



The Phylogeny and Functional Characterization of Peanut Acyl-ACP Thioesterases

Zhenying Peng^{1,2} · Hui Zhang³ · Haiying Tian¹ · Lei Shan^{1,2} · Zhimeng Zhang⁴ · Hong Ding⁴ · Wenwei Gao³ · Xinguo Li^{1,2}

Received: 1 July 2019 / Accepted: 14 January 2020 / Published online: 28 January 2020
© The Author(s) 2020

Abstract

Fatty acyl-acyl thioesterases (FATs), which hydrolyze the thioester bond linking acyl chains to an acyl carrier protein, thereby terminating their elongation, contribute significantly to the fatty acid (FA) content and composition of seed storage lipids. The peanut (*Arachis hypogaea* L.) genome was found to harbor 21 *FAT* (*AhFAT*) genes, distributed over 12 of the 20 chromosomes. The length of their predicted translation products varied from 74 to 415 residues, and all but one included the 1–2 Acyl-ACP_TE conserved domains. All of the coding sequences were interrupted by at least one intron, with the exon number ranging from two to 12, and five of the genes were liable to alternative splicing. When the RNA-Seq platform was used to assess the transcriptional behavior of the 21 *AhFAT* genes, transcription of only 13 was detectable in samples of root, leaves, and developing seed; among these, six were transcribed throughout the plant, three were root-specific and one was leaf-specific. A detailed analysis of a pair of homologous *AhFATs* showed that the coding region of each was split into six exons and that both were transcribed in all of the plant organs surveyed (although the intensity of their transcription was not the same in immature seed). The product of both genes was deposited in the chloroplast outer membrane. The constitutive expression of these genes in either yeast or *Arabidopsis thaliana* increased the FA content, especially that of saturated FAs. In peanut genome, 21 *AhFAT* genes were found and two of them were transformed into yeast and *Arabidopsis* for function identification. Results showed that overexpression of these two genes could increase the FA content, especially the saturated FAs content.

Keywords Fatty acid · Transcriptional level · Gene structure · Sub-cellular localization · Saturated fatty acids content · Genetic transformation

Abbreviations

AS	Alternative splicing
C14:0	Myristic acid
C16:0	Palmitic acid
C16:1	Palmitoleic acid

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00344-020-10073-9>) contains supplementary material, which is available to authorized users.

✉ Zhenying Peng
pengzhenying2005@126.com

✉ Wenwei Gao
gww@xjau.edu.cn

✉ Xinguo Li
xinguol@163.com

Hui Zhang
562483125@qq.com

Haiying Tian
310303568@qq.com

Lei Shan
shlei1025@163.com

Zhimeng Zhang
qinhdao@126.com

Hong Ding
dingpeanut@163.com

¹ Bio-Tech Research Center, Shandong Academy of Agricultural Science/Shandong Provincial Key Laboratory of Genetic Improvement, Ecology and Physiology of Crops, Jinan 250100, China

² Shandong Normal University, Jinan 250100, China

³ Xinjiang Agricultural University, Urumchi 830052, China

⁴ Shandong Peanut Research Institute, Qingdao 266100, China

C18:0	Stearic acid
C18:1	Oleic acid
C18:2	Linoleic acid
C18:3	Linolenic acid
C20:0	Arachidic acid
C20:1	Eicosenoic acid
C20:2	Eicosadienoic acid
C22:1	Erucic acid
FA	Fatty acid
FAT	Fatty acyl-acyl thioesterase
RT-PCR	Real-time quantitative polymerase chain reaction
Taqman-PCR	Taqman-polymerase chain reaction

Introduction

Fatty acyl-acyl carrier protein thioesterases (FATs) are pivotal enzymes in the synthesis of fatty acids (FAs). Following the termination of acyl chain elongation, FAs are hydrolyzed by FATs and transported to the cytosol (Beisson et al. 2003; Li-Beisson et al. 2010; Sandelius and Aronsson 2008). As a result, FATs determine the acyl chain length of FAs. Two groups of FAT have been recognized: while FATAs prefer unsaturated fatty acids as a substrate, FATBs prefer saturated fatty acids (Jones 1995). Within the FATBs, subgroup 1 enzymes act on long-chain fatty acids (Dörmann et al. 1995; Filichkin et al. 2006; Jha et al. 2010; Moreno-Pérez et al. 2011; Sánchezgarcía et al. 2010), while subgroup 2 enzymes target medium-length chain fatty acids (Filichkin et al. 2006; Jing et al. 2011; Voelker 1996; Voelker et al. 1997). The genetic manipulation of various plant FAT genes has been shown to influence FA composition: for example, the overexpression of the *Arabidopsis thaliana* gene *AtFATB1* results in the accumulation of C16:0 (palmitic acid) (Dörmann et al. 2000). The expression in *Chlamydomonas reinhardtii* of a FATA gene isolated from the algal species *Dunaliella tertiolecta* has the effect of producing 63% and 94% more neutral lipids than the wild type, without compromising growth (Tan and Yuan 2017). When a FATB gene isolated from the cigar plant *Cuphea lanceolata* was expressed in *C. reinhardtii*, the production of C14:0 (Myristic acid)-containing triacylglycerols increased by up to 1.6-fold (Inaba et al. 2017). The heterologous expression of a *Brassica napus* FATB gene in yeast raised the content of the saturated C16:0 and C18:0 (stearic acid) by, respectively, 46% and 22%, at the same time reducing that of both the unsaturated C16:1 (palmitoleic acid) and C18:1 (oleic acid) by, respectively, 15% and 31% (Tan et al. 2015).

The peanut (*Arachis hypogaea* L.) is a leading oil seed crop. Several genes influencing its seed oil content have been identified, among which are a number of FATs. The expression of an *AhFATA* in both *Escherichia coli* and certain

algal species has been shown to be effective in both altering the FA profile and increasing the content of a number of FAs (Chen et al. 2017). Manipulation of *AhFATB1* has been effective in increasing the content of both saturated and unsaturated FAs (Chen et al. 2012, 2017; Wen et al. 2012). With the acquisition of the genome sequence of peanut (Bertioli et al. 2016; Bertioli et al. 2019; Zhuang et al. 2019), it is now possible to identify the species' full complement of *AhFAT* genes and to characterize the functionality of each of them. The aim of the current research was document the family of *AhFAT* genes, to clarify their phylogeny and gene structure, and to profile their transcription in different organs of the plant. Moreover, two not previously isolated FATs have been transformed and functionally analyzed in *Saccharomyces cerevisiae* and *A. thaliana* hosts.

Materials and Methods

Plant and Yeast Materials

The peanut cultivar (cv.) 'Fenghua 1' was used for gene isolation. 'Fenghua 1' was planted in greenhouse with 16-h light/8-h dark photoperiod and 30 °C day/22 °C night temperature cycle. When it got to 12 days, the young mainstem, mainstem leaves, and root tips of ten plants were taken down and frozen in liquid nitrogen. In the full podding bearing stage, seeds of different developmental stages were taken down and frozen in liquid nitrogen. These materials were kept at – 80 °C until use. Genetic transformation experiments involved *A. thaliana* ecotype Col-0 and yeast (*S. cerevisiae*) strain W303. Arabidopsis plants were grown in flower pots in a temperature controlled incubator, with 16-h light/8-h dark photoperiod and 21 °C day/16 °C night temperature cycle. W303 yeast strains were kept on YPD medium (Qingdao Hope Bio-Technology Co. Ltd, China) and grown at 30 °C.

Sequence Analysis

The query sequences used to reveal the family of *AhFAT* genes represented in the species' genome sequence (<https://www.peanutbase.org/>) were *AhFATA* (GenBank accession GU324446.1) and *AhFATB* (EF117305.1). Gene structures were analyzed using GSDS software (gsds.cbi.pku.edu.cn/) and transmembrane domain analysis was conducted using TMHMM software (<http://www.cbs.dtu.dk/services/TMHMM/>). Sub-cellular localization was predicted using the ProtComp 9.0 package (<https://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc>). Conserved domains were identified using SMART software (<https://smart.embl-heidelberg.de/>).

Phylogenetic Analysis

The AhFATA and AhFATB1 sequences were used to scan the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>) for plant FATs (Supplemental Table S1). The resulting sequences were used to construct a FAT phylogeny based on ClustalW software (<https://www.genome.jp/tools-bin/clustalw>), with parameters set to default values. Mega6 software (<https://www.megasoftware.net/>) was used to construct a plant FAT phylogeny, based on the neighbor-joining method; statistical confidence in the clade branching points was obtained by running 1,000 bootstrap replicates (Tamura et al. 2013).

Transcription Profiling Analysis

RNA was extracted from the all roots and leaves of 12-day old cv. ‘Fenghua 1’ seedlings and from developing seeds harvested either 30 or 50 days after flowering (respectively “seed1” and “seed2”). RNA was extracted using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) then treated with RNase-free DNase I (New England Biolabs, USA) for 30 min at 37 °C to degrade any contaminating DNA present. The concentration and purity of the resulting RNA preparations were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and its integrity was checked using an RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA). A 1.5 µg aliquot of RNA was processed with a Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA) to remove the rRNA component, and the subsequently prepared sequencing libraries based on the residual RNA, following treatment with an NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (New England Biolabs, USA). Index codes were added to enable each sequence to be attributed its organ provenance. Paired-end sequences were generated by an Illumina HiSeq2500 platform. The resulting sequence data, stripped of adapter sequence and low-quality reads, have been deposited with NCBI (SRA ID: SRP093901). The sequences were aligned with the genome sequence of the peanut cv. ‘Tifrunner’ (https://www.peanutbase.org/peanut_genome) using TopHat2 software (Kim et al. 2013). Aligned reads were assembled into a full transcriptome using the Cufflinks v2.2.1 program (Trapnell et al. 2010). FPKM values for each transcript were used to estimate transcript abundance. The expression levels of *FATBs* from *Physcomitrella patens*, *Populus trichocarpa*, and *Glycine max* are obtained from Phytozome v12.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>).

AhFATB2 Isolation

Seed1 cDNA was provided as the template for PCRs driven by either primer pair AhFATB2-1-F/R or AhFATB2-2-F/R (Supplemental Table S2). Each 25 µL PCR contained 1 µL cDNA (100 ng/µL), 2 µL dNTP (10 mM each), 2.5 µL 10× *TransTaq* HiFi buffer (<https://www.transgenbiotech.com>), 1 µL *TransTaq* HiFi DNA polymerase (<https://www.transgenbiotech.com>), 1 µL of each primer (10 µM), and 16.5 µL ddH₂O. The amplification program comprised an initial denaturation (95 °C/5 min), followed by 35 cycles of 95 °C/30 s, 60 °C/30 s, 72 °C/35 s, with a final elongation of 72 °C/10 min. The amplicons were resolved by electrophoresis through a 1% agarose gel, and sequenced by Sangon Biotech (Shanghai, China).

Isolation of RNA and Semi-quantitative PCR

Total RNA was extracted from the root, stem, leaf, flower, and seeds at 10, 20, 30, 40, 50, and 60 days after flowering of cv. ‘Fenghua 1’ plants, using a DP441 RNAprep Pure Plant Kit (Tiangen, Beijing, China), and was reverse transcribed using a Thermo Scientific RevertAid First Strand cDNA Synthesis kit (<https://www.thermofisher.com/cn/zh/home.html>). A Taqman-PCR assay (Agilent Technologies, California, USA) was used to determine transcript abundance, based on the primer pairs FATB2F/R (Supplemental Table S2) for the amplification of a fragment of *AhFATB2* and the fluorescently labeled pairs FATB2-1-P/2-P (Supplemental Table S2) for the detection of, respectively, *AhFATB2-1* and *AhFATB2-2*. The reference sequence was *Actin* (GenBank number GO264911), assayed with the primer pair AhActin-S/-A (Supplemental Table S2). Each 20 µL PCR contained 10 µL TaqMan Fast qPCR Master Mix, 0.4 µL of each of AhFATB2F/R (10 µM each), 0.4 µL AhFATB2-1-P or AhFATB2-2-P (10 µM), 2 µL cDNA (100 ng/µL), and 6.8 µL ddH₂O. The amplification program comprised an initial denaturation (94 °C/3 min), followed by 45 cycles of 95 °C/5 s, 57 °C/15 s, 72 °C/30 s, with a final elongation of 72 °C/10 min. The relative transcript abundance of each *FATB2* was calculated using the 2^{-ΔΔCT} method (Livak and Schmittgen 2001). Each experiment was run in triplicate.

Sub-cellular Localization

The open reading frames (lacking the stop codon) of *AhFATB2-1* and *AhFATB2-2* were PCR-amplified using the primer pairs FATB2-1-GFP-F/R and FATB2-2-GFP-F/R (Supplemental Table S2), after which the amplicons were treated with *Hind* III and *Bam* HI, then cloned into the multiple cloning site of the pBSK+-35S-EGFP plant transient expression vector (kindly provided by Weicai Yang, Institute

of Genetics and Developmental Biology, Chinese Academy of Science). The constructs were transformed into *A. thaliana* mesophyll protoplasts using the PEG-calcium transfection method (Yoo et al. 2007). Fluorescence generated by the expression of *GFP* was observed by confocal laser scanning microscopy.

Yeast Transformation

AhFATB2 fragments were PCR-amplified using the primer pairs FATB2-1-BamHI-F/-KpnI-R or FATB2-2-BamHI-F/-KpnI-R (Supplemental Table S2), after which the amplicons were treated with *Kpn* I and *Bam*H I, then inserted into the multiple cloning site of the galactose-inducible yeast expression vector pESC-URA (Agilent Technologies). The resulting constructs plasmids were transformed into *S. cerevisiae* strain W303 competent cells which were grown on a selective SD-URA medium. Pooled colonies from each strain were grown in a selective liquid medium, then diluted into a liquid medium containing galactose and raffinose to induce protein expression. Cells harboring an empty plasmid served as the negative control. The cells were dried, then analyzed for their fatty acid content, following a gas chromatography protocol described by Zheng et al. (2017).

A. thaliana Transformation

The *AhFATB2* open reading frames were PCR-amplified using the primer pairs FATB2-1-KpnI-F/-SacI-R or FATB2-2-KpnI-F/-SacI-R (Supplemental Table S2) and the resulting amplicons inserted first into the binary vector PRI101 (kindly provided by Minqin Wang, Shandong University, China) and thence into *Agrobacterium tumefaciens* strain LBA4404 (Poirier et al. 2000). *A. thaliana* plants were transformed using inflorescence infection method (Clough and Bent 1998). Progeny of putative transgenics were plated on half-strength MS medium (Murashige and Skoog 1962) containing 100 mg/L kanamycin in order to select for transgene-positive T₁ individuals, and this procedure was repeated to obtain the T₂ and T₃ generations. The progeny of T₃ plants were analyzed for their fatty acid content, following a gas chromatography protocol described by Zheng et al. (2017).

Results

The *AhFAT* Gene Content of the Cultivated Peanut Genome

In total, 21 potential *AhFATs* were uncovered from *A. hypogaea* genome (Supplemental Table S3); four loci mapped to each of chromosomes Arahy.12 and Arahy.13, three to Arahy.01, two to Arahy.02, and the other eight each

mapped to different chromosomes. Two of the 21 genes were *FATAs*, namely *Arahy.YI95N7.1* (on chromosome Arahy.04) and *Arahy.QR9F0J.1* (Arahy.14); the 19 *AhFATB* genes were distributed across ten chromosomes. On the basis of their sequences, *Arahy.YI95N7.1* was deemed to be a synonym of the previously described gene *AhFATA*, as was *Arahy.KT8BMX.1* of *AhFATB1* (Chen et al. 2012; Wen et al. 2012). The length of the predicted translation products ranged from 74 (Arahy.L512HM) to 415 (Arahy.SJ2WML) residues. By sequence alignment in NCBI, we found Arahy.L512HM appeared to be an incomplete protein. By comparing the gene sequences with their genomic sequences, we found five of the genes produced more than one isoform: while both *Arahy.QR9F0J* and *Arahy.L8EKIA* produced three splicing variants, *Arahy.H8W5PF*, *Arahy.LCII34*, and *Arahy.3BV5AA* each produced two. Alternative splicing (AS) led to the formation of one transmembrane domain of Arahy.QR9F0J.3. Arahy.QR9F0J.3 was the only one which had transmembrane domain, whereas the other *AhFATs* had none. It had no effect on the products' sub-cellular localization, with all 21 products predicted to be deposited in the chloroplast. With the exception of the very short Arahy.L512HM.1, all of the proteins possessed a 1–2 Acyl-ACP_TE domain, with six also possessing an Acyl-thio_N domain.

There was extensive variation with respect to the gene structures of the *AhFATs*, with the number of exons ranging from two to twelve (Fig. 1). The shorter genes *Arahy.L512HM.1*, *Arahy.C2SFDP.1*, *Arahy.0VM2VP.1*, *Arahy.QR9F0J.3*, and *Arahy.5L01PW.1* each featured only two or three exons, while the others' coding sequences were split into between six and 12 fragments. The *Arahy.L8EKIA.1* coding sequence was interrupted by 11 introns and featured three alternative splicing variants (one complete, one comprising eight exons, and one of seven exons); these isoforms derived from exon skipping, alternative transcription start sites and alternative transcription termination sites. The three isoforms generated by *Arahy.QR9F0J* were of length eight, six, and four exons. Intron retention was observed for both *Arahy.QR9F0J.1* and *Arahy.QR9F0J.2*.

Phylogeny of the *AhFATs*

After removing gene fragments and incomplete sequences, the scan of plant *FATs* delivered a set of the 179 sequences from 26 species (Supplemental Table S1) used to construct a *FAT* phylogeny (Fig. 2). The representative from *C. reinhardtii* was located closest to the root of the tree, while the remainder formed two large groups (the *FATAs* and the *FATBs*). The two subgroups *FATB1* and *FATB2* were recognized: the former consisted entirely of genes harbored by angiosperm species, and comprised four branches (I through IV). The dicotyledoneae *FATBs* (including

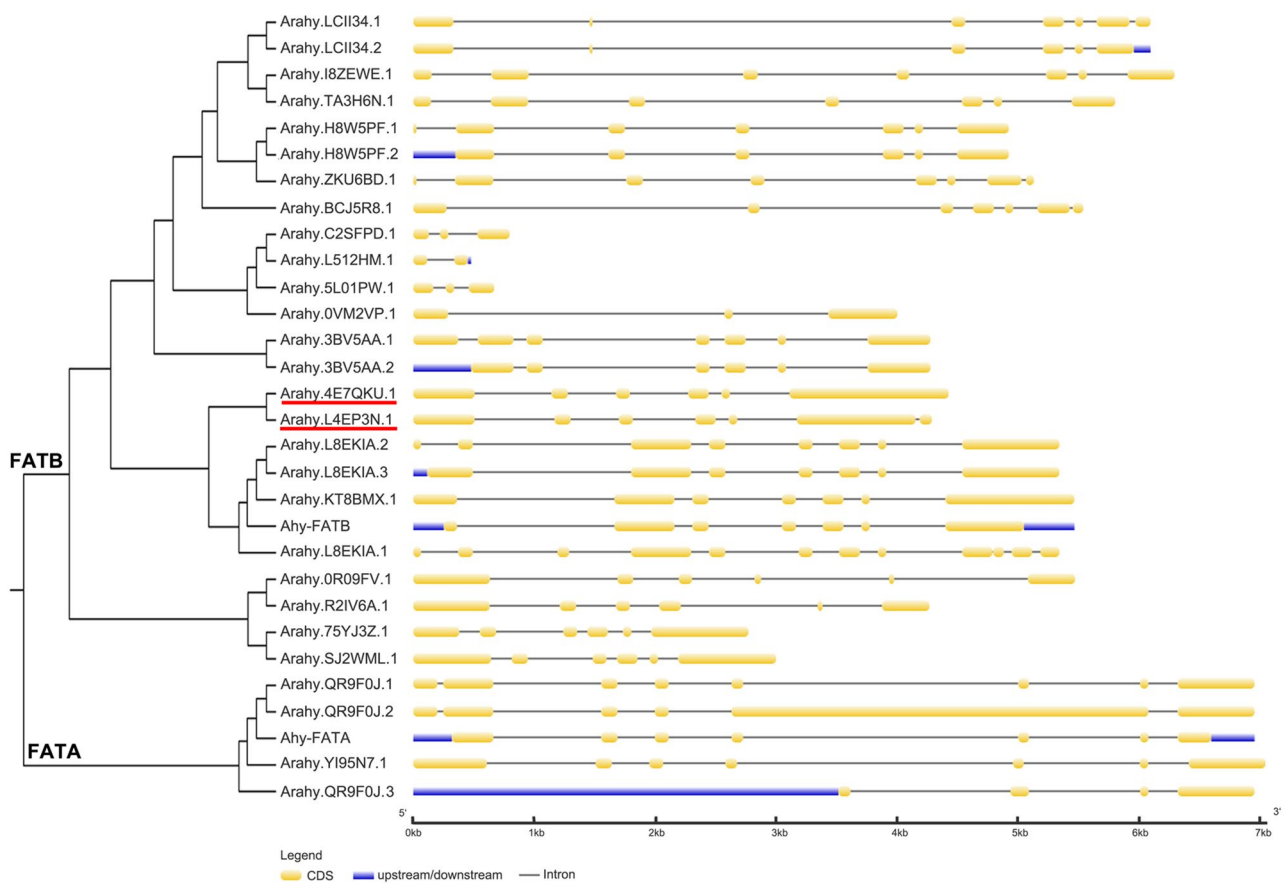


Fig. 1 Gene structure analysis of AhFATs. *Note* Ahy.FATA and Ahy.FATB were identified in other papers (Wen et al. 2012). The two genes with red lines are the genes identified in this paper

some of the peanut FATBs) clustered within branch I, whereas branch II consisted entirely of monocotyledoneae FATBs, branch III only of peanut FATBs, and branch IV of FATBs from *Prunus persica* and *Populus trichocarpa*. With respect to the FATB2 subgroup, genes harbored by primitive land plant species located close to the root, and the remainder clustered into two groups (branches V and VI); and branch V contained all of the other primitive land plant genes, and branch VI the angiosperm genes. Peanut FATBs were distributed over the four branches I, III, V, and VI, each branch consisted of four, seven, two, and two AhFATBs, respectively. The FATAs produced by *Marchantia polymorpha* (a liverwort), *Selaginella moellendorffii* (a lycophyte), and *Physcomitrella patens* (a bryophyte) were located close to the root, and the remainder formed branches VII and VIII of the phylogenetic tree. Branch VII members were from monocotyledoneae and some dicotyledoneae species, while branch VIII clustered genes exclusively from dicotyledoneae (including the AhFATAs). The most closely related sequences to the AhFATs were from other Leguminosae, such as soybean and *Medicago truncatula*.

Transcription Analysis

When the abundance of the various AhFAT transcripts was assessed using RNA-Seq data, only 13 of the 21 genes were represented (Fig. 3). Both AhFATAs (*Arahy.YI95N7.1* and *Arahy.QR9F0J.1*) were transcribed in all four samples, and especially in the seed1 and leaf samples. Four of the AhFATB genes (*Arahy.KT8BMX.1*, *Arahy.L8EKIA.1*, *Arahy.L4EP3N.1*, and *Arahy.4E7QKU.1*, all belong to branch I) were also ubiquitously transcribed; the abundance of *Arahy.KT8BMX.1* transcript was particularly high in the seed1 sample. Branch III had seven AhFATBs, of which three (*Arahy.OVM2VP.1*, *Arahy.3BV5AA.1*, and *Arahy.BCJ5R8.1*) were expressed. *Arahy.OVM2VP.1* and *Arahy.3BV5AA.1* were root-specific, whereas *Arahy.BCJ5R8.1* had a low expression level (with FPKM = 0.81) specifically in the leaf. Branch V had two members, namely, *Arahy.OR09FV.1* and *Arahy.R21V6A.1*. *Arahy.OR09FV.1* was leaf-specific, whereas *Arahy.R21V6A.1* was expressed in seed1 and leaf. Branch VI had two members, namely, *Arahy.75YJ3Z.1* and *Arahy.SJ2WML.1*. *Arahy.75YJ3Z.1* was root-specific, whereas *Arahy.SJ2WML.1* had a low expression level in

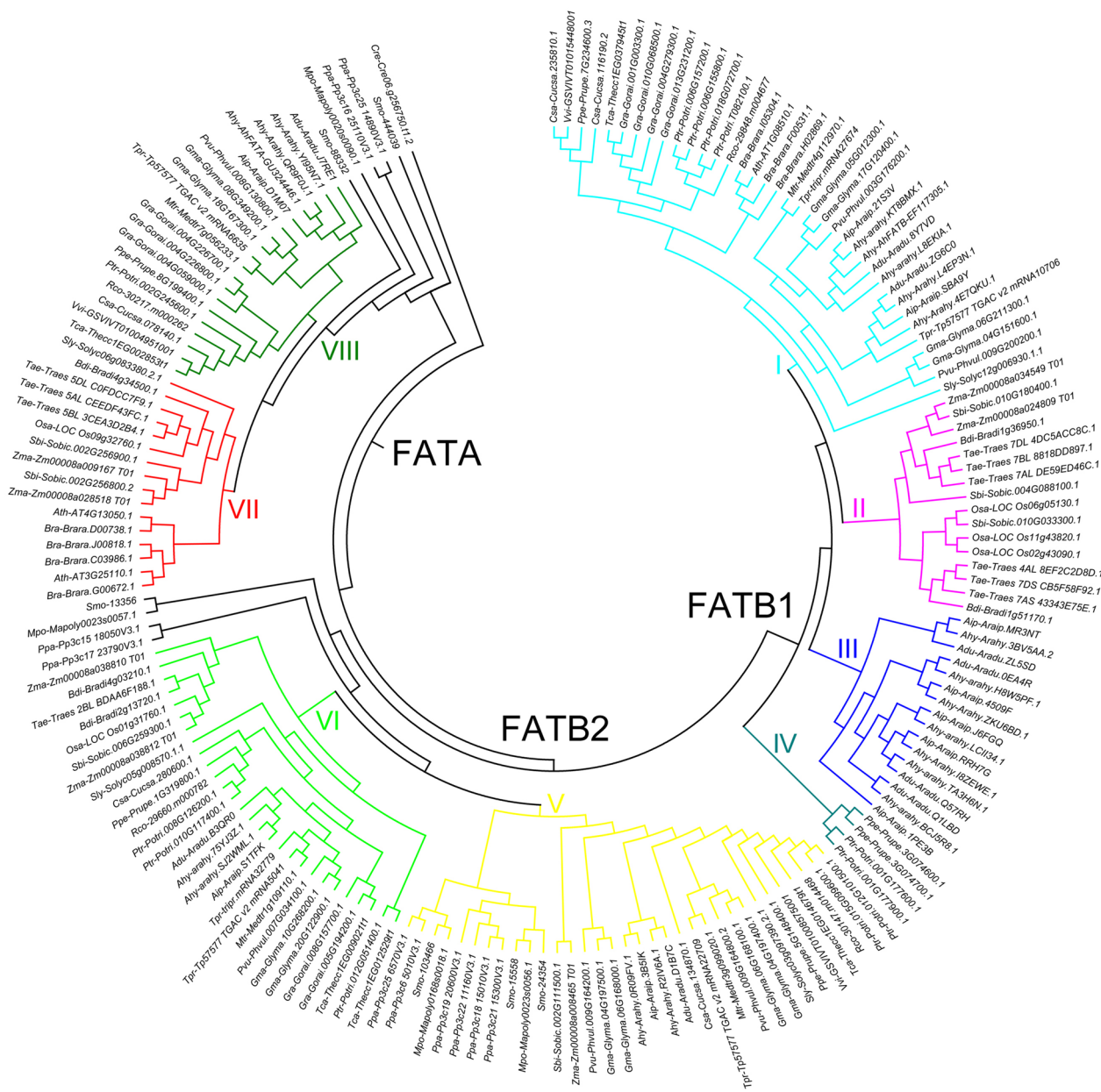


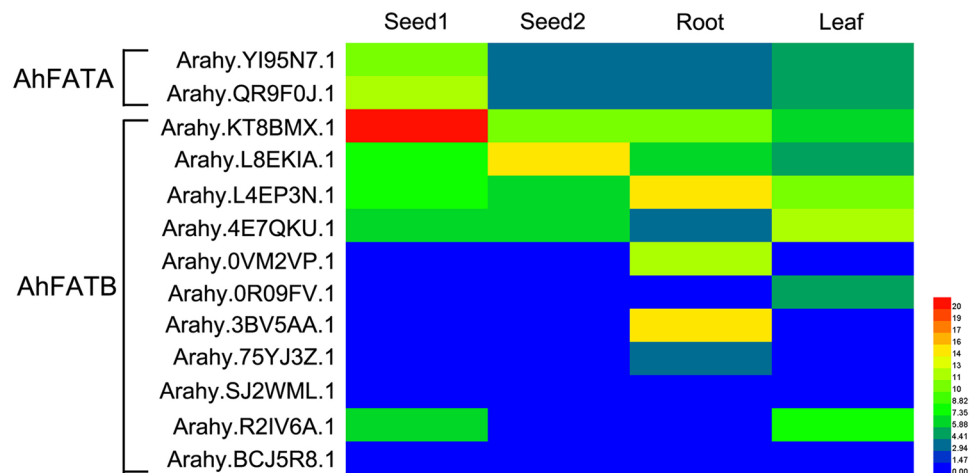
Fig. 2 Phylogenetic tree analysis of plant FATA and FATB1, FATB2 genes. The tree was conducted with Mega6 software based on the neighbor-joining method. Statistical confidence in the clade branching points was obtained by running 1,000 bootstrap replicates

seed1, root, and leaf (with FPKM values of 0.5, 0.24, and 0.95, respectively) but no expression in seed2. Combined with RNA-Seq data (Fig. 3) and phylogenetic tree analysis (Fig. 2), the results demonstrated that the *AhFATB* genes in branch I were ubiquitously transcribed in all tissues, whereas the *FATB* genes in branches III–VI were expressed with tissue-specific patterns.

To verify the universality of this phenomenon, we analyzed the expression patterns of *FATBs* from *P. patens*, *P. trichocarpa*, and *G. max* (Fig. S1A–C). The three species have many *FAB* genes in their genomes, similar to peanut.

Moreover, they exhibit different levels of evolution. The *P. patens* genome harbored eight *FATBs*, two of which (i.e., *Pp3c15_18050V3.1* and *Pp3c17_23790V3.1*) were located closest to the root of the *FATB* group, while the others were located in branch V (Fig. 2). *Pp3c19_20600V3.1* and *Pp3c18_15010V3.1* were ubiquitously transcribed in all tissues at high levels, and the others were expressed in different tissues (Fig. S1A). The *P. trichocarpa* genome harbored 11 *FATBs*, which were distributed over branches I, IV, V, and VI, with each branch consisting of four, two, two, and three *FATBs*, respectively (Figs. 2, S1B). Four members in branch

Fig. 3 Expression pattern analysis of AhFATs. FPKM values for each transcript were used to estimate transcript abundance



I were ubiquitously transcribed in all tissues at high levels, while the others were differentially expressed (Fig. S1B). The *G. max* genome harbored 10 *FATBs*, which were distributed over branches I, V, and VI, with each branch consisting of four, four, and two *FATBs*, respectively (Figs. 2, S1C). Four members in branch I were ubiquitously transcribed at high levels in all tissues, while the others were differentially expressed (Fig. S1C). The sub-functionalization of *FABs* started early, such as in the *P. patens* period. Comparatively, the evolution of the *FATA* genes was conservative. Most plant genomes harbored one to four *FATA* genes (Fig. 2), but only one or two expressed ubiquitously in all tissues. The rest were expressed with different expression patterns (Fig. S1).

AhFATB2 Isolation and Analysis of Its Sequence

When cDNA prepared from seeds harvested 30 days after flowering was amplified using the primer pairs FATB2-1-F/R and FATB2-2-F/R (Supplemental Table S2), the two sequences thereby generated proved to be highly similar (99.4%) to those of *Arahy.LAEP3N.1* and *Arahy.4E7QKU.1*. (Their sequences, designated *AhFATB2-1* and *AhFATB2-2*, have been archived in GenBank under accession numbers MH105081 and MH105082). The two sequences, differing from one another at just eight nucleotides, comprised a 1,245 nt open reading frame, encoding the identical 414 residue polypeptide. Both shared the three conserved catalytic residues Asp311, Asp313, and Cys348, as well as the Acyl-ACP_TE domain. Their level of similarity with the sequence of *AhFATB* (GenBank number EF117305.1) was 91.4%. *AhFATB2-1* mapped to chromosome Arahy.10 and *AhFATB2-2* to chromosome Arahy.20, a pair of chromosomes recognized as homologous; they shared a different gene structure (six and seven exons, respectively) and variation in their intronic sequence (Fig. 1).

Semi-quantitative RT-PCR Analyses of *AhFATB2-1* and *AhFATB2-2*

The very high similarity between the *AhFATB2-1* and *AhFATB2-2* sequences made it impossible to distinguish their transcripts from one another using conventional RT-PCR. To achieve the necessary discrimination, a Taqman-PCR method was employed. A comparison of their transcript abundance in the root, stem, leaf, and flower, sampled at 10, 20, 30, 40, 50, and 60 days after flowering, showed that they shared a very similar transcriptional profile. Both were more abundantly transcribed in the flower than elsewhere (Fig. 4a). In seeds, the expression level of *AhFATB2-1* and *AhFATB2-2* increased with the seeds development from 10 to 30 d, then peaked at 30 days, and decreased along with the seeds maturation. The period from 20 to 40 d was the most important stage of peanut seed development, the high expression level of *AhFATBs* was in accordance with the rapid oil accumulation in this stage (Fig. 4b).

Sub-cellular Localization of *AhFATB2-1* and *AhFATB2-2*

When the two genes were expressed in *A. thaliana* protoplasts, the transgene was expressed most strongly in the vicinity of the chloroplasts, specifically in their outer membranes (Fig. 5).

FA Analysis of Transgenic Yeast

An analysis of the FA content of non-transformed yeast cells highlighted four major fatty acid types, namely C18:1, C16:1, C16:0, and C18:0; these accounted for, respectively, 37.5%, 34.6%, 15.1%, and 8.9% of the total FA content. An equivalent analysis of the transgenic yeast cells showed that cells expressing *AhFATB2-1* accumulated 35.0% of the wild type level of C18:1, 50.1% of C16:1, 197.9% of C16:0, and 319.6%

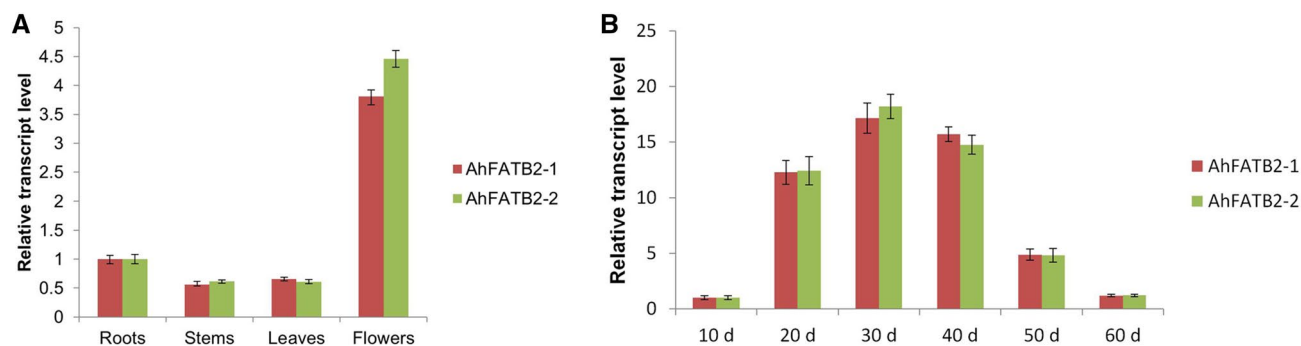
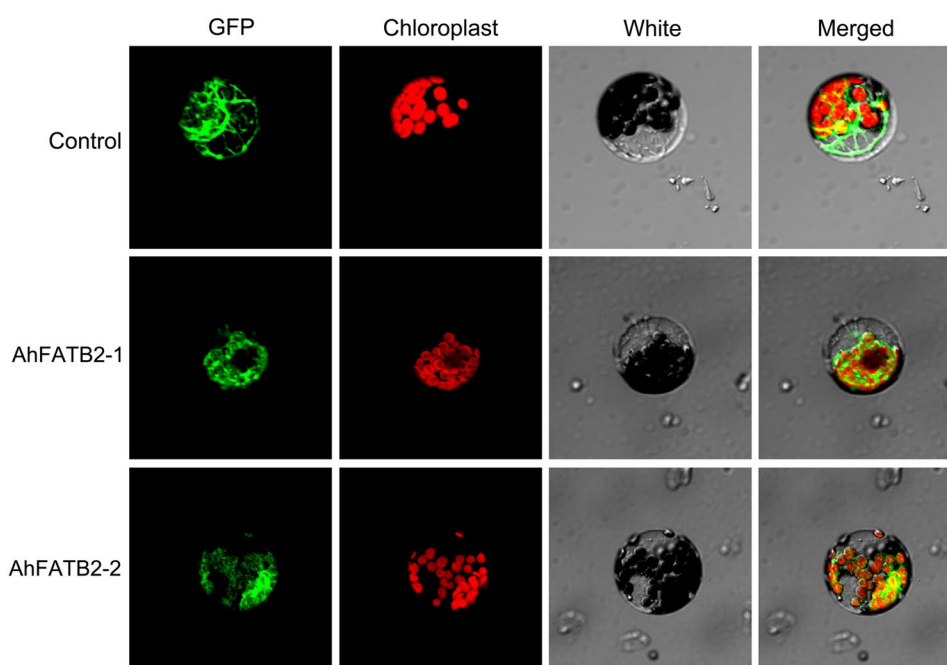


Fig. 4 Expression pattern analyses of AhFATB2s in different organs (a) and different development stages of peanut seeds (b). From twelve-day-old plants, the young mainstem, mainstem leaves,

and root tips of about ten plants were taken for RT-PCR. Seeds of 15 days, 30 days, 45 days, and 60 days after flowering were used. Three replicates were conducted

Fig. 5 Sub-cellular localization of AhFATB2s in *Arabidopsis* protoplasts. Note GFP indicating green fluorescent signal, Chloroplast indicating the autofluorescence of chloroplast, White indicating white light images, Merged indicating fused images



of C18:0, while the *AhFATB2-2* transformants behaved similarly (40.1% of the wild type level of C18:1, 45.5% of C16:1, 170.0% of C16:0, and 277.4% of C18:0). The overall increase in FA content was 92.1% for the *AhFATB2-1* transformants and 85.4% for the *AhFATB2-2* transformants (Fig. 6a). The constitutive expression of both *AhFATB2s* in yeast lines clearly altered the composition of FAs (Fig. 6b): the content of saturated FAs (C16:0 and C18:0) increased, whereas that of unsaturated FAs (C16:1 and C18:1) decreased.

The effect of *AhFATB2* constitutive expression in *A. thaliana*

When the two *AhFATB2* genes were constitutively expressed in *A. thaliana*, the FA content of the seeds was raised from

167.6 mg/g dry weight to, respectively, 283.3 mg/g dry weight and 281.2 mg/g dry weight (Fig. 7a). In the seed of non-transgenic *A. thaliana*, nine fatty acid types were detected, namely C16:0, C18:0, C18:1, C18:2 (linoleic acid), C18:3 (linolenic acid), C20:0 (arachidic acid), C20:1 (eicosenic acid), C20:2 (eicosadienoic acid), and C22:1 (erucic acid). The effect of the transgene was to raise the content of six of these, and in particular that of C16:0 and C18:0; meanwhile, the content of both C18:1 and C20:1 was reduced. In non-transgenic *A. thaliana* seeds, the predominant fatty acids were unsaturated (C18:2, C18:3, C20:1, C18:1), accounting for 67.5% of the total fatty acid content (Fig. 7b). However, in the seed of the transgenic *A. thaliana* lines, C16:0 became the most abundant fatty acid, representing around 31% of the total fatty acid content; at the

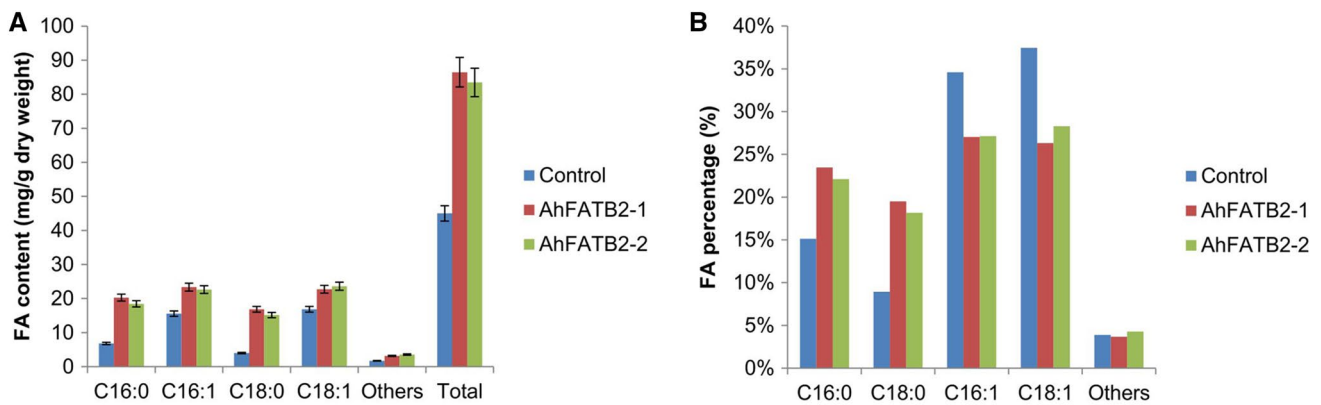
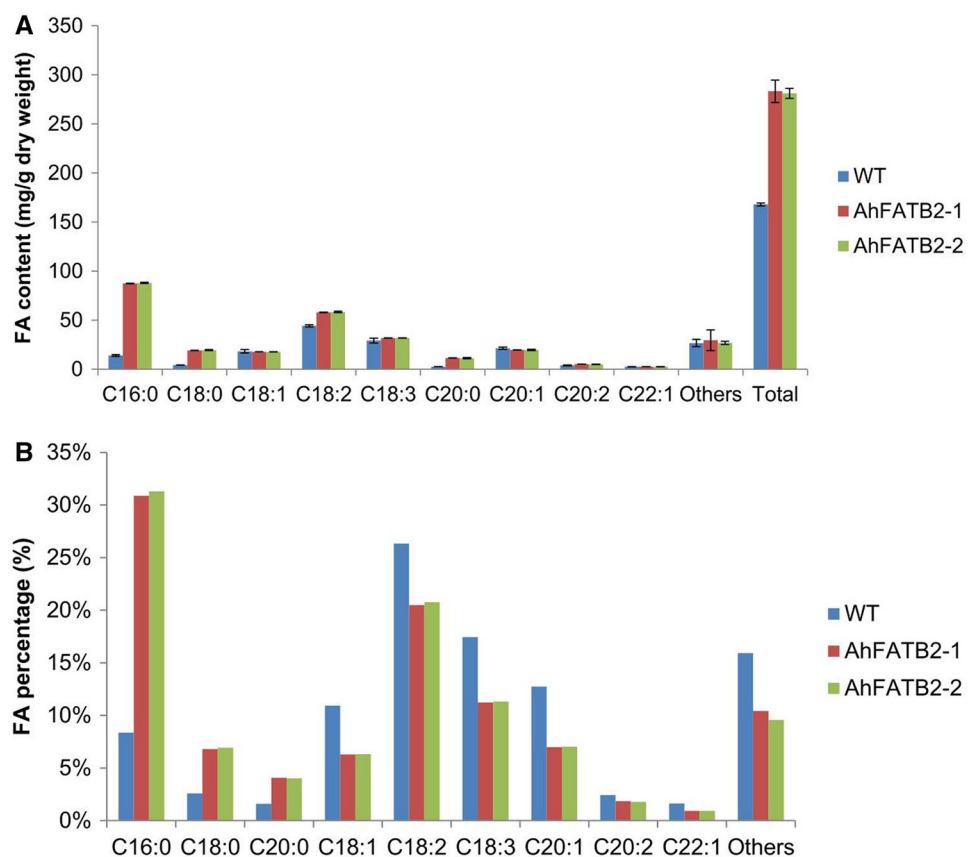


Fig. 6 Analysis of FA content in transgenic and wild type yeast lines. Note: compared with control, * $P < 0.01$. Three replicates were conducted

Fig. 7 Analysis of the FA content and percentage in wild and transgenic Arabidopsis seeds. **a** FA content; **b** FA percentage. Note compared with WT, * $P < 0.05$, ** $P < 0.01$. Three replicates were conducted



same time, the content of the saturated fatty acids C18:0 and C20:0 was also significantly increased.

Discussion

Thioesterases are ubiquitous in both prokaryotic and eukaryotic organisms. Based on their catalytically active site, their oligomerization, and their substrate specificity, they

have been allocated EC numbers from 3.1.2.1 to 3.1.2.27, with many remaining unclassified (EC 3.1.2.-) (Cantu et al. 2010). The FATs (EC 3.1.2.14) hydrolyze FAs synthesized by the FA synthase complex. The FATs found in bacteria and plants have been classified into ten subfamilies (Jing et al. 2011); here, a collection of 179 documented plant FATs, which included both FATAs and FATBs, fell into eight clades (Fig. 2). Peanut FATBs were distributed over the four branches I, III, V, and VI. Of particular interest

was the finding that one of these (branch III) was composed entirely of FATBs synthesized by *Arachis* spp. (seven by *A. hypogaea*, five by *A. ipaensis* and four by *A. duranensis*) (Fig. 2). A likely explanation for this rather surprising outcome is that these genes have evolved from a single ancestral copy by duplication. Of the seven *A. hypogaea* branch III FATBs, only two (*Arahy.3BV5AA.1* and *Arahy.BCJ5R8.1*) appeared to be active (Fig. 3): transcripts of the former gene were particularly abundant in the root, whereas the latter was transcribed—at a rather low level—exclusively in the leaf.

The large number of *FAT* genes present in the genome varies widely between species. This variation cannot be accounted for by suggesting that their number simply accumulates during evolution, since although the algal species *C. reinhardtii* encodes just one *FAT*, the genomes of the primitive land species *M. polymorpha* (a liverwort), *S. moellendorffii* (a lycophyte), and *P. patens* (a bryophyte) harbor, respectively, four, six, and ten *FATs*, a greater number than is harbored by most angiosperm species (Supplemental Table S1). Nor can the large number of *FAT* genes harbored by peanut (21*AhFATs*) does not be explained by its being a tetraploid, since the hexaploid species bread wheat harbors just ten genes. Finally, there is no discernible relationship between the number of *FAT* genes and genome size, as can be seen from a comparison between the small genome species *A. thaliana* and the large genome species bread wheat. A possibility which remains to be tested is that selection for high oil content in the seed tends to favor the accumulation of *FAT* genes.

Based on RNA-Seq data, only 13 of the 21 peanut *FATs* were transcribed in root, leaf, and seed tissue (Fig. 3), and only six of these were ubiquitously transcribed. This observation implies a degree of sub-functionalization among the genes, as well as suggesting a high rate of gene silencing. However, the present experiment did not cover many other organs in the plant, such as the flower, the stem, and the pod. Furthermore, some *FATs* appear to be involved in the stress response (Li et al. 2017). Given the importance of lipid synthesis in the seed, it was unexpected that only one in three of the active genes were transcribed in the seed, while almost half were active in the root and leaf. The implication is that a wider spectrum of *FAs* is required by vegetative organs than is required in the seed. Lipids account for about half of the peanut seed's dry weight, from which about 80% is in the form of C18:1 or C18:2 compounds. The remarkable gene *Arahy.KT8BMX.1* expressed at extremely high level in seed1 stage may be responsible for the major lipids synthesis. Lipid storage is of minor importance in both the root and the leaf, but a wider range of *FATs* may be required to support the complexity of physiological reactions relying on membrane lipids.

Based on substrate preference, the *FATs* have been divided into two groups: while the *FATAs* prefer unsaturated

FAs, *FATBs* prefer saturated *FA*, as shown by observation that co-expressing in *E. coli* the *Cinnamomum camphora* gene *CcFatB1* with *E. coli fadR* and *Bacillus megaterium P450_{BM3}* boosts the content of ω -3-OH-C14:1 (Xiao et al. 2018). Three *FATs* produced by coconut share a similar substrate specificity, producing predominantly C14:0 and C16:1 compounds (Jing et al. 2011). At the same time, different *FATs* produced by a given same species do not necessarily differ in their function, since three *FATAs* and two *FATBs* encoded by *Sorghum* appeared to show a very similar substrate specificity when heterologously expressed in *E. coli* (Jing et al. 2011). Similarly, when *AhFATA* and *AhFATB1* were expressed in the cyanobacterium *Synechocystis* sp. PCC6803, both increased the content of C16:1, C18:0, C18:2, and C18:3 compounds (Chen et al. 2017). Here, the two *AhFATB2s* constitutively expressed in both yeast and *A. thaliana* appeared to prefer saturated over unsaturated *FAs* (Figs. 6, 7).

In response to the world energy crisis and global climate change, biodiesel has attracted renewed interest in the exploration for sustainable development (Zinoviev et al. 2010). Free fatty acids (FFAs) have been extensively applied in the manufacture of biofuels, cosmetics, and pharmaceutical drugs (Shin et al. 2016). Given that the overexpression of *FATs* has considerably increased the production of FFAs in prokaryotes and eukaryotes (Chen et al. 2017; Li et al. 2017; Lin and Lee 2017; Shin et al. 2016; Tan and Yuan 2017), *FATs* remain the first choice in the production of oleochemicals. The substrate specificity of *FATs* often determines the composition of oils. Thus, many plant *FATs* have been isolated and remolded to produce oleochemicals with desired chain lengths (Lozada et al. 2018). Peanut is an important oil crop (with seed oil content up to 50%), and its genome possesses 21 *FATs*, which are far more than any other species (Supplemental Tables S1, S3). Not all *AhFATs* are suitable for oleochemical production. At least four *AhFATs*, namely, *Arahy.KT8BMX.1*, *Arahy.L8EKIA.1*, *Arahy.LAEP3N.1*, and *Arahy.4E7QKU.1*, with high expression levels in branch I are good choices for biofuel production. Studies have demonstrated that all these *AhFATs* can increase the oil content of the transformed microalgae (Chen et al. 2012, 2017; Li et al. 2017; Wen et al. 2012), yeast, and plants (Figs. 6, 7). In addition, the *FATs* in branch I from other plants are good candidates for biofuel production because of their high expression level (Supplemental Fig. S1).

Conclusions

In summary, we analyzed the peanut *FAT* gene family in peanut genome, and analyzed their phylogeny relationship with other plant *FATs*. Here, two new *AhFATB* genes from cultivated peanut were cloned and transformed into *S.*

cerevisiae and *A. thaliana* for function verification. Results showed that both of them could increase the FA content significantly, especially that of saturated FAs.

Acknowledgements This work was supported by the National Key R&D Program of China (2018YFD1000900), Shandong Province Germplasm Innovation (2017LZN035), Major Basic Research Project of Natural Science Foundation of Shandong Province (2018GHZ007), Major Scientific and Technological Innovation Projects of Shandong Province (2018YFJH0601), the Earmarked Fund for Modern Agroindustry Technology Research System (CARS-13), and Agricultural Science and Technology Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2018B05, CXGC2018D04). The founders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Hui Zhang, Haiying Tian, Lei Shan, Zhimeng Zhang and Hong Ding. The first draft of the manuscript was written by Zhenying Peng, Wenwei Gao, Xinguo Li and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Beisson F et al (2003) Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol* 132:681. <https://doi.org/10.1104/pp.103.022988>
- Bertioli DJ et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet* 48:438–446. <https://doi.org/10.1038/ng.3517>
- Bertioli DJ et al (2019) The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet* 51:877–884. <https://doi.org/10.1038/s41588-019-0405-z>
- Cantu DC, Chen Y, Reilly PJ (2010) Thioesterases: a new perspective based on their primary and tertiary structures. *Protein Sci* 19:1281–1295. <https://doi.org/10.1002/pro.417>
- Chen G et al (2012) Functional expression analysis of an acyl-ACP thioesterase FatB1 from *Arachis hypogaea* L. seeds in *Escherichia coli*. *J Food Agric Environ* 10:332–336
- Chen G et al (2017) Functional expression of the *Arachis hypogaea* L. Acyl-ACP thioesterases AhFatA and AhFatB enhances fatty acid production in *Synechocystis* sp. PCC6803. *Energies*. <https://doi.org/10.3390/en10122093>
- Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–742. <https://doi.org/10.1046/j.1365-313x.1998.00343.x>
- Dörmann 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. <https://doi.org/10.1006/abbi.1995.1081>
- Dörmann P, Voelker TA, Ohlrogge JB (2000) Accumulation of palmitate in Arabidopsis mediated by the acyl-acyl carrier protein thioesterase FATB1. *Plant Physiol* 123:637–644. <https://doi.org/10.1104/pp.123.2.637>
- Filichkin SA, Slabaugh MB, Knapp SJ (2006) New FatB thioesterases from a high-laurate *Cuphea* species: functional and complementation analyses. *Eur J Lipid Sci Technol* 108:979–990. <https://doi.org/10.1002/ejlt.200600158>
- Inaba Y, Nakahigashi K, Ito T, Tomita M (2017) Alteration of fatty acid chain length of *Chlamydomonas reinhardtii* by simultaneous expression of medium-chain-specific thioesterase and acyl carrier protein. *Phycol Res* 65:94–99. <https://doi.org/10.1111/pre.12161>
- Jha SS, Jha JK, Chattopadhyaya B, Basu A, Sen SK, Maiti MK (2010) Cloning and characterization of cDNAs encoding for long-chain saturated acyl-ACP thioesterases from the developing seeds of *Brassica juncea*. *Plant Physiol Biochem* 48:476. <https://doi.org/10.1016/j.plaphy.2010.02.006>
- Jing F, Cantu D, Tvaruzkova J, Chipman J, Nikolau B, Yandea-Nelson M, Reilly P (2011) Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochem* 12:44. <https://doi.org/10.1186/1471-2091-12-44>
- Jones A (1995) Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell* 7:359–371. <https://doi.org/10.1105/tpc.7.3.359>
- Kim D et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>
- Li HY et al (2017) Cloning and expression analysis of fatty acyl-ACP thioesterase (FATB2) gene in peanut. *J Peanut Sci*. <https://doi.org/10.14001/j.issn.1002-4093.2017.04.002>
- Li-Beisson Y et al (2010) Acyl-lipid metabolism. *Arabidopsis Book* 8:e0133. <https://doi.org/10.1199/tab.0161>
- Lin H, Lee YK (2017) Genetic engineering of medium-chain-length fatty acid synthesis in *Dunaliella tertiolecta* for improved biodiesel production. *J Appl Phycol* 29:2811–2819. <https://doi.org/10.1007/s10811-017-1210-7>
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>
- Lozada NJH, Lai R-Y, Simmons TR, Thomas KA, Chowdhury R, Maranas CD, Pflieger BF (2018) Highly active C8-Acyl-ACP thioesterase variant isolated by a synthetic selection strategy. *ACS Synth Biol* 7:2205–2215. <https://doi.org/10.1021/acssynbio.8b00215>
- Moreno-Pérez AJ, Sánchez-García A, Salas JJ, Garcés R, Martínez-Force E (2011) Acyl-ACP thioesterases from macadamia (*Macadamia tetraphylla*) nuts: cloning, characterization and their impact on oil composition. *Plant Physiol Biochem* 49:82. <https://doi.org/10.1016/j.plaphy.2010.10.002>
- Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassay with tobacco tissue culture. *Physiol Plant* 15:473–497. <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>

- Poirier Y, Ventre G, Nawrath C (2000) High-frequency linkage of co-expressing T-DNA in transgenic *Arabidopsis thaliana* transformed by vacuum-infiltration of *Agrobacterium tumefaciens*. *Theor Appl Genet* 100:487–493. <https://doi.org/10.1007/s001220050063>
- Sánchezgarcía A, Morenopérez AJ, Muropastor AM, Salas JJ, Garcés R, Martínezforce E (2010) Acyl-ACP thioesterases from castor (*Ricinus communis* L.): an enzymatic system appropriate for high rates of oil synthesis and accumulation. *Phytochemistry* 71:860–869. <https://doi.org/10.1016/j.phytochem.2010.03.015>
- Sandelius A, Aronsson H (2008) The chloroplast: interactions with the environment, vol 8. Chloroplast membrane lipid biosynthesis and transport. Springer, Berlin.
- Shin KS, Kim S, Lee SK (2016) Improvement of free fatty acid production using a mutant acyl-CoA thioesterase I with high specific activity in *Escherichia coli*. *Biotechnol Biofuels* 9:208. <https://doi.org/10.1186/s13068-016-0622-y>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Tan KWM, Yuan KL (2017) Expression of the heterologous *Dunaliella tertiolecta* fatty acyl-ACP thioesterase leads to increased lipid production in *Chlamydomonas reinhardtii*. *J Biotechnol* 247:60–67. <https://doi.org/10.1016/j.jbiotec.2017.03.004>
- Tan XL et al (2015) Cloning and functional characterization of a fatty acyl-acyl carrier protein thioesterase gene (BnFatB) in *Brassica napus* L. *J Agric Sci Technol* 17:987–997
- Trapnell C et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515. <https://doi.org/10.1038/nbt.1621>
- Voelker T (1996) Plant Acyl-ACP thioesterases: chain-length determining enzymes in plant fatty acid biosynthesis. *Genet Eng* 18:111–133. https://doi.org/10.1007/978-1-4899-1766-9_8
- 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. <https://doi.org/10.1104/pp.114.2.669>
- Wen Q, Lei Y, Huang J, Jiang H, Liao B (2012) Molecular cloning and characterization of an acyl-ACP thioesterase gene (AhFatB1) from allotetraploid peanut (*Arachis hypogaea* L.). *Afr J Biotechnol*. <https://doi.org/10.5897/AJB12.778>
- Xiao K et al (2018) Metabolic engineering for enhanced medium chain omega hydroxy fatty acid production in *Escherichia coli*. *Front Microbiol* 9:139. <https://doi.org/10.3389/fmicb.2018.00139>
- Yoo SD, Cho YH, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* 2:1565–1572. <https://doi.org/10.1038/nprot.2007.199>
- Zhuang W et al (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet* 51:865–876. <https://doi.org/10.1038/s41588-019-0402-2>
- Zheng L et al (2017) Discovery of a new mechanism for regulation of plant triacylglycerol metabolism: The peanut diacylglycerol acyltransferase-1 gene family transcriptome is highly enriched in alternative splicing variants. *J Plant Physiol* 219:62–70. <https://doi.org/10.1016/j.jplph.2017.09.009>
- Zinoviev S et al (2010) Next-generation biofuels: Survey of emerging technologies and sustainability issues. *Chem Sus Chem* 3:1106–1133. <https://doi.org/10.1002/cssc.201000052>

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