



***In silico* characterization of 1,2-diacylglycerol cholinephosphotransferase and lysophosphatidylcholine acyltransferase genes in *Glycine max* L. Merrill**

C.S. Sousa^{1,2}, B.A. Barros³, D. Barh⁴, P. Ghosh⁵, V. Azevedo², E.G. Barros^{6,7} and M.A. Moreira¹

¹Departamento de Bioquímica e Biologia Molecular - BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, Brasil

²Laboratório de Genética Celular e Molecular, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

³Empresa Brasileira de Pesquisa Agropecuária, Sete Lagoas, MG, Brasil

⁴Institute of Integrative Omics and Applied Biotechnology, Medinipur, WB, India

⁵Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA

⁶Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, Brasil

⁷Universidade Católica de Brasília, Brasília, DF, Brasil

Corresponding author: C.S. Sousa

E-mail: cassissousa@gmail.com

Genet. Mol. Res. 15 (3): gmr.15038974

Received July 14, 2016

Accepted August 9, 2016

Published August 26, 2016

DOI <http://dx.doi.org/10.4238/gmr.15038974>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. The enzymes 1,2-diacylglycerol cholinephosphotransferase (CPT) and lysophosphatidylcholine acyltransferase (LPCAT) are important in lipid metabolism in soybean seeds. Thus, understand-

ing the genes that encode these enzymes may enable their modification and aid the improvement of soybean oil quality. In soybean, the genes encoding these enzymes have not been completely described; therefore, this study aimed to identify, characterize, and analyze the *in silico* expression of these genes in soybean. We identified two gene models encoding CPT and two gene models encoding LPCAT, one of which presented an alternative transcript. The sequences were positioned on the physical map of soybean and the promoter regions were analyzed. Cis-elements responsible for seed-specific expression and responses to biotic and abiotic stresses were identified. Virtual expression analysis of the gene models for CPT and LPCAT indicated that these genes are expressed under different stress conditions, in somatic embryos during differentiation, in immature seeds, root tissues, and calli. Putative amino acid sequences revealed the presence of transmembrane domains, and analysis of the cellular localization of these enzymes revealed they are located in the endoplasmic reticulum.

Key words: Lipid metabolism enzymes; Gene models; Cis-element identification; Virtual expression; Putative amino acid sequences; Cellular localization

INTRODUCTION

The identification of genes within the soybean genome that govern metabolic steps involved in triacylglycerol (TAG) biosynthesis represents an important step in the modification of oil content and quality of soybean seeds. The availability of the soybean genome sequence (<http://www.phytozome.net/soybean>) has greatly accelerated the identification, isolation, and functional analysis of such genes (Schmutz et al., 2010).

Due to the wide use of seeds as a source of TAG for human and animal consumption, the synthesis of these compounds has attracted a lot of interest (Li et al., 2010). Among the enzymes involved in lipid biosynthesis in soybean seeds, two have been highlighted as being important for TAG biosynthesis, namely, 1,2-diacylglycerol cholinephosphotransferase (CPT, EC 2.7.8.2) and lysophosphatidylcholine acyltransferase (LPCAT, EC 2.3.1.23) (Li et al., 2010).

In seeds, TAG biosynthesis occurs in the membrane of the endoplasmic reticulum (ER). TAG accumulates in lipid bodies, which in turn are generated by budding from the ER membrane (Li et al., 2010). During TAG biosynthesis in oilseeds, acyl groups are sequentially added at positions 1 and 2 of glycerol 3-phosphate by the enzyme glycerol 3-phosphate acyltransferase, producing phosphatidic acid (PA). PA is then converted to diacylglycerol (DAG) by PA phosphatase. DAG represents a branching point leading to the subsequent biosynthesis of membrane or storage lipids. To produce membrane lipids, the enzyme CPT reversibly converts DAG into phosphatidylcholine (PC). PC is a substrate for polyunsaturated fatty acid biosynthesis, which can replenish the pool of acyl-CoA through the action of cytoplasmic LPCAT. DAG can also be converted to TAG by acylation at position 3 of the glycerol molecule in a reaction catalyzed by diacylglycerol acyltransferase (EC 2.3.1.20) (Li et al., 2010).

CPT is a key enzyme in the metabolism of oil in soybean seeds as it catalyzes both the synthesis of PC from DAG as well as the reverse reaction, which permits polyunsaturated fatty

acids to be incorporated into TAGs (Vogel and Browse, 1996).

LPCAT is another important enzyme involved in lipid metabolism in soybean seeds, which acts to replenish the pool of acyl-CoA polyunsaturated fatty acids in the cytoplasm. This enzyme catalyzes the acyl-CoA-dependent acylation of lysophosphatidylcholine (LPC) to produce PC and CoA. LPCAT activity affects the incorporation of fatty acids into the sn-2 position of PC where polyunsaturated fatty acids are formed; therefore, it indirectly influences the composition of seed TAGs (Fukurawa-Stoffer et al., 2003). The exchange of acyl groups between CoA and acyl-phosphatidylcholine is important for regulating the quality of polyunsaturated fatty acids in the acyl-CoA pool for oil synthesis in developing seeds (Lanna et al., 2005).

There is little information in the literature on the genes encoding these enzymes and on how CPT and LPCAT contribute to the accumulation of TAG in oilseeds. Thus, the objective of this study was to identify, characterize, and analyze the *in silico* expression of genes encoding CPT and LPCAT in soybean.

MATERIAL AND METHODS

Identification and *in silico* characterization

Genes encoding the enzymes CPT and LPCAT were identified using the BLASTn algorithm (Altschul et al., 1997), with an E-value cutoff of $1e^{-100}$, against the soybean genome (Glyma 1.01) (<http://www.phytozome.net>). For CPT, a sequence available in GenBank (accession No. U12735.1) was used, while for LPCAT, a contig made with soybean expressed sequence tags (ESTs), produced by Lima (2005) was used because the complete soybean LPCAT sequence was not available.

SignalScan software (Prestridge, 1991) and the PLACE database (Higo et al., 1999) (<http://www.dna.affrc.go.jp/PLACE>) were used to identify the regulatory cis-elements in the promoter region. A 1500-bp region upstream of the putative translation start site (+1) of the genes encoding CPT and LPCAT was used.

Expression of the genes encoding both enzymes was evaluated *in silico*. For this, the transcript of the predicted sequence of each gene was used to search the soybean EST bank on NCBI by means of BLASTn, which consisted of 394,070 sequences. The E-value cutoff was $1e^{-100}$. Virtual expression analysis from RNASeq data was also performed (<http://www.soybase.org/soyseq>). This analysis can be used to verify tissue-specific expression and also allows the mRNA levels of each analyzed gene to be assessed quantitatively. Each read (about 36 bp) mapped in a given gene model represents a molecule of mRNA under the conditions studied (Severin et al., 2010).

Peptide sequences were aligned using the ClustalW software (<http://www.ebi.ac.uk/Tools/clustalw2>). The conserved domains in protein sequences corresponding to genes of interest were identified with InterPro and InterProScan DB (www.ebi.ac.uk/interpro). Transmembrane helices were predicted with TMHMM Server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). Signal peptides were identified using the software SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) and the probable subcellular location was determined by the prediction software PREDOTAR (<http://urgi.versailles.inra.fr/predotar/predotar.html>), TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>) and WoLF PSORT (<http://wolfpsort.org/>). All identified genes were positioned and anchored in the physical map based on the consensus sequence of the

soybean genome available in the SoyBase database (<http://soybeanphysicalmap.org/index.php>) (Grant et al., 2010).

RESULTS AND DISCUSSION

In silico identification and analysis of nucleotide and amino acid sequences

To identify genes encoding the enzymes CPT and LPCAT, the nucleotide sequences were analyzed using BLASTn against the predicted soybean genome (Glyma1.01). This analysis revealed the existence of two gene models that potentially encode CPT (Glyma12g08720 and Glyma02g14210). These two gene models are 97% identical to each other. Similarly, two other gene models were detected that may potentially encode LPCAT, Glyma17g36670 and Glyma14g08400, showing 96% identity to each other. Glyma14g08400 has an alternative transcript (Glyma14g08400.2), which differs from the primary transcript (Glyma14g08400.1) only at the 5'UTR.

The putative amino acid sequences of both enzymes were then aligned. For CPT, the proteins showed 86% similarity between Glyma12g08720 and Glyma02g14210. The alignment is shown in Figure 1. For LPCAT, protein alignment revealed 95% homology between both Glyma17g36670 and Glyma14g08400.1, and between Glyma17g36670 and Glyma14g08400.2. There was 100% homology between Glyma14g08400.1 and Glyma14g08400.2. The result of this alignment is shown in Figure 2.

```

Glyma12g08720      MGYIGTHGVAALHRYKYSGVDHSYVAKYVLQPFWSRFVNFPLWMPSEYTNLYLFLFIL 60
Glyma02g14210     MGYIGAHGVAALHRYKYSGVDHSYVAKYVLQPFWSRFVNFPLWMP----- 46
*****

Glyma12g08720      ISWHSSNLCSFNMITLMGFMLLLSALLGYIYSPQLDTAPPRVWFHAGLLLLFLYQTFDA 120
Glyma02g14210     -----PNMITLMGFMLLVSALLGYIYSPQLDTPPRVWFHAGLLLLFLYQTFDA 96
*****

Glyma12g08720      VDGKQARRTNSSSPLGELFDHGCDALACTFEALAFGSTAMCGRITFWWWLISAITFYGAT 180
Glyma02g14210     VDGKQARRTNSSSPLGELFDHGCDALACTFEALAFGSTAMCGRNTFWWWLISAITFYGAT 156
*****

Glyma12g08720      WEHYFTNIIILPVIINGPTEGLMIYICHFFTAIVGAENWVQQFGKSLPFLNWLPLYLGG-- 238
Glyma02g14210     WEHYFTNIIILPVIINGPTEGLMIYICHFFTAIVGAENWVQQFGKSLPFLNWLPLYLAANV 216
*****

Glyma12g08720      ----IPTFKAILCLM----IAFGVPIPIV-----CNVSNVYKVV 269
Glyma02g14210     KLVDIPNMDLVYCLKRVHPLLRYICSLMSGMTIFFSQIFCCVLQYVLYFNVCVSNVYKVV 276
*.:. : * : : . :

Glyma12g08720      KGKNGSMPLALAMLYPFVVLVGGVLVWDYLSPSDIMGKYPHLVWIGTGLTFGYLVGRMIL 329
Glyma02g14210     KAKNGSMPLALAMLYPFVVLVGGVLVWDYLSPLDIMGRYPHLVWIGTGLTFGYLVGRMIL 336
*_*

Glyma12g08720      AHLCDPEPKLKTGMCSLMFLPLAIAIVLASRLNDGVPLVDERLVLLGYCAFVILYLHF 389
Glyma02g14210     AHLCDPEPKLKTGMCSLMFLPLAIAIVLASRLNDGVPLVDERLVLLGYCAFVILYLHF 396
*****

Glyma12g08720      ATSVIHEITNALGIYCFRITRKEA 413
Glyma02g14210     ATSVIHEITNALGIYCFRITRKEA 420
*****

```

Figure 1. Alignment of predicted amino acid sequences of the enzyme 1,2-diacylglycerol cholinephosphotransferase (Glyma02g14210 and Glyma12g08720) encoded in the soybean genome.

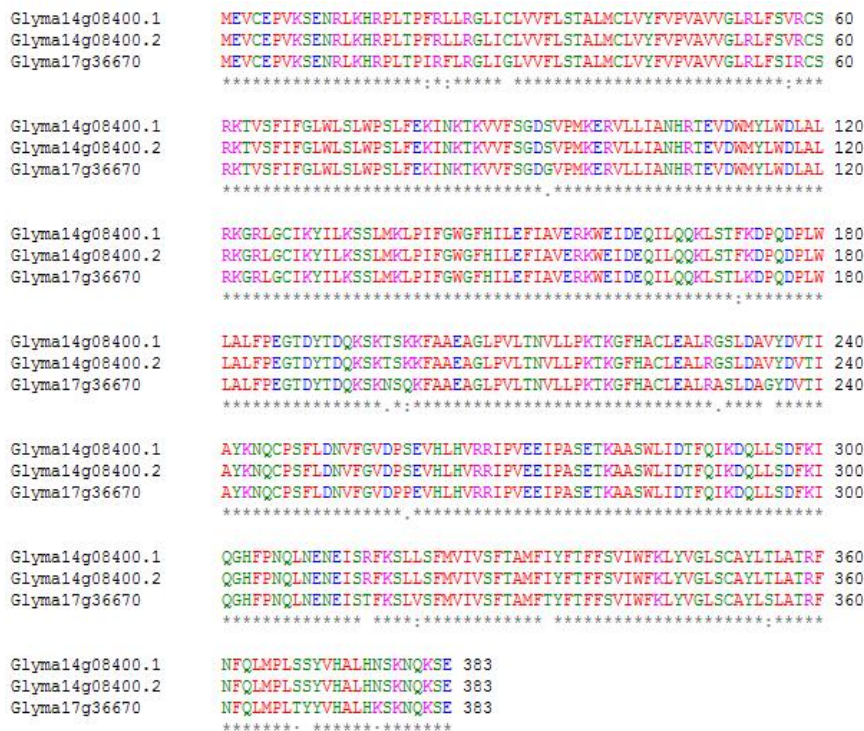


Figure 2. Alignment of the predicted amino acid sequences of the enzyme lysophosphatidylcholine acyltransferase (Glyma14g08400 and Glyma17g36670) encoded in the soybean genome.

The identified gene models were positioned in the genome using a map based on the consensus sequence of the soybean genome obtained from the USDA SoyBase database. Each gene was mapped on a specific chromosome 1) on chromosome 12 (LG H): Glyma12g08720 (CPT), flanked by microsatellites Satt 442 (43.4 cM) and Sat_034 (55.7 cM); 2) on chromosome 2 (LG D1b): Glyma02g14210 (CPT), flanked by the microsatellites Satt 634 (55.6 cM) and Sat_034 (61.4 cM); 3) on chromosome 17 (LG D2): Glyma17g36670 (LPCAT), flanked by microsatellites Sat_086 (104.8 cM) and Satt 386 (110.4 cM); and 4) on chromosome 14 (LG B2): Glyma14g08400 (LPCAT), flanked by microsatellites Sat_086 (104.8 cM) and Satt 386 (110.4 cM). The chromosome, LG, position, and the microsatellite markers that flank each gene model are illustrated in Figure 3.

Analysis of gene expression

To characterize the promoter region of the identified gene models, the 1500-bp sequence upstream of the translation start site (+1) was analyzed in order to evaluate the presence of known regulatory motifs.

The promoter regions of the gene models analyzed contained elements responsible for seed-specific expression, the expression of storage proteins, and responses to biotic and abiotic stress. Table 1 lists the cis-elements found within the promoter regions of the four gene models that potentially encode CPT and LPCAT.

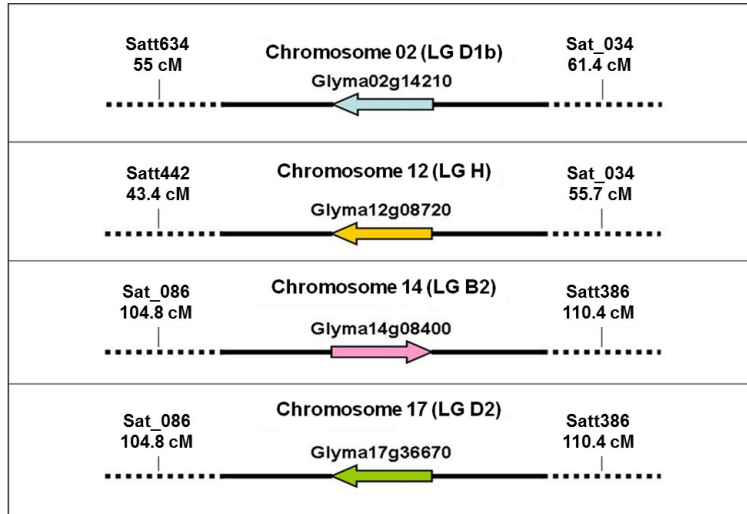


Figure 3. Positioning of gene models encoding 1,2-diacylglycerol cholinephosphotransferase (CPT) (Glyma02g14210 and Glyma12g08720) and lysophosphatidylcholine acyltransferase (LPCAT) (Glyma14g08400 and Glyma17g36670) on the soybean physical map.

Table 1. Common cis-elements found within the promoter regions of gene models that encode 1,2-diacylglycerol cholinephosphotransferase (CPT) and lysophosphatidylcholine acyltransferase (LPCAT) in soybean: 1) Glyma12g08720; 2) Glyma02g14210; 3) Glyma17g36670; 4) Glyma14g08400.1; 5) Glyma14g08400.2.

| PLACE ID and Motif | Position in relation to start codon | Description |
|------------------------|--|---|
| ACGTATERD1 (ACGT) | 1) -74; -779; -1007 2) -1229; -1285 3) -1352 4) -105; -156 5) -225; -1281; -1410 | Sequence required for response to water stress |
| CAAT BOX1 (CAAT) | 1) -126; -136; -245; -335; -687; -750; -785; -896; -1204; -1259; -1435 2) -41; -219; -228; -267; -304; -408; -474; -597; -701; -705; -957; -1004 3) -80; -118; -153; -171; -324; -369; -663; -736; -786; -835; -852; -941; -1338; -1438 4) -174; -255; -400; -605; -1031; -1046; -1109; -1156; -1207; -1217; -1377; -1482 5) -533; -553; -837; -1116; -1216; -1223; -1406; -1417; -1941; -1618 | Seed-specific activity |
| DOFCOREZM (AAAAG) | 1) -94; -188; -891 2) -4; -205; -718; -829; -880; -993; -1076; -1250; -1308; -1363; -1522 3) -234; -534; -575; -597; -609; -902; -982; -1410; -1487; -1550; -1586 4) -304; -357; -393; -1052; -1105; -1143; -1512 5) -282; -521; -667; -746; -1062; -1105; -1296; -1319; -1341; -1395; -1646; -1661; -1668 | Binding site of DOF transcription factors induced by salicylic acid |
| EBOXBNNAPA (CANNTG) | 1) -1; -67; -1414 2) -666; -1254; -1265 3) -80; -258; -443 4) -277; -521; -731; -911; -1161; -1408 5) -169 | Essential for the expression of storage proteins |
| GT1CONSENSUS (GRWAAW) | 1) -307; -798; -817; -834; -875; -945 2) -1028; -1083; -1156 3) -74; -85; -273; -318; -531; -606; -1382 4) -572; -682; -771; -779; -835; -1140 5) -30; -210; -518; -784; -879; -890; -1179; -1292; -1303; -1664 | Activity regulated by light |
| INRNTPSADB (YTCANTTY) | 1) -1573; -920; -980; -1465; -1470 2) -1569 3) -367; -828; -1436 4) -253; -398; -1336 5) -1431; -835; -1611 | Activity regulated by light |
| OSE2ROOTNODULE (CTCTT) | 1) -928 2) -53; -186; -905; -1383; -1402; -1550; -1592 3) -710; -1043 4) -430 5) -814; -987; -1035; -1435; -1605 | Sequence found in promoters activated in infected cells of root nodules |
| TATA BOX (TATAAAT) | 1) -342; -763 2) -736 3) -290; -341; -403; -549 4) -410; -453 5) -270 | Positionally important element for the accuracy of transcription initiation |
| WRK4710S (TGAC) | 1) -133; -149; -743; -782; -1005; -1119 2) -1298; -1513 3) -156; -491; -540; -732; -739; -938; -1535 4) -339 5) -811; -1350; -1487; -1494 | Transcriptional repressor of the gibberellin signaling pathway |

DOF = DNA-binding with one finger proteins.

Similar results from studies investigating the *in silico* expression of genes in soybean were reported by Soares (2011) and de Almeida Barros et al. (2012). de Almeida Barros et al. (2012) examined the expression of the multigene family encoding Bowman-Birk type soybean inhibitors, while Soares (2011) examined the expression of genes involved in the biosynthesis of raffinose oligosaccharides in soybean.

Vandepoele et al. (2009) performed similar *in silico* analyses in *Arabidopsis*. In their study, elements responsible for the circadian expression of genes were identified. They also identified elements that are responsive to biotic stress as well as differentially expressed genes involved in processes such as protein biosynthesis and photosynthesis. Evolutionary analysis suggested that regulatory modules are conserved between species belonging to different plant families (Vandepoele et al., 2009).

In the present study, we detected cis-elements that respond to biotic and abiotic stress. According to Upchurch (2008), plants can respond to these stresses by remodeling and adjusting plasma membrane fluidity. Fluidity of the lipid membrane is adjusted through alterations in the levels of unsaturated fatty acids, thereby maintaining the proper environment for the function of integral membrane proteins of the photosynthetic apparatus during stress.

This pattern of differential gene expression is due to transcriptional regulation, and for this reason, many studies have been carried out to improve our understanding of gene expression in soybean under different conditions, including excess water (Komatsu et al., 2009), soybean cyst nematode infection (Yeckel et al., 2007), circadian expression, and flowering controlled by photoperiod (Quecini et al., 2007).

Table 1 shows the presence of binding sites for DNA-binding with one finger (DOF) proteins in promoter regions of the gene models encoding CPT and LPCAT. The DOF proteins are involved in the response to salicylic acid. According to Kang and Singh (2000), DOF proteins, which are induced by salicylic acid, interact and stimulate the activity of DNA-binding bZIP proteins that are also responsive to stress. Table 1 also shows the presence of a transcriptional repressor cis-element of the gibberellin signaling pathway. This cis-element contains a WRKY binding site that, according to Eulgem et al. (2000), constitutes a family of transcription factors that are unique to plants. The DNA-binding activity of members of this family is induced by pathogens and wounding.

We also carried out an *in silico* analysis to determine the conditions and tissues in which the plant gene models for CPT and LPCAT were expressed. To achieve this, the predicted sequence of each transcript was used to search a soybean EST database and to analyze RNA Seq data.

The results of EST and RNA Seq data analysis were consistent with the cis-elements presented in Table 1. The results of the EST search suggested that the genes of interest are expressed under conditions of water or salinity stress, during infection with pathogens, in response to salicylic acid signaling, during somatic embryo differentiation, in immature seeds, in root tissues, and in soybean calli grown in the dark.

The results obtained from RNA Seq analyses (Table 2) demonstrated that the gene models for CPT and LPCAT are differentially expressed in different tissues and at different stages of development. Furthermore, expression of the gene model Glyma17g36670 (LPCAT) appears to be specific to young leaves, flowers, 1-cm pods, and roots, and therefore, it is not expressed in seeds or more developed pods. The gene model Glyma14g08400 (LPCAT) was expressed in different tissues of soybean, including seeds, and was also shown to be more highly expressed than that by the gene model Glyma17g36670. Gene models Glyma02g14210 and Glyma12g08720 (CPT) were shown to be expressed in different tissues of soybean, including seeds.

Table 2. *In silico* gene expression analysis using RNA-seq data of gene models encoding the enzymes CPT (Glyma02g14210 and Glyma12g08720) and LPCAT (Glyma14g08400 and Glyma17g36670) in different soybean tissues.

| | Glyma02g14210 | Glyma12g08720 | Glyma14g08400 | Glyma17g36670 |
|------------|---------------|---------------|---------------|---------------|
| Young leaf | 4 | 10 | 7 | 1 |
| Flower | 4 | 14 | 29 | 3 |
| Pod 1 cm | 3 | 11 | 1 | 1 |
| Pod 10 DAF | 3 | 8 | 1 | 0 |
| Pod 14 DAF | 4 | 6 | 1 | 0 |
| Pod 10 DAF | 2 | 5 | 2 | 0 |
| Pod 14 DAF | 3 | 6 | 1 | 0 |
| Pod 21 DAF | 4 | 6 | 1 | 0 |
| Pod 25 DAF | 2 | 4 | 1 | 0 |
| Pod 28 DAF | 1 | 2 | 0 | 0 |
| Pod 35 DAF | 1 | 5 | 2 | 0 |
| Pod 42 DAF | 0 | 2 | 2 | 0 |
| Root | 5 | 9 | 8 | 1 |
| Nodule | 2 | 4 | 2 | 0 |

Numbers represent mRNAs mapped to the respective gene models. DAF = days after flowering.

The pattern of differential expression may be due to the cis elements present in the promoter being responsible for the induction of tissue-specific expression, which is consistent with the results shown in Table 1 and with those obtained by EST analysis. An example of a cis-element responsible for tissue-specificity is the CAAT box1 present in the promoter regions of the genes encoding CPT and LPCAT (Table 1).

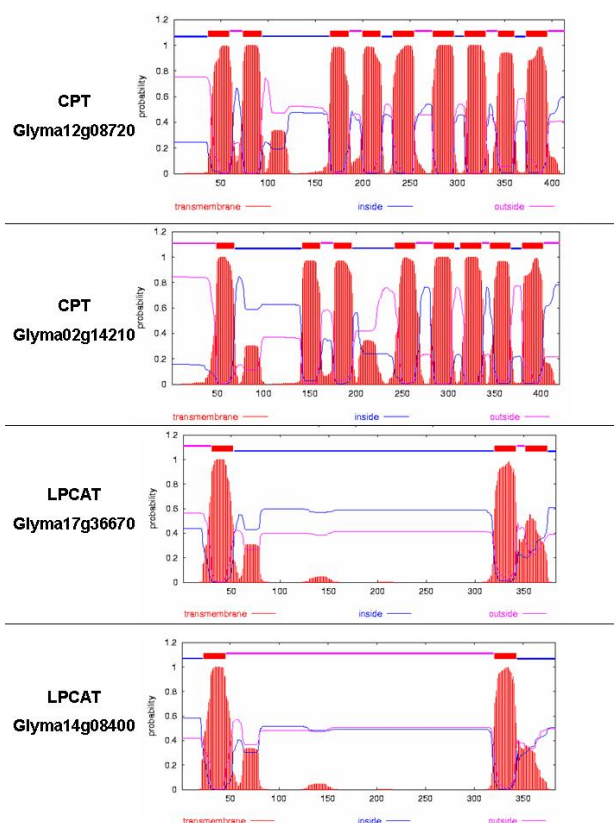
To determine whether the primary transcript and the alternative transcript of the gene model Glyma14g08400 are differentially expressed, specific sequences of the 5'UTR region of Glyma14g08400.1 (54 bp) and Glyma14g08400.2 (256 bp) were used as probes to screen the cDNA and RNA seq databases. The results indicated that Glyma14g08400.1 (54 bp) is expressed under conditions of water stress, salinity, *Pseudomonas* infection in etiolated seedlings, and in soybean calli when grown in the dark, as well as in the apical meristem, flowers, and green seeds. The alternative transcript Glyma14g08400.2 is expressed under the same conditions; however, this transcript is also detected in soybean plants infected with cyst nematode. This suggests that expression of the alternative transcript may be related to the response to cyst nematode infection, and may represent an adaptation to pathogen attack.

Analysis of amino acid sequences, identification of domains, and subcellular localization

While searching for conserved domains in this class of proteins, the predicted peptide sequences of CPT and LPCAT were used to identify the conserved regions. As the domains present in a protein/enzyme are closely related to function, it is important to identify and characterize these regions in order to determine protein structure and correlate this with enzyme function. The results are described in the text and illustrated in Table 3 and Figure 4.

Table 3. Domains identified in the enzymes CPT and LPCAT through the analysis of amino acid sequences.

| Enzyme | Domains |
|---|--|
| CPT (Glyma12g08720) (Glyma02g14210) | 1) CDP-alcohol phosphatidyltransferase |
| | 2) Choline/ethanolamine phosphotransferase |
| | 3) Signal peptide and transmembrane regions (nine transmembrane regions in the protein encoded by Glyma12g08720 and eight transmembrane regions in the protein encoded by Glyma02g14210) |
| LPCAT (Glyma17g36670) (Glyma14g08400) | 1) Phospholipid/glycerol acyltransferase |
| | 2) Acyltransferase regions, signal peptide, transmembrane regions (three transmembrane regions in the protein encoded by Glyma17g36670 and two transmembrane regions in the protein encoded by Glyma14g08400) and a region with no description |

**Figure 4.** Prediction of transmembrane helices in 1,2-diacylglycerol cholinephosphotransferase (CPT) and lysophosphatidylcholine acyltransferase (LPCAT). The peptide sequence was used in the TMHMM database.

The two CPT amino acid sequences used in the present analyses generated comparable results, and were found to contain a CDP-alcohol phosphatidyltransferase domain, a choline/ethanolamine phosphotransferase domain, a signal peptide, and transmembrane regions. Between the sequences, the only significant difference was that the peptide sequence obtained from Glyma12g08720 indicated the presence of nine transmembrane regions while that for Glyma02g14210 indicated the presence of eight transmembrane regions (Figure 4). The same result was obtained when we used a database to predict transmembrane helices in proteins. Since transmembrane domains are uniformly distributed along the entire extension of the polypeptide chain, this finding suggests that this enzyme is an integral membrane protein

(Figure 4). According to the database used to predict subcellular localization, CPT is located in the membrane of the ER, consistent with this suggestion. According to Schmutz et al. (2010), soybean CPT is involved in the synthesis of membrane lipids of the ER, which is consistent with the results of the subcellular localization analysis performed here.

The LPCAT sequences used revealed a phospholipid/glycerol acyltransferase domain, a domain containing acyltransferase regions, a signal peptide, transmembrane regions, and a region with no description (Table 2). According to the database used to predict transmembrane helices in proteins, the peptide sequence obtained from Glyma17g36670 contains three transmembrane helices and the peptide sequence obtained from Glyma14g08400 contains two transmembrane helices. In both peptide sequences, transmembrane domains are located at the terminals, indicating that they are membrane-anchored proteins. Furthermore, the results of these analyses show that the protein with three helices transmembrane is anchored in the membrane and faces towards the ER, whilst the protein with two helices is anchored in the membrane and is directed out of the ER (Figure 4). Akermoun et al. (2000) investigated the role of LPCAT in the import of extraplasmidial precursors of lipids, which are required for lipid synthesis. Those authors found that LPCAT is membrane associated, because solubilization with detergent was necessary to detect enzymatic activity in the solubilized protein fraction.

The results obtained in the present study from *in silico* analysis of the promoter regions of gene models encoding the enzymes CPT and LPCAT, and expression analysis of these gene models, revealed that these enzymes are expressed in several tissues, at different stages of development, and under different biotic and abiotic stress conditions in soybean. This indicates that these enzymes have an important role in lipid metabolism.

The domains identified and the subcellular localization of the enzymes suggested their involvement in fatty acid and triacylglycerol biosynthesis, as they are likely to be located within the ER membrane, which is the site of TAG synthesis, and possess transferase activity domains.

As these genes have been studied for the improvement of soybean oil quality (Lanna et al., 2005; Lima, 2005; Martins, 2007; Sousa, 2007), the presence of at least one seed-specific gene model was expected. However, this was not observed, and the gene models were shown to be expressed in different tissues. It may be possible to use such genes in order to genetically manipulate lipid metabolism, for example by gene silencing, if seed-specific promoters are used to control the expression of the transgene.

In summary, the results of the present study confirmed that the soybean genome possesses two loci encoding the enzyme CPT and two loci encoding the enzyme LPCAT. In addition to being important for maintaining the structure of the plasma membrane and for the composition of fatty acids in oilseeds, these enzymes are also involved in protection against biotic and abiotic stress, as their promoters were found to possess response elements associated with these conditions. In addition to the primary transcripts identified by the gene models, one gene model encoding LPCAT presents an alternative transcript that is expressed when soybean is infected by cyst nematode; hence, this enzyme may be related to adaptation against this pathogen.

The position of these enzymes as integral ER or anchored proteins, the presence of the phosphotransferase domain of CPT, and the acyltransferase domain in LPCAT are consistent with their role in triacylglycerol biosynthesis. The CPT and LPCAT genes represent potential targets for genetic improvement of the quality of soybean oil through genetic manipulation if seed-specific promoters are used.

In conclusion, in this study, we characterized the genes encoding two important enzymes involved in the lipid biosynthesis pathway in soybean seeds, namely CPT and

LPCAT. Knowledge of these genes is essential for genetic improvement of soybean seed with the aim of producing oil with greater oxidative stability, which is of considerable importance for both feeding and biofuel production.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors wish to acknowledge Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), of Universidade Federal de Viçosa, Minas Gerais, Brasil, and the financial support of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

- Akermoun M, Testet E, Cassagne C and Bessoule JJ (2000). Solubilization of the plastidial lysophosphatidylcholine acyltransferase from *Allium porrum* leaves: towards plants devoid of eukaryotic plastid lipids? *Biochem. Soc. Trans.* 28: 713-715. <http://dx.doi.org/10.1042/bst0280713>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402. <http://dx.doi.org/10.1093/nar/25.17.3389>
- de Almeida Barros B, da Silva WG, Moreira MA and de Barros EG (2012). *In silico* characterization and expression analysis of the multigene family encoding the Bowman-Birk protease inhibitor in soybean. *Mol. Biol. Rep.* 39: 327-334. <http://dx.doi.org/10.1007/s11033-011-0742-1>
- Eulgem T, Rushton PJ, Robatzek S and Somssich IE (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5: 199-206. [http://dx.doi.org/10.1016/S1360-1385\(00\)01600-9](http://dx.doi.org/10.1016/S1360-1385(00)01600-9)
- Furukawa-Stoffer TL, Boyle RM, Thomson AL, Sarna MA, et al. (2003). Properties of lysophosphatidylcholine acyltransferase from *Brassica napus* cultures. *Lipids* 38: 651-656. <http://dx.doi.org/10.1007/s11745-003-1110-0>
- Grant D, Nelson RT, Cannon SB and Shoemaker RC (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38: D843-D846. <http://dx.doi.org/10.1093/nar/gkp798>
- Higo K, Ugawa Y, Iwamoto M and Korenaga T (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27: 297-300. <http://dx.doi.org/10.1093/nar/27.1.297>
- Kang HG and Singh KB (2000). Characterization of salicylic acid-responsive, *arabidopsis* Dof domain proteins: overexpression of OBP3 leads to growth defects. *Plant J.* 21: 329-339. <http://dx.doi.org/10.1046/j.1365-3113x.2000.00678.x>
- Komatsu S, Yamamoto R, Nanjo Y, Mikami Y, et al. (2009). A comprehensive analysis of the soybean genes and proteins expressed under flooding stress using transcriptome and proteome techniques. *J. Proteome Res.* 8: 4766-4778. <http://dx.doi.org/10.1021/pr900460x>
- Lanna AC, José IC, Oliveira MGA, Barros EG, et al. (2005). Effect of temperature on polyunsaturated fatty acid accumulation in soybean seeds. *Braz. J. Plant Physiol.* 17: 213-222, 2205. <http://dx.doi.org/10.1590/S1677-04202005000200004>
- Li R, Yu K and Hildebrand DF (2010). DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. *Lipids* 45: 145-157. <http://dx.doi.org/10.1007/s11745-010-3385-4>
- Lima ABP (2005). Construção de cassete para a co-supressão do gene da oleoil dessaturase e transformação genética de embriões somáticos de soja. Doctoral Dissertation, Universidade Federal de Viçosa.
- Martins PK (2007). Transformação genética em nós cotiledonares de soja visando a co-supressão do gene da lisofosfatidilcolina aciltransferase. Doctoral Dissertation, Universidade Federal de Viçosa.
- Prestridge DS (1991). SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. *Comput. Appl. Biosci.* 7: 203-206.
- Quecini V, Zucchi MI, Baldin J and Vello NA (2007). Identification of soybean genes involved in circadian clock mechanism and photoperiodic control of flowering time by *in silico* analyses. *J. Integr. Plant Biol.* 49: 1640-1653.

- <http://dx.doi.org/10.1111/j.1774-7909.2007.00567.x>
- Schmutz J, Cannon SB, Schlueter J, Ma J, et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178-183. <http://dx.doi.org/10.1038/nature08670>
- Severin AJ, Woody JL, Bolon YT, Joseph B, et al. (2010). RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10: 160. <http://dx.doi.org/10.1186/1471-2229-10-160>
- Soares APG (2011). Caracterização *in silico* de genes envolvidos na biossíntese de oligossacarídeos de rafinose em soja e construção de cassetes de silenciamento do gene da estaquiase sintase, via interferência por RNA. Doctoral Dissertation, Universidade Federal de Viçosa, Viçosa.
- Sousa CS (2007). Construção de cassetes de expressão para silenciamento de genes da biossíntese de ácidos graxos em soja via RNA de interferência. Master's Thesis, Universidade Federal de Viçosa, Viçosa.
- Upchurch RG (2008). Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnol. Lett.* 30: 967-977. <http://dx.doi.org/10.1007/s10529-008-9639-z>
- Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, et al. (2009). Unraveling transcriptional control in *Arabidopsis* using cis-regulatory elements and coexpression networks. *Plant Physiol.* 150: 535-546. <http://dx.doi.org/10.1104/pp.109.136028>
- Vogel G and Browse J (1996). Cholinephosphotransferase and diacylglycerol acyl transferase - substrate specificities at a key branch point in seed lipid metabolism. *Plant Physiol.* 110: 923-931.
- Yeckel G, Ithal N, Nettleton D, Recknor J, et al. (2007). Characterization of soybean genes involved in soybean cyst nematode (SCN) resistance. In: 2007 Undergraduate Research and Creative Achievements Forum (MU), 163.