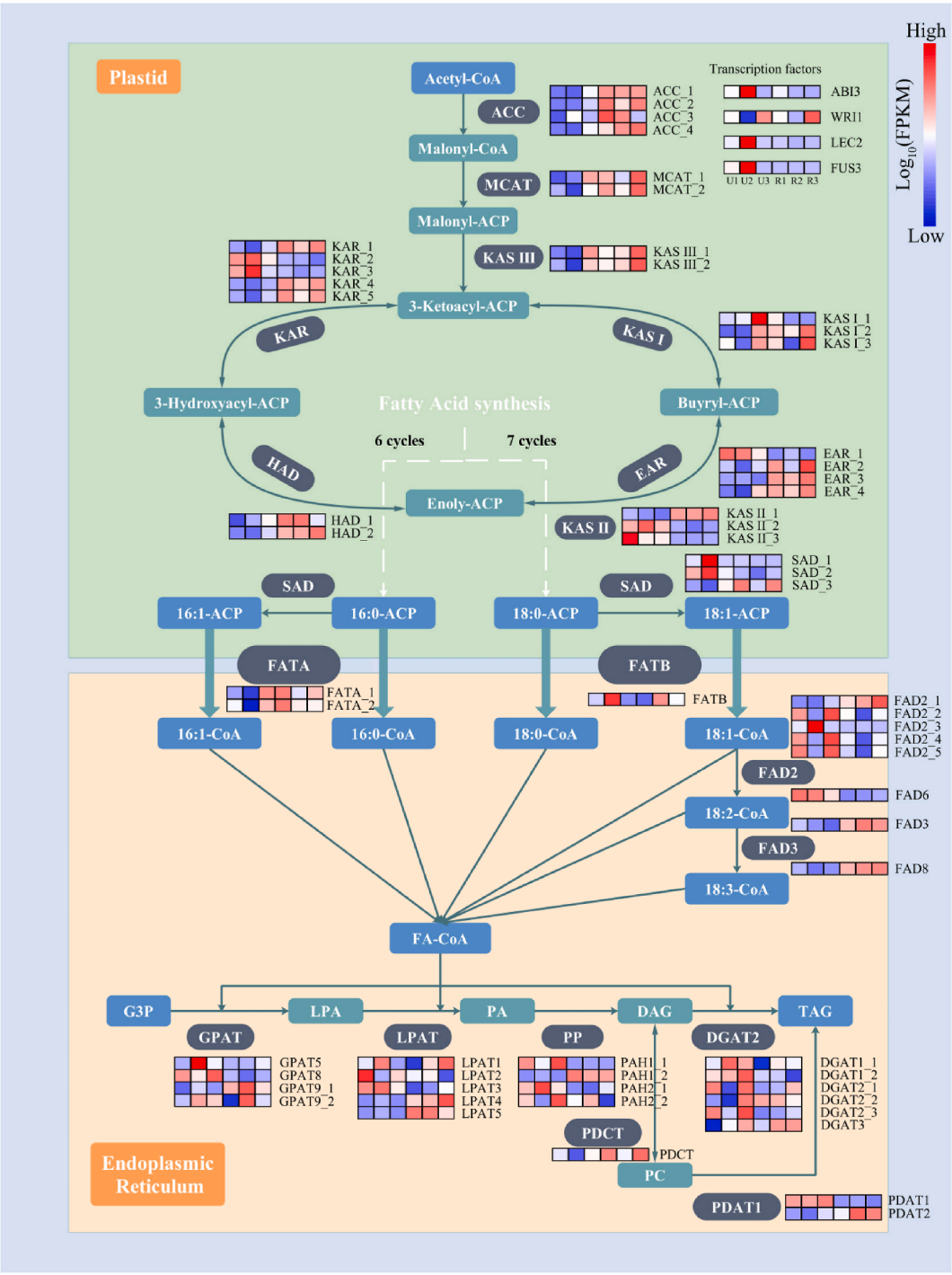


Fig. 8. Predicted UFA and TAG biosynthesis pathways in *I. polycarpa*. The rectangles of different colors delineate biosynthesis reactions in the plastid and endoplasmic reticulum, respectively. Abbreviations of enzyme genes: ACC, acetyl-CoA carboxylase; MCAT, malonyl-CoA: ACP transferase; KAS, 3-ketoacyl-ACP synthetase; KAR, 3-ketoacyl-ACP reductase; HAD, 3-hydroxyacyl-ACP dehydratase; EAR, 2-enoyl-ACP reductase; SAD, stearoyl-ACP 9-desaturase; FATA, acyl-ACP thioesterase A; FATB, acyl-ACP thioesterase B; FAD2, omega-6 fatty acid desaturase; FAD3, omega-3 fatty acid desaturase; GPAT, glycerol 3-phosphate acyltransferase; LPAT, lysophosphatidic acid acyltransferase; PP, phosphatidate phosphatase; PDCT, phospholipid:diacylglycerol cholinephospho transferase; PDAT, phospholipid:diacylglycerol acyltransferase; DGAT, diacylglycerol acyltransferase. Abbreviations of intermediates: ACP, acyl carrier protein; G3P, glycerol-3-phosphate; LPA, lysophosphatidic acid; PA, phosphatidic acid; DAG, diacylglycerol; PC, phosphatidylcholine; TAG, triacylglycerol.

higher plants, fruits are used to attract birds and animals for seed transportation, hence aiding seed dispersal. High-lipid fruits can provide more energy to the animal and are therefore favoured by smaller-sized birds [46]. Fruits therefore need to accumulate lager amounts of lipids or UFA before maturation. At the beginning of preparation stage 70 DAP, they constantly absorb external nutrients to increase their lipid contents. After a period of rapid growth, the lipid content tends to be stable, and the fruit begins to turn red. When the fruit turns completely red, the lipid content will increase again and reach the maximum at the end of the mature stage at 130 DAP, without any further changes. The bright colour of the ripened oil grape attracts the animals, and the wax in the seed coat protects the seeds during the passage through the intestinal tract. We also observed a significant change of the FA relative ratio in oil grape during all growth periods. These findings are similar to those for other woody oil plants, such as oil palm, which show a dramatic change in FA composition during development [47].

4.2. Candidate genes related to TGA biosynthesis

Through transcriptome *de novo* assembly and DEGs analysis of the KEGG pathway, we successfully screened out candidate genes related to PUFA synthesis. In the TAGs assembly, *AhDGAT2* from peanut significantly increased the total FA content, especially that of LA and ALA. In *Ricinus communis*, the *RcDGAT2* expressed in *Arabidopsis and* tobacco seeds promoted hydroxy FA accumulation [42]. The *AtDGAT2* preferentially incorporated PA, and *JcDGAT2* exhibited a significant LA preference in yeast expression studies (Zhou et al., 2013). In our previous study, heterologous expression of *IpDGAT2* resulted in an increased LA accumulation in transformed *A. thaliana* seeds [5]. According to these results, *IpDGAT2* might be highly important during TAG biosynthesis.

4.3. Candidate genes related to UFA biosynthesis

Fatty acid desaturation occurs in two steps in the plastid and the ER. The first step is the biosynthesis of MUFA from SFA in the plastid. The *KASII* is a key enzyme in the 16:0-ACP to 18:0-ACP acyl protein synthesis, which is similar to our results [17]. Over-expression of *AtKASII* could decrease the content of PA and increase that of SA [48]. The *FATA* and *FATB* release FAs from the ACP molecule to transport them to the ER. In plants, *FATA* encodes C18 FAs and *FATB* encodes C16 FAs. The specific expression of *AtFATB* in *A. thaliana* seeds leads to the accumulation of a large amount of C16 FAs (Dormann et al., 2000). Transferring the *GmFATA* of *Garcinia mangostana* into rape increased the C18:0 content to 22 % (Facciotti et al., 1999). The 18:0-ACP FA may be desaturated by *SAD* to produce OA. In another study, the *ZmSAD* of *Zea mays* was overexpressed in *A. thaliana* and *Z. mays*, and the contents of stearic acid were significantly lower than those of the control group (Du et al., 2016). In this study, the *IpSAD* was overexpressed in *A. thaliana*, and the contents of OA and eicosenoic acid were significantly higher in the transgenic group. The OA is a precursor to the formation of LA, although the amount of OA in oil grape is less than 10 %. It is also an important bifurcation point of FA synthesis; its content directly determines the total amount of UFA in vegetable oil and the relationship between SFA and UFA (Craig et al., 2008; Kachroo et al., 2007).

The second step is the biosynthesis of specific unsaturated bonds on the MUFAs, which is catalysed by enzymes located on the ER membrane (*FAD2* and *FAD3*) and the chloroplast membrane (*FAD6, FAD7* and *FAD8*) (Mikkilineni and Rocheford, 2003); *FAD2* and *FAD6* desaturate OA to form LA, and *FAD3*, *FAD7* and *FAD8* desaturate LA to form ALA. Overexpression of the *P. frutescens FAD2* gene in *Saccharomyces cerevisia* induces the OA to generate LA [19]. Similar functional verifications have been reported in cotton, soybean and olive ([40]; Li et al., 2007; [49]). In addition, in some plants, such as *Elaeis guineensis* (Sun et al., 2016), *Perilla Perilla* (Lee et al., 2016) and *Flax mustard* (Rodriguez et al., 2016), *FAD2* not only induces OA to generate LA but also can induces POA to generate hexadecadienoic acid (C16:2), this phenomenon may be related to different plant *FAD2* substrate biases (Guo et al., 2013). The function of *IpFAD2* in *I. polycarpa* is that the isolated IpFAD2 proteins could catalyse linoleic synthesis in transformed *A. thaliana* seeds [9]. In a previous study, overexpression of the tomato *LeFAD3* ALA content was significantly increased in both leaves and fruits (Dominguez et al., 2010). In addition, the *FAD3* gene of *Sapium sebiferum* overexpressed in *S. cerevisia* can produce ALA (Niu et al., 2008). In our study, the *IpFAD3* was overexpressed in *A. thaliana*, the contents of ALA were significantly higher, and those of LA were significantly higher in the transgenic group.

Overall, the unigenes *IpGPAT*, *IpDGAT2*, *IpKAS II*, *IpSAD*, *IpFAD2*, *IpFAD3*, *IpFAD7* and *IpFAD8* showed different expression levels between 70 DAP and 130 DAP, making them candidate genes for their differences in UFA compositions in oil grape. The most notable candidate gene was *IpFAD2*, which showed a high expression level in all periods. The function of *IpFAD2* in *I. polycarpa* is that the isolated IpFAD2 proteins could catalyse linoleic synthesis in transformed *A. thaliana* seeds [9], which is the main component of oil grape. In addition, *IpDGAT2*, *IpFAD3* and *IpSAD* also have a significant influence on PUFA biosynthesis, determining PUFA yield and composition.

5. Conclusions

The UFA of oil grape has high nutritional and industrial value, but our understanding of UFA accumulation and biosynthesis is still poor. In this study, the accumulation and biosynthesis of UFA in oil grape development periods was systematically researched. Two growth period transcriptomes of oil grape were sequenced and annotated, and seven key candidate genes, namely *IpGPAD*, *IpDGAT2*, *IpKAS II, IpSAD*, *IpFAD2*, *IpFAD3* and *IpFAD8*, were verified by qRT-PCR and identified by RACE. Therefore, we could infer that the high UFA content in oil grape can put down to the abundance and activity levels of *IpDGAT2*, *IpSAD*, *IpFAD2* and *IpFAD3*. In particular, there is no report of *I. polycarpa* transformation so far, which means that characterising the gene function in *I. polycarpa* is difficult. Further genetic research, such as silencing, heterologous expression or overexpression, will help to understand the biosynthetic pathway of

UFA. This study lays the foundation for the molecular genetic breeding of *I. polycarpa*.

Funding

This work was supported by Science and technology planning project of Shaanxi province(2020NY-075).

Data availability statement

Raw data in this study is available and has been published on NCBI SRA, at: https://www.ncbi.nlm.nih. gov/sra/PRJNA694636.

CRediT authorship contribution statement

Ruishen Fan: Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Boheng Wang:** Formal analysis, Data curation. **Hang Yu:** Resources, Project administration. **Yiran Wang:** Formal analysis, Data curation. **Yanpeng Kui:** Data curation, Conceptualization. **Minmin Chen:** Methodology, Investigation, Data curation. **Yibin Wang:** Software, Data curation, Conceptualization. **Xiaoming Jia:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Funding acquisition.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgment

We thank Mrs. Xiuping Yang for their her in GC analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38015.

References

[1] D. Eriksson, A. Merker, Cloning and functional characterization of genes involved in fatty acid biosynthesis in the novel oilseed crop Lepidium campestre L, Plant Breed. 130 (3) (2011) 407–409, https://doi.org/10.1111/j.1439-0523.2010.01825.x.
[2] J.A.M. Lummiss, K.C. Oliveira, A.M.T. Pranckevicius, A.G. Santos, E.N. dos Santos, D.E. Fogg, Chemical plants: high-value molecules from essential oils, J. Am. Chem. Soc. 134 (46) (2012) 18889–18891, https://doi.org/10.1021/ja310054d.
[3] S.H. Kim, S.H. Sung, S.Y. Choi, Y.K. Chung, J. Kim, Y.C. Kim, Idesolide: a new spiro compound from *Idesia polycarpa*, Org. Lett. 7 (15) (2005) 3275–3277, https://doi.org/10.1021/ol051105f.
[4] S. Wang, Y. Li, Z. Li, L. Chang, L. Li, Identification of an SCAR marker related to female phenotype in *Idesia polycarpa* Maxim, Genet. Mol. Res. 14 (1) (2015) 2015–2022, https://doi.org/10.4238/2015.March.20.11.
[5] R.S. Fan, G. Cai, X.Y. Zhou, Y.X. Qiao, J.B. Wang, H.M. Zhong, J.X. Bo, F. Miao, W. Tu, F.Y. Long, Z.Q. Li, Characterization of diacylglycerol acyltransferase 2 from *Idesia polycarpa* and function analysis, Chem. Phys. Lipids 234 (2021) 105023, https://doi.org/10.1016/j.chemphyslip.2020.105023.
[6] F. Yang, Y. Su, X. Li, Q. Zhang, R. Sun, Preparation of biodiesel from *Idesia polycarpa var. vestita* fruit oil, Ind. Crop. Prod. 29 (2–3) (2009) 622–628, https://doi.org/10.1016/j.indcrop.2008.12.004.
[7] D. Jin, H. Huang, R. Tang, The investigation on the variety resources of Sapium sebiferum in China, Guihaia 17 (1997) 345–362 (in Chinese).
[8] R. Li, X. Gao, L. Li, X. Liu, Z. Wang, S. Lu, De novo assembly and characterization of the fruit transcriptome of *Idesia polycarpa* teveals candidate genes for lipid biosynthesis, Front. Plant Sci. 7 (2016), https://doi.org/10.3389/fpls.2016.00801.
[9] R.S. Fan, L. Li, G. Cai, J. Ye, M.H. Liu, S.H. Wang, Z.Q. Li, Molecular cloning and function analysis of *FAD2* gene in *Idesia polycarpa*, Phytochemistry 168 (2019) 112114, https://doi.org/10.1016/j.phytochem.2019.112114.
[10] D.L. Nelson, M.M. Cox, Lehninger Principles of Biochemistry, Worth Publishing, New York, 2004, pp. 69–70.
[11] J. Ohlrogge, J. Browse, Lipid biosynthesis, Plant Cell 7 (7) (1995) 957–970, https://doi.org/10.1105/tpc.7.7.957.
[12] L.M. Hill, E.R. Morley-Smith, S. Rawsthorne, Metabolism of sugars in the endosperm of developing seeds of oilseed rape, Plant Physiol. 131 (1) (2003) 228–236, https://doi.org/10.1104/pp.010868.
[13] B.J. Nikolau, J.B. Ohlrogge, E.S. Wurtele, Plant biotin-containing carboxylases, Arch. Biochem. Biophys. 414 (2) (2003) 211–222, https://doi.org/10.1016/s0003-9861(03)00156-5.
[14] E.P. Kennedy, Metabolism of lipides, Annu. Rev. Biochem. 26 (1957) 119–148, https://doi.org/10.1146/annurev.bi.26.070157.001003.
[15] K. Stobart, M. Mancha, M. Lenman, A. Dahlqvist, S. Stymne, Triacylglycerols are synthesised and utilized by transacylation reactions in microsomal preparations of developing safflower (*Carthamus tinctorius L*) seeds, Planta 203 (1) (1997) 58–66, https://doi.org/10.1007/s004250050165.
[16] Z. Yang, C. Li, Q. Jia, C. Zhao, D.C. Taylor, D. Li, M. Zhang, Transcriptome analysis reveals candidate genes for petroselinic acid biosynthesis in fruits of *Coriandrum sativum L*, J. Agric. Food Chem. 68 (19) (2020) 5507–5520, https://doi.org/10.1021/acs.jafc.0c01487.
[17] H. Hakozaki, J. Park, M. Endo, Y. Takada, T. Kazama, Y. Takeda, G. Suzuki, M. Kawagishi-Kobayashi, M. Watanabe, Expression and developmental function of the *3-ketoacyl-ACP synthase2* gene in *Arabidopsis thaliana*, Genes Genet. Syst. 83 (2) (2008) 143–152, https://doi.org/10.1266/ggs.83.143.
[18] M. Cai, W. Li, J. Wang, Cloning and expression characteristics of *GhSAD2* in Gossypium hirsutum, Acta Bot. Boreali Occident. Sin. 36 (9) (2017) 1713–1720, https://doi.org/10.11882/j.issn.0254-5071.2015.07.030.

[19] Y. Xue, N. Yin, B. Chen, F. Liao, A. Win, J. Jiang, R. Wang, X. Jin, N. Lin, Y. Chai, Molecular cloning and expression analysis of two *FAD2* genes from chia (*Salvia hispanica*), Acta Physiol. Plant. 39 (4) (2017), https://doi.org/10.1007/s11738-017-2390-0.

[20] T. Nishiuchi, M. Nishimura, V. Arondel, K. Iba, Geno nucleotide-sequence of a gene encoding a microsomal omega-3-fatty-acid desaturase from *Arabidopsis thaliana*, Plant Physiol. 105 (2) (1994) 767–768, https://doi.org/10.1104/pp.105.2.767.

[21] E.P. Heppard, A.J. Kinney, K.L. Stecca, G. Miao, Developmental and growth temperature regulation of two different microsomal omega-6 desaturase genes in soybeans, Plant Physiol. 110 (1) (1996) 311–319, https://doi.org/10.1104/pp.110.1.311.

[22] A.A. Wendel, T.M. Lewin, R.A. Coleman, Glycerol-3-phosphate acyltransferases: rate limiting enzymes of triacylglycerol biosynthesis, Biochim. Biophys. Acta-Mol. Cell Biol. Lipids 1791 (6) (2009) 501–506, https://doi.org/10.1016/j.bbalip.2008.10.010.

[23] E.P. Kennedy, Biosynthesis of complex lipids, Fed. Proc. 20 (1961) 934–940.

[24] S. Lung, R.J. Weselake, Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis, Lipids 41 (12) (2006) 1073–1088, https://doi.org/10.1007/s11745-006-5057-y.

[25] A.C. Turchetto-Zolet, F.S. Maraschin, G.L. de Morais, A. Cagliari, C.M.B. Andrade, M. Margis-Pinheiro, R. Margis, Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT): a key enzyme in neutral lipid biosynthesis, BMC Evol. Biol. 11 (2011), https://doi.org/10.1186/1471-2148-11-263.

[26] C.L. Yen, S.J. Stone, S. Koliwad, C. Harris, R.V.Jr Farese, DGAT enzymes and triacylglycerol biosynthesis, JLR (J. Lipid Res.) 49 (11) (2008) 2283–2301, https://doi.org/10.1194/jlr.R800018-JLR200.

[27] K.D. Lardizabal, J.T. Mai, N.W. Wagner, A. Wyrick, T. Voelker, D.J. Hawkins, *DGAT2* is a new diacylglycerol acyltransferase gene family - purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity, J. Biol. Chem. 276 (42) (2001) 38862–38869, https://doi.org/10.1074/jbc.M106168200.

[28] R. Li, K. Yu, T. Hatanaka, D.F. Hildebrand, Vernonia *DGATs* increase accumulation of epoxy fatty acids in oil, Plant Biotechnol. J. 8 (2) (2010) 184–195, https://doi.org/10.1111/j.1467-7652.2009.00476.x.

[29] S. Cases, S.J. Stone, P. Zhou, E. Yen, B. Tow, K.D. Lardizabal, T. Voelker, R.V. Farese, Cloning of *DGAT2*, a second mammalian diacylglycerol acyltransferase, and related family members, J. Biol. Chem. 276 (42) (2001) 38870–38876, https://doi.org/10.1074/jbc.M106219200.

[30] A. Turkish, S.L. Sturley, Regulation of triglyceride metabolism. I. Eukaryotic neutral lipid synthesis: "many ways to skin ACAT or a DGAT", Am. J. Physiol. Gastrointest. Liver Physiol. 292 (4) (2007) G953–G957, https://doi.org/10.1152/ajpgi.00509.2006.

[31] K.D. Lardizabal, K.A. Bennett, N.W. Wagner, Diacylglycerol Acyltransferase Nucleic Acid Sequences and Associated Products, Monsanto Technology LLC, 2008.

[32] K. Lardizabal, R. Effertz, C. Levering, J. Mai, M.C. Pedroso, T. Jury, E. Aasen, K. Gruys, K. Bennett, Expression of Umbelopsis ramanniana *DGAT2A* in seed increases oil in soybean, Plant Physiol. 148 (1) (2008) 89–96, https://doi.org/10.1104/pp.108.123042.

[33] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns (vol 95, pg 14863, 1998), Proc. Natl. Acad. Sci. U. S. A. 96 (19) (1999), 10943-10943.

[34] Y. Zuo, H.B. Liu, B. Li, H. Zhao, X.L. Liu, J.T. Chen, L. Wang, Q.B. Zheng, Y.Q. He, J.S. Zhang, M.X. Wang, C.Z. Liang, L. Wang, The *Idesia polycarpa* genome provides insights into its evolution and oil biosynthesis, Cell Rep. 43 (2024) 113909, https://doi.org/10.1016/j.celrep.2024.113909.

[35] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, https://doi.org/10.1016/s0022-2836(05)80360-2.

[36] R. Li, K. Yu, D.F. Hildebrand, *DGAT1, DGAT2* and *PDAT* expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants, Lipids 45 (2) (2010) 145–157, https://doi.org/10.1007/s11745-010-3385-4.

[37] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010), https://doi.org/10.1186/gb-2010-11-10-r106.

[38] N. Korotkova, J.V. Schneider, D. Quandt, A. Worberg, G. Zizka, T. Borsch, Phylogeny of the eudicot order Malpighiales: analysis of a recalcitrant clade with sequences of the petD group II intron, Plant Systemat. Evol. 282 (3–4) (2009) 201–228, https://doi.org/10.1007/s00606-008-0099-7.

[39] Y. Li-Beisson, B. Shorrosh, F. Beisson, M.X. Andersson, V. Arondel, P.D. Bates, S. Baud, D. Bird, A. Debono, T.P. Durrett, R.B. Franke, I.A. Graham, K. Katayama, A.A. Kelly, T. Larson, J.E. Markham, M. Miquel, I. Molina, I. Nishida, O. Rowland, L. Samuels, K.M. Schmid, H. Wada, R. Welti, C. Xu, R. Zallot, J. Ohlrogge, Acyl-lipid metabolism, Arabidopsis Book 8 (2010), https://doi.org/10.1199/tab.0133 e0133-e0133.

[40] M. Zhang, J. Fan, D.C. Taylor, J.B. Ohlrogge, DGAT1 and PDAT1 Acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development, Plant Cell 21 (12) (2009) 3885–3901, https://doi.org/10.1105/tpc.109.071795.

[41] C. Jako, A. Kumar, Y.D. Wei, J.T. Zou, D.L. Barton, E.M. Giblin, P.S. Covello, D.C. Taylor, Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight, Plant Physiol. 126 (2) (2001) 861–874, https://doi.org/10.1104/pp.126.2.861.

[42] J. Burgal, J. Shockey, C. Lu, J. Dyer, T. Larson, I. Graham, J. Browse, Metabolic engineering of hydroxy fatty acid production in plants: *RcDGAT2* drives dramatic increases in ricinoleate levels in seed oil, Plant Biotechnol. J. 6 (8) (2008) 819–831, https://doi.org/10.1111/j.1467-7652.2008.00361.x.

[43] C. Zhang, U. Iskandarov, E.T. Klotz, R.L. Stevens, R.E. Cahoon, T.J. Nazarenus, S.L. Pereira, E.B. Cahoon, A thraustochytrid diacylglycerol acyltransferase 2 with broad substrate specificity strongly increases oleic acid content in engineered *Arabidopsis thaliana* seeds, J. Exp. Bot. 64 (11) (2013) 3189–3200, https://doi.org/10.1093/jxb/ert156.

[44] M. Luisa-Hernandez, L. Whitehead, Z. He, V. Gazda, A. Gilday, E. Kozhevnikova, F.E. Vaistij, T.R. Larson, I.A. Graham, A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued arabidopsis seed oil catabolism mutants, Plant Physiol. 160 (1) (2012) 215–225, https://doi.org/10.1104/pp.112.201541.

[45] E. Gazave, E.E. Tassone, M. Baseggio, M. Cryder, K. Byriel, E. Oblath, S. Lueschow, D. Poss, C. Hardy, M. Wingerson, J.B. Davis, H. Abdel-Haleem, D.M. Grant, J. L. Hatfield, T.A. Isbell, M.F. Vigil, J.M. Dyer, M.A. Jenks, J. Brown, M.A. Gore, D. Pauli, Genome-wide association study identifies acyl-lipid metabolism candidate genes involved in the genetic control of natural variation for seed fatty acid traits in *Brassica napus L*, Ind. Crop. Prod. 145 (2020), https://doi.org/10.1016/j.indcrop.2019.112080.

[46] J.P. Simpson, J.B. Ohlrogge, 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 28 (1) (2016) 248–264, https://doi.org/10.1105/tpc.15.00900.

[47] S. Dussert, C. Guerin, M. Andersson, T. Joet, T.J. Tranbarger, M. Pizot, G. Sarah, A. Omore, T. Durand-Gasselin, F. Morcillo, Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition, Plant Physiol. 162 (3) (2013) 1337–1358, https://doi.org/10.1104/pp.113.220525.

[48] J. Beld, D.J. Lee, M.D. Burkart, Fatty acid biosynthesis revisite: structure elucidaation and metabolic engineering, Mol. BioSyst. 11 (2015) 38–59.

[49] M.L. Hernandez, M. Mancha, J.M. Martinez-Rivas, Molecular cloning and characterization of genes encoding two microsomal oleate desaturases (FAD2) from olive, Phytochemistry 66 (12) (2005) 1417–1426, https://doi.org/10.1016/j.phytochem.2005.04.004.