



Genome-wide identification and analysis of soybean acyl-ACP thioesterase gene family reveals the role of GmFAT to improve fatty acid composition in soybean seed

Zhou Zhou^{1,2} · Naoufal Lakhssassi¹ · Dounya Knizia¹ · Mallory A. Cullen¹ · Abdelhalim El Baz¹ · Mohamed G. Embaby³ · Shiming Liu¹ · Oussama Badad¹ · Tri D. Vuong⁴ · Amer AbuGhazaleh³ · Henry T. Nguyen⁴ · Khalid Meksem¹

Received: 12 April 2021 / Accepted: 13 July 2021 / Published online: 28 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Key message Soybean acyl-ACP thioesterase gene family have been characterized; *GmFATA1A* mutants were discovered to confer high oleic acid, while *GmFATB* mutants presented low palmitic and high oleic acid seed content.

Abstract Soybean oil stability and quality are primarily determined by the relative proportions of saturated versus unsaturated fatty acids. Commodity soybean typically contains 11% palmitic acid, as the primary saturated fatty acids. Reducing palmitic acid content is the principal approach to minimize the levels of saturated fatty acids in soybean. Though high palmitic acid enhances oxidative stability of soybean oil, it is negatively correlated with oil and oleic acid content and can cause coronary heart diseases for humans. For plants, acyl–acyl carrier protein (ACP) thioesterases (TEs) are a group of enzymes to hydrolyze acyl group and release free fatty acid from plastid. Among them, *GmFATB1A* has become the main target to genetically reduce the palmitic acid content in soybean. However, the role of members in soybean acyl-ACP thioesterase gene family is largely unknown. In this study, we characterized two classes of TEs, *GmFATA*, and *GmFATB* in soybean. We also denominated two *GmFATA* members and discovered six additional members that belong to *GmFATB* gene family through phylogenetic, syntenic, and in silico analysis. Using TILLING-by-Sequencing⁺, we identified an allelic series of mutations in five soybean acyl-ACP thioesterase genes, including *GmFATA1A*, *GmFATB1A*, *GmFATB1B*, *GmFATB2A*, and *GmFATB2B*. Additionally, we discovered mutations at *GmFATA1A* to confer high oleic acid (up to 34.5%) content, while mutations at *GmFATB* presented low palmitic acid (as low as 5.6%) and high oleic acid (up to 36.5%) phenotypes. The obtained soybean mutants with altered fatty acid content can be used in soybean breeding program for improving soybean oil composition traits.

Keywords Acyl-ACP thioesterase gene family · FATA · FATB · Oleic acid · Palmitic acid · *Glycine max* · EMS mutagenesis · TILLING-by-sequencing⁺

Introduction

Soybean oil is an important edible resource of vegetable oil that makes up 53% in the U.S. vegetable oil consumption in 2017 (American Soybean Association 2018). As the predominant saturated fatty acid, palmitic acid (16:0) typically accounts for 11% in conventional soybean oil (Fehr 2007). Although elevated palmitic acid content improves oxidative stability of soybean oil, it also causes the decrease in oleic acid and oil contents (Stoltzfus et al. 2000). On the contrary, reducing palmitic acid content has been reported to reduce the risks of developing cardiovascular diseases for humans (Hu et al. 1997). To produce edible oil with < 7% total saturates required by U.S. Food and Drug Administration,

✉ Khalid Meksem
meksem@siu.edu

¹ Department of Plant, Soil, and Agricultural Systems, Southern Illinois University, Carbondale, IL 62901, USA

² Present Address: Plant Science Department, McGill University, Montreal, QC H9X 3V9, Canada

³ Department of Animal Science, Food, and Nutrition, Southern Illinois University, Carbondale, IL 62901, USA

⁴ Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

several soybean lines with reduced palmitic acid phenotype have been identified as potential genetic resources for developing low palmitic acid cultivars (Rebetzke et al. 1998). In fatty acid biosynthetic pathway, 16:0-ACP fatty acid thioesterase (FATB) is the major target to genetically reduce the level of palmitic acid in soybean seeds.

Plant acyl–acyl carrier protein (ACP) thioesterase (TE), an enzyme terminates plastidial fatty acid biosynthesis, catalyzed acyl-ACP thioester bond hydrolysis to release free fatty acids and ACP. The substrate specificity of individual TEs is essential for the chain length of fatty acids exported from the plastid (Hills 1999; Voelker 1996). In the online database named ThYme, there are 25 TE families, from which Family TE14 include bacterial and plant acyl-ACP TEs (Cantu et al. 2010). Based on amino acid sequence alignment and substrate specificity, the plant TEs have been categorized into two classes, FATA and FATB (Voelker et al. 1997). The FATA class primarily hydrolyze 18:1-ACP with minor activity toward saturated acyl-ACP substrates, while FATB class shows preference for acyl-ACP with saturated fatty acyl chains (Dormann et al. 1995; Salas and Ohlrogge 2002). They both contain two helix/multi-stranded sheet motifs (hotdog domains), in which residues in the N-terminal domain were found to affect substrate specificity of enzymes and highly conserved residues in the C-terminal domain involved in catalysis (Mayer and Shanklin 2005). Two thioesterases maintain the saturated/unsaturated balance of membrane fatty acids for normal plant growth under critical conditions (Bonaventure et al. 2003).

As an allotetraploid crop species, soybean possesses a highly duplicated genome that ~75% of its genes are present with multiple copies. Two whole-genome duplication events have occurred in soybean genome, including one shared by legume species 59 million years ago and another glycine-specific one around 13 million year ago. The number of genes involved in acyl lipid biosynthesis in soybean is almost doubled compared to *Arabidopsis* (Schmutz et al. 2010). The gene families involved in fatty acid synthesis are generally much larger in soybean, such as omega-6 fatty acid desaturase (FAD2) with seven members (Lakhssassi et al. 2021). Such genetic redundancy drastically increases the complexity of genetic basis behind agronomical important traits but provide an invaluable resource for breeding desired phenotypes.

From mutagenized soybean lines, five quantitative trait loci (QTLs) have been associated with low palmitic acid phenotype, including *fap1* in C1726, *fap** in ELLP2, *fap3* in A22, *sop1* in J3, and *fap_{nc}* in N79-2077-12 (Burton et al. 1994; Cardinal et al. 2014; De Vries et al. 2011; Rahman et al. 1996; Stijšín et al. 1998). With an exception of *fap_{nc}* allelic with *fap3*, *fap1*, *fap3*, and *fap** are independent alleles conferring low palmitic acid content (Primomo 2000; Schnebly et al. 1994). At *fap1*, a disrupted splicing mutation

in a 3-ketoacyl-ACP synthase enzyme III (GmKASIIIA) has been associated with reduced palmitic acid phenotype (Cardinal et al. 2014). At *fap3*, a single nucleotide polymorphism (SNP) has caused loss-of-function for *GmFATB1A* (De Vries et al. 2011). *Fap_{nc}* represents the second allele of *GmFATB1A*, in which a deletion is responsible for low palmitic acid phenotype (Cardinal et al. 2007). Thapa et al. (2016) have identified two additional alleles of *GmFATB1A* from soybean mutant lines with 30% reduction in palmitic acid content. More recently, a 254-kb genomic deletion, including the *GmFATB1A* gene, has been reported to result in reduced palmitic acid content in soybean seeds (Bachleda et al. 2016; Goettel et al. 2016). Alternatively, downregulation of *GmFATB* gene expression can reduce palmitic acid content in soybean seeds (Buhr et al. 2002; Wilson et al. 2001). In *Arabidopsis*, a *FATB* knockout mutant has shown not only the low saturated fatty acid content, but also slow seedling growth and low-viable seed development (Bonaventure et al. 2003). However, no reports on *FATB* soybean mutants with negative impact on soybean growth and seed quality have been published so far.

TILLING (Targeting Induced Local Lesions IN Genomes) has been developed to screen-induced mutations from a chemical mutagenized population in early 2000s (McCallum et al. 2000). It combines traditional chemical mutagenesis with a high-throughput mutation screening method. Ethylmethane sulfonate (EMS) is widely used as the most common chemical mutagen to randomly create point mutations in plant genome (Anderson et al. 2018; Kandath et al. 2017; Koornneef et al. 1982; Lakhssassi et al. 2020a). A large number of TILLING populations have been developed in a variety of plant species, such as barley, legume, maize, rice, sorghum, and wheat (Till et al. 2009). Using reverse genetic methods like TILLING, scholars have been studying the gene function for economically important traits in soybean, such as disease resistance and seed oil composition traits. Two missense mutations in the *GmSHMT08* gene are identified in soybean cv. ‘Forrest’ mutant populations and result in alternation of SCN-resistant phenotype (Liu et al. 2012). Three missense mutations in individual soybean lines were detected in the *GmFAD2-1A*, and one of them leads to high oleic acid and low linoleic acid contents in the seed oil (Dierking and Bilyeu 2009). However, the complex traits resulted from duplicated soybean genome dramatically lower the efficiency of mutation screening in soybean. Using gel-based TILLING, a recent study shows that no mutations were found in either *GmFAD2-1A* or *GmFAD2-1B* from 2,000 EMS-mutagenized soybean lines, but five mutants in either of targeted genes were identified using forward phenotypic screening, which was followed by targeted sequencing analysis (Lakhssassi et al. 2017). The adoption of exome capture sequencing enabled the high-throughput screening for hidden mutations in multiple homologous wheat genes

controlling one trait (Krasileva et al. 2017). More recently, we developed a versatile TILLING-by-Sequencing⁺ technology and discovered novel genes associated with improved seed stearic acid content in soybeans (Lakhssassi et al. 2020b).

In the current study, we characterized the soybean acyl-ACP thioesterase gene family through a comprehensive analysis of phylogeny, gene structure and expression, synteny, and conserved domain variations and identified six additional members belonging to *GmFATB* gene family. Using TILLING-by-Sequencing⁺, we discovered for the first time that EMS-induced mutations in *GmFATA1A* resulted in high oleic acid content in soybean seed. Mutations at four *GmFATB* members also are associated with low palmitic acid and high oleic acid contents. These *GmFAT* mutants are the valuable sources to breed new soybean cultivars with low saturated and high monounsaturated fatty acid contents.

Materials and methods

Identification of FATA and FATB from soybean and other plant species

The putative soybean acyl-ACP thioesterase genes were identified by BLASTP searches against soybean reference genome (*Glycine max*, Wm82.a2.v1) at Phytozome (v12.1) using *Arbidopsis thaliana* acyl-ACP thioesterase protein sequences as queries (<https://phytozome.jgi.doe.gov>). Using the same approach, the putative acyl-ACP thioesterases were identified from reference genome of *Phaseolus vulgaris* (v2.1), *Medicago truncatula* (Mt4.0v1), *Brassica rapa* FPsc (v1.3), *Oryza sativa* (v7_JGI), *Lotus japonicas* genome assembly build 3.0 (<http://www.kazusa.or.jp/lotus/>), *Elaeis guineensis* assembly EG5 (<https://www.ncbi.nlm.nih.gov/genome/2669>), and *Cocos nucifera* assembly ASM812446v1 (<https://www.ncbi.nlm.nih.gov/genome/?term=Cocos+nucifera>). The total of 50 identified protein sequences with accession numbers were included in this study.

Phylogenetic analysis

Multiple sequence alignments of the full-length acyl-ACP thioesterase protein sequences from nine plant species were performed with MULTIPLE Sequence Comparison by Log-Expectation (MUSCLE). An unrooted phylogenetic tree was then constructed by maximum likelihood (ML) method in MEGA X using Jones-Taylor-Thornton Gamma Distributed (JTT + G) model for all *FAT* genes and JTT + G + I model with Invariant Sites (I) for soybean *FAT* genes (Hall 2013; Kumar et al. 2018).

Gene structure, expression profiling, and conserved domain analysis

The genomic and coding sequences of soybean acyl-ACP thioesterase genes retrieved from Phytozome v12.1 were aligned to generate the gene exon–intron structure diagram using the Gene Structure Display Server (Hu et al. 2015). To analyze the tissue-specific expression of soybean acyl-ACP thioesterase genes, normalized transcript data in six different tissues were downloaded from Soybase (<https://www.soybase.org/soyseq/>). The expressions profiling was visualized through heatmap using Heatmapper (Babicki et al. 2016). Followed by multiple sequence alignment between FATA and FATB in soybean and *A. thaliana*, the residues for substrate specifying have been proposed based on the criteria described by Jing et al. (2018). Catalytic residues in conserved motifs of soybean acyl-ACP thioesterases were identified from NCBI Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/cdd>).

Chromosomal localization and syntenic analysis

The locations of soybean acyl-ACP thioesterase genes and their corresponding chromosomes were drawn based on soybean genome annotation a2.v1 on SoyBase. Syntenic analysis was performed using soybean acyl-ACP thioesterase genes as locus identifier in plant genome duplication database (PGDD) (Lee et al. 2012). Nonsynonymous (Ka) versus synonymous substitution (Ks) rates were calculated based on their values retrieved from PGDD. For gene pairs whose information are not available at PGDD, PAL2NAL program was used to estimate Ka and Ks (Suyama et al. 2006). Given the Ks values and a rate of 6.1×10^{-9} substitutions per site per year, the divergence time (T) was equal to $Ks / (2 \times 6.1 \times 10^{-9}) \times 10^{-6}$ Mya for each gene pair (Chen et al. 2014).

Development of EMS-mutagenized soybean populations

EMS mutagenesis was performed as described in the past (Meksem et al. 2008). The soybean cv. Forrest and PI88788 seeds were used to generate M2 population in the greenhouse at SIUC Horticulture Research Center (HRC). Forrest is a Peking type and PI88788 is a PI88788 type in SCN resistance with the major loci being the *Rhg4 + rhg1-a* and the *rhg1-b*, respectively. A total of 4032 M2 lines were advanced to M3 generations by single-seed descent in the field between 2012 and 2015. M3 seeds from each mutant line were harvested, thrashed, and stored at -20 °C.

Mutation detection and validation

The mutations in five soybean acyl-ACP thioesterase genes were detected using TILLING-by-Sequencing⁺ method. A subset of mutations at *GmFATA1A*, *GmFATB1A*, *GmFATB1B*, *GmFATB2A*, and *GmFATB2B* were confirmed by Sanger sequencing. PCR primers were designed to amplify the fragments covering the exons of three soybean acyl-ACP thioesterase genes using Primer3 (Koressaar and Remm 2007). The PCR program was set up with 30 cycles of amplification at 94 °C for 30 s, 52 °C for 30 s, and 72 °C for 1 min. The PCR products were then purified using QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA, USA). The purified samples were sent for sequencing at GENEWIZ (<https://www.genewiz.com/>). The putative mutations were identified by alignment sampled sequences to reference using Unipro UGENE (Okonechnikov et al. 2012).

Fatty acid analysis of seeds from *GmFAT* mutants

Five major fatty acids' content was measured from selected M2/M3 lines according to the two-step methylation procedure (Kramer et al. 1997). At least three seeds per line were crushed in 16 mm × 200 mm tube with Teflon-lined screw cap individually. 2 mL sodium methoxide was added into tube followed by 50 °C incubation for 10 min. After cooling for 5 min, the samples were mixed with 3 mL of 5% (v/v) methanolic HCl, incubated at 80 °C for 10 min, and cooled for 7 min. Each tube was then added with 7.5 mL of 6% (w/v) potassium carbonate and 1 mL of hexane and centrifuged at 1200 g for 5 min. The upper layer was transferred to vials, from which the individual fatty acid content was determined as a percentage of the total fatty acids of soybean seed by gas chromatography. A Shimadzu GC-2010 (Columbia, MD) gas chromatograph fitted with a flame ionization

detector was equipped with a Supelco 60-m SP-2560-fused silica capillary famewax column (0.25 mm i.d. × 0.25 μm film thickness). The standard fatty acids were run first to create calibration reference.

Results

Identification of plant acyl-ACP thioesterase gene family members in soybean

Four *FATB* genes have previously been identified in soybean, from which *GmFATB1A* is associated with reducing palmitic acid content (Cardinal et al. 2007). To identify the putative members of TE family in soybean, a BLASTP search against the soybean genome database (Wm82.a2.v1) was performed by using *A. thaliana* TE protein sequences as queries. Combined with soybean TEs from Family TE14 in the ThYme database, a total of 12 TEs have been found in soybean genome, including ten *GmFATB* and two *GmFATA*. Based on nomenclature proposed previously, additional six *GmFATB* genes are denominated as *GmFATB3A* (Glyma.04G197400), *GmFATB3B* (Glyma.06G168100), *GmFATB4A* (Glyma.04G197500), *GmFATB4B* (Glyma06g17625), *GmFATB5A* (Glyma.10G268200), *GmFATB5B* (Glyma.20G122900), as well as two *GmFATA* genes, *GmFATA1A* (Glyma.18G167300) and *GmFATA1B* (Glyma.08G349200) (Table 1).

Amino acid sequence alignment has shown that two genes in each subfamily of *GmFATB* and *GmFATA* share high similarity, such as *GmFATB1A*/*FmFATB1B* (96%), *GmFATB3A*/*GmFATB3B* (94%), and *GmFATA1A*/*GmFATA1B* (93%). The coding DNA sequence (CDS) lengths of the *GmFATB* range from 1140 to 1269 bp with an average of 1203 bp while that of *GmFATA* averages 1140 bp. The

Table 1 The list of soybean acyl-ACP thioesterase genes with their corresponding gene ID, nucleotide sequence characteristics, and protein sequence properties

Gene name	Gene ID	Gene length (bp)	CDS (bp)	Exons	Protein sequence (aa)	Mol. wt. (kDa)	pI
FATB1A	Glyma.05G012300	4195	1251	6	416	45.9	6.48
FATB1B	Glyma.17G120400	4382	1251	6	416	46.0	6.28
FATB2A	Glyma.04G151600	7394	1269	6	422	46.1	8.12
FATB2B	Glyma.06G211300	4959	1260	6	419	46.3	7.08
FATB3A	Glyma.04G197400	2627	1149	6	382	43.9	9.30
FATB3B	Glyma.06G168100	3108	1152	6	383	44.0	8.91
FATB4A	Glyma.04G197500	3391	1158	6	385	44.8	9.27
FATB4B	Glyma.06G168000	3838	1140	6	379	43.9	8.89
FATB5A	Glyma.10G268200	2701	1203	6	400	45.6	8.41
FATB5B	Glyma.20G122900	3197	1200	6	399	45.4	8.13
FATA1A	Glyma.18G167300	6650	1125	7	374	42.0	8.11
FATA1B	Glyma.08G349200	5070	1155	8	384	43.3	8.86

sizes and predicted molecular weight of GmFATB1 and GmFATB2 subfamilies are larger than 400 amino acids and 45.8 kDa, respectively. GmFATB1A, GmFATB1B, and GmFATB2B show acidic isoelectric point (pI) values, whereas the rest of soybean TEs presented basic pI values (Table 1).

Phylogenetic analysis of plant acyl-ACP thioesterase gene family

13 FATA and 25 FATB proteins from other three legumes, two dicot species, and three monocot species have been identified through BLAST searches using *A. thaliana* TE protein sequences. A maximum likelihood (ML) tree was constructed with 50 protein sequences to elucidate the phylogenetic relationships among TEs from nine plant species (Fig. 1). As expected, two distinct clusters are formed to separate 15 FATA members from 35 FATB ones. In FATB cluster, all 35 FATB members could be classified into four subgroups. In subgroup I, GmFATB1A, GmFATB1B, GmFATB2A, and GmFATB2B are grouped together with AtFATB, BrFATB, and eight FATB members from other three legume species. Subgroup II contains all FATB members from three monocot species except one OsFATB. There are seven FATB members in subgroup III, including GmFATB3A, GmFATB3B, GmFATB4A, and GmFATB4B. And subgroup IV has GmFATB5A, GmFATB5B, and three FATB members from two legumes and one monocot species (Fig. 1). On the other hand, FATA members from all legume species are clustered apart from ones in monocot species. However, AtFATA are grouped with BrFATA in two different branches. The phylogenetic analysis also shows a close evolutionary relationship within each of six soybean TEs gene pairs with $\geq 88\%$ reliability (Fig. 1).

Gene structure and expression profiling of soybean acyl-ACP thioesterase genes

Given the two whole-genome duplication events, the soybean TE gene family consists of 12 members, which is four times more than those in *Arabidopsis* and double compared to the number of TEs in common bean, palm, and rice. Compared to an average of 5860 bp for *GmFATA*, the gene lengths of *GmFATB1* and *GmFATB2* subfamilies are more than 4195 bp while that of *GmFATB3*, *GmFATB4*, and *GmFATB5* subfamilies are 3143 bp on average (Table 1). The *GmFATB2A* has the longest gene length among soybean TEs due to its extended 3'-UTR region. The gene structures of *GmFATB* are highly conserved with six exons for all ten members; on the contrary, *GmFATA1A* and *GmFATA1B* have seven and eight exons, respectively (Fig. 2).

GmFATB1A, *GmFATB1B*, and *GmFATB2A* show relatively high expression in soybean seeds while the transcripts

of two *GmFATB2* genes were abundant in soybean flowers. *GmFATA1A* and *GmFATA1B* were also highly expressed in soybean seeds. Two *GmFATB1* genes expressed relatively high levels in soybean root and nodule. Additionally, the expression of *GmFATB1*, *GmFATB2*, and *GmFATA* exhibits similar patterns in leaves and pod. The expression of *GmFATB3A*, *GmFATB3B*, and *GmFATB5A* is recorded as 0 in most of tested tissues and no RNA-seq data are available for *GmFATB4A*, *GmFATB4B*, and *GmFATB5B* in Soybase (Figure S1).

Chromosomal distribution and gene duplication

Based on the physical locations, 12 soybean TE genes are unevenly distributed on eight soybean chromosomes (Fig. 3). Chromosome 4 and 6 contain three *GmFATB* genes each while only one *GmFATB* gene each is present on chromosomes 5, 10, 17, and 20. Two *GmFATA* genes are located at chromosome 8 and 18, respectively. Among the *GmFATB* subfamilies, *GmFATB1* and *GmFATB5* are evenly distributed on four chromosomes. Nevertheless, the other three subfamilies, *GmFATB2*, *GmFATB3*, and *GmFATB4*, are concentrated on two chromosomes with three on each (Fig. 3).

The duplication analyses have shown that all soybean TE genes are located within eight duplicated blocks (Table 2). The gene pair, *GmFATB1A* and *GmFATB1B*, belongs to a large duplicated segment containing 62 anchor genes, while *GmFATB3A/GmFATB4B* and *GmFATB5A/GmFATB5B* are presented in huge syntenic regions with 711 and 884 anchor genes, respectively (Figure S2). Two gene pairs, *GmFATB3A/GmFATB4A* and *GmFATB3B/GmFATB4B*, are regarded as the outcomes of tandem duplication events due to their tight physical distance of less than 7 kb (without any genes in between). The ratio of nonsynonymous to synonymous substitutions (K_a/K_s) was calculated for each gene pair to determine the types of natural selection acting on coding sequences. The K_a/K_s of soybean TEs gene pairs is less than 0.5, which suggests that the evolution of soybean TEs is under purifying selection (Juretic et al. 2005; Li et al. 1981). The duplication of eight gene pairs is estimated to have occurred between 7.38 and 76.23 Mya based on 6.161029 synonymous mutations per synonymous site per year for soybean (Table 2).

Conserved domain variations among plant TEs

The protein sequences of 15 soybean and *Arabidopsis thaliana* TEs, four FATA and eleven FATB, were aligned to compare residues within two conserved hotdog domains. A residue that is conserved within one plant TE class but differs between FATA and FATB classes may contribute to the difference in substrate specificity (Jing et al. 2018). Based on these criteria, a total of 13 residues were selected,

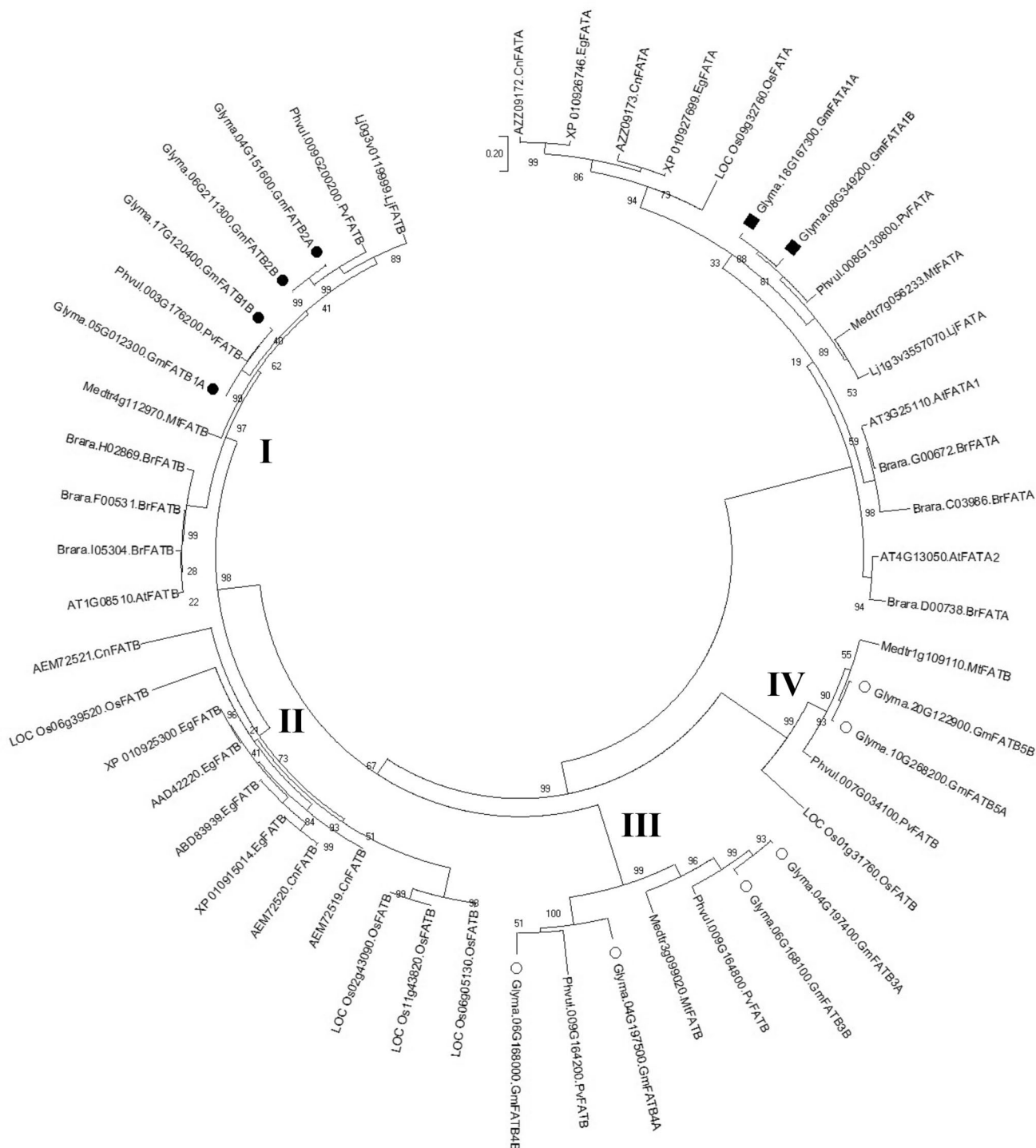


Fig. 1 Phylogenetic tree of acyl-ACP thioesterases gene family from nine plant species. The protein sequences of all acyl-ACP thioesterases were subjected to a MUSCLE multiple alignment and phylogenetic tree was constructed by maximum likelihood (ML) method using Mega X. The two members of soybean *GmFATA* gene subfamily were labeled with filled squares. The previously identified four members of soybean *GmFATB1/2* subfamilies were marked with

filled circles, while the newly identified members of *GmFATB3/4/5* subfamilies in this study were labeled with empty circles. The name and abbreviation of plant species used for the analysis are *Arabidopsis thaliana* (At); *Glycine max* (Gm); *Phaseolus vulgaris* (Pv); *Medicago truncatula* (Mt); *Lotus japonicus* (Lj); *Brassica rapa* (Br); *Cocos nucifera* (Cn); *Elaeis guineensis* (Eg); *Oriza sativa* (Os)

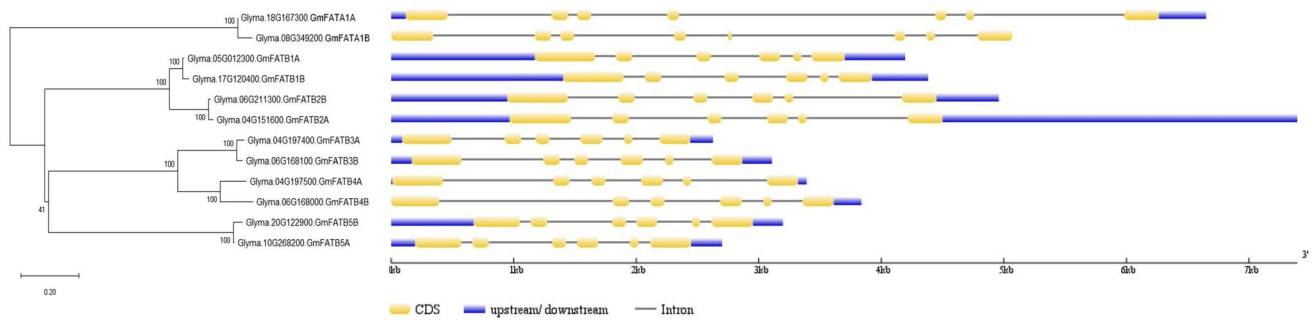


Fig. 2 Phylogenetic relationships and gene structures of *GmFATA* and *GmFATB*. The protein sequences of all soybean acyl-ACP thioesterases were subjected to a MUSCLE alignment, and phylogenetic gene tree was constructed using Mega X. The structures of 12 soybean acyl-ACP thioesterase genes were plotted with yellow boxes repre-

sented exons (coding DNA sequence, CDS), black lines illustrating introns, and blue boxes indicating 5'-UTR and 3'-UTR regions. The size of gene structures could be measured by the scale in the unit of base pair (bp) at the bottom. The gene structure was drawn using the Gene Structure Display Server

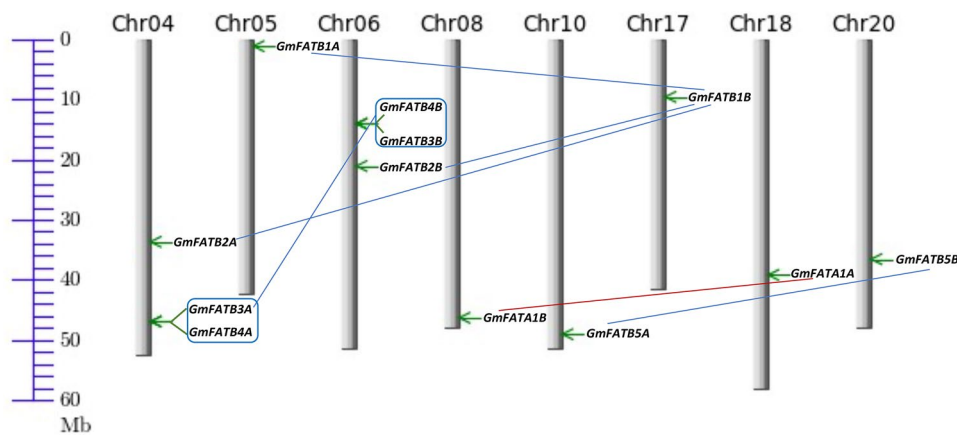


Fig. 3 Chromosomal locations and duplications of soybean acyl-ACP thioesterase genes. Each chromosome number is indicated above bar by Roman number and the scale (on the left) is in mega base (Mb). The size of chromosome and gene locations are based on soybean

genome annotation a2.v1 on SoyBase. Each pair of segmental duplication in *GmFATA* and *GmFATB* subfamilies is connected by red and blue lines, respectively. The tandem duplicated genes are shown in rectangle box

Table 2 Divergence and duplication of acyl-ACP thioesterase gene pairs in soybean

Acyl-ACP thioesterase gene pairs	Ka	Ks	Ka/Ks	Duplication time (Mya)	Duplication type
GmFATB1A-GmFATB1B	0.02	0.09	0.222	7.38	Segmental
GmFATB1B-GmFATB2A	0.16	0.86	0.186	70.49	Segmental
GmFATB1B-GmFATB2B	0.17	0.93	0.183	76.23	Segmental
GmFATB3A-GmFATB4A	0.28	0.71	0.392	58.20	Tandem
GmFATB3A-GmFATB4B	0.22	0.56	0.393	45.90	Segmental
GmFATB3B-GmFATB4B	0.27	0.65	0.414	53.28	Tandem
GmFATB5A-GmFATB5B	0.02	0.1	0.2	8.20	Segmental
GmFATA1A-GmFATA1B	0.04	0.1	0.457	8.20	Segmental

from which A194G, T208V, and D276E have previously been reported as specificity determining positions (Mayer and Shanklin 2007). Additionally, seven residues are found as completely different between FATA and FATB classes

in hotdog domain I, including K150R, N163D, V178T, H212Q, I213V, R236K, and K246R. In hotdog domain II, another three residues, T347K, D362E, and D372E, meet the same criteria (Table 3). Among the ten newly identified

Table 3 The predicted 13 substrate specificity sites and three conserved catalytic sites of soybean acyl-ACP thioesterase genes

Hotdog domain	I										II					
	150 ^a	163	178	194 ^b	208 ^b	212	213	236	246	276 ^b	340 ^c	342 ^c	347	362	372	377 ^c
FATA																
AT3G25110 AtFATA1	K	N	V	A	T	H	I	R	K	D	N	H	T	D	D	C
AT4G13050 AtFATA2	K	N	V	A	T	H	I	R	K	D	N	H	T	D	D	C
Glyma.18G167300 GmFATA1A	K	N	V	A	T	H	I	R	K	D	N	H	T	D	D	C
Glyma.08G349200 GmFATA1B	K	N	V	A	T	H	I	R	K	D	N	H	T	D	D	C
FATB																
AT1G08510 AtFATB	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.05G012300 GmFATB1A	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.17G120400 GmFATB1B	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.04G151600 GmFATB2A	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.06G211300 GmFATB2B	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.04G197400 GmFATB3A	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.06G168100 GmFATB3B	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.04G197500 GmFATB4A	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.06G168000 GmFATB4B	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.10G268200 GmFATB5A	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C
Glyma.20G122900 GmFATB5B	R	D	T	G	V	Q	V	K	R	E	N	H	K	E	E	C

^aNumbering of the residues is according to the sequence of GmFATA1A

^bPreviously reported substrate specificity sites of acyl-ACP thioesterases

^cThe catalytic sites of acyl-ACP thioesterases

residues, three residues, V178T, H212Q, and T347K, present non-conservative difference in amino acid between FATA and FATB, while the rest of seven residues contain conservative changes. In addition, three conserved catalytic residues, N340, H342, and C377, may form a papain-like catalytic triad across the FATA and FATB classes (Mayer and Shanklin 2005). From Conserved Domain Database (CDD) at National Center for Biotechnology Information (NCBI), seven active sites of plant TEs have been revealed in hotdog domain I. Among them, two residues, T208V and R236K, overlap with ones identified as substrate specifying residues, while the rest are highly conserved between FATA and FATB classes except two mismatches at positions 237 and 238 (Figure S3).

Identification of new alleles of *GmFAT* to improve fatty acid composition in soybean seed

Five soybean acyl-ACP thioesterase genes, *GmFATA1A*, *GmFATB1A/1B*, and *GmFATB2A/2B*, have been included in screening mutations through TILLING-by-Sequencing⁺. The estimated mutation density of these five genes is 1/232 kb using the formula as the total number of mutations divided by the total number of base pairs (amplicon size x individuals screened) (Table 4) (Cooper et al. 2008). Among the 280 identified mutations in these five *GmFAT* genes, the typical EMS-type mutations are the majority of base changes with

45.7% in G to A and 37.1% in C to T, while the other types of mutations only took up 17.1% (Table 4). In the coding regions of these five *GmFAT* genes, a total of 118 amino acid changes are detected, from which 71.2% are missense mutations, 26.3% are silent mutations, and 2.5% are nonsense mutations. Nonsense mutations are found in *GmFATA1A* and *GmFATB2A/2B* genes, whereas no nonsense mutations are present in *GmFATB1A/1B* genes (Table 4).

A subset of *GmFAT* mutants has been confirmed by Sanger sequencing, and their novel alleles have been associated with altered fatty acid profiles. Six missense mutations (S37F, A55T, T146I, A231V, G277E, and V310I) are identified from *GmFATA1A* mutants, in which two mutants, F243 and F393, present > 30% high oleic acid content. Another four mutants, F636, F1305, F740, and F1188, display moderately high oleic acid content (> 24%) compared to Forrest wild type (18.0%) (Table 5). Five *GmFATB1A* mutants (F1040, F1129, F1539, F1200, and F1166) carry the missense mutations, P18L, G128R, R138H, G223E, and A371T, respectively. The palmitic acid content of these five mutants ranges from 8.9 to 10.7%, but an increase in oleic acid content (23.9–34.0%) is also found in *GmFATB1A* mutants (Table 5). Likewise, five missense mutations (P118S, G128E, A174T, D284N, and R348K) are detected at *GmFATB1B*, from which one mutant (F3) shows a decreased palmitic acid content by 51.7% when compared to Forrest wild type.

Table 4 A summary of mutations in five soybean acyl-ACP thioesterase genes identified by TILLING-by-sequencing⁺

Gene ID	Ampli-con size (bp)	Base changes	Type of base changes			InDel	AA substitutions	Missense mutations	Nonsense mutations	Silent mutations	Mutation density (kb)
			G>A	C>T	Others						
Glyma.05G012300	2960	42	19	14	9	1	17	12	0	5	1/284
Glyma.17G120400	2960	38	18	18	2	3	13	11	0	2	1/314
Glyma.04G151600	4840	93	40	41	12	1	30	20	1	9	1/210
Glyma.06G211300	2760	69	33	18	18	3	41	29	1	11	1/161
Glyma.18G167300	2600	38	18	13	7	0	17	12	1	4	1/276
Total	16,120	280	128	104	48	8	118	84	3	31	1/232
Percentage of each type of base change			45.7%	37.1%	17.1%						
Percentage of each type of amino acid change								71.2%	2.5%	26.3%	

Table 5 A summary of mutants in *GmFATA* and *GmFATB* genes identified by TILLING-by-sequencing⁺ and their fatty acids phenotype

Gene ID	Plant ID	Nucleotide change	Amino acid substitution	C16:0	C18:0	C18:1	C18:2	C18:3
<i>GmFATA1A</i> (Glyma.18G167300)	F636	C110T	S37F	10.2	5.1	27.6	51.3	5.7
	F1305	G163T	A55T	10.5	5.5	25.4	52.8	5.8
	F740	C437T	T146I	10.5	3.8	25.3	52.4	6.0
	F243	C692T	A231V	9.1	3.6	32.4	44.7	8.8
	F393	G830A	G277E	9.9	4.1	34.5	45.2	4.1
	F1188	G928A	V310I	9.7	5.1	24.1	54.8	6.3
<i>GmFATB1A</i> (Glyma.05G012300)	F1040	C52T	P18L	9.3	4.7	32.2	48.0	5.9
	F1129	G382A	G128R	10.2	4.8	34.0	46.0	5.1
	F1539	G413A	R138H	10.7	4.0	24.8	54.3	6.2
	F1200	G668A	G223E	9.2	4.2	23.9	54.7	8.0
	F1166	G1111A	A371T	8.9	7.1	26.2	49.1	5.2
<i>GmFATB1B</i> (Glyma.17G120400)	F359	C352T	P118S	10.6	7.6	25.0	48.2	6.7
	F43	G383A	G128E	9.5	4.5	31.9	46.2	6.2
	F1726	G520A	A174T	8.5	4.3	33.4	47.1	6.7
	F3	G850A	D284N	5.6	4.3	26.2	56.2	6.6
	F858	G1043A	R348K	12.0	4.1	26.5	51.8	5.6
<i>GmFATB2A</i> (Glyma.04G151600)	F1345	C47T	P16L	9.6	3.5	36.5	45.5	5.0
	F1576	G153A	A153T	9.8	4.2	25.0	53.4	7.6
	F1104	G1117A	A373T	9.9	4.2	34.3	46.2	5.3
	F1220	G1154A	R385Q	9.6	6.1	26.5	52.0	5.9
	F854	G1184A	G395D	11.1	4.3	24.6	51.4	8.6
<i>GmFATB2B</i> (Glyma.06G211300)	F317	G136A	G46S	8.9	10.0	34.8	41.3	4.9
	F225	G763A	V255M	11.9	3.4	19.9	56.0	8.9
	F564	C988T	Q330*	10.1	3.0	21.1	57.0	8.7
	F-WT			11.6	3.3	18.0	54.5	6.2

All five *GmFATB1B* mutants also present an elevated oleic acid content up to 33.4% (Table 5). Another five missense mutations, P16L, A153T, A373T, R385Q, and G395D, are identified at *GmFATB2A*, from which four show a

reduction in palmitic acid content. Two missense and one nonsense mutations are discovered at *GmFATB2B*. All *GmFATB2A* and *GmFATB2B* mutants show an increase in oleic acid content (Table 5).

Discussion

Among type II fatty acid synthases (FAS), the plant TE is a major contributing factor in determining the carbon chain length of fatty acids through their substrate specificity. The large number of soybean TE genes implied genome expansion of the soybean compared to counterparts in other plant species. Previous studies have identified four unique *GmFATB* genes, from which mutations at *GmFATB1A* result in low palmitic acid content in soybean seed (Cardinal et al. 2007). In this study, we performed a genome-wide search for soybean TE genes with the aid of Phytozome and ThYme databases. Additional six *GmFATB* genes have been identified and named according to previously proposed nomenclature, as well as two *GmFATA* genes (Table 1). The total number of TE genes is four times higher in soybean compared to *A. thaliana*.

Here, we conducted an overall phylogenetic analysis of plant TEs gene families from nine plant species using maximum likelihood (ML) method (Fig. 1). Interestingly, *GmFATB1* and *GmFATB2*, *GmFATB3* and *GmFATB4*, and *GmFATB5* subfamilies are in three different subgroups under *FATB* cluster, respectively. For plant species with high palmitic acid levels, such as coconut and palm, their *FATBs* appear to evolve independently from dicot species. Although two *FATA* and one *FATB* genes are presented in *Arabidopsis* genome, members from *FATA* class are generally much fewer than ones from *FATB* class in other higher plants. Two *GmFATA* are grouped with *FATAs* in other three legume species but apart from ones in other dicot species (Fig. 1).

The gene structures are similar within *GmFATB1/GmFATB2* subfamilies, *GmFATB3/GmFATB4/GmFATB5* subfamilies, and *GmFATA*. Although the *GmFATA* have the longer gene lengths due to the extended intron length, the coding sequence lengths of *GmFATA* are generally shorter than those of *GmFATB*. The gene lengths of *GmFATB1* and *GmFATB2* subfamilies are longer than those of *GmFATB3*, *GmFATB4*, and *GmFATB5* subfamilies, as are their coding sequences (Table 1). With the advent of intron gain/loss events, all *GmFATB* lose at least one intron when they evolve divergently from *GmFATA* (Fig. 2).

The expression profiling data reveal various expression patterns of eight soybean TE genes in six soybean tissues. The similar expression patterns point to functional redundancy during soybean evolution, which could lead to neofunctionalization and subfunctionalization within soybean TE gene family (Figure S1). As expected, the high transcript level of *GmFATB1* subfamilies, *GmFATB2A*, and *GmFATA* has been detected in soybean seeds, which indicated that these genes play a major role in releasing

free fatty acids to cytosol. Thus, they should be the main targets to genetically modify fatty acids composition in soybean seed. For the newly identified *GmFATB* members, the very low expression level of *GmFATB3A*, *GmFATB3B*, and *GmFATB5A* in all six tissues suggests that their functions need to be explored further (Figure S1).

The distribution of 12 soybean TE genes has been shown to be on eight chromosomes (Fig. 3). Chromosomes 4 and 6 contain the largest number of TE genes (3), whereas chromosomes 5, 8, 10, 17, 18, and 20 only have one TE gene on each. The majority of soybean TE genes are found toward the chromosome ends, suggesting potential interchromosomal crossovers due to the high genetic recombination rates. Plant species acquired novel traits and adapted to various environments through gene duplication (Bowers et al. 2003; Eckardt 2004). There are three main gene duplication patterns, including segmental duplication, tandem duplication, and transposition (Kong et al. 2007). Our syntenic analysis shows that soybean TE gene family expands through both segmental and tandem duplications (Table 2). It is also well known that two whole-genome duplication events have occurred in soybean genome, including one shared by legume species 59 million years ago and another glycine-specific one around 13 million year ago (Schmutz et al. 2010,2014; Young and Bharti 2012). The duplication time of soybean TE gene pairs is estimated to match with either of these two time periods. *GmFATB1B/GmFATB2A*, *GmFATB1B/GmFATB2B*, *GmFATB3A/GmFATB4A*, *GmFATB3A/GmFATB4B*, and *GmFATB3B/GmFATB4B* are formed between 45.90 and 76.23 Mya while the duplication of *GmFATB1A/GmFATB1B*, *GmFATB5A/GmFATB5B*, and *GmFATA1A/GmFATA1B* has occurred between 7.38 and 8.20 Mya (Table 2).

Within two hotdog domains, 10 newly identified residues that are completely different between *GmFATA* and *GmFATB* could be the candidate positions to determine the differences in substrate specificity of TEs in soybean. Compared to other seven residues, three residues (V178T, H212Q, and T347K) may play a more important role in substrate specifying due to their non-conservative difference in amino acid (Table 3). In the current study, one *GmFATA1A* mutant (F740) has been identified to possess non-conservative amino acid changes at a previously reported specificity determining position (T208V) and confer an increased level of oleic acid in soybean seed (Tables 3 and 5) (Mayer and Shanklin 2007).

The mutations at *GmFATB1A* have been repeatedly associated with low palmitic acid phenotype in soybean. The expression level of *GmFATA* and *GmFATB* may have significant impact on soybean seed fatty acid composition, however, future studies are required to elucidate the role of soybean acyl-ACP thioesterases in controlling seed oleic acid content (Byfield and Upchurch 2007). This is the first

time to discover that the novel alleles of *GmFATA1A* confer an elevated oleic acid content in soybean seeds. The increase in oleic acid content in *GmFATA1A* mutants, F243 and F393, is comparable to the high oleic acid content in either *GmFAD2-1A* or *GmFAD2-1B* mutants with the same genetic background (Table 5) (Lakhssassi et al. 2017). Meanwhile, the novel alleles of *GmFATB1A*, *GmFATB1B*, *GmFATB2A*, *GmFATB2B* are identified to confer low palmitic acid content in soybean seeds. Interestingly, these *GmFATB* mutants also present an elevated oleic acid content, which is consistent with the significant increase in oleic acid content from other previously reported *GmFATB1A* mutants (Table 5) (Bachleda et al. 2016; Cardinal et al. 2007). Zhou et al. (2019) indicate that a negative correlation exists between palmitic acid and oleic contents in both natural and mutagenized soybean populations. The identified *GmFAT* mutants are the new sources of seed high oleic acid and low palmitic acid contents for soybean breeding.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-021-03917-9>.

Author contributions statement ZZ wrote the manuscript draft, made the corresponding figures and tables, performed phylogenetic, gene structure, expression profiling, chromosome and syntenic, and *in silico* analysis. NL proofread, edited the manuscript, and directed the work together with KM. ZZ, NL, and SL developed the EMS-mutagenized populations. ZZ, NL, DK, MC, and OB did TILLING-by-Sequencing⁺ analysis and identified *GmFAT* mutations. ZZ, MC, AEB, MGE, TDV, and HTN performed fatty acid phenotyping and edited the manuscript. KM conceived and designed the experiments, supervised the work, and edited the manuscript. All authors edited, reviewed, and approved the final version of the manuscript.

Funding This research was supported in part from the United Soybean Board, project USB-2020-162-0127 to K.M. and N.L.

Availability of data and material Not applicable.

Code availability Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- American Soybean Association (2018) A Reference guide to soybean facts and figures
- Anderson J, Lakhssassi N, Kantartzi SK, Meksem K (2018) Non-hypothesis analysis of a mutagenic soybean (*Glycine max*[L.] population for protein and fatty-acid composition. *J Am Oil Chem Soc* 95:461–471
- Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS (2016) Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res* 44:W147–W153
- Bachleda N, Pham A, Li Z (2016) Identifying *FATB1a* deletion that causes reduced palmitic acid content in soybean N87–2122-4 to develop a functional marker for marker-assisted selection. *Mol Breed* 36:45
- Bonaventure G, Salas JJ, Pollard MR, Ohlrogge JB (2003) Disruption of the *FATB* gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell* 15:1020–1033
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438
- Buhr T, Sato S, Ebrahim F, Xing A, Zhou Y, Mathiesen M, Schweiger B, Kinney A, Staswick P, Clemente T (2002) Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *Plant J* 30:155–163
- Burton J, Wilson R, Brim C (1994) Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. *Crop Sci* 34:313
- Byfield GE, Upchurch RG (2007) Effect of temperature on delta-9 stearoyl-ACP and microsomal omega-6 desaturase gene expression and fatty acid content in developing soybean seeds. *Crop Sci* 47:1698–1704
- Cantu DC, Chen Y, Lemons ML, Reilly PJ (2010) ThYme: a database for thioester-active enzymes. *Nucleic Acids Res* 39:D342–D346
- Cardinal AJ, Burton JW, Camacho-Roger AM, Yang JH, Wilson RF, Dewey RE (2007) Molecular analysis of soybean lines with low palmitic acid content in the seed oil. *Crop Sci* 47:304–310
- Cardinal AJ, Whetten R, Wang S, Auclair J, Hyten D, Cregan P, Bachlava E, Gillman J, Ramirez M, Dewey R (2014) Mapping the low palmitate *fap1* mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations. *Theor Appl Genet* 127:97–111
- Chen X, Chen Z, Zhao H, Zhao Y, Cheng B, Xiang Y (2014) Genome-wide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. *PLoS ONE* 9:e87156
- Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N (2008) TILLING to detect induced mutations in soybean. *BMC Plant Biol* 8:9
- De Vries BD, Fehr WR, Welke GA, Dewey RE (2011) Molecular characterization of the mutant *fap3(A22)* allele for reduced palmitate concentration in soybean. *Crop Sci* 51:1611–1616
- Dierking EC, Bilyeu KD (2009) New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biol* 9:89
- Dormann P, Voelker TA, Ohlrogge JB (1995) Cloning and expression in *Escherichia coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long-chain acyl–acyl carrier proteins. *Arch Biochem Biophys* 316:612–618
- Eckardt NA (2004) Two genomes are better than one: widespread paleopolyploidy in plants and evolutionary effects. *Plant Cell* 16:1647–1649
- Fehr WR (2007) Breeding for modified fatty acid composition in soybean. *Crop Sci* 47:S72–S87
- Goettel W, Ramirez M, Upchurch RG, Y-qC An (2016) Identification and characterization of large DNA deletions affecting oil quality traits in soybean seeds through transcriptome sequencing analysis. *Theor Appl Genet* 129:1577–1593
- Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol* 30:1229–1235
- Hills MJ (1999) Improving oil functionality by tuning catalysis of thioesterase. *Trends Plant Sci* 4:421–422

- Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, Hennekens CH, Willett WC (1997) Dietary fat intake and the risk of coronary heart disease in women. *N Engl J Med* 337:1491–1499
- Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296–1297
- Jing F, Zhao L, Yandeu-Nelson MD, Nikolau BJ (2018) Two distinct domains contribute to the substrate acyl chain length selectivity of plant acyl-ACP thioesterase. *Nat Commun* 9:1–10
- Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE (2005) The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res* 15:1292–1297
- Kandoth PK, Liu S, Prenger E, Ludwig A, Lakhssassi N, Heinz R, Zhou Z, Howland A, Gunther J, Eidson S, Dhroso A, LaFayette P, Tucker D, Johnson S, Anderson J, Alaswad A, Cianzio SR, Parrott WA, Korkin D, Meksem K, Mitchum MG (2017) Systematic mutagenesis of serine hydroxymethyltransferase reveals an essential role in nematode resistance. *Plant Physiol* 175:1370–1380
- Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, DePamphilis CW (2007) Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J* 50:873–885
- Koornneef M, Dellaert L, Van der Veen J (1982) EMS-and relation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat Res Fundam Mol Mech Mutagen* 93:109–123
- Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23:1289–1291
- Kramer JK, Fellner V, Dugan ME, Sauer FD, Mossoba MM, Yurawecz MP (1997) Evaluating acid and base catalysts in the methylation of milk and rumen fatty acids with special emphasis on conjugated dienes and total trans fatty acids. *Lipids* 32:1219–1228
- Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P (2017) Uncovering hidden variation in polyploid wheat. *Proc Natl Acad Sci* 114:E913–E921
- Kumar S, Stecher G, Li M, Nuyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
- Lakhssassi N, Zhou Z, Liu S, Colantonio V, AbuGhazaleh A, Meksem K (2017) Characterization of the FAD2 gene family in soybean reveals the limitations of gel-based TILLING in genes with high copy number. *Front Plant Sci* 8:324
- Lakhssassi N, Piya S, Knizia D, El Baze A, Cullen MA, Meksem J, Lakhssassi A, Hewezi T, Meksem K (2020a) Mutations at the serine hydroxymethyltransferase impact its interaction with a soluble nsf attachment protein and a pathogenesis-related protein in soybean. *Vaccines (basel)* 8:349
- Lakhssassi N, Zhou Z, Liu S, Piya S, Cullen MA, El Baze A, Knizia D, Patil GB, Badad O, Embaby MG, Meksem J, Lakhssassi A, AbuGhazaleh A, Hewezi T, Meksem K (2020b) Soybean TILLING-by-Sequencing⁺ reveals the role of novel GmSACPD members in the unsaturated fatty acid biosynthesis while maintaining healthy nodules. *J Exp Bot* 71:6969–6987
- Lakhssassi N, Lopes-Caitar VS, Knizia D, Cullen MA, Badad O, El Baze A, Zhou Z, Embaby MG, Meksem J, Lakhssassi A, Chen P, AbuGhazaleh A, Vuong TD, Nguyen HT, Hewezi T, Meksem K (2021) TILLING-by-sequencing⁺ reveals the role of novel fatty acid desaturases (GmFAD2-2s) in increasing soybean seed oleic acid content. *Cells* 10:1245
- Lee T-H, Tang H, Wang X, Paterson AH (2012) PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res* 41:D1152–D1158
- Li W-H, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. *Nature* 292:237–239
- Liu S, Kandoth PK, Warren SD, Yeckel G, Heinz R, Alden J, Yang C, Jamai A, El-Mellouki T, Juvalle PS (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492:256–260
- Mayer KM, Shanklin J (2005) A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. *J Biol Chem* 280:3621–3627
- Mayer KM, Shanklin J (2007) Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *BMC Plant Biol* 7:1
- McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. *Nat Biotechnol* 18:455
- Meksem K, Liu S, Liu XH, Jamai A, Mitchum MG, Bendahmane A, El-Mellouki T (2008) TILLING: a reverse genetics and a functional genomics tool in soybean. In: Kahl G, Meksem K (eds) *The handbook of plant functional genomics: concepts and protocols*. Wiley, Weinheim, pp 251–265
- Okonechnikov K, Golosova O, Fursov M, Team U (2012) UniPro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28:1166–1167
- Primomo V (2000) Inheritance and stability of palmitic acid alleles in soybeans (*Glycine max* L Merr.). Master's thesis Univ of Guelph, Guelph, ON, Canada
- Rahman SM, Takagi Y, Kinoshita T (1996) Genetic analysis of palmitic acid contents using two soybean mutants, J3 and J10. *Jpn J Breed* 46:343–347
- Rebetzke GJ, Burton JW, Carter TE Jr, Wilson RF (1998) Genetic variation for modifiers controlling reduced saturated fatty acid content in soybean. *Crop Sci* 38:303–308
- Salas JJ, Ohlrogge JB (2002) Characterization of substrate specificity of plant FataA and FatB acyl-ACP thioesterases. *Arch Biochem Biophys* 403:25–34
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet* 46:707
- Schneibly SR, Fehr WR, Welke GA, Hammond EG, Duvick DN (1994) Inheritance of reduced and elevated palmitate in mutant lines of soybean. *Crop Sci* 34:829–833
- Stijšín D, Ablett GR, Luzzi BM, Tanner JW (1998) Use of gene substitution values to quantify partial dominance in low palmitic acid soybean. *Crop Sci* 38:1437–1441
- Stoltzfus DL, Fehr WR, Welke GA (2000) Relationship of elevated palmitate to soybean seed traits. *Crop Sci* 40:52–54
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612
- Thapa R, Carrero-Colón M, Hudson KA (2016) New alleles of FATB1A to reduce palmitic acid levels in soybean. *Crop Sci* 56:1076–1080
- Till B, Afza R, Bado S, Huynh O, Jankowicz-Cieslak J, Matijevic M, Mba C (2009) Global TILLING projects. In: *Induced plant mutations in the genomics era* Food and Agriculture Organization of the United Nations, Rome, pp 237–239
- Voelker T (1996) Plant acyl-ACP thioesterases: chain-length determining enzymes in plant fatty acid biosynthesis. In: Setlow JK (ed) *Genetic engineering*. Springer, Berlin, pp 111–133
- Voelker TA, Jones A, Cranmer AM, Davies HM, Knutzon DS (1997) Broad-range and binary-range acyl-acyl-carrier-protein thioesterases suggest an alternative mechanism for medium-chain production in seeds. *Plant Physiol* 114:669–677

- Wilson R, Marquardt T, Novitzky W, Burton J, Wilcox J, Kinney A, Dewey R (2001) Metabolic mechanisms associated with alleles governing the 16:0 concentration of soybean oil. *J Am Oil Chem Soc* 78:335–340
- Young ND, Bharti AK (2012) Genome-enabled insights into legume biology. *Annu Rev Plant Biol* 63:283–305
- Zhou Z, Lakhssassi N, Cullen MA, El Baz A, Vuong TD, Nguyen HT, Meksem K (2019) Assessment of phenotypic variations and

correlation among seed composition traits in mutagenized soybean populations. *Genes* 10:975

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.