Proteomics
www.proteomics-journal.com

Quantitative Proteomic Analysis of Low Linolenic Acid Transgenic Soybean Reveals Perturbations of Fatty Acid Metabolic Pathways

*Nazrul Islam, Philip D. Bates, K. M. Maria John, Hari B. Krishnan, Zhanyuan J. Zhang, Devanand L. Luthria, and Savithiry S. Natarajan**

To understand the effect of fatty acid desaturase gene (*GmFAD3*) silencing on perturbation of fatty acid (FA) metabolic pathways, the changes are compared in protein profiling in control and low linolenic acid transgenic soybeans using tandem mass tag based mass spectrometry. Protein profiling of the transgenic line unveiled changes in several key enzymes of FA metabolism. This includes enzymes of lower abundance; fabH, fabF, and thioestrase associated with FA initiation, elongation, and desaturation processes and LOX1_5, ACOX, ACAA1, MFP2 associated with β-oxidation of α-linolenic acids pathways. In addition, the *GmFAD3* silencing results in a significant reduction in one of the major allergens, Gly m 4 (C6T3L5). These results are important for exploring how plants adjust in their biological processes when certain changes are induced in the genetic makeup. A complete understanding of these processes will aid researchers to alter genes for developing value-added soybeans.

1. Introduction

Soybean has multifarious usages for human and animal consumption that has made it a prime target for genetic manipulation. Such manipulations have been more prominent since the inception of genome sequences in 2010. Among these modifications, gene manipulations for oil composition were intensively investigated[1,2] due to consumer health awareness and increased demand for vegetable oil. Soybean oil constitutes about 60% of total world seed oil (http://www.soystats.com) and most of it is used for human consumption.

The fatty acid composition of the soybean oil is approximately 13% of palmitic acid (16:0), 4% stearic acid (18:0), 20% oleic acid (18:1), 55% linoleic acid (18:2), and 7–10% linolenic acid (18:3).[3] The palmitic and stearic acids are saturated fatty acids, and the remaining are unsaturated fatty acids. The lower concentration of polyunsaturated fatty acids (18:3) is always desirable as it reduces the shelf life due to oxidation, which causes an unpleasant odor. Because of the demand from the end users, "The Better Bean Initiative" (BBI) was launched in 2000 by the United Soybean Board (USB) to improve the composition of soybean.[4] As per BBI, the targeted composition of fatty acids in soybean oil includes a higher level of oleic and lower levels of linolenic and saturated fatty acids.[2]

To improve better shelf life and stability of soybean oil at higher temperatures, we targeted on fatty acid desaturase (*FAD3*) gene, responsible for α-linolenic acid synthesis controlled by the three active members: *GmFAD3A*, *GmFAD3B*, and *GmFAD3C*.[5] To reduce the linolenic acid level in soybean, the enzymes that catalyze the conversion of linoleic acids (18:2) to α-linolenic acids (18:3) during fatty acid biosynthesis needed to be repressed. Using RNAi, we silenced the omega-3 FAD gene family in soybean genome and found a lower amount of the α-linolenic acids in the seeds as compared to the control. We also confirmed that the silencing of these genes is inheritable.[5]

Transgenic approach is frequently targeted on single class of component/s. However, this targeted approach may result in an increase or decrease of other traits, which may be desirable/undesirable for end users. In addition, the expression of a gene specifically in a polyploidy like soybean and wheat is not totally independent; rather it is the product of interactions among the genes or diploid genomes in the polyploidy constitutions, as was reported in the case of wheat.[6] Krishnan et al.[7] successfully introduced a leginsulin gene (cysteine-rich protein) in soybean. However, the total amount of sulfur-containing amino acids did

Dr. N. Islam, Dr. S. S. Natarajan
Soybean Genomics and Improvement Laboratory
USDA-ARS
Beltsville, MD 20705, USA
E-mail: savi.natarajan@ars.usda.gov

Dr. P. D. Bates
Institute of Biological Chemistry
Washington State University
Pullman, WA 99164, USA

Dr. K. M. M. John, Dr. D. L. Luthria
Food Composition and Methods Development Laboratory
BHNRC, USDA-ARS
Beltsville, MD 20705, USA

Dr. H. B. Krishnan
Plant Genetics Research Unit
USDA-ARS, University of Missouri
Columbia, MO 65211, USA

Dr. H. B. Krishnan, Dr. Z. J. Zhang
Division of Plant Sciences
University of Missouri
Columbia, MO 65211, USA

DOI: 10.1002/pmic.201800379

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

not change as compared to the control. Similarly, efforts have been made to improve rice grain protein quality by inserting a sunflower seed albumin gene encoding a sulfur-rich protein.[8] A proteomic analysis demonstrated the total amount of sulfur content in the transgenic seeds did not improve because of the competition of sulfur with the endogenous sulfur-rich proteins.[9] Therefore, any disruption in the constituents of a genome either by gene silencing or introducing a foreign gene needs to be investigated thoroughly to elucidate and understand the underlying metabolomic mechanism.

Several successful attempts have been made to improve soybean seed oil compositions by silencing genes that are responsible for encoding fatty acids such as *GmFAD3* which encodes α-linolenic acid content in soybean seed oil.[5,10] In addition, in previous studies, limited information is available on how silencing of *GmFAD3* disrupts the metabolic pathways of FA biosynthesis. In the present study, we investigated the impact of silencing *FAD3* gene in soybean using mass spectrometry approaches to identify changes in protein and fatty acid profiles in transgenic soybeans. Because of the complexity of the seed proteome, we adopted a high-throughput and sensitive tandem mass tag (TMT) technique that can quantify more proteins as compared to classical gel separation techniques.

Significance Statement

Soybean has been a vital source of human food and animal feed due to its protein and oil content. Recently, there has been a significant interest in modifying the fatty acid composition of the soybean oil due to its nutritional, health, and processing significance. Classical breeding and genetic manipulation using biotechnical approaches have been used to improve the quantity and quality of soybean oil. In the present study, we investigated the impact of silencing FAD3 gene in soybean using mass spectrometry approaches to identify changes in protein profile. Because of the complexity of the seed proteome, we adopted a high-throughput and sensitive tandem mass tag (TMT) technique that can quantify more proteins to understand the mechanism of genetic manipulation. The results demonstrated that the fatty acid desaturase gene (*GmFAD3*) silencing affected the key enzymes in FA metabolism. A complete understanding of these processes will aid researchers to alter genes for developing value-added soybeans and help address potential biosafety issue of genetically modified soybean.

2. Experimental Section

2.1. Chemicals

The TMT 6plex was purchased from Thermo Scientific (https://www.thermofisher.com). For fatty acid methyl esters (FAMEs) analyses, the HPLC-grade chemicals (methanol, and acetonitrile) were purchased from Burdick and Jackson (Muskegon, MI, USA). FAMEs standards were purchased from NcChek prep (www.nu-chekprep.com/). All extracts were filtered through PVDF (0.4 μm filter).

2.2. Plant Materials

Soybean cultivar, Jack (control), and four independent FAD3-silenced transgenic lines (*FAD-A*, *FAD-B*, *FAD-C*, and *FAD-E*) derived from the Jack were grown in 13 L pots. The plants were grown in PRO-MIX (Premier Horticulture, Quebec, Canada) medium and fertilized with Osmocot 14-14-14 (Hummert International, Earth City, USA) in a greenhouse (University of Missouri, MO) with supplemental light (intensity 50–90 klux). Greenhouse settings were 16-h day length with 30/18 °C day/night temperatures. Dry seeds were harvested by hand and stored in cold seed storage room.

2.3. Fatty Acid Methyl Ester Extraction and Analysis

The ground transgenic soybean seeds (*FAD-A*, *FAD-B*, *FAD-C*, *FAD-E*) and control were weighed (100 mg) separately and were extracted twice with hexane (5 mL). The extraction was performed under an ultrasonic bath (power 600 W) for a period of 15 min. The extracts were centrifuged at 5000 rpm for 10 min and the supernatant was collected and evaporated to dryness under a slow stream of nitrogen gas. The concentrated soybean oil was resuspended with 2 mL hexane and from that 1 mL was separated and evaporated to dryness for the preparation of FAMEs derivatives by transesterification of extracted soybean oil. The derivatization was performed using 5 mL of acidified methanol (10 mL of acetyl chloride to 90 mL of cold methanol). The mixture kept at ambient temperature overnight with continuous stirring, and then 3 mL of water was added. The FAMEs were extracted with 2 mL of hexane. The hexane layer was separated and analyzed with GC.[11]

2.4. Protein Extraction

Protein extraction from soybean was performed using a phenol extraction protocol as described by Hurkman.[12] Initially, 200 mg of the ground soybean from each line were defatted using hexane.[13] The residue was extracted with approximately 1 mL of the buffer containing sucrose (0.7 M), tris (0.5 M), EDTA (50 mM), KCl (0.1 M), DTT (25 mM), and PMSF (2 mM). The mixture was incubated for 30 min at room temperature with shaking. The mixture was centrifuged at 8000 *g* for 30 min and the supernatant was separated and an equal amount of water saturated phenol was added to it and the mixture was mixed well for 10 min. The phenol phase which contains the proteins were collected after centrifugation at 4 °C for 30 min. The proteins (phenol phase) were precipitated using five times the volume of ammonium acetate (0.1 M) in methanol. This mixture was incubated overnight at −20 °C. Next day, protein pellets were collected after centrifugation at 15 000 *g* for 30 min. The protein pellets were further

© 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

washed with cold acetone (three times) and resuspended in 6 M urea, 100 mM Tris-HCl. The protein concentration was estimated using bicinchoninic acid assay (Pierce, Rockford, IL).

2.5. Protein Digestion and Peptide Labeling

For proteomic analysis, the samples were prepared as described earlier.[14] The urea solubilized proteins were further cleaned by methanol/chloroform precipitation and the protein pellets were suspended in lysis solution which contains 8 M Urea, 1% SDS, 50 mM Tris pH 8.5, protease and phosphatase inhibitors. The protein concentration was measured by microBCA method (Pierce, Thermo Fisher Scientific, MA, USA). Proteins (1 mg) were reduced with DTT and alkylated with iodoacetamide. Digestion was performed overnight at 37 °C using LysC (1:50) followed by trypsin (1:100; enzyme:protein). The resultant peptides were purified by reverse phase as previously described and the peptide concentration was measured using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific). An equal amount (100 µg) of peptide was labeled with TMT 6-plex reagents per manufacturer instructions (Thermo Fisher Scientific). A ratio test was performed using equal amount for each sample. The normalized intensities for each sample from the ratio check were mostly similar. However, based on the ratio check some adjustments were made before mixing (Thermo Fisher Scientific). The mixed labeled samples were further fractionated. LC-MS3 data collection strategy was used to analyze 12 of the 24 peptide fractions from the basic reverse phase step (every other fraction).[15] An Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) equipped with a Proxeon Easy nLC 1000 were used for online sample handling and peptide separations.

2.6. Mass Spectrometry and Data Analysis

The peptides were (≈5 µg) suspended in 5% formic acid with 5% acetonitrile for further analysis. It was loaded onto a fused-silica micro capillary (100 µm inner diameter) with a needle tip pulled to an internal diameter less than 5 µm. The column was packed to a length of 35 cm with a C18 reverse phase resin (GP118 resin 1.8 µm, 120 Å, Sepax Technologies). The peptides separation was achieved using a 180-min linear gradient from 3% to 25% buffer B containing 100% ACN + 0.125% formic acid equilibrated with buffer A containing 3% ACN + 0.125% formic acid with 400 nL min^{-1} flow rate. Fusion Orbitrap with an MS1 spectrum (Orbitrap analysis, resolution 120 000, 400−14 000 m/z scan range with quadrupole isolation, AGC target 1 × 106, maximum injection time 100 ms, dynamic exclusion of 60 s) was used for scanning sequence. The top ten fragment ion precursors from MS1 and MS2 scan were selected for MS3 analysis (synchronous precursor selection), in which precursors were fragmented by HCD prior to Orbitrap analysis (NCE 55, max AGC 1.5 × 105, maximum injection time 150 ms, MS2 quadrupole isolation was set to 2.5 Da, resolution 50 000).

Data analysis was done with an in-house software tool. This includes file processing (RAW file); controlling peptide; protein level false discovery rates; assembling proteins from peptides; protein quantification from peptides. These steps were described earlier.[14] A Uniprot soybean database (2016) was used for MS/MS spectral search with both the forward and reverse sequences. Database search criteria used are as follows: tryptic with two missed cleavages, a precursor mass tolerance of 50 ppm, fragment ion mass tolerance of 1.0 Da, static alkylation of cysteine (57.02146 Da), static TMT labeling of lysine residues and N-termini of peptides (229.162932 Da), and variable oxidation of methionine (15.99491 Da). The intensity of TMT reporter ion was calculated with a 0.003 Da window around the theoretical m/z for each reporter ion in the MS3 scan. For quantification, poor quality of MS3 spectra were discarded (<100 summed signal-to-noise across six channels and <0.5 precursor isolation specificity). These mass spectrometry data files were submitted to massive.ucsd.edu (Accession # MassIVE MSV000082278).

Peptide spectra were considered for quantitative analyses if they are detected in the three independent biological replicates. The protein quantitative values for each channel was initially normalized based on the total spectral count in each channel and the normalized value was then scaled to 100. A t-test was performed to identify differentially expressed proteins and the Benjamini and Hochberg correction was applied to limit false discovery to ≤0.05.

Annotations for the uncharacterized proteins were extracted by submitting the full-length sequences to Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg) databases and to the MEta Sever for Sequence Analysis (MESSA) (http://prodata.swmed.edu/MESSA/MESSA.cgi) to determine their functional annotation.[16] With a given protein sequence, MESSA utilizes several tools to predict the local sequence properties such as signal peptides and transmembrane helices. In addition, it predicts homologous proteins along with their functional annotations.

2.7. Mapping the Differentially Expressed Protein in Metabolic Pathways

The full-length sequences of the differentially expressed proteins were retrieved from the NCBI (https://www.ncbi.nlm.nih.gov/protein/). The exported FASTA files were submitted to KEGG pathways to annotate sequence (http://www.kegg.jp/kegg/tool/annotate_sequence.html) using 3803 as family and the subsequent pathways were visualized for enrichment using KEGG identifier. Proteins/enzymes with no equivalent KEGG identifier could not be mapped on the metabolic pathways.

3. Results

3.1. Effect of *GmFAD3* Silencing on the Polyunsaturated Fatty Acids

To affirm our previous finding, we estimated the fatty acid profiles of the oil extracted from the transgenic lines and the control soybean seeds. The oil was transesterified with methanolic-HCl and the FAMEs were extracted and analyzed using GC-FID (gas chromatography–flame ionization detector). Nine replicate extractions and analyses were carried out with each set of sample. The results are presented in **Figure 1**A. Five predominant fatty

© 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

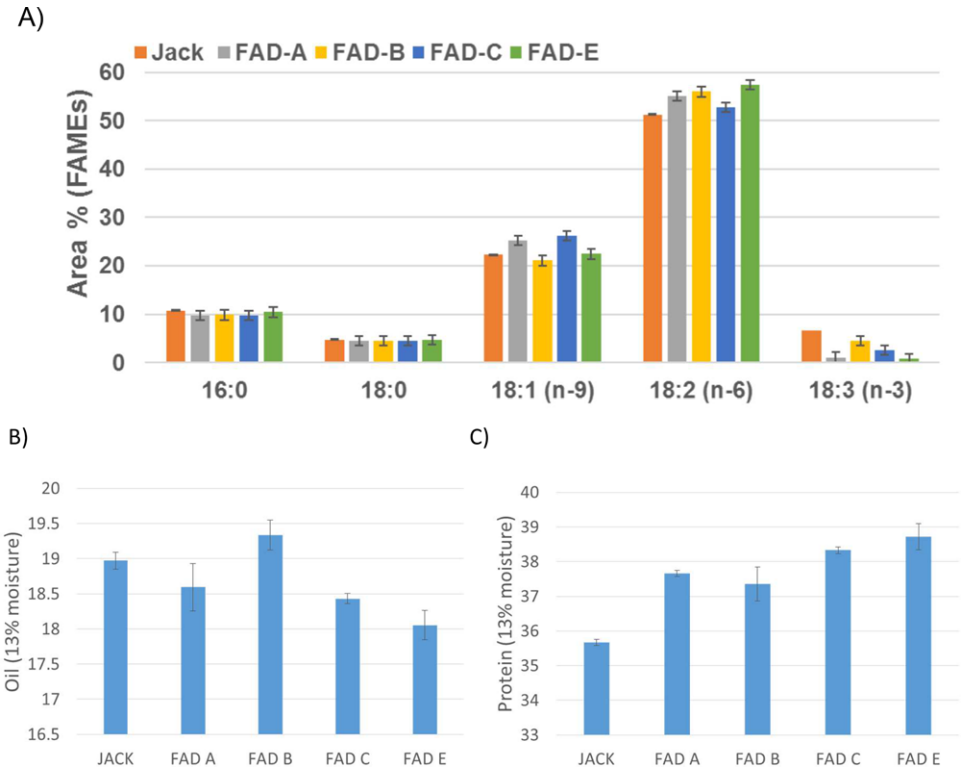

Figure 1. A) Fatty acid methyl esters analysis by gas chromatography–flame ionization detector of the transesterified oil extracted from five sets of soybean samples; B) total oil content of the five soybean samples and; C) total protein content of the five soybean samples determined by NIR.

acids which constituted over 95% of the fatty acids were identified by comparing the retention time with an authentic commercial standard as palmitic, stearic, oleic, linoleic, and linolenic acids. Results revealed that except linolenic acid, all other acids showed insignificant variation between the transgenic lines and the control soybean samples. The palmitic, stearic, oleic, and linoleic acid content varied in ranges 9.7–10.9, 4.6–4.9, 20.9–26.4, and 51.1–57.6%, respectively. However, the linolenic acid content varied significantly between control (6.7%) and transgenic samples (0.9–4.5%). The average linolenic acid content in control soybean sample was 6.7% and the average linolenic acid content in transgenic lines *FAD-A*, *FAD-B*, *FAD-C*, and *FAD-E* was determined as 1.1, 4.5, 2.5, and 0.9%, respectively.

3.2. Effect of GmFAD3 Silencing on the Seed Oil and Protein Accumulation

In addition to a change in fatty acid composition, the *GmFAD3* silencing caused a variable response to total fatty acid accumulation as oil in the seeds. The *FAD-B* line had a slight increase in total oil, and the *FAD-E* line had the largest decrease in total oil which was reduced from approximately 19% in the control to approximately 18% in *FAD-E* (Figure 1B). All lines had an increase in protein content, and the *FAD-E* line had the largest increase in protein content, which increased from approximately 35% in the control to 38% in *FAD-E* (Figure 1C). Unlike the oil content, all lines exhibited the increased content of proteins with *FAD-E* being the highest (Figure 1C).

3.3. Effect of *GmFAD3* Silencing on the Protein Profile

To identify the effect of *GmFAD3* silencing on the protein profile, a non-gel quantitative proteomic approach was applied using TMT tag. Isolated proteins from both the control and transgenic soybean (*FAD-E*) seeds were subjected to trypsin and lysC digestion and labeled with TMT reagents followed by mass spectrometry analyses. The *FAD-E* line was chosen for detailed analysis because it had the significant changes in fatty acid composition and protein content as compared to other transgenic events (Figure 1). Hereafter, *FAD-E* line designated as transgenic line throughout the manuscript. Using the SEQUEST algorithm against a Uniprot composite database, we detected 59822 peptides with less than 1% false discovery rate (FDR). These peptides were then grouped into 7429 proteins, of which 6074 proteins were quantified from peptides with a summed (SN) threshold of ≥100 and isolation specificity of 0.5. To eliminate outliers because of the technical issues, we performed a Spearman correlation coefficient analyses and found a correlation between the control and the transgenic line (**Figure 2**A). The control (Jack) and the transgenic (*FAD-E*) soybean samples showed a distinct variation. The blue color indicates the three replicates of the control (J1, J2, J3) samples and the transgenic samples (F1, F2, F3) are shown in red. This analysis showed no outliers between the two sets of soybean samples. A separate heat map was plotted to visualize the protein changes between the control and the transgenic line (Figure 2B). The high (red) and low (green) abundance proteins were represented in the 30% relative abundance scale. Correlation and clustering showed differential protein abun-

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

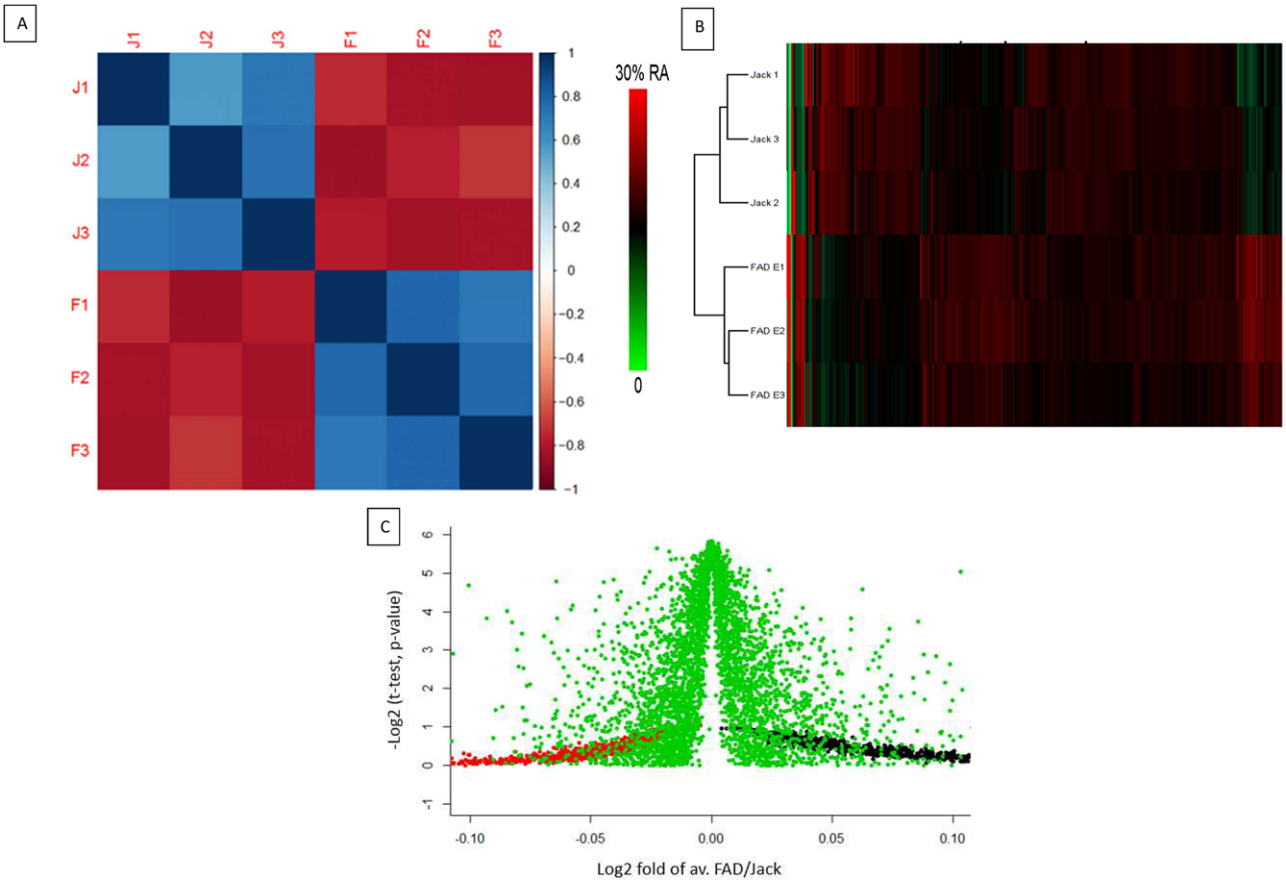

Figure 2. Changes in protein profile of transgenic soybean F(*FAD-E*) compared to control J (Jack). A) Correlation heat map between control and transgenic, B) hierarchical clustering analyses showing distinct variations between control and transgenic, C) scatter plot of total proteins identified: Green represents non-significant proteins, red represents higher abundance, and black represents lower abundance.

dance between the control and the transgenic line (Figure 2A,B). Of the 6074 proteins identified (Table S1, Supporting Information), 1036 showed higher abundance (Table S2, Supporting Information). These proteins were associated with photosynthetic pathways, glycolysis, etc. However, 1079 proteins exhibited significantly ($\leq$0.05) lower abundance (Table S3, Supporting Information) in the transgenic line when compared to the control. These proteins are linked with fatty acids and secondary metabolites. The changes of proteins were also mapped on the global metabolic pathways (**Figure 3**). These proteins are also shown in the scatter plot (Figure 2C). The diverse changes of protein expression on the global metabolic pathways in the transgenic line imply that the silencing of *GmFAD3* (FAD) has perturbed carbon partitioning from oil pool to the protein pool. In addition to the significant proteins as described above, the non-significant proteins are shown in the Table S4, Supporting Information.

The silencing of *GmFAD3* produced diverse changes to many central metabolic pathways including glycolysis, pentose phosphate pathway, and oxidative phosphorylation (**Table 1**). Each pathway had enzymes which had significant increases or decreases in abundance which likely reflects compensating regulation of all of metabolism as a response to the reduction of linolenic acid production.

The abundance of several key enzymes on fatty acid metabolic pathways were also changed due to the gene silencing (**Table 2**). The largest-fold increase in protein abundance for fatty acid metabolism was two of the four subunits of acetyl-CoA carboxylase. Acetyl-CoA carboxylase (ACCase) is the committed step and major regulatory enzyme that controls de novo fatty acid synthesis and total oil accumulation. The heteromeric ACCase is composed of four subunits: alpha- and beta-carboxyltransferase, biotin carboxyl carrier protein, and biotin carboxylase.[17] These subunits also associate as two subcomplexes made up of the alpha- and beta-carboxyltransferases together, and the biotin carboxyl carrier protein with the biotin carboxylase.[18,19] Quantitative proteomics of *Arabidopsis thaliana* ACCase demonstrated non-stochiometric accumulation of the four subunits, and indicated that modulation of functional subcomplex levels is part of ACCase regulation.[20] Here we found that the biotin carboxyl carrier protein (accB) and the biotin carboxylase (accC) were specifically increased in the *FAD-E* line (accB 2.75-fold, and accC 1.67-fold) (Table 2). Recently, new components of the ACCase complex (biotin attachment domain containing proteins, BADCs) were identified in *Arabidopsis thaliana* which do not contain biotin and act as negative regulators of ACCase by replacing the biotin containing accB protein in the complex with a BADC protein which produces a non-functional enzyme.[19] *Arabidopsis*

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

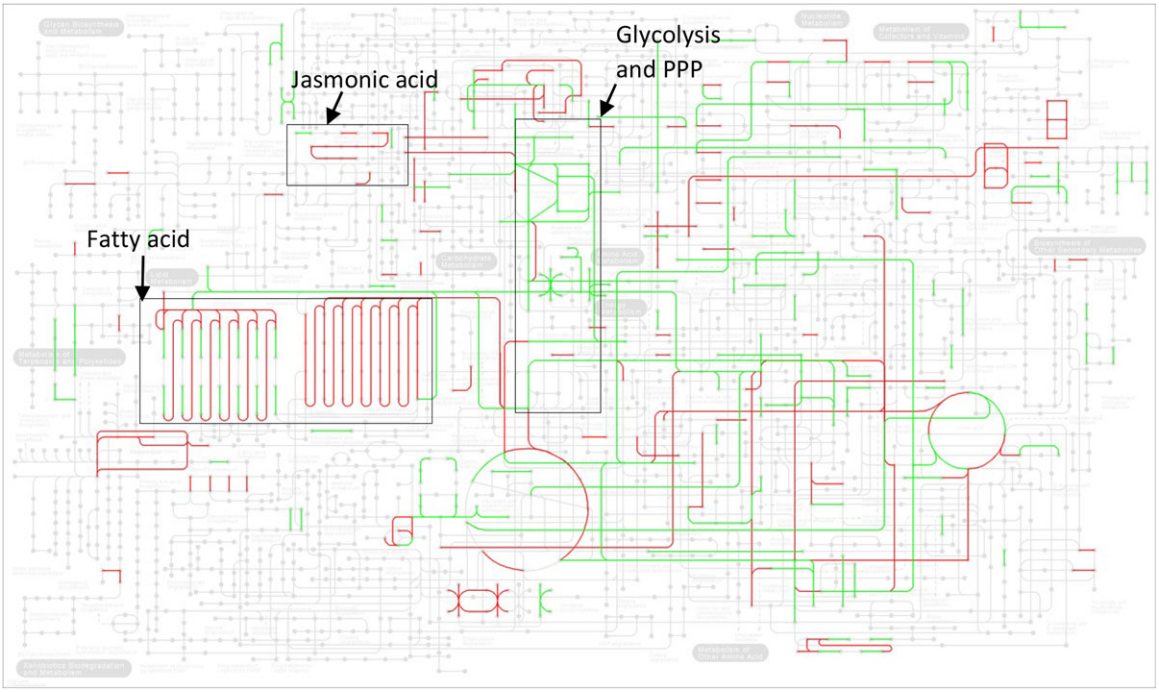

Figure 3. Mapping of differentially expressed proteins on the global metabolic pathways. Red represents downregulated and green represents upregulated.

encodes three genes for these negative regulators which have ≈24–29% amino acid identity to *Arabidopsis* accB. We found homology of all the BADC genes in the soybean database. The NCBI accession for each protein was submitted to UniProt to retrieve UniProt accession. The UniProt accession was searched among the statistically significant proteins from our experiments. But only one gene match was found annotated as "biotin carboxyl carrier protein of acetyl-CoA carboxylase (XP_006590336.1 = I1LMV7)" which has 64% similarity with BADC1 of *Arabidopsis*. This protein was increased by 1.51-fold in transgenic compared to the control line.

Enzymes downstream from ACCase that are part of the type II fatty acid synthase complex of plastids involved in de novo fatty acid synthesis were both increased (fabG, fabI, fabD) and decreased (fabH and fabF) (Table 2). We also mapped the differentially expressed enzymes on FA elongation and α-linolenic acid metabolic pathways. As shown in **Figure 4**, two enzymes; very-long-chain 3-oxoacyl-CoA reductase [EC: 1.1.1.330] and very-long-chain enoyl-CoA reductase [EC:1.3.1.93] exhibited increased abundance and acyl-coenzyme A thioesterase 1/2/4 [EC:3.1.2.2] showed decreased abundance in the transgenic line when compared to the control.

We also observed the differential expression of several key enzymes in the α-linolenic acid metabolism (**Figure 5**). These include lipoxygenase [EC:1.13.11.12] and alcohol dehydrogenase class-P [EC:1.1.1.1] which exhibited higher abundance and, several enzymes of β-oxidation such as 12-oxophytodienoic acid reductase [EC:1.3.1.42], acetyl-CoA acyltransferase [EC:2.3.1.16], ACX-acyl-CoA oxidase [EC:1.3.3.6], and MFP2-enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.171.1.1.35 1.1.1.211], exhibited significantly lower abundance when compared to the control.

Seven proteins known to be associated with the storage (glycinin G4 precursor, β-conglycinin), antinutritional (Kunitz-type elastase inhibitor; Kunitz family trypsin, and protease inhibitor protein), and allergens (P34 probable thiol protease precursor) were upregulated in transgenic soybeans as compared to the control (Table S5, Supporting Information). One soybean protein (accession number, C6T3L5) showed a significant reduction (≈98%) in transgenic soybean as compared to the control (Table S3, Supporting Information). The peptide spectrum counts were 1279.93 in control compared to 21.01 in transgenic soybeans. This protein was initially not characterized in UniProt database but showed sequence similarities (confident) to Gly m 4 and belongs to the Bet_V_1 (birch pollen) homologous superfamily.

4. Discussion

In our previous study, we confirmed that silencing *GmFAD3* significantly reduced α-linolenic acids in soybean.[5] In this study, we affirmed the results by analyzing fatty acids methyl ester (FAMEs) profiles of the oil extracted from soybean seeds (Figure 1). The GC-FID profiles of the FAMEs content were similar to those published earlier.[5] The percentage of linolenic acid in transgenic soybean was significantly lower than the control soybean (Jack). Based on the FAMEs analysis, we selected *FAD-E* transgenic line which showed lowest linolenic acid content (0.9%) as compared to the control soybean samples (6.7%). We also demonstrated that the lines with the lowest level of α-linolenic acid had reduced total oil content (Figure 1B). Total protein content in the five soybean lines as determined by NIR analysis is shown in Figure 1C. Detailed proteomic profile comparison was carried out for two sets of samples in three

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Abundance of enzymes mapped in the glycolytic, pentose phosphate pathways, and oxidative phosphorylation pathways.

EC no.	Gene	Enzyme name	FAD/jack (fold change)	*p*-value
Glycolytic pathway				
1.1.1.1	frmA	S-Glutathione dehydrogenase/alcohol dehydrogenase	0.59	0.0056
1.2.1.9	gapN	Glyceraldehyde-3-phosphate dehydrogenase NADP+	0.81	0.0023
1.2.1.12	GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	1.23	0.0082
1.2.4.1	PDHA	Pyruvate dehydrogenase E1 component alpha subunit	1.44	0.0015
2.3.1.12	DLAT	Pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase)	1.20	0.0115
2.7.1.11	pfkA	6-Phosphofructokinase 1	0.82	0.0036
2.7.1.90	pfp	Diphosphate-dependent phosphofructokinase	1.13	0.0127
2.7.2.3	PGK	Phosphoglycerate kinase	1.21	0.0141
5.1.3.3	galM	Aldose 1-epimerase	0.82	0.0036
5.1.3.15		Glucose-6-phosphate 1-epimerase	1.39	0.0073
5.3.1.1	TPI	Triosephosphate isomerase	1.29	0.0108
5.3.1.9	GPI	Glucose-6-phosphate isomerase	1.22	0.0050
5.4.2.2	pgm	Phosphoglucomutase	1.20	0.0093
5.4.2.12	gpmI	2,3-Bisphosphoglycerate-independent phosphoglycerate mutase	1.09	0.0174
1.1.1.1	ADH1	Alcohol dehydrogenase class-P	1.35	0.0028
Pentose phosphate pathway				
1.1.1.49	G6PD	Glucose-6-phosphate 1-dehydrogenase	1.38	0.0044
1.2.1.9	gapN	Glyceraldehyde-3-phosphate dehydrogenase NADP+	0.81	0.0023
2.2.1.1		transketolase	1.40	0.0097
2.2.1.2		Transaldolase	1.60	0.0030
2.7.1.11	pfkA	6-Phosphofructokinase 1	0.79	0.0066
2.7.1.15	rbsK	Ribokinase	0.89	0.0152
2.7.1.90	pfp	Diphosphate-dependent phosphofructokinase	1.13	0.0127
5.3.1.9	GPI	Glucose-6-phosphate isomerase	1.22	0.0050
5.4.2.2	pgm	Phosphoglucomutase	1.20	0.0093
Oxidative phosphorylation pathway				
1.10.2.2	UQCRFS1	Ubiquinol-cytochrome c reductase iron–sulfur subunit	0.79	0.00826
	QCR7	Ubiquinol-cytochrome c reductase subunit 7	0.85	0.01085
3.6.3.6	PMA1	H+-transporting ATPase	1.22	0.00777
3.6.3.14	ATPeF1B	F-type H+-transporting ATPase subunit beta	2.07	0.00790
	ATPF1G	F-type H+-transporting ATPase subunit gamma	1.96	0.00674
3.6.3.14	ATPeF1B	F-type H+-transporting ATPase subunit beta	0.90	0.01162
	ATPeF0O	F-type H+-transporting ATPase subunit O	0.88	0.01079
	ATPeF0D	F-type H+-transporting ATPase subunit D	0.89	0.00714
	ATPeFG	F-type H+-transporting ATPase subunit G	0.89	0.00765
	ATPeV1C	V-type H+-transporting ATPase subunit C	0.87	0.00952
	ATPeV0A	V-type H+-transporting ATPase subunit A	0.93	0.01483
1.6.5.3	NDUFS1	NADH dehydrogenase (ubiquinone) Fe–S protein 1	0.86	0.00286
	NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit	0.89	0.01101

biological replicates to investigate the possible correlations of proteins involved in fatty acid biosynthesis as affected by *GmFAD3* silencing.

Fatty acid synthesis is initiated by the carboxylation of acetyl-CoA to malonyl-CoA by acetyl-CoA caroboxylase (ACCase) which is a regulatory point in fatty acid biosynthesis. The active enzyme is made of four subunits,[19,21] the abundance of two of these subunits (accB and accC) which make up a subcomplex were increased by silencing *GmFAD3* (Table 2). In addition, a possible negative regulator (BADC1) of the soybean ACCase which replaces accB in the ACCase holoenzyme was also increased. In the subsequent two reactions of fatty acid synthesis, the protein abundances of the enzymatic steps were both increased and decreased. The first malonate is transferred from a thioester of CoA to a thioester of acyl carrier protein (ACP) by acyl-carrier-protein s-malonyltransferase (fabD), this enzyme was increased by 1.55-fold (Table 2). Subsequently, the malonyl-ACP is condensed with the growing acyl chain by a 3-keto-acyl carrier protein synthase (KAS). Plants have three KAS enzymes, of which KASIII (fabH) and KASI (fabF) are responsible for the condensation steps up

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. List of enzymes associated with fatty acid metabolism.

Accession	Protein name	FAD/jack (fold change)	*p*-value
A0A0R0F5Q2	fabG; 3-oxoacyl-[acyl-carrier protein] reductase4	1.24	0.0100
I1N1×0	fabI; enoyl-[acyl-carrier protein] reductase I	1.10	0.0155
I1MZV4	fabD; [acyl-carrier-protein] S-malonyltransferase	1.55	0.0022
O81273	accC; acetyl-CoA carboxylase, biotin carboxylase subunit	1.67	0.0019
I1LMV7	Biotin carboxyl carrier protein of acetyl-CoA carboxylase 64% homology to AT BADC1	1.51	0.0065
Q9GE06	Biotin carboxyl carrier protein/accB, acetyl-CoA carboxylase	2.75	0.0087
C6TME1	HSD17B12; long-chain 3-oxoacyl-CoA reductase	1.16	0.0114
I1JDE1	TER; very-long-chain enoyl-CoA reductase	1.45	0.0102
B0M1A9	Peroxisomal 3-ketoacyl-CoA thiolase	0.70	0.0017
I1KIP9	Peroxisomal fatty acid beta-oxidation multifunctional protein AIM1	0.71	0.0073
Q945U3	ACOX; acyl-CoA oxidase	0.72	0.0039
Q9M4R7	fabH; 3-oxoacyl-[acyl-carrier-protein] synthase III/SOYBN Beta-ketoacyl-acyl carrier protein synthase III	0.93	0.0141
Q9M508	fabF; 3-oxoacyl-[acyl-carrier-protein] synthase/SOYBN Beta-ketoacyl-ACP synthetase I	0.91	0.0041

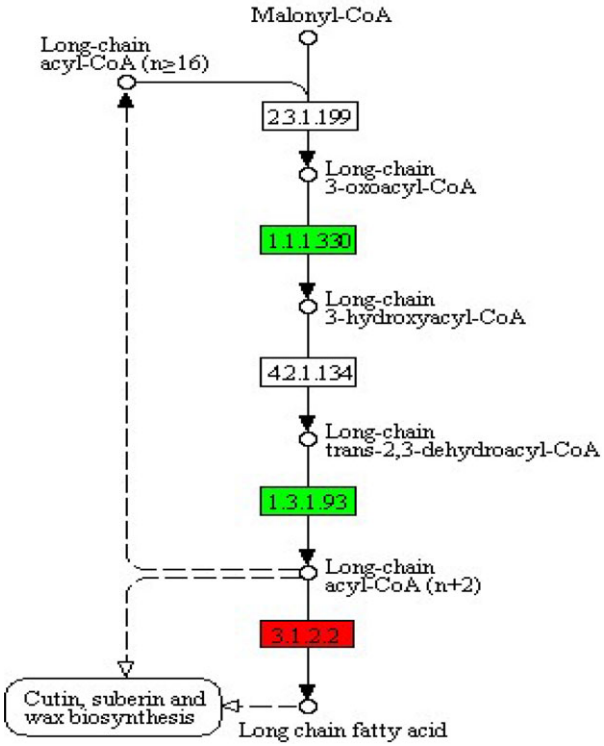

Figure 4. Effect of *Gm FAD3* silencing on enzymes in the FA elongation in the ER based on KEGG narration. Proteins are indicated by enzyme numbers. Red represents enzymes that are significantly downregulated ($p < 0.05$) in the transgenic line, while green are upregulated. Very-long-chain 3-oxoacyl-CoA reductase [EC:1.1.1.330], very-long-chain enoyl-CoA reductase [EC:1.3.1.93], acyl-coenzyme A thioesterase 1/2/4 [EC:3.1.2.2].

to a 16 carbon chain.[5,21] The abundance of both fabH and fabF were decreased by 0.93 and 0.91-fold, respectively. Considering that abundance of only part of the ACCase enzyme complex was increased and so was a negative inhibitor of the complex, it is unlikely that the increase in accB and accC led to more production of malonyl-CoA for fatty acid synthesis, and the increase in BADC1 may have reduced total ACCase activity. Together, with the reduction in KAS enzyme abundance, this change may have contributed to the reduced oil content of the *FAD-E* line.

The silencing of *GmFAD3* specifically reduces production of α-linolenic acid. While this fatty acid is a significant component of wild-type soybean oils, it also has other metabolic functions in plants. One function is the metabolic precursor to the stress response hormone jasmonic acid (JA).[22] We measured both increases and decreases in the abundance of enzymes involved in JA synthesis (Figure 4). The lipoxygenase [EC:1.13.11.12] involved in the first step of linolenic acid utilization for JA biosynthesis was increased, whereas many of the latter steps of JA biosynthesis involving β-oxidation of intermediates were decreased (for example, MFP2 and ACOX; Table 2, Figure 4). Together, these results suggest that the plant may be responding to the lack of α-linolenic acid for JA biosynthesis by upregulating the initial steps, but with limited substrate flux through JA biosynthesis the latter steps are downregulated.

From our investigation, it was obvious that silencing *GmFAD3* shifted the carbon-substrate flux from oil to proteins (Figure 1B,C). While the phenomenon of increased protein causes decreased oil content is reported elsewhere (negative correlation),[23] we demonstrated that the silencing of *GmFAD3* is directly related to a shift of the carbon flux. Increased protein/oil content in soybean have been an area of interest for developing value added soybeans. This research has been carried out using genetic gains strategies through conventional breeding or detailed molecular mapping of the genes associated with protein and oil modification.[24] The major quantitative trait loci, chromosomes (Chr.) 20 (LG-I), and 15 (LG-E) contained (QTL), for soybean proteins are reported.[24] While some improvement has been achieved through genetic gains strategy it has several challenges. For instance, the domesticated soybean has 75% of the genes with multiple copies.[25] The epigenetic complexity might result in epistatic interaction (non-additive),[26] leading to the limit of gene expression when genetic gains strategy through conventional breeding is considered. From our study, we have demonstrated that the manipulation of the carbon metabolic

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

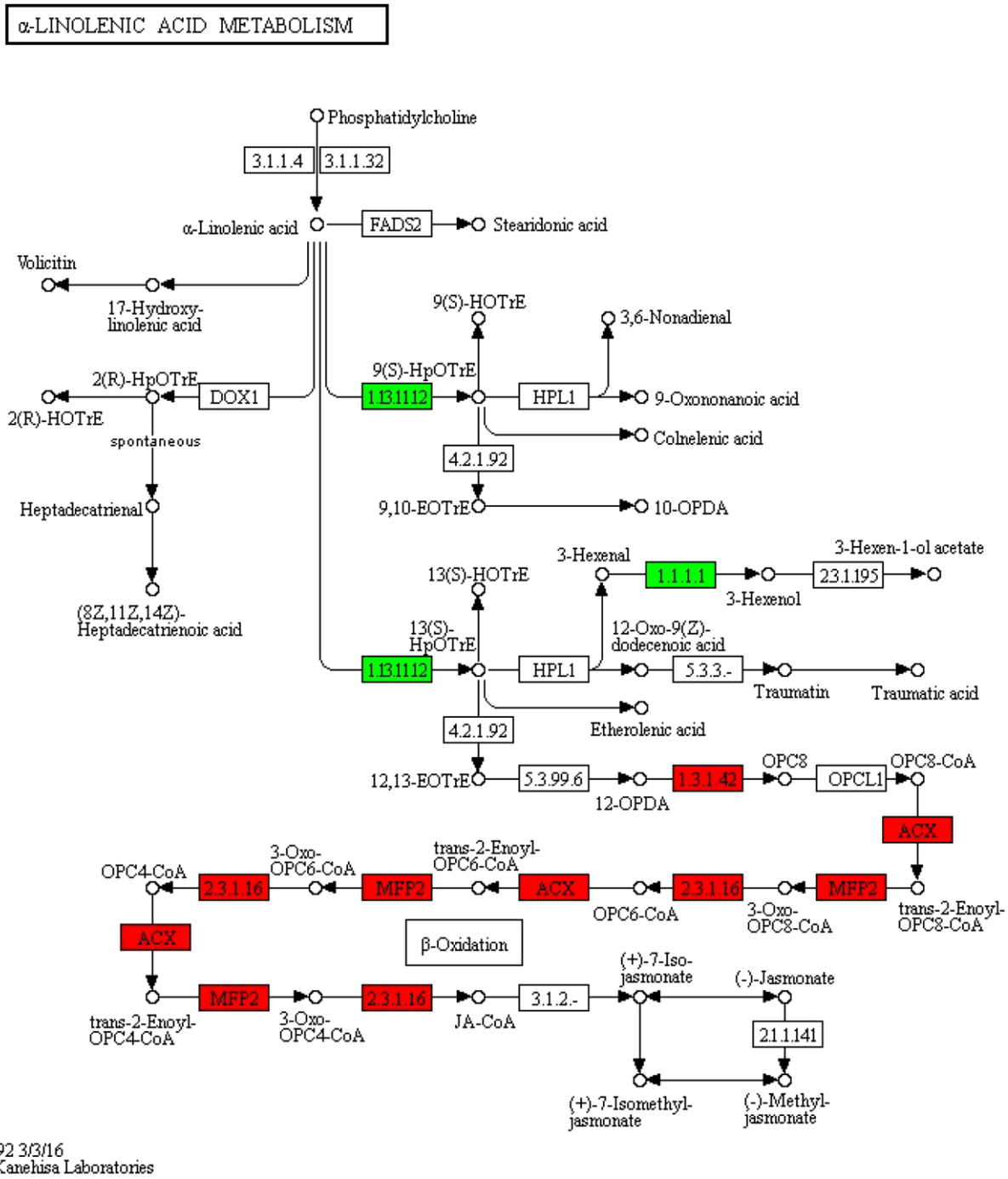

Figure 5. Effect of *Gm FAD3* silencing on enzymes in the α-linolenic metabolism based on KEGG narration. Proteins are indicated by enzyme numbers. Red represents enzymes that are significantly downregulated ($p < 0.05$) in the transgenic line, while green represents upregulated. lipoxygenase [EC:1.13.11.12], alcohol dehydrogenase class-P [EC:1.1.1.1], 12-oxophytodienoic acid reductase [EC:1.3.1.42], ACX [E1.3.3.6, ACOX1, ACOX3], MFP2 [enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [EC:4.2.1.17 1.1.1.35 1.1.1.211]], acetyl-CoA acyltransferase [EC:2.3.1.16], acyl-CoA oxidase [EC:1.3.3.6], acetyl-CoA acyltransferase [EC:2.3.1.16], acetyl-CoA acyltransferase [EC:2.3.1.16].

pathways may be utilized to achieve either the higher protein or oil content in soybean. Although silencing *GmFAD3* shifted the carbon-substrate flux from oil to proteins, we did not observe a significant difference in the total biomass between the control and the transgenic soybean.

Although soybean seed is considered as an abundant source of protein, it also contains several antinutritional factors (ANF) such as agglutinin, protease inhibitors, phytic acid, tannins, and allergenic proteins.[27] We identified storage/allergen/ANF proteins including β-conglycinin, glycinin, 2S albumin, P34

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

probable thiol protease precursor, and Kunitz trypsin protease inhibitor that were upregulated in the transgenic soybeans (Table 2). The abundance and variation of these proteins found in the transgenic line were similar to the variation found in wild and cultivated soybeans and soybeans cultivars grown under different environments.[28] In addition, one protein showed over 90% reduction in the transgenic soybean as compared to the control. This protein was initially uncharacterized using Uniprot database. However, a Blast search revealed sequence similarities to Gly m 4 allergen which is the member of Bet_V_1 family domain.[29] Gly m 4 is also reported as a member of the pathogenesis-related proteins which expressed in stress-related conditions such as in response to wounding, pathogen infection, plant hormones, and several environmental factors.[29] The reduction of Gly m 4 (C6T3L5) allergen in the transgenic lines might have better food quality value for soybean consumers sensitive to soybean allergens. Recently, in another study, Geng et al. reported level of five soybean allergens such as Gly m 4, Gly m 5, Gly m 6, Gly m Bd 28K, and Gly m Bd 30K from 604 soybean samples collected from North/South Americas over five growing seasons (2009–13/2014).[30] The authors reported 5- to 19-fold variations among the five allergens. Geng et al.(2015) developed a Sandwich ELISA technique and quantified soybean allergen, Gly m 4 in 128 soybean samples that were grown at eight different geographical locations.[31] The authors reported a 13-fold difference in the amounts of Gly m 4 and concluded that significant sources of variability in Gly m 4 levels in the conventional varieties were related to location and variety. The differentially expressed proteins in our study could be from positional integration events as described in earlier studies.[5,30,31] However, additional research is needed to further investigate this observation.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors would like to thank Dr. Ryan Kunz, Thermo Center at Harvard Medical School for performing the quantitative protein analyses reported in this study. Funding for this research was provided by Agricultural Research Service, USDA. Mention of trade name, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

low linolenic, mass spectrometry, proteome, soybean, transgenic

Received: October 26, 2018
Revised: February 6, 2019

[1] K. Roesler, B. Shen, E. Bermudez, C. Li, J. Hunt, H. G. Damude, K. G. Ripp, J. D. Everard, J. R. Booth, L. Castaneda, L. Feng, K. Meyer, *Plant Physiol.* **2016**, *171*, 878.
[2] A.-T. Pham, J.-D. Lee, J. G. Shannon, K. D. Bilyeu, *Theor. Appl. Genet.* **2011**, *123*, 793.
[3] W. Goettel, E. Xia, R. Upchurch, M.-L. Wang, P. Chen, Y.-Q. C. An, *BMC Genomics* **2014**, *15*, 299.
[4] D. Durham, *AgBioForum* **2003**, *6*, 23.
[5] T. Flores, O. Karpova, X. Su, P. Zeng, K. Bilyeu, D. A. Sleper, H. T. Nguyen, Z. J. Zhang, *Transgenic Res.* **2008**, *17*, 839.
[6] N. Islam, H. Tsujimoto, H. Hirano, *Proteomics* **2003**, *3*, 549.
[7] H. B. Krishnan, W.-S. Kim, N. W. Oehrle, A. A. Alaswad, I. Baxter, W. J. Wiebold, R. L. Nelson, J. Agric. *Food Chem.* **2015**, *63*, 2862.
[8] N. D. Hagan, N. Upadhyaya, L. M. Tabe, T. J. Higgins, *Plant J.* **2003**, *34*, 1.
[9] N. Islam, N. M. Upadhyaya, P. M. Campbell, R. Akhurst, N. Hagan, T. J. V. Higgins, *Phytochemistry* **2005**, *66*, 2534.
[10] T. Hoshino, S. Watanabe, Y. Takagi, T. Anai, *Breeding Sci.* **2014**, *64*, 371.
[11] Q. Chen, D. L. Luthria, H. Sprecher, *Arch. Biochem. Biophys.* **1998**, *349*, 371.
[12] W. J. Hurkman, C. K. Tanaka, *Plant Physiol.* **1986**, *81*, 802.
[13] R. C. Hill, T. J. Oman, X. Wang, G. Shan, B. Schafer, R. A. Herman, R. Tobias, J. Shippar, B. Malayappan, L. Sheng, A. Xu, J. Bradshaw, J. Agric. *Food Chem.* **2017**, *65*, 5531.
[14] M. P. Weekes, P. Tomasec, E. L. Huttlin, C. A. Fielding, D. Nusinow, R. J. Stanton, E. C. Wang, R. Aicheler, I. Murrell, G. W. Wilkinson, P. J. Lehner, S. P. Gygi, *Cell* **2014**, *157*, 1460.
[15] G. C. McAlister, D. P. Nusinow, M. P. Jedrychowski, M. Wühr, E. L. Huttlin, B. K. Erickson, R. Rad, W. Haas, S. P. Gygi, *Anal. Chem.* **2014**, *86*, 7150.
[16] Q. Cong, N. V. Grishin, *BMC Biol.* **2012**, *10*, 82.
[17] P. D. Bates, S. Stymne, J. Ohlrogge, *Curr. Opin. Plant Biol.* **2013**, *16*, 358.
[18] M. J. Salie, J. J. Thelen, *Biochim. Biophys. Acta* **2016**, *1861*, 1207.
[19] M. J. Salie, N. Zhang, V. Lancikova, D. Xu, J. J. Thelen, *Plant Cell* **2016**, *28*, 2312.
[20] R. S. Wilson, J. J. Thelen, *J. Proteome Res.* **2018**, *17*, 1773.
[21] J. Ohlrogge, J. Browse, *Plant Cell* **1995**, *7*, 957.
[22] A. J. Koo, *Phytochem. Rev.* **2018**, *17*, 51.
[23] F. R. Wilson, in *Soybeans: Improvement, Production, and Uses* Ed: J. R. Wilcox, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI **2004**, p. 621.
[24] G. Patil, R. Mian, T. Vuong, V. Pantalone, Q. Song, P. Chen, G. J. Shannon, T. C. Carter, H. T. Nguyen, *Theor. Appl. Genet.* **2017**, *130*, 1975.
[25] J. Schmutz, S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T. Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.-C. Zhang, K. Shinozaki, H. T. Nguyen, R. A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R. C. Shoemaker, S. A. Jackson, *Nature* **2010**, *463*, 178.
[26] L. Comai, *Nat. Rev. Genet.* **2005**, *6*, 836.
[27] M. H. Pedersen, T. Holzhauser, C. Bisson, A. Conti, L. B. Jensen, P. S. Skov, C. Bindslev-Jensen, D. S. Brinch, L. K. Poulsen, *Mol. Nutr. Food Res.* **2008**, *52*, 1486.

 © 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[28] S. Natarajan, F. Khan, Q. Song, S. Lakshman, P. Cregan, R. Scott, E. Shipe, W. Garrett, J. Agric. *Food Chem.* **2016**, *64*, 1433.
[29] C. Radauer, H. Breiteneder, *J. Allergy Clin. Immunol.* **2007**, *120*, 518.
[30] T. Geng, D. Stojšin, K. Liu, B. Schaalje, C. Postin, J. Ward, Y. Wang, Z. L. Liu, B. Li, K. Glenn, J. Agric. *Food Chem.* **2017**, *65*, 463.
[31] T. Geng, K. Liu, R. Frazier, L. Shi, E. Bell, K. Glenn, J. M. Ward, J. Agric. *Food Chem.* **2015**, *63*, 4947.

© 2019 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table S5. Spectral count of protein associated with food values between control(Jack) and a transgenic event (FAD). Proteins were sorted based on functional categories as food value associated.

Accession	Protein Name	Normalized spectral count (no.)		
		Jack	FAD	FAD/Jack
Q9SB11	glycinin G4 precursor	17006.74	24906.90	1.46
F7J077	beta-conglycinin	12810.94	17624.72	1.38
I1LE33	beta-conglycinin	23305.35	28539.47	1.22
Q4LER6	beta-conglycinin	164764.64	212947.97	1.29
K7MIV6	kunitz-type elastase inhibitor BrEI-like	1292.66	1529.25	1.18
B1ACD3	kunitz family trypsin and protease inhibitor protein	1310.91	1678.69	1.28
O64458	P34 probable thiol protease precursor	66730.40	78930.66	1.18
C6T3L5	Soybean Allergen Gly M 4	1279.93	21.01	0.02