



Article

# RNA Sequencing and Coexpression Analysis Reveal Key Genes Involved in $\alpha$ -Linolenic Acid Biosynthesis in *Perilla frutescens* Seed

Tianyuan Zhang <sup>1,2</sup> , Chi Song <sup>3</sup>, Li Song <sup>2</sup>, Zhiwei Shang <sup>1</sup>, Sen Yang <sup>1</sup>, Dong Zhang <sup>3</sup>, Wei Sun <sup>3</sup>, Qi Shen <sup>1,\*</sup> and Degang Zhao <sup>1,2,\*</sup>

<sup>1</sup> Rapeseed Research Institute, Guizhou Academy of Agricultural Sciences, Guiyang 550008, China; zhangtianyuan@foxmail.com (T.Z.); szw1225@163.com (Z.S.); yangsen.90@163.com (S.Y.)

<sup>2</sup> The Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Guizhou University, Guiyang 550025, China; lpsssl@126.com

<sup>3</sup> Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China; Csong@icmm.ac.cn (C.S.); dzhang1987@icmm.ac.cn (D.Z.); wsun@icmm.ac.cn (W.S.)

\* Correspondence: shenqi-gz1983@163.com (Q.S.); dgzhao@gzu.edu.cn (D.Z.); Tel.: +86-15085087045 (Q.S.); +86-13885012693 (D.Z.)

Received: 26 October 2017; Accepted: 15 November 2017; Published: 16 November 2017

**Abstract:** *Perilla frutescens* is used as traditional food and medicine in East Asia. Its seeds contain high levels of  $\alpha$ -linolenic acid (ALA), which is important for health, but is scarce in our daily meals. Previous reports on RNA-seq of perilla seed had identified fatty acid (FA) and triacylglycerol (TAG) synthesis genes, but the underlying mechanism of ALA biosynthesis and its regulation still need to be further explored. So we conducted Illumina RNA-sequencing in seven temporal developmental stages of perilla seeds. Sequencing generated a total of 127 million clean reads, containing 15.88 Gb of valid data. The de novo assembly of sequence reads yielded 64,156 unigenes with an average length of 777 bp. A total of 39,760 unigenes were annotated and 11,693 unigenes were found to be differentially expressed in all samples. According to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, 486 unigenes were annotated in the “lipid metabolism” pathway. Of these, 150 unigenes were found to be involved in fatty acid (FA) biosynthesis and triacylglycerol (TAG) assembly in perilla seeds. A coexpression analysis showed that a total of 104 genes were highly coexpressed ( $r > 0.95$ ). The coexpression network could be divided into two main subnetworks showing over expression in the medium or earlier and late phases, respectively. In order to identify the putative regulatory genes, a transcription factor (TF) analysis was performed. This led to the identification of 45 gene families, mainly including the AP2-EREBP, bHLH, MYB, and NAC families, etc. After coexpression analysis of TFs with highly expression of FAD2 and FAD3 genes, 162 TFs were found to be significantly associated with two FAD genes ( $r > 0.95$ ). Those TFs were predicted to be the key regulatory factors in ALA biosynthesis in perilla seed. The qRT-PCR analysis also verified the relevance of expression pattern between two FAD genes and partial candidate TFs. Although it has been reported that some TFs are involved in seed development, more direct evidence is still needed to verify their function. However, these findings can provide clues to reveal the possible molecular mechanisms of ALA biosynthesis and its regulation in perilla seed.

**Keywords:** *Perilla frutescens*; RNA sequencing (RNA-seq);  $\alpha$ -linolenic acid (ALA); triacylglycerol (TAG) biosynthesis; herbgenomics

## 1. Introduction

*Perilla frutescens*, a traditional food and medicinal plant, belongs to the family Lamiaceae [1]. It is widely cultivated in East Asian countries, especially in China, Korea, and Japan. There are two cultivated

varieties of perilla—*Pf. var. frutescens*, mainly used for edible and health purposes, and *Pf. var. crispa*, mainly used as medicine, vegetable, and spice [2]. The seeds of *Pf. var. frutescens* contain approximately 45–55% oil, which comprises 54–64%  $\alpha$ -linolenic acid (ALA) (C18:3) [3]. The ALA content of perilla seeds is significantly higher than that of the other major oil seed crops, e.g., soybean (5–13%) [4], rapeseed (1–8%), maize (0.5–2%), and sunflower (0.5–2%) [5]. Linolenic acid (C18:3) plays an important role in the maintenance of brain nerve system and has an obvious positive effect on intelligence, memory, and eyesight, but is usually scarce in our daily meals [6–8]. Therefore, the oils of perilla seed have attracted wide attention in research and health domain. However, the mechanism of ALA and its regulation in perilla remains unclear.

With the development of high-throughput sequencing technology, RNA-seq has become an effective method to analyze the spatiotemporal expression of a gene and to obtain more comprehensive information about gene transcription and regulation [9]. Some oil crop plant seed contain specific fatty acid (FA) contents. Understanding those specific lipid formation process is extremely important to unravel the unique lipid metabolism and regulation mechanisms [10]. Recently, transcriptome sequencing and characterization based on the Illumina second-generation sequencing technology has enabled the rapid identification and profiling of mechanisms involving oil content and FA composition in various oil plants, such as soybean [11,12], peanut [13,14], and palm oil [15]. Similarly with perilla, the flax [16], sea buckthorn [17], *Camelina sativa* [18], *Plukenetia volubilis* L. [19], and tree peony [20] also rich in ALA in seed oil. In sequencing of flax seed, the key embryogenesis regulators were mined [14]. And more biosynthesis genes contributing in ALA accumulation have been identified in those plant seeds [11–20]. For perilla, 540 unique genes involved in acyl-lipid metabolism from four developmental stages in perilla seeds, and characterized the expression profiles of 43 genes involved in FA and triacylglycerol (TAG) synthesis [21].

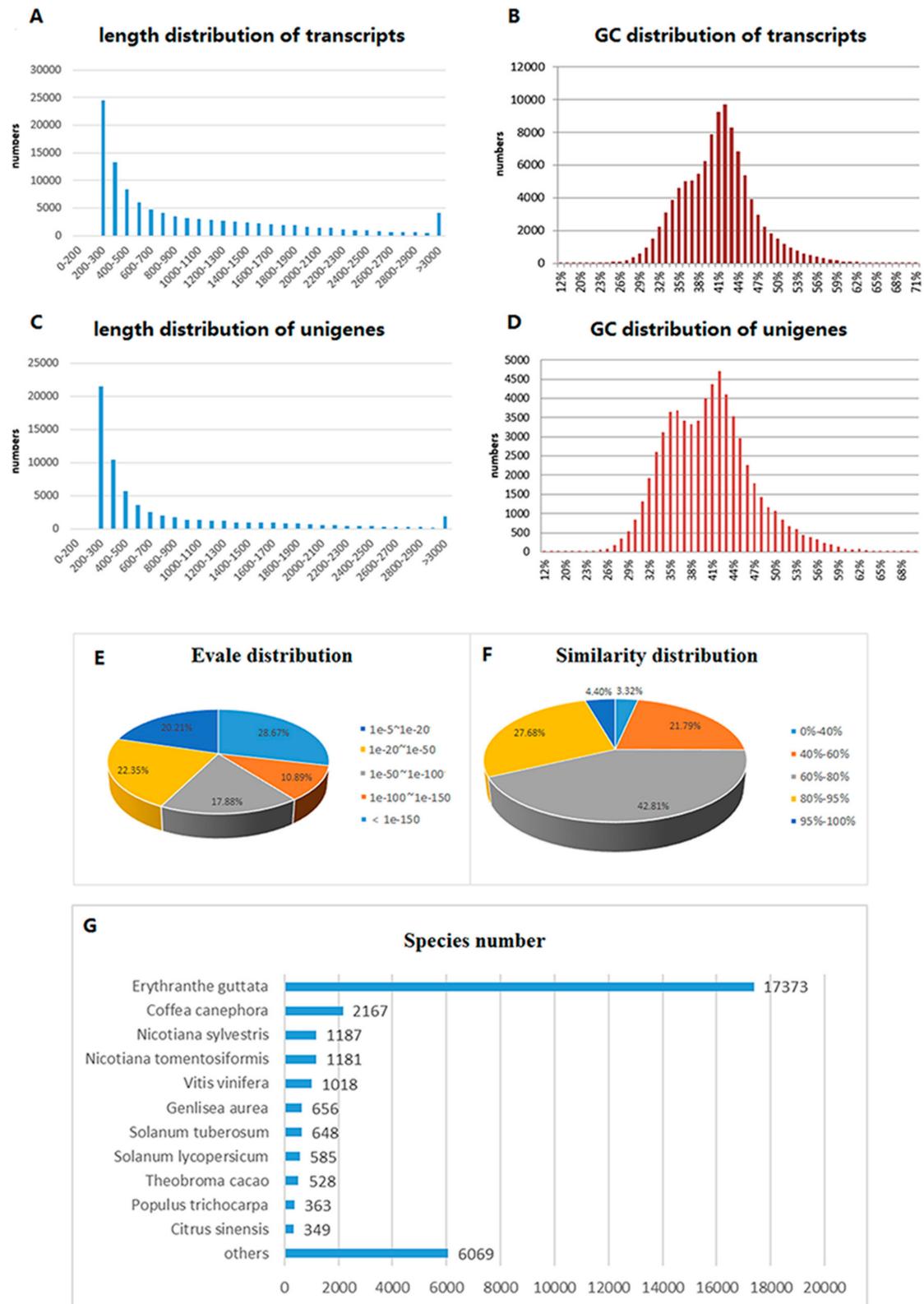
Coexpression analysis, based on the “guilt-by-association” principle, is used to identify the genes that have similar expression patterns and are more likely to be functionally associated [22]. Recently, coexpression analysis has proved to be a powerful tool for identifying genes and regulatory factors in transcriptional networks in human [23,24], plants [25–28], and animals [29,30]. In *Arabidopsis*, a large number of transcription factors (TFs) involved in seed development, FA and protein biosynthesis, flavone metabolism, and several important metabolic pathways were identified by using the coexpression method [31–34].

In this study, we conducted an Illumina RNA sequencing in seven developmental stages of perilla seeds. The coexpression of lipid metabolism genes and TFs identification provided more genes and regulation involving in ALA biosynthesis. Overall, the transcriptome-wide identification and coexpression analysis provide a foundation for possible molecular mechanisms of ALA biosynthesis and its regulation in perilla seed.

## 2. Results

### 2.1. Transcriptome Sequencing and Assembly

In order to explore the possible molecular mechanism of ALA biosynthesis and its regulatory in perilla seed, seven different seed developmental stages after flowering (2, 6, 10, 14, 18, 22, and 26 DAF) were collected and subjected to Illumina paired-end sequencing. After removing adaptor sequences and filtering low-quality and ambiguous reads, a total of 127 million clean reads, containing 15.88 Gb of valid data were acquired. All sequencing statistics are shown in Table 1. Subsequently, we obtained 104,638 transcripts and 64,156 unigenes by de novo assembly using the Trinity program (Table 2). The average length of transcripts was 968 bp and that of N50 was 1608 bp. Furthermore, the average length of unigenes was 777 bp and that of N50 was 1417 bp. Of all unigenes, 20,965 (32.68%) possibly contained complete open reading frames (ORFs), which were predicted by the program TransDecoder. The GC content and length distribution of all assembled unigenes and contigs are presented in Figure 1. The sequencing data have been deposited in the National Center for Biotechnology Information (NCBI) database (Accession number: SRP111892).



**Figure 1.** Overview of the de novo assembly of transcriptome sequencing in *Perilla frutescens* and annotation based on a non-redundant (NR) protein database. Length (A) and GC distribution (B) of transcripts; length (C) and GC distribution (D) of unigenes are shown; (E) e-value distribution of BLAST hits for the assembled unigenes; (F) Similarity score distribution of the top BLAST hits for the assembled unigenes; (G) Species distribution of the top BLAST hits for the assembled unigenes.

**Table 1.** Summary of perilla seed transcriptome data sequenced by the Illumina platform.

Sample ID	Total Reads	Total Bases	GC Content	Q20	Q30
2DAF	18,094,914	2,261,864,250	46.64%	94.64%	89.76%
6DAF	17,885,482	2,235,685,250	47.88%	94.36%	89.70%
10DAF	17,075,076	2,134,384,500	49.31%	94.38%	89.77%
14DAF	16,747,764	2,093,470,500	49.61%	94.25%	89.58%
18DAF	14,929,122	1,866,140,250	51.42%	93.70%	88.05%
22DAF	20,751,538	2,593,942,250	52.52%	93.74%	88.71%
26DAF	21,517,792	2,689,724,000	47.89%	95.39%	91.21%
Total	127,001,688	15,875,211,000			

DAF: days after flowering.

**Table 2.** Statistics of de novo assembly of sequence reads.

Item	Total Number (bp)	N50 (bp)	Median Length (bp)	Average Length (bp)	Total Length (bp)
Transcripts	104,638	1608	600	968	101,378,085
Unigenes	64,156	1417	402	777	49,883,108

## 2.2. Functional Annotation and Classification

For functional annotation, all assembled unigenes were searched against the a non-redundant (NR), Swiss-Prot, TrEMBL, and Clusters of Orthologous Groups (COG), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Table 3).

**Table 3.** Statistics of annotations of assembled unigenes.

Database	Account	Percentage <sup>c</sup>
NR <sup>a</sup>	32,132	50.08%
KEGG classified unigenes	10,904	17.00%
COG classified unigenes	8654	14.47%
GO classified unigenes	22,263	34.70%
Blast_hit <sup>b</sup>	31,287	48.77%
Pfam classified unigenes	19,340	30.15%
Eggnog classified unigenes	9425	14.69%
TmHMM classified unigenes	6719	10.47%
SignalP classified unigenes	2354	3.67%
All annotated unigenes	39,760	61.97%
All	64,156	100.00%

<sup>a</sup> NCBI non-redundant database; <sup>b</sup> SWISSPORT and TREMBLE database; <sup>c</sup> Percentage of all assembled unigenes.

In NR database, a total of 32,132 (50.08%) unigenes showed matched items. Among them, 9211 (28.67%) unigenes showed an *e*-value less than  $1e^{-150}$ , and 3500 (10.89%) unigenes had an *e*-value between  $1e^{-100}$ , and  $1e^{-150}$  (Figure 1E). In addition, 1413 (4.40%) unigenes exhibited alignment identities greater than 95%, and 8,895 (27.68%) exhibited alignment identities between 80% and 95% (Figure 1F).

The species annotation results showed that 17,373 (54.08%) unigenes were highly matched with *Erythranthe guttata*. Other unigenes matched with *Coffea canephora* (6.75%), *Nicotiana sylvestris* (3.70%), *Nicotiana tomentosiformis* (3.68%), and *Vitis vinifera* (3.17%) (Figure 1G).

However, only 86 unigenes matched with *P. frutescens* sequences in the NR database. It is possible that the genomic and transcriptomic information is currently lacking for perilla in the NR database (only 105 sequences).

Based on the alignment of paralogs or orthologs, 8654 unique sequences matched in the COG database were clustered into 25 functional categories (Figure S1). Among the 25 categories, the largest category was general function prediction only (2702; 31.22%), followed by transcription (1727; 29.96%) and replication, recombination and repair (1652; 19.09%). A GO functional classification was used to classify unigene functions on the basis of NR annotation. Furthermore, 22,263 (34.70%) unigenes were assigned to one or more GO terms, which contained cellular component (18,113; 81.36%), molecular function (18,847; 84.66%), and biological process (18,119; 81.39%) (Figure S2).

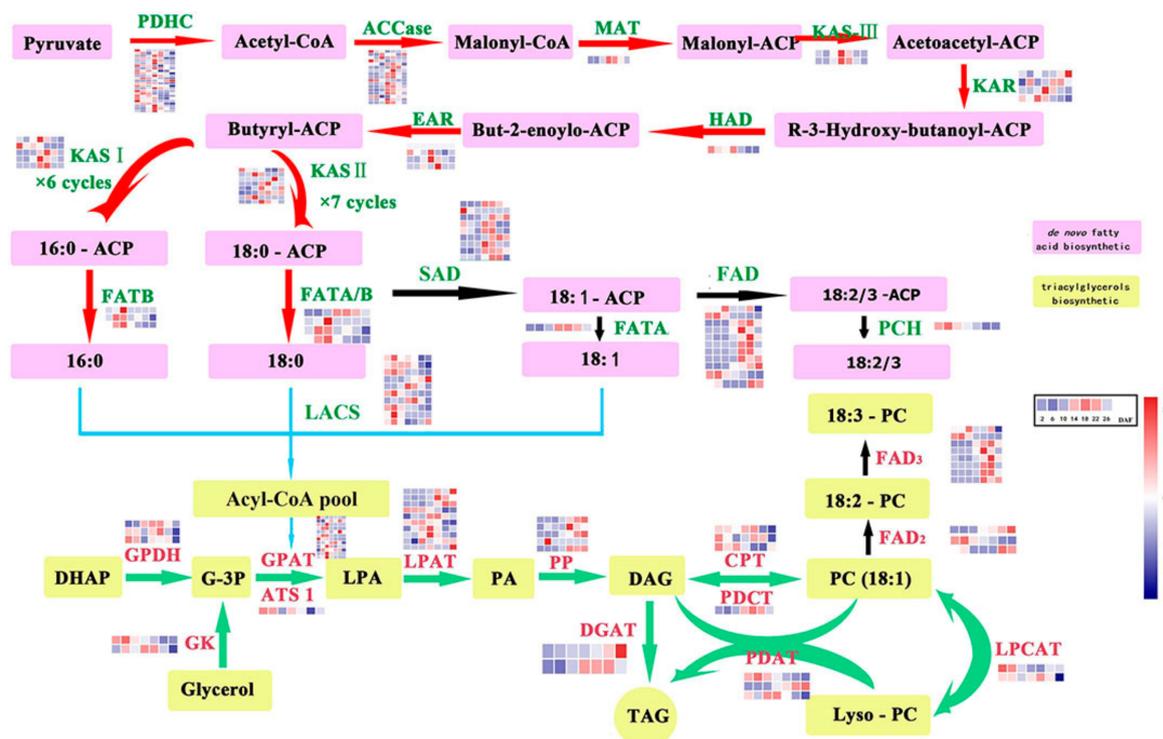
A KEGG pathway analysis was performed to identify the biological metabolism pathways. A total of 10,904 unigenes were grouped into 301 KEGG pathways, mainly including the organismal systems (1661), metabolism (4818), genetic information processes (2479), environmental information processing (1098), and cellular processes (1133) (Figure S3).

### 2.3. Genes Related to Lipid Biosynthesis in *Perilla* Seed

Based on the annotation results, 371 (4.01%) unigenes in COG annotation were assigned to lipid transport and metabolism, 2624 (11.79%) unigenes were assigned to the GO term “developmental process,” and 486 (4.3%) unigenes in the KEGG annotation were assigned to lipid metabolism pathway, mainly including “fatty acid degradation” (67 unigenes; ko00071), “glycerolipid metabolism pathway” (78 unigenes; ko00561), “biosynthesis of unsaturated fatty acids” (42 unigenes; ko01040), “linoleic acid metabolism pathway” (21 unigenes; ko00591), and “ $\alpha$ -linolenic acid metabolism pathway” (64 unigenes; ko00592) (Table S1).

For acquiring more knowledge on FA biosynthesis and TAG assembly pathway genes in perilla seed, more attention was paid to the 150 unigenes that were annotated as FA biosynthesis and TAG assembly genes (Figure 2 and Table S2). In this study, we found 40 FA biosynthesis unigenes, of which 25 encoded pyruvate dehydrogenase complex (pdh A, B, C, and D) subunits, 14 encoded acetyl-CoA carboxylase (ACCase) subunits, and one encoded malonyl-CoA ACP transacylase (MAT). A total of 39 unigenes were involved in fatty acid acyl chain elongation; of these, 15 unigenes encoded 3-ketoacyl-ACP synthases (KAS), one unigene encoded hydroxyacyl-ACP Dehydrase (HAD), four unigenes encoded ketoacyl-ACP reductase (KAR), four unigenes encoded enoyl-ACP reductase (EAR), five unigenes encoded acyl-ACP thioesterase (FAT) and palmitoyl-CoA hydrolase (PCH), and 10 unigenes encoded long-chain acyl-CoA synthetases (LACS). Of the 20 unigenes involved in fatty acid desaturase, nine encoded stearoyl-ACP desaturase (SAD), three encoded oleate desaturase (FAD2), and eight unigenes encoded linoleate desaturase (FAD3).

The Kennedy pathway in perilla seed showed the involvement of 36 unigenes in the TAG assembly pathway. Of these unigenes, three encoded glycerol-3-phosphate dehydrogenase (GPDH), 13 encoded glycerol-3-phosphate acyltransferase (GPAT) and homologous gene (glycerol-3-phosphate-O-acyltransferase ) ATS1, two encoded glycerol kinase (GK), 10 encoded lysophosphatidic acid acyltransferase (LPAT) and homologous gene phospholipase A2 (PLA2), five encoded phosphatidic acid phosphatase (PAP). Two unigenes encoded acyl-CoA: diacylglycerol acyltransferase (DGAT, which transfers an acyl group from acyl-CoA to the sn-3 position of sn-1,2-diacylglycerol to form TAG). Three unigenes encoded phospholipid: diacylglycerolacyltransferase (PDAT), which preferentially transfers an acyl group from the sn-2 position of a phospholipid to diacylglycerol. A unigene encoding phosphatidylcholine: diacylglycerol choline phosphotransferase (PDCT) and three unigenes encoding diacylglycerol cholinephosphotransferase (CPT) were identified in the perilla seed.



**Figure 2.** Perilla sequences associated with fatty acid biosynthetic pathway. Each row represents a gene, and each column represents a specimen (stage). Depths of color in the red and blue rectangles indicate higher and lower represents the Z-score RNA expression level. Identified enzymes include: PDHC: pyruvate dehydrogenase complex; ACCase: acetyl-CoA carboxylase; MAT: malonyl-CoA ACP transacylase; ACP: acyl carrier protein; KAS I, II, III: ketoacyl-ACP synthase I, II, III; KAR: ketoacyl-ACP reductase; HAD: hydroxyacyl-ACP dehydrase; EAR: enoyl-ACP reductase; SAD: stearyl-ACP desaturase; FAD: fatty acid desaturase; FATA/B: fatty acyl-ACP thioesterase A/B; PCH: palmitoyl-CoA hydrolase; LACS: long-chain acyl-CoA synthetase; FAD2: oleate desaturase (endoplasmic reticulum); FAD3: linoleate desaturase; GK: glycerol kinase; GPDH: glycerol-3-phosphate dehydrogenase; GPAT: glycerol-3-phosphate acyltransferase; LPAT: 1-acylglycerol-3-phosphate acyltransferase; ATS1: glycerol-3-phosphate O-acyltransferase; PP: phosphatidate phosphatase LPIN; DGAT, acyl-CoA: diacylglycerolacyltransferase; PDAT: phospholipid:diacylglycerol acyltransferase; LPCAT: 1-acylglycerol-3-phosphocholine acyltransferase; CPT: diacylglycerol cholinephosphotransferase; PDCT: phosphatidylcholine:diacylglycerol cholinephosphotransferase.

#### 2.4. Differentially Expressed Genes (DEGs)

To identify the DEGs in a developed seed of perilla, we calculated the FPKM (Fragments Per Kilobase of exon per Million mapped fragments) values of the assembled unigenes at each developmental stage of perilla seed and performed the pairwise comparison of different development stages. A total of 17,416 unigenes were found to be widely expressed in each samples (FPKM > 1 in all samples). Among them 11,693 unigenes showed more than two-fold different expression level. Among samples, the 6DAF vs. 10DAF had 294 up-regulated and 558 down-regulated unigenes, and the 10DAF vs. 14DAF had 353 up-regulated and 368 down-regulated unigenes, which were the two most highly different adjacent stages (Table 4).

Subsequently, we performed GO analyses to identify the function of these DEGs (Figure S2). In the GO classification analysis, 7501 unigenes were assigned to three main GO functional categories. The largest sub-category in the “molecular function” category was ion binding (2928 unigenes, accounting for 25.04% of all DEGs), followed by DNA binding (911, 7.62%), and oxidoreductase activity (878; 7.51%). In “cellular component”, the largest sub-category was cell (5237; 44.79%),

followed by intracellular (4363; 58.17%), and organelle (3792; 37.31%). In “biological process”, the largest sub-category was biosynthetic process (2214; 18.93%), followed by cellular nitrogen compound metabolic process (1594; 13.63%), and response to stress (1275; 10.90%).

**Table 4.** Differentially Expressed Genes (DEGs) between two different developmental stages.

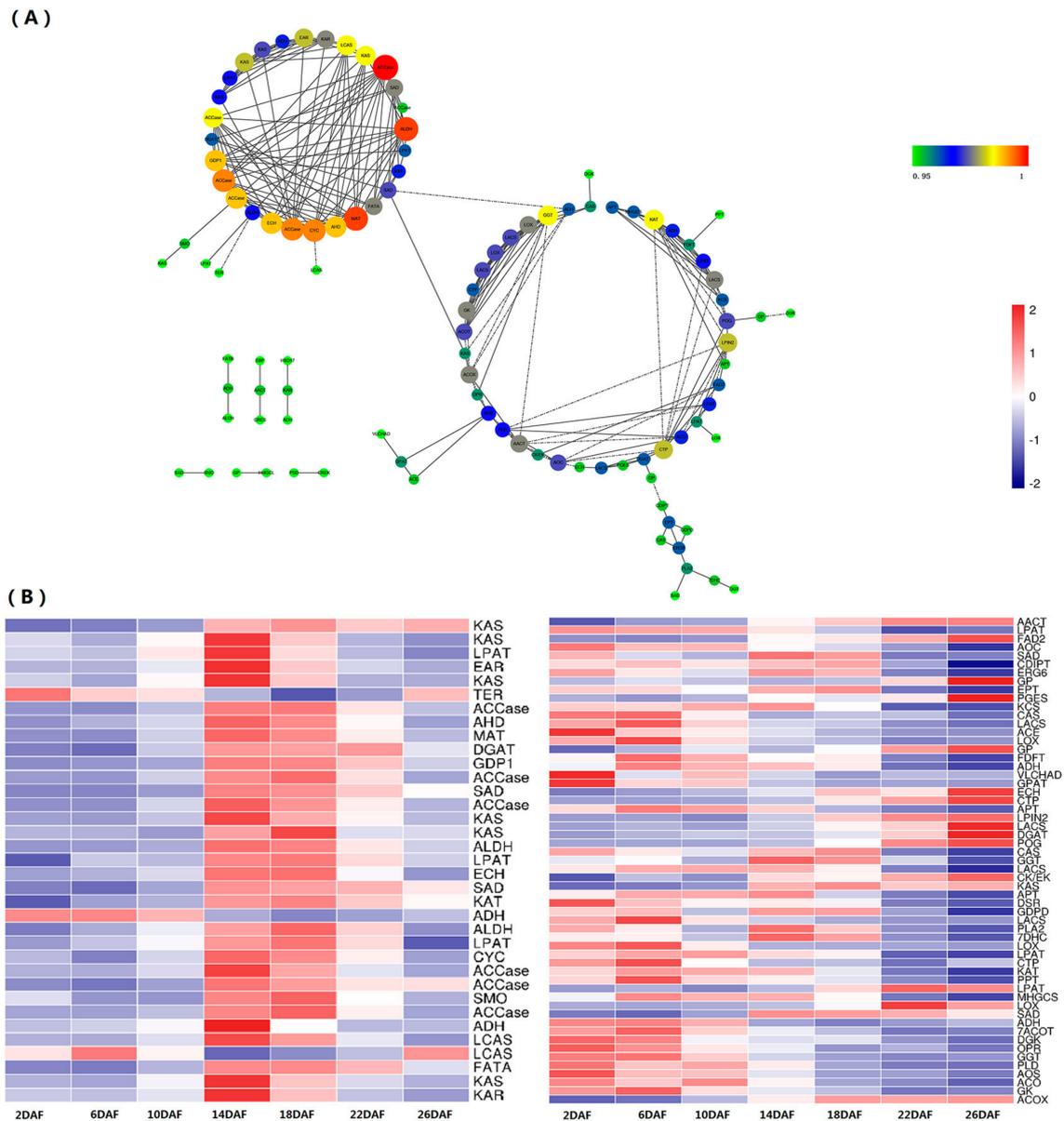
	2DAF	6DAF	10DAF	14DAF	18DAF	22DAF	26DAF
2DAF	0						
6DAF	137↑ 263↓	0					
10DAF	876↑ 1316↓	294↑ 558↓	0				
14DAF	1755↑ 1399↓	1352↑ 1227↓	353↑ 368↓	0			
18DAF	3022↑ 1813↓	2370↑ 1583↓	2001↑ 955↓	257↑ 85↓	0		
22DAF	3666↑ 2031↓	3188↑ 1816↓	2941↑ 1363↓	1350↑ 674↓	47↑ 83↓	0	
26DAF	4751↑ 2221↓	4730↑ 2119↓	4445↑ 1692↓	2615↑ 1268↓	1253↑ 949↓	360↑ 122↓	0

DAF: days after flowering.

### 2.5. Coexpression of Lipid Metabolism Genes

The transcriptome-wide coexpression analysis showed that out of the 486 perilla lipid metabolism genes, 104 were highly correlated ( $r > 0.95$ ). The coexpression network presenting two main links (containing 89 unigenes) could be divided into two relatively independent subnetworks (I and II). Subnetwork I contained 35 unigenes, mainly including the TAG biosynthesis genes, such as SAD, LACS, PP, LPAT, and DGAT. Whereas subnetwork II contained 57 unigenes, mainly including the de novo FA synthesis genes, such as ACCase, PDCH, KAS, SAD, and FATA, and some transferase and peroxidase. Fifteen unigenes, mainly including aldehyde dehydrogenase (NAD<sup>+</sup>, ALDH), ceramide kinase (CKER), ceramide synthetase (CERS), and so on, showed relatively less correlation with others.

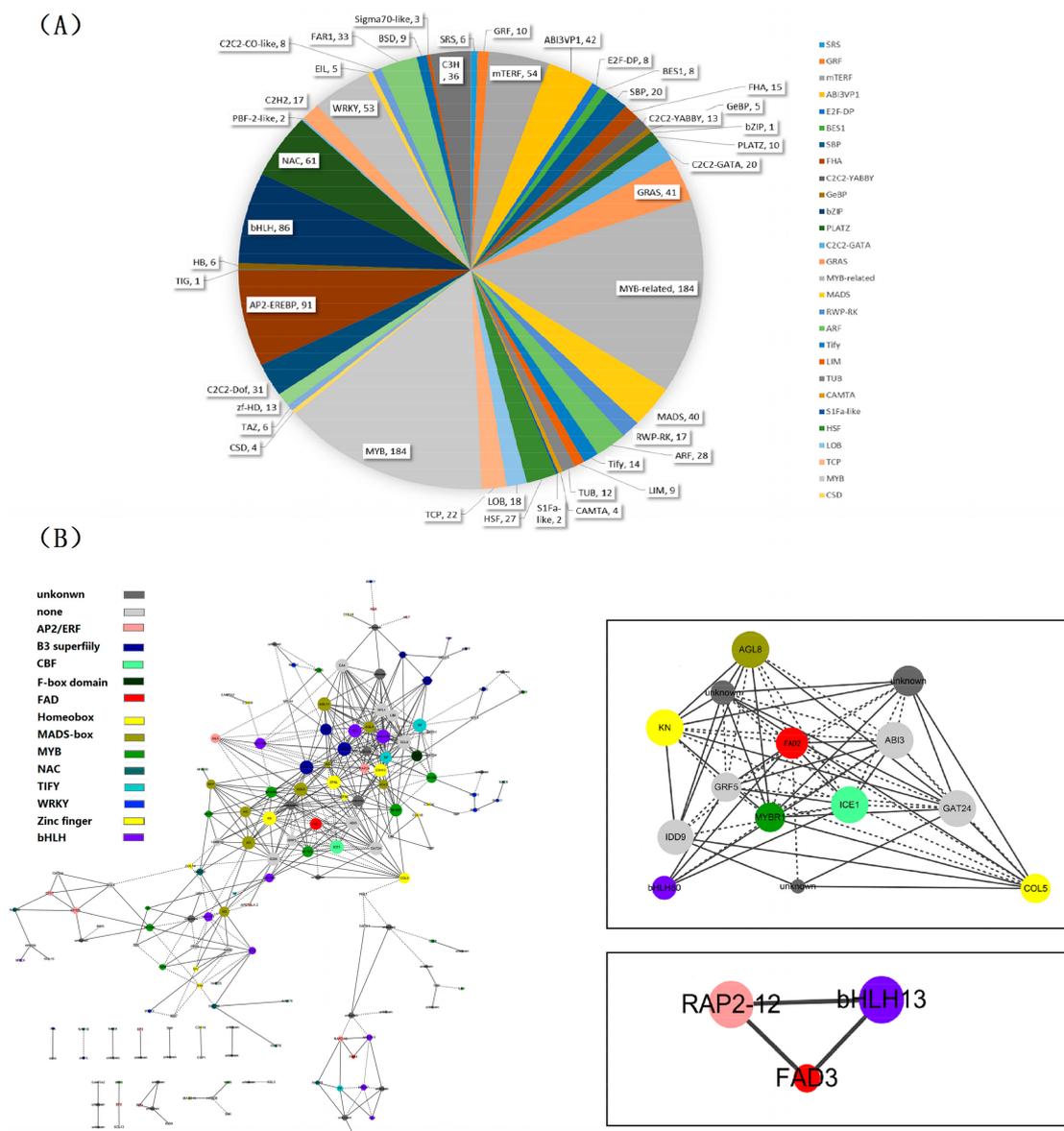
Surveys and summaries of the expression patterns of these genes in the seeds at various developmental stages showed that subnetwork I genes were preferentially highly expressed in 14–18DAF, which showed consistent coexpressed tendency. Whereas subnetwork II guide genes showed not only consistent but also negative correlation coexpressed tendency. The de novo FA synthesis genes preferentially highly expressed in the early stages of development, and some transferase and peroxidase related lipid degradation or regulation preferentially highly expressed in the late stages of development (Figure 3).



**Figure 3.** (A) A lipid metabolism-enriched module is presented with the degree-sorted circle layout of Cytoscape v3.4.10, with the sizes and colors of nodes reflecting the level of connectivity within the network. The bigger the node, the greater the number of connections it has. For clarity, the edges with correlation values smaller than 0.95 were removed; (B) Heat maps of the coexpression genes of lipid metabolism; The gene in left heat maps is correspond with the subnetwork I, and The genes in right heat maps is correspond with the subnetwork II. Each line represents a gene, and each column represents a specimen (stage). Depths of color in the red and blue rectangles indicate higher and lower represents the Z-score RNA expression lever.

## 2.6. Identification of Transcription Factor Families

For more information about the regulation of genes involved in lipid biosynthesis during the development of perilla seeds, the database PlnTFDB was used for identifying TFs in all unigenes. We identified a total of 1279 TFs, belonging to 45 TF families, among which the top five families were related to MYB (184), MYB-related (184), AP2-EREBP(91), bHLH(86), and NAC(61) (Figure 4A).



**Figure 4.** Transcription factor analysis. **(A)** Distribution of transcription factor (TF) families; **(B)** Coexpression network of transcript factors and fatty acid desaturase (FAD). The TF module was presented by Cytoscape v3.4.10. The rectangle indicates the TFs directly related to the FAD in the network. The solid line represents positive correlation, and the dotted line represents negative correlation.

### 2.7. Identification and Functional Annotation of the Genes Coexpressing with Perilla Fatty Acid Desaturase Genes

ALA is abundantly accumulated in mature perilla seeds. It is important for health, but is scarce in our daily meals. For identifying the TFs involved in ALA biosynthesis and its regulation, the omega-3 and omega-6 fatty acid desaturase genes (DN26137\_c0\_g1 and DN25736\_c0\_g1) were defined as the “guide genes”. Based on these guide genes, a total of 162 TFs related to the two FADs ( $r > 0.95$ ) were identified (Figure 4B and Table S3).

Most of these TFs belonged to six gene families, including MYB (16), AP2/ERF (12), bHLH (11), MADS-box (10), Zinc finger (10), and NAC (10) (Table S3). The *RAP2-12* (DN23390\_c0\_g1,  $r = 0.981$ ) and *bHLH13* (DN35936\_c1\_g1,  $r = 0.962$ ) TFs were found to be significantly correlated with omega-3 fatty acid desaturase. Thirteen TFs, including *ABI3* (DN33484\_c3\_g1,  $r = 0.986$ ), *ICE1*

(DN34648\_c0\_g1,  $r = 0.9877491$ ), *IDD5* (DN33027\_c2\_g5,  $r = -0.965$ ), *GRF5* (DN29710\_c0\_g1,  $r = 0.995$ ), *KN* (DN24707\_c1\_g1\_i1,  $r = -0.964$ ), *COL5* (DN28351\_c0\_g1,  $r = -0.974$ ), *GAT24* (DN30716\_c0\_g1,  $r = -0.963$ ), *bHLH80* (DN31044\_c0\_g1,  $r = 0.982$ ), *MYBR1* (DN27387\_c0\_g1,  $r = -0.989$ ), *AGL8* (DN20914\_c0\_g1,  $r = -0.962$ ) and three unknown genes, were significantly correlated with omega-6 fatty acid desaturase gene. Moreover, we identified the *WRI1* gene in the coexpression network.

### 2.8. The qRT-PCR Analysis of the Lipid Synthesis Genes in the Perilla

To confirm the gene expression results in our research, seven genes related to lipid synthesis and its regulation were selected for qRT-PCR analysis. The qRT-PCR results significantly correlated with RNA-Seq result were verified ( $r = 0.697$ ) (Figure S4). Among them, *FAD3* (DN26137\_c0\_g1), *FAD2* (DN25736\_c0\_g1), *RAP2-12* (DN23390\_c0\_g1), *bHLH13* (DN35936\_c1\_g1), *WRI1* (DN28398\_c0\_g1), and *GNA1* (DN19692\_c12\_g4) showed highly expression levels in 18DAF; but *GRF5* (DN29710\_c0\_g1), gene were mainly expressed in 26DAF (Figure S5). Further, the remarkable correlation were verified between *FAD3* and *RAP2-12* ( $r = 0.975$ ) and *bHLH13* ( $r = 0.979$ ) in qRT-PCR analysis, and the other genes also showed remarkable correlation relationship with *FAD3* gene (Table S4). These candidate TFs identified provided clues to clone and reveal the key molecular mechanisms underlying unsaturated FA biosynthesis and its regulation in perilla seeds.

## 3. Discussion and Conclusions

The transcriptome of perilla seed have been reported by Kim et al. [21], involving in four developmental stages in perilla seeds, and the genes related to acyl-lipid metabolism and FA and TAG synthesis were analysis. On the basis of research, we reported the RNA sequencing of seven different development phases in perilla seed. The number of unigenes and N50 slightly higher than previously reported [21] and support that homology analysis and identification of acyl-lipid genes. In total, we obtained 64,156 unigenes and 11,693 differentially expressed unigenes from our seed samples. The 8654, 10,904, and 22,263 unigenes with the COG, GO, and KEGG annotation, respectively were identified (Table 3). Among them, 150 unigenes were annotated as FA biosynthesis and TAG assembly genes (Figure 2 and Table S2). In the acyl-lipid genes, *PDAT* gene showed the relatively higher expression in perilla developed seeds that suggested it might be a priority pathway for the synthesis of perilla oil (Figure 3B). The TAG molecules can be stored in the form of an oil body (OB) surrounded by a layer of phospholipid membrane and embedded with several molecules of a protein called oleosin in mature plant seeds [35–37]. The main function of oleosin is to help stabilize OBs and prevent the fusion of OBs [38]. Five unigenes encoding oleosin were identified in the perilla seed transcriptome. It is noteworthy that three unigenes (DN25414\_c0\_g1, DN24171\_c0\_g1, and DN18941\_c0\_g1) showed high-level transcription in the late development, implying that they were involved in the formation of oil bodies. The analysis and identification of acyl-lipid genes effective supported the previous results [21,39]. Next, we analyzed the coexpression of lipid metabolism genes and identification of TFs. In our study, gene coexpression networks were identified from lipid biosynthesis genes of perilla. Two subnetworks, mainly representing the TAG biosynthesis and de novo FA synthesis genes, were obtained (Figure 3A). The results are similar with the transcriptome of *Arabidopsis* seeds from the globular to mature embryo stage, which reported three coexpression subnetworks involved in FA biosynthesis, oleosins and seed storage proteins, and TAG assembly pathways [40]. In perilla seed, the expression patterns of these genes also showed a coincident trend that the subnetwork I genes were preferentially highly expressed in the middle stage and the subnetwork II genes were preferentially highly expressed in the early or late stages of development (Figure 3B).

The perilla seed transcriptome and acyl-lipid metabolism genes have been systematically analyzed according to Kim et al. [21]. The *FAD2* and *FAD3* were showed obviously highly expressed in interim development phase of perilla seed [21]. As is known, the omega-6 (*FAD2*) and omega-3 (*FAD3*) fatty acid desaturase are important enzymes that catalyze the formation of ALA. In our study, *FAD2* and *FAD3* also showed some expression tendency, which provides key clues of the high ALA accumulation

in perilla seeds. Therefore, we used the two *FADs* as inquiry to execute the co-expression analysis, and found 162 TFs related to these two *FADs*. Acquired TFs belonged to important regulatory gene families of lipid biosynthesis and seed development, such as AP2/ERF (12), bHLH (11), MADS-box (10), B3 (9), MYB (16), Zinc finger (10), and so on (Figure 4A). As reported previously, the AP2/ERF-family TFs are involved in the control of primary and secondary metabolism, growth and development, and responses to environmental stimuli [41]. The *WRI* gene, encoding a 48.4-kDa protein with two AP2 binding domains, directly regulates oil accumulation and storage in seeds, leading to the production of enough oil needed for seedling establishment [42]. The main target genes of *WRI1* are involved in glycolysis and FA synthesis in *Arabidopsis* [43]. In perilla *FAD* coexpression results, *WRI1* and three other AP2/ERF family members (*ERF*, *AIL*, and *RAP*) were identified. The basic helix-loop-helix (bHLH) proteins are a large superfamily of TFs. bHLH TF plays a key role in the transcriptional regulation of genes related to storage lipid biosynthesis and accumulation during seed development [44]. The bHLH proteins were associated with fruit development and ripening in tomato recently report by Sun et al. [45]. More bHLH and bHLH-related genes were identified is a key regulator of *FAD2* expression in *Arabidopsis*, sesame and cotton [44,46,47]. The SebHLH protein mediates transactivation of the Se*FAD2* gene promoter through binding to E- and G-box elements [44]. The *bHLH* ( $r = 0.962$ ) showed coexpression with perilla *FAD3* implies the potential function of ALA biosynthesis and its regulation (Figure 4B). Next, growth-regulating factor (GRF4) is an important TF, which was reported to regulate grain size and yield by OsmiR396 controls in rice [48]. It was found to coexpress with two *FADs*. This TF gene from coexpression analysis gave more clue for future research to uncover ALA biosynthesis and its regulation mechanism. The possible TFs regulating ALA biosynthesis and co-expression research showed more information to understand ALA underlying biosynthesis and regulation than previous reports [21].

In conclusion, we focused on the key genes and TFs involved in ALA biosynthesis in perilla seed. The results will serve as an important basis for perilla seed development analysis and provide more critical information and characterization of ALA synthesis mechanism and TAG biosynthesis in oilseeds.

## 4. Materials and Methods

### 4.1. Plant Materials, Library Construction, and Transcriptome Sequencing

The plants of *Pf. var. frutescens* were grown in a farm of the Guizhou Rapeseed Institute, Guizhou Province, China (31°39' N, 119°19' E). For transcriptome samples, the seeds at seven development stages, that is, 2, 6, 10, 14, 18, 22, and 26 days after flowering (DAF), were collected from the same individual plant and were quickly frozen in liquid nitrogen. The total RNA from all seed samples was extracted separately using the RNAPrep Pure Plant Kit (Tiangen, Beijing, China) following the manufacturer's instructions. An amount of 1–2 µg total RNA per sample with 28S/18S RNA ratio  $\geq 1.8$  used for library preparation. The mRNA sequencing library was constructed using the NEBNext® Ultra RNA Library Prep Kit (New England Biolabs Inc., Ipswich, MA, USA). The sequencing library was analyzed using the Agilent 2100 Bioanalyzer with a minimum integrity number (RIN) value of 7 and the insertion element are 250–300 bp. The library was sequenced using an Illumina HiSeq™ 2000 sequencing system (Illumina Inc., San Diego, CA, USA).

### 4.2. Data Filtering and De Novo Assembly

After removing the adapter sequences, low-quality sequences with N percentage (i.e., the percentage of nucleotides in read which could not be sequenced) over 5%, and those containing more than 50% bases with  $q$ -value  $\leq 5$ , and sequences shorter than 35 bases, the clean reads from seven different developmental-phase seeds were obtained. The clean reads were de novo assembled into transcripts using the Trinity software with min\_kmer\_cov set to 4 and all other parameters set to

default values [49]. The complete open reading frames (ORFs) of unigenes were predicted using the program TransDecoder (<https://transdecoder.github.io/>).

#### 4.3. Functional Annotation of Unigenes

For functional annotation, all assembled unigenes were searched against the non-redundant protein database of the National Center for Biotechnology Information (NCBI) (NR, <http://www.ncbi.nlm.nih.gov>), Swiss-Prot (<http://www.expasy.ch/sprot>), TrEMBL (<http://www.ebi.ac.uk/trembl>), KOG (<http://www.ncbi.nlm.nih.gov/COG/>) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg>) using the basic local alignment search tool (BLAST) program with a cut-off *e*-value of  $<1e^{-5}$ . Subsequently, gene ontology (GO) annotation was performed using the Blast2GO and ClusterProfiler [50,51] programs to determine the GO functional classifications according to molecular function, biological process, and cellular component.

#### 4.4. Identification and Coexpression Analysis of Lipid Metabolism-Related Genes.

The genes related to lipid metabolism were selected according to the KEGG annotation results. For coexpression analysis, the Pearson correlation coefficient (*r*) based on FPKM was calculated using Python [52]. Coexpression networks were analyzed across all developmental stages. The genes that were not expressed in 60% of the samples were excluded from the analysis. The paired genes with a Pearson correlation coefficient (*r*) greater than 0.95 were considered as significantly coexpressed genes and were selected to build a coexpression network using the Perl script, the Perl script available at <https://git.coding.net/zhangtianyuan/coexpression-analysis.git>.

#### 4.4. Identification of Differentially Expressed Genes

The abundance of unigenes was normalized using the FPKM (Fragments Per Kilobase of exon per Million mapped fragments) values. The differential gene expression analysis was performed on the paired samples of the seven developmental stages of perilla seeds using the edgeR package [53]. The absolute value of  $\log_2$ Foldchange  $> 1$  and the false discovery rate (FDR)  $< 0.05$  was used to identify the significance of different gene expression. Subsequently, the GO functional enrichment and KEGG pathway enrichment analyses of the DEGs were performed using the software tools Goseq and KOBAS, respectively [54,55].

#### 4.5. Transcription Factor Identification and Co-Expression with Fatty Acid Desaturase

Transcription factor (TF) families were identified using the known plant TF database PlnTFDB (<http://plntfdb.bio.uni-potsdam.de/v3.0>). To select the possible TFs regulating ALA biosynthesis, two highly expressed fatty acid desaturase genes—*FAD3* (catalyzes linoleic acid to  $\alpha$ -linolenic acid) and *FAD2* (catalyzes oleic acid to linoleic acid)—were constructed for a co-expressed analysis with a default value 0.6 [51]. As described above, the paired genes showing a Pearson correlation coefficient (*r*) greater than 0.95 were considered as significantly coexpressed genes and selected to build a coexpression network using the Perl script. Data correlation and visualization were performed using the program Cytoscape v3.4.10 [56].

#### 4.6. Quantitative Real-Time PCR Analysis

The selected genes were confirmed by qRT-PCR using Rotor-GeneQ (Qiagen, Hilden, Germany) with SYBR Green qPCR SuperMix (Transgene, Beijing, China). Plant materials were planted and collected in same condition as transcriptome sample, RNA were extracted and transcribed to cDNA as sample template. The gene primers were designed using primer premier 3.0. The qPCR using two-step method, and the genes expression quantity were analyzed by  $2^{-\Delta\Delta Ct}$  method using the perilla actin sequence as the internal gene [57]. We all used the four stages (2, 10, 18, and 26 DAF) in samples from four biologic repetition by qRT-PCR.

The qPCR reactions as followed: Each 20  $\mu$ L reaction mixture contained 10  $\mu$ L of SYBR Green qPCR SuperMixUDG TaqTM, 1.5  $\mu$ L of diluted cDNA, 0.4  $\mu$ L of each primer (10  $\mu$ M), 0.4  $\mu$ L of ROX

Reference Dye (50\*) and 7.7 µL of double distilled water. The qPCR cycling conditions were as follows: 50 °C for 2min; followed by 40 cycles of 94 °C for 5 s, and 60 °C 30 s in PCR strip tubes.

**Supplementary Materials:** Supplementary materials can be found at [www.mdpi.com/1422-0067/18/11/2433/s1](http://www.mdpi.com/1422-0067/18/11/2433/s1).

**Author Contributions:** Tianyuan Zhang and Chi Song were responsible for analyzed the data and wrote draft the manuscript. Qi Shen, Zhiwei Shang and Sen Yang prepared the material for sequencing. Li Song, Degang Zhao, Dong Zhang, and Wei Sun revising the manuscript, Degang Zhao and Qi Shen designed the experiments and revised the manuscript.

**Acknowledgments:** This work was funded by Science & Technology offers and academy of agricultural science of Guizhou province (Grant No. LH[2015]7062) and National Science Foundation of China (Grant No. 31360067), Science-technology Support Projects of Guizhou province (Grant No. NY[2016]3052) and also thanks to the Shilin Chen and Yujun Zhang of the Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences for helpful discussion and language polishing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Peiretti, P.G. Fatty Acid Content and Chemical Composition of Vegetative Parts of Perilla (*Perilla frutescens* L.) after Different Growth Lengths. *Res. J. Med. Plants* **2011**, *5*, 72–78. [CrossRef]
- Sa, K.J.; Choi, S.H.; Ueno, M.; Park, K.C.; Park, Y.J.; Ma, K.H.; Lee, J.K. Identification of genetic variations of cultivated and weedy types of perilla, species in korea and japan using morphological and ssr markers. *Genes Genom.* **2013**, *35*, 649–659. [CrossRef]
- Asif, M. Health effects of omega-3,6,9 fatty acids: *Perilla frutescens*, is a good example of plant oils. *Orient. Pharm. Exp. Med.* **2011**, *11*, 51–59. [CrossRef] [PubMed]
- Akond, M.; Liu, S.; Boney, M.; Kantartzi, S.K.; Meksem, K.; Bellaloui, N.; Lightfoot, D.A.; Kassen, M.A. Identification of Quantitative Trait Loci (QTL) Underlying Protein, Oil, and Five Major Fatty Acids' Contents in Soybean. *Am. J. Plant Sci.* **2014**, *5*, 158–167. [CrossRef]
- Ramos, M.J.; Fernández, C.M.; Casas, A.; Rodríguez, L.; Pérez, A. Influence of fatty acid composition of raw materials on biodiesel properties. *Bioresour. Technol.* **2009**, *100*, 261. [CrossRef] [PubMed]
- Connor, W.E. Importance of n-3 fatty acids in health and disease. *Am. J. Clin. Nutr.* **2000**, *71*, 171S. [PubMed]
- Schuchardt, J.P.; Huss, M.; Stauss-Grabo, M.; Hahn, A. Significance of long-chain polyunsaturated fatty acids (PUFAs) for the development and behaviour of children. *Eur. J. Pediatr.* **2010**, *169*, 149–164. [CrossRef] [PubMed]
- Janssen, C.I.; Kiliaan, A.J. Long-chain polyunsaturated fatty acids (LCPUFA) from genesis to senescence: The influence of LCPUFA on neural development, aging, and neurodegeneration. *Prog. Lipid Res.* **2014**, *53*, 1–17. [CrossRef] [PubMed]
- Trapnell, C.; Hendrickson, D.G.; Sauvageau, M.; Goff, L.; Rinn, J.L.; Pachter, L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **2013**, *31*, 46. [CrossRef] [PubMed]
- Pradhan, S.; Bandhiwal, N.; Shah, N.; Kant, C.; Gaur, R.; Bhatia, S. Global Transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Front. Plant Sci.* **2014**, *5*, 698. [CrossRef] [PubMed]
- Severin, A.J.; Woody, J.L.; Bolon, Y.; Joseph, B.; Diers, B.W.; Farmer, A.D.; Muehlbauer, G.J.; Nelson, R.T.; Grant, D.; Specht, E.J.; et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biol.* **2010**, *10*, 160. [CrossRef] [PubMed]
- Jang, Y.E.; Kim, M.Y.; Shim, S.; Li, J.; Lee, S. Gene expression profiling for seed protein and oil synthesis during early seed development in soybean. *Genes Genom.* **2015**, *37*, 409–418. [CrossRef]
- Yin, D.; Wang, Y.; Zhang, X.; Li, H.; Lu, X.; Zhang, J.; Zhang, W.; Chen, S. De novo assembly of the peanut (*Arachis hypogaea* L.) seed transcriptome revealed candidate unigenes for oil accumulation pathways. *PLoS ONE* **2013**, *8*, e73767. [CrossRef] [PubMed]
- Gupta, K.; Kayam, G.; Faigenboimdoron, A.; Clevenger, J.; Oziasakins, P.; Hovav, R. Gene expression profiling during seed-filling process in peanut with emphasis on oil biosynthesis networks. *Plant Sci.* **2016**, *248*, 116–127. [CrossRef] [PubMed]
- Dussert, S.; Guerin, C.; Andersson, M.; Joët, T.; Tranbarger, T.J.; Pizot, M.; Sarah, G.; Omoro, A.; Durand-Gasselín, T.; Morcillo, F. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.* **2013**, *162*, 1337–1358. [CrossRef] [PubMed]

16. Venglat, P.; Xiang, D.; Qiu, S.; Stone, S.L.; Tibiche, C.; Cram, D.; Alting-mees, M.; Nowak, J.; Cloutier, S.; Deyholos, M.; et al. Gene expression analysis of flax seed development. *BMC Plant Biol.* **2011**, *11*, 74. [[CrossRef](#)] [[PubMed](#)]
17. Fatima, T.; Snyder, C.L.; Schroeder, W.; Cram, D.; Datla, R.; Wishart, D.S.; Weselake, R.; Krishna, P. Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. *PLoS ONE* **2012**, *7*, e34099. [[CrossRef](#)] [[PubMed](#)]
18. Liang, C.; Liu, X.; Yiu, S.; Lim, B.L. De novo assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. *BMC Genom.* **2013**, *14*, 146. [[CrossRef](#)] [[PubMed](#)]
19. Wang, X.; Xu, R.; Wang, R.; Liu, A. Transcriptome analysis of Sacha Inchi (*Plukenetia volubilis* L.) seeds at two developmental stages. *BMC Genom.* **2012**, *13*, 716. [[CrossRef](#)] [[PubMed](#)]
20. Li, S.; Wang, L.; Shu, Q.; Wu, J.; Chen, L.; Shao, S.; Yin, D. Fatty acid composition of developing tree peony (*Paeonia section Moutan* DC.) seeds and transcriptome analysis during seed development. *BMC Genom.* **2015**, *16*, 208. [[CrossRef](#)] [[PubMed](#)]
21. Kim, H.U.; Lee, K.; Shim, D.; Lee, J.H.; Chen, G.Q.; Hwang, S. Transcriptome analysis and identification of genes associated with  $\omega$ -3 fatty acid biosynthesis in *Perilla frutescens* (L.) var. *frutescens*. *BMC Genom.* **2016**, *17*, 474. [[CrossRef](#)] [[PubMed](#)]
22. Yonekura-Sakakibara, K.; Saito, K. Transcriptome Coexpression Analysis Using ATTED-II for Integrated Transcriptomic/Metabolomic Analysis. *Methods Mol. Biol.* **2013**, *1011*, 317–326. [[CrossRef](#)] [[PubMed](#)]
23. Willsey, A.J.; Sanders, S.J.; Li, M.; Dong, S.; Tebbenkamp, A.T.; Muhle, R.A.; Reilly, S.K.; Lin, L.; Fertuzinhos, S.; Miller, J.A.; et al. Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* **2013**, *155*, 997–1007. [[CrossRef](#)] [[PubMed](#)]
24. Banerjee, N.; Chothani, S.P.; Harris, L.; Dimitrova, N. Identifying RNAseq-based coding-noncoding co-expression interactions in breast cancer. In Proceedings of the 2013 IEEE International Workshop on Genomic Signal Processing and Statistics, Houston, TX, USA, 17–19 November 2013; pp. 11–14. [[CrossRef](#)]
25. Du, J.; Wang, S.; He, C.; Zhou, B.; Ruan, Y.L.; Shou, H. Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* **2017**, *68*, 1955. [[CrossRef](#)] [[PubMed](#)]
26. Song, X.; Liu, G.; Huang, Z.; Duan, W.; Tan, H.; Li, Y.; Hou, X. Temperature expression patterns of genes and their coexpression with lncRNAs revealed by RNA-Seq in non-heading Chinese cabbage. *BMC Genom.* **2016**, *17*, 297. [[CrossRef](#)] [[PubMed](#)]
27. Iorizzo, M.; Ellison, S.; Senalik, D.; Zeng, P.; Satapoomin, P.; Huang, J.; Bowman, M.J.; Lovene, M.; Sanseverion, W.; Cavagnaro, P.F.; et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **2016**, *48*, 657. [[CrossRef](#)] [[PubMed](#)]
28. Yang, Z.; Jiang, Y.; Ma, C.; Silvestri, G.; Bosinger, S.E.; Li, B.; Jong, A.; Zhou, Y.; Huang, S. Coexpression Network Analysis of Benign and Malignant Phenotypes of SIV-Infected Sooty Mangabey and Rhesus Macaque. *PLoS ONE* **2016**, *11*, e0156170. [[CrossRef](#)] [[PubMed](#)]
29. Chen, F.; Zhu, H.; Zhou, L.; Li, J.; Zhao, L.; Wu, S.; Wang, J.; Liu, W.; Chen, Z. Genes related to the very early stage of ConA-induced fulminant hepatitis: A gene-chip-based study in a mouse model. *BMC Genom.* **2010**, *11*, 240. [[CrossRef](#)] [[PubMed](#)]
30. Filteau, M.; Pavey, S.A.; St Cyr, J.; Bernatchez, L. Gene Coexpression Networks Reveal Key Drivers of Phenotypic Divergence in Lake Whitefish. *Mol. Biol. Evol.* **2013**, *30*, 1384–1396. [[CrossRef](#)] [[PubMed](#)]
31. Hirai, M.Y.; Sugiyama, K.; Sawada, Y.; Tohge, T.; Obayashi, T.; Suzuki, A.; Araki, R.; Sakurai, N.; Suzuki, H.; Aoki, K.; et al. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6478. [[CrossRef](#)] [[PubMed](#)]
32. Yonekura-Sakakibara, K.; Tohge, T.; Matsuda, F.; Nakabayashi, R.; Takayama, H.; Niida, R.; Watanabe-Takahashi, A.; Inoue, E.; Saito, K. Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* **2008**, *20*, 2160–2176. [[CrossRef](#)] [[PubMed](#)]
33. Yonekura-Sakakibara, K.; Tohge, T.; Niida, R.; Saito, K. Identification of a flavonol 7-Orhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.* **2007**, *282*, 14932–14941. [[CrossRef](#)] [[PubMed](#)]

34. Albinsky, D.; Sawada, Y.; Kuwahara, A.; Nagano, M.; Hirai, A.; Saito, K.; Hirai, M.Y. Widely targeted metabolomics and coexpression analysis as tools to identify genes involved in the side-chain elongation steps of aliphatic glucosinolate biosynthesis. *Amino Acids* **2010**, *39*, 1067–1075. [[CrossRef](#)] [[PubMed](#)]
35. Chen, E.C.; Tai, S.S.; Peng, C.C.; Tzen, J.T. Identification of Three Novel Unique Proteins in Seed Oil Bodies of Sesame. *Plant Cell Physiol.* **1998**, *39*, 935. [[CrossRef](#)] [[PubMed](#)]
36. Kim, H.U.; Hsieh, K.; Ratnayake, C.; Huang, A.H. A novel group of oleosins is present inside the pollen of Arabidopsis. *J. Biol. Chem.* **2002**, *277*, 22677. [[CrossRef](#)] [[PubMed](#)]
37. Huang, A.H. Oleosins and oil bodies in seeds and other organs. *Plant Physiol.* **1996**, *110*, 1055–1061. [[CrossRef](#)] [[PubMed](#)]
38. Vindigni, J.D.; Wien, F.; Giuliani, A.; Erpapazoglou, Z.; Tache, R.; Jaqic, F.; Chardot, T.; Gohon, Y.; Forissard, M. Fold of an oleosin targeted to cellular oil bodies. *Biochim. Biophys. Acta* **2013**, *1828*, 1881–1888. [[CrossRef](#)] [[PubMed](#)]
39. Le, B.H.; Cheng, C.; Bui, A.Q.; Wagmaister, J.A.; Henry, K.F.; Pelletier, J.; Kwong, L.; Belmonte, M.; Kirkbride, R.; Horvath, S.; et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8063–8070. [[CrossRef](#)] [[PubMed](#)]
40. Peng, F.Y.; Weselake, R.J. Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in arabidopsis. *BMC Genom.* **2011**, *12*, 286. [[CrossRef](#)] [[PubMed](#)]
41. Huang, Z.; Zhong, X.J.; He, J.; Jiang, M.Y.; Yu, X.F.; Li, X. Identification and characterization of AP2/ERF transcription factors in moso bamboo (*Phyllostachys edulis*). *Mol. Biol.* **2016**, *50*, 785. [[CrossRef](#)]
42. An, D.H.; Michung, S. Overexpression of Arabidopsis WR1 enhanced seed mass and storage oil content in Camelina sativa. *Plant Biotechnol. Rep.* **2015**, *9*, 137–148. [[CrossRef](#)]
43. Baud, S.; Wuillème, S.; To, A.; Rochat, C.; Lepiniec, L. Role of wrinkled1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in arabidopsis. *Plant J.* **2009**, *60*, 933. [[CrossRef](#)] [[PubMed](#)]
44. Kim, M.J.; Kim, J.K.; Shin, J.S.; Suh, M.C. The SebHLH transcription factor mediates trans-activation of the SeFAD2 gene promoter through binding to E- and G-box elements. *Plant Mol. Biol.* **2007**, *64*, 453–466. [[CrossRef](#)] [[PubMed](#)]
45. Sun, H.; Fan, H.J.; Ling, H.Q. Genome-wide identification and characterization of the bHLH, gene family in tomato. *Front. Plant Sci.* **2015**, *16*, 9. [[CrossRef](#)] [[PubMed](#)]
46. Park, S.J. Analysis and expression of the cotton gene for the  $\Delta$ -12 fatty acid desaturase 2-4 (FAD2-4). Ph.D. Thesis, University of North Texas, Denton, TX, USA, August 2003.
47. Makkena, S.; Labm, R.S. The bhlh transcription factor spatula is a key regulator of organ size in arabidopsis thaliana. *Plant Signal. Behav.* **2013**, *8*, e24140. [[CrossRef](#)] [[PubMed](#)]
48. Duan, P.; Ni, S.; Wang, J.; Zhang, B.; Xu, R.; Wang, Y.; Chen, H.; Zhu, X.; Li, Y. Regulation of osgrf4 by osmir396 controls grain size and yield in rice. *Nat. Plants* **2015**, *2*, 15203. [[CrossRef](#)] [[PubMed](#)]
49. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2013**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
50. Conesa, A.; Götz, S.; Garcíagómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674. [[CrossRef](#)] [[PubMed](#)]
51. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **2012**, *16*, 284. [[CrossRef](#)] [[PubMed](#)]
52. Ariani, A.; Gepts, P. Genome-wide identification and characterization of aquaporin gene family in common bean (*Phaseolus vulgaris* L.). *Mol. Genet. Genom.* **2015**, *290*, 1771. [[CrossRef](#)] [[PubMed](#)]
53. Robinson, M. D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
54. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L.P. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–W322. [[CrossRef](#)] [[PubMed](#)]
55. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [[CrossRef](#)] [[PubMed](#)]

56. Smoot, M.; Ono, K.; Ideker, T.; Maere, S. PiNGO: A Cytoscape plugin to find candidate genes in biological networks. *Bioinformatics* **2011**, *27*, 1030–1031. [[CrossRef](#)] [[PubMed](#)]
57. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).