Industrial Crops & Products 145 (2020) 112051

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Comparative transcriptome analysis of the genes involved in lipid biosynthesis pathway and regulation of oil body formation in *Torreya grandis* kernels

Mingzhu Ding[a], Heqiang Lou[a,*], Wenchao Chen[a], Ying Zhou[a], Zihao Zhang[a], Minghui Xiao[a], Zhanqi Wang[b], Yi Yang[a], Lei Yang[a], Feicui Zhang[a,c], Jiasheng Wu[a,c,*], Lili Song[a,c,*]

[a] State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China
[b] Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, College of Life Sciences, Huzhou University, Huzhou, 313000, China
[c] NFGA Engineering Research Center for Torreya grandis 'Merrillii', Zhejiang A&F University, Hangzhou, 311300, China

ARTICLE INFO

Keywords:
Torreya grandis
Lipid
Oil biosynthesis
Oil body-associated proteins
Transcriptome

ABSTRACT

Torreya grandis (*T. grandis*) is an important economic tree species because of the high content of unsaturated fatty acids in its kernel oil. However, different *T. grandis* landraces vary significantly in their fatty acid and oil contents, and their molecular regulatory mechanisms remain unclear. To investigate the molecular basis of the fatty acid biosynthesis pathway and oil body formation in *T. grandis* kernels, transcriptome sequencing were performed based on the Illumina platform in ten different *T. grandis* landraces. In total, 112,699 unigenes were identified, among them, 175 unigenes related to lipid biosynthesis were found, including 126 unigenes for fatty acid biosynthesis, 37 unigenes for triacylglycerol assembly and 12 unigenes for oil body proteins. The correlation analysis between the oil content and the expression levels of unigenes suggested that biotin carboxylase, acyl-ACP thioesterase A and lysophosphatidic acid acyltransferase may play key roles in oil accumulation. Three candidate genes (*TgOLEO1*, *TgCLO1* and *TgSLO1*) encoding oil body-associated proteins were further identified through phylogenetic analysis and amino acid sequence alignment. Further subcellular localization assay indicated that all of the proteins were localized in both of endoplasmic reticulum and oil body. Compared with wild type, oil body proteins transiently over-expressed tobacco leaves had a higher abundance of oil bodies. Furthermore, the candidate regulatory genes (including *TgWRI1, TgFUS3,* etc.) that may be involved in the regulation of lipid biosynthesis were also identified. This study will be critical for the molecular assisted screening and breeding of high content of oil and unsaturated fatty acids cultivars for *T. grandis*.

1. Introduction

Vegetable oils are synthesized during seed development in oilseed plants and are stored in the form of triacylglycerol (TAG) (Baud et al., 2008). TAGs are involved in diverse physiological processes and multiple metabolic reactions in plants (Xu and Shanklin, 2016). For example, they afford the major energy for seedling development after germination, produce carbohydrates during seed germination and the early development of seedling and are critical for normal growth in plants (Penfield et al., 2004; Theodoulou and Eastmond, 2012; Zhang et al., 2009). Vegetable oils, as an important source of edible oils and industrial materials, play vital roles in human life.

In higher plants, TAGs are synthesized from glycerol-3-phosphate (G3P) and fatty acids (FAs). The synthesis of FA first occurs in plastid by a set of reactions, including condensation, reduction and dehydration with the participation of FA synthases (Xu and Shanklin, 2016). The FA synthases including acetyl-CoA carboxylase (ACCase), 3-ketoacyl-ACP synthase (KAS), 3-ketoacyl-ACP reductase (KAR), 3-hydroxyacyl-ACP dehydratase (HAD) and enoyl-ACP reductase (ENR), acyl-ACP thioesterases (FATA and FATB), and acyl-CoA sythetase (LACS). After that,

Abbreviations: DAG, diacylglycerol; TAGs, triacylglycerols; ACCase, acetyl-CoA carboxylase; ACP, acyl-carrier protein; KAS, 3-ketoacyl-ACP synthase; KAR, 3-ketoacyl-ACP reductase; HAD, 3-hydroxyacyl-ACP dehydratase; ENR, enoyl-ACP reductase; FAT, acyl-ACP thioesterases; LACS, acyl-CoA sythetase; G3P, glycerol-3-phosphate; GPAT, glycerol-3-phosphate acyltransferase; LPAT, lysophosphatidic acid acyltransferase; PAP, phosphatidic acid phosphatase; DGAT, diacylglycerol acyltransferase; PDAT, diacylglycerol acyltransferase; FAME, fatty acid methyl esters; FPKM, the fragments per kilobase per million fragments; DAF, days after flowering; LEC1, LEAFY COTYLEDON1; WRI1, WRINKLED1; FUS3, FUSCA3

* Corresponding authors at: State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China.
E-mail addresses: 20170030@zafu.edu.cn (H. Lou), wujs@zafu.edu.cn (J. Wu), lilisong@zafu.edu.cn (L. Song).

https://doi.org/10.1016/j.indcrop.2019.112051
Received 13 August 2019; Received in revised form 6 October 2019; Accepted 15 December 2019
Available online 30 December 2019

the nascent FAs were transferred to the endoplasmic reticulum (ER) from plastid for assembling TAG with glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAT), diacylglycerol acyltransferase (DGAT), and the phospholipid: diacylglycerol acyltransferase (PDAT) (Bates et al., 2013; Furmanek et al., 2014). During the final progress, some major oil body-associated proteins were bond to TAG to form oil bodies.

In mature seeds, TAG can be stored as oil bodies which are bounded by phospholipids implanted with some intrinsic proteins including oleosin, caleosin and steroleosin (Lin and Tzen, 2004). Oleosin is an alkaline protein that specifically bind to oil bodies, and is found ubiquitously in seeds (Siloto et al., 2006; Ischebeck, 2016). Amino acid sequence comparison revealed that the central hydrophobic domains of different oleosins are highly conserved (Pasaribu et al., 2017), particularly in a distinct 12-amino acid hydrophilic region named the Proline knot. The facilitation of targeting oleosins to oil bodies is drived by the Proline knot (Chen and Tzen, 2001). Studies showed that oleosins are vital for providing stability to discrete oil-bodies during seed desiccation, preventing fusion of oil bodies and maintaining seed germination (Miquel et al., 2014; Shimada and Hara-Nishlmura, 2010). Caleosin is a protein involved in lipid mobilization. Caleosin contains three structural regions: an N-terminal hydrophilic calcium (Ca^{2+})-binding region, a central hydrophilic anchoring motif, and a C-terminal hydrophilic motif contains a few phosphorylation sites (Chen et al., 1999; Næsted et al., 2000). Steroleosin consists of an N-terminal lipid anchoring domain and a dissolvable sterol-binding dehydrogenase motif which is related to signal transduction (Li-Jen et al., 2002). In addition, steroleosin was also demonstrated to possess the functions in the formation or degradation of seed oil bodies (Lin and Tzen, 2004).

Although genes involved in oil accumulation have been widely cloned and well characterized in many species, the operating manners of these genes showed some differences, especially with the four acyltransferases, GPAT, LPAT, DGAT and PDAT, involved in TAG assembly. It has been identified that *Arabidopsis* GPAT9 are responsible for the synthesis of plant membrane lipid and TAG, which located in ER (Shockey et al., 2016). In transgenic *Arabidopsis* and *Brassica napus*, overexpression of LPAT genes could increase oil content (Rao and Hildebrand, 2009). Studies also showed that DGAT plays a dominant role on TAG accumulation in many oilseeds (Meng et al., 2009; Shiu-Cheung and Weselake, 2006). For example, it was shown that a phenylalanine insertion in DGAT is responsible for the increased oil and oleic acid contents in maize (*Zea mays* L.) (Zheng et al., 2008). However, it was found that PDAT shows significant higher activity in synthesizing unsaturated FAs in safflower and linseed though its contribution has not been established (Banaś et al., 2013; Pan et al., 2013). These results reveal that the key genes limiting the oil biosynthesis varied in different species.

Furthermore, several transcription factors (TFs) like *LEAFY COTYLEDON1* (*LEC1*), *WRINKLED1*(*WRI1*), *FUSCA3* (*FUS3*) and *ABSCISIC ACID3* (*ABI3*) are reported to be involved in regulating seed development (Wang et al., 2007). Overexpression of *LEC1* could increase oil content and the expression levels of genes concerned with FA biosynthesis in *Arabidopsis* (Jinye et al., 2008). Overexpression of *LEC1* and a *LEC1* downstream TF, *WRI1*, leading to a significant increase of the oil content in maize seeds (Bo et al., 2010). When up-regulating the expression of *FUS3,* the expression level of genes involved in TAG assembly is higher than that involved in FA biosynthesis (Zhang et al., 2016). These studies illustrated that TFs play very important roles in regulating oil accumulation in plant.

T. grandis, which belongs to the gymnospermous yew family (Taxaceae), is a large, evergreen economic species, normally cultivated in southeast China (Dong et al., 2014). As one of the rare nuts in China, the kernels of *T. grandis* not only contain high nutritional values and special flavour, but also have multiple biological functions including anti-oxidative, antitussive, anti-inflammatory, antifungal, antibacterial and antitumor activities (Chen et al., 2006; Dong et al., 2014). Due to the environmental impact and cultivation management, *T. grandis* has developed into different cultivars and characters, thus, it displays diverse qualities (Wu et al., 2018). *T. grandis* is well known for their exceptionally high oil content (approximately up to 50 % dry mass at the mature stage). However, different landraces vary considerably in their oil content and FA composition and the molecular regulatory mechanisms of the oil biosynthesis are still unclear. Thus, the molecular basis for the regulatory mechanisms of the high levels of oil biosynthesis, including identification of oil biosynthesis pathway and prediction of the key genes in the pathway, needs to be excavated primarily.

Here, a quantitative comparative transcriptome analysis of *T. grandis* kernels from ten different landraces were performed, the genes involved in lipid biosynthesis pathway were identified and the key genes regulating the oil accumulation were detected. In addition, three candidate oil body-associated proteins in TAG assembly were also identified. Furthermore, the subcellular localization of these proteins were carried out to illustrate where they function. The results will provide novel insights into the regulatory mechanisms of high levels of oil accumulation in *T. grandis*, which will be helpful to develop high oil contents genotypes.

2. Materials and methods

2.1. Plant materials

Samples were taken from ten *T. grandis* landraces (named as C06, R08, R20, R24, S18, X05, X08, X18, Z07 and Z11, respectively) in early September 2016, which were distributed in different places (listed in Additional file 1). Kernels were gathered at the mature stage, with other parts of the seeds removed. After that, the kernels were flash-frozen in liquid nitrogen and then stored at −80 °C for further analysis.

2.2. Oil extraction and lipid analysis

The kernels of *T. grandis* were dried to a constant weight at 60 °C. Total oil was extracted from 10 g of dried powder at 45 °C for 12 h with petroleum ether as a solvent in a Soxhlet apparatus. The oil contents were calculated as a percent of dry sample. Fatty acid methyl esters (FAME) were obtained using the method provided by Wu et al. with some modifications and analyzed by using a gas chromatograph (Thermo Scientific TRACE-1300, Italy) (Wu et al., 2018). The relative content of each fatty acid was determined by the retention time of the fatty acid methyl ester gas phase standard (Sigma). For each sample, the oil content and lipid analysis were performed with three biological replicates under a completely random experimental design.

2.3. RNA extraction, cDNA library construction and Illumina RNA sequencing

Total RNA was extracted from the ten *T. grandis* kernels by the RNA extraction Kit (TIANGEN, DP441) with an additional DNase I (TIANGEN) to digest any genomic DNA contamination. The purified RNA concentration was examined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). The NEBNext®Ultra™ RNA Library Prep Kit was used to generate cDNA libraries. The library quality was evaluated by the Agilent Bioanalyzer 2100 system. Paired-end sequencing was carried out for each library using a HiSeq PE Cluster Kit v4 cBot (Illumina, San Diego, CA, USA). Sequencing was conducted by an Illumina HiSeq™ 4000 paired-end sequencing system.

2.4. Sequence data assembly and annotation

To obtain high quality data, in-house perl scripts was applied to process the Raw data in FASTQ format. In this step, adapters of sequences were cut and the low quality reads with ≥5 uncertain bases or

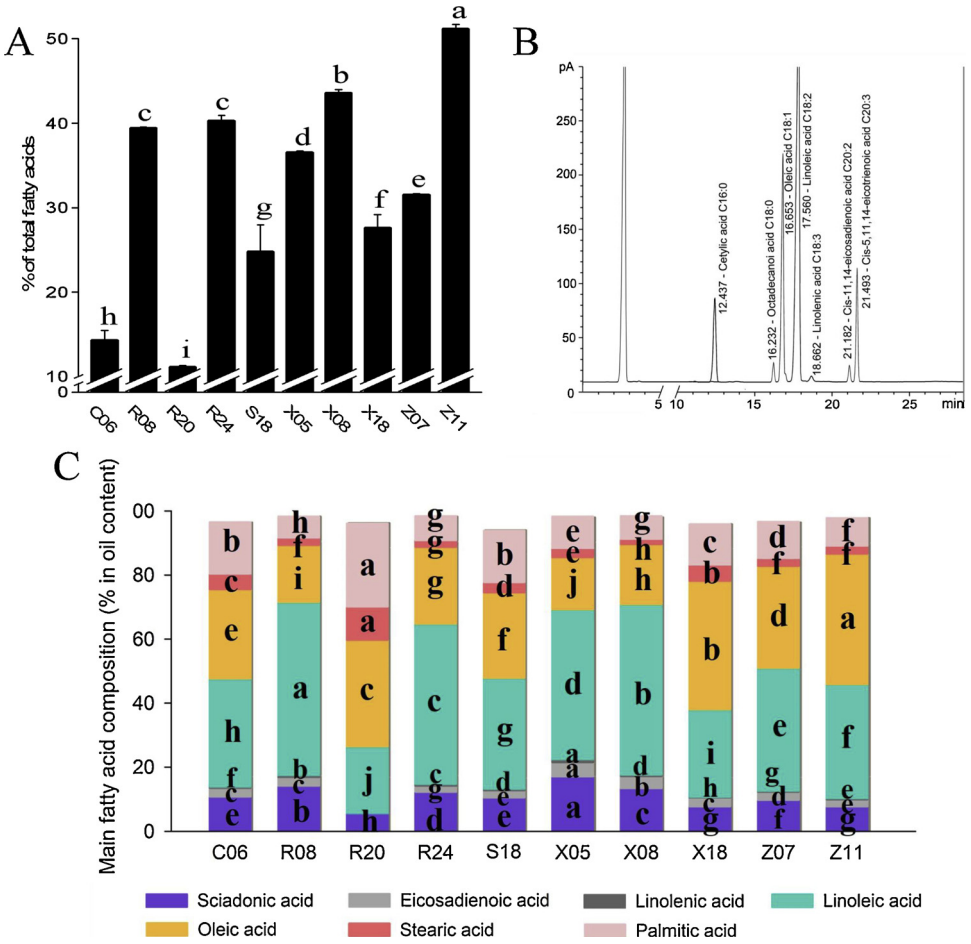


Fig. 1. The oil content and fatty acid composition of different *T. grandis*, Bars topped by the same letter were not significantly different according to one-way ANOVA (P value). Significant differences were determined using Duncan's new multiple range test at $P = 0.05$. (A) The oil content of kernels in different *T. grandis*. (B) Gas chromatograms of main fatty acids standards. (C) Main fatty acid compositions of different *T. grandis*.

with over 50 % of Qphred < = 20 bases were removed from the raw data. Meanwhile, the clean data containing Q20, Q30, GC-content, and sequence duplication levels were calculated, providing a basis for all the downstream analyses. All clean reads from ten *T. grandis* landraces were put together for transcripts assembly using Trinity 2.0 (Grabherr et al., 2011). Functional annotation of genes was based on the transcripts with the comparison against the following public protein databases: NCBI non-redundant protein (Nr) database; NCBI non-redundant nucleotide sequences (Nt); SwissProt database; Clusters of Orthologous Groups of proteins (COG), the Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO).

2.5. RNA-Seq data analysis

The transcript abundances of each gene were calculated by the fragments per kilobase per million fragments (FPKM) method (Mortazavi et al., 2008). The significance of gene expression difference was judged with '$|\log_2$FPKM ratio$| \geq 1$ and FDR (false discovery rate) ≤ 0.001' as the threshold (Benjamini et al., 2001). GO (Gene Ontology) enrichment analysis of the unigenes was performed using GOseq R packages based on the Wallenius noncentral hyper-geometric distribution (Young et al., 2010). GO was used to classify unigenes based on molecular function, biological processes and cellular components for kernels (Conesa and Gotz, 2008).

2.6. qRT-PCR analysis

qRT-PCR was performed to identity the expression patterns of candidate genes in the kernels of ten different *T. grandis* landraces or different development stages of X08 landrace. cDNA was synthesized from total RNA using the PrimeScript™RT Master Mix (Takara). Diluted cDNA was amplified with specific primers (listed in Additional file 2) and reacted with ChamQ SYBR qPCR Master Mix (Vazyme) on a C1000 Touch™ Thermal Cycler (Bio-Rad) by the following condition: 45 cycles of 95 ˚C for 10 s, 57 ˚C for 10 s and 72 ˚C for 20 s. Each sample was normalized using *Actin* as internal control, and the formula $2^{-\triangle\triangle Cp}$ was uesd to calculate the results. Three biological replicates were performed.

2.7. Microscopy

After cloning the full-length cDNA (without stop codon) from the kernel of X08 landrace, the binary vector *35S ::GFP* (modified from pCAMBIA1300) was applied for recombination. The recombinant plasmids were subsequently transformed into *Agrobacterium tumefaciens* strain GV3101 and transiently expressed in tobacco *(N. benthamiana)* leaves. Oil bodies were stained with 2 μg/mL Nile Red (Sigma-Aldrich) in 50 mM PIPES buffer (pH 7.0) for 20 min. After then, the samples were washed three times (10 min each time) using 50 mM PIPES buffer immediately (Gidda et al., 2016). Fluorescence was observed via confocal laser scanning microscopy (LSM510: Karl Zeiss).

2.8. Statistical and sequence analyses

The data of the experiment were collected and analyzed using the statistical analysis systems software (SPSS, version 21.0). Pearson's correlation coefficient (r) was applied to calculate the data correlations, and significant differences were applied using Duncan's new multiple range test at $P = 0.05$. For phylogenetic analysis, the unrooted phylogenetic tree was constructed using MEGA 7.0 with the neighbor joining method and bootstrap values from 1000 replicates were indicated at each branch (Tamura et al., 2011). Amino acid sequence comparison was analyzed by DNAMAN 8.0.

3. Results

3.1. The oil content and fatty acid composition of mature kernels in different T. grandis landraces

The oil content and fatty acid composition were examined in *T. grandis* kernels from ten landraces. Comparing to C06 (14.27 %) and R20 (11.15 %), the oil content are much higher in S18, X18, Z07, R08, R24, X05, X08 and Z11, up to 22.49 %, 26.50 %, 31.42 %, 39.29 %, 40.27 %, 36.54 %, 43.58 % and 51.16 %, respectively (Fig. 1A). This result indicates that the oil content in mature kernels among the different landraces were varied and abundant.

The fatty acid composition was identified by gas chromatography, with the retention time of authentic standards already preserved (Fig. 1B). There were seven dominant fatty acids in *T. grandis*, which were palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1$\triangle^{9}$), linoleic acid (C18:2$\triangle^{9,12}$), linolenic acid (C18:3$\triangle^{9,12,15}$), eicosadienoic acid (C20:2$\triangle^{11,14}$) and sciadonic acid (C20:3$\triangle^{5,}$11,14). The accumulation of these main fatty acids was also various in different landraces. The most abundant fatty acids in these landraces were oleic acid and linoleic acid (Fig. 1C). In contrast with other samples, the content of linoleic acid in R20, X18 and Z11 was less than that of oleic acid. The results indicated that the oleic and linoleic acid content showed a reversible conversion, which appeared to be one decreased while the other increased. Interestingly, the content of linoleic acid and sciadonic acid exhibited similar trends in different landraces (Fig. 1C; Additional file 3), which suggested that linoleic acid and sciadonic acid may be co-regulated in *T. grandis*.

3.2. Transcriptome sequencing, similarity analysis, and functional annotation

Based on RNA-Seq sequencing, the statistical summary of the ten samples were measured. The low-quality reads, adapter reads and sequences containing N ratios higher than 10 % were filtered out firstly. Then, an average number of 62.76 million clean reads were generated. The total number of clean nucleotides sequences per sample was 7.4 G or more. The average GC content were 46.24 % and the Q20 and Q30 values were greater than 96.84 % and 91.95 %, respectively, indicating that our transcriptome data was reliable and of high quality (Additional file 4). Subsequently, 112,699 unigenes were producted from these high-quality reads.

The transcripts were annotated by BLASTing against the Nr, Nt, SwissProt, COG, GO and KEGG databases with an E-value threshold of 10^{-5}. The overall functional annotation was shown in Additional file 5. Results showed that there are 87,172, 11,511, 62,107, 37,477, 53,937, and 44,339 unigenes were annotated by BLASTing against the Nr, Nt, SwissProt, COG, GO, and KEGG databases, respectively. In order to investigate the differentially expressed genes (DEGs) among ten *T. grandis* landraces, expression level of the unigenes was compared with one of the landraces, C06, and the number of up-regulated and down-regulated genes was counted. Finally, over 8030 DEGs were identified with more than 3047 upregulated and more than 3387 downregulated unigenes (Fig. 2). In order to further explore the functions of DEGs, GO enrichment analysis was performed. Among the down-regulated unigenes, 239 genes were categorized to the biological process, 132 and 90 were categorized to the molecular function and the cellular component, respectively (Fig. 3A). Among the up-regulated unigenes, the corresponding genes of the biological process, the molecular function and the cellular component were 47, 61 and 23, respectively (Fig. 3B).

3.3. Identification of genes involved in lipid biosynthesis from different T. grandis

Although previous studies have identified the lipid biosynthesis pathway in other plant species (Li-Beisson et al., 2013), the oil body formation pathway and how the lipid biosynthesis pathway genes involves in regulating oil accumulation in *T. grandis* are still unclear. Moreover, limited number of studies have exemplified how the transcription of genes in the pathway is coordinated during oil accumulation. Therefore, the lipid biosynthesis pathway including oil body formation pathway were identified based on previous studies and the transcriptome data in this study (Fig. 4), and the roles of the pathway genes in oil accumulation in *T. grandis* were further analyzed. In total, 175 unigenes related to lipid biosynthesis, including 126 responsible for FAs biosynthesis, 37 for TAG assembly and 12 for oil bodies, were identified. The FPKM value of each unigene in the ten landraces was shown in the form of heatmap (Fig. 4).

The initial process involved in FAs biosynthesis in plastids is complex. Researchers have reported that genes related to fatty acid synthesis were expressed the highest at 120 days after flowering (DAF) in the embryo of pecan (*Carya illinoinensis*) (Huang et al., 2017), which indicated that the expression levels of genes responsible for the nascent FAs are highly related to the developmental stages of seeds. In this study, although there was no significant correlation between those gene expressions and oil contents in different landraces, some notable clues concerned with the composition of fatty acids were inferred. According to the correlation analysis between the content of the main fatty acids and the FPKMs of candidate genes involved in FA biosynthesis of the ten landraces, 2 unigenes (*BRD_TGR20275, TR56397-c0_g1_i1*) of PDH are significantly and negatively correlated with oleic acid ($r = -0.734$) and eicosadienoic acid ($r = -0.760$), respectively (Additional file 6). The unigenes of biotin carboxylase (BC) which belongs to ACCase showed high relations to several main fatty acids, with 1, 3 and 6 unigenes significantly correlated with palmitic acid, oleic acid and linoleic acid, respectively. Besides, only the unigene, *BRD_TGR42355*, exhibited an extremely significant correlation with linolenic acid ($r = 0.801$), while another unigene, *BRD_TGR23527*, had an independent correlation with the total oil content ($r = 0.636$) (Additional file 6). These results imply that the BC genes play key roles in fatty acid biosynthesis. Interestingly, It was found that 1 unigene (*BRD_TGR51105*) of FATA displayed a remarkably correlation with oleic acid (18:1) ($r = 0.797$) (Additional file 6), consistent with previous result that FATA is broadly accepted to hydrolyze 18:1-ACP (Moreno-Pérez et al., 2012).

In ER, apparently most unigenes of LPAT were highly correlated to unsaturated fatty acids, especially to oleic acid (18:1), eicosadienoic acid (20:2) and sciadonic acid (20:3) (Additional file 6), which means LPAT may be an important enzyme for improving the synthesis of unsaturated fatty acids. To further determine which genes exert the most important role in oil accumulation, correlation analysis between the expression levels of the unigenes of LPAT, GPAT, PDAT, and DGAT and total oil contents in the ten landraces was conducted. The result showed that the unigenes of PDAT, BRD_TGR47946 and BRD_TGR65782, were highly and positively correlated with oil contents (Additional file 6). This result is consistent with previous research in safflower and linseed (Banaś et al., 2013; Pan et al., 2013), implying that PDAT has a great contribution to the oil synthesis in *T. grandis*.

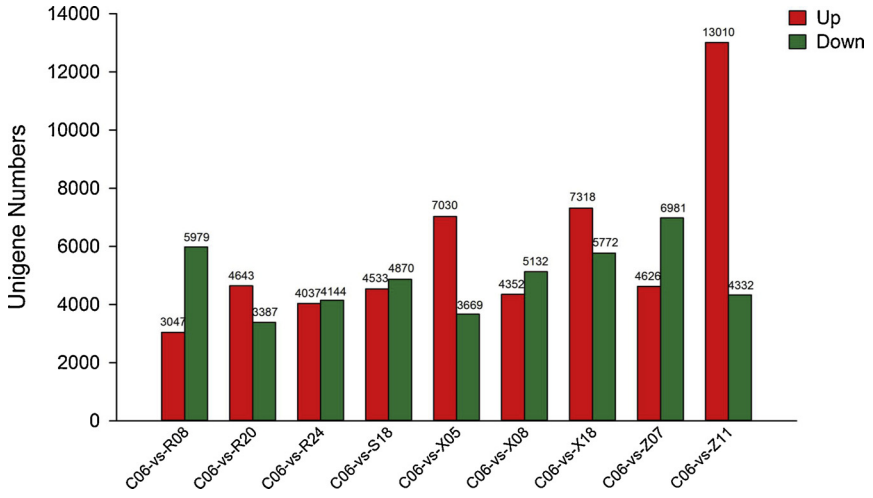


Fig. 2. Analysis of differentially expressed genes (DEGs) in different *T. grandis* cultivars.

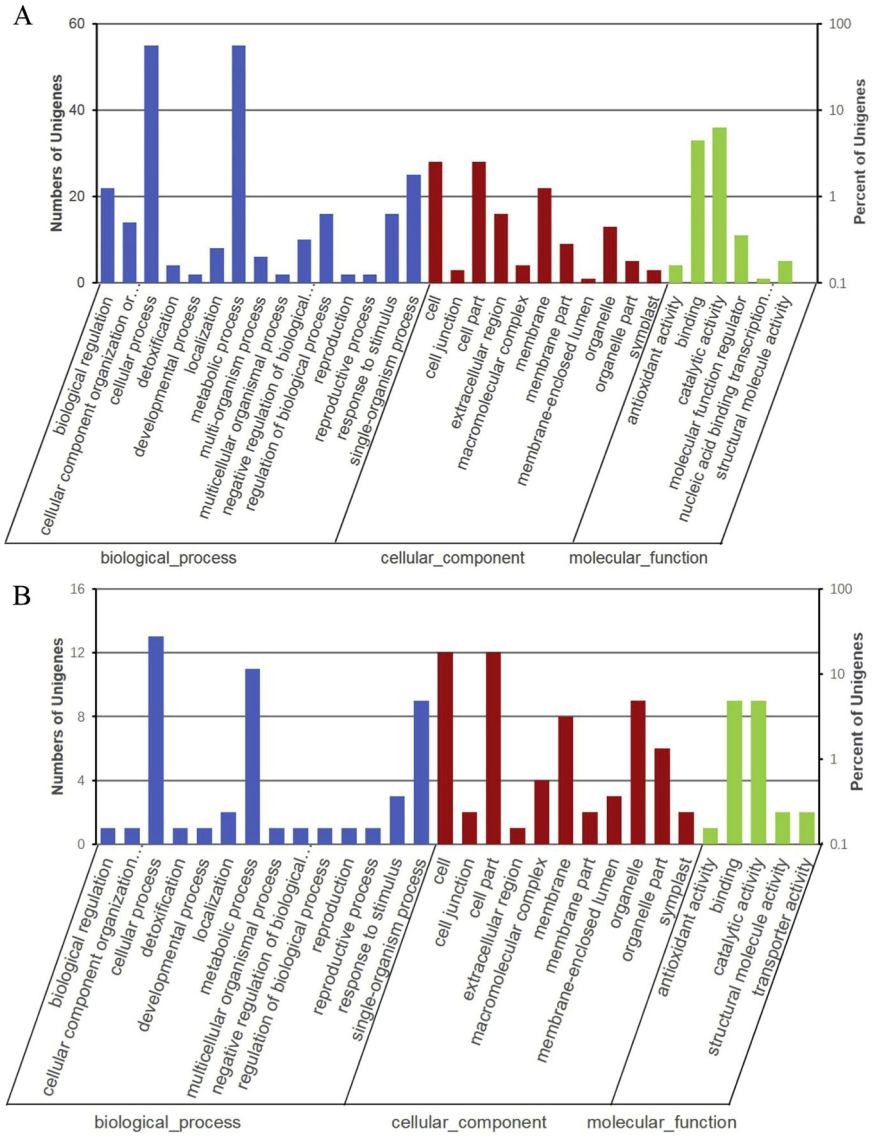


Fig. 3. Classification of full-length transcripts in *T. grandis*: (A) Downregulated Gene Ontology (GO) classification; (B) Upregulated Gene Ontology (GO) classification.

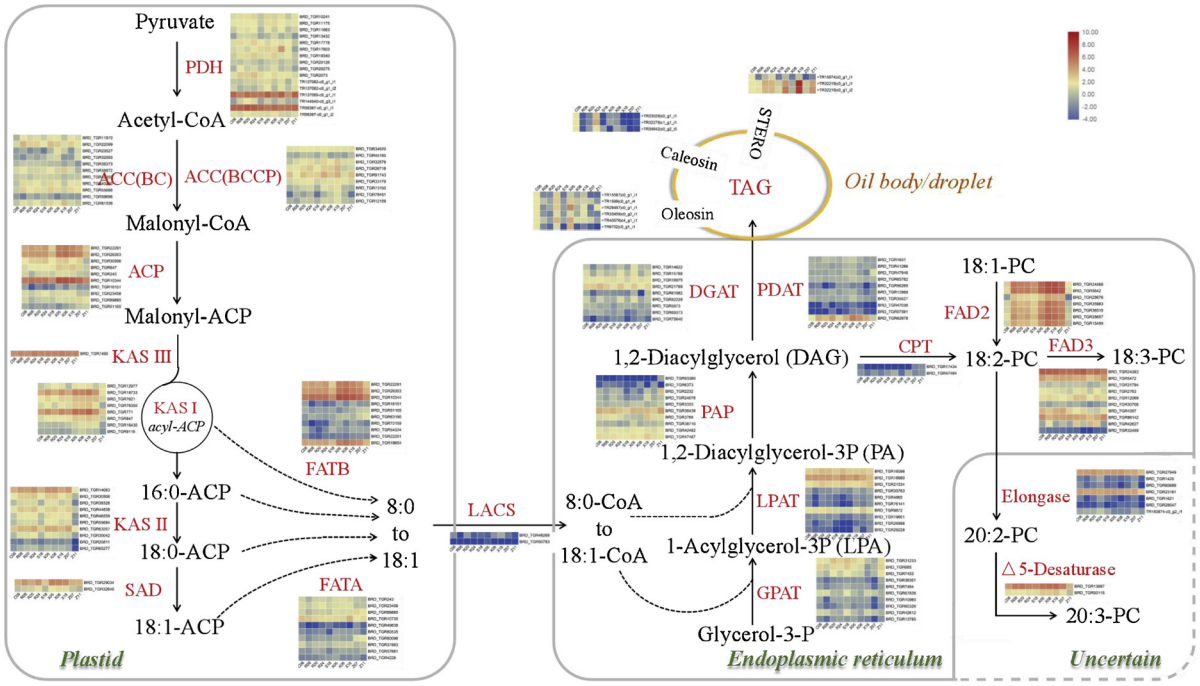


Fig. 4. Transcriptional specialization of lipid-related genes in the mature stage of embryo of *T. grandis*. Enzymes found in this pathway are marked in red words. The ten squares of each enzymes correspond to various landraces. Full names of the unigenes are presented in the right of squares. Abbreviations: PDH, pyruvate dehydrogenase; ACC (BC), Biotin carboxylase subunit of heteromeric acetyl-CoA carboxylase (ACCase); ACC(BCCP), biotin carboxyl carrier protein of heteromeric ACCase; KAS, ketoacyl-ACP synthase; SAD, stearoyl-ACP desaturase; FATA, acyl-ACP thioesterase A; FATB, acyl-ACP thioesterase B; LACS, long-chain acyl-CoA sythetase; GPAT, glycerol-3-phosphate acyltransferase; LPAT, 1ysophosphatidic acid acyltransferase; PAP, phosphatidic acid phosphatase; DGAT, diacylglycerol acyhransferase; PDAT, phospholipid: diacylglycerol acyltransferase; CPT, diacylglycerol cholinephosphotransferase; FAD2, v-6 desaturase; FAD3, v-3 desaturase; TAG, triacylglycerol.

3.4. Characterization of candidate genes encoding proteins related to oil bodies

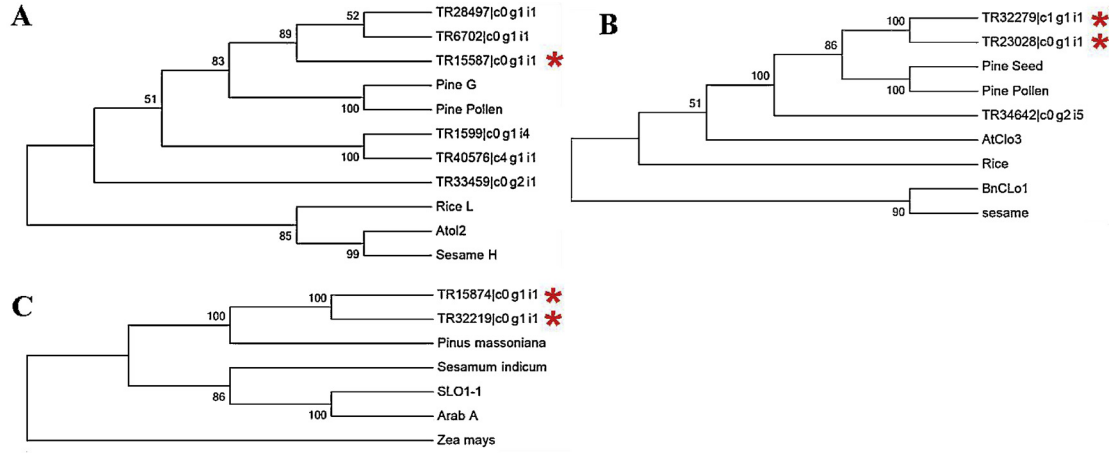
In mature seeds, oil is present mainly in the form of oil bodies, which are stable due to the functions of steric hindrance and electronegative repulsion supplied by their integral proteins (Jiang et al., 2009). Thus, it is necessary to identify the candidate proteins involved in oil body in *T. grandis*. From the sequences of the transcripts, 6, 3 and 3 unigenes which encoded oleosin, caleosin and steroleosin were selected, respectively. Phylogenetic analysis was carried out to characterize the evolutionary relationships of those oil body-associated proteins from *T. grandis* and other known species. As shown in Fig. 5A–C, a unigene of oleosin (*TR15587|c0_g1_i1*, named as *TgOLEO1*) in *T. grandis* had high homology with *Pine G* (pine oleosin presented in gymnosperm) and *Pine pollen*, while two unigenes of caleosin (*TR23028|c0_g1_i1* and *TR32279|c1_g1_i1*, named as *TgCLO1* and *TgCLO2*, respectively) and two unigenes of steroleosin (*TR32219|c0_g1_i1* and *TR15874|c0_g1_i1*, named as *TgSLO1* and *TgSLO2*, respectively) had high homology with that of *Pine* and *Pinus massoniana*, respectively. Besides, the correlation analysis between oil contents and the expression levels of the unigenes for these proteins in five developmental stages of *T. grandis* from X08 revealed that both of *TgOLEO1* and *TgCLO1* were significantly correlated with oil contents (Additional file 7). Furthermore, the alignment analysis of amino acid sequence was also conducted between TgOLEO1, TgCLO1 and TgSLO1, and their homologs in other plants, respectively. As expected, the result showed that the domain of *TgOLEO1* were highly homologous with other oleosins, and possesses the conserved motif known as Proline knot (Fig. 5D); The position was also found in *TgCLO1*, along with a Ca^{2+}-binding region and four phosphorylation sites (one tyrosine kinase and three casein kinase II phosphorylation sites) (Fig. 5E); Three structural regions: Membrane anchoring, NADPH binding, and Sterol binding regions in *TgSLO1* were proposed after comprising with other plants (Fig. 5F). These results suggested that *TgOLEO1, TgCLO1* and *TgSLO1* are probably the candidate genes encoding the proteins involved in oil bodies.

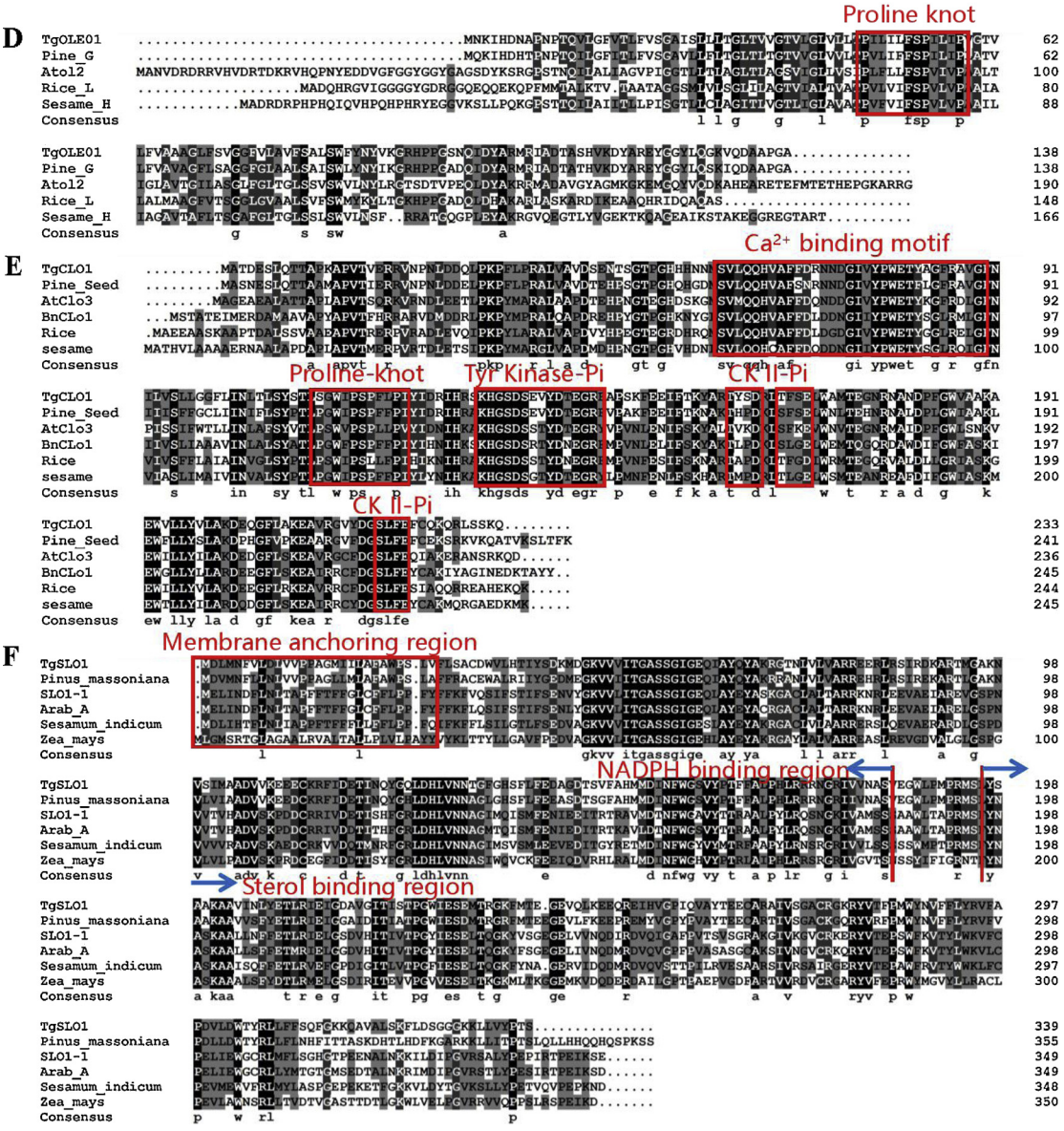
3.5. Subcellular localization of oil body-associated proteins

To further investigate where those oil body-associated proteins function, the three candidate proteins, *TgOLEO1, TgCLO1* and *TgSLO1* were selected to observe their localization. Thus, the coding sequences of those proteins were fused to the C-terminus of GFP driven by the CaMV35S promoter. The fusion constructs then transiently expressed into tobacco (*N. benthamiana*) leaves. As shown in Fig. 6, each of *TgOLEO1, TgCLO1* and *TgSLO1* was co-localized to both the ER marker and Nile red-strained oil body. This indicating that oleosin, caleosin and steroleosin represent bona fide proteins in lipid mobilization from ER to oil bodies. Most importantly, comparing with the controls, there was a higher abundance oil bodies in all the oil body proteins transiently over-expressed tobacco leaves. Notably, the *TgSLO1*-transgeneic leaves had a significant larger size of oil bodies.

3.6. Identification of regulatory genes potentially involved in oil biosynthesis

Due to the biological value and potential economic importance of vegetable oil, the regulation of oil accumulation has been widely studied to increase the oil yield in plants. Some TFs, such as *LEC1*, *WRI1*, *FUS3* and *ABI3*, involved in plant oil metabolism has been identified from several species. In our transcriptomic data in *T. grandis*, 4 unigenes for WRI1 and 1 unigene for FUS3 were detected. The transcripts for *LEC1* and *ABI3* were not detected in any *T. grandis* landrace, suggesting the lipid biosynthesis is not regulated by *LEC1* and *ABI3* in mature kernels of *T. grandis*. However, some unigenes for ABI3 interacting proteins that may be involved in lipid biosynthesis were identified. To further investigate the correlation between the expression level of the





(caption on next page)

Fig. 5. Phylogenetic analysis and amino acid sequence alignment of oil body related proteins from *T. grandis* and other plants. The accession numbers of the sequences used for phylogenetic analysis (A–C) are as follows: Pine G (KJ415242), Atol2 (*Arabidopsis thaliana* Z54164.1), Rice L (U43930) and Sesame H (AF302807). Pine Seed(KJ415240.1), AtClo3 (*Arabidopsis thaliana* At2g33380), BnCLo1 (*Brassica napus* AAY40837.1), Rice(X89891.1), Sesame (AF109921.1). *Pinus massoniana* (KT731102.1), SLO1-1 (*Brassica napus* EU678274), Arab A (*Arabidopsis thaliana* BAA96983.1), *Sesamum indicum* (AF302806) and *Zea mays* (NM_001159142). (D) Alignment of the deduced amino acid sequences of TgOLEO1 with Pine G (KJ415242), Atol2 (*Arabidopsis thaliana* Z54164.1) Rice L (U43930) and Sesame H (AF302807) amino acid sequences. The position of the proline-knot motif is indicated on top of the sequences. (E) Alignment of the deduced amino acid sequences of TgCLO1 with Pine Seed(KJ415240.1), AtClo3 (*Arabidopsis thaliana* At2g33380), BnCLo1 (*Brassica napus* AAY40837.1), Rice(X89891.1) and Sesame (AF109921.1) amino acid sequences. The positions of a calcium-binding motif, a proline-knot motif and four phosporylation sites (one tyrosine kinase and three casein kinase II phosphorylation sites) are indicated on tops of their sequences. (F) Alignment of the deduced amino acid sequences of TgSL01 with *Pinus massoniana* (KT731102.1), SLO1-1 (*Brassica napus* EU678274), Arab A (*Arabidopsis thaliana* BAA96983.1), *Sesamum indicum* (AF302806) and *Zea mays* (NM_001159142) amino acid sequences. The proposed structural regions (membrane anchoring, NADPH binding, and sterol binding) are indicated on the tops of the sequences.

transcripts for WRI1, FUS3, and ABI3 interacting proteins and the expression level of the transcripts for lipid biosynthesis pathway enzymes, Pearson's correlations were calculated (Fig. 7). The results showed that the ABI3 interacting protein transcripts, BRD_TGR51101 and BRD_TGR93779, were positively and significantly correlated with the most of the BCCP transcripts; BRD_TGR74197, BRD_TGR62856, BRD_TGR50862, BRD_TGR44617, BRD_TGR43511, and BRD_TGR36404 were negatively and significantly correlated with the most of the KASII transcripts. However, the ABI3 interacting protein transcript, BRD_TGR40616 was positively and significantly correlated with the most of the KASII transcripts. It was also found that the WRI1 transcript (BRD_TGR14987) and the ABI3 interacting protein transcripts (BRD_TGR25574, BRD_TGR50862 and BRD_TGR62856) were negatively and significantly correlated with almost all the FAD2 transcripts. While, the positive and extremely significant correlations between the ABI3 interacting protein transcripts (BRD_TGR93779 and BRD_TGR40616) and all the FAD2 transcripts were found. From the whole results of Fig. 7, it was found that the ABI3 interacting protein transcripts (BRD_TGR36404, BRD_TGR43511, BRD_TGR44617, BRD_TGR50862 and BRD_TGR62856) play negative regulatory roles in lipid biosynthesis; while the ABI3 interacting protein transcripts, BRD_TGR30687, BRD_TGR40616 and BRD_TGR93779, play positive regulatory roles in lipid biosynthesis. The most of WRI1 transcript (BRD_TGR128) negatively and significantly correlated transcripts are fatty acid biosynthesis related genes, while the most of WRI1 transcript (BRD_TGR128) positively and significantly correlated transcripts are oil assembly related genes (Fig. 7). Those regulators would be helpful for further research in the regulatory mechanism of oil biosynthesis in *T. grandis*.

3.7. Sequence variation analysis and validation of RNA-Seq data by qRT-PCR

To look for sequence variation that could be associated with oil content, cDNAs of oil body assembly and oil-body proteins related genes were sequenced and analyzed in the different *T. grandis* landraces (Additional file 8). A total of 30 SNPs from unigenes of PDAT, GPAT, DGAT, LPAT, *TgOLEO1, TgCLO1* and *TgSLO1* were obtained. Of these 30 SNPs, nine are synonymous. It was found that SNP M2 and M3 from*TgSLO1*, which is unique to the two high oil landraces (R24 and Z11) may be contributed to their high oil content (Additional file 8).

To confirm the data from the RNA-seq, the transcription of the genes encoding PDAT, WRI1 and ABI3 interacting proteins in ten *T. grandis* landrace kernels was analyzed using qRT-PCR. The correlation between the RNAseq and qRT-PCR data was highly significant ($r = 0.7743$, $p < 0.001$), indicating that the expression data obtained by RNA-Seq was reliable (Additional file 9).

4. Discussion

In this study, ten *T. grandis* landraces which are different in their oil content and fatty acid composition were selected. Presumably, the genetic varieties and environmental factors might be responsible for those differences. Then, transcriptome sequencing in ten different *T. grandis*

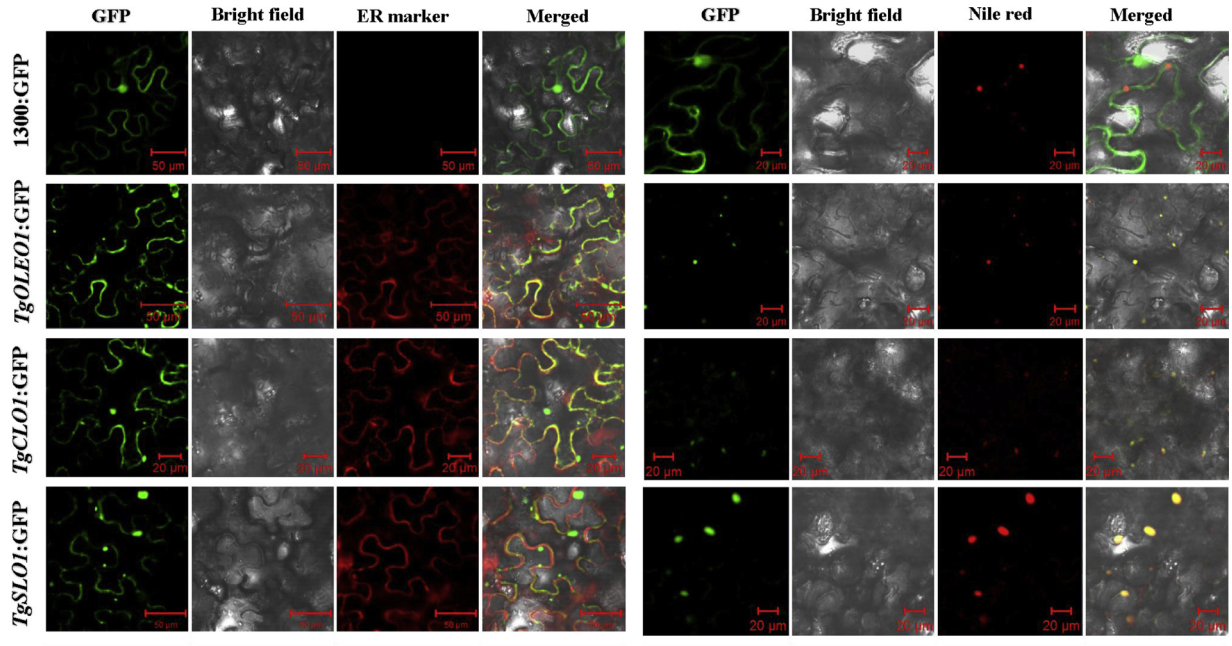


Fig. 6. Representative confocal laser-scanning microscopy images of tobacco leaf cells transiently expressing *TgOLEO1, TgCLO1* and *TgSLO1* with GFP, stained with the neutral lipid-selective dye Nile red. For each set of images, the corresponding merged and bright field images are shown. Bars = 50/20 μm.

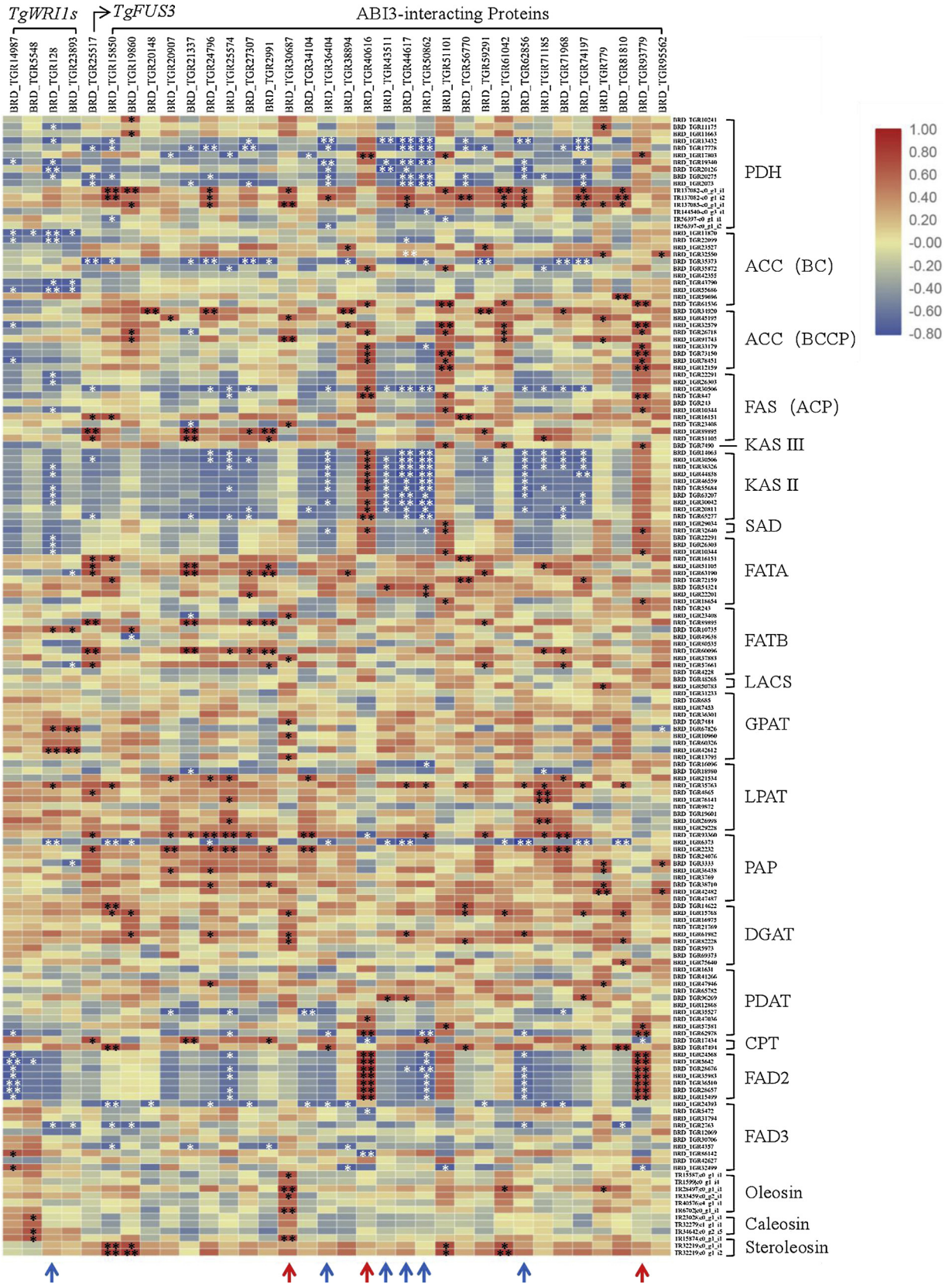


Fig. 7. Correlation analysis between the expression level of the transcripts for WRI1, FUS3, and ABI3 interacting proteins and the expression level of the transcripts for lipid biosynthesis pathway enzymes in the 10 selected landraces. Different colors indicate different Pearson's correlation coefficients. Asterisk (*) means $P < 0.05$; Asterisk (**) means $P < 0.01$. The red arrow indicates that the corresponding column have more significantly and positively correlations; The blue arrow indicates that the corresponding column have more significantly and negatively correlations.

landraces were performed. From the transcriptome data, 112,699 unigenes were obtained. Among these unigenes, 175 unigenes related to lipid biosynthesis, including 126 unigenes for fatty acid biosynthesis, 37 unigenes for TAG assembly and 12 unigenes for oil body proteins, were further identified. Then, the metabolic pathways of lipid biosynthesis were further proposed. Moreover, some key genes for regulating the synthesis of lipid in *T. grandis* were also identified. The correlation analysis between the oil content and the expression levels of unigenes suggested that BC, FATA and PDAT play key roles in oil accumulation. The correlation analysis between the fatty acids and the expression levels of unigenes indicated that PDH, BC, FAS, KASII, SAD, FATA and LACS play important roles in the accumulation of fatty acid components. These results suggest that the regulatory mechanism of lipid biosynthesis in *T. grandis* is different from other plants. This conclusion is supported by the following lines of evidence.

It is reported that increasing activity of key enzymes such as PDH could promote lipid accumulation in oilseed (Rajalakshmi and Natraj, 1991), but there are no positively correlations between the expression of PDH unigenes and lipid components in *T. grandis*. Instead, the PDH unigenes, BRD_TGR20275 and TR56397-c0_g1_i1, were strongly and negatively correlated with oleic acid and eicosadienoic acid, respectively, suggesting that PDH may decrease lipid accumulation in *T. grandis*. It is worth noting that the chain length of saturated FAs was determined by the specific substrates of plant FATA and FATB acyl-ACP thioesterases (Pollard et al., 1991). It is reported that the lack of FATA transcripts in oil palm embryo still contributes to high accumulation of unsaturated FAs (Dussert et al., 2013). In *Arabidopsis*, FATB was demonstrated as a decisive component of FA synthesis and was critical for seed development (Bonaventure et al., 2003). The AtFatA recognized both 14:1-ACP and 16:1-ACP as substrates, while the AtFatB showed the highest activity toword 16:0-ACP but also appeared a significant substrate activity to 18:1-ACP (Salas and Ohlrogge, 2002). However, in *T. grandis*, the high correlation between the level of transcription of FATA and the content of unsaturated FAs suggests that high oil deposition, especially in unsaturated FAs, requires the considerable upregulation of FATA.

It has been shown that *PDAT1* and *DGAT1* are the two main acyltransferase enzymes contributing to TAG biosynthesis in *Arabidopsis* seeds (Banaś et al., 2013). However, the relative contribution of both enzymes in TAG biosynthesis is uncertain. From this study, it is obvious that PDAT showed a tighter connection with the content of oil in *T. grandis* than DGAT, suggesting PDAT contributed more in oil accumulation than DGAT. Nevertheless, the importance of DGAT, which is essential to TAG biosynthesis in many oilseeds should not be ignored. On the one hand, RNAi strategy was used to prove that *DGAT1* and *PDAT1* have supplementary effects during embryonic development in *Arabidopsis* seeds. On the other hand, it was observed that low level expression of *DGAT1* was sufficiently to maintain normal oil body size (Zhang et al., 2009). Therefore, to dig out the potential value of the two genes by genetic engineering is important not only for laying the foundation for further understanding of plant oil synthesis, but also for improving strategies for regulating oil synthesis in *T. grandis*.

In this study, the lipid body proteins, oleosin, caleosin and steroleosin, in *T. grandis* were evaluated by phylogenetic analysis, protein alignment, correlation analysis and subcellular localization. The current studies have focused on the identification of plant lipid body proteins while research of the mechanism about how they function in the physiological aspects remains elusive. In *Arabidopsis*, Oleo1 was the most abundant isoform of seed oleosin, its suppression resulted in a larger form of lipid bodies (Siloto et al., 2006). It is suggested that the accumulation of oleosins might regulate the size of oil body in vivo by correlation analysis between oil body size and oleosin levels (Ting et al., 1996; Tzen et al., 1993). In soybean (*Glycine max*), RNA interference of oleosin resulted in the formation of giant oil body and disrupted seed germination (Schmidt and Herman, 2008). The expression of oleosins was spatially and developmentally regulated by the transactivator, ABI3 (Crowe et al., 2000; Keddie et al., 1994). This study also provides evidence of the overall increased expression of oleosin during the development of *T. grandis* kernels. However, overexpression of steroleosin formed oil droplets larger than native pine oil bodies, particularly for the recombinant steroleosin which lack of the sterol binding domain (Pasaribu et al., 2016). Consistently, microscopic observation also showed that overexpression of the *TgSLO1* resulted in bigger size of oil bodies in this study. Besides, overexpressed these oil body-associated proteins resulted in increased amount of oil bodies, suggesting their crucial role on the developing *T. grandis* seeds.

Previous studies have identified a set of TFs, such as LEC1, LEC2, ABI3, FUS3 and WRI1, which exhibit extremely important roles in regulating synthesis and accumulation of lipid in plant seeds. In *T. grandis*, the transcripts of ABI3 interacting proteins, FUS3 and WRI1 were preliminarily identified, while the regulatory mechanism of those TFs needs further verification. The LEC1, LEC2 and ABI3 transcripts were not detected, suggesting that a lack of importance for the three TFs in the mature stage of kernels in *T. grandis*. Up-regulated or down-regulated expression of these genes will make a great change in the oil accumulation. Study showed that the ectopic expression of *LEC1* in *Arabidopsis* can activate the expression of *ABI3* and *FUS3*, and further resulted in the accumulation of seed storage protein (Kagaya et al., 2005). LEC2, FUS3 and ABI3 cannot only interact with genes involved in fatty acid biosynthesis, but also participate in oil accumulation by more sophisticated ways (Baud and Lepiniec, 2009). For instance, LEC2 can quickly and directly activate the expression of seed specific genes such as S3 *oleosin* in *Arabidopsis* leaves (Santos Mendoza et al., 2005). It was also found that *WRI1* is regulated by LEC2 and is essential for regulating LEC2 in fatty acid metabolism (Baud et al., 2007). In this study, the expression of *LEC1*, *LEC2* and *ABI3* were not detected in the mature kernels, suggesting the lipid biosynthesis is not regulated by *LEC1*, *LEC2* and *ABI3* in the mature kernels of *T. grandis*. *LEC1*, *LEC2* and *ABI3* may be expressed in the early or/and middle development stages of *T. grandis* kernels to regulate the lipid biosynthesis, which requires further study. WRI1 is a unique TF, which can regulate glycolytic and FA metabolism in plant. Researches showed that WRI1 functions by driven downstream genes of pyruvate kinase (*PKp-β1*), ACCase (*BCCP2* and *BC*), ACP (*ACP1*), ENR (*ENR*) and KAS (*KASI*) (Baud et al., 2009; Fukuda et al., 2013; Maeo et al., 2009; Qu et al., 2012). Some studies showed that WRI1 play positive regulatory roles in oil accumulation (Cemac and Benning, 2004; Baud et al., 2007; Liu et al., 2010). In this study, it was found that the TgWRI1 transcripts negatively and significantly correlated with most of the transcripts of fatty acid biosynthesis pathway enzymes, suggesting that WRI1 may play negative regulatory roles in fatty acid biosynthesis. In addition, there are still positive and significant correlations between WRI1 transcripts and oil assembly related transcripts. Therefore, whether WRI1 plays a positive or negative regulatory role in oil accumulation needs further verification. In future, the functions of these candidate genes and the regulatory mechanisms of lipid biosynthesis and oil accumulation should be further characterized, which might be important for breeding new cultivars with high qualities.

5. Conclusion

In summary, comparative transcriptome analysis were carried out based on the Illumina platform in ten different *T. grandis* landraces and 112,699 transcripts were obtained which provide important genetic resource in our research. The lipid biosynthetic pathway in *T. grandis* were proposed and all the genes involved in lipid biosynthesis were identified from the transcriptome data. In *T. grandis,* BC, FATA, and PDAT play key roles in oil accumulation, which differed from other plants, suggesting that a unique regulatory mechanism of oil accumulation in *T. grandis* might exist. The candidate genes (*TgOLEO1, TgCLO1* and *TgSLO1*) for oil body-associated proteins were further identified. Subcellular localization assay indicated that all of the proteins were

localized in both of ER and oil body. Overexpression of the proteins could increase the amount of oil bodies. Furthermore, some candidate regulatory transcripts (*TgWRI1* and *TgFUS3*, etc.) that may be involved in the regulation of lipid biosynthesis were also identified. Taken together, this study preliminarily revealed the unique molecular mechanism of lipid biosynthesis pathway in *T. grandis*, which will be critical for breeding new cultivars with high oil contents.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31670687); the Special Fund for Forest Scientific Research in the Public Welfare (201504708); the Young Elite Scientists Sponsorship Program by China Academy of Space Technology (CAST) (2018QNRC001); the State Key Laboratory of Subtropical Silviculture (ZY20180312 and ZY20180209); and the Open fund of Zhejiang Key Laboratory for Modern Silvicultural Technology (201607); the Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, Huzhou University (HUZUL201905).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.indcrop.2019.112051.

References

Banaś, W., Garcia, A.S., Banaś, A., Stymne, S., 2013. Activities of acyl-CoA: diacylglycerol acyltransferase (DGAT) and phospholipid: diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. Planta 237, 1627–1636. https://doi.org/10.1007/s00425-013-1870-8.

Bates, P.D., Stymne, S., Ohlrogge, J., 2013. Biochemical pathways in seed oil synthesis. Curr. Opin. Plant Biol. 16, 358–364. https://doi.org/10.1016/j.pbi.2013.02.015.

Baud, S., Dubreucq, B., Miquel, M., Rochat, C., Lepiniec, L., 2008. Storage reserve accumulation in Arabidopsis: metabolic and developmental control of seed filling. Arabidopsis Book 6, e0113. https://doi.org/10.1199/tab.0113.

Baud, S., Lepiniec, L., 2009. Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. Plant Physiol. Biochem. 47, 448–455. https://doi.org/10.1016/j.plaphy.2008.12.006.

Baud, S., Mendoza, M.S., To, A., Harscoet, E., Lepiniec, L., Dubreucq, B., 2007. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant J. 50, 825–838. https://doi.org/10.1111/j.1365-313X.2007.03092.x.

Baud, S., Wuilleme, S., To, A., Rochat, C., Lepiniec, L., 2009. Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in Arabidopsis. Plant J. 60, 933–947. https://doi.org/10.1111/j.1365-313X.2009.04011.x.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., Golani, I., 2001. Controlling the false discovery rate in behavior genetics research. Behav. Brain Res. 125, 279–284. https://doi.org/10.1016/S0166-4328(01)00297-2.

Bo, S., Allen, W.B., Peizhong, Z., Changjiang, L., Kimberly, G., Jerry, R., Douglas, N., Tarczynski, M.C., 2010. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant Physiol. 153, 980–987. https://doi.org/10.1104/pp.110.157537.

Bonaventure, G., Salas, J.J., Pollard, M.R., Ohlrogge, J.B., 2003. Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. Plant Cell 15, 1020–1033. https://doi.org/10.1105/tpc.008946.

Cemac, A., Benning, C., 2004. WRINKLEDl encodes an AtE/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. Plant J. 40, 575–585. https://doi.org/10.1111/j.1365-313X.2004.02235.x.

Chen, B.Q., Cui, X.Y., Xin, Z., Zhang, Y.H., Piao, H.S., Kim, J.H., Lee, B.C., Pyo, H.B., Yun, Y.P., 2006. Antioxidative and acute antiinflammatory effects of *Torreya grandis*. Fitoterapia 77, 262–267. https://doi.org/10.1016/j.fitote.2006.03.019.

Chen, J.C., Tsai, C.C., Tzen, J.T., 1999. Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds. Plant Cell Physiol. 40, 1079. https://doi.org/10.1093/oxfordjournals.pcp.a029490.

Chen, J.C.F., Tzen, J.T.C., 2001. An in vitro system to examine the effective phospholipids and structural domain for protein targeting to seed oil bodies. Plant Cell Physiol. 42, 1245–1252. https://doi.org/10.1093/pcp/pce160.

Conesa, A., Gotz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008, 619832. https://doi.org/10.1155/2008/619832.

Crowe, A.J., Abenes, M., Plant, A., Moloney, M.M., 2000. The seed-specific transactivator, ABI3, induces oleosin gene expression. Plant Sci. 151, 171–181. https://doi.org/10.1016/S0168-9452(99)00214-9.

Dong, D.D., Wang, H.F., Xu, F., Xu, C., Shao, X.F., Li, H.S., 2014. Supercritical carbon dioxide extraction, fatty acid composition, oxidative stability, and antioxidant effect of *Torreya grandis* seed oil. J. Am. Oil Chem. Soc. 91, 817–825. https://doi.org/10.1007/s11746-014-2419-0.

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

Fukuda, N., Ikawa, Y., Aoyagi, T., Kozaki, A., 2013. Expression of the genes coding for plastidic acetyl-CoA carboxylase subunits is regulated by a location-sensitive transcription factor binding site. Plant Mol. Biol. 82, 473–483. https://doi.org/10.1007/s11103-013-0075-7.

Furmanek, T., Demski, K., Banaś, W., Haslam, R., Napier, J., Stymne, S., Banaś, A., 2014. The utilization of the acyl-CoA and the involvement PDAT and DGAT in the biosynthesis of erucic acid-rich triacylglycerols in Crambe seed oil. Lipids 49, 327. https://doi.org/10.1007/s11745-014-3886-7.

Gidda, S.K., Park, S., Pyc, M., Yurchenko, O., Cai, Y., Wu, P., Andrews, D.W., Chapman, K.D., Dyer, J.M., Mullen, R.T., 2016. Lipid droplet-associated proteins (LDAPs) are required for the dynamic regulation of neutral lipid compartmentation in plant cells. Plant Physiol. 170, 2052–2071. https://doi.org/10.1104/pp.15.01977.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Huang, R., Huang, Y., Sun, Z., Huang, J., Wang, Z., 2017. Transcriptome analysis of genes involved in lipid biosynthesis in the developing embryo of pecan (*Carya illinoinensis*). J. Agr. Food Chem. 65, 4223–4236. https://doi.org/10.1021/acs.jafc.7b00922.

Ischebeck, T., 2016. Lipids in pollen-They are different. Biochim. Biophys. Acta 1861, 1315–1328. https://doi.org/10.1016/j.bbalip.2016.03.023.

Jiang, P.L., Chen, J.C., Chiu, S.T., Tzen, J.T., 2009. Stable oil bodies sheltered by a unique caleosin in cycad megagametophytes. Plant Physiol. Biochem. 47, 1009–1016. https://doi.org/10.1016/j.plaphy.2009.07.004.

Jinye, M., Helin, T., Qi, Z., Fuyou, F., Yan, L., Jian, Z., Xiaohui, Y., Tai, W., Kang, C., Xiu-Jie, W., 2008. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol. 148, 1042–1054. https://doi.org/10.1104/pp.108.126342.

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T., 2005. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiol. 46, 399–406. https://doi.org/10.1093/pcp/pci048.

Keddie, J.S., Tsiantis, M., Piffanelli, P., Cella, R., Hatzopoulos, P., Murphy, D.J., 1994. A seed-specific *Brassica napus* oleosin promoter interacts with a G-box-specific protein and may be bi-directional. Plant Mol. Biol. 24, 327–340. https://doi.org/10.1007/BF00020171.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Ohlrogge, J., 2013. Acyl-Lipid Metabolism. Arabidopsis Book 11. https://doi.org/10.1199/tab.0161.

Li-Jen, L., Tai, S.S.K., Chi-Chung, P., Tzen, J.T.C., 2002. Steroleosin, a sterol-binding dehydrogenase in seed oil bodies. Plant Physiol. 128, 1200–1211. https://doi.org/10.1104/pp.010982.

Lin, L.J., Tzen, J.T.C., 2004. Two distinct steroleosins are present in seed oil bodies☆. Plant Physiol. Biochem. 42, 601–608. https://doi.org/10.1016/j.plaphy.2004.06.006.

Liu, J., Hua, W., Zhan, G., Wei, F., Wang, X., Liu, G., Wang, H., 2010. Increasing seed mass and oil content in transgenic Arabidopsis by the overexpression of wri1-like gene from *Brassica napus*. Plant Physiol. Biochem. 48, 9–15. https://doi.org/10.1016/j.plaphy.2009.09.007.

Maeo, K., Tokuda, T., Ayame, A., Mitsui, N., Kawai, T., Tsukagoshi, H., Ishiguro, S., Nakamura, K., 2009. An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J. 60, 476–487. https://doi.org/10.1111/j.1365-313X.2009.03967.x.

Meng, Z., Jilian, F., Taylor, D.C., Ohlrogge, J.B., 2009. DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21, 3885–3901. https://doi.org/10.1105/tpc.109.071795.

Miquel, M., Trigui, G., d'Andréa, S., Kelemen, Z., Baud, S., Berger, A., Deruyffelaere, C., Trubuil, A., Lepiniec, L., Dubreucq, B., 2014. Specialization of oleosins in oil body dynamics during seed development in Arabidopsis seeds. Plant Physiol. 164, 1866–1878. https://doi.org/10.1104/pp.113.233262.

Moreno-Pérez, A.J., Mónica, V.C., Vaistij, F.E., Salas, J.J., Larson, T.R., Rafael, G., Graham, I.A., Enrique, M.F., 2012. Reduced expression of FatA thioesterases in Arabidopsis affects the oil content and fatty acid composition of the seeds. Planta 235, 629–639. https://doi.org/10.1007/s00425-011-1534-5.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Næsted, H., Frandsen, G.I., Jauh, G.Y., Hernandez-Pinzon, I., Nielsen, H.B., Murphy, D.J., Rogers, J.C., Mundy, J., 2000. Caleosins: Ca^{2+}-binding proteins associated with lipid bodies. Plant Mol. Biol. 44, 463–476. https://doi.org/10.1023/A:1026564411918.

Pan, X., Siloto, R.M.P., Wickramarathna, A.D., Mietkiewska, E., Weselake, R.J., 2013. Identification of a pair of phospholipid: diacylglycerol acyltransferases from developing flax (*Linum usitatissimum L.*) seed catalyzing the selective production of

trilinolenin. J. Biol. Chem. 288, 24173–24188. https://doi.org/10.7939/R3891245V.
Pasaribu, B., Chung, T.Y., Chen, C.S., Jiang, P.L., Tzen, J.T.C., 2016. Identification of steroleosin in oil bodies of pine megagametophytes. Plant Physiol. Biochem. 101, 173–181. https://doi.org/10.1016/j.plaphy.2016.02.008.
Pasaribu, B., Wang, M.M., Jiang, P.L., 2017. Identification of an oleosin-like gene in seagrass seeds. Biotechnol. Lett. 39, 1757–1763.
Penfield, S., Rylott, E.L., Gilday, A.D., Graham, S., Larson, T.R., Graham, I.A., 2004. Reserve mobilization in the Arabidopsis endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. Plant Cell 16, 2705–2718. https://doi.org/10.1105/tpc.104.024711.
Pollard, M.R., Anderson, L., Fan, C., Hawkins, D.J., Davies, H.M., 1991. A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of *Umbellularia californica*. Arch. Biochem. Biophys. 284, 306–312. https://doi.org/10.1016/0003-9861(91)90300-8.
Qu, J., Ye, J., Geng, Y.F., Sun, Y.W., Gao, S.Q., Zhang, B.P., Chen, W., Chua, N.H., 2012. Dissecting functions of KATANIN and WRINKLED1 in cotton fiber development by virus-induced gene silencing. Plant Physiol. 160, 738–748. https://doi.org/10.1104/pp.112.198564.
Rajalakshmi, K.J., Natraj, C.V., 1991. Kinetic analyses of the pyruvate dehydrogenase complex from *Candida* 107 (NCYC 911). Indian J. Biochem. Bio. 28, 389. https://doi.org/10.1016/0009-3084(91)90027-9.
Rao, S.S., Hildebrand, D., 2009. Changes in oil content of transgenic soybeans expressing the yeast gene. Lipids 44, 945–951. https://doi.org/10.1007/s11745-009-3337-z.
Schmidt, M.A., Herman, E.M., 2008. Suppression of soybean oleosin produces micro-oil bodies that aggregate into body/ER complexes. Mol. Plant 1, 910–924. https://doi.org/10.1093/mp/ssn049.
Salas, J.J., Ohlrogge, J.B., 2002. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Arch. Biochem. Biophys. 403, 25–34. https://doi.org/10.1016/S0003-9861(02)00017-6.
Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M., Lepiniec, L., 2005. LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. FEBS Lett. 579, 4666–4670. https://doi.org/10.1016/j.febslet.2005.07.037.
Shimada, T.L., Hara-Nishlmura, I., 2010. Oil-body-Membrane proteins and their physiological functions in plants. Biol. Pharm. Bull. 33, 360–363. https://doi.org/10.1248/bpb.33.360.
Shiu-Cheung, L., Weselake, R.J., 2006. Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. Lipids 41, 1073. https://doi.org/10.1007/s11745-006-5057-y.
Shockey, J., Regmi, A., Cotton, K., Adhikari, N., Browse, J., Bates, P.D., 2016. Identification of Arabidopsis GPAT9 (At5g60620) as an essential gene involved in triacylglycerol biosynthesis. Plant Physiol. 170, 163–179. https://doi.org/10.1104/pp.15.01563.
Siloto, R.M.P., Kim, F., Arturo, L.V., Yeung, E.C., Nykiforuk, C.L., Moloney, M.M., 2006. The accumulation of oleosins determines the size of seed oilbodies in Arabidopsis. Plant Cell 18, 1961–1974. https://doi.org/10.1105/tpc.106.041269.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. https://doi.org/10.1093/molbev/msr121.
Theodoulou, F.L., Eastmond, P.J., 2012. Seed storage oil catabolism: a story of give and take. Curr. Opin. Plant Biol. 15, 322–328. https://doi.org/10.1016/j.pbi.2012.03.017.
Ting, J.T., Lee, K., Ratnayake, C., Platt, K.A., Balsamo, R.A., Huang, A.H., 1996. Oleosin genes in maize kernels having diverse oil contents are constitutively expressed independent of oil contents. Size and shape of intracellular oil bodies are determined by the oleosins/oils ratio. Planta 199, 158–165.
Tzen, J., Cao, Y., Laurent, P., Ratnayake, C., Huang, A., 1993. Lipids, proteins, and structure of seed oil bodies from diverse species. Plant Physiol. 101, 267–276. https://doi.org/10.1104/pp.101.1.267.
Wang, H., Guo, J., Lambert, K.N., Lin, Y., 2007. Developmental control of Arabidopsis seed oil biosynthesis. Planta 226, 773–783. https://doi.org/10.1007/s00425-007-0524-0.
Wu, J.S., Huang, J.D., Hong, Y.W., Zhang, H.Z., Ding, M.Z., Lou, H.Q., Hu, Y.Y., Yu, W.W., Song, L.L., 2018. De novo transcriptome sequencing of *Torreya grandis* reveals gene regulation in sciadonic acid biosynthesis pathway. Ind. Crop. Prod. 120, 47–60. https://doi.org/10.1016/j.indcrop.2018.04.041.
Xu, C., Shanklin, J., 2016. Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. Annu. Rev. Plant Biol. 67, 179–206. https://doi.org/10.1146/annurev-arplant-043015-111641.
Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11, R14. https://doi.org/10.1186/gb-2010-11-2-r14.
Zhang, M., Cao, X., Jia, Q., Ohlrogge, J., 2016. FUSCA3 activates triacylglycerol accumulation in Arabidopsis seedlings and tobacco BY2 cells. Plant J. 88, 95–107. https://doi.org/10.1111/tpj.13233.
Zhang, M., Fan, J., Taylor, D.C., Ohlrogge, J.B., 2009. DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21, 3885–3901. https://doi.org/10.1105/tpc.109.071795.
Zheng, P., Allen, W.B., Roesler, K., Williams, M.E., Zhang, S., Li, J., Glassman, K., Ranch, J., Nubel, D., Solawetz, W., Bhattramakki, D., Llaca, V., Deschamps, S., Zhong, G.Y., Tarczynski, M.C., Shen, B., 2008. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat. Genet. 40, 367–372.