Quantitative proteomic and lipidomics analyses of high oil content GmDGAT1-2 transgenic soybean, illustrates the regulatory mechanism of lipoxygenase and oleosin

Yang Xu（✉ 317337335@qq.com ）

    Jilin University

Fan Yan

    Jilin University

Yajing Liu

    Jilin University

Ying Wang

    Jilin University

Han Gao

    Jilin University

Shihui Zhao

    Jilin University

Qingyu Wang

    Jilin University

Jingwen Li

    Jilin University

Abstract

The main goal of soybean breeding is to increase the oil content. Diacylglycerol acyltransferase (DGAT) is a key rate-limiting enzyme in fatty acid metabolism, and may regulate oil content. Herein, 10 *GmDGAT* genes were isolated from soybean and transferred into wild-type (WT) *Arabidopsis*. The total fatty acid was 1.2 times higher in T_3 *GmDGAT1-2* transgenic *Arabidopsis* seeds than in WT. Therefore, *GmDGAT1-2* was transferred into WT soybean (JACK), and four T_3 transgenic soybean lines were obtained. The results of High performance gas chromatography and Soxhlet extractor showed that, compared with those of JACK, oleic acid (18:1), and total fatty acid levels in transgenic soybean plants were much higher, but linoleic acid (18:2) was lower than WT. Palmitic acid (16:0), stearic acid (18:0), and linolenic acid (18:3) were not significantly different. For mechanistic studies, 436 differentially expressed proteins (DEPs) and 180 differentially expressed metabolites (DEMs) were identified between WT (JACK) and transgenic soybean pods using proteomic and lipidomics analyses. Four lipoxygenase proteins were down-regulated in linoleic acid metabolism while four oleosin proteins were up-regulated in the final oil formation. The results showed an increase in the total fatty acid and 18:1 composition, and a decrease in the 18:2 composition of fatty acid. Our study brings new insights into soybean genetic transformation and the deep study of molecular mechanism that changes the total fatty acid, 18:1, and 18:2 compositions in *GmDGAT1-2* transgenic soybean.

Key Message

Proteomic and lipidomics analyses of WT and *GmDGAT1-2* transgenic soybeans showed that *GmDGAT1-2* over-expression induced *lipoxygenase* down-regulatation and *oleoin* up-regulatation, which significantly changed the compositions and total fatty acid.

1. Introduction

Soybean (*Glycine max*) is an important oil and protein crop worldwide. It is one of the main sources of vegetable oil for humans and animals, and an important bioenergy resource for sustainable development. Soybean oil plays an important role in the growth, development, and reproduction of soybean. The genetic variation that determines soybean seed oil content is small compared with that which determines protein content; thus, it is difficult to enhance oil content genetically. The biosynthesis of soybean fat includes the synthesis, termination, release, and desaturation of the fatty acid chain, and synthesis of triglycerides (TAGs), including the formation of polyunsaturated fatty acids and liposomes. Synthesized soybean fat is stored in two ways: 1) as glycerin and phospholipid used to form biofilms, and 2) in seeds (often in the form of TAG) (Mekhedov et al. 2000). In plants, TAG synthesis is vital for the formation of seed oil (Barthole et al. 2012).

Vegetable oil synthesis is complex and involves several enzymes. Changes in the expression or regulations of genes that encode these enzymes may affect lipid synthesis. Fatty acid is an important component of oil, which is mainly synthesized in plastids. FAT enzymes are divided into two categories: A and B. FATA catalyzes 18:1-ACP, while FATB catalyzes 16:0-ACP and 18:0-ACP (Saha et al. 2006). SAD catalyzes 18:0-ACP to 18:1-ACP and KASII catalyzes 16:0-ACP to 18:0-ACP. Under catalysis by FAD2 and FAD3, the 18:1 fatty acid is converted to 18:2 and 18:3, respectively (Okuley et al. 1994 ; Reed et al. 2000). On one hand, this complex enzymes and synthesis make it difficult to study the genetic mechanism of vegetable oil concentrations and composition, but on the other hand, it provides an opportunity to improve vegetable oil content and fatty acid composition (Liu et al. 2013). Recent studies have focused on identifying and verifying an increasing number of genes related to lipid metabolism (Barthole et al. 2012). Genetic engineering, overexpression, or interference in the synthesis of key enzyme-encoding genes can stimulate or inhibit the corresponding enzyme activity, thus improving the quality of vegetable oils.

DGAT (diacylglycerol acyltransferase) is the only rate-limiting enzyme responsible for transferring an acyl group from acyl-CoA at the sn-3 position of sn-1,-2-diacylglycerol to form triacylglycerol (TAG), the final product in the formation of storage lipids in the Kennedy pathway. To date, four types of DGATs have been identified in plants (Yen et al. 2008). *DGAT* subfamilies include *DGAT1*, *DGAT2*, *DGAT3*, and WS/DGAT. The *DGAT1* subfamily exists in both animals and plants. Cases et al. (1998) cloned the first *DGAT1* gene in mice, and then Hobbs et al. (1999) cloned it in *Arabidopsis*. Subsequently, the *DGAT1* gene was successfully cloned in plants such as nasturtium (Xu et al. 2008), castor (He et al. 2004), maize (Zheng et al. 2008), tobacco (Bouvier-Nave et al. 2000), and rape (Nykiforuk et al. 1999). The *DGAT2* gene was cloned in *Mortierella ramanniana* (Lardizabal et al. 2001), then *Arabidopsis* (Salanoubat et al. 2000), castor (Kroon et al. 2006), and tung tree (Shockey et al. 2006). Members of the *DGAT3* subfamily have

been identified only in peanuts (Saha et al. 2006) and *Arabidopsis* (Hernandez et al. 2012), and their protein sequences have low homology with those of the other *DGAT* subfamilies. *WS/DGAT* is a member discovered in recent years and cloned from *Acinetobacter calcoaceticus*. This gene is mainly involved in the synthesis of wax ester and some TAG in *Arabidopsis* (Li et al. 2008).

Ten *DGAT* genes in soybean belong to the *DGAT1*, *DGAT2*, or *DGAT3* subfamilies and are distributed across six chromosomes. *DGAT1* can be expressed in many tissues but is mainly found in germinated and mature seeds. Members of the *DGAT2* subfamily are found in nodules, leaves, flower green pods, and mature seeds. Members of *DGAT3* are expressed in roots, leaves, and seeds (Liu et al. 2013). TAG in plants is mainly synthesized and accumulated in the oil bodies. Most studies have shown that DGAT is located in oil bodies or the endoplasmic reticulum. Kamisaka et al. found that the main site of DGAT activity in the subcells of oleaginous fungus is the oil body (Kamisaka et al. 1997). Using confocal and transmittance electron microscopies, Lardizabal et al (2008) transformed soybean seed cells to express *UrDGAT2A* and found that UrDGAT2A is mainly located on the endoplasmic reticulum and oil body membrane (Lardizabal et al. 2008). This demonstrates that *DGAT* expression is closely related to changes in oil content. Chen et al. found that *GmDGAT2D* transgenic hairy roots tend to synthesize 18:1 or 18:2 TAG, whereas those with *GmDGAT1A* prefer 18:3 acyl-CoA for TAG synthesis. Overexpression of *GmDGAT2D* promotes 18:2 TAG in wild-type (WT) seeds but increases 18:1 TAG production and decreases 18:3 TAG production in *rod1* mutant seeds (Chen et al. 2016). However, to our knowledge, there have been no studies reporting the stability of *DGAT* gene inheritance in soybean lines or verification of their further functions.

High-throughput molecular biological techniques, such as transcriptomic, proteomic, and metabolomic approaches, are widely used to investigate various molecular mechanisms in plants (Guo, R et al. 2017 ; Padula, et al. 2017). Tandem Mass Tags (TMT) is a relative and absolute quantitative technique for in vitro isotopic tagging (Ross, PL et al. 2004).This technique uses a various isotope reagents to label the N-terminal or lysine side chain groups of protein polypeptides, and analyzes them in tandem using a high precision mass spectrometer to quantify the high-throughput screening techniques commonly used in proteomics. Lipidomics is a high-throughput analysis technique based on LC-MS that systematically analyzes the patterns of lipid compositions and expressions in organisms. Liposome analysis can efficiently assess the changes and functions of lipid compounds in various biological processes, and elucidate the associated underlaying mechanisms (Tang HY, et al. 2018). Several studies have demonstrated an improved oil composition of soybean seed by silencing genes that are responsible for encoding fatty acids (Nazrul Islam, et al. 2019). Similarly, other studies have focused on improving rice grain protein quality by inserting a special sunflower gene encoding a novel protein using a proteomic analysis.

In this study, we isolated 10 *DGAT* family genes from soybean and over expressed *GmDGAT1-2* in transgenic soybean to investigate its role in regulating total fatty acids and their compositions. Furthermore, using proteomic and lipidomics analyses, eight key differentially expressed proteins (DEPs), namely four lipoxygenases and four oleosin proteins, were screened to provide a basis for further understanding of its function in soybean fatty acid synthesis.

2. Results

2.1. Isolation and characterization of GmDGAT family genes

Ten *GmDGAT* family genes were obtained from the Soybase database according to the amino acid sequences of AtDGAT1, AtDGAT2, and AhDGAT3. There were three genes homologous to AtDGAT1 (Glyma09g07520, Glyma17g06120, and Glyma13g16560), five homologous to AtDGAT2 (Glyma11g09411, Glyma01g36011, Glyma16g21970, Glyma09g32790 and Glyma16g21960), and two homologous to AhDGAT3 (Glyma13g17860, and Glyma17g04650). The alignment of cDNA with the sequence in the soybean GmGDB database demonstrated that all genes in the GmDGAT1 subfamily contained an MBOAT domain, except for Glyma09g07520. DGAT1 is usually predicted to possess six or more transmembrane domains and belongs to the membrane-bound O-acyltransferase (MBOAT) family (He et al. 2004). The GmDGAT2 subfamily contains a DAGAT domain. DGAT2 proteins possess only two to three predicted transmembrane domains and belong to the monoacylglycerol acyltransferase family. The GmDGAT3 subfamily contains a domain whose function is unknown (Fig. 1). Compared with Williams82, there were some mutations, deletions, and increases in amino acids in the cloning of *GmDGATs* (except in *GmDGAT2-1, GmDGAT2-3, GmDGAT3-1,* and *GmDGAT3-2*) in JD5 owing to the differences between soybean varieties.

2.2. Measurement of oil content in transgenic Arabidopsis

All *GmDGAT* genes were transferred into *Arabidopsis* Col-0 to verify the function of GmDGATs in seed oil production. T_3 transgenic lines in the 10 *GmDGAT* family genes were used to measure oil content. Mature *Arabidopsis* seeds (0.02–0.05 g) of T_3 plants from more than 10 independent lines were sampled with three biological replicates, including the WT and transgenic *Arabidopsis* that were used for oil content analysis. High performance gas chromatography revealed peaks and peak time that correlated with palmitic acid (16:0, 11.672 min), stearic acid (18:0, 16.011 min), oleic acid (18:1, 16.471 min), linoleic acid (18:2, 17.779 min), and linolenic acid (18:3, 20.345 min).

The results showed that the 18:1 content was significantly higher in the overexpressing transgenic *Arabidopsis* seeds than in the WT seeds, especially in *GmDGAT1-2* whose contents were 1.5 times higher, respectively, than in WT, (Fig. 2a). However, the contents of 16:0, 18:0 and 18:2 were slightly lower in the transgenic lines than in the WT. Nevertheless, the total TAG content in these *GmDGAT* transgenic lines was greatly enhanced to different degrees. In all the transgenic lines overexpressing *GmDGAT* subfamilies, the variation trends in the five components of total oil content were approximately the same (Fig. 2).

2.3. Generation and detection of GmDGAT1-2 transgenic soybean

The four T_3 transgenic soybean lines were confirmed using bar strip, southern blot, and western blot. The results showed that *GmDGAT1-2* was transformed into the WT (JACK) cultivar (Fig. 3), while ELISA analysis indicated that the expression was higher in transgenic soybeans than in JACK (Table. 1).

The 1497bp *GmDGAT1-2* open reading frame (ORF) was cloned into the *Xba* Ⅰ/ *Sac* Ⅰ restriction site of the vector pTF101-35s in the sense direction (Fig. 3c). The bar gene and *GmDGAT1-2* target gene were used as probes in a southern blot to evaluate the copy number in the T_3 generation of transgenic soybean strains. To eliminate interference from endogenous genes, the gene base sequence containing the mutation was selected for the target gene probe. There were approximately 9 kb between the bar gene probe and left boundary of T-DNA, approximately 2 kb between the bar gene probe and right boundary, approximately 9.6 kb between the target gene and left boundary of T-DNA and approximately 1 kb between the target gene and right boundary (Fig. 3c). Following *Hind* III digestion, the enzyme site was approximately 8.7 kb from the left boundary and 1.7 kb from the right boundary of T-DNA. The size of the hybridization fragment was consistent with the expected results, which showed that OE-1, OE-2, and OE-3 transgenic strains were detected with only one hybridization band, OE-4 transgenic strains with two bands, and the WT with no bands. This suggests an exogenous T-DNA in the form of single-copy DNA integrated into the soybean genome in OE-1, OE-2, and OE-3, and exogenous DNA in double-copy form in OE-4 strains (Fig. 3b and d). However, the specific insertion sites warrant further investigations using re-sequencing the genome of transgenic soybean.

DGAT1 was predicted to possess six or more transmembrane domains and a high protein molecular weight. Specific antibodies against DGAT1 remain difficult to synthesize unless the corresponding structure is knocked out or several polypeptide chains are spliced and synthesized; however, this method is not guaranteed to produce effective antibodies. Therefore, we translated the herbicide bar gene on the pTF101 vector into the phosphinothricin-N-acetyltransferase (PAT) protein and synthesized monoclonal antibodies to detect positive transgenic plants. Western blot results showed an apparent hybridization signal at 21 KD, suggesting the successfully expression of the bar gene at the protein level in transgenic plants. We further speculated that the GmDGAT1-2 protein was also successfully expressed in four transgenic lines at the qualitative level (Fig. 3e).

Roots, stems, leaves, flowers, and different stages of pods were collected from the WT (JACK) and transgenic soybeans, then analyzed using qRT-PCR to verify gene over-expression in the transgenic lines. The results showed *GmDGAT1-2* expression in the leaf, 10- and 40- day-old pods was significantly higher in transgenic soybeans than in WT ($p < 0.01$). *GmDGAT1-2* expression in the 20- and 30- day-old pods was significantly higher in transgenic than in WT soybeans ($p < 0.05$). However, there was no significant difference in *GmDGAT1-2* expression in the roots (Fig. 3f).

ELISA results showed that PAT protein expression in negative control WT soybeans was very low, such that its value was below the effective detection range of the standard curve. PAT protein was expressed in all four transgenic lines, while its expression level in the vegetative growth stage was lower than in the reproductive growth stage. PAT expression was highest in mature soybean seeds, approximately 10 times higher than in nutritious organs (root, stem, leaf, and flower). Since seeds are important storage sites for

soybean oil, this result validates the oil content determination results, and indicates that GmDGAT1-2 proteins are expressed in four transgenic soybean lines at the quantitative level (Table 1).

Table 1

PAT protein content of different soybean tissues in wild type and four different transgenic lines (OE-1 to OE-4). Data represent the average values of three biological replicates. NA indicates that the value is lower than the effective detection range of the standard curve.

Material	JACK	OE-1 (ng/g)	OE-2 (ng/g)	OE-3 (ng/g)	OE-4 (ng/g)
Root	NA	235.15 ± 8.52	244.12 ± 9.22	204.91 ± 7.45	185.34 ± 8.75
Stem	NA	155.27 ± 11.62	160.81 ± 18.2	138.62 ± 12.9	100.85 ± 2.39
Leave	NA	623.7 ± 2.14	652.41 ± 1.05	614.2 ± 2.88	599.32 ± 3.18
Flower	NA	100.5 ± 2.58	115.31 ± 8.41	104.98 ± 3.42	95.36 ± 10.4
10-day Pod	NA	852.3 ± 10.51	864.36 ± 8.94	835.28 ± 1.06	789.63 ± 4.65
20-day Pod	NA	612.3 ± 7.81	702.51 ± 9.98	601.2 ± 12.8	586.3 ± 7.14
30-day Pod	NA	796.3 ± 10.94	782.5 ± 5.88	833.69 ± 20.98	725.36 ± 15.23
40-day Pod	NA	906.3 ± 5.88	988.14 ± 19.25	900.17 ± 20.31	872.69 ± 15.61
Seed	NA	1152.3 ± 13.9	1281.5 ± 16.9	1096.9 ± 23.1	893.9 ± 13.7

Table 2

The result of differentially expressed proteins (DEPs) between WT (JACK) and transgenic soybean

Compared Samples	Num. of Total Quant.	regulated Type	fold-change > 1.2
GmDGAT1_2.vs.JACK	6876	up-regulated	243
		down-regulated	193

Table 3

The result of differentially expressed metabolites (DEMs) between WT (JACK) and transgenic soybean

Compared Samples	Num. of Total Ident.	Num. of Total Sig.	Num. of Sig.Up	Num. of Sig.down
GmDGAT1_2.vs.JACK_pos	527	107	12	95
GmDGAT1_2.vs.JACK_neg	249	73	39	34

2.4. GmDGAT1-2 overexpression enhances oil content in transgenic soybean

Three biological replicates of mature soybean seeds (0.01 g), including WT and *GmDGAT1-2* transgenic seeds used for oil content analysis, were sampled. High performance gas chromatography showed no significant difference in palmitic acid (16:0), stearic acid (18:0), or linolenic acid (18:3) contents between transgenic soybeans and the WT (JACK). However, significant difference was observed for oleic acid (18:1), with the transgenic line having 29% higher content, compared with that in WT. The levels of 18:2 were significantly lower in the four transgenic lines than in the WT. The total fatty acid in T_3 *GmDGAT*1-2 transgenic seeds was 3.1 % higher than that in WT seeds (Fig. 4a and b).

2.5. Proteomic analyses between WT (control) and 40- day-old GmDGAT1-2 transgenic soybean pods

The 40-day old pods of *GmDGAT1-2* transgenic soybean (OE-3) were assessed by TMT quantitative proteomics. Results showed that 35912 unique peptides corresponding to 6876 proteins were identified via LC-MS/MS determination; the data were searched in the UniProt database. In the analyses, up-regulated proteins were selected with an adjusted fold change ≥ 1.2 and a P-value of ≤

0.05, while the down-regulated proteins were selected with an adjusted fold change ≤ 0.83 and a P-value of ≤ 0.05 as the threshold to screen significant difference in the abundance of differentially expressed proteins (DEPs) between WT (JACK) and transgenic soybean. Finally, we identified 243 up-regulated DEPs and 193 down-regulated DEPs (Table. 2, Fig. 5a and. b).

To further investigate the functional features of these DEPs, Gene Ontology (GO) term enrichment analysis was conducted using the GO and UniProt databases. The 436 DEPs were classified into three categories, namely biological process (BP), cellular component (CC), and molecular function (MF) (Fig. 5c). In the BP category, the most abundant groups included peptide metabolic process, cellular nitrogen compound biosynthetic process, and gene expression. The CC category of DEPs was significantly enriched for the intracellular non-membrane-bounded organelle, membrane part, and intracellular ribonucleoprotein complex. For the MF category, structural molecule activity, structural constituent of ribosome, and enzyme inhibitor activity were the most abundant groups.

Furthermore, the proteomic results via KEGG analysis showed that 44 pathways were enriched. These DEPs were involved primarily in the following top 20 metabolic pathways: Base excision repair; Fatty acid elongation; Isoflavonoid biosynthesis; ABC transporters; Linoleic acid metabolism; Cutin, suberine and wax biosynthesis; Photosynthesis; Porphyrin and chlorophyll metabolism; Photosynthesis – antenna proteins; One carbon pool by folate; Protein export; Thiamine metabolism; Ribosome; Galactose metabolism; Glyoxylate and dicarboxylate metabolism; Pentose and glucuronate interconversions; Carbon fixation in photosynthetic organisms; Fatty acid metabolism; Amino sugar and nucleotide sugar metabolism; and Carbon metabolism (Fig. 5d). Based on these results, we focused on the DEPs in the linoleic acid metabolic pathway for further study.

2.6. Expression analysis of four key DEPs involved in the linoleic acid metabolic pathway in the seeds of WT and GmDGAT1-2 transgenic soybean

To verify our results, 4 key DEPs (13s-lipoxygenase-1, 9s-lipoxygenase-1, 9s-lipoxygenase-4, and 9s-lipoxygenase-5) involved in the linoleic acid metabolic were selected for qRT-PCR between WT (control) and *GmDGAT1-2* transgenic soybean seeds (OE-3). These four differentially expressed genes belonged to the lipoxygenase family according to the GO analyses results. The expression of the four genes showed similar trends between the qRT-PCR results and the TMT analysis. 9s-lipoxygenase-1, 9s-lipoxygenase-4, and 9s-lipoxygenase-5 analyzed using qRT-PCR were down-regulated in linoleic acid metabolic. However, 13s-lipoxygenase-1 analyzed using qRT-PCR showed that the expression was not significantly higher than WT (Fig. 6a). These results suggest that the TMT analysis was highly reliable.

Furthermore, another four DEPs (P24 oleosin isoform B, P24 oleosin isoform A, and two oleosin-5) were screened, oleosin protein family genes are involved in the later stages of the fatty acid synthesis, and influence the total oil content. These four oleosin proteins were up-regulated as illustrated by the proteomics analysis. The results of oleosin proteins analyzed using qRT-PCR showed that the expressions were all higher than WT (Fig. 6b). The result was consistent with proteomics analysis.

2.7. Lipidomics analyses between WT (control) and 40- day-old GmDGAT1-2 transgenic soybean pods

In our study, 776 types of lipid metabolites (527 positive ions and 249 negative ions) were identified via LC−MS/MS, and the data were searched against the Lipidmaps and Lipiblast spectrogram database. In total, 180 differentially expressed metabolites (DEMs) including 107 positive ions which 12 metabolites were up-regulated and 95 were down-regulated, and 73 negative ions, which 39 metabolites were up-regulated and 34 were down-regulated (Table. 3).

There were more than 20 kinds of DEMs phosphatidyl ethanolamines (PE), phosphatidic acids (PA), and phosphatidyl glycerols (PG) DEMs that passed the criteria of an OPLS-DA model VIP > 1.0, FC > 1.5, or had a P-value of < 0.05 and FC < 0.667 (Heischmann S, et al. 2016 ; Haspel JA, et al. 2014) between WT (control) and 40- day-old *GmDGAT1-2* transgenic soybean pods were selected (Fig. 7a and b).

To select marker metabolites accurately and investigate changes in the related metabolic processes, the selected DEMs were analyzed using hierarchical clustering based on the expression of significantly different metabolites in each sample groups (Fig. 7c and d).

2.8. Integrated proteomic and lipidomics analyses of 40- day-old pods in GmDGAT1-2 transgenic soybean

By conducting a KEGG mapping analysis based on the differences in proteins and metabolites before, we focused on the integrated analysis of four DEPs of one screened family and top DEMs. The four DEPs include four lipoxygenase those are 13s-lipoxygenase-1 (B3TDK4), 9s-lipoxygenase-1 (A7LCD5), 9s-lipoxygenase-4 (B3TDK9), and 9s-lipoxygenase-5 (I1LBB9). The top forty DEMs include different components of total oil content, such as Com-119-pos,Com-48-pos⬜Com-2231-pos represented the 16:0, 17:0, 18:0, 18:1, 18:2, 18:3 and 20:0. Code names and corresponding lipid metabolites are shown in Table 4.

Table 4
The code names and corresponding lipid metabolites of all DEMs.

Code names (pos and neg)	Lipid metabolites names
Com-119-pos	18:0, 18:1
Com-48-pos	16:0, 18:1
Com-2231-pos	17:0, 18:1
Com-4363-pos	20:0, 18:1
Com-195-pos	18:2
Com-119-pos	18:0, 18:1
Com-2231-pos	17:0, 18:0
Com-536-pos	18:1, 18:2
Com-31-pos	18:1, 18:2
Com-4497-pos	18:1, 20:1
Com-48-pos	16:0, 18:1
Com-6987-pos	18:3, 20:0
Com-8064-pos	16:0
Com-4363-pos	20:0, 18:1
Com-6390-pos	18:1, 19:1
Com-3418-pos	18:0
Com-7555-pos	18:0
Com-5080-pos	18:1
Com-3704-pos	18:2, 19:1
Com-45-pos	18:1
Com-9817-pos	18:0, 18:3
Com-63-neg	18:1, 20:0
Com-79-neg	16:0, 18:1
Com-1416-neg	16:0
Com-961-neg	16:0, 18:1
Com-278-neg	18:2
Com-610-neg	18:1
Com-490-neg	18:1
Com-1189-neg	18:1
Com-51-neg	18:, 20:1
Com-1367-neg	18:0, 18:2
Com-31-neg	18:0, 18:1
Com-667-neg	18:1
Com-1083-neg	16:0, 18:1
Com-63-neg	18:1, 20:0

Code names (pos and neg)	Lipid metabolites names
Com-51-neg	18:1, 20:1
Com-1189-neg	18:1
Com-136-neg	18:3
Com-639-neg	18:2
Com-1510-neg	18:1,18:2
Com-9-neg	18:2
Com-30-neg	18:2, 18:3
Com-905-neg	18:2
Com-79-neg	16:0, 18:1
Com-1369-neg	18:2

Different proteins and metabolites obtained from proteomics and metabolomics analyses were analyzed to measure the degree of correlation between proteins and metabolites based on Pearson correlation coefficient. The value range of correlation coefficient is -0.99 to + 0.99). When the correlation coefficient is less than 0, it is referred to as negative correlation, while positive correlation refers a coefficient greater than 0; coefficient equals 0, is referred to as zero correlation, i.e., no correlation. In the map, if the expression level of DEPs changed, several DEMs would change accordingly. The down-regulated expression of the three lipoxygenase proteins (9s-lipoxygenase-1, 9s-lipoxygenase-4, and 9s-lipoxygenase-5) resulted in an increase in the 18:1 and a decrease in the 18:2, the DEMs are shown in Table 4 and Fig. 8a-c.

3. Discussion

Soybean is one of the most important global food crops, accounting for most of the world's vegetable oil (Broun et al. 1999). Improving its plant oil is important for crop breeding. Fatty acid metabolism is closely associated with plant stress resistance. Increasing plant fatty acid content, especially polyunsaturated fatty acid levels, plays an important role in maintaining cell membrane homeostasis.

Recent studies on improving seed oil content have mainly focused on the synthesis pathway of plant fatty acids and the assembly pathway of TAG. Some enzymes and their catalytic reactions have been elucidated, but the detailed mechanism of action of other enzymes remains to be determined. Synthesizing vegetable fat involves sugar metabolism, pyruvate metabolism, fatty acid metabolism and several other pathways and enzymes; the interaction between these enzymes still remains unclear. Fatty acid synthesis and TAG assembly are regulated at the transcription, post-transcription, and metabolic levels, however, their regulatory networks remain to be characterized. In addition, some transcription factors (e.g., LEC1, LEC2, WRI1, Dof, ABI3, and FUS3) have been found to play an important role in lipid synthesis (Libeisson et al. 2010). Multi-vector construction and its genetic transformation technology are effective techniques to ensure the coordinated gene expression of multiple metabolic pathways (Sun et al. 2011).

However, there are limited studies on the *DGAT* gene in soybean. DGAT catalyzes the last step in the synthesis of TAG, and is a key enzyme in the regulation of fatty acid accumulation and lipid synthesis in plant seeds (Jako et al. 2001). Seed oil content is correlated with quantitative inheritance, and regulated by several factors, including fatty acid synthesis, lipid accumulation, and seed development (Bao et al. 1999; Yun et al. 2000). Over-expression of *DGAT1* has been shown to promote TAG accumulation in transgenic *Arabidopsis* and tobacco (Bouvier-Nave et al. 2000). Lardizabal et al. (2008) had successfully transferred the *UrDGAT2* gene of *Umbelopsis ramanniana* into soybean, which expressed the gene in its seeds, and bred a new soybean variety with significantly higher oil content. In addition, studies on *WS/DGAT* bifunctional genes and cytoplasmic peanut *AhDGAT* genes have demonstrated the functional diversity of DGAT in regulating lipid synthesis and accumulation, which broadens the research in this field. Increasing DGAT expression in oil seed crops not only increases seed oil content, but also plays an important role in improving seed fatty acid composition.

Compared with animal fats, vegetable oils contain more polyunsaturated fatty acids (18:1, 18:2 and 18:3), which are more beneficial to human health, especially 18:1. As living standards improve, the demand for vegetable oils and fats will increase. When expressed in *Arabidopsis* seeds, *GmDGAT1-2* caused the transgenic seeds to accumulate significantly more TAG 18:1 at the expense of TAG 16:0. This is consistent with previous findings (Chen et al. 2016). However, no previous study has transferred *GmDGAT* genes into soybeans to investigate other functions. Therefore, the present study used genetic engineering as a way to improve the content and quality of fatty acids to regulate the metabolic process of seed oils.

Data from proteomic analysis showed that four oleosin proteins (P24 oleosin isoform B, P24 oleosin isoform A, and two oleosin-5) in the DEPs were up-regulated in the mature transgenic soybeans. Previous studies have shown that over expression of the oleosin genes increased total fatty acid content (Shimada et al. 2010). Although these four oleosin proteins were not enriched in the KEGG pathway, we verified the results using qRT-PCR between the WT and *GmDGAT1-2* transgenic soybeans (Fig. 6b). We speculated that the increased content of fatty acid was associated with the up-regulated expression of oleosin in *GmDGAT1-2* transgenic soybean.

The DEPs were enriched into several oil-related pathways as shown by the KEGG enrichment analysis, including linoleic acid metabolism, fatty acid elongation, fatty acid metabolism, and fatty acid degradation several of which were uncharacterized using Gene Ontology. The over-expression of *GmDGAT1-2* changed the fatty acid components that increased the 18:1 and decrease the 18:2 in the mature seeds in linolenic acid metabolism (Fig. 9). The *lipoxygenase* played an important role in many steps of the pathway.

In addition, most recent studies on *DGAT1* have primarily focused on lipid metabolism in seeds, but only limited studies focused on the relationship between *DGAT1* and plant resistance to growth or the effects of *DGAT1* on other crop traits besides seed oil content. Moreover, few studies have investigated the key physiological roles and influences of *DGAT1* in plants. In our future studies, we aim to address these knowledge gaps.

In the present study, we have characterized the partial functions of 10 *GmDGATs* in *Arabidopsis*, and confirmed the major functions of *GmDGAT1-2* in transgenic soybean using proteomic and lipidomics analyses. Besides, with over-expression of *GmDGAT1-2*, the expression levels of *FATA*, *FATB*, and *SAD* were increased, and the fatty acid content (especially 18:1) of the *GmDGAT1-2*-overexpressed transgenic soybean was significantly higher than that of WT (Fig. 10). Therefore, we speculated that the *GmDGAT1-2* gene was synergistically expressed with the key genes *FATA*, *FATB*, and *SAD* in the fatty acid metabolism pathway and regulated the accumulation of unsaturated fatty acids. Besides, we mapped all the genes screened and those of the fatty acid synthesis pathway (Fig. 11). Through over-expressing the *GmDGAT1-2* gene, the expression levels of the related genes was changed to different degrees (up-regulated or down-regulated), resulting in an increase in 18:1, a decrease in 18:2, and an increase in total fatty acid content. However, the mechanism underlaying the internal interaction warrants further investigated. Compared with traditional breeding approaches, molecular-assisted breeding has the advantages of orientation and short duration. Breeders in the future should apply a combination of molecular-assisted and traditional breeding strategies to production. Our results provide a theoretical support for molecular design breeding.

4. Materials And Methods

4.1. Plant materials

Three soybean *(Glycine max)* varieties were selected in this study. JD5 plants were used for cloning and expression analyses of *GmDGATs* in the 10 *DGAT* family genes of various soybean tissues using qRT-PCR. WT (JACK) mature seeds were used as receptors for soybean genetic transformation and contrast with transgenic soybeans. T_3 *GmDGAT1-2* transgenic soybeans were sampled for expression specificity analysis of various soybean tissues and analysis of fatty acid components using high performance gas chromatography. All 50- day- old pods of Jack and T_3 *GmDGAT1-2* transgenic soybean were sampled for proteomic and lipidomics analyses at the same time. JD5 is a high-oil variety in northeastern China, with an oil content of 24.1%. JACK is a receptor of transgenic soybean strains with high genetic transformation efficiency. All soybeans were planted under natural conditions at the Jilin University Crop Experimental Station of China in spring. Tissues including roots, stems, leaves, flowers, 10-, 20-, 30-, and 40- day-old pods, and mature seeds were sampled in autumn for further experimentation.

Arabidopsis ecotype Col-0 was used to transform 10 target *GmDGAT* genes and determine fatty acid content of WT and transgenic *Arabidopsis*. Two hundred seeds from WT and transgenic *Arabidopsis* plants were sterilized using sodium hypochlorite. After 3 – 5 days of vernalization at 4°C, seeds were planted and germinated in Murashige and Skoog (MS) solid medium in a plant incubator at 22°C under a 16:8 h light:dark cycle regime with 70% humidity. After 2 weeks, seedlings were transplanted into soil for subsequent experiments. The seeds of the WT and 10 transgenic *Arabidopsis* plants were used to extract fatty acids and determine fatty acid content. All tissues were collected from three random, different plants and stored at – 80°C.

4.2. Isolation and sequence analysis of GmDGAT genes from soybean

The full-length sequences of 10 *GmDGAT* genes were obtained using the NCBI Nucleotide Blast tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The conserved regions of these genes were analyzed using SMART (http://smart.embl-heidelberg.DE). Alignments of 10 cDNAs in the sequence were demonstrated in the soybean GmGDB database (http://www.plantgdb.org/GmGDB/cgi-bin/blastGDB). Total RNA was extracted from 40-day-old mature embryos of JD5 to clone five cDNAs (*GmDGAT1-1, GmDGAT1-2, GmDGAT1-3, GmDGAT2-1*, and *GmDGAT2-2*) and total RNA was extracted from young leaves of JD5 to clone five other cDNAs (*GmDGAT2-3, GmDGAT2-4, GmDGAT2-5, GmDGAT3-1*, and *GmDGAT3-2*). The cDNAs were synthesized using the PrimeScript™ 1st Strand cDNA Synthesis Kit (Takara, Changchun, China). The 10 full-length cDNAs were cloned with RT-PCR (the specific primers are listed in Table 5) according to the cDNA sequence in the Primer Premier 5.0 software. The RT-PCR reaction contained 17.5 μL double distilled water (ddH$_2$O), 1 μL of the first strand of cDNA, 1 μL of each primer, 2.5 μL of 10x PCR reaction buffer, 2 μL of dNTP mix, and 0.5 μL *ExTaq* DNA polymerase (Takara, Changchun, China). The specific PCR product was inserted into the pMD18-T cloning vector (Takara, Changchun, China) and sequenced for its accuracy by Biosciences Company (Comate, Jilin, China).

Table 5
Gene specific primer pairs used for RT-PCR, qRT-PCR, and detection.

Purpose	Gene name	Target gene	Forward primer (5'-3')	Reverse primer (5'-3')
RT-PCR	GmDGAT1-1	Glyma09g07520	TGATTCTGATTCTGATTTCTTTCTC	AACCCAGAGGACTCACTATCAAA
qRT-PCR	GmDGAT1-2	Glyma17g06120	GGAGAAGAGAAGACTGAGTTAGAA	TTAGAGTGTTCCTTGGCAGTTATC
	GmDGAT1-3	Glyma13g16560	GAGAAGACTGAGTTAGTAAACACGC	AGGTGGAAATCCTTCAGAGCA
	GmDGAT2-1	Glyma11g09411	CATCACACACTCAGTCAGCGTT	CAAATACTTGACAACATCCAGCATA
	GmDGAT2-2	Glyma01g36011	ATAACTGAACCTTTCCCAGCCTT	GAAATACTTTGACAACATCCAGCA
	GmDGAT2-3	Glyma16g21970	ATAAATTCCATGTCAACCAACTACC	ATTCAAATGTATGACTTCACCCG
	GmDGAT2-4	Glyma09g32790	AACAACACTTCCTCCTCTTACATTC	ACCAATCACTCCACTCCATCTT
	GmDGAT2-5	Glyma16g21960	ACACAACACTTCCTTCTTACATTCG	CCTGACCTGAGAGGAGATAAGTGTA
	GmDGAT3-1	Glyma13g17860	AACAGAGTAGTTTGCATCGCC	AATAGCATTCCCAATAAACAAGG
	GmDGAT3-2	Glyma17g04650	ATAATAGCCACCAACCAGAATCAA	ATAATGTTTCCTCGGCATCCTG
	GmDGAT1-2	Glyma17g06120	CAAGTTGTGGGCTTTCGG	CCAAGTTTGCCTTTCCTA
	13s-lipoxygenase-1	Glyma08g189600	GCAACAAACAGGCATCTTA	AATACCCTCAGCATTCACC
		Glyma07g034800	CCTTCCTGCGGATCTTATC	CACCGTGACCCTTCTCAAC
	9s-lipoxygenase-1	Glyma15g026400	ACTACCACAAAGACTTACGCTAC	ACTTCGCTAACAGGACCAT
		Glyma10g153900	AATAAGGGAGGTGGGTCAT	TCGCCACAAGTTCATCGTA
	9s-lipoxygenase-4			
	9s-lipoxygenase-5			
	KASII	Glyma13g112700	AATCTTCCGCTCCTCCAA	CACCCAATCCTGTCACAACTA
	SAD	Glyma14g121400	TGGTGCTGGTGGGTGATA	AAATGAGGTAATGTACGGTCTT
	FATA	Glyma18g167300	GCAACAAGCAAATGGGTA	GCTATTACTATCTGCCTCTGG
	FATB	Glyma05g012300	TTTTCCCTGTTACTTCACC	GACGCAGATTTGGATTTT
	FAD2	Glyma03g144500	CATAGCCTTCTGCCTCTA	GCACAAATACTTCATCCC
	FAD3	Glyma14g194300	TGTTATGTGGCTGGACTT	ATGAAATGGTAATGGTGC
	β-Tubulin	GMU12286	GGAAGGCTTTCTTGCATTGGTA	AGTGGCATCCTGGTACTGC
	Actin	J01298	GTCCTTTCAGGAGGTACAACC	CCTTGAAGTATCCTATTGAGC
Detection	bar		GTCTGCACCATCGTCAACCACTACA	AGACGTACACGGTCGACTCGGCCGT
	GmDGAT1-2	Glyma17g06120	CAAGTTGTGGGCTTTCGG	CCAAGTTTGCCTTTCCTA

4.3. Plasmid vector construction of GmDGAT genes from soybean

The ORFs of 10 *GmDGAT* genes were each inserted into a separate expression vector pTF101-35s. The resultant pTF101-GmDGATs were heat-shocked into *Agrobacterium tumefaciens* EHA105. The resultant pTF101-GmDGAT1-2 vector was transferred into the JACK soybean plants using genetic transformation of the soybean cotyledon nodes. The main components of the pTF101 vector included Tvsp, bar ORF, TEV enhancer, 2 x p35s, CaMV 35s promoter, NOS terminator, and a series of polyclonal sites (Fig. 3c).

4.4. Transformation of GmDGAT genes in Arabidopsis

Ten resultant pTF101-GmDGATs were heat shocked into *A. tumefaciens* EHA105 and then transferred into *Arabidopsis* Col-0 using the floral-dip method (Clough et al. 1998). The transformed *Arabidopsis* plants were selected on the MS medium containing 4 mg L^{-1} glufosinate-ammonium at 22°C under a 16:8 h light:dark cycle regime with 70% humidity in the plant incubator. Then the T_0 transgenic *Arabidopsis* were screened using PCR and bar strip to confirm bar and target gene expression, according to the manufacturer's instruction. The bar strip (catalog number A07-13-413, Beijing Artron Biotech, Ltd, Beijing, China) coded PAT protein. Finally, T_3 transgenic *Arabidopsis* seeds with the 10 *GmDGAT* genes were harvested for further experiments. Samples were checked to make sure that there were at least three transgenic positive plants for each gene.

4.5. GmDGAT1-2 expression in soybean

The pTF101-*GmDGAT1-2* construct was heat shocked into *A. tumefaciens* EHA105 and then transferred into cotyledon nodes of the soybean JACK. The T_0 transformed soybean cotyledon nodes were selected on MS and B5 media containing 8 mg L^{-1} glufosinate-ammonium and grown at 24°C:22°C under an 18:6 h light:dark cycle and at a relative humidity of 60%. After a series of positive identifications, the T_1 transformed soybeans were planted in the plant incubator. T_2 and T_3 transgenic soybeans were planted under natural conditions at the Jilin University Crop Experimental Station of China. The JACK receptors were planted simultaneously as controls. Finally, four T_3 transgenic soybean lines were obtained and harvested for further experimentation.

4.6. Detection of GmDGAT1-2 transgenic soybean by molecular methods

The regenerated plantlets were transplanted into a plant incubator. The T_0 transgenic plants were screened using PCR and bar strip to confirm bar expression. Transgenic plants were tested for the target gene by RT-PCR using DNA of soybean leaves as the template. The primer pairs were used to amplify the bar sequences and the target genes (Table 5). The bar sequence for PCR was 220 bp with the target *GmDGAT1-2* gene sequence for RT-PCR being 1497 bp. To prevent interference from endogenous genes, we designed the upstream primer in the pTF101 vector and downstream primer in the ORF. The bar strip test is a simple and accurate method for detecting positive plants, and two lines appearing in the bar strip showed that the plants were positive for the target gene.

Southern blot hybridization was performed to detect transgenic plants using DIG High Prime DNA Labeling and Detection Starter Kit II (Roche, Germany). DNA was extracted from the young leaves of T_3 transgenic and WT (JACK) plants using the Plant Genomic DNA Kit (Cowin Bio, China). Approximately 10 μg of genomic DNA was digested with *Hind* III (Takara, China). The PCR primers of the bar gene (220 bp) and target gene (140 bp) were used as probes for hybridization; they are given in Table 5. Agarose electrophoresis, washing gel, transferring membrane, probe labeling, prehybridization, hybridization, membrane washing, and signal detection were performed according to the manufacturer's instructions (Roche Applied Science).

In this study, we synthesized mouse monoclonal antibodies against PAT protein from a biotechnology company (Changchun, China) as the primary antibody, and purchased corresponding secondary antibodies (HRP Goat Anti-Mouse IgG) from the ABclonal Antibody Development Company (Wuhan, China). The titers of both antibodies were 1:5000.

The western blot was used to detect the expression of the target protein in transgenic soybeans using the antigen-antibody specific binding principle. Negative control and target proteins were extracted from JACK and transgenic soybean leaves. The protein was extracted from 0.1 g of young leaves of T_3 transgenic and WT plants (JACK) using the Plant Total Protein Extraction Kit (Sangon Biotech, China). Specific hybridization steps included SDS-PAGE gel electrophoresis, transferring membrane (nitrocellulose membrane), blocking, primary antibody incubation (one hour), washing and secondary antibody incubation (one hour), and exposure imaging in a dark case. A pre-stained protein marker was purchased from Genview (Ding Guo Biotechnology Company, Beijing, China).

ELISA analysis was performed to determine the expression of PAT protein in transgenic plants. The 0.1 g soybean leaves were fully ground with liquid nitrogen, 100 μL coating buffer was added, the mixture was centrifuged in 12000 rpm, at 4°C for 10 min, and the supernatant was then absorbed. All samples (100 μL per well) were added to the 96-well operating plate for three biological replicates and three technical replicates. The PAT protein was the positive control. Negative control and target proteins were

extracted from JACK and transgenic soybean leaves. The sealing buffer was 3% bovine serum protein, the washing buffer was PBS with 0.05% Tween 20, and the color buffer was Soluble TMB Substrate Solution. The final protein results were determined in OD_{450} and performed in the Eon High Performance Microplate Spectrophotometer (BioTek, VT, USA). Protein expressions were calculated using the standard curve. The standard curve was drawn from the concentration gradient of the five proteins from the positive control (0.5, 1, 2, 4, and 8 μg/mL).

4.7. Quantitative RT-PCR analysis

Total RNA of the WTs (JACK and JD5) and *GmDGAT1-2* transgenic soybean tissues (roots; stems; leaves; and 10-, 20-, 30-, and 40-day-old pods) were extracted using RNAiso Plus and synthesized to cDNA using an M-MLV Kit (Takara, Changchun, China) according to the manufacturer's instructions. qRT-PCR was performed using SYBR Premix Ex Taq and analyzed using Applied Biosystems Step One Plus™ software. Each reaction mixture was composed of SYBR Green I, a cDNA sample, ROX reference Dye II, gene-specific primers, and distilled water to a total volume of 20 μL. All the primer pairs used to amplify the 10 *GmDGAT* genes, *β-Tubulin*, and *Actin* are given in Table 5. Two internal reference genes (*β-Tubulin* and *Actin*) were chosen to standardize the data (Jian et al. 2008; Hu et al. 2009; Le et al. 2012).

The qRT-PCR conditions were as follows: 95°C for 30 s, 95°C for 5 s (40 cycles), 60°C for 34 s, and 72°C for 30 s. The double delta CT method was used to evaluate gene expression. All qRT-PCR reactions were run for three biological replicates and three technical replicates. These methods rigorously followed the MIQE guidelines for qRT-PCR. The expression results of different tissues were predicted using Phytozome online software.

4.8. Measurement of fatty acid content in Arabidopsis and soybean

All the fatty acids were extracted from 10 mature seeds of *GmDGAT* transgenic *Arabidopsis* using the methyl ester method. Samples (0.03 g) of *Arabidopsis* seeds were ground into powder with liquid nitrogen; 1 mL 2.5% concentrated sulfuric acid-methanol solution was added and mixed in by inverting. The mixture was kept at 70°C for 30 min in a water bath. After the sample was cooled, the supernatant was absorbed and 600 μL NaCl and 300 μL *n*-hexane were added; the mixture was shaken by inverting for several minutes, then centrifuged at 5000 rpm for 10 min. Finally, the supernatant was absorbed, and the sample was tested and analyzed using an Agilent 7890A GC system with flame ionization detector (FID) according to conditions previously described (Browse et al. 1986). An Agilent capillary column PEG-20M (30m × 320 μm × 0.33 μm) was used as the chromatographic column. The detector was maintained at a constant temperature of 290°C, the inlet temperature at 250°C, and the column temperature at 200°C. The injection volume was 1 μL, and the split injection method as used with a split ratio of 5:1. The carrier gas was nitrogen at a flow rate of 12.602 mL·min^{-1}. The velocity of the hydrogen was 30 mL·min^{-1}. The air velocity was 400 mL·min^{-1}. The following heating procedure was used: 180°C for 4 min, increased to 220°C over 10 min at a rate of 4°C·min^{-1}. Using five fatty acid standards (16:0, 18:0, 18:1, 18:2, and 18:3) as the control material, serial dilutions of the standard solution were prepared with a concentration gradient to measure and record the corresponding peak values at different peak times. The slope and intercept were also calculated and used to draw the standard curve, which was used to calculate the levels and compositions of fatty acids and oils.

The total fatty acids in mature seeds of *GmDGAT1-2* transgenic soybean were extracted by the hydrolysis-extraction and Soxhlet extractor methods (NY/T-1982), which are very common normalization methods for determining fatty acids in food in China (GB5009.168-2016). The following four chromatographic determination steps were taken: hydrolysis, fat extraction, fat saponification, and fatty acid methylation. Finally, the samples were tested and analyzed with a Trace GC 1300 system (YQ472, Thermo Fisher, USA) with FID. The chromatographic column was a Thermo Fisher chromatographic column TR-FAME (100 m × 0.25 mm × 0.2 μm). The detector was maintained at 260°C, and the inlet temperature at 250°C. The injection volume was 1 μL; the split ratio for the split injection method was 8:1. The carrier gas was nitrogen, at a flow rate of 2 mL·min^{-1}. The following heating procedure was used: 100°C for 15 min, 225°C for 1 min, then increased to 250°C at a rate of 5°C·min^{-1}. A standard composed of 37 fatty acids (BW4368-01) was used, and five common fatty acids (16:0, 18:0, 18:1, 18:2, and 18:3) were used as the control material. Serial dilutions of the standard were prepared along a concentration gradient to measure and record the corresponding peak values at different peak times, and the slope and intercept were calculated to draw the standard curve. The contents and compositions of fatty acids and oils were calculated using the standard curve.

4.9. Analysis of differentially expressed proteins (DEPs)

The 40-day-old pods of WT (control) and *GmDGAT1-2* transgenic soybeans were sampled at the same time and place. Then, the soybean samples of one line were mixed and then divided (at least 250mg/sample) into 2 mL centrifuge tubes. After labeling, it was quickly placed into liquid nitrogen and frozen for at least 15min. The specific methods of total protein extraction refer to relevant instructions (Kachuk et al. 2015, Liang et al. 2018). All proteins were tested for quality according to the instructions of Bradford protein quantitative kit and each protein peptides was TMT labeled (Zhang et al. 2016). For transition library construction, shotgun proteomics analyses were performed using an EASY-nLCTM 1200 UHPLC system (Thermo Fisher) coupled with a Q Exactive™ HF-X mass spectrometer (Thermo Fisher) operating in the data-dependent acquisition (DDA) mode using LC-MS/MS. After data analysis, the protein began to be identified and quantitated. The proteins whose quantitation significantly different between experimental and control groups, (p < 0.05 and |log2FC| > * (FC > * or FC < * [fold change, FC]), were defined as differentially expressed proteins (DEP). Gene Ontology (GO) and InterPro (IPR) functional analysis were conducted using the interproscan program against the non-redundant protein database (including Pfam, PRINTS, ProDom, SMART, ProSite, PANTHER) (Jones et al. 2014), and the databases of COG (Clusters of Orthologous Groups) and KEGG (Kyoto Encyclopedia of Genes and Genomes) were used to analyze the protein family and pathway. DPEs were used for Volcanic map analysis, cluster heat map analysis and enrichment analysis of GO, IPR and KEGG (Huang et al. 2014). The probable protein-protein interactions were predicted using the STRING-db server (Franceschini et al. 2012).

4.10. Analysis of differentially metabolites (DEMs)

For lipid extraction, Methanol (0.75 mL) was added to a 100 μL sample (or 100 mg tissue powder), which was placed into a glass tube with a Teflon lined cap, and the tube was vortexed. A 2.5 mL MTBE was added and the mixture was incubated for 1 h at room temperature on a shaker. After a 10-min incubation at room temperature, the sample was centrifuged at 5000 rpm for 10 min. The upper (organic) phase was collected, and the lower phase was re-extracted with 1 mL of the solvent mixture (MTBE/methanol/water (10:3:2.5, v/v/v)), and collecting the upper phase. Combined organicphases were dried and dissolved in 100 μL of isopropanol for storage (Matyash et al. 2008), then analyzed by LC-MS/MS.

UHPLC-MS/MS analyses were performed using a Vanquish UHPLC system (Thermo Fisher, Germany) coupled with an Orbitrap Q Exactive™ HF mass spectrometer (Thermo Fisher, Germany) in Novogene Co., Ltd. (Beijing, China). Samples were injected onto a Thermo Accucore C30 column (150 × 2.1mm, 2.6 μm) using a 20 min linear gradient at a flow rate of 0.35 mL/min. The column temperature was set at 40°C. Mobile phase buffer A was acetonitrile / water (6/4) with 10 mM ammonium acetate and 0.1% formic acid, whereas buffer B was acetonitrile/ isopropanol (1/9) with 10 mM ammonium acetate and 0.1% formic acid. The solvent gradient was set as follows: 30% B, initial; 30% B, 2 min; 43% B, 5 min; 55% B, 5.1 min; 70% B, 11 min; 99% B, 16 min; 30% B, 18.1 min. Q Exactive™ HF mass spectrometer was operated in positive and negative polarity mode with sheath gas :20 arbitrary units, sweep gas: 1 arbitrary units, auxiliary gasrate: 5, spray voltage: 3 kV, capillary temperature: 350°C, heater temperature: 400°C, S-Lens RF level: 50, resolving power (full scan): 120000, scan range: 114−1700 m/z, automatic gain control target: 1e6, resolving power (MS2): 30000 (Top20), normalized collision energy: 25; 30 [20; 24; 28], Injection time: 100 ms, Isolation window: 1m/z, automatic gaincontrol target (MS2): 1e5, dynamic exclusion: 15s.

The raw data files generated by UHPLC-MS/MS were processed using the Compound Discoverer 3.01 (CD3.1, Thermo Fisher) to perform peak alignment, peak picking, and quantitation for each metabolite. After that, peak intensities were normalized to the total spectral intensity. The normalized data was used to predict the molecular formula based on additive ions, molecular ion peaks and fragment ions. Peaks were then matched with the Lipidmaps, Lipidblast and HMDB database to obtained the accurate qualitative and relative quantitative results.

Principal components analysis (PCA) and Partial least squares discriminant analysis (PLS-DA) were performed at metaX (a flexible and comprehensive software for processing metabolomics data). We applied univariate analysis (t-test) to calculate the statistical significance (P-value). The metabolites with VIP > 1 and P-value < 0.05 and fold change ≥ 2 or FC ≤ 0.5 were considered to be differential metabolites. Volcano plots were used to filter metabolites of interest which based on Log2 (FC) and -log10 (P-value) of metabolites. For clustering heat maps, the data were normalized using z-scores of the intensity areas of differential metabolites and were ploted by Pheatmap package in R language. The correlation between differentially expressed metabolites were analyzed in R language. Statistically significant of correlation between DEMs were calculated in R language. P-value < 0.05 was considered as statistically significant and correlation plots were ploted by corrplot package in R language.

4.11. Statistical analysis

All statistical analyses were performed using SPSS 19.0 (IBM Corp, USA). Differences between samples were considered significant at $p < 0.05$ or $p < 0.01$, based on an independent-sample t-test. Unless indicated otherwise, the mean values and standard deviation are shown.

4.12. Accession number

Sequence data of six GmDGAT proteins: GmDGAT1-1, GmDGAT1-2, GmDGAT1-3, GmDGAT2-2, GmDGAT2-4, and GmDGAT2-5 were deposited in the DDBJ Database under accession LC533831, LC533832, LC533833, LC533834, LC533835, and LC533836.

Abbreviations

DGAT   Diacylglycerol acyltransferase

MS   Murashige and Skoog

NCBI   National Center for Biotechnology Information

ORF   Open reading frame

qRT-PCR   Quantitative real time PCR

SD   Standard Deviation

ACP   Acyl carrier protein

PAT   Phosphinothricin-N-acetyltransferase

MBOAT   Membrane-bound O-acyltransferase

TAG   triacylglycerol

WT   Wild type.

Declarations

References

1. Barthole G, Lepiniec L, Rogowsky PM et al (2012) Controlling lipid accumulation in cereal grains [J]. Plant Sci 185–186:33–39

2. Broun P, Gettner S, Somerville C (1999) Genetic engineering of plant lipids [J]. Annu Rev Nutr 19:197–216

3. Bouvier-Navé P, Benveniste P, Oelkers P et al (2000) Expression in yeast and tobacco of plant c DNAs encoding acyl-CoA:diacylglycerol acyltransferase [J]. Eur J Biochem 267(1):85–96

4. Bao X, Ohlrogge J (1999) Supply of triacylglycerol in develo-ping embryos [J]. Plant Physiol 120:1057–1062

5. Browse J, McCourt PJ, Somerville CR (1986) Fatty acid composition of leaf lipids determined after combined digestion and fatty acid methyl ester formation from fresh tissue. Anal Biochem 152:141–145

6. Chen BB, Wang JJ, Zhang GY (2016) Two types of soybean diacylglycerol acyltransferases are differentially involved in triacylglycerol biosynthesis and response to environmental stresses and hormones [J]. Sci Rep 6:28541

7. Cases S, Smith SJ, Zheng YW (1998) Identification of a gene encoding an acyl-Co A: diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis [J]. *Proceedings of the National Academy of Sciences* 95(22): 13018–13023

8. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16(6):735–743

9. Franceschini A, Szklarczyk D, Frankild S et al (2012) STRING V9.1: Protein-Protein Interaction Networks, with Increased Coverage and Integration [J]. *Nucleic Acids Research* 41

10. Guo R et al (2017) Ionomic and metabolic responses to neutral salt or alkaline salt stresses in maize (Zea mays L.) seedlings [J]. BMC Plant Biol 17:41–53

11. Hobbs HD, L CF, Hills M (1999) Cloning of a c DNA encoding acyltransferase from Arabidopsis thaliana and its functional expression [J]. FEBS Lett 452:145–149

12. He XH, Grace QC, Lin JT et al (2004) Regulation of diacylglycerolacy transferase in developing seeds of castor [J]. Lipids 39(9):865–871

13. Hernandez ML et al (2012) A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose- rescued Arabidopsis seed oil catabolism mutants. Plant Physiol 160:215–225

14. Hu RB, Fan CM, Li HY, Zhang QZ (2009) Evaluation of Putative Reference Genes for Gene Expression Normalization in Soybean by Quantitative Real-Time RT-PCR. BMC Mol Biol 10(1):93

15. Heischmann S et al (2016) Exploratory Metabolomics Profiling in the Kainic Acid Rat Model Reveals Depletion of 25-Hydroxyvitamin D3 during Epileptogenesis. *Scientific Reports* 6

16. Haspel JA et al (2014) Circadian rhythm reprogramming during lung inflammation. *Nature communications* 5

17. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists [J]. Nucleic Acids Res 37(1):1–13

18. Jian Bo, Liu B, Bi YR, Hou WS, Wu CX (2008) Validation of Internal Control for Gene Expression Study in Soybean by Quantitative Real-Time PCR. BMC Mol Biol 9(1):59

19. Jako C, Kumar A, Wei YD et al (2001) Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight [J]. Plant Physiol 126:861–874

20. Jones P, Binns D, Chang HY et al (2014) InterProScan 5: genome-scale protein function classification[J]. Bioinformatics 30(9):1236–1240

21. Kroon JT, Wei W, Simon WJ et al (2006) Identification and functional expression of a type 2 acyl-Co A:diacylglycerolacyl transferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals [J]. Phytochemistry 67(23):2541–2549

22. Kamisaka Y, Mishra S, Nakahara T (1997) Purification and characterization of diacylglycerol acyltransferase from the lipid body fraction of an oleagious fungus [J]. The Journal of Biological Chemistry 121:1107–1114

23. Kachuk, Carolyn S et al (2015) Comparison of sodium dodecyl sulfate depletion techniques for proteome analysis by mass spectrometry [J]. *Journal of chromatography*, A: Including electrophoresis and other separation methods

24. Liangjie, Niu H et al (2018) Modified TCA/acetone precipitation of plant proteins for proteomic analysis [J]. *PLoS ONE*

25. Liu ZJ (2013) Cloning of genes involved in seed oil content and transgenic research of cotton [D]. *Beijing: China Agricultural University, Doctor Dissertation*

26. Lardizabal KD, Mai JT, Wagner NW et al (2001) DGAT2 is a new diacylglycerolacyl transferase gene family: purification, cloning, and expression in insect cells of two polypeptides from Mortierella ramanniana with diacylglycerolacyl transferase activity [J]. The Journal of Biological Chemistry 276(42):38862–38869

27. Li F, Wu X, Lam P, Bird D, Zheng H, Samuels L et al (2008) Identification of the was ester synthase/acyl-coenzyme A: Diacylglycerol acyltransferase WSD1 required for stem was ester biosynthesis in Arabidopsis. Plant Physiol 148:97–107

28. Liu GQ, Shao Q, Huang R, Feng XZ (2013) Characterization and Expression Analysis of *DGAT* Gene Family in Soybean. Chin Agric Sci Bull 29(12):55–61

29. Lardizabal K, Effertz R, Levering C et al (2008) Expression of Umbelopsis ramanniana DGAT2A in seed increases oilin soybean [J]. Plant Physiol 148:89–96

30. Libeisson Y, Shorrosh B, Beisson F et al (2010) Acyl-lipidmetabolism [J]. Arabidopsis Book 8(8):e0133

31. Le Dung, Tien DL, Aldrich, Babu, Valliyodan, Yasuko, Watanabe C, Van H, Rie N, Satish K, Guttikonda (2012) Evaluation of Candidate Reference Genes for Normalization of Quantitative RT-PCR in Soybean Tissues under Various Abiotic Stress Conditions. Edite par Christian Schönbach. *PLoS ONE* 7 (9)

32. Mekhedov S, De Ilárduya OM, Ohlrogge J (2000) Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis [J]. Plant Physiol 122(2):389–402

33. Matyash V, Liebisch G, Kurzchalia TV et al (2008) Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics [J]. J Lipid Res 5:1137–1146

34. Nykiforuk CL, Laroche A, Weselake RJ (1999) Isolation and characterization of a c DNA encoding a second putative diacylglycerolacyl transferase from a microspore-derived cell suspension culture of Brassica napus L. cv Jet Neuf [J]. Plant Physiol 121(3):1957–1959

35. Nazrul I, Philip DB, Maria KM J, et al (2019) Quantitative Proteomic Analysis of Low Linolenic Acid Transgenic Soybean Reveals Perturbations of Fatty Acid Metabolic Pathways. [J]. Proteomics 19:1800379

36. Okuley J, Lightner J, Feldmann K et al (1994) Arabidopsis *FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis [J]. Plant Cell 6:147–158

37. Padula MP, Berry IJ, O'Rourke M (2017) A comprehensive guide for performing sample preparation and top-down protein analysis. Proteomes [J] 5:11

38. Reed DW, Schafer UA, Covello PS et al (2000) Characterization of the Brassica napus extraplastidial linoleate desaturase by expression in Saccharomyces cerevisiae [J]. Plant Physiol 122:715–720

39. Ross PL, Huang YN, Marchese JN et al (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.[J]. Molecular Cellular Proteomics Mcp 3(12):1154–1169

40. Salanoubat M, Lemcke K, Rieger M et al (2000) Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana [J]. Nature 408(6814):820–822

41. Shockey JM, Gidda SK, Chapital DC et al (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum [J]. Plant Cell 18(9):2294–2313

42. Saha S, Enugutti B, Rajakumari S et al (2006) Cytosolic triacylglycerol biosynthetic pathway in oil seeds. Molecular cloning and expression of peanut cytosolic diacylglycerolacyl transferase [J]. Plant Physiol 141(4):1533–1543

43. Sun QX (2011) Cloning and characterization of genes involved in the biosynthesis of very long chain polyunsaturated fatty acid and the reconstitution of this pathway in crop plants [D]. *Shandong Taian: Shandong Agricultural University, Doctor Dissertation*

44. Shimada TL, Hara-Nishimura I (2010) Oil-body-membrane proteins and their physiological functions in plants [J]. Biological Pharmaceutical Bulletin 33:360–363

45. Tang HY et al (2018) Lipidomics reveals accumulation of the oxidized cholesterol in erythrocytes of heart failure patients. Redox Biology 14:499–508

46. Xu JY, Francis T, Mietkiewska E et al (2008) Cloning and characterization of an acyl Co

47. A-dependent diacylglycerolacyl transferase 1 (*DGAT1*) gene from Tropaeolummajus, and a

48. study of the functional motifs of the DGAT protein using site-directed mutagenesis to modify

49. enzyme activity and oil content[J]. *Plant Biotechnology Journal* 6(8):799–818

50. Yen CE, Stone SJ, Koliwad S et al (2008) DGAT enzymes and triacylglyrol biosynthesis [J]. J Lipid Res 49:2283–2301

51. Yun SL, Isleib TG (2000) Genetic analysis of fatty acids in American peanuts [J]. Chinese journal of oil crops 22(1):35–37

52. Zhang H, Liu T, Zhang Z et al (2016) Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer [J] *Cell*

53. Zheng P, Allen WB, Roesler K et al (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize [J]. Nat Genet 40(3):267–372
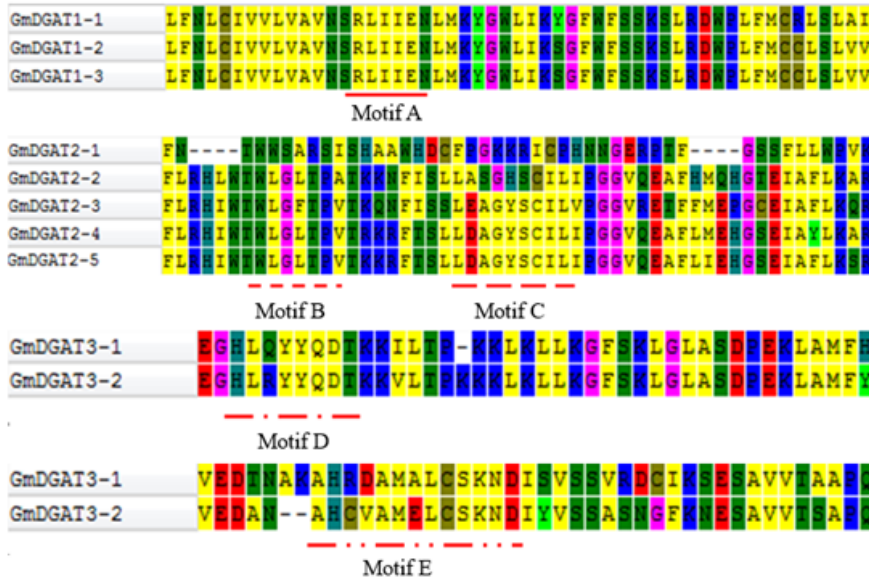
Figures

Figure 1

Partial alignment of deduced amino acids of 10 GmDGATs in soybean. Sequences were aligned using MEGA 6. The different lines represent five special domains. RLIIE (A), WLGLT (B), LDAGYSC (C), HVQYYGD (D), and AHCVAMELCSKND (E) are five conserved motifs. GenBank accession numbers of the 10 genes are: GmDGAT1, LC533831, LC533832, and LC533833; GmDGAT2, XM_028384419.1, LC533834, XM_003547843.4, LC533835, and LC533836; GmDGAT3, XM_003542355.4 and XM_028353784.1
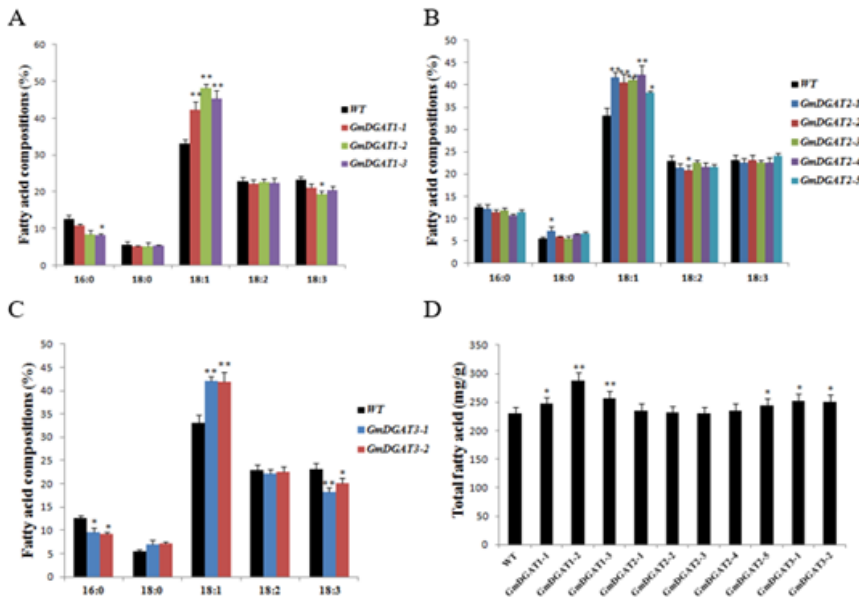

Figure 2

(a) Compositions of five fatty acids in triglyceride (TAG) from wild type (WT) and three GmDGAT1 over-expressing transgenic Arabidopsis lines using high performance gas chromatography. (b) Fatty acid composition of TAG from WT and five GmDGAT2 over-expressing transgenic Arabidopsis lines using high performance gas chromatography. (c) Fatty acid composition of TAG from WT and two GmDGAT3 over-expressing transgenic Arabidopsis lines using high performance gas chromatography. (d) Total TAGs

in 10 GmDGATs overexpressing Arabidopsis seeds. The data represent the average of at least three independent experiments ± SD. Error bars indicate the SE of three replicates. *P < 0.05 and **P < 0.01 by independent-sample t-test for significant difference.
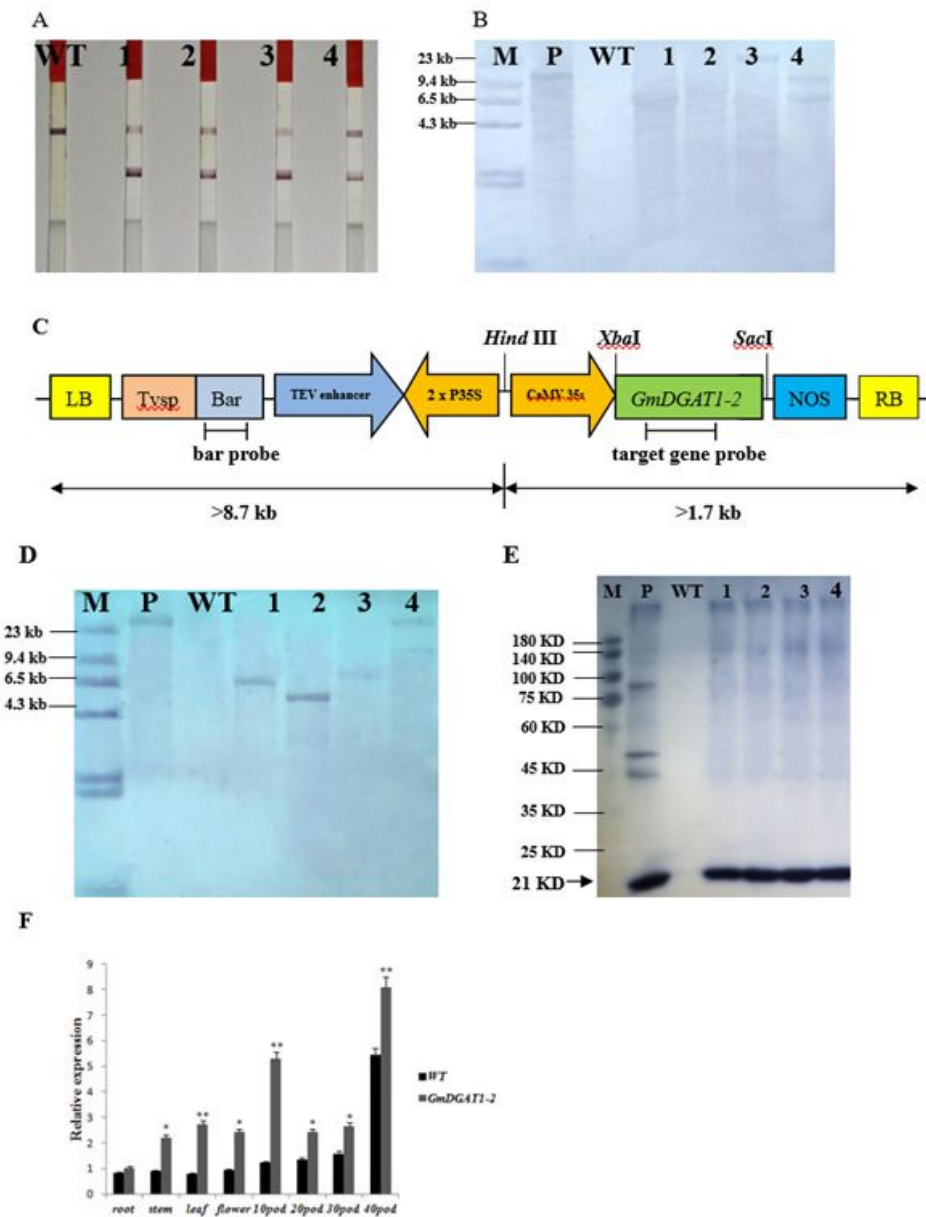
Figure 3

Detecting GmDGAT1-2 in transgenic soybean. (a) Bar strip analysis of four GmDGAT1-2 transgenic soybean lines. (b) Southern blot of T3 transgenic soybeans using the bar gene fragment as a probe. The genomic DNA of transgenic plants was digested with Hind III. M, DNA Marker; P, recombinant plasmid as positive control; WT, wild-type plants as negative control; numbers 1−4 are transgenic lines. (c) Map of pTF101-GmDGAT1-2 vector used for GmDGAT1-2 transgenic soybean lines. (d) Southern blot of T3 transgenic soybeans using the target gene fragment as a probe. The genomic DNA of transgenic plants was digested with Hind III. M, DNA Marker; P, recombinant plasmid as positive control; WT, wild-type plants as negative control; numbers 1−4 are transgenic lines. (e) Western blot of T3 transgenic soybeans. M, pre-stained protein marker; P, PAT protein as positive control; WT, wild-type as negative control; numbers 1−4 are transgenic lines. (f) GmDGAT1-2 expression in different soybean tissues (root, stem, leaf, flower, and pods) was determined using quantitative RT-PCR. The data represent the average of three independent experiments ± SD. Error bars indicate the SE of three replicates. Values were normalized against the result for β-tubulin and soybean Actin. Statistical significance was determined using independent-sample t-test (*p < 0.05, **p < 0.01).
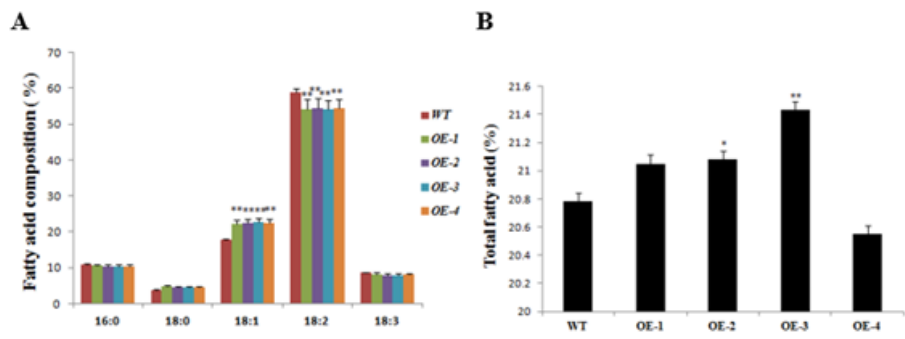
Figure 4

a) Seed oil content and composition in wild type (JACK) and four different GmDGAT1-2 transgenic lines (OE-1 to 4) of soybean, measured using high performance gas chromatography. (b) Total fatty acid of WT and four GmDGAT1-2 overexpressing soybean seeds. Data are the average values of three biological replicates. Error bars indicate the SE of three replicates. Statistical significance was determined by independent-sample t-test (*p < 0.05, **p < 0.01).
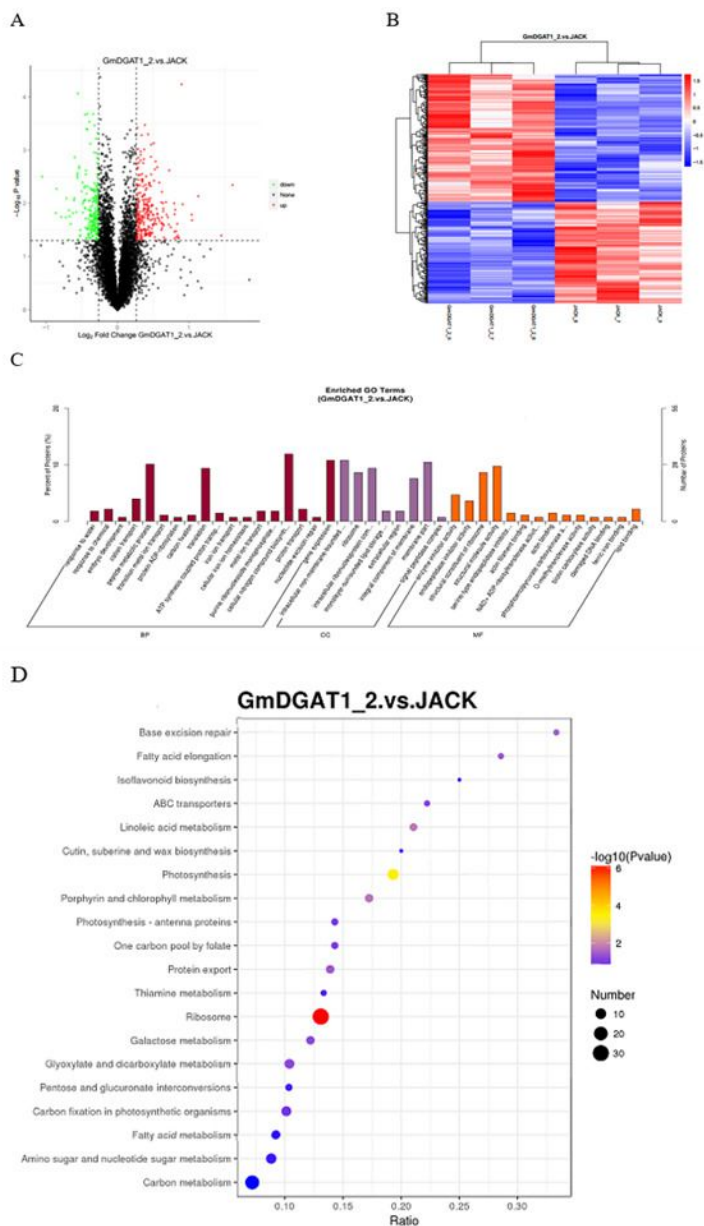
Figure 5

(a) scatter plot of total proteins identified: Black represents non-significant proteins, red represents up-regulated DEPs, and green represents down-regulated DEPs. (b) hierarchical clustering analyses showing distinct variations between control and GmDGAT1-2 transgenic soybean (c) a GO classification of the identified proteins between control and transgenic soybean. The results for the three main GO categories are summarized: BP, CC, and MF. (d) KEGG pathways of the differentially expressed proteins between Jack (control) and the 40- day-old GmDGAT1-2 transgenic soybean pods.
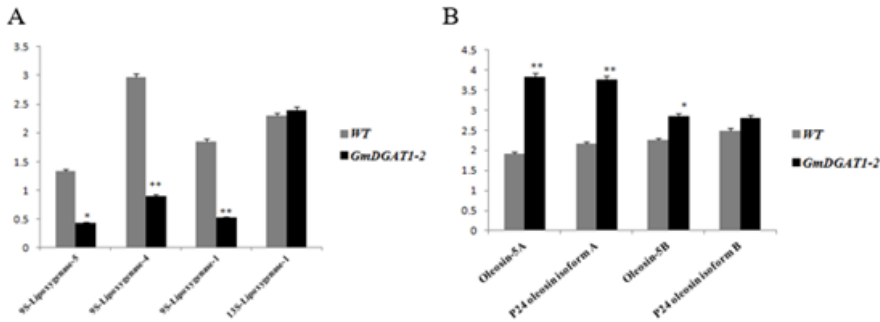
Figure 6

(a) Expression of four differentially expressed lipoxygenase genes between WT (control) and GmDGAT1-2 transgenic soybean seeds was determined using quantitative RT-PCR. (b) Expression of four differentially expressed oleosin genes between WT (control) and GmDGAT1-2 transgenic soybean seeds was determined using quantitative RT-PCR. The data represent the average of three independent experiments ± SD. Error bars indicate the SE of three replicates. Values were normalized against the result for β-tubulin and soybean Actin. Statistical significance was determined via an independent-sample t-test (*p < 0.05, **p < 0.01).
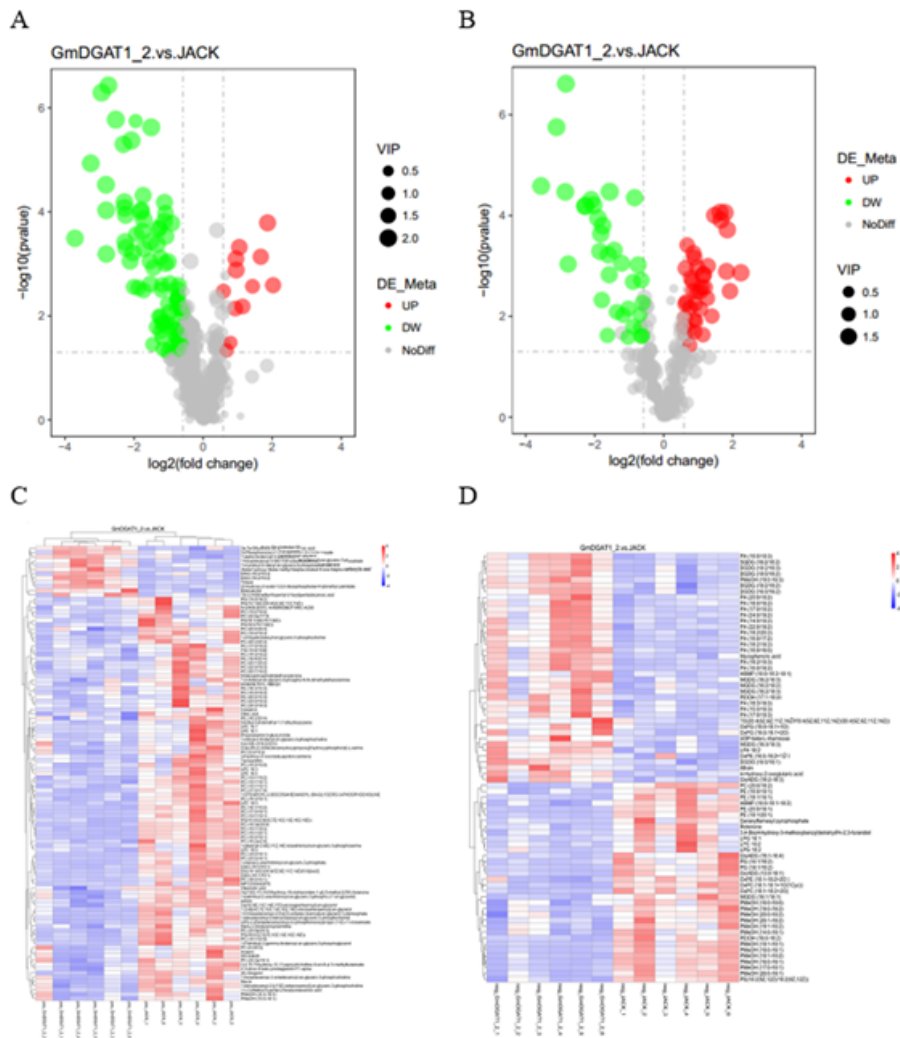
Figure 7

a and b) Volcanic map of differential lipid compounds. Each dot in the volcanic map represents a lipid compound (positive and negative ions mode). Significant up-regulated lipid compounds are represented by red dots, while significantly down-regulated lipid compounds are represented by green dots. The size of the dot represents the VIP value. (c and d) Clustering heat map of total differential lipid compounds. Vertical is the sample clustering, horizontal is the lipid clustering; the shorter the cluster branch represents the higher similarity.
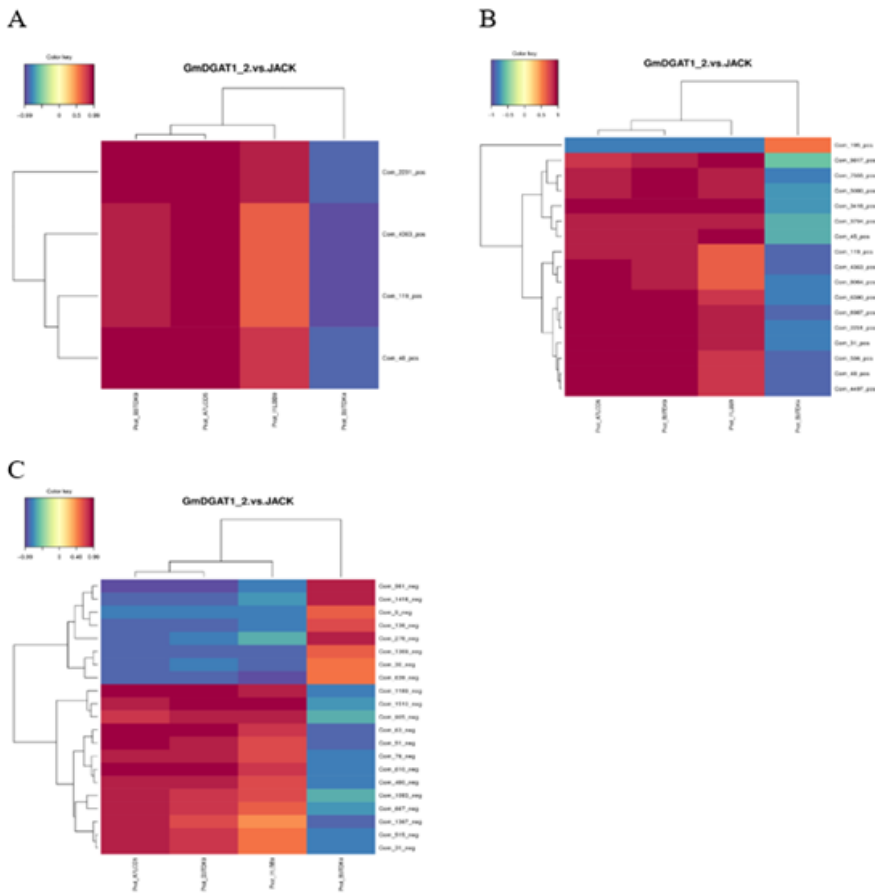
Figure 8

(a) Correlation analysis of protein and metabolism between top four DEPs and top four DEMs (b and c) Correlation analysis of protein and metabolism between top four DEPs and some other top DEMs. The redder the color, the stronger the positive correlation, while the bluer the color, the stronger the negative correlation.
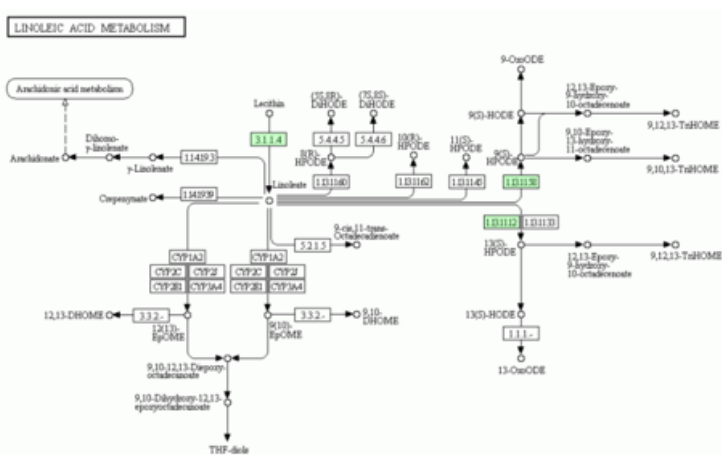

Figure 9

Effect of GmDGAT1-2 over-expression on enzymes in linoleic acid metabolism in the cytoplasm based on KEGG analysis. Proteins are indicated by enzyme numbers. The lipoxygenases [EC:3.1.1.4], [EC:1.13.11.12] and [EC:1.13.11.58] are the key points of DEPs in the pathway.
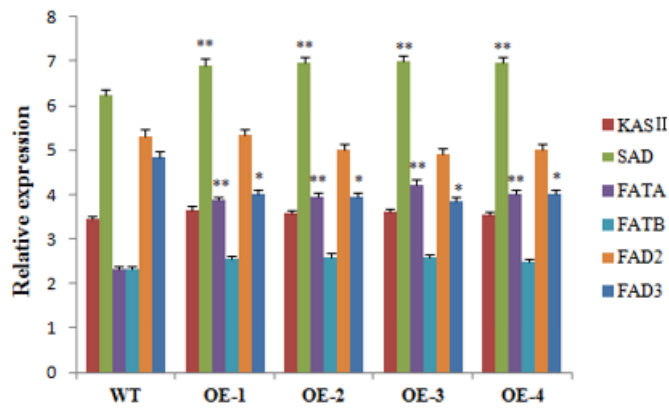
Figure 10

The expression of key genes KASII, SAD, FATA, FATB, FAD2, and FAD3, involved in the fatty acid metabolic pathway in the seeds of T3 GmDGAT1-2 transgenic lines (OE-1 to OE-4) of soybean using quantitative RT-PCR. The data represent the average of three independent experiments ± SD. Error bars indicate the SE of three replicates. Values were normalized against the result for β-tubulin and soybean Actin. Statistical significance was determined using independent-sample t-test (*$p < 0.05$, **$p < 0.01$).
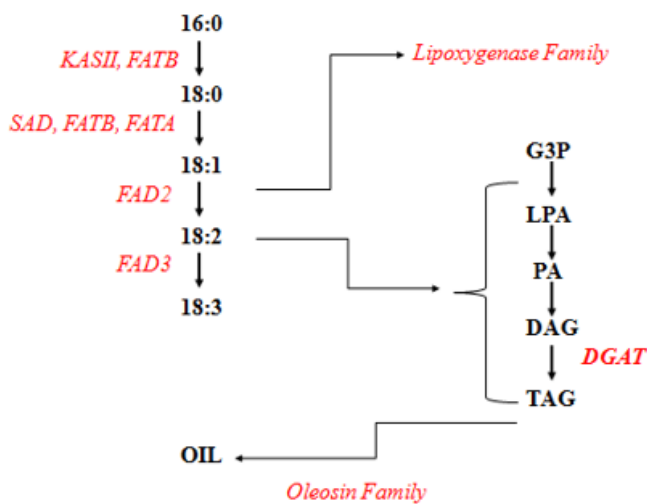

Figure 11

A schematic diagram on fatty acid synthesis. The red markers are the key genes in the pathway including six genes: KASII, SAD, FATA, FATB, FAD2, FAD3, and two families: lipoxygenase family and oleosin family.