Theor Appl Genet (2006) 112: 1086–1097
DOI 10.1007/s00122-006-0210-9

ORIGINAL PAPER

Hui-Wen Wang · Jin-Song Zhang · Jun-Yi Gai
Shou-Yi Chen

Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean

Received: 8 September 2005 / Accepted: 28 December 2005 / Published online: 24 January 2006

Abstract Diacylglycerol acyltransferase (DGAT), as an important enzyme in triacylglycerol synthesis, catalyzes the final acylation of the Kennedy pathway. In the present study, the *GmDGAT* gene was cloned from *Glycine max* by using *AtDGAT* as a query to search against the soybean EST database and the rapid amplification of cDNA ends (RACE) method. Allelic genes were also isolated from 13 soybean accessions and the divergence of the deduced amino acid sequences were compared. The comparison reveals that although GmDGAT is a highly conserved protein, several differences of insertion/deletion were identified in the N-terminal region of the GmDGATs from various soybean accessions. In the C-terminal regions, a single amino acid mutation specific to both *G. max* and *G. soja* was also found. The *GmDGAT* genomic sequences were further cloned and the number and size of exons in the DGAT genomic sequence were very similar among different plant species, whereas the introns were more diverged. These results may have significance in elucidating the genetic diversity of the GmDGAT among the soybean subgenus.

Communicated by F. J. Muehlbauer

H.-W. Wang · J.-S. Zhang (✉) · S.-Y. Chen
The National Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 100101 Beijing, China
E-mail: jszhang@genetics.ac.cn
Tel.: +86-10-64886859
Fax: +86-10-64873428
E-mail: sychen@genetics.ac.cn

H.-W. Wang
Graduate School of the Chinese Academy of Sciences, 100039 Beijing, China

J.-Y. Gai
National soybean Institute for Improvement, Nanjing University, 210095 Nanjing, China

Keywords Diacylglycerol acyltransferase gene · *Glycine max* · Diversity

Introduction

Soybean is one of the most important crops, accounting for 48% of the world market in oil crops (Singh and Hymowitz 1999). China is the place of origin for soybean (*Glycine max* L. Merr) and it has rich soybean germplasm. Our laboratory has been focusing on soybean genomic research for years and several genes related to agronomic traits have been cloned and characterized (Zhang et al. 1997; Liu et al. 2000; Wu et al. 2001; He et al. 2002, 2003; Tian et al. 2004a, b; Wang et al. 2005; Luo et al. 2005). Oil content is the major trait of soybean. Triacylglycerol (TAG) is the main component of soybean oil whose content ranged from 13 to 22% in the soybean cultivars (Singh et al. 1999). In plants, TAG biosynthesis mainly contributes to the accumulation of seed oil. TAG is an important agricultural commodity and acts as a major energy source for the growth of new seedling (Lu and Hills 2002). TAG synthesis is via the Kennedy pathway, a biochemical process of the acylation of the glycerol backbone (Kennedy 1961). Diacylglycerol acyltransferase (DGAT) is a membrane-bound enzyme that transfers an acyl group from acyl-coenzyme-A to the sn-3 position of 1,2-diacylglycerol in the final acylation step of the Kennedy pathway. It is the only step that is unique to TAG synthesis. DGAT-catalyzed esterification of TAG was proposed to be a rate-limiting step in the control of plant TAG synthesis (Settlage et al. 1998). A strong positive correlation was found between the rate of oil accumulation and DGAT activity in mature seeds of soybean (Settlage et al. 1998) and rape (Perry and Harwood 1993). In the *Arabidopsis thaliana* mutant line AS11, which has an insertion mutation in the DGAT, the DGAT activity and seed TAG content were reduced (Jako et al. 2001). Over-expression of the *AtDGAT* cDNA in wild-type *A. thaliana* enhanced oil deposition and average seed weight (Jako et al. 2001).

The first reported sequence of a *DGAT* gene was from mouse (AF078752) (Cases et al. 1998). Then homologous *DGAT* genes have been cloned from various plants, such as *A. thaliana* (AJ131831) (Hobbs et al. 1999), *Brassica napus* (AF164434) (Nykiforuk et al. 1999) and *Ricinus communis* (AY366496) (He et al. 2004). Several DGATs were purified from various plants and the acyl-CoA-dependent DGAT activity was tested (Hobbs et al. 1999; Triki et al. 2000; Hobbs and Hills 2000; Jako et al. 2001). The DGAT was proved to be a component of the endoplasmic reticulum (Cao and Huang 1986; Settlage et al. 1995; Lacey and Hill 1996). Transcript levels of the *DGAT* were also examined. The *DGAT* was highly expressed in maturing seeds (Kaup et al. 2002; Lu et al. 2003) and its expression was also found in flowers and germinating seeds (Zou et al. 1999). Investigation of the *DGAT* gene would allow us to ascertain the role of the DGAT in regulating the plant TAG synthesis.

Soybean is an important oil crop and its *DGAT* gene has not been identified. To understand the features of this gene and compare the phylogenetic relationship of the *DGAT* between wild type and cultivars of soybean, the full-length cDNA of *GmDGAT* was cloned from cultivated soybean accession using the rapid amplification of cDNA ends (RACE) method (Frohman 1993). The *DGAT* cDNA sequences were further amplified from different accessions of soybean. The phylogenetic relationships of these *DGAT* proteins and the exon/intron structure of the *DGAT* genomic sequences between plant species were analyzed.

Materials and methods

Plant materials and DNA and RNA extractions

Soybean (*Glycine max* L. Merr.) cultivar 8904 and 17 other accessions from the subgenera *Soja* and *Glycine* were used in this study (Table 1). The bulked soybean seeds, which were harvested in the experimental station last year, were grown in the experimental station from May to September in Beijing. Leaves, flowers and pods were harvested at the indicated stages and stored at −70°C for RNA isolation. RNA was extracted according to the method of Zhang et al. (1996). The tissues were ground to fine powder in liquid nitrogen. Isolation of total RNA was performed using guanidine thiocyanate, and then purified with phenol–chloroform extraction. Mature leaves were harvested and DNA extraction was performed as described previously (Chen et al. 1991). Genomic DNA was extracted with SDS extraction solution and then purified with phenol–chloroform extraction and ethanol precipitation.

Cloning of the full-length cDNA of *GmDGAT*

Five μg total RNA from pods of 8904 at 20 days after flowering (DAF) was reverse-transcribed into first-strand cDNAs with the cDNA synthesis kit (Promega, Madison, WI, USA) in a 20 μl reaction volume. For screening of the *GmDGAT* from soybean, the cDNA sequence of *AtDGAT* was used as a query for a Basic Local Alignment Search Tool (BLAST) search against soybean ESTs in the GenBank, and two putative GmDGAT ESTs were obtained. The PCR primers, DGAT P1 and DGAT P2 (Table 2, Fig. 1), were designed according to the two EST sequences obtained above. The total volume of the PCR reaction mixture was 25 μl, containing 1 μl cDNA, 0.5 μM of each primer, 1×PCR buffer, 0.4 mM dNTPs, and 1 unit of long and accurate (LA) DNA polymerase (Takara, Kyoto, Japan) which is a proofreading enzyme; hence the fidelity is significantly better than that of rTaq polymerase (Takara). The reaction was denatured at 94°C for 5 min, and then followed by 30 cycles of 1 min at 94°C, 1 min at 56°C and 2 min at 72°C, then 10 min at 72°C. PCR reaction was performed by using a GeneAmp PCR System 9600 (Perkin Elmer, Boston, MA, USA). The partial *GmDGAT* fragment was purified using DNA Purification Kit (Dingguo, Beijing, China), cloned into pMD18-T vector (Takara) and sequenced.

For cloning of the full-length of *GmDGAT* from *G. max*, a SMART^{TM} RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA) was used. RACE is a method widely used to isolate the cDNA of unknown 5′ or 3′ flanking sequences. One μg total RNA from pods of 8904 at 20 DAF was used to synthesize cDNA as described in the protocol. The RACE primers (Table 2, Fig. 1) were designed based on the sequence of the partial *DGAT* fragment described above. The PCR reaction was performed as the protocol.

Based on the full-length sequence of the *GmDGAT* gene obtained above, the full length ORF of the *DGAT* gene from other accessions was amplified from corresponding cDNAs, with the primers DGAT F-P1 and DGAT F-P2 (Table 2, Fig. 1). The PCR reactions were performed as described above and three individual clones were selected and sequenced. The putative *LcDGAT* and *OsDGAT* ESTs from *Lotus corniculatus* and *Oryza sativa*, respectively, were obtained by searching GenBank, and the complete sequence of *LcDGAT* and *OsDGAT* were assembled by ContigExpress, a component of Vector NTI Suite 6.0 (InforMax Inc., Carlsbad, CA, USA). The alignments of the DNA or amino acid sequences were performed with MegAlign of DNASTAR (DNASTAR Inc., Madison, WI, USA) by using clustal method.

Southern hybridization analysis

About 10 μg of genomic DNA was digested with 50 units of restriction enzyme *Taq*I in a final volume of 50 μl at 37°C for 18 h. The cleaved DNA fragments were run on a 0.8% (w/v) agarose gel and transferred to Hybond-N+ membranes (Amersham Pharmacia, Little Chalfont, Buckinghamshire, England). The membranes

Table 1 A list of different accessions of soybean used in this article

Accessions	Latin name	Seed color	Origin area[a]	Oil content (% w/w)	Usage in this study
8904	*G. max*	Yellow	–	21.50	Cloning the cDNA and genomic sequence. Analysis of *GmDGAT* expression
Xiangchundou15	*G. max*	Yellow	Hunan	15.98	Cloning the cDNA sequence
DianYD028	*G. max*	Yellow	Yunnan	18.25	Cloning the cDNA sequence
Jidou12	*G. max*	Yellow	Heibei	17.14	Cloning the cDNA sequence
Jilin30	*G. max*	Yellow	Jilin	16.61	Cloning the cDNA sequence
Heilong38	*G. max*	Yellow	Heilongjiang	19.60	Cloning the cDNA sequence
Wandou16	*G. max*	Yellow	Anhui	17.30	Cloning the cDNA sequence
ZYD4617	*G. soja*	Black	Jiangxi	11.50	Cloning the cDNA sequence
ZYD3708	*G. soja*	Black	Shanxi	9.65	Cloning the cDNA sequence
ZYD4174	*G. soja*	Black	Jiangsu	11.70	Cloning the cDNA sequence
ZYD4433	*G. soja*	Black	Zhejiang	10.48	Cloning the cDNA sequence
ZYD3235	*G. soja*	Black	Shandong	9.54	Cloning the cDNA sequence
PW0063	*G. tomentella*	Black	–	–	Cloning the cDNA and genomic sequence
PW0031	*G. latifolia*	Black	–	–	Cloning the cDNA and genomic sequence. Analysis of *GmDGAT* expression
Heihuangdou	*G. max*	Black	–	22.00	Cloning the genomic sequence
PI547844	*G. max*	Yellow	–	22.10	Cloning the genomic sequence
y43	*G. gracilis*	Black	–	–	Cloning the genomic sequence
y74	*G. gracilis*	Black	–	–	Cloning the genomic sequence

[a]Origin area is the province name of China

were hybridized with [α-^{32}p]dCTP-labeled (Amersham Pharmacia) 0.8 kbp *GmDGAT* fragment mentioned above. The hybridization procedures followed the standard protocols (Chen et al. 1991). The membrane was washed once with 2×SSC plus 0.1% (w/v) SDS at 65°C for 15 min, then 1×SSC plus 0.1% (w/v) SDS at 65°C for 8 min and exposed to FUJI Medical X-ray film at −70°C.

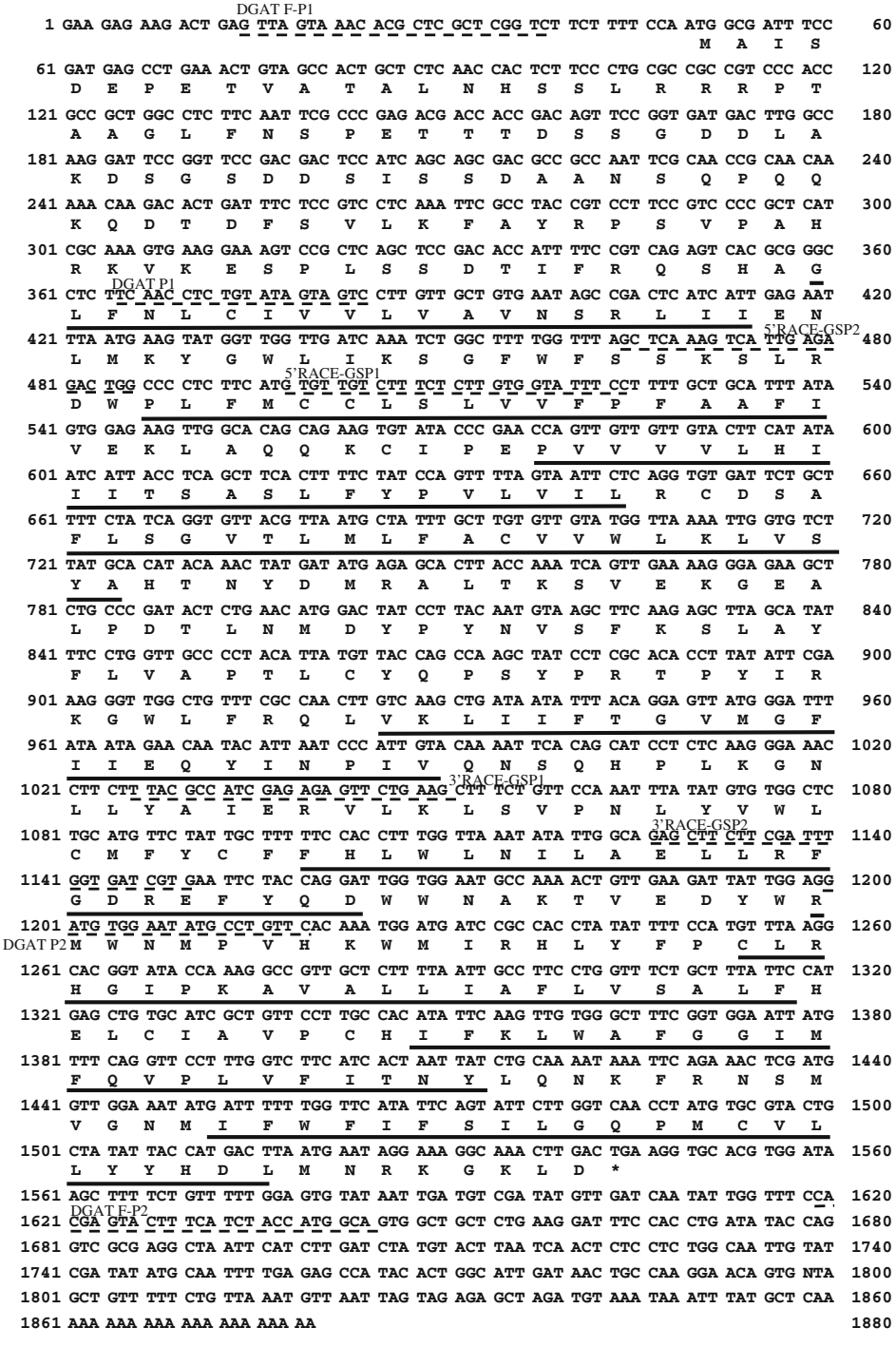
Cloning of the *DGAT* genomic sequences

Partial genomic sequences containing the *DGAT* gene from the genomic DNAs of six accessions, *G. max* PI547844, *G. max* Heihuangdou, *G. gracilis* y43, *G. gracilis* y74, *G. latifolia* PW0031 and *G. tomentella* PW0063, were amplified with a pair of primer (DGAT P1 and DGAT PG-2) (Table 2). The full-length

Table 2 Primers used in this study

Name	Sequence (5′ → 3′)	Usage in this study
DGAT P1	TCA ACC TCT GTA TAG TAG TC	Cloning the partial *GmDGAT* cDNA and genomic DNA
DGAT P2	GAA CAG GCA TAT TCC ACA TC	Cloning the partial *GmDGAT* cDNA and genomic DNA
5′RACE-GSP1	GGA AAT ACC ACA AGA GAA AGA CAA CAC	RACE
5′RACE-GSP2	CCA GTC TCT CAA TGA CTT TGA GC	RACE
3′RACE-GSP1	TTA CGC CAT CGA GAG AGT TCT GAA GC	RACE
3′RACE-GSP2	GAG CTT CTT CGA TTT GGT GAT CGT G	RACE
DGAT F-P1	GTT AGT AAA CAC GCT CGC TCG GTC	Cloning the full-length *GmDGAT* cDNA
DGAT F-P2	CTG CCA TGG TAG ATG AAA GTA CTC GTG	Cloning the full-length *GmDGAT* cDNA
DGAT PG-2	TT CCA CAT CAA AGA TAT ATG	Cloning the partial *GmDGAT* genomic DNA
G-primer1	TGA GTT AGT AAA CAC GCT CGC TCG GTC	Cloning the *GmDGAT* genomic sequence
G-primer1′	CCA GTC TCT CAA TGA CTT TGA GC	Cloning the *GmDGAT* genomic sequence
G-primer2	GAT TAT TCA ACC TCT GTA TAG TAG TC	Cloning the *GmDGAT* genomic sequence
G-primer2′	GAA CAG GCA TAT TCC ACA TC	Cloning the *GmDGAT* genomic sequence
G-primer3	CAA GAG CTT AGC ATA TTT CCT GGT TGC	Cloning the *GmDGAT* genomic sequence
G-primer3′	CAG CCA ACC CTT TCG AAT ATA AGG TGT GCG AG	Cloning the *GmDGAT* genomic sequence
G-primer4	CAT TGT ACA AAA TTC ACA GCA TC CTC	Cloning the *GmDGAT* genomic sequence
G-primer4′	ATG GAA AAT ATA GGT GGC GGA TCA TCC	Cloning the *GmDGAT* genomic sequence
G-primer5	CAC CTA TAT TTT CCA TGT TTA AGG CAC	Cloning the *GmDGAT* genomic sequence
G-primer5′	AAT ATA GCA GTA CGC ACA TAG GTT GAC	Cloning the *GmDGAT* genomic sequence
GmDGAT-RT-P-1	GTG GAG AAG TTG GCA CAG CAG AAG TG	Analysis of *GmDGAT* expression
GmDGAT-RT-P-2	GTG GCA AGG AAC AGC GAT GCA CAG CTC	Analysis of *GmDGAT* expression
β-Tubulin P1	AAC CTC CTC CTC ATC GTA CT	Analysis of *GmDGAT* expression
β-Tubulin P2	GAC AGC ATC AGC CAT GTT TCA	Analysis of *GmDGAT* expression

A



B

Fig. 1 *GmDGAT* gene sequence and predicted protein structure. **a** Nucleotide and deduced amino acid sequence of the *GmDGAT*. The deduced amino acid sequence was presented under the DNA sequence in *bold capital letters* and the stop codon was marked with an *asterisk*. The positions of the potential transmembrane domains are indicated by *black lines*. The *broken lines* under the nucleotide sequence indicate the position of primers. **b** Schematic representation of the GmDGAT structure. The *black rectangles* represent the potential transmembrane domains predicted by SMART

fragment of *DGAT* was amplified from the genomic DNA of *G. max* 8904 and *G. tomentella* PW0063 respectively with five pairs of primers (Table 1). The total volume of the PCR reaction mixture was 25 μl, containing 100 ng of the genomic DNA, 0.5 μM of each primer, 1×PCR buffer, 0.4 mM dNTP, and 1 unit of LA DNA polymerase (Takara, Kyoto, Japan). The reaction was denatured at 94°C for 5 min, and followed by 35 cycles of 1 min at 94°C, 1 min at 56°C, 2 min at 72°C and then 10 min at 72°C. The amplified fragments of the *DGAT* genomic sequence were recovered, cloned into pMD18-T vector (Takara, Kyoto, Japan) and three individual clones were selected and sequenced.

Analysis of *GmDGAT* expression

Five microgram of the total RNA isolated from young leaves, mature leaves, flowers, and pods from the cultivar 8904 or wild-type *G. latifolia* PW0031 was used for cDNA synthesis by using a first strand cDNA synthesis kit (Promega, Madison, WI, USA) in a 20 μl reaction volume. The first strand cDNA mix was used as template for RT-PCR.

The *GmDGAT*-RT-P-1 and *GmDGAT*-RT-P-2 were used as primers (Table 2). The total volume of the PCR reaction mixture was 25 μl, containing 1 μl cDNA, 0.5 μM of each primer, 1×PCR buffer, 0.4 mM dNTP, and 1 unit of rTaq enzyme (Takara). The reaction was denatured at 94°C for 2 min, and then followed by 30 cycles of 1 min at 94°C, 1 min at 56°C, 1 min at 72°C and concluded with one step of 10 min at 72°C.

Amplified fragments were separated on a 1% (w/v) agarose gel and stained with ethidium bromide. The agarose gel was scanned by using Gel Doc GS670 (Bio-Rad, Hercules, CA, USA). A soybean β-Tubulin gene, amplified with primers β-Tubulin P1 and β-Tubulin P2 (Table 2), was used as a control in the experiments.

Results and discussion

Cloning and structural analysis of the soybean *DGAT* gene

Arabidopsis *AtDGAT* gene was used as a query to search against the soybean EST database and two homologous ESTs were identified. The two putative *GmDGAT* ESTs were 559 bp (BM178620) and 572 bp (BM309669) in length, respectively, and non-overlapping, possibly representing the 5′- and 3′- sequence of *GmDGAT* when compared with other plant *DGAT*s. Two specific primers (DGAT P1 and DGAT P2) were thus designed from these two ESTs and a 0.8 kbp fragment of the *GmDGAT* gene were obtained from soybean cultivar 8904. Using the RACE method, the 5′- and 3′- ends of the *GmDGAT* cDNA were further cloned from soybean pod cDNAs, and the full-length cDNA of *GmDGAT* was obtained under the accession number of AY496439. The full-length cDNA of *GmDGAT* was 1,880 bp in length containing 48 bp of the 5′- leader sequence and 335 bp of the 3′- untranslated region. The complete open reading frame (ORF) of 1,497 bp encoded a protein of 498 amino acids (Fig. 1a), with a predicted molecular weight of 57.3 kDa and a calculated isoelectric point of 8.89 (Protparam: http://www.expasy.ch). By BLAST against the ESTs in the GenBank, we also assembled the complete ORF of *LcDGAT* (AY859489) from *L. corniculatus* and *OsDGAT* (AY858584) from rice.

In plants, DGAT has been shown to localize in the endoplasmic reticulum (Cao and Huang 1986; Settlage et al. 1995; Lacey and Hill 1996) where the Kennedy pathway mainly occurs. Using SMART program (Simple Modular Architecture Research Tool: http://www.smart.embl-heidelberg.de/), nine potential transmembrane domains were identified in the GmDGAT, suggesting that the protein is also localized in the membrane systems (Fig. 1a, b). A putative signal peptide peptidase domain (amino acids 10–232) and an acid phosphatase homolog region (amino acid 108–216) were also recognized. However, these domains may not be active since some of the required catalytic sites were not detected in these regions. A WWE domain (amino acid 214–279) was predicted, which is named after three of its conserved residues and is predicted to mediate specific protein–protein interactions in ubiquitin and ADP ribose conjugation systems (Aravind 2001). By comparing the soybean DGAT with other plant DGATs, an MBOAT (membrane bound O-acyltransferase) domain (amino acid 204–489) was identified in the GmDGAT (Fig. 2a). This domain is possibly involved in acyl transfer (Hofmann 2000). Compared with other proteins, the full-length GmDGAT exhibited 78.2% similarity to LcDGAT from *L. corniculatus* (AAW51456) and 66.3% to AtDGAT (AAF19262). The highest similarity was mainly within the C-terminal region, which contained the MBOAT domain. The MBOAT domains shared a similarity of 91.9% between GmDGAT and LcDGAT, and 78.4% between GmDGAT and AtDGAT. The N-terminal region showed a high degree of variation among the DGATs compared (Fig. 2a). Cluster analysis was also performed and the GmDGAT was grouped with that of *L. corniculatus*, a legume plant, and it also appeared to be closely related to the DGATs from *Arabidopsis* and *Brassica*. However, the GmDGAT may be more divergent when compared with the proteins from tobacco and rice (Fig. 2b).

Comparative analysis of the DGATs from 14 soybean accessions

Soybean germplasm is abundant and highly diversified in China, where several wild *Glycine* species are native. There are many differences between wild-type species and cultivated soybean with regard to traits such as oil content, seed color, and plant morphology. It would be

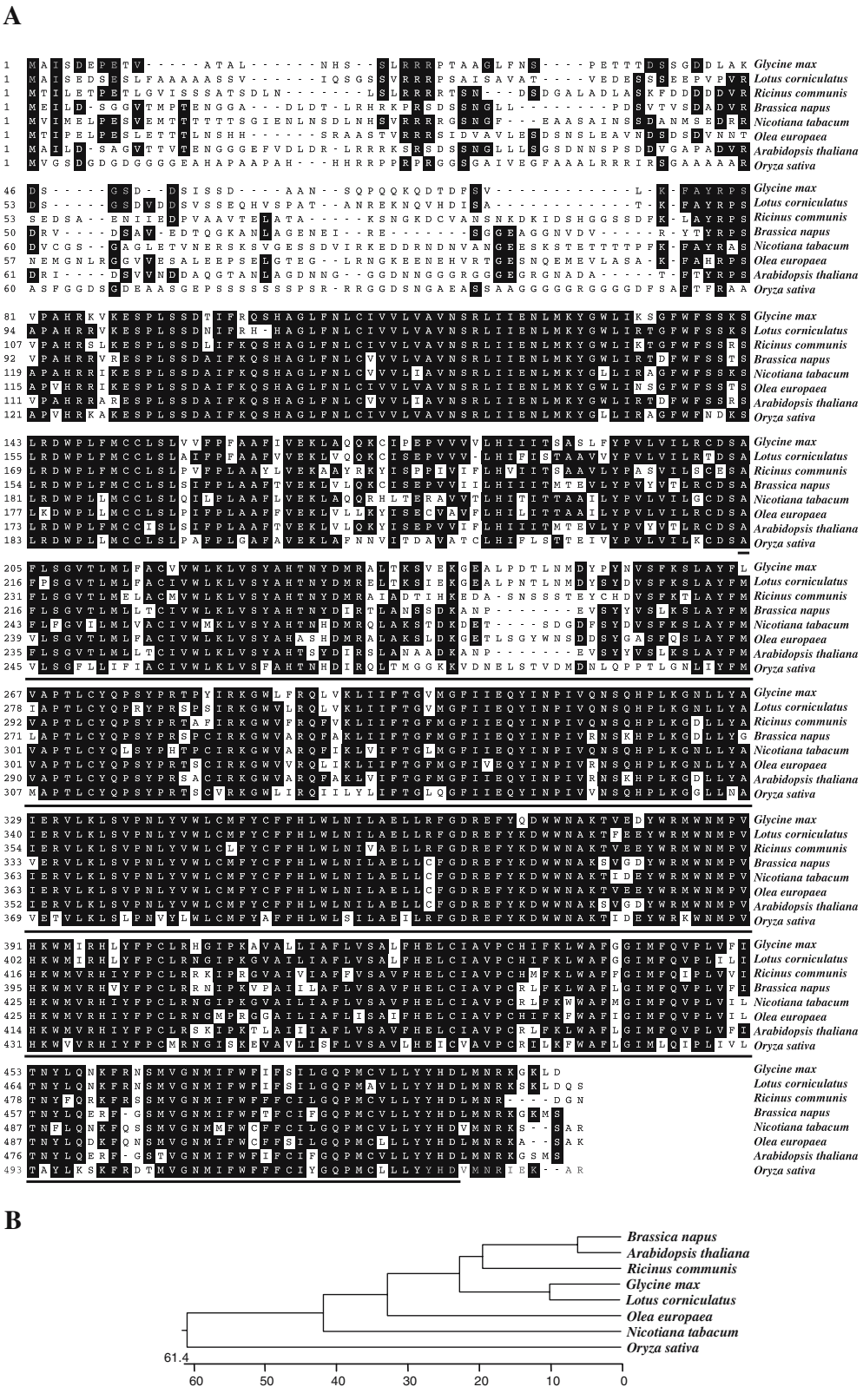


Fig. 2 Comparison of different plant DGATs. **a** Alignment of multiple plant DGATs. The DGATs from *Arabidopsis thaliana* (AAF19262), *Brassica napus* (AAF64065), *Nicotiana tabacum* (AAF19345), *Olea europaea* (AAS01606), *Ricinus communis* (AAR11479), and the two assembled sequences from *Oryza sativa* (AAW47581) and *Lotus corniculatus* (AAW51456) were compared. Conserved sequences were shaded in *black*. The *bold line* under the sequence represents the MOBAT domain in different plants. **b** The phylogenetic tree of different plant DGATs. The length of each pair branches represents the distance between sequence pairs. The units at the bottom of the tree indicate the number of substitution events

interesting to know if there is any difference in the amino acid sequence of the DGAT protein from different soybean accessions. The full length of the *DGAT* cDNA was cloned from 13 accessions and compared with the *GmDGAT* gene of accession 8904. Among these 14 accessions, 7 were from *G. max*, 5 were from *G. soja*, 1 was from *G. latifolia* (PW0031) and 1 was from *G. tomentella* (PW0063) (Table 1). Alignment of the deduced amino acid sequences of these DGATs revealed that the C-terminal regions were almost the same except for a few amino acid changes, whereas a number of differences in the N-terminal region were identified among these proteins (Fig. 3a, b). Seven DGATs from *G. max* showed similarities from 98.2 to 100% among the members. When compared with the five DGATs from *G. soja*, the DGATs from *G. max* showed similarities from 97.8 to 100% (Fig. 3b). The DGATs from cultivar 8904 and Xiangchundou15 were the same as those in accessions ZYD4174 and ZYD3235 from *G. soja*. The DGATs from the two wild-type accessions, PW0031 (*G. latifolia*) and PW0063 (*G. tomentella*), showed a distinct divergence in comparison with the other proteins. The high similarity of the DGATs between members from both *G. max* and *G. soja* indicates that *G. soja* is an ancestor of *G. max*. The *G. latifolia* and *G. tomentella* have a more distant relationship to *G. max* and *G. soja*. Other cytological, morphological and molecular evidence also suggest that *G. soja* is the ancestor of *G. max* (Hymowitz 1970; Doyle and Beachy 1985; Shoemaker et al. 1986; Doyle 1988).

In the N-terminal region, four locations were recognized where amino acid residues were missing when a comparison was made between the soybean accessions. The first difference occurred in the 26th residue where three residues (Thr-Ser-Ala) were missing in the 12 accessions of subgenus *Soja* in comparison with other wild-type accessions PW0031 and PW0063. The second was found at the 58th position where one or two residues were missing in all accessions of subgenus *Soja* but not in *G. latifolia* (PW0031) and *G. tomentella* (PW0063). The third occurred in the 63rd amino acid where an Asn was missing in all subgenus *Soja* accessions and wild-type accessions PW0031. The last difference at the 101th position showed an interesting variation (Fig. 3a). Both cultivars DianYD028 and Wandou16 had the same sequence Gln-Leu-Gln as that of the wild type PW0031 (*G. latifolia*), although they were from different subgenus, whereas the other 11 accessions have a single Gln residue at this position. Moreover, there were 20 single amino acid changes in at least one of these 14 accessions of soybean (Fig. 3). Considering the properties of polarity and charge of the changed amino acids, ten changes are synonymous and ten changes are nonsynonymous.

In the C-terminal regions, only five single-residue changes were found (Fig. 3). There are two synonymous single amino acid changes and three non-synonymous single amino acid changes. The amino acid change at the 371th position should be noted. In the two wild-type accessions, PW0063 and PW0031, this position is a Lys whereas in the other 12 accessions, 7 from *G. max* cultivar and 5 from *G. soja*, a Gln is present instead of Lys (Fig. 3a). In DGATs from other plants such as *Lotus corniculatus*, *Ricinus communis*, *Brassica napus*, *Nicotiana tabacum*, *Olea europaea*, *Arabidopsis thaliana* and *Oryza sativa*, there is also a Lys at this position (Fig. 2a), which is the same as the ones in the two wild-type soybean accessions. It seems that Gln^{371} is specific for *G. max* and *G. soja* but not for *G. latifolia* and *G. tomentella* and the other plant species examined. It is possible that the DGATs from the accessions of *G. max* and *G. soja* are derived from the Lys to Gln mutation, and it might be additional molecular evidence that the cultivated soybean is derived from *G. soja*. The other functions of this mutation need to be further investigated.

We further examined if the protein sequence variation accounts for different oil levels between the soybean accessions. We find that although the DGATs from these 14 accessions are highly homologous, the oil contents in these accessions are more variable (Table 1). Transcript abundance of the *DGAT* between the soybean accessions had no significant difference (data not shown). This may indicate that the DGATs do not directly correlate with the oil content in soybean seeds. Other factors may play roles in determination of the oil content.

Genomic architecture of the *GmDGAT* gene in various soybean accessions

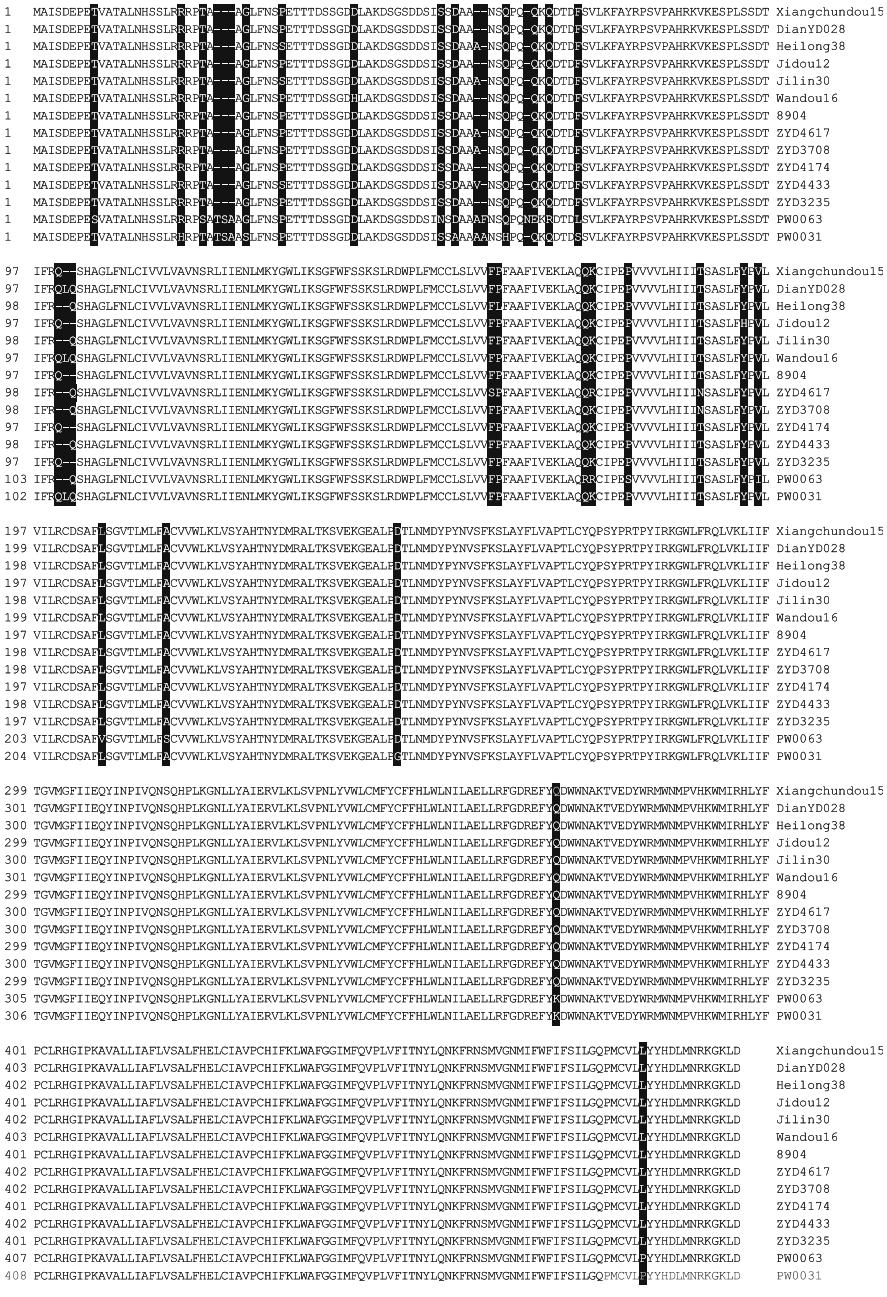
We further investigated the difference in the genomic architecture of the *DGAT* gene between soybean accessions. The genomic DNAs from 27 accessions were digested with *Taq*I and subjected to Southern analysis. The 0.8 kbp *GmDGAT* fragment derived from two original ESTs was used as a probe. The results are shown in Fig. 4. It can be seen that the hybridization patterns are the same in all accessions from the subgenus *Soja*, with two major bands and a few weaker bands. It is also possible that minor variations in fragment length existed among these different accessions. However, the hybridization pattern was completely different in the subgenus *Glycine* genomes, with some accessions having two bands and others having up to seven bands. Therefore, the genomic architecture of the *DGAT* gene is very similar among the accessions in the subgenus *Soja*, but is more diversified among the accessions in subgenus *Glycine*.

Comparison of the *DGAT* genomic sequences from various soybean accessions

Although the amino acid sequences of the soybean DGATs are highly similar, their genomic architectures are more varied between subgenus *Soja* and *Glycine*. It probably suggests that there are major differences in the introns of the genomic sequence of the *DGAT* genes.

Fig. 3 Comparison of the deduced amino acid sequence of the DGATs from different accessions of soybean. **a** Alignment of the soybean DGAT sequences. The names of accessions are shown on the *right* and amino acid sequence numbers are show on the *left*. Differences in amino acids are shaded in *black*. **b** The percent divergence and similarity between the soybean DGATs

A



B

Percent Similarity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14			
1	—	99.4	99.0	99.8	99.2	99.2	100.0	98.8	99.2	100.0	99.2	100.0	95.4	96.6	**1**	**Xiangchundou15**	
2	0.0	—	98.2	99.2	98.4	99.8	99.4	98.0	98.4	99.4	98.4	99.4	94.4	97.2	**2**	**DianYD028**	
3	0.4	0.4	—	98.8	99.8	98.0	99.0	99.0	99.4	99.0	99.6	99.0	95.0	96.2	**3**	**Heilong38**	
4	0.2	0.2	0.6	—	99.0	99.0	99.8	98.6	99.0	99.8	99.0	99.8	95.2	96.4	**4**	**Jidou12**	***G.max***
5	0.2	0.2	0.2	0.4	—	98.2	99.2	99.2	99.6	99.2	99.8	99.2	95.2	96.4	**5**	**Jilin30**	
6	0.2	0.2	0.6	0.4	0.4	—	99.2	97.8	98.2	99.2	98.2	99.2	94.2	97.0	**6**	**Wandou16**	
7	0.0	0.0	0.4	0.2	0.2	0.2	—	98.8	99.2	100.0	99.2	100.0	95.4	96.6	**7**	**8904**	
8	0.6	0.6	1.0	0.8	0.8	0.8	0.6	—	99.6	98.8	99.0	98.8	95.2	96.0	**8**	**ZYD4617**	
9	0.2	0.2	0.6	0.4	0.4	0.4	0.2	0.4	—	99.2	99.4	99.2	95.2	96.4	**9**	**ZYD3708**	
10	0.0	0.0	0.4	0.2	0.2	0.2	0.0	0.6	0.2	—	99.2	100.0	95.4	96.6	**10**	**ZYD4174**	***G. soja***
11	0.2	0.2	0.4	0.4	0.2	0.4	0.2	1.0	0.6	0.2	—	99.2	95.0	96.2	**11**	**ZYD4433**	
12	0.0	0.0	0.4	0.2	0.2	0.2	0.0	0.6	0.2	0.0	0.2	—	95.4	96.6	**12**	**ZYD3235**	
13	2.9	2.9	3.3	3.1	3.1	3.1	2.9	3.1	3.1	2.9	3.3	2.9	—	95.0	**13**	**PW0063**	***G. tomentella***
14	1.6	1.6	2.0	1.8	1.8	1.8	1.6	2.2	1.8	1.6	2.0	1.6	3.7	—	**14**	**PW0031**	***G. latifolia***
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**			

Percent Divergence

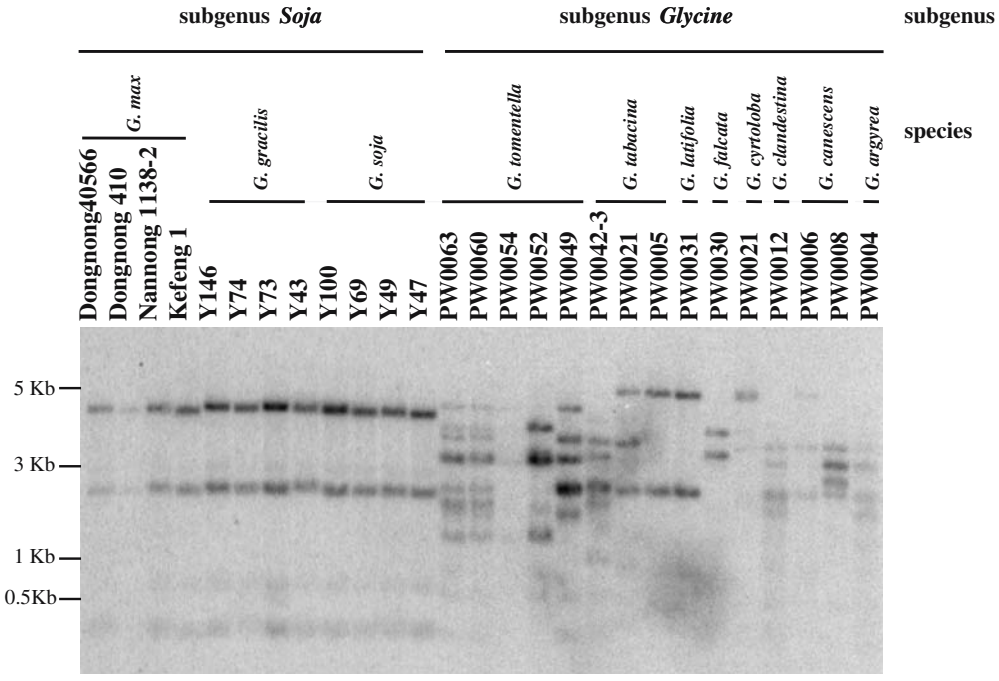


Fig. 4 Southern analysis of the *GmDGAT* gene in the different accessions of soybean. The genomic DNA was digested with *Taq*I and then subjected to Southern analysis. Names of accessions are listed on the top of the figure and subgenera classifications are marked above the names

A pair of specific primers was then designed from the 0.8 kbp cDNA fragment and partial genomic sequences of the *DGAT* were amplified from six accessions, which belong to *G. max*, *G. gracilis*, *G. tomentella* and *G. latifolia*, respectively (Fig. 5a). Comparative and cluster analysis indicated that the two genomic sequences from *G. max* were the same and very similar to that from the two accessions of *G. gracilis* (Fig. 5a). This relationship is consistent with the fact that *G. gracilis* is classified in the subgenus *Soja*. However, the sequences from *G. tomentella* and *G. latifolia* were more diverged (Fig. 5a, b). Differences in the intron between cultivated accessions and wild-type accessions were found (Fig. 5b). The largest difference was an insertion of 130 bp at the 1,833th position of *G. tomentella* (PW0063) compared to others. Moreover, there were 17 small differences among the three sequences. These differences were insertions/deletions from 3 to 14 bp in length. In addition, there were many minor differences of 1 or 2 bp among these sequences. We did not find major differences between the exons of six accessions. Therefore, the introns are the most variable part of the *DGAT* genomic sequences. Through comparison of the partial genomic sequences from the subgenus *Soja* and *Glycine* accessions, we found that the differences in genomic architecture were mainly due to the difference in intron sequence and/or size, which was possibly affected by the deletion or insertion event.

Based on the cDNA sequences, the full-length of the *DGAT* genomic sequences from the cultivated accession 8904 (*GmDGAT*) and the wild-type accession PW0031 (*GlDGAT*) were cloned by PCR. Their intron/exon structures were analyzed and compared with those from other plants (Fig. 6). The full-length genomic sequences of *AtDGAT* from Arabidopsis, *OsDGAT* from rice and *LcDGAT* from Lotus were obtained from Genebank, and the sizes of *AtDGAT*, *OsDGAT*, *LcDGAT*, *GmDGAT* and *GlDGAT* were 3,020, 6,220, 5,762, 7,575 and 6,614 bp, respectively. Although there were major differences in length among them, they shared high similarity in the deduced amino acid sequences (Figs. 2, 3). Both the genomic DNAs of *AtDGAT* and *OsDGAT* had 16 exons and most exon lengths were very similar (Fig. 6), while most of the introns (11 out of 15) were longer in *OsDGAT* than those in *AtDGAT*. Unlike the *DGAT* genes in rice and Arabidopsis, all three genes from legume had 15 exons, 12 of these were exactly the same among the three genes. The only difference in length lies in the first, second and the last exon (Fig. 6). The difference in the exon number between *AtDGAT* and *OsDGAT* and three legume genes may be due to the combination of the last two exons in the legume genes. The lengths of the last two exons in *AtDGAT* and *OsDGAT* showed much similarity with the last exon in the three legume genes. The introns of five genomic DNAs showed great variation in length, ranging from 75 to 141 bp in *AtDGAT* and 98 to 1,299 bp in *GmDGAT*. The introns of the three legume genes also showed much difference in length. The intron size may determine the gene size and the genome size of each plant. However, the biological significance of the variation of the intron size remains to be further studied.

Expression of the soybean *DGAT* gene

Transcript abundance of the *DGAT* was examined in different organs from the cultivar 8904 and wild-type *G. latifolia* PW0031 by RT-PCR. The same cDNAs were also used to amplify the β-Tubulin gene as an internal

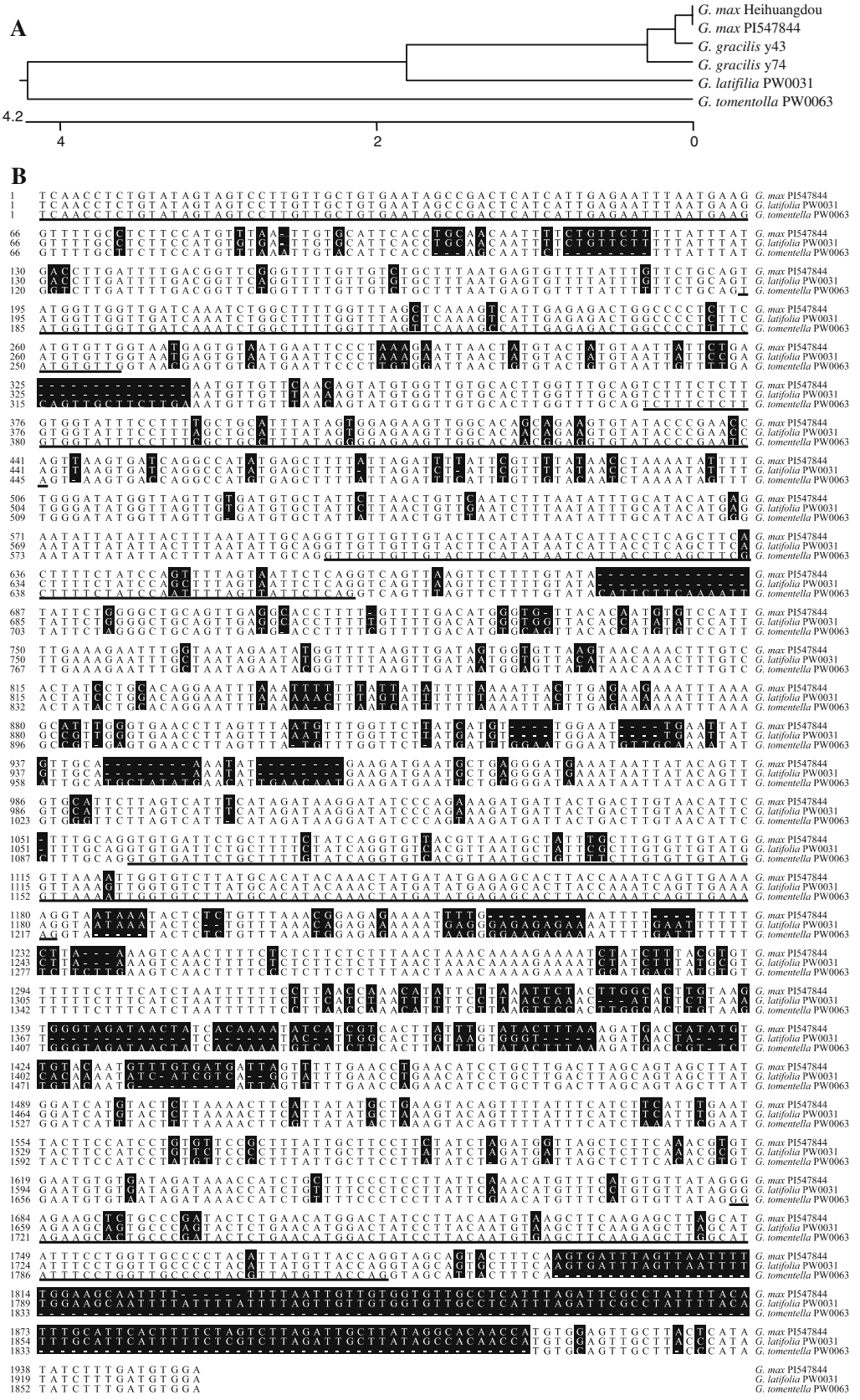


Fig. 5 Comparison of the partial genomic sequences of the *DGATs* from six soybean accessions. **a** The phylogenetic tree of the *DGATs* from six accessions of soybean. The length of each pair branches represents the distance between sequence pairs. The units at the bottom of the tree indicate the number of substitution events. **b** Alignment of the *DGAT* partial genomic sequences. The names of accessions are shown on the *right* and nucleotide sequence numbers were shown on the *left*. Conserved sequences are shown on a *white background* and the differences are shaded in *black*. The *bold lines* under the sequence indicate the positions of the exons. Because the sequences from *G. max* Heihuangdou, *G. gracilis* y43 and *G. gracilis* y74 are highly homologous to the sequence from *G. max* PI547844, only the sequences from *G. max* PI547844, *G. latifolia* PW0031 and *G. tomentella* PW0063 were compared

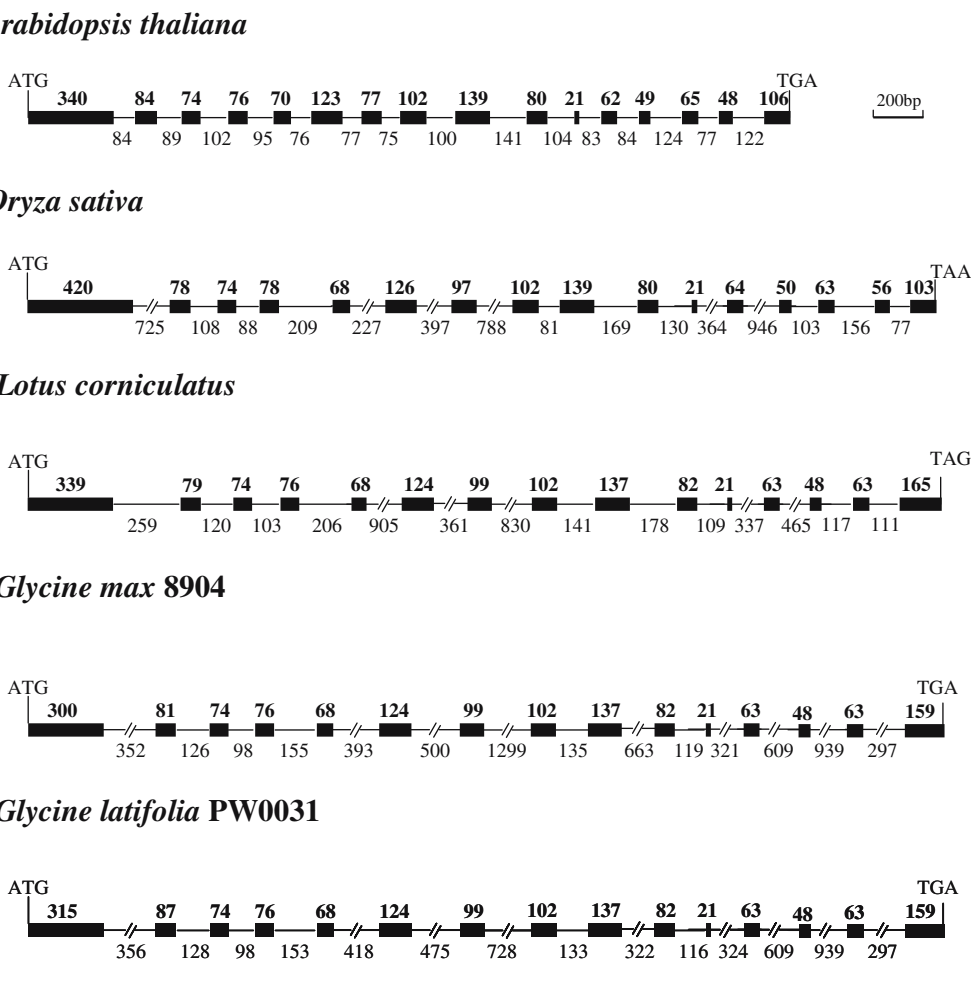


Fig. 6 Genomic DNA structure of the *DGATs* from soybean and other plants. The *filled boxes* indicate exons and the *lines* indicate introns

control. The results showed that the *DGAT* expression was relatively higher in flower and pods at 30 DAF, but lower in mature leaves and the pods at 20 DAF in both cultivar 8904 and PW0031 (Fig. 7). However, almost no expression was detected in young leaves and in pods at 10 DAF (Fig. 7). This pattern resembled the RNA expression pattern in Arabidopsis (Kaup et al. 2002). These results suggest that the *DGAT* gene is expressed at the later stage of the pod development, and that the *DGAT* expression patterns have not significantly changed during the evolution of legume plants.

Although the soybean *DGAT* gene has been cloned from a number of accessions, it seems that this gene is not closely related to the oil content based on the structural and expression analysis (Table 1; Figs. 3a, 7, and data not shown). This is due to the fact that many steps are involved in the lipid biosynthesis and many oil QTLs have been detected. Currently, more than 60 oil QTLs have been detected (http://www.soybase.ncgr.org; Zhang et al. 2004) and it is not known if the soybean *DGAT* gene is linked to or belongs to any oil QTL. Because no polymorphism was detected between the two parents from which a mapping population has been derived, the soybean *DGAT* gene cannot be mapped onto the linkage map available (Zhang et al. 2004; data

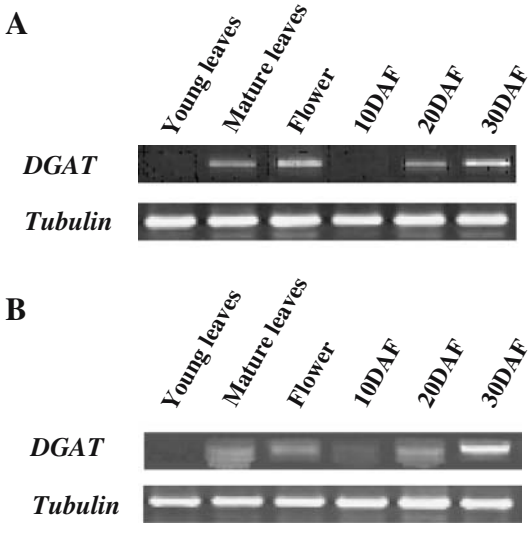


Fig. 7 RT-PCR analysis of the *DGAT* gene expression in different organs of soybean plants. Total RNA from cultivated soybean 8904 and wild-type soybean PW0031 were extracted from young leaves, mature leaves, flowers, and pods at 10 DAF, 20 DAF and 30 DAF. The first-strand cDNA was used as template to amplify the *GmDGAT*. *β*-Tubulin gene was amplified as an internal control

not shown). Further detection of the soybean *DGAT* gene polymorphism in other parents and its mapping in the corresponding population may facilitate our understanding of the relationship between the *DGAT* gene and oil QTLs in soybean.

In conclusion, we cloned *DGAT* genes from different accessions of soybean plants and compared the difference at both the amino acid level and the nucleotide level. We find that the difference at the amino acid level is relatively small whereas the difference at the genomic sequence level is large. These results may have significance in elucidation of the gene evolution between wild-type soybean plants and the cultivated ones.

Acknowledgements This work was supported by the Major Basic Research Program of China (2002CB111303), National Nature Science Foundation (30392100) and National 863 Program (2002AA211051), and CAS Project (KSCX2-SW-328)

References

Aravind L (2001) The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. Trends Biochem Sci 26:273–275

Cao YZ, Huang AHC (1986) Diacylglycerol acyltransferase in maturing oil seeds of maize and other species. Plant Physiol 82:813–820

Cases S, Smith SJ, Zheng YW, Myers HM, Lear SR, Sande E, Novak S, Collins C, Welch CB, Lusis AJ, Erickson SK, Farese RV (1998) Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proc Natl Acad Sci USA 95:13018–13023

Chen SY, Zhu LH, Hong J, Chen SL (1991) Molecular biological identification of a salt-tolerant rice line. Acta Bot Sin 33:569–573

Doyle JJ (1988) 5S ribosomal gene variation in the soybean and its progenitor. Theor Appl Genet 75:621–624

Doyle JJ, Beachy RN (1985) Ribosomal gene variation in soybean (*Glycine*) and its wild relatives. Theor Appl Genet 70:369–376

Frohman MA (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. Methods Enzymol 218:340–358

He CY, Zhang JS, Chen SY (2002) A soybean gene encoding a proline-rich protein is regulated by salicylic acid, an endogenous circadian rhythm and by various stresses. Theor Appl Genet 104:1125–1131

He CY, Tian AG, Zhang JS, Zhang ZY, Chen SY (2003) Isolation and characterization of a full-length resistance gene homolog from soybean. Theor Appl Genet 106:786–793

He X, Turner C, Chen GQ, Lin JT, McKeon TA (2004) Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. Lipids 39:311–318

Hobbs DH, Hills MJ (2000) Expression and characterization of diacylglycerol acyltransferase from *Arabidopsis thaliana* in insect cell cultures. Biochem Soc Trans 28:687–689

Hobbs DH, Lu C, Hills MJ (1999) Cloning of a cDNA encoding diacylglycerol acyltransferase from *Arabidopsis thaliana* and its functional expression. FEBS Lett 452:145–149

Hofmann K (2000) A superfamily of membrane-bound O-acyltransferases with implications for wnt signaling. Trends Biochem Sci 25:111–112

Hymowitz T (1970) On the domestication of the soybean. Econ Bot 24:408–421

Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM, Covello PS, Taylor DC (2001) Seed-specific over-expression of an arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol 126:861–874

Kaup MT, Froese CD, Thompson JE (2002) A role for triacylglycerol acyltransferase during leaf senescence. Plant Physiol 129:1616–1626

Kennedy EP (1961) Biosynthesis of complex lipids. Fed Proc 20:934–940

Lacey DJ, Hill MJ (1996) Heterogeneity of the endoplasmic reticulum with respect to lipid synthesis in developing seeds of *Brassica napus* L. Planta 199:545–551

Liu F, Zhuang BC, Zhang JS, Chen SY (2000) Construction and analysis of soybean genetic map. Acta Bot Sin 27:1018–1026

Lu C, Hills MJ (2002) Arabidopsis mutants deficient in diacylglycerol acyltransferase display increased sensitivity to abscisic acid, sugars, and osmotic stress during germination and seedling development. Plant Physiol 129:1352–1358

Lu C, de Noyer SB, Hobbs DH, Kang J, Wen Y, Krachtus D, Hills MJ (2003) Expression pattern of diacylglycerol acyltransferase-1, an enzyme involved in triacylglycerol biosynthesis, in *Arabidopsis thaliana*. Plant Mol Biol 52:31–41

Luo GZ, Wang HW, Huang J, Tian AG, Wang YJ, Zhang JS, Chen SY (2005) A putative plasma membrane cation/proton antiporter from soybean confers salt tolerance in Arabidopsis. Plant Mol Biol 59:809–820

Nykiforuk CL, Laroche A, Weselake RJ (1999) Isolation and sequence analysis of a novel cDNA encoding a putative diacylglycerol acyltransferase from a microspore-derived cell suspension culture of *Brassica napus* L. cv Jet Neuf. Plant Physiol 120:99–123

Perry HJ, Harwood JL (1993) Changes in the lipid content of developing seeds of *Brassica napus*. Phytochemistry 32:1411–1415

Settlage SB, Kwanyuen P, Wilson RF (1998) Relation between diacylglycerol acyltransferase activity and oil concentration in soybean. JAOCS 75:775–781

Settlage SB, Wilson RF, Kwanyien P (1995) Localization of diacylglycerol acyltransferase to oil body associated endoplasmic reticulum. Plant Physiol Biochem 33:399–407

Shoemaker RC, Hatfield PM, Palmer RG, Atherly AA (1986) Chloroplast DNA variation and evolution in the genus *Glycine* subgenus *soja*. J Hered 77:26–30

Singh RJ, Hymowitz T (1999) Soybean genetic resources and crop improvement. Genome 42:605–616

Tian AG, Zhao JY, Zhang JS, Gai JY, Chen SY (2004a) Genomic characterization of the S-adenosylmethionine decarboxylase genes from soybean. Theor Appl Genet 108:842–850

Tian AG, Wang J, Cui P, Yu J, Xu HH, Cong LJ, Huang XG, Wang XL, Jiao YZ, Wang BJ, Wang YJ, Zhang JS, Chen SY (2004b) Characterization of soybean genomic features by analysis of its expressed sequence tags. Theor Appl Genet 108:903–913

Triki S, Ben Hamida J, Mazliak P. (2000) Diacylglycerol acyltransferase in maturing sunflower seeds. Biochem Soc Trans 28:689–692

Wang YJ, Li YD, Luo GZ, Tian AG, Wang HW, Zhang JS, Chen SY (2005) Cloning and characterization of an HDZip I gene GmHZ1 from soybean. Planta 221:831–843

Wu XL, He CY, Wang YJ, Zhang ZY, Dongfang Y, Zhang JS, Chen SY, Gai JY (2001) Construction and analysis of a genetic linkage map of soybean. Acta Genet Sin 28:1051–1061

Zhang DS, Dong W, Hui DW, Chen SY, Zhuang BC (1997) Construction of a soybean linkage map using an F2 hybrid population from a cultivated variety and a semi-wild soybean. Chin Sci Bull 42:1326–1330

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139

Zhang JS, Zhou JM, Zhang C, Chen SY (1996) Differential gene expression in a salt-tolerance rice mutant and its parental variety. Sci China (Series C) 39:310–319

Zou J, Wei Y, Jako C, Kumar A, Selvaraj G, Taylor DC (1999) The *Arabidopsis thaliana TAG1* mutant has a mutation in a diacylglycerol acyltransferase gene. Plant J 19:645–653