Identification and characterization of a gene encoding a putative lysophosphatidyl acyltransferase from *Arachis hypogaea*

Si-Long Chen[1,2,3,†], Jia-Quan Huang[1,†], Yong Lei[1], Yue-Ting Zhang[1], Xiao-Ping Ren[1], Yu-Ning Chen[1], Hui-Fang Jiang[1], Li-Ying Yan[1], Yu-Rong Li[2] and Bo-Shou Liao[1,*]

[1]*Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, China*

[2]*Institute of Food and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Laboratory of Crop Genetics and Breeding of Hebei Province, Shijiazhuang 050031, China*

[3]*Graduate School of Chinese Academy of Agriculture Sciences, Beijing 100081, China*

**Corresponding author (Email, lboshou@hotmail.com)*
[†]These authors contributed equally to this work.

Lysophosphatidyl acyltransferase (LPAT) is the important enzyme responsible for the acylation of lysophosphatidic acid (LPA), leading to the generation of phosphatidic acid (PA) in plant. Its encoding gene is an essential candidate for oil crops to improve oil composition and increase seed oil content through genetic engineering. In this study, a full-length *AhLPAT4* gene was isolated via cDNA library screening and rapid amplification of cDNA ends (RACE); our data demonstrated that *AhLPAT4* had 1631 nucleotides, encoding a putative 43.8 kDa protein with 383 amino acid residues. The deduced protein included a conserved acyltransferase domain and four motifs (I–IV) with putative LPA and acyl-CoA catalytic and binding sites. Bioinformatic analysis indicated that AhLPAT4 contained four transmembrane domains (TMDs), localized to the endoplasmic reticulum (ER) membrane; detailed analysis indicated that motif I and motifs II–III in AhLPAT4 were separated by the third TMD, which located on cytosolic and ER luminal side respectively, and hydrophobic residues on the surface of AhLPAT4 protein fold to form a hydrophobic tunnel to accommodate the acyl chain. Subcellular localization analysis confirmed that AhLPAT4 was a cytoplasm protein. Phylogenetic analysis revealed that AhLPAT4 had a high homology (63.7–78.3%) with putative LPAT4 proteins from *Glycine max*, *Arabidopsis thaliana* and *Ricinus communis*. *AhLPAT4* was ubiquitously expressed in diverse tissues except in flower, which is almost undetectable. The expression analysis in different developmental stages in peanut seeds indicated that *AhLPAT4* did not coincide with oil accumulation.

[Chen S-L, Huang J-Q, Lei Y, Zhang Y-T, Ren X-P, Chen Y-N, Jiang H-F, Yan L-Y, Li Y-R and Liao B-S 2012 Identification and characterization of a gene encoding a putative lysophosphatidyl acyltransferase from *Arachis hypogaea*. *J. Biosci.* **37** 1029–1039] **DOI** 10.1007/s12038-012-9277-4

1. Introduction

Triacylglycerols (TAGs) are a group of nonpolar lipids representing the major components of vegetable oils. It is an efficient and important form of storage energy and carbon resource in seeds (Ohlrogge and Browse 1995). In higher plants, fatty acids, the precursors of TAGs, are predominantly synthesized *de novo* from acetyl-coenzyme A (CoA) in plastid. They

Keywords. *Arachis hypogaea*; expression profile; gene cloning; lysophosphatidyl acyltransferase; structure and function

Abbreviations used: CTAB, hexadecyltrimethylammonium bromide; DAF, days after flowering; ER, endoplasmic reticulum; G3P, *sn*-glycerol-3-phosphate; GFP, green fluorescent protein; GPAT, glycerol-3-phosphate acyltransferase; LPA, lysophosphatidic acid; LPAT, lysophosphatidyl acyltransferase; ORF, open reading frame; PA, phosphatidic acid; RACE, rapid amplification of cDNA ends; TAG, triacylglycerol

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/dec2012/supp/Chen.pdf*

Published online: 30 October 2012

are then transported to endoplasmic reticulum (ER) or cytoplasm to form acyl-CoA (Ohlrogge 1994). TAG is subsequently synthesized by three sequential acyl-CoA-dependent acylations of the glycerol backbone beginning with *sn*-glycerol-3-phosphate (G3P); it is known as the classical Kennedy pathway which occurs primarily in the ER (Ohlrogge *et al.* 1991). In the second step of TAG bioassembly, lysophosphatidic acid (LPA) is acylated to generate phosphatidic acid (PA) catalysed by lysophosphatidyl acyltransferase (LPAT, EC 2.3.1.51). PA can be either converted to phospholipids that is directly integrated in diverse cellular membranes, or converted to TAGs for storage in seeds (Weselake *et al.* 2009). LPAT plays a critical role in the regulation of many physiological processes (Yu *et al.* 2004). It also determines TAGs acyl composition at the *sn*-2 position because of its high substrate specificity (Baud and Lepiniec 2010). Therefore, LPAT is a candidate gene for genetic engineering to change fatty acids composition.

Metabolic control analysis suggested that reactions in TAG assembly can provide increased sink strength that boost fatty acid production and stimulate increased carbon flux for TAG formation (Weselake *et al.* 2008). Recent studies have also shown that the conversion of LPA to PA catalysed by LPAT is a potential rate-limiting step in TAG biosynthesis in plants (Maisonneuve *et al.* 2010). Apparently, the increase of LPAT activity can result in the enhancement of storage lipid sink size by feedforward effect. And many studies have demonstrated that *LPAT* overexpression in plants led to a significant increase in seed oil content. For instance, when a yeast (*Saccharomyces cerevisiae*) long-chain *sn*-2 acyltransferase gene *SLC1-1*, a variant *LPAT*, was overexpressed in rapeseed (*Brassica napus*) and *Arabidopsis* (*Arabidopsis thaliana*), the proportions and content of very-long-chain fatty acyl at the *sn*-2 position of seed TAG increased, and seed oil content elevated from 8% to 48% (Zou *et al.* 1997). When microsomal *LPAT* genes (*BAT1.13* and *BAT1.5* which encodes rapeseed microsomal LPATs) from rapeseed were overexpressed in *Arabidopsis* in a seed-specific manner: the average seed weight and the total fatty acid content of seed storage lipids increased 6% and 13% respectively compared with nontransformed plants (Maisonneuve *et al.* 2010). *LPAT* is thus a potential gene in the oil content improvement.

LPAT is a gene family with several members; in *Arabidopsis*, at least nine distinct genes encode LPAT-like proteins (Roscoe 2005). LPAT is associated with a variety of membrane systems, including chloroplasts, ER and the outer membrane of mitochondria, suggesting that several different isoforms were present in plants (Yu *et al.* 2004). Previous analysis indicated that LPAT1 from *Arabidopsis* is a plastidic isoform whose transcript is present at a substantially higher level in leaves than in other organs (Kim and Huang 2004), whereas LPAT2 is a typical ER-localized LPAT ubiquitously expressed in various organs (Kim *et al.* 2005). It is thus important to isolated different isomers in order to know their biological function.

Peanut (*Arachis hypogaea*) is one of the most important oilseed crops and a good source of edible oil; which and how different genes regulate fatty acid composition and TAG formation has aroused much interest. Recently, based on EST sequencing and homologous analysis, several genes in fatty acid biosynthesis and TAG formation has been isolated and characterized, such as *FAS* (Li *et al.* 2009), *FAD* (Chi *et al.* 2011) and *DGAT* (Saha *et al.* 2006). However, a large number of genes involved in peanut TAG biosynthesis is still unknown, and no *LPAT* genes were isolated and characterized in peanut except one *LPAT* cDNA belonging to *Arabidopsis LPAT2* orthologue (Chen *et al.* 2012). In this study, we isolated an *AhLPAT4* gene from peanut, sequence and putative structure-function relationships were analysed, and the expression profile of *AhLPAT4* gene was investigated in different organs as well as in seeds at different developmental stages. Our results will facilitate to understand the biochemical role of *AhLPAT4* gene in peanut.

2. Materials and methods

2.1 *Plant material and treatments*

Peanut cultivar CR-1 with high oil content, grown in National Wild *Arachis* Nursery Garden (Wuchang, China), was used to construct a seed full-length cDNA library, to isolate genomic DNA and to amplify *LPAT* gene. Five organs including root, stem, leaf, flower, gynophore and developing seeds (15, 20, 25, 30, 40, 50, 60, 70, 80 and 90 days after flowering, or DAF) were used for gene expression analysis. Each experiment repeated three times in which at least two plants were bulked. All samples were frozen in liquid nitrogen immediately and stored at −80°C before DNA and RNA extraction. Immature seeds from different seed developmental stages were air-dried for seed oil content analysis.

2.2 *Seed oil content measurement*

Peanut seed oil content was determined using NMR analysis with a bench-top Minispec mq-20 NMR analyser (Bruker Optik GmbH, Germany). The standard curve was produced using pure peanut oil. Oil content of 15 selected samples using this method was evaluated by standard Soxhlet extraction method as described previously (Tan *et al.* 2011), and these two methods matched well with each other (r^2=0.978, p<0.01).

2.3 *Genomic DNA and total RNA extraction*

Genomic DNA was extracted from young leaves of peanut seedlings using the hexadecyltrimethylammonium bromide (CTAB) method as described previously (Allen *et al.* 2006). Total RNA was isolated from different organs (root, stem, leaf, flower and gynophore) with Trizol Reagent (Invitrogen, San Diego, CA) according to the manufacturer's instructions. Total RNA from immature seeds at different developmental stages was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Total RNA was treated with RNase-free DNase A (Fermentas, USA).

2.4 *Cloning of full-length AhLPAT4 cDNA and genomic DNA*

Previously, we have constructed a cDNA library using developing seed of peanut and sequenced more than 20000 EST sequences. The partial sequence of *AhLPAT4* was obtained via cDNA library screening. In order to amplify full-length *AhLPAT4* gene, the 5′ RACE method was performed to obtain the 5′-terminal regions of *AhLPAT4* cDNA using 5′ RACE kit (TaKaRa, Dalian, China). A full-length cDNA containing an intact open reading frame (ORF) was identified using ORF Finder (Wheeler *et al.* 2003) and the prediction result was verified by BLASTX at NCBI database (*http://ncbi.nlm.ncih.gov/*). In order to obtain the full-length genomic *AhLPAT4* gene sequence, the exon–intron junction sites were firstly identified by comparing peanut *AhLPAT4* full-length cDNA sequence with *Arabidopsis AtLPAT4* (AT1G75020) homologous sequence. Three gene-specific primer pairs were designed based on the known cDNA sequence and intron regions were amplified by PCR using PrimeSTAR HS DNA Polymerase (TaKaRa). Three fragments containing more than 120 bp overlapping sequences with each other were obtained. Finally, full-length peanut *AhLPAT4* genomic DNA sequence was assembled and obtained. All primers used in this study are listed in supplementary table 1.

2.5 *Bioinformatic analysis*

Genomic DNA and cDNA alignments as well as splice signal identification were performed using NCBI Splign (*http://www.ncbi.nlm.nih.gov/sutils/splign/*). Conserved domains analysis was carried out using Conserved Domain Databases (NCBI CDD). Active sites in AhLPAT4 were predicted by PROSITE in Predictprotein (*http://www.predictprotein.org/*). Hydrophobicity profile, charge density distribution and secondary structure of AhLPAT4 were predicted by the Protean 7.1 software package (DNASTAR, USA). The putative subcellular localization was estimated by subcell4 (*http://chemdata.shu.edu.cn/subcell/subcell4.jsp*) and MultiLoc2-HighRes (*http://abi.inf.uni-tuebingen.de/Services/MultiLoc2*). AhLPAT4 tertiary structural model was constructed using the Phyre Version 0.2 (Kelley and Sternberg 2009) and visualized using PyMOL software Version1.3 (Schrödinger 2010). Alignments of LPAT protein sequences were achieved using ClustalW1.83 (Larkin *et al.* 2007) with default settings. The alignment output was used to generate a cladogram based on the neighbour-joining method (Saitou and Nei 1987), as implemented in the MEGA 5.0 (Tamura *et al.* 2011). Sequences were shaded using GeneDoc (Nicholas *et al.* 1997). The transmembrane domains (TMDs) of AhLPAT4 were predicted by TMPred (Hofmann and Stoffel 1993) and TopPred (von Heijne 1992).

2.6 *Protein subcellular localization analysis*

The *AhLPAT4* ORF without the stop codon was inserted into the *Kpn*I–*Xma*I sites of pEGFP vector, generating a construct with GFP at the C-terminal of *AhLPAT4* under the control of the 35S promoter from cauliflower mosaic virus (CaMV). This construct was transferred into onion epidermal cells by microprojectile bombardment following the procedures described by Singsit *et al.* (1997). Cells harbouring the empty pEGFP vector were used as a control. The transformed onion epidermal cells were incubated in the dark at 25°C. After 24 h, the cell layers were soaked in 4′, 6-diamidine-2′-phenylindole dihydrochloride (DAPI; Roche, Germany) methanol solution (1 μg/ml) for visualization of nuclei. The GFP signals were observed using a Nikon's A1 confocal laser microscope system (Nikon, Japan).

2.7 *Gene expression analysis*

Two micrograms of total RNA was reverse-transcribed using First Strand cDNA Synthesis Kit ReverTra Ace –α– (TOYOBO, Osaka, Japan) with oligo $(dT)_{20}$ primer. Real-time quantitative PCR (qRT-PCR) was carried out in a reaction system containing 10 μL 2× SYBR Green Realtime PCR Master Mix (TOYOBO), 0.4 μL (10 μM) of each primer (supplementary table 1), 5 μL of an 1: 20 dilution of the synthesized cDNA and 4.5 μL ddH_2O. PCR amplification was conducted in an iQ™ 5 Multicolor Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). The PCR condition was as follows: pre-denaturation at 95°C for 5 min, followed by 40 cycles of 95°C for 15 s, 60°C for 20 s, 72°C for 20 s and a final extension at 72°C for 5 min. Each PCR reaction was performed in triplicate. Peanut *Ubiquitin* gene (Luo *et al.* 2005) was used as the internal control. Expression data of the target gene was normalized with internal control using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

3. Results

3.1 *Cloning and nucleotide sequence analysis of AhLPAT4*

A truncated segment of *AhPLAT4* was obtained through EST sequence screening. The 5′-end *AhLPAT4* was cloned by 5′ RACE. Sequencing analysis indicted that the length of the 5′ RACE PCR product was 907 bp including 137 bp overlapped with the previous cDNA sequence. The assembled sequences contained 1631 bp nucleotides with 1152 bp ORF. The 5′-untranslated region (UTR) included 383 nucleotides with one TATA box. An ATG initiation codon was found 383 nucleotides downstream of the 5′ start site, and there was an in frame stop codon located 99 nucleotides upstream the end of 3′-UTR of *AhLPAT4*. In addition, the sequence of the predicted ORF start site was quite fit the Kozak rule. Although the canonical polyadenylation signal (AATAAA) was absent, a 19 nucleotide poly (A) tail was found in 3′-UTR (figure 1).

The genomic DNA sequence of this gene was 3274 bp with a putative 1648 bp intron sequences. Two introns were present in the coding region, and there was a 1035 bp intron in the 5′-UTR. The organization of introns and exons of the *AhLPAT4* gene was identical to the reported *LPAT4* gene from other species such as *Arabidopsis*, soybean (*Glycine max*) and cast bean (*Ricinus communis*) (data not shown). All the exon–intron junction sequences had conserved GT–AG dinucleotide and a high degree of sequence conservation (supplementary figure 1).

3.2 *Conserved regions in the deduced protein sequence of AhLPAT4*

The ORF of *AhLPAT4* gene was deduced to encode a protein of 383 amino acid residues with a calculated molecular mass of 43.8 kDa and isoelectric point of 9.23. Conserved domain analysis detected a conserved 1-acylglycerol-3-phosphate acyltransferase-related (PTHR10983) or lysocardiolipin acyltransferase-related (PTHR10983: SF10) domain and a conserved glycerol acyltransferase domain (PF01553). Another glycerol-3-phosphate (1)-acyltransferase domain (SSF69593) as well as the putative acyl-acceptor binding pocket was also predicted using NCBI CDD (figure 1).

Multi-alignment of AhLPAT4 with other 17 LPATs from 11 plant species showed that LPAT2–5 shared 22–78% sequence identity and 39–89% sequence similarity. And the predicted functional domains among plant LPATs were also conserved (figure 2). There were four conserved key motifs I–IV, i.e. NH(X4) D (residue position from 105 to 111, motif I), LPVIGW (from 139 to 144, motif II), WLAFFPEGTD (from 180 to 189, motif III), and NVLLPKTKGF (from 212 to 221, motif IV): they were conserved among all LPATs in various plant species. Specifically, motifs I and III was highly conserved with the invariant residues, such as H106, D111, E186, G187 and T188. In addition, the potential biologically functional sites, such as a putative N-glycosylation site, six protein kinase C phosphorylation sites, three casein kinase II phosphorylation sites, two tyrosine kinase phosphorylation sites, five N-myristoylation sites as well as one leucine zipper motif, were also found in AhLPAT4.

3.3 *The predicted structure of AhLPAT4 protein*

Hydrophobicity prediction using Kyte-Doolittle algorithm revealed that the total number of hydrophobic residues (216 amino acids) were more than those of hydrophilic residues (166 amino acids), suggesting that AhLPAT4 was a hydrophobicity protein. Two strongly hydrophobic stretches with about 40 amino acids were distributed in N- and C-terminal regions separately (supplementary figure 2A). Secondary structure prediction demonstrated that putative AhLPAT4 protein contained 45% of α-helix (amount 13), 12% of β-sheet (amount 9) and 43% of random coil (supplementary figure 2C). A dilysine motif sequence KXKXX was present in the C-terminal to retrieve the type I transmembrane proteins to ER as previously suggested (Jackson *et al.* 1993). In addition, Subcell4 and MultiLoc2 predicted that this protein located on the ER membrane. Further prediction of AhLPAT4 transmembrane topology using TMPred and TopPred revealed that AhLPAT4 contained four TMDs, and its N- and C-termini were cytoplasmically oriented. Unexpectedly, motif I was separated from motif III by the third TMD and the partial or entire motif II was buried in the third TMD (supplementary figure 3), suggesting they might be located on opposite side of the ER membrane.

Phyre homology modelling of AhLPAT4 showed a similar folding mode and spatial configuration to CmGPAT (PDB code 1IUQ) (Tamada *et al.* 2004), a stereo molecular model from G3P acyltransferase (GPAT) of squash (*Cucurbita maschata*). The stereo diagrams indicated that the secondary structure elements of AhLPAT4 were organized into a compact domain consisting seven α-helices and four parallel β-sheets with the former distributed in the N-terminus (figure 3A). The hydrophobic residues also fold to form a hydrophobic tunnel to accommodate the acyl chain (figure 3B). And the essential residues, such as H106, D111, E186, G187 and T188, lay close to each other and formed a small cluster on the molecular surface, which was consistent with the conserved residues in the G3P binding sites of squash GPAT.

3.4 *Phylogenetic analysis of the LPAT*

Phylogenetic analysis using 18 LPATs from peanut and other species was demonstrated (figure 4). The results indicated that all LPATs could be classified into three clusters. As

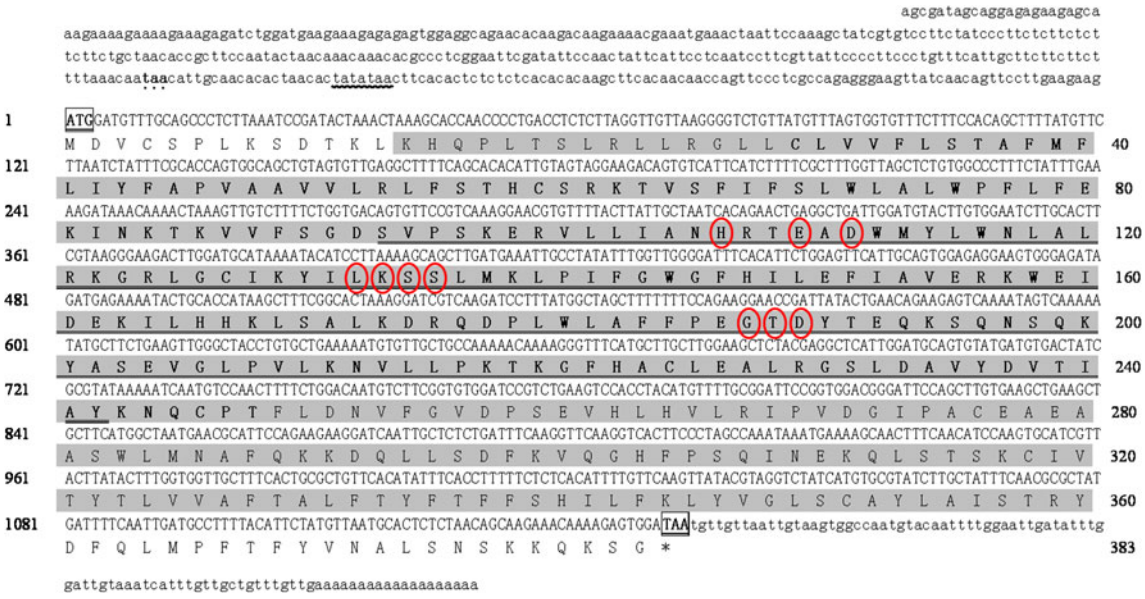

Figure 1. Nucleotide and deduced amino acid sequence of *AhLPAT4* cDNA from peanut. The untranslated terminal regions are in lowercase, whereas the open reading frame is in uppercase letters. The start codon (ATG) and stop codon (TAA) are marked by boxes. The bold letters (taa) with under dots present an in frame stop codon. The underlined nucleotides (tatataa) indicate the TATA box. The residues in grey background indicate conserved 1-acylglycerol-3-phosphate acyltransferase-related domain. The bold letters in grey background indicate a glycerol-3-phosphate (1)-acyltransferase domain. The letters with underline represent a glycerol acyltransferase domain. The putative acyl-acceptor binding pocket sites are marked by red circles.

expected, AhLPAT4 protein appeared to be more closely related to GmLPAT4 (from soybean) than other LPAT4 isoforms, suggesting peanut and soybean LPAT4 had a common origin. AhLPAT4 shared 78.3%, 63.7% and 63.7% amino acid sequence identities (88.5%, 80.4% and 77.5% similarities) with those from soybean, castor bean and *Arabidopsis* respectively. Both AhLPAT4 and GmLPAT4 were evolutionarily related to *Arabidopsis* LPAT4. The branch distribution of the phylogenetic tree also indicated that the evolutionary split between monocots and dicots occurred prior to the divergence of *LPAT4* and *LPAT5* genes.

3.5 *Subcellular localization of AhLPAT4 protein*

In order to determine the subcellular localization of the AhLPAT4 protein, fusion protein expression vector pEGFP-*AhLPAT4*-GFP (figure 5) was transformed into onion epidermal cells by microprojectile bombardment using empty vector pEGFP as control. The results showed that green fluorescent signal from AhLPAT4–GFP fusion protein was detected exclusively around the cell in cytoplasm. Unlike the fusion protein, the fluorescent signal from cell expression control pEGFP vector demonstrated the expected GFP distribution in both of cytoplasm and nucleus (figure 5), and these results indicated that AhLPAT4 was a cytoplasm protein, which was consistent with the subcellular localization prediction result.

3.6 *Expression pattern of the AhLPAT4 gene*

The expression of *AhLPAT4* in different peanut organs was determined using qRT-PCR. The results showed that *AhLPAT4* was constitutively expressed in peanut, with higher expression in leaf, stem and root, lower expression level in seed and gynophore, and it was almost undetectable in the flower (figure 6A). *LPAT4* expression has not been reported in *Arachis* species in previous studies. But its homologue expression has been observed in *Arabidopsis*: similar to our results, its expression was not restricted to typical oil storage organs like seeds, and was expressed ubiquitously with low abundance (Kim *et al.* 2005).

The expression of *AhLPAT4* gene was checked in seeds in different developmental stages (figure 6B). The results demonstrated that *AhLPAT4* expression had a biphasic pattern. Moderate expression was observed at 15 and 20 DAF, followed by a decrease at 25 and 30 DAF. It was then increased at 40 and 50 DAF, coinciding with the stage with the highest rate of oil accumulation in peanut seed. At 60 DAF, its expression declined, and then increased from 70 to 80 DAF, with the highest expression at 80 DAF, and at that time, the rate of seed oil increased slightly. After that, a gradual reduction occurred.

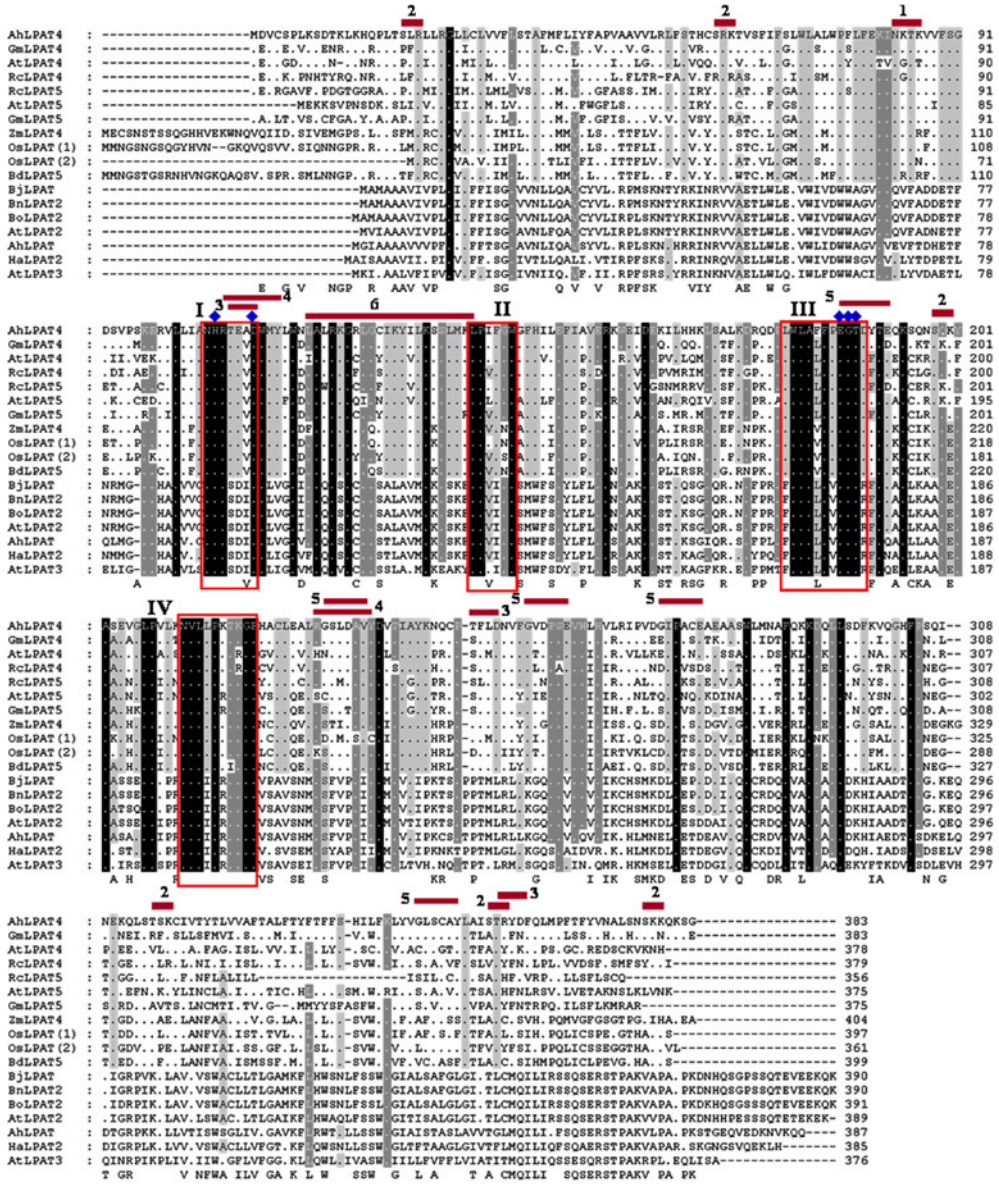

Figure 2. Multiple sequence alignment of LPAT homologues. Two shading levels are set: black for 100% identical amino acids and grey for 80% similarity groups. The catalytically important sites in acyltransferase family are indicated in box I to IV. Sites 1 to 6 represent amino-glycosylation site, protein kinase C phophorylation site, casein kinase II phosphorylation site, tyrosine kinase phosphorylation site, N-myristoylation site and leucine zipper pattern, respectively. GenBank accession number are as follows: ZmLPAT4, NP_001151948; RcLPAT4, ACC59199; BjLPAT, ABM92334; RcLPAT5, XP_002519095; HaLPAT2, ABP93351; GmLPAT4, XP_003544428; GmLPAT5, XP_003541079; OsLPAT(1), AAU90234; OsLPAT(2), AD53283; BnLPAT2, Q9XFW4; BoLPAT2, Q6IWY1; AtLPAT4, Q8L4Y2; AtLPAT5, Q9LHN4; AtLPAT2, Q8LG50; AtLPAT3, Q9SYC8; BdLPAT5, XP_003568161. Abbreviations: Ah, *Arachis hypogaea*; Zm, *Zea mays*; Rc, *Ricinus communis*; Bj, *Brassica juncea*; Ha, *Helianthus annuus*; Gm, *Glycine max*; Os, *Oryza sativa*; Bn, *Brassica napus*; Bo, *Brassica oleracea*; At, *Arabidopsis thaliana*; Bd, *Brachypodium distachyon*.

The expression of *AhLPAT4* gene was not in complete agreement with seed oil accumulation rate, especially in the earlier stages of seed development like the period from 15 to 30 DAF. However, these results indicated that *AhLPAT4* is an important component in lipid biosynthesis process.

4. Discussion

LPATs enzymes belong to a large acyltransferase family contained at least two classes of LPATs according to the different subcellular localization, i.e. plastid LPATs and

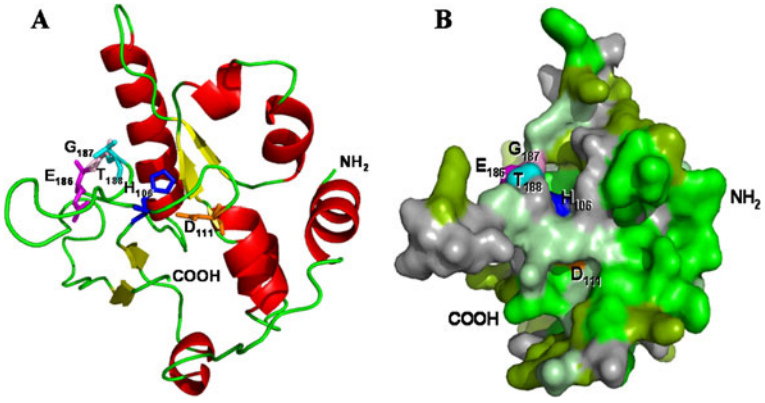

Figure 3. Tertiary structure characteristics of partial AhLPAT4 protein. (**A**) Cartoon depiction of AhLPAT4 three-dimensional structure. α-helices are indicated as red ribbons. β-sheets are indicated as yellow arrows. Random coils are indicated as green lines. The key conserved residues are schematically drawn as stick models at the corresponding locations of the protein sequences. (**B**) Space fill model of AhLPAT4 three-dimensional structure. The hydrophobicity of surface residues are indicated in different green colors ranging from dark green to light green as the hydrophobicity varies from strong to weak. The positions of the key conserved residues H106, D111, E186, G187 and T188 are represented in blue, orange, magentas, cygan and pink, respectively.

cytoplasmic LPATs. In cytoplasm, the synthesized PA can be converted to TAG for storage in oilseeds. There are several genes encoding cytoplasmic LPATs. For instance, four genes were speculated to encode putative cytoplasmic LPATs (LPAT2–5) in *Arabidopsis* (Kim and Huang 2004). In this study, we cloned and identified one *LPAT* gene encoding a LPAT-like protein from peanut, *AhLPAT4*. The full-length cDNA of *AhLPAT4* was 1631 bp containing 1152 bp ORF, which encoded a putative protein of 383 residues. Alignment of the deduced polypeptide sequences of cytoplasmic LPAT proteins demonstrated that AhLPAT4 was similar to cytoplasmic LPATs from other plant species. *AhLPAT4* was similar to identified *Arabidopsis* cytoplasmic *LPAT4* in introns location (Kim *et al.* 2005), which differed remarkably from that of *Arabidopsis* plastid *LPAT1*. Moreover, the AhLPAT4 protein lacked the N-terminal intracellular

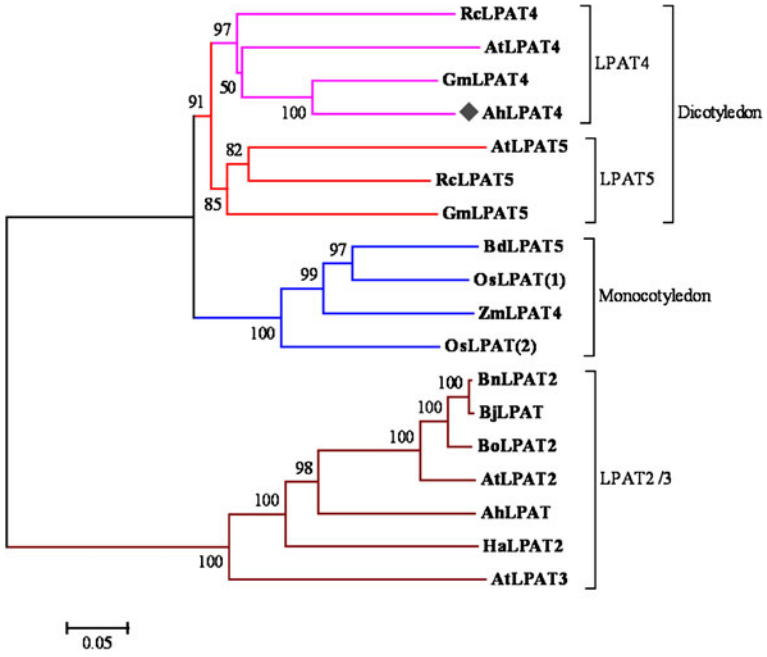

Figure 4. Phylogenetic analysis of the selected LPAT members. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown at each node. The scale bar represents the number of substitutions per amino acid site. Species and enzyme names are abbreviated as listed in figure 2.

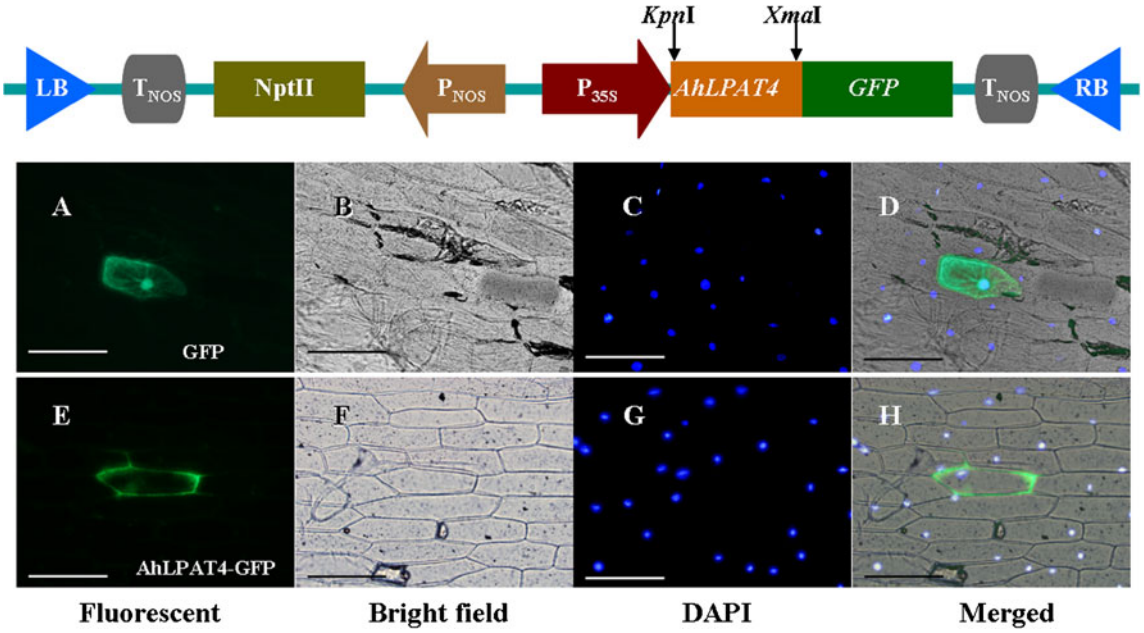

Figure 5. Subcellular localization of *AhLPAT4–GFP* fusion protein in onion epidermal cells. The top panels show the expression vector fusing the *AhLPAT4* and the *GFP* reporter gene. *AhLPAT4* tagged with *GFP* in the C-terminus was inserted into pEGFP vector between both restriction sites of *Kpn*I and *Xma*I. LB, left border; RB, right border; *PNOS* and *TNOS* are promoter and polyadenylation signal of the nopaline synthase gene, respectively; NptII neomycin phosphotransferase II, 35S promoter of Cauliflower mosaic virus. Bottom panels show expression of GFP and AhLPAT4-GFP. (**A**) *GFP* fluorescence of onion epidermal cells expressing *GFP*. (**B**) Bright field image of (**A**). **(C)** Cell nuclei counterstained with DAPI. **(D)** Merged image of **(A)**, **(B)** and **(C)**. (**E**) *GFP* fluorescence of onion epidermal cells expressing *AhLPAT4–GFP*. (**F**) Bright field image of (**E**). **(G)** Cell nuclei counterstained with DAPI. **(H)** Merged image of **(E), (F)** and **(G).** Scale bars=50 μm.

targeting signal motifs, but it possessed a putative dilysine-like ER retention signal ($KQKSG_{383}$) located at C-terminus; these features were similar to the microsomal LPATs from rapeseed (Maisonneuve *et al.* 2010) and the cytoplasmic LPAT4 from *Arabidopsis* (Kim *et al.* 2005). Phylogenetic analysis further indicated that there was close evolutionary relationship between AhLPAT4 and LPAT4 members from other plant species. The sequence of the putative AhLPAT4 protein was highly conserved for at least 11 different plant species. AhLPAT4 shared 88.5%, 80.4% and 77.5% amino acid sequence similarity to that of soybean, castor bean and *Arabidopsis* respectively, suggesting that this gene might play a crucial role in plant lipid metabolism. However, AhLPAT4 only shared 41.9% amino acid sequence similarity with the AhLPAT which was also derived from peanut (Chen *et al.* 2012), suggesting that these two may be analogous genes. Multiple gene copies existing in peanut genome could be because of the allotetraploid nature of cultivated peanut, or because the acyltransferase gene families have undergone frequent duplication as has been suggested previously (Cagliari *et al.* 2010). Multiple genes with similar functions made the glycerolipid synthesis process more sophisticated.

Analysis of the encoded polypeptide sequence of AhLPAT4 revealed that there were four conserved motifs (I to IV) important for acyltransferase activity. Notably, NH (X4)D motif and EGT motif were conserved in diverse organisms, such as plant, bacteria, yeast, animal and human (Eberhardt *et al.* 1997; Lewin *et al.* 1999), which have been shown to be the catalytic site and substrate binding site respectively (Heath and Rock 1998). The existence of the same domains suggested that AhLPAT4 also had acyltransferase activity and might play a role in glycerolipid metabolism in plant cells. Additionally, AhLPAT4 contained an invariant His residue in the NH(X4)D consensus sequence, which was essential in the membrane-associated glycerolipid acyltransferases (Heath and Rock 1998; Yamashita *et al.* 2007). The highly conserved Asp residue that was assumed to be important for proper folding and membrane insertion of the acyltransferases (Heath and Rock 1998) was also present in AhLPAT4, suggesting that AhLPAT4 had an acyltransferase activity. Moreover, AhLPAT4 also contained the specific residues in protein, such as E186, G187 and T188 in motif III, and P216 in motif IV, which were known to play a critical role in enzyme activity.

It has been confirmed that all the acyltransferases in acylation pathways for TAGs formation were membrane bound (Saha *et al.* 2006). Typically, all LPATs derived from plants had at least two putative TMDs. But the structure of

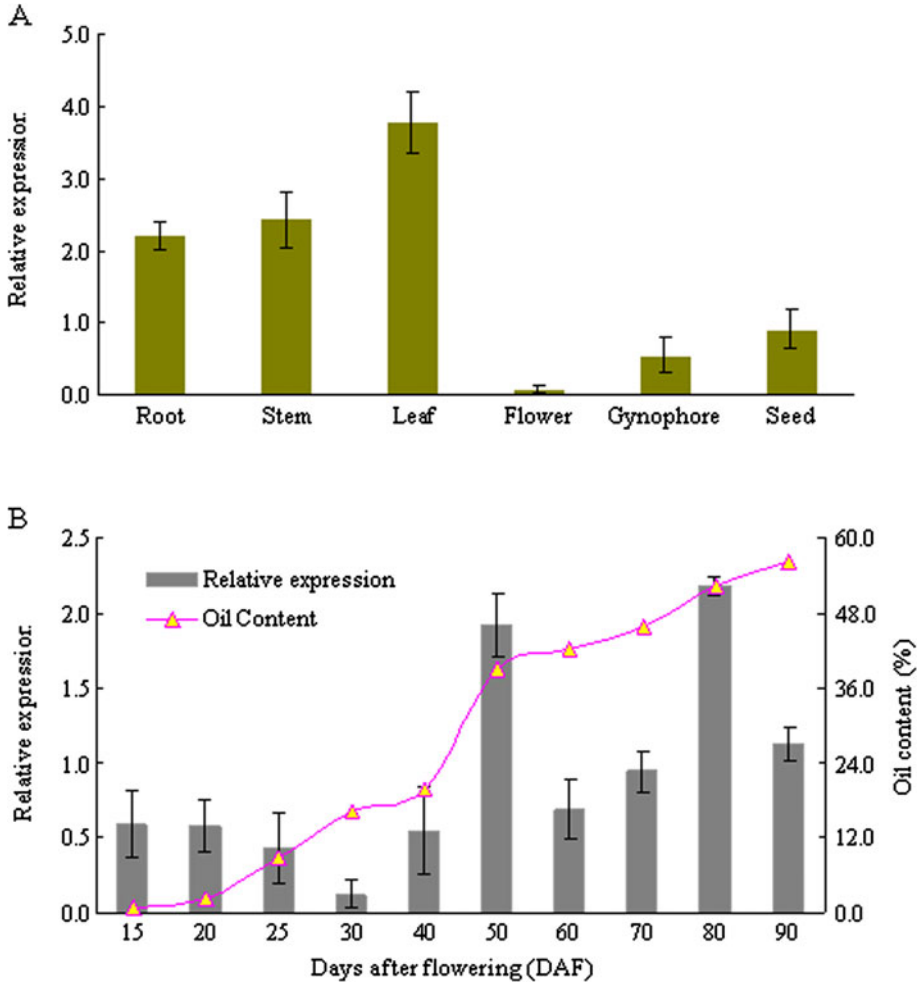

Figure 6. Expression profile of peanut *AhLPAT4* gene. **(A)** Tissue-specific expression of *AhLAPT4*. **(B)** *AhLPAT4* gene expression and seed oil accumulation during peanut seed development. Error bar represents the standard deviation of three replicates. The peanut *Ubiquitin* gene was used as an internal control for normalization.

LPATs in peanut has not been demonstrated before. In this study, using bioinformatic analysis, it is suggested that AhLPAT4 might be anchored to the ER membrane and may be an integral transmembrane protein, the NH(X4)D motif and EGT motif were separated by the third putative TMD (residues S134–R156), and these two motifs were predicted to locate on the cytosolic and luminal sides of the ER membrane, respectively. However, LPAT enzymatic catalysis domain requires the NH(X4)D and EGT motifs to be together forming a very tight spatial relationship. Interestingly, this predicted topology model seemed to be inconsistent with the fact that these motifs interact in converting LPA and acyl-CoA to PA. Similar findings have been reported for acyltransferases from other non-plant sources, such as GPAT1 from mouse (Gonzalez-Baro *et al.* 2001), LPATs from human (Leung 2001; Schmidt *et al.* 2010) and Slc1p from yeast (Pagac *et al.* 2011). The predicted tertiary structure of AhLPAT4 in our study demonstrated that the essential residues H106 and D111 in NH(X4)D motif were in the close proximity with the E186, G187 and T188 in EGT motif, which were consistent with the explanation that motifs I–IV and may penetrate into the membrane surface from the cytosolic or lumenal side to act in concert in close proximity (Yamashita *et al.* 2007).

From qRT-PCR analyses, it was shown that *AhLPAT4* was expressed constitutively although the expression was pretty low in flower. These findings were consistent with the previous studies in *Arabidopsis* and rapeseed, in which expression of *AtLPAT4* (Kim *et al.* 2005) and *BAT1.13* (Maisonneuve *et al.* 2010) were detected constitutively. Thus, we proposed that *AhLPAT4* might be involved in the production of PA for a constitutive requirement in most tissues. *AhLPAT4* gene in our study showed that gene expression was higher in vegetative organs (leaf, stem and root) than that in developing seed. In contrast, *AhLPAT* as reported before demonstrated a distinct tissue distribution, with higher expression in developing seed and lower

expression in vegetative organs (Chen *et al.* 2012). The expression of *AhLPAT4* was detected at all stages in developing seed (15 DAF to maturity). The accumulation of the *AhLPAT4* mRNA was found not to coincide with the seed oil increase rate among the stages measured in the present study. However, *AhLPAT4* mRNA was found to be obviously elevated at 50 DAF, corresponding to the drastical increase in oil content in peanut seed. Furthermore, *AhLPAT4* expression was also up-regulated at 80 DAF, when the rate of seed oil accumulation was slightly raised. These results suggested that *AhLPAT4* might play a role in glycerolipid accumulation occurring in developing seed. In contrast, *AhLPAT* had a distinct pattern of expression during the same seed developing stages in our previous studies (Chen *et al.* 2012). Whether the difference in preferences of fatty acid exists in *AhLPAT* and *AhLPAT4* needs further evidence.

In conclusion, bioinformatic analysis suggested that *AhLPAT4* encoded a LPAT enzyme isoform. The structure–function relationships and gene expression profiles in this study, provided a foundation for further exploration the roles of the ER-localized LPATs, which probably play an important role in peanut glycerolipid biosynthesis, TAG assembly and seed oil accumulation.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (31201239, 31071456), the Natural Science Foundation of Hebei Province, China (C2010001594), a grant from Modern Agro-industry Technology Research System (nycytx–19) and the Key Basic Research Program of Hebei Province Applied Basic Research Plan (10960122D). We gratefully thank Professor Shengyi Liu (OCRI, CAAS) and his laboratory staff for provision of plasmid pEGFP and technical assistance.

References

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S and Thompson WF 2006 A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **1** 2320–2325

Baud S and Lepiniec L 2010 Physiological and developmental regulation of seed oil production. *Prog. Lipid. Res.* **49** 235–249

Cagliari A, Margis-Pinheiro M, Loss G, Mastroberti AA, de Araujo Mariath JE and Margis R 2010 Identification and expression analysis of castor bean (*Ricinus communis*) genes encoding enzymes from the triacylglycerol biosynthesis pathway. *Plant Sci.* **179** 499–509

Chen SL, Huang JQ, Lei Y, Ren XP, Wen QG, Chen YN, Jiang HF, Yan LY, *et al.* 2012 Cloning and expression analysis of lysophosphatidic acid acyltransferase (LPAT) encoding gene in peanut. *Acta. Agron. Sin.* **38** 245–255 (in Chinese with English abstract)

Chi X, Yang Q, Pan L, Chen M, He Y, Yang Z and Yu S 2011 Isolation and characterization of fatty acid desaturase genes from peanut (*Arachis hypogaea* L.). *Plant Cell Rep.* **30** 1393–1404

Eberhardt C, Gray PW and Tjoelker LW 1997 Human lysophosphatidic acid acyltransferase. cDNA cloning, expression, and localization to chromosome 9q34.3. *J. Biol. Chem.* **272** 20299–20305

Gonzalez-Baro MR, Granger DA and Coleman RA 2001 Mitochondrial glycerol phosphate acyltransferase contains two transmembrane domains with the active site in the N-terminal domain facing the cytosol. *J. Biol. Chem.* **276** 43182–43188

Heath RJ and Rock CO 1998 A conserved histidine is essential for glycerolipid acyltransferase catalysis. *J. Bacteriol.* **180** 1425–1430

Hofmann K and Stoffel W 1993 TMBASE-a database of membrane spanning protein segments. *Biol. Chem. HoppeSeyler.* **374** 166

Jackson MR, Nilsson T and Peterson PA 1993 Retrieval of transmembrane proteins to the endoplasmic reticulum. *J. Cell Biol.* **121** 317–333

Kelley LA and Sternberg MJ 2009 Protein structure prediction on the web: a case study using the Phyre server. *Nat. Protoc.* **4** 363–371

Kim HU and Huang AH 2004 Plastid lysophosphatidyl acyltransferase is essential for embryo development in *Arabidopsis*. *Plant Physiol.* **134** 1206–1216

Kim HU, Li Y and Huang AH 2005 Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in *Arabidopsis*. *Plant Cell* **17** 1073–1089

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, *et al.* 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* **23** 2947–2948

Leung DW 2001 The structure and functions of human lysophosphatidic acid acyltransferases. *Front. Biosci.* **6** D944–953

Lewin TM, Wang P and Coleman RA 1999 Analysis of amino acid motifs diagnostic for the *sn*-glycerol-3-phosphate acyltransferase reaction. *Biochemistry* **38** 5764–5771

Li MJ, Li AQ, Xia H, Zhao CZ, Li CS, Wan SB, Bi YP and Wang XJ 2009 Cloning and sequence analysis of putative type II fatty acid synthase genes from *Arachis hypogaea* L. *J. Biosci.* **34** 227–238

Livak KJ and Schmittgen TD 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2 (−Delta Delta C(T)) Method. *Methods* **25** 402–408

Luo M, Dang P, Bausher MG, Holbrook CC, Lee RD, Lynch RE and Guo BZ 2005 Identification of transcripts involved in resistance responses to leaf spot disease caused by *Cercosporidium personatum* in Peanut (*Arachis hypogaea*). *Phytopathology* **95** 381–387

Maisonneuve S, Bessoule JJ, Lessire R, Delseny M and Roscoe TJ 2010 Expression of rapeseed microsomal lysophosphatidic acid acyltransferase isozymes enhances seed oil content in *Arabidopsis*. *Plant Physiol.* **152** 670–684

Nicholas KB, Nicholas Jr HB and Deerfield II DW 1997 GeneDoc: analysis and visualization of genetic variation. *EMBNET.news* **4** 1–4

Ohlrogge JB 1994 Design of new plant products: engineering of fatty acid metabolism. *Plant Physiol.* **104** 821–826

Ohlrogge JB and Browse J 1995 Lipid biosynthesis. *Plant Cell* **7** 957–970

Ohlrogge JB, Browse J and Somerville CR 1991 The genetics of plant lipids. *Biochim. Biophys. Acta* **1082** 1–26

Pagac M, de la Mora HV, Duperrex C, Roubaty C, Vionnet C and Conzelmann A 2011 Topology of 1-acyl-*sn*-glycerol-3-phosphate acyltransferases SLC1 and ALE1 and related membrane-bound *O*-acyltransferases (MBOATs) of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **286** 36438–36447

Roscoe TJ 2005 Identification of acyltransferases controlling triacylglycerol biosynthesis in oilseeds using a genomics-based approach. *Eur. J. Lipid. Sci. Technol.* **107** 256–262

Saha S, Enugutti B, Rajakumari S and Rajasekharan R 2006 Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiol.* **141** 1533–1543

Saitou N and Nei M 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4** 406–425

Schmidt JA, Yvone GM and Brown WJ 2010 Membrane topology of human AGPAT3 (LPAAT3). *Biochem. Biophys. Res. Commun.* **397** 661–667

zSchrödinger L 2010 The PyMOL molecular graphics system, Version 1.3r1. http://www.pymol.org.

Singsit C, Adang MJ, Lynch RE, Anderson WF, Wang A, Cardineau G and Ozias-Akins P 1997 Expression of a *Bacillus thuringiensis cryIA*(c) gene in transgenic peanut plants and its efficacy against lesser cornstalk borer. *Transgenic Res.* **6** 169–176

Tamada T, Feese MD, Ferri SR, Kato Y, Yajima R, Toguri T and Kuroki R 2004 Substrate recognition and selectivity of plant glycerol-3-phosphate acyltransferases (GPATs) from *Cucurbita moscata* and *Spinacea oleracea*. *Acta Crystallogr. D* **60** 13–21

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28** 2731–2739

Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, *et al.* 2011 Enhanced seed oil production in canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol.* **156** 1577–1588

von Heijne G 1992 Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225** 487–494

Weselake RJ, Shah S, Tang M, Quant PA, Snyder CL, Furukawa-Stoffer TL, Zhu W, Taylor DC, *et al.* 2008 Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. *J. Exp. Bot.* **59** 3543–3549

Weselake RJ, Taylor DC, Rahman MH, Shah S, Laroche A, McVetty PB and Harwood JL 2009 Increasing the flow of carbon into seed oil. *Biotechnol. Adv.* **27** 866–878

Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, *et al.* 2003 Database resources of the National Center for Biotechnology. *Nucleic. Acids. Res.* **31** 28–33

Yamashita A, Nakanishi H, Suzuki H, Kamata R, Tanaka K, Waku K and Sugiura T 2007 Topology of acyltransferase motifs and substrate specificity and accessibility in 1-acyl-*sn*-glycero-3-phosphate acyltransferase 1. *Biochim. Biophys. Acta* **1771** 1202–1215

Yu B, Wakao S, Fan J and Benning C 2004 Loss of plastidic lysophosphatidic acid acyltransferase causes embryo-lethality in *Arabidopsis*. *Plant Cell Physiol.* **45** 503–510

Zou J, Katavic V, Giblin EM, Barton DL, MacKenzie SL, Keller WA, Hu X and Taylor DC 1997 Modification of seed oil content and acyl composition in the brassicaceae by expression of a yeast *sn*-2 acyltransferase gene. *Plant Cell* **9** 909–923

MS received 14 February 2012; accepted 04 September 2012

Corresponding editor: INDRANIL DASGUPTA