

In Silicon Cloning and Analysis of a LACS Gene from *Glycine Max* (L.)

Lili Yu

Institute of Life Sciences, Jiangsu University

Zhenjiang 212013, China

E-mail: merrys222@126.com

Xiaoli Tan (Corresponding author)

Institute of Life Sciences, Jiangsu University

Zhenjiang 212013, China

Tel: 86-511-8291-4326 E-mail: xltan@ujs.edu.cn

Weiwei Yuan

Institute of Life Sciences, Jiangsu University

Zhenjiang 212013, China

E-mail: jsuni020219@163.com

Fuge Zhu

Institute of Life Sciences, Jiangsu University

Zhenjiang 212013, China

E-mail: zfg12@163.com

The research is financed by the National Basic Research Program (973 Program) of China (No.2006CB101600) and Talent Fund of Jiangsu University (05JDG003).

Abstract

Long chain acyl-coenzyme A synthetases (LACSS) activate free fatty acids to acyl-CoA thioesters, and play important roles in the biosynthesis and degradation of lipids. In this study, a *Glycine max*(L.) LACS gene, designated as *GmLACS*, has been isolated through *in silicon* cloning. The gene is 2,219 bp with an open reading frame (ORF) of 1,989 bp, which encodes a LACS with 662 amino acid residues, with the isoelectric point of 6.42 and the calculated molecular mass of 65.6 kDa. Sequence analysis showed that *GmLACS* possessed typical domains of LACSSs. Real-time quantitative PCR data analysis suggested that *GmLACS* was highly expressed in leaves and young pods.

Keywords: *Glycine max*(L.), *GmLACS*, *Silicon* cloning, Real-time quantitative PCR

1. Introduction

Acyl-coenzyme A synthetases (ACSSs) are generally classified by their specificities for fatty acids of varying chain length. The Commission on Enzymes of the International Union of Biochemistry has classified these enzymes as acetyl-CoA synthetase (EC 6.2.1.1), medium-chain ACS (EC 6.2.1.2) and long-chain ACS (EC 6.2.1.3). A unifying feature of all acyl-CoA synthetases is the presence of an “AMP-binding domain signature” (PROSITE PS00455)(Watkins, P.A., 1997, pp.55-83). In all of LACSSs, long chain acyl-coenzyme A synthetases (LACSSs) play a key role in metabolism of fatty acid (Shrago, E., 2000, pp.290-293). LACS esterifies free fatty acids to acyl-CoAs, a key activation step that is necessary for the utilization of fatty acids by most lipid metabolic enzymes(Bradford, M.M., 1976, pp.248-254). LACS catalyzes the formation of acyl-CoAs by a two-step mechanism. In the first step, the free

fatty acid is converted to an acyl-AMP intermediate, called an adenylate, through the pyrophosphorolysis of ATP. The activated carbonyl carbon of the adenylate is then coupled to the thiol group of CoA, releasing AMP and final product, acyl-CoA (Groot, P.H., 1976, pp.75-126). Another special molecular characteristic is that eukaryotic LACSs contain a linker domain with the length about 30 to 40 amino acid residues(Shockley, J.M., 2002, pp.1710-1722). The linker domain does not exist in the other acyl-CoA synthetases. Though the detailed utility is unknown, the linker domain seems to be necessary for eukaryotic LACSs' function(Iijima, H., 1996, pp.186-190). The length of linker domain is also important for the activity of LACSs. Two LACS-like proteins At4g14070 and At3g23790 in *Arabidopsis thaliana* have high identity with AtLACSs, but they do not encode LACS activity. The abnormal length of their linker domain, about 70 amino acid residues, is a putative reason for their non-LACS function(Shockley, J.M., 2002, pp.1710-1722).

This important class of enzymes affect prominently in several fatty acid-derived metabolic pathways, including phospholipid, triacylglycerol, jasmonate biosynthesis and fatty acid β -oxidation. Oil accumulation in oilseeds has significant economic interest for food, feed, cosmetics and detergents etc., and triacylglycerol (TAG) is the main component of plant oil(Shen, B., 2006, pp.377-387). Fatty acyl-CoA thioesters and glycerol 3-phosphate are substrates for acyltransferases to synthesize TAG via Kennedy cycle. During the processes, LACSs play a pivotal role by providing fatty acyl-CoA and link fatty acid *de novo* synthesis and TAG assembly(Ohlrogge, J.B., 1997, pp.109-113). Another important role LACSs played is in fatty acids transport. This process has been studied in detail in bacteria, yeast (*Saccharomyces cerevisiae*), and mammalian cells. *Escherichia coli* contains a single LACS, encoded by *FadD* gene, which was proved to transport the fatty acids(Black, P.N., 1992, pp.25513-25520). LACS also initiates the process of fatty acid β -oxidation. In oilseeds, carbon reserves are stored as triacylglycerol (TAG). With the onset of germination, lipases release free fatty acids from the TAG molecules(Hills, M.J., 1986, pp.671-674; Lin, Y.H., 1986, pp.346-356). LACS activates the free fatty acids to acyl-CoAs that enter the β -oxidation pathway in the glyoxysomes of the germinating seedling.

In *Arabidopsis thaliana*, it has been established that nine *LACS* genes exist and were named *LACS1-9*. Nine LACSs could be classified into 3 distinct clades. *LACS1* is supposed to be involved in the syntheses of lipids and *LACS2* is supposed to be involved in the pathway of cutin synthesis(Schnurr, J., 2004, pp.629-642). *LACS6* and *LACS7* are localized in peroxisome, and both of them are involved in peroxisomal fatty acid β -oxidation(Fulda, M., 2004, pp.394-405). In *Ricinus communis*, three *LACS* genes have been cloned. *RcLACS2* is likely to be a peroxisomal ACS isoform. *RcLACS4* is supposed to be involved in the syntheses of lipids(He, X., 2007, pp.931-938). *Capsicum annuum* also has been found *GaLACS*. In this paper, we found a *LACS* gene from *Glycine max* through *in silicon* cloning, designated *GmLACS*. Sequence analysis indicated that the deduced protein possessed AMP-binding motifs and a linker domain.

2. Materials and methods

2.1 Bioinformatics analysis

Blast in NCBI(<http://www.ncbi.nlm.nih.gov/>) was used for genes searching in network. DNA and amino acid sequence manipulation was performed with EditSeq program of DNASTar 5.0 package. Domain prediction was done at ExPASy Proteomics Server(<http://au.expasy.org/>). Sequence alignments were carried out by GeneDoc. Phylogenetic tree was constructed using MEGA4.0 and TreeView was used for exhibition of phylogenetic tree.

2.2 Plant materials

Glycine max(L.)cv Willimas was utilized for expression analysis of *GmLACS* at different reproductive stages.

2.3 RNA extraction

RNA samples extracted from different soybean tissues including leaf, root, flower and pod. RNA samples were extracted by plant Trizol reagent (Invitrogen Biotech Co. Ltd., U.S.). All steps were carried out following the instruction of manufacturer. RNA samples were DNase treated with DNA-free(TaKara, Japan) according to the manufacturers directions. RNA concentrations were determined using spectrophotometer (GeneSpecIII, U.S.) at absorbance 260 nm. Aliquots of RNA (free of genomic DNA) were diluted to 50 ng μL^{-1} in RNase-free water and stored at -70°C until use. To verify RNA integrity, 500 ng of total RNA of each sample was examined on a 1% agarose gel following electrophoresis and staining with ethidium bromide.

2.4 RT-PCR assay of *GmLACS* expression profiles

2 μg total RNAs were used for the first strand cDNA synthesizing with M-MLV Reverse Transcriptase (Takara, Japan) according to the manufacture's protocol. Real-time reverse transcription-polymerase chain reaction(RT-PCR) was performed with the iCycler using the SYBR Green RT-PCR kit (Takara, Japan) to quantify *GmLACS* in different tissues. *GmLACS* gene specific primers *GmLACS* (5'-AGCTGGGTAAAGCGACTG -3' and 5'-CCTATTGAAGCACCATGCCATA -3') were designed in the non-conservative regions. Primers used for the housekeeping gene soybean actin (5'-GAGCTATGAATTGCCTGATGG-3' and

5'-CGTTTCATGAATTCCAGTAGC-3') were designed by Byfield (Byfield, G.E., 2006, pp.840–846) based on the GenBank accession number U60500(Monizde, S.M., 1996, pp.1198-1212). Real-time quantitative polymerase chain reactions with the *GmLACS* and actin primers produced the following amplicons: *GmLACS*, 138 base pairs (bp); and actin, 188 bp. A typical reaction done in duplicate, contained 10 uL of 2X SYBR Green PCR master mix, 250 nM each primer, and 1uL RT mix, followed by 95°C for 5 min and then 40 cycles of 15 s at 94°C, 30 s at the annealing temperature 56°C, and 30 s at 72°C. A melt curve analysis over a 10°C temperature gradient at 0.05°C s⁻¹ from 78 to 88°C was done after amplification to verify that a single product was produced in each reaction.

3. Results

3.1 In silicon cloning of *GmLACS*

Arabidopsis LACS1-9 (GeneBank Acc. No. AAM28868-AAM28876) were used as querys to blast *Glycine max* (L.) database, two *Glycine max* cDNAs(GeneBank Acc. No. AK245419 and AK245622) with high sequence identity were obtained. From this group, one cDNA(AK245419) which share 75% identity with *AtLACS4*, 74% identity with *AtLACS5* and 73% identity with *AtLACS3* was selected. It is a 2,219 bp long fragment. Sequence aligning with *GmLACS* sequence suggested that the whole fragment contained intact ORF of 1,989 bp, designated as *GmLACS*. Translated by DNASTar software, *GmLACS* was predicted to encode a protein of 662 amino acid residues (Fig. 1), with the theoretical pI of 7.11 and calculated molecular weight of 74.06 KDa.

3.2 Sequence analysis of *GmLACS* cDNA

Sequence alignment detected that the *GmLACS* ORF shared 74% sequence identity with *AtLACS5* (GeneBank Acc. No. AF503755) and the deduced amino acid sequence shared 74% sequence identity with AtLACS5 (GeneBank Acc. No. AAM28872). The deduced peptide sequence of *GmLACS* shared high identity with other LACSSs from plants which was 81%, 81%, 78%, and 75% sequence identity with RcLACS4 (GeneBank Acc. No. ACB30545), RcLACS1(GeneBank Acc. No. ABC02880), CaLACS (GeneBank Acc. No ACF17663), and AhLACS (GeneBank Acc. No. ACC91252), respectively. *GmLACS* shared much lower identity with LACSSs from mammal and microbe (data not shown).

Multiple sequence alignment of *GmLACS* and some other LACSSs from plants showed that there was some considerable conserved amino acid sequences existed in the form of blocks. Three blocks appeared among these proteins, and three AMP-binding motifs—I[MCV]TSG[TS][ST]GXPK, GYGXTE and GW[FL][HK]TG— orderly located in Block I- III. The conserved tyrosine residue at the position 481 on *GmLACS* was assumed to be involved in adenylate formation (Fig. 2. A). Multiple sequence alignment of central sequences of *GmLACS*, *AtLACS1*, *AtLACS2*, *RcLACS4* and *Arabidopsis* LACS-like protein *At4g14070* demonstrated that *GmLACS* contained a linker domain of 31 amino acid residues as well as other LACSSs (Fig. 2. B), and this length of the linker domain suggested it might encode LACS activity.

3.3 Molecular evolution analysis

A neighbor-joining phylogenetic tree was generated to exhibit the distances among *GmLACS* and other plant LACSSs(Fig. 3). Bootstrap analysis was performed for the reliability of phylogenetic tree. The phylogenetic tree demonstrated that the LACSSs derived from a common ancestor and diverged into two separate clades. *AtLACS3*, *AtLACS4*, *AtLACS5*, *AtLACS6*, *AtLACS7* and *AtLACS8* were in the first clade; *RcLACS4*, *GmLACS*, *CaLACS*, *AhLACS*, *AtLACS1*, *AtLACS2* and *AtLACS9* were in the second clade. The phylogenetic tree revealed that there was remarkable species specificity among LACSSs. The sequence homology among those proteins was probably coincident with their function or subcellular location. A phylogenetic tree was drawn based on the deduced amino acid sequence and other LACSSs (Fig. 3). In the phylogenetic tree, it showed that *GmLACS* had a higher homology with the second clade, especially with the *RcLACS4* in the second clade.

3.4 Analysis of *GmLACS* expression profile

To investigate the *GmLACS* expression profile in different tissues, total RNA from root, leaf, flower and pod tissues were used as templates to detect the transcription of *GmLACS* by real-time quantitative PCR. The actin gene was used as the internal control to ensure that the amount of RNA used is equal. The results from real-time PCR assay indicated that *GmLACS* genes were expressed in all tissues tested, but the patterns is quite different, stronger expression exhibited in leaves and young pods (Fig. 4). The most distinctive result was the high accumulation of *GmLACS* in young pod. Based on the high sequence similarity of *GmLACS* to *RcLACS4* which was supposed to be related with the syntheses of lipids, *GmLACS* was likely to be involved in the syntheses of lipids (Fig. 3).

4. Discussions

The long chain acyl-coenzyme A synthetases (LACSSs), such as *AtLACS1*, *AtLACS2*, *AtLACS6* and *AtLACS7*, play essential roles in the biosynthesis and degradation of lipids in model plant *Arabidopsis thaliana*(Shockley, J.M., 2002, pp.1710-1722; Fulda, M., 2004, pp.394-405; Schnurr, J., 2004, pp.629-642). But their functions in *Glycine max*(L.) are

not reported. In this study, we successfully identified a novel gene, named as *GmLACS*, which might be involved in lipids metabolism in soybean. We obtained a complete cDNA of *GmLACS* from *Glycine max*(L.) using *in silico* cloning. Sequence analysis indicated that *GmLACS* belonged to AMP-binding super-family and contained a linker domain of 31 amino acid residues as well as other LACSs, it is suggested *GmLACS* presumably encoded LACS activity. *GmLACS* was highly homologous to *RcLACS4*, a gene involved in the syntheses of lipids, in the second clade in phylogenetic tree. (He, X., 2007, pp.931–938). Real-time quantitative PCR analysis showed that the *GmLACS* was strongly expressed in leaves and young pods. This indicated that *GmLACS* may be involved in the syntheses of lipids in soybean seed development like *AtLACS1* or *AtLACS2* play roles in cutin biosynthesis. Based on the sequence similarity of *GmLACS* to *RcLACS4*, *GmLACS* is likely to be involved in the syntheses of lipids (Fig. 3). To identify its functional characterization, more evidences need to get.

References

- Black, P. N., DiRusso, C. C., Metzger, A. K., and Heimert, T. L. (1992). Cloning, sequencing, and expression of the fadD gene of *Escherichia coli* encoding acyl coenzyme A synthetase. *J Biol Chem*, 267(35), 25513-25520.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72, 248-254.
- Byfield, G. E., Xue, H., and Upchurch, R. G. (2006). Two genes from soybean encoding soluble Δ9 stearoyl-ACP desaturases. *Crop Science*, 46(2), 840–846.
- Fulda, M., Schnurr, J., Abbadi, A., Heinz, E., and Browse, J. (2004). Peroxisomal Acyl-CoA Synthetase Activity Is Essential for Seedling Development in *Arabidopsis thaliana*. *Plant Cell*, 16(2), 394-405.
- Groot, P. H., Scholte, H. R., and Hulsmann, W. C. (1976). Fatty acid activation: specificity, localization, and function. *Adv Lipid Res*, 14, 75-126.
- He, X., Chen, G. Q., and Kang, S. T. (2007). *Ricinus communis* Contains an Acyl-CoA Synthetase that Preferentially Activates Ricinoleate to Its CoA Thioester. *Lipids*, 42, 931–938.
- Hills, M. J., and Beevers, H. (1986). ATPase in Lipid Body Membranes of Castor Bean Endosperm. *Plant Physiol*, 82(3), 671-674.
- Iijima, H., Fujino, T., Minekura, H., Suzuki, H., Kang, M. J., and Yamamoto, T. (1996). Biochemical studies of two rat acyl-CoA synthetases, ACS1 and ACS2. *Eur J Biochem*, 242(2), 186-190.
- Lin, Y. H., Yu, C., and Huang, A. H. (1986). Substrate specificities of lipases from corn and other seeds. *Arch Biochem Biophys*, 244(1), 346-356.
- Monizde, S. M., and Drouin, G. (1996). Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol*, 13(9), 1198-1212.
- Ohlrogge, J. B., and Jaworski, J. G. (1997). Regulation of fatty acid synthesis. *Annual Review of Plant Physiology & Plant Molecular Biology*, 48(1), 109-113.
- Schnurr, J., Shockley, J., and Browse, J. (2004). The Acyl-CoA Synthetase Encoded by LACS2 Is Essential for Normal Cuticle Development in *Arabidopsis*. *Plant Cell*, 16(3), 629-642.
- Shen, B., Sinkevicius, K. W., Selinger, D. A., and Tarczynski, M. C. (2006). The homeobox gene GLABRA2 affects seed oil content in *Arabidopsis*. *Plant Mol Biol*, 60(3), 377-387.
- Shockley, J. M., Fulda, M. S., and Browse, J. A. (2002). *Arabidopsis* Contains Nine Long-Chain Acyl-Coenzyme A Synthetase Genes That Participate in Fatty Acid and Glycerolipid Metabolism. *Plant Physiology*, 129, 1710-1722.
- Shrago, E. (2000). Long-Chain Acyl-CoA as a Multi-effector Ligand in Cellular Metabolism. *Journal of Nutrition*, 130(2), 290-293.
- Watkins, P. A. (1997). Fatty acid activation. *Prog Lipid Res*, 36(1), 55-83.

Figure 1. The nucleotide sequence and deduced amino acid sequences of GmLACS

First line: nucleotide sequence. The initiation codon (ATG) is underlined. The stop codon (TGA) is asterisked. Second line: deduced amino acid sequence.

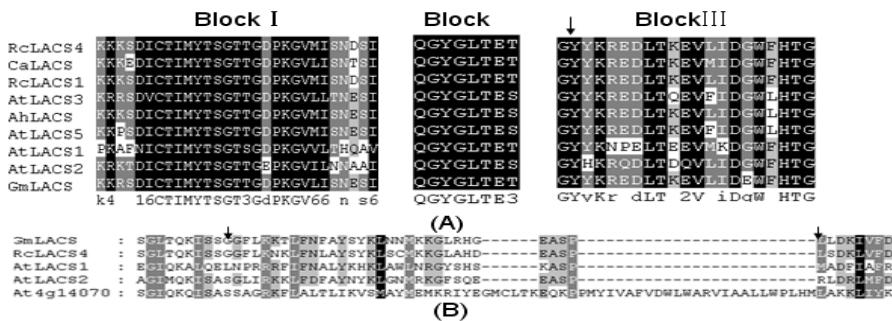


Figure 2. Alignment of GmLACS with other LACSSs

(A). Multiple amino acid sequence alignment of three blocks located in GmLACS with other homologous plant LACSSs. Black shading presented strictly conserved residues, and gray presented less strictly conserved residues. The conserved tyrosine which was indicated with black arrow was assumed to be involved in the adenylate formation. RcLACS1, RcLACS4 (*Ricinus communis* LACS1, LACS4, ABC02880, ACB30545), CaLACS(*Capsicum Annum*LACS, ACF17663), AtLACS1,AtLACS2, AtLACS3,AtLACS5(*Arabidopsis thaliana* LACS1, 2, 3, 5, AAM28868, AAM28869, AAM28870, AAM28872, AhLACS (*Arabidopsis halleri*LACS, ACC91252). (B). Comparison of the central sequences of At4g14070, AtLACS1, AtLACS2, RcLACS4 and GmLACS. Black triangles indicated the rough borders of the linker domain.

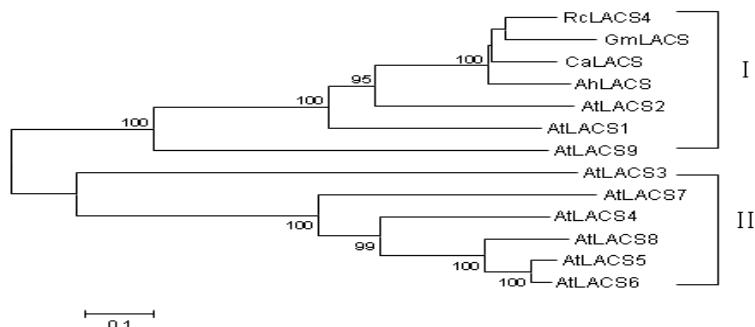


Figure 3. Phylogenetic analysis of *Glycine max* LACS, *Ricinus communis* LACS4, *Capsicum Annum* LACS, *Arabidopsis halleri* LACS and *Arabidopsis thaliana* LACSS

Protein sequences of LACSSs were aligned by ClustalX and Treeview. The bars stand for evolutionary distance. Bar = 0.1.

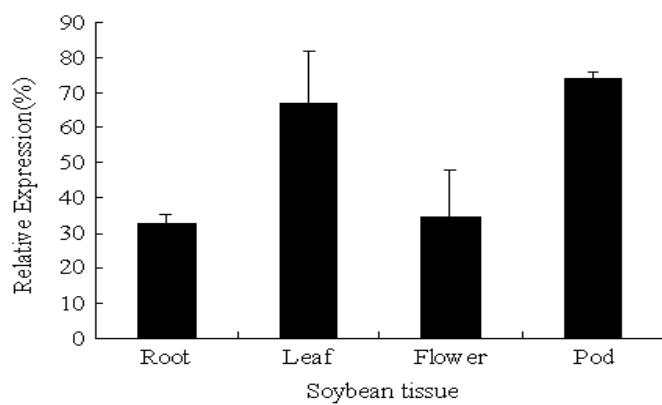


Figure 4. Expression of *GmLACS* genes in the soybean plant

Aliquots of total RNA were analyzed for expression of the gene in different tissues by real-time quantitative PCR using *GmLACS* gene-specific primers. Values were normalized to actin and represent mean \pm SE ($p < 0.05$).