MOLECULAR CLONING AND ANALYSIS OF THE GENES FOR COTTON PALMITOYL-ACYL CARRIER PROTEIN THIOESTERASE (PATE) AND Δ-12 FATTY ACID DESATURASE (FAD2-3) AND CONSTRUCTION OF SENSE AND ANTI-SENSE *PATE* PLASMID VECTORS FOR ALTERING OILSEED COMPOSITION OF TRANSGENIC COTTON PLANTS

Mongkol Nampaisansuk, B.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHIOLOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2002

APPROVED:

Robert M. Pirtle, Major Professor
Robert C. Benjamin, Committee Member
Kent D. Chapman, Committee Member
Douglas D. Root, Committee Member
John E. Knesek, Committee Member
Earl G. Zimmerman, Chair of the Department of Biological Sciences
C. Neal Tate, Dean of the Robert B. Toulouse School of Graduate Studies

Nampaisansuk, Mongkol, Molecular cloning and analysis of the genes for cotton palmitoyl-acyl carrier protein thioesterase (PATE) and Δ-12 fatty acid desaturase (FAD2-3) and construction of sense and anti-sense *PATE* plasmid vectors for altering oilseed composition of transgenic cotton plants. Doctor of Philosophy (Molecular Biology), May 2002, 241 pp., 1 table, 52 illustrations, reference list, 126 titles.

A cotton *PATE* cDNA clone has a 1.7-kb insert with an coding region for 410 amino acids, lacking codons for the three N-terminal amino acids. The predicted amino acid sequence of the PATE preprotein has a characteristic stromal-targeting domain and a 63% identity to the *Arabidopsis* FatB1 thioesterase sequence. A cotton genomic clone containing a 17.4-kb DNA segment was found to encompass a palmitoyl-ACP thioesterase (*FatB1*) gene. The gene spans 3.6 kb with six exons and five introns. The six exons are identical in nucleotide sequence to the open reading frame of the corresponding cDNA, and would encode a preprotein of 413 amino acids. The preprotein is identified as a FatB thioesterase from its deduced amino acid sequence similarity to those of other FatB thioesterase preproteins. A 5'-flanking region of 914 bp was sequenced, with the potential promoter/enhancer elements including basic helix-loop-helix elements (E box). Alkaline blot hybridization of cotton genomic DNA suggests the presence at least two *FatB1* thioesterase genes in cotton. Four plasmid constructs for both constitutive and seed-specific anti-sense RNA suppression and gene-transgene co-suppression of *PATE* gene expression were successfully generated. Two overlapping cotton genomic clones were found to encompass a Δ-12 fatty acid desaturase (*FAD2-3*) gene. The continuous *FAD2-3* coding region is 1,155 bp and would encode a protein of

384 amino acids. The *FAD2-3* gene has one large intron of 2,967 bp entirely within its 5'-untranslated region. Several potential promoter/enhancer elements, including several light responsive motifs occur in the 5'-flanking region. Yeast cells transformed with a plasmid construct containing the cotton *FAD2-3* coding region accumulate an appreciable amount of linoleic acid (18:2), not normally present in wild-type yeast cells, indicating that the gene encodes a functional FAD2 enzyme.

ACKNOWLEDGMENTS

This research was supported in part by Agreements 97-471 and 99-652 from Cotton Incorporated overseen by Dr. Gay M. Jividen, by Grant 003594-014 from the Texas Advanced Technology Program, Award 98-35503-6339 from the NRI Competitive Grants Program/USDA, and University of North Texas Organized Research Funds.

I thank Drs. Robert and Irma Pirtle for their support and encouragement throughout my study at the University of North Texas. I also thank Dr. Kent Chapman, Dr. Robert Benjamin, Dr. Douglas Root and Dr. John Knesek for serving on my research committee and for their advice.

I thank my parents for their encouragement and financial support throughout my years in the USA.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF ILLUSTRATIONS

GENERAL INTRODUCTION

Cottonseed oil is commercially used in many industries, including the food and plastics industries. Cottonseed oil is primarily composed of 22% saturated fatty acid (palmitic acid), 15% monounsaturated fatty acid (oleic acid) and 58% polyunsaturated fatty acid (linoleic acid) (Jones and King, 1996). Oils with high monounsaturated fatty acid (oleic acid) have a higher value in the edible vegetable oil market due to health benefits and also high stability when used in food processes. The synthesis of palmitic, oleic, and linoleic acids and their incorporation into triacylglycerols are metabolically interelated (Browse and Ohlrogge, 1995), as shown in Figure 2. Their relative proportions are determined by the regulation of carbon flux through these complex metabolic pathways. The metabolic relationship between palmitic, oleic, and linoleic acid biosynthesis makes it possible to manipulate the relative proportions of these fatty acids in the cottonseed oil profile. To be capable of manipulating the fatty acids profile of cottonseed oil toward a higher oleic acid level, it is important to understand *de novo* fatty acid biosynthesis pathways and the subsequent assembly of those fatty acids into triacylglycerol seed oil. The crux to manipulating the proportions of these fatty acids in seed oil is to understand the regulatory points in these pathways and to specifically alter them. As can be seen in Figures 1 and 2, palmitoyl-acyl carrier protein (ACP) thioesterase (PATE) and oleoyl-phosphatidylcholine desaturase (FAD2) may be two key enzymes regulating the cottonseed fatty acid composition toward higher oleic acid levels. PATE is an enzyme that has been shown to be responsible for hydrolyzing the palmitoyl-

ACP thioester bond, and thus terminating acyl chain elongation, as shown in Figure 2. PATE has been shown to be responsible for formation of palmitic acid in palmitate-rich oils (Dörmann et al., 1995; Jones et al., 1995). Therefore, PATE may be a key enzyme controlling carbon flux via palmitic acid. Based on the metabolic relationship between palmitic and oleic acid biosynthesis, as shown in Figure 2, we hypothesize that the inhibition of mRNA translation of this PATE enzyme by a cognate antisense RNA or by gene co-suppression will increase the flux of carbon in the pathway toward oleic acid formation. In plants with polyunsaturated fatty acids in their triacylglycerols, the acyl moieties of phosphatidylcholine are the substrates for the introduction of additional double bonds. The FAD2 and FAD3 enzymes desaturate extrachloroplast lipids (Browse and Ohlrogge, 1995), as shown in Figure 2, and are responsible for the synthesis of linoleic acid (18:2) and linolenic acid (18:3), respectively. Oleoyl-phosphatidylcholine (PC) desaturase (FAD2) selectively prefers 18:1-PC as its substrate (Okuley, 1994) resulting in the formation of linoleoyl-PC (18:2-PC). Thus, FAD2 may be a key enzyme controlling carbon flux through linoleic acid. Based on the metabolic relationship between linoleic acid and oleic acid biosynthesis, as shown in Figure 2, we hypothesize that the inhibition of mRNA translation of the FAD2 enzyme by the cognate antisense RNA or, alternatively, by gene co-suppression will increase the flux of carbon in the pathway toward oleic acid formation. To test these hypotheses, the structure and function of the *Gossypium hirsutum PATE* and *FAD2* genes must be analyzed. Understanding the structure and function of these *PATE* and *FAD2* genes may provide the necessary information to alter fatty acid composition of cottonseed oil. One goal of this dissertation research project was to clone and analyze the cotton *PATE* (Chapter 1) and *FAD2* genes

(Chapter 3), and to design and construct sense and antisense vectors for altering the oilseed composition of transgenic cotton plants (Chapter 2).

In Chapter 1, the isolation and analysis of a cotton *PATE* cDNA and *PATE* gene are described. The architecture of the cotton *PATE* gene is described and the potential regulatory (promoter/enhancer) elements of this cotton gene are identified.

In Chapter 2, the construction of *PATE* sense and antisense vectors for altering cotton oilseed composition is described. Dr. Tu T. Huynh (Huynh, 2001), of the laboratory of Dr. Kent Chapman at the University of North Texas, used these constructs to examine the *in planta* role of the cotton *PATE* in transgenic cotton plants harboring the PATE sense and antisense constructs.

In Chapter 3, the isolation and analysis of the cotton *FAD2-3* gene is described. The architecture of the cotton FAD2-3 is also described and the potential regulatory (promoter/enhancer) elements of this cotton gene are identified. The functional expression of the cotton *FAD2-3* gene in transformed *Saccharomyces cerevisiae* cells were examined in order to address if the isolated cotton *FAD2-3* cDNA clone encodes a functional protein product.

CHAPTER 1

MOLECULAR CLONING AND ANALYSIS OF THE GENE FOR A COTTON PALMITOYL-ACYL CARRIER PROTEIN THIOESTERASE (PATE)

Introduction

Fatty Acid Biosynthesis in Plant

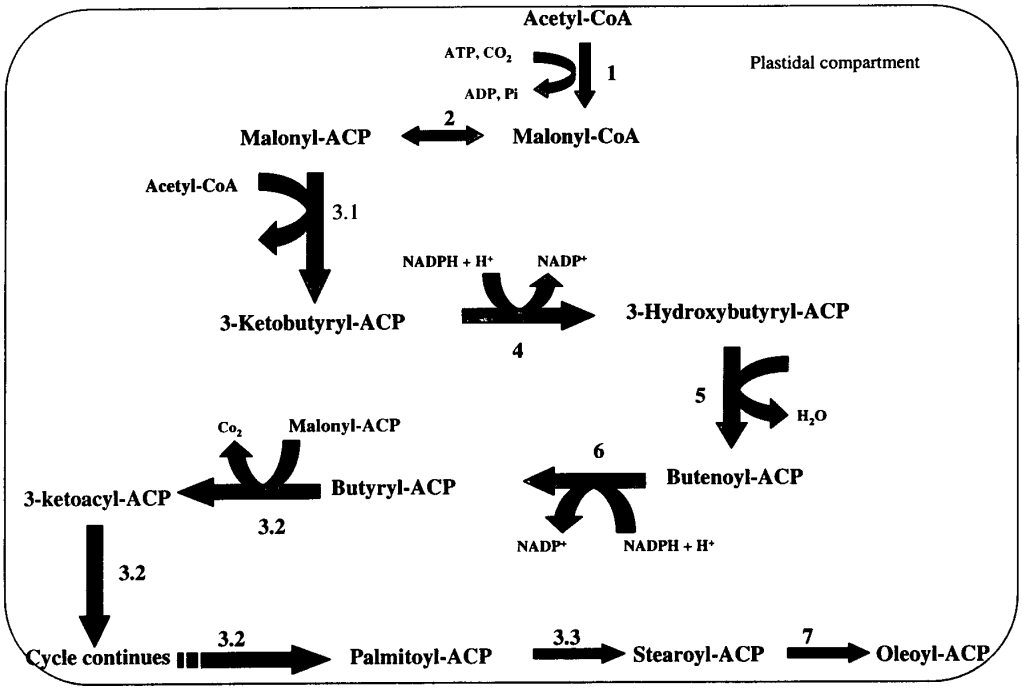
Fatty acids in plants, as in all other organisms, are the major structural components of phospholipids of biological membranes and triacylglycerols of storage oils. Bilayer membranes form boundaries between the inside and outside of cells and also between cellular compartments. For example, chloroplast membranes are the sites where the light and dark reactions of photosynthesis and the electron transport reactions of photosynthesis occur. Triacylglycerols serve as the major storage form of carbon in seed oils for use as an energy source during germination (Ohlrogge and Browse, 1995). Therefore, it is important to understand the mechanisms underlying the control of fatty acid amounts and compositions in the phospholipids of biological membranes and in the triacylglycerols of storage oils in plants.

De novo fatty acid biosynthesis occurs primarily in the stromal compartment of plastids. A type II, fully dissociable fatty acid synthase complex is utilized in this complex pathway (Ohlrogge et al., 1993; Ohlrogge and Browse, 1995; Ohlrogge and Jaworski, 1997; Harwood, 1996). As shown in Figure 1, at least 30 enzymatic reactions are required to produce the common 16- or 18-carbon fatty acids from acetyl-CoA and malonyl-CoA. Malonyl-CoA is a carbon donor for fatty acid biosynthesis and is

produced from acetyl-CoA and carbon dioxide by acetyl-CoA carboxylase in plastids. The malonyl group is then transferred from a malonyl-CoA to an acyl carrier protein (ACP) to form malonyl-ACP. Once malonyl-ACP is formed, it enters into a series of reactions including acylation, condensation and reduction reactions. These reactions result in the formation of a carbon-carbon bond and the release of one molecule of carbon dioxide. In each round of the synthesis, a two-carbon unit from malonyl-ACP is added to the growing acyl-ACP chain. There are three separate condensing enzymes, 3-ketoacyl-ACP synthases (KAS) I, II, and III, required in fatty acid biosynthesis (Jaworski et al., 1994). KAS III is responsible for catalyzing the first condensation of acetyl-CoA and malonyl-ACP. KAS I is responsible for producing acyl chains of from 6 to 16 carbons, and KAS II is required for chain elongation of palmitoyl-ACP (C16) to stearoyl-ACP (C18) (Browse et al., 1994). After each condensation, three additional reactions occur to form a saturated fatty acid. First, 3-ketoacyl-ACP is reduced by 3-ketoacyl-ACP reductase at the carbonyl group using NADPH as an electron donor. The second reaction is dehydration by hydroxyacyl-ACP dehydratase. The last reaction is reduction of a double bond by enoyl–ACP reductase using NADH or NADPH to form a saturated fatty acid. Repeated cycles of the chain elongation of fatty acyl-ACP continue until the carbon length is 16 or 18, which is normal for a fatty acid found in phospholipid membranes and in triacylglycerols of seed oils. Chain elongation of fatty acids in plastids is terminated when an acyl group is hydrolyzed from an acyl carrier protein (ACP) by an acyl-ACP thioesterase.

Figure 1. Simplified scheme of plant fatty acid biosynthesis Modified from Figure 3 of Ohlrogge, J. and Browse, J. The Plant Cell 7, 957-970 (1995).

1-	Acetyl-CoA carboxylase	4-	3-ketoacyl-ACP reductase
2-	Malonyl-CoA: ACP transferase	5-	3-hydroxyacyl-ACP dehydrase
3.1	3-ketoacyl-ACP synthase III	6-	enoyl-ACP reductase
3.2	3-ketoacyl-ACP synthase I	7-	stearoyl-ACP desaturase
3.3	3-ketoacyl-ACP synthase II		


Acetyl-CoA
ATP, CO_2
ADP, Pi
1
Plastidal compartment
2
Malonyl-ACP
Malonyl-CoA
Acetyl-CoA
3.1
NADPH + H+
NADP+
3-Ketobutyryl-ACP
3-Hydroxybutyryl-ACP
4
5
H_2O
Co_2
Malonyl-ACP
6
3-ketoacyl-ACP
Butyryl-ACP
Butenoyl-ACP
3.2
NADP+
NADPH + H+
3.2
Cycle continues
3.2
Palmitoyl-ACP
3.3
Stearoyl-ACP
7
Oleoyl-ACP
Extraplastidal compartment

As shown in Figure 2, there are two possible ways for terminating acyl-chain elongation. In most cases, an acyl-ACP thioesterase hydrolyses an acyl-ACP and releases a free fatty acid. This has been termed the eukaryotic pathway of glycerolipid biosynthesis (Ohlrogge and Browse, 1995). This free fatty acid is then converted into a corresponding acyl-CoA in an unknown manner and is probably transported out of plastids by simple diffusion. Thus, acyl-ACP thioesterases play a central role in fatty acid biosynthesis because they are important in determining the chain length of fatty acids transported to the cytosol (Chasan, 1995; Voelker, 1996). Alternatively, acyltransferases in plastids transfer fatty acids from ACP to glycerol-3-phosphate or to monoacylglycerol-3-phosphate to produce phospholipids for plastidal membranes. This process is called the prokaryotic pathway of glycerolipid synthesis. The departure of fatty acids from plastids determines which of the two pathways is used (Ohlrogge and Browse, 1995).

The focus of this part of this dissertation is on the way in which fatty acids are incorporated into the triacylglycerols of oil bodies. These fatty acids have to leave the plastids as free fatty acids and then are re-esterified with CoA. These acyl-CoAs in turn are the substrates for triacylglycerol formation in oil bodies and for membrane phospholipid synthesis in the endoplasmic reticulum.

Plant Acyl-ACP Thioesterases

Thioesterases are enzymes that have been shown to be responsible for releasing acyl groups from ACP and thus terminating chain elongation during *de novo* fatty acid biosynthesis. This free acyl group is then able to leave plastids and be incorporated into

Figure 2. Simplified scheme of plant glycerolipid synthesis

Modified from Figures 4 and 5 of Ohlrogge, J. and Browse, J. The Plant Cell 7, 957-970 (1995).

1- lauroyl-ACP thioesterase

2- palmitoyl-ACP thioesterase

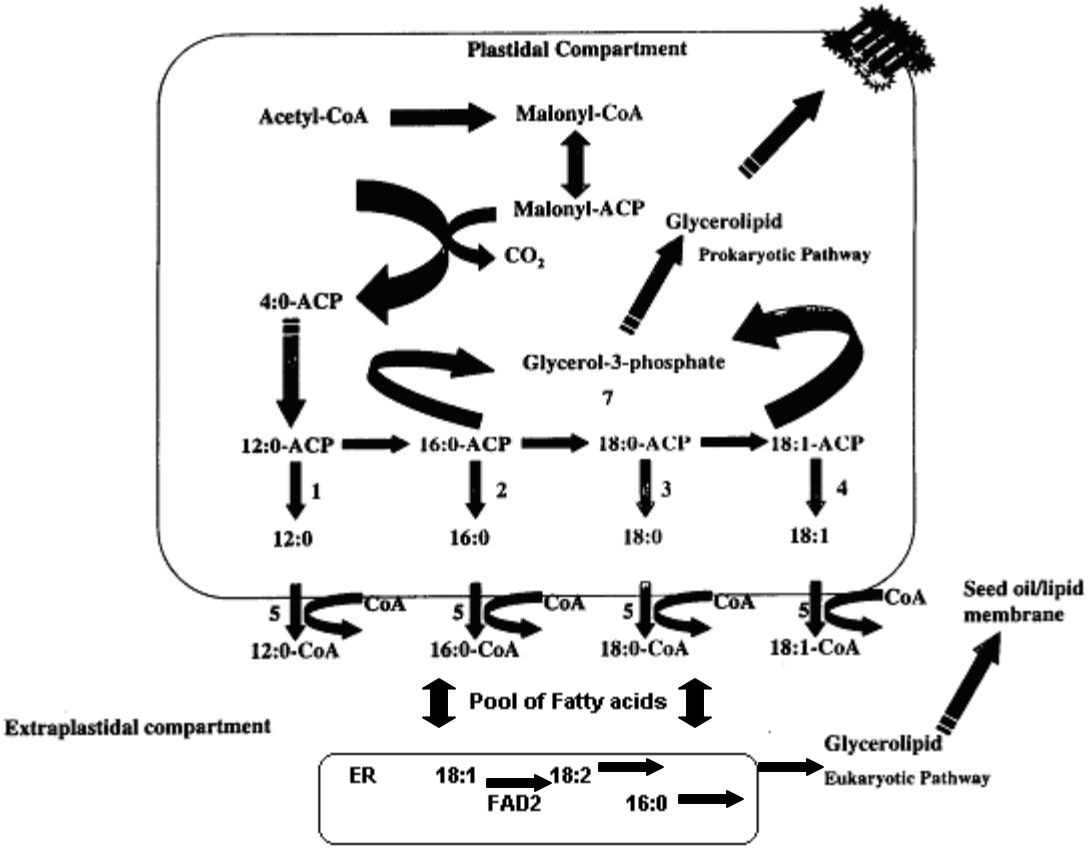
3- stearoyl-ACP thioesterase

4- oleoyl-ACP thioesterase

5- acyl-CoA synthase

6- endoplasmic reticulum acyltransferase

7- plastidal acyl transferase


Plastidal Compartment
Acetyl-CoA
Malonyl-CoA
Malonyl-ACP
CO_2
Glycerolipid
Prokaryotic Pathway
4:0-ACP
Glycerol-3-phosphate
7
12:0-ACP
16:0-ACP
18:0-ACP
18:1-ACP
1
2
3
4
12:0
16:0
18:0
18:1
5
CoA
12:0-CoA
16:0-CoA
18:0-CoA
18:1-CoA
Seed oil/lipid membrane
Pool of Fatty acids
Extraplastidal compartment
Glycerolipid
Eukaryotic Pathway
ER
18:1
18:2
FAD2
16:0

triacylglycerols. Plant acyl-ACP thioesterase activity was first detected in avocado mesocarp extracts by Shine et al. (1976). However, the first plant acyl-ACP thioesterase was partially purified and characterized by Ohlrogge et al. (1978). The acyl-ACP thioesterase activity from this mesocarp tissue was shown to be most active on 18:1-ACP (oleate), less active on 18:0-ACP (stearate) and 16:0-ACP (palmitate), and least active on 12:0-ACP and 14:0-ACP. The acyl-ACP substrate preferences are in good correlation since oleate (18:1) and its derivatives are the major acyl substrates supplied to the eukaryotic pathway of glycerolipid biosynthesis, while palmitate (16:0) and its derivatives represent the remainder (between 5 and 20%) in normal plant tissues (Hilditch and Williams, 1964). Until now, oleoyl-ACP thioesterases specific for the 18:1-ACP substrate have been the most commonly found. Again, this might be because oleate and its derivatives represent the major portion of *de novo* fatty acyl chain flux through the eukaryotic pathway in all plant cells (Hilditch and Williams, 1964). Therefore, oleoyl-ACP thioesterase is considered to be an essential housekeeping enzyme (Gibson et al., 1994).

Voelker and coworkers classified plant acyl-ACP thioesterases by their preferences for acyl-ACP substrates, specified by the degree of saturation and chain length of the acyl group (Jones et al., 1995; Voelker, 1996). All *Fat A* enzymes appear to prefer mostly 18:1-ACP as substrates, with minor activities with 18:0- and 16:0-ACPs, suggesting that they are involved predominantly in hydrolysis of unsaturated fatty acyl-ACPs. Since all thioesterase isolates from several plant families have a near identical hydrolytic specificity, it is likely that *Fat A* thioesterases are the conserved 18:1-ACP thioesterase activity in higher plants. This thioesterase activity is a critical function in

plant fatty acid biosynthesis because 18:1 is an immediate precursor of most fatty acyl groups found in phospholipids and triacylglycerols synthesized in the eukaryotic pathway. In contrast, the *Fat B* thioesterase enzymes involved in hydrolysis of saturated fatty acyl-ACPs showed a broad array of substrate specificities. Thus, the *Fat B* enzymes can be grouped into two classes. One group preferentially hydrolyzes acyl-ACP substrates of medium-chain length (8:0-12:0) corresponding to those accumulating in seed triacylglycerols such as California bay thioesterase (*Uc Fat B1*), *Cuphea hookeriana* (*Ch Fat B2*), *Cuphea lanceolate* (*Cl Fat B3*), or *Cuphea palustris* (*Cp Fat B1*). *Fat B* thioesterases with hydrolytic specificities for the predominant medium-chain fatty acids in the respective seed oils, have been found in species of taxonomically disperse angiosperm families (*Lauraceae, Lythraceae,* and *Ulmaceae*). Therefore, *FatB* thioesterases with preferences for medium chain length acyl-ACPs are common components in fatty acid biosynthesis.

The expression and enzymatic properties of the medium chain-specific *Fat B* subgroup contrasts markedly with the other group represented by the *Ch Fat B1* thioesterase from *Cuphea hookeriana* and the *At Fat B1* thioesterase of *Arabidopsis thaliana*. This group was classified as long-chain (14:0-18:0) acyl-ACP thioesterases. Both groups of thioesterase enzymes cover the C14-C18 acyl-ACP range, with a preference for 16:0-ACP. Also, expression of these genes was not restricted to seeds. The typical 16:0-ACP *Fat B* is probably present in all major plant parts and is not restricted to species producing medium chain-length fatty acids. Based on the expression pattern, enzyme specificity, and presence in all angiosperm species investigated, this type of *Fat B* thioesterase is considered to be the evolutionary ancestor of the specialized

medium chain-length thioesterase. Therefore, all higher plant tissues apparently not only contain an oleoyl-ACP thioesterase but also a somewhat less specific, 16:0-centered thioesterase, designated as a palmitoyl-ACP thioesterase (Jones et al., 1995; Voelker, 1996).

Using oligonucleotide probes deduced from peptide sequences, Knutzon et al. (1992) was able to isolate two oleoyl-ACP thioesterase cDNA clones (*Ct Fat A1* and *Ct Fat A2*) from a maturing seed safflower (*Carthamus tinetorius*) cDNA library. The full-length cDNAs are about 1.5 kb, encoding polypeptides of 385 and 390 amino acids, respectively, with an identity of about 80%. The N-terminus of the purified protein is located at residue 61 of the derived sequence of *Ct Fat A1*, most likely defining a transit peptide of 60 amino acids and a mature protein of 325 amino acids, with a calculated molecular weight of 37,254 that matched the approximate molecular weight of the partially purified safflower thioesterase. Activity profiles of *Ct Fat A1* and *A2* expressed in *E. coli* were rather similar, with both having a preference for 18:1-ACP, indicating that the two cDNAs encoded oleoyl-ACP thioesterases (Knutzon et al., 1992). Helleyer et al. (1992) purified and characterized an oleoyl-ACP thioesterase from maturing seeds of *Brassica napus*. Loader et al. (1993), using oligonucleotide probes deduced from internal peptide sequences of Helleyer et al. (1992), isolated two classes of oleoyl-ACP thioesterase cDNA clones from maturing seeds of *Brassica napus*. The cDNA open reading frames (ORFs) are nearly identical (about 98%). Both cDNAs encode polypeptides of 366 amino acids. Sequence comparisons with *Fat A* thioesterase cDNAs of safflower (Knutzon et al., 1992) showed high sequence conservation, indicating that

these two cDNAs (*Bn Fat A1* and *Bn Fat A2*) very likely also encode oleoyl-ACP thioesterases (Loader et al., 1993).

In the case of specialized plant tissues (such as maturing seeds of California bay trees), medium-chain fatty acids are exclusively produced for incorporation in triacylglycerols, the major carbon storage. Medium chain-producing seeds of several plant species were investigated, but no evidence of medium-chain thioesterase activity was detected. Pollard et al. (1991) and Davies et al. (1991) showed that extracts from maturing seeds of *Umbellularia californica* exhibited hydrolytic activity toward 12:0-ACP when compared to immature seeds of the same species. They also showed that this 12:0-ACP hydrolytic activity was different from the hydrolytic activity of 18:1-ACP hydrolase and 12:0-CoA hydrolase which were also present in the extracts, thus indicating specificity for 12:0-ACP. These results suggested the role of this 12:0-ACP thioesterase in the production of medium-chain fatty acids *in vivo* (Pollard et al., 1991; Davies et al., 1991).

Using a protein sequence-based cloning method, Voelker et al. (1992) were able to isolate this 12:0-ACP thioesterase (*Uc Fat B1*) cDNA clone from seed tissues of Undomestical California Bay (*Umbellularia californica*). *Uc Fat B1* cDNA has an open reading frame encoding a polypeptide of 382 amino acids, matching most of the short peptide sequences obtained from the purified protein (Voelker et al., 1992). GenBank database searches yielded no significant matches with the *Brassica napus* 18:1-ACP thioesterase cDNA sequences, indicating that the bay thioesterase belongs to a novel class of genes. Expression of this cDNA clone under control of a napin seed storage protein promoter in transgenic *Arabidopsis* plants resulted in a dramatically elevated

12:0-ACP hydrolytic activity as compared to that of control plants, and medium-chain fatty acids accumulated up to 20% of the acyl groups in the mature seeds of these plants.

Jones et al. (1995) unexpectedly obtained a 16:0-ACP thioesterase cDNA clone from seeds of *Cuphea hookeriana*, in which mainly 8:0 and 10:0 fatty acids accumulated. Expression of this cDNA clone in transgenic *Brassica napus* plants led to production of a 16:0-rich oil (Jones et al., 1995). Dörmann et al. (1995) isolated a thioesterase cDNA coding for a long-chain acyl-ACP thioesterase from *Arabidopsis thaliana*. The deduced amino acid sequence of this protein has an identity of 51% with the bay 12:0-ACP thioesterase and has an identity of 39% to the predicted sequence of the safflower 18:1-ACP thioesterase. The gene product, when expressed in *E. coli* showed thioesterase activity for the long-chain acyl-ACPs (14:0, 16:0, 18:0, and 18:1). When expressed in β-oxidation in *E. coli*, the transformed cells had a high level of fatty acids that mainly consisted of 16:0 and some 14:0 fatty acids (Dörmann et al., 1995). These results suggest that this might be a novel thioesterase, different from the medium-chain thioesterases, and that is responsible for synthesis of long-chain fatty acids in plants accumulating long-chain fatty acids in their carbon storage.

One of our long-term goals is to alter the fatty acid composition of cottonseed oil. In order to be capable of manipulating the fatty acid profiles of cottonseed oil toward a higher level of monounsaturated fatty acid (oleate), an understanding of the regulation of *de novo* fatty acid biosynthesis and assembly of those fatty acids into seed triacylglycerols is required. As can be seen in Figures 1 and 2, palmitoyl-ACP thioesterase and oleoyl-PC desaturase may be two key enzymes regulating cottonseed fatty acid composition. Oleoyl-phosphatidylcholine (PC) desaturases (fatty acid

desaturases) have been shown to be responsible for formation of linoleoyl-PC in linoleate (18:2)-rich oil (Okuley et al., 1994). Thus, this desaturase may be a key enzyme controlling carbon flux through linoleic acid. Palmitoyl-ACP thioesterases have been shown to be responsible for formation of palmitic acid in palmitate (16:0)-rich oil (Dörmann et al., 1995; Jones et al., 1995). Thus, the palmitoyl-ACP thioesterase (PATE) may be a key enzyme controlling carbon flux through palmitic acid. By controlling expression of these two enzymes, it may be possible to shunt the carbon flux through the formation of oleic acids, resulting in accumulation of increased levels of oleic acid for oil biosynthesis. As an initial step, we have isolated and characterized a cotton *Fat B1* cDNA clone for palmitoyl-ACP thioesterase (Pirtle et al., 1999), and demonstrated that a *FatB1* acyl-ACP thioesterase activity was present in cotton and probably responsible for the high levels of palmitic acids incorporated in cottonseed oils.

A primary and poorly understood level of control of plant lipid metabolism is that of the regulation of expression of the genes encoding enzymes of lipid biosynthesis such as promoter/enhancer elements of individual genes of lipid biosynthesis, and the more complex question of the likelihood of trans-acting transcription factors simultaneously controlling the expression of many genes of lipid synthesis (Topfer and Martini, 1994; Ohlrogge and Browse, 1995; Ohlrogge and Jaworski, 1997). Numerous cDNAs encoding plant thioesterases have been analyzed but only one thioesterase gene, a *Brassica napus Fat A* gene has been completely sequenced (Loader and Safford, 1993; GenBank accession number X87842). To date, the structure of a *Fat B* thioesterase gene has not been reported, and the types of potential promoter/enhancer elements in the 5' flanking region of this gene have not been identified. Thus, the regulation of expression of both

Fat A and *Fat B* acyl-ACP thioesterase genes and their roles in lipid biosynthesis in plant tissues are still largely unclear. However, research described in this dissertation should provide information about the structure and the potential regulatory (promoter/enhancer) elements of the cotton palmitoyl acyl-ACP thioesterase gene, and this could be a start point to delineate the role of fatty acyl-ACP thioesterase gene expression in lipid biosynthesis (Yoder et al., 1999).

The focus of the first chapter of this dissertation is on the molecular cloning and characterization of a cotton palmitoyl-ACP thioesterase cDNA (Pirtle et al., 1999) and gene (Yoder et al., 1999), which will permit the analysis of potential regulatory elements in expression of this class of genes for lipid biosynthesis. Understanding the structure and function of this *Gh Fat B1* gene may provide the necessary information to alter fatty acid composition of cottonseed oil by genetic engineering. Due to the difficulty in purifying to homogeneity acyl-ACP thioesterase specific for only long-chain acyl-ACP substrate and the high percentage of sequence homology among the plant acyl-ACP thioesterases, a DNA sequence-based cloning method was used in this research. Using this approach, Knutzon et al. (1992) was able to obtain a *Fat A* cDNA clone from *Brassica rapa* using a safflower *Fat A1* cDNA heterologous hybridization probe. The major advantage of the DNA sequence-based method is the ability to clone genes encoding thioesterases with unique specificities (Voelker, 1996). Thus, this research has involved isolation of a cotton *PATE* cDNA and gene by screening cotton cDNA and genomic libraries using the *At Fat B1* cDNA (Dörmann et al., 1995) as a heterologous hybridization probe. The structure of the *PATE* cDNA and gene have been characterized by physical mapping and DNA sequencing (Pirtle et al., 1999; Yoder et al., 1999). In

addition, the gene copy number of the *PATE* gene in the cotton genome has been determined by genomic blot hybridization.

Materials and Methods

A cotton (*Gossypium hirsutum*, *cv. Acala SJ-5*) genomic library, generated from cotton genomic DNA that was partially digested with *Sau3AI* and ligated into the lambda vector EMBL3 was generously provided by Dr. David M. Anderson of Phytogen Seeds, Placentia, CA (Grula et al., 1995). A cotton cDNA library, constructed from cotyledon mRNA of 48-hour dark grown cotton seedings (*Gossypium hirsutum cv. Deltapine 62*) and harbored in the Stratagene UniZAP lambda vector, was kindly provided by Dr. R. N. Trelease of Arizona State University (Ni and Trelease, 1991). An *Arabidopsis thaliana* cDNA clone, designated TE 3-2 encoding a long-chain acyl-ACP thioesterase, was generously provided by Dr. John Ohlrogge of Michigan State University (Dörmann et al., 1995). Two cotton genomic DNAs, cultivars *Pima* and *TM1*, were kindly provided by Dr. Sukumar Saha of the USDA Agriculture Research Station, Mississippi State University, MS. The plasmid vector pGEM7Zf(+) was purchased from Promega Corp. The *E. coli* strain *DH5α* was from Invitrogen. *E. coli XL1-Blue MRF'* and ExAssist Interference-Resistant Helper Phage with the *E. coli SOLR* strain were from Stratagene. Restriction endonucleases, T4 DNA ligase, and calf intestinal alkaline phosphatase were from Invitrogen. Agarose (NuSieve, LE, and Sea Plaque GTG) was from FMC Corp. Oligonucleotides for polymerase chain reaction (PCR) amplification and sequencing primers were from Biosynthesis, Inc. Nitrocellulose and Hybond N^+ nylon membranes for DNA transfer were from Schleicher and Schuell and Amersham Life Sciences, respectively. Wizard Plus Miniprep DNA purification kits were from Promega Corp.,

and QIAquick gel extraction and PCR purification kits were from Qiagen Corp. Sequenase Version 2.0 DNA sequencing kits and ThermoSequenase cycle sequencing kits were from Amersham Life Sciences. Radioactive materials, including [α-^{33}P]-ddNTPs and [α-^{32}P]-dCTP, were purchased from Amersham Life Sciences and NEN Life Sciences, Inc., respectively. Various other reagents, chemicals and materials were purchased from Kodak, Fisher Scientific, Sigma Chemical Co., and Millipore Corp.

Large Scale Isolation and Purification of an *Arabidopsis* Thioesterase TE 3-2 Clone

An *Arabidopsis thaliana* palmitoyl-ACP thioesterase cDNA clone, designated TE 3-2, encoding a long-chain acyl-ACP thioesterase was generously provided by Dr. John Ohlrogge of Michigan State University (Dörmann et al., 1995). A small culture of 50 ml of Terrific Broth (TB) containing amplicillin (100 μg/ml) was inoculated with three loops of *E. coli* containing the recombinant plasmid designated TE 3-2 and incubated with aeration at 37°C overnight. The absorbance at 600 nm (A_{600}) was monitored until it was 1.96. Then, 2.0 ml of the overnight culture were added to 100 ml of TB with an amplicillin-containing overday culture. This overday culture was grown until the A_{600} reached 0.685 (late-log phase). Then, 80 ml of this overday culture was added to two liters of TB containing amplicillin (100 μg/ml) and incubated with aeration at 37°C overnight. This was done in five 2-liter flasks containing 400 ml of culture each. Bacterial cells were harvested by centrifugation in a Sorvall GS-3 rotor at 4,000 rpm (2,300xg) for 30 minutes at 4°C. The supernatants were removed and the cells were weighed. The cells were then subjected to several washes in ice-cold STE buffer (0.1 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

The plasmid DNA was isolated from the bacterial cells by an alkaline lysis procedure (Sambrook et al., 1989). The bacterial cells were resuspended in Solution I (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8.0) and lysed in a freshly prepared solution of lysozyme (10 mg/ml) in 10 mM Tris-HCl (pH 8.0) and Solution II (0.2 N NaOH, 1% SDS). Solution III (5 M potassium acetate, pH 4.8) was then added to the lysed cells to precipitate the bacterial chromosomal DNA and proteins. The cellular debris was pelleted at 4,000 rpm (2,300xg) for 30 minutes at 4°C. Supernatants were collected and precipitated by addition of 0.6 volume of isopropanol at room temperature for 10 minutes. The precipitated DNA was then pelleted by centrifugation in a Sorvall GS-3 rotor at 6,000 rpm (5,200xg) for 15 minutes at room temperature. The pellets were resuspended in 10 ml of TE buffer. RNase A was added to make a final concentration of 10 μg/ml and incubated at room temperature for one hour. A phenol/chloroform extraction was done by addition of one volume of phenol/chloroform/isoamyl alcohol (24:24:1) with vigorous vortexing. The samples were subsequently centrifuged in a Sorvall SS-34 rotor at 6,000 rpm (5,200xg) for 15 minutes. The aqueous phase was transferred to a new tube and one volume of chloroform:isoamyl alcohol (24:1) was added to the aqueous phase. The samples were vigorously vortexed for two minutes and centrifuged in the SS-34 rotor at 6,000 rpm (5,200xg) at room temperature for 15 minutes.

After transfer of the aqueous phase to a clean tube, 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.0 volumes of 100% ethanol were added and mixed with the aqueous phase for overnight precipitation in a -90°C freezer. The ethanol-precipitated sample was warmed to 4°C on ice and then centrifuged in the Sorvall SS-34 rotor at

12,500 rpm (14,400xg) for 30 minutes at 4°C. Pellets were washed with 10 ml of 70% ethanol, dissolved in an appropriate volume of 1X HPLC starting buffer (25 mM Tris-HCl, 1 mM EDTA, pH 8.0) and then purified by HPLC by Dr. Irma Pirtle of this laboratory with a Protein-Pak DEAE 5PW anion exchange HPLC chromatography using the general conditions of Merion and Warren (1989). The absorbance of the HPLC-purified plasmid DNA was measured at 260 nm and 280 nm in order to estimate the amount and purity of plasmid DNA. The recombinant plasmid DNA was analyzed by cleavage with several restriction endonucleases and electrophoresis on a 1% agarose gel to confirm the expected fragment sizes and to verify the identity of the particular recombinant plasmid DNA. The *Arabidopsis* TE 3-2 plasmid DNA was digested with *Sal*I and *Eco*RV and fractionated on a 0.8% agarose gel, and the gel slice containing the 1.4 kb *Sal*I/*Eco*RV restriction fragment was recovered from the gel slice by electroelution (Sambrook et al., 1989). This purified 1.4 kb *Sal*I/*Eco*RV fragment was then used by Dr. Irma Pirtle as a template to generate a heterologous *Arabidopsis* ^{32}P-labeled *PATE* hybridization probe by the random-priming procedure of Feinberg and Vogelstein (Feinberg and Vogelstein, 1983; Feinberg and Vogelstein, 1984).

Isolation of the Cotton Palmitoyl-ACP Thioesterase cDNA

A cotton cDNA library generated from cotyledon mRNA of 48-h dark-grown cotton seedlings (*Gossypium hirsutum cv. Deltapine62*) and harbored in the Stratagene UniZap lambda vector was kindly provided by Dr. R. N. Trelease of Arizona State University (Ni and Trelease, 1991). The 1.4-kb *Sal*I/ *Eco*RV restriction fragment from the *Arabidopsis* TE 3-2 cDNA clone was used as hybridization probe. In collaboration with Mr. David Yoder and Drs. Robert and Irma Pirtle of our laboratory, approximately

380,000 plaques of the cotton cDNA library were screened by the plaque hybridization procedure of Benton and Davis (1977), using the host strain *E. coli XL1-blue MRF'* (Stratagene). Prehybridization of the positively-charged nylon filter (Hybond N^+, Amersham) replicas was done at 55°C in a solution consisting of 6xSSC (1XSSC is 150 mM NaCl and 15 mM sodium citrate, pH 7.0), 0.5% SDS, 5X Denhardt's reagent and denatured sheared salmon sperm (100 μg/ml) for four hours. Hybridization was done overnight in a solution consisting of 6XSSC, 0.5% SDS, 5X Denhardt's reagent, 20 mM Tris-HCl (pH 8.0), 2 mM Na_2 EDTA (pH 7.5), 2.5 mM sodium pyrophosphate (pH 8.0), denatured sheared salmon sperm (100 μg/ml), and ^{32} P-labeled probe (Anderson and Young, 1985; Ausubel et al., 1987; Sambrook et al., 1989). After hybridization, the filters were rinsed once at room temperature with 2XSSC. Subsequently, the filters were washed three times at 55°C (once with 2XSSC and 0.1% SDS for 30 minutes, and twice with 1XSSC and 0.1% SDS for 30 minutes). Several plaques that exhibited intense positive signals with the heterologous *Arabidopsis* TE 3-2 hybridization probe were subjected to successive rounds for plaque purification.

A pBluescript SK(-) phagemid derivative designated SKCPc115b containing a 1.7-kb cotton cDNA insert was excised *in vivo* using Stratagene ExASSIST Interference-Resistant helper phage and the *SOLR* Strain (Stratagene) of *E. coli* as host for excision plating with the manufacturer's protocols. Plasmid excision was done by Mr. David Yoder. The SKCPc115b plasmid DNA then was isolated and purified by HPLC (Sambrook et al., 1989; Merion and Warren, 1989). Plasmid purification was done by Dr. Irma Pirtle. Both strands of the 1,694-bp insert of the SKCPc115b plasmid DNA were sequenced predominantly by Mr. David Yoder and Dr. Irma Pirtle using a primer-

based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, TX). The intact SKCPc115b plasmid DNA was used as the template DNA for either conventional dideoxynucleotide chain termination (Tabor and Richardson, 1987) sequencing with [α-^{35}S]dCTP and T7 Sequenase 2.0 (Amersham) or terminator cycle sequencing (Fan et al., 1996) with ThermoSequenase and [α-^{33}P]-labeled dideoxynucleotside triphosphates (Amersham). Alignment and analysis of the DNA sequences were done with DNASIS software from Hitachi.

Screening of a Cotton Genomic Library

In collaboration with Mr. David Yoder and Drs. Robert and Irma Pirtle, a cotton (*Gossypium hirsutum, cv. Acala SJ-5*) genomic library harbored in the lambda vector EMBL3, generously provided by Dr. David M. Anderson of Phytogen Seed, Placentia, CA (Grula et al., 1995) was screened by the plaque purification method of Benton and Davis (1977) using the heterologous probe derived from the *Arabidopsis* cDNA clone TE 3-2 as described above. A small culture (50 ml) of the lambda-sensitive host bacterial stain *E. coli K802* in LB broth supplemented with 0.2% maltose and 10 mM $MgSO_4$ was grown in a shaker/incubator at 37°C overnight. In order to over-represent the entire cotton genome, approximately 1 x 10^7 plaques of lambda EMBL3 containing the cotton genomic library were screened. The bacterial host cells were centrifuged down and resuspended in 10 mM $MgSO_4$. They were then infected with lambda EMBL3 containing the cotton genomic library to achieve a plaque density of about 3 x 10^5 plaque-forming units (pfu) per plate and incubated for 15 minutes at 37°C.

Four ml of top agarose (48°C) was mixed with the infected host cells and spread evenly onto NZY bottom agar plates (0.5% yeast extract, 1% NZY amine, 0.5% NaCl

and 0.02 M $MgSO_4$). The recombinant phages were incubated on the bacterial lawn at 37°C overnight. The lysed plaques reached a diameter of about 1.5 mm and made contact with one another overnight. The plates were then chilled at 4°C prior to making membrane replicas. Hybond N^+ nylon membranes (Amersham) pre-wetted in 1 M NaCl were overlayed on each plate for four minutes to adhere the phage. The membranes were successively transferred to denaturing solution (1.5 M NaCl and 0.5 M NaOH), neutralizing solution (0.5 M NaCl, 0.1 M Tris-HCl, pH 7.3), and rinsing solution (2xSSC, 0.2 mM Tris-HCl, pH 7.5). The membranes were then dried and baked in a vacuum oven at 80°C for two hours.

Prehybridization of nylon membrane filter replicas was done at 55°C for four hours in a solution consisting of 6XSSC (1XSSC is 15 mM sodium citrate, pH 7.0, 150 mM NaCl), 0.5% SDS, 5X Denhardt 's reagent and denatured sheared salmon sperm DNA (100 μg/ml). Hybridization was done overnight at 55°C in a solution consisting of 20 mM Tris-HCl, 2 mM EDTA, 2.5 mM sodium pyrophosphate, pH 8.0, 5X Denhardt's reagent, 6XSSC, 0.5% SDS, denatured sheared salmon sperm DNA (100 μg/ml) and the ^{32}P-labeled probe (Anderson and Young, 1985; Ausubel et al., 1987; Sambrook et al., 1989). After hybridization, the membranes were rinsed once at room temperature with 2XSSC. Subsequently the membranes were washed three times at 55°C (once with 2XSSC and 0.1% SDS for 30 minutes and twice with 1XSSC and 0.1% SDS for 30 minutes). Then the membranes were air-dried and exposed to Kodak X-OMAT film with an intensifying screen at -90°C. Several plaques exhibiting intense positive signals were subjected to successive rounds of plaque purification until all plaques on a single plate

hybridized to the labeled probe, indicative of plaque purity. The DNAs from promising genomic clones were initially analyzed by a minilysate phage DNA preparation (Sambrook et al., 1989).

Minilysis Preparation of Lambda Phage DNA

Selected phage clones giving strong positive signals were mixed with SM buffer (20 mM Tris-HCl (pH 7.5), 20 mM $MgSO_4 \cdot 7H_2O$, 100 mM NaCl and 2% gelatin) and *E. coli K802* host cells and incubated at 37°C for 30 minutes. The cultures were transferred to 25 ml of LB media supplemented with 0.2 maltose and 10 mM $MgSO_4$ and incubated at 39°C with shaking until lysis occurred (usually overnight). Chloroform (1 ml) was added to the lysates and they were shaken at 39°C for 15 minutes to ensure complete lysis. Following centrifugation at 7,500 rpm (6,600xg) in a SA-600 rotor to remove bacterial debris, DNase I and RNase A, each at a concentration of 2 μg/ml, were added to the supernatants and incubated at 37°C for 40 minutes. An equal volume of a 20% PEG (polyethyleneglycol) and 2 M NaCl solution were added to the supernatants and incubated on ice for one hour. The precipitated phage particles were recovered by centrifugation at 12,000 rpm (16,800xg) in the SA-600 rotor for 20 minutes at 4°C. Pellets were collected and resuspended in 400 μl STE buffer (20 mM Tris-HCl, 250 mM NaCl, and 1 mM Na_2EDTA, pH 7.5).

Phenol/chloroform extractions were done to remove the proteins. The phage DNAs were precipitated by two volumes of 95% ethanol and 0.1 volume of 2.5 M sodium acetate (pH 5.2), and the DNA pellets were redissolved in TE buffer. The phage DNAs from the minilysate preparations were further analyzed by physical mapping

procedures involving restriction endonuclease digestion, agarose gel electrophoresis, and alkaline blot hybridization (Reed and Mann, 1985). Based upon the results from alkaline blot hybridization of the phage mini-lysates, a single clone designated as LCPg59 was chosen for large scale phage preparation (Tiemeir et al., 1977; Blattner et al., 1978).

Large Scale Preparation of Lambda Phage DNA

A volume of 2.0 liters of LB broth supplemented with 0.2% maltose and 10 mM $MgSO_4$ in five 2-liter flasks, each containing 400 ml, were prepared for inoculation. The lambda genomic clone designated as LCPg59 encompassing the putative *PATE* gene at a M.O.I (multiplexity of infection) of 0.3 was mixed with 1 ml SM buffer and *E. coli* host *K802* cells and the phage allowed to adhere at 39°C with shaking for 30 minutes. The 400 ml cultures were inoculated and incubated at 39°C with aeration overnight for complete lysis. To ensure complete lysis, 8 ml of chloroform was added to each 400 ml culture and the mixture was shaken at 39°C for 10 minutes. RNase A and DNase I were added to each 2-liter flask to a final concentration of 2 μg/ml each and the mixture incubated at room temperature for 30 minutes. To precipitate the bacterial debris, NaCl was added to each 400 ml culture to a final concentration of 1 M NaCl, shaken for 10 minutes, and then held on ice for two hours. Following centrifugation in the Sorvall GS-3 rotor at 6,000 rpm (5,200xg) at 4°C for 30 minutes to remove the bacterial debris and chloroform, polyethylene glycol (PEG8000) was added to a final concentration of 10% (w/v) in order to precipitate the phage at 4°C overnight. The phage pellets were centrifuged in the Sorvall GS-3 rotor at 6,000 rpm (5,200xg) and redissolved in SM buffer (20 mM Tris-HCl (pH 7.5), 20 mM $MgSO_4 \cdot 7H_2O$, 100 mM NaCl and 2% gelatin).

The phage particles were isolated and purified by CsCl gradient ultracentrifugation (Sambrook et al., 1989) in a Beckman LS-65 ultracentrifuge using a Ti 75 rotor at 38,000 rpm (22,000xg) for 24 hours. The lambda phage band was removed and the CsCl was dialyzed out against cold dialysis buffer (0.1 M Tris-HCl (pH 8.0), 0.3 NaCl) at 4°C overnight, with several buffer changes to dilute the CsCl. The phage DNA was isolated by phenol/chloroform extraction and subsequent ethanol precipitation. The yield of the purified LCPg59 phage DNA was determined to be about 1.5 mg by A_{260} measurement.

Physical Mapping of Phage containing Putative *PATE* Gene and Restriction Fragment Selection

Lambda LCPg59 DNA was digested with various restriction endonucleases with minimal cleavage of the lambda EMBL3 arms. Both single and double restriction endonuclease digestions were required. The restriction enzymes were used as specified by the manufacturer. To set up digestion of DNA with a restriction enzyme in a volume of 20 μl, 2.0 μl of the appropriate 10X buffer was added to the calculated amount of water in a sterile Eppendorf tube. The DNA sample (1-4 μg) and 10-20 units of restriction enzyme were added into the reaction and mixed by vortexing. The reaction was incubated at the appropriate temperature for 3-24 hours, depending on the activity of the enzyme. The DNA fragments were subsequently resolved by electrophoresis on agarose gels of various gel compositions (0.5%, 0.8%, 1.0%, and 2.0%) to maximize the optimal separation. To determine the sizes of very small restriction fragments (less than 1.0 kb), a 4% agarose gel consisting of NuSieve:agarose LE (3:1) was used. The gels were electrophoresesed in a horizontal gel apparatus submerged in TAE buffer (40 mM Tris-acetate and 2 mM EDTA, pH 8.5). The restriction enzyme digests were mixed with

loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 1% SDS, 0.1 M Na_2EDTA (pH 8.0), and 20% Ficoll 400) and loaded into wells in the gels. Electrophoresis was usually carried out in TAE buffer at 2.3 volts/cm. The resolved lambda LCPg59 DNA fragments were transferred onto positively charged Hybond N^+ nylon membranes (Amersham) in 0.4 N NaOH by the alkaline blotting procedure, with a 0.25 M HCl depurination step (Reed and Mann, 1985).

Prehybridization was done at 55°C for four hours in a solution consisting of 6X SSC, 0.5% SDS, 5X Denhardt's reagent and denatured sheared salmon sperm DNA (100 μg/ml) (Sambrook et al., 1989). Hybridization was done overnight in a solution of 20 mM Tris-HCl, 2 mM EDTA, 2.5 mM sodium pyrophosphate (pH 8.0), 5X Denhardt's Reagent, 6XSSC, 0.5%SDS, denatured sheared salmon sperm DNA (100μg/ml) and ^{32}P-labeled heterologous *Arabidopsis* probe generated from the 1.4 kb *Sal*I/*Eco*RV fragment of the *Arabidopsis* cDNA clone designated TE 3-2. The *Arabidopsis* DNA fragment was used as the template to generate ^{32}P-labeled DNA fragments by random priming. After hybridization, the nylon membrane replicas were washed at 55°C in 2XSSC, 0.1% SDS, once for 30 minutes and twice in 1X SSC, 0.1% SDS for 30 minutes. Then the membranes were air-dried for autoradiography.

Subcloning a DNA fragment encompassing the *PATE* gene

To select an appropriate-sized restriction fragment for subcloning, it is necessary to construct the physical map of a genomic clone. Based upon the physical mapping data, a 5.2-kb *Xho*I fragment was selected for subcloning because it seemed to encompass the entire coding and flanking regions of the putative *PATE* gene. Therefore, this 5.2-kb *Xho*I fragment from the LCPg59 clone, which hybridized to the *Arabidopsis PATE* probe,

was selected for isolation and subsequent cloning into the multiple cloning site of the plasmid vector pGEM7Zf(+). The LCPg59 phage DNA and pGEM7Zf (+) vector DNA were both digested to completion with *Xho*I. Agarose gel electrophoresis on a 0.8% gel was done to check the completeness of digestion and to estimate the amount of DNA. The linearized vector was then extracted with phenol/chloroform (1:1) and precipitated in 100% ethanol. The linear pGEM7Zf(+) vector was dephosphorylated by treatment with calf intestinal alkaline phosphatase (CIAP) in order to minimize occurrence of vector self-ligation. The 20 µl (4.5 µg) of phenol/chloroform-extracted linear pGEM7Zf(+) vector DNA was mixed with 2.5 µl of 10X CIAP buffer (500 mM Tris-HCl (pH 8.5), 1 mM EDTA) and 1 µl (0.1 unit) of CIAP and 1.5 µl of autoclaved water for a total reaction volume of 25 µl. The reaction mixture was incubated for one hour at 37°C, followed by extraction with phenol and chloroform and precipitation with 100% ethanol. The 5.2-kb *Xho*I fragment of the LCPg59 genomic clone and the linearized and dephosphorylated pGEM7Zf(+) vector fragment were purified by electrophoresis on a 0.8% agarose gel and extracted from the gel using the QIAquick Gel Extraction Kit (QIAGEN) using the manufacturer's protocol. Briefly, the 5.2-kb *Xho*I fragment and the linear pGEM7Zf(+) vector fragment were cut from the gel, the gel slice was weighed, and then three volumes of Buffer QX1 (QIAGEN) to one volume of gel were added. The mixture then was incubated at 50°C for 10 min until the gel slice completely melted. The mixture was applied to the QIAquick column and centrifuged in a microfuge for one minute in order to bind DNA to the QIAquick membrane. The flow-through was discarded and then 750 µl of Buffer PE (QIAGEN) was added to the column to wash out

all undesired trace materials and centrifuged for one minute. The flow-through was discarded and the column was centrifuged for an additional minute. Water was added to the center of the QIAquick column and centrifuged for one minute in order to elute the DNA. Water was dispersed directly onto the QIAquick membrane for complete elution of bound DNA. The purity of the purified 5.2-kb *Xho*I fragment and the linear pGEM7Zf(+) DNA were determined by electrophoresis on a 0.8% agarose test gel. The fragment yields were estimated by comparison with standard marker bands with known sizes and amounts of DNA. The 5.2-kb *Xho*I fragment was subcloned into the *Xho*I site of the pGEM7Zf(+) plasmid vector. The 5.2-kb *Xho*I fragment and dephosphorylated linear pGEM7Zf(+) DNA were mixed in a 3:1 ratio (insert:vector). Four (4) μl of 5x ligase buffer (250 mM Tris-HCl (pH. 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethene glycol-8000), 1.0 μl (0.1 unit) T4 DNA ligase, and autoclaved water in a final volume of 20 μl were added to the mixture. The ligation mixture was then incubated at room temperature for about one hour. The resulting recombinant plasmid DNAs were used for transforming electrocompetent *E. coli DH5α* cells.

Preparation and Transformation of Electrocompetent *E. coli DH5α* cells

For preparation of electrocompetent cells, 200 μl of a *E. coli DH5α* glycerol stock was inoculated into 50 ml of LB media (Luria-Bertani media; 10 grams of Bactotryptone, 5 grams of Bactoyeast Extract, and 5 grams of NaCl per liter), and the culture was grown overnight at 37°C with shaking at 250 rpm on a New Brunwick platform shaker/incubator. A large-scale (1 liter) culture in LB was then done in two 2-liter flasks containing 500 ml each by inoculation with 15 ml of the overnight culture.

The large scale cultures were grown at 37°C with aeration, and the A_{600} readings of the cultures were monitored until they were between 0.5 and 1.0, and the cells were then chilled on ice. The bacterial cells were harvested by centrifugation in a Sorvall GS-3 rotor at 5,500 rpm (2,790xg) for 15 minutes at 4°C. The supernatants were removed and the cells were resuspended in one volume of sterile cold water. The cells were then resuspended twice in a 0.5 volume of sterile ice-cold water and once in a 0.02 volume of sterile ice-cold water, with centrifugation in a Sorvall GS-3 rotor at 5,500 rpm (2,790xg) for 15 minutes at 4°C. The final resuspension was in a 0.003 volume of cold filter-sterilized 10% glycerol. The electrocompetent cells were then aliquoted into microfuge tubes and frozen at -90°C.

To transform *E. coli DH5α* cells, approximately 100 ng of the recombinant plasmid DNA in the vector pGEM7Zf(+) was mixed with 35 µl of the electrocompetent cells. A BTX disposable cuvette (1 mm gap) was then filled with the above transformation mixture, and chilled on ice for one minute prior to electroporation. An electroporator apparatus (BTX Electro Cell Manipulator model 395) was set on HV mode/3kV, and a single 1.5 kV electrical pulse was applied, with a resulting a field strength of 15.0 kV/cm with an exponential decay constant of approximately 5-6 milliseconds. Immediately following the electrical pulse, 960 µl of LB media was added to the cuvette and gently pipetted up and down for through mixing. The cell suspension was then transferred to a polypropylene tube and incubated at 37°C for one hour in a shaker/incubator at 225 rpm. Aliquots of the cell suspension (50, 100 and 200 µl portions) were plated on LB/amplicillin plates containing 10 µl of 10 mM IPTG

(isopropylthio-β-D-galactoside) and 50 µl of 2% X-gal (5-bromo-4-chloro-3-indoyl- β-D-galactoside) and incubated overnight at 37°C. After overnight incubation, it was possible to distinguish white colonies containing the recombinant plasmids from the blue colonies containing non-recombinant plasmids. The white colonies were picked and streaked onto fresh LB/amplicillin plates containing IPTG and X-gal to ensure that they were recombinant clones and to serve as master plates. The master plates, which served as the primary source of the transformed cells, were incubated at 37°C overnight and then stored at 4°C.

The recombinant cells were screened by several methods to confirm that the clones contained the desired recombinant plasmids (Sambrook et al., 1989). One method was to use a small-scale plasmid DNA preparation procedure ("miniprep") and identify the desired clones by restriction digestion and gel electrophoresis analysis of the plasmid DNA. A single white colony on a master plate was picked and inoculated in five ml of LB media containing amplicillin. The inoculated cultures were incubated overnight at 37°C with shaking at 250 rpm. The overnight cultures were placed into a 1.5 ml microfuge tubes and centrifuged at 12,000xg for one minute in a microcentrifuge. The supernatants were removed and bacterial pellets were left as dry as possible. The pellets were resuspended by vortexing after adding 100 µl of ice-cold cell resuspension solution (25 mM Tris-HCl (pH 8.0), 10 mM EDTA and 50 mM glucose) and incubated at room temperature for five minutes. A 200 µl volume of freshly prepared cell lysis solution (0.2 N NaOH and 1% SDS) was added, mixed by inversion, and incubated for five minutes on ice. This step denatured the bacterial chromosomal DNA. Ice-cold potassium acetate

solution, pH 4.8 (150 µl) was added to neutralize the lysate and mixed by inversion for 10 seconds and then incubated on ice for five minutes. After incubation, the bacterial DNA was pelleted by centrifugation at 12,000xg for five minutes, leaving the plasmid DNA in the supernatants. The supernatants were transferred to fresh tubes and 0.5 µl of DNase-free pancreatic RNase A (100 µg/ul) was added and incubated for five minutes at room temperature. An equal volume of phenol/chloroform/isoamyl alcohol (24/24/1) was added and vortexed for one minute and centrifuged at 12,000xg for five minutes.

The upper aqueous phases were transferred to fresh tubes and equal volumes of chloroform/isoamyl alcohol (24:1) were added and vortexed for 30 seconds and then centrifuged at 12,000xg for two minutes. The upper aqueous phases were transferred to fresh tubes and 0.1 volume of 2.5 M sodium acetate (pH 5.2) and 2.5 volumes of ice-cold 100% ethanol were added. The supernatants were mixed and allowed to precipitate at -90°C for one hour. The pellets were collected by centrifugation at 12,000 xg for five minutes and rinsed with ice-cold 70% ethanol and dried under vacuum. The isolated plasmid DNAs were digested with the restriction enzyme *Xho*I. The reaction volumes of 20 µl contained 10 µl of plasmid DNA, 1 µl of *Xho*I (10 units/µl) and 2 µl of the appropriate 10x buffer. The digested plasmid DNAs were then fractionated by agarose gel electrophoresis to identify those subclones containing the 5.2-kb *Xho*I fragment encompassing the *PATE* gene. One subclone, designated pCPg59, was selected for a large scale plasmid preparation as previously described. Both DNA strands of the cotton insert in pCPg59 were sequenced by Dr. David Yoder and Dr. Irma Pirtle using a primer-based approach with synthetic oligonucleotide primers from Biosynthesis Inc., Lewisville, TX. The intact pPCg59 plasmid DNA was used as the template for

terminator cycle sequencing by the method of Fan et al. (1996) using ThermoSequenase and ^{33}P-labeled dideoxy nucleoside triphosphates from Amersham. Analysis of the DNA sequences and the deduced amino acid sequence of the presumptive PATE preprotein and comparisons with other acyl-ACP thioesterase genes, cDNAs, and predicted amino acid sequences was done by Dr. Irma Pirtle of our laboratory with DNASIS software from Hitachi.

Determination of Gene Copy Number

The copy number (reiteration frequency) of the *PATE* gene in the allotetraploid cotton genome was analyzed by genomic blot hybridization. Two cotton genomic DNAs, from cultivars *Pima* and *TM1*, were kindly provided by Dr. Sukuma Saha of the USDA Agricultural Research Station, Mississippi State University, MS. These two genomic DNA samples were digested with several restriction enzymes. The lambda DNA from the genomic clone LCPg59 containing the *PATE* gene was also digested with several restriction enzymes. The resulting fragments were analyzed by agarose gel electrophoresis and alkaline blotting for comparison of the fragment sizes between the cloned LCPg59 DNA fragments and the actual cotton genomic DNA fragments. The gel was then soaked in depurination solution (0.25 M HCl) for 15 minutes. After depurination, a membrane replica of the gel was made by blotting onto Hybond N^+ nylon membrane (Amersham).

Prehybridization was done at 55°C for two hours in a solution consisting of 6XSSC (1XSSC is 150 mM NaCl and 15 mM sodium citrate, pH 7.0), 0.5% SDS, 5X Denhardt's reagent and denatured sheared salmon sperm (100 μg/ml). Hybridization was done overnight in a solution of 6XSSC, 0.5%SDS, 5X Denhardt's Reagent, 20 mM Tris-

HCl (pH 8.0), 2 mM Na_2 EDTA (pH 7.5), 2.5 mM sodium pyrophosphate (pH 8.0), denatured sheared salmon sperm (100 μg/ml) and ^{32} P-labeled homologous cotton *PATE* cDNA probe. After hybridization, the nylon membranes were rinsed at room temperature in 2XSSC. Subsequently, the filters were washed three times at 55°C (once with 2XSSC and 0.1% SDS for 30 minutes and twice with 1XSSC and 0.1% SDS for 30 minutes).

Results

Identification of the cotton *PATE* cDNA and gene

A cotton cDNA library generated from cotyledon mRNA of 48-hour dark-grown cotton seedlings (*Gossypium hirsutum, cv. Deltapine 62*) was screened with a heterologous *Arabidopsis* probe by the plaque hybridization procedure of Benton and Davis (1977). A 1.4-kb *Sal*I/*Eco*RV restriction fragment was isolated from an *Arabidopsis* cDNA clone designated TE 3-2 encoding a long-chain acyl-ACP thioesterase (Dörmann et al., 1995). The fragment was used as a template to generate ^{32}P-labeled DNA fragments by random priming as previously described. A pBluescript SK(-) phagemid derivative designated pSKCPc115b, containing a 1.7-kb cotton *PATE* cDNA insert, was excised *in vivo* using Stratagene ExAssist Interference-Resistant helper phage and the *SOLR* strain of *E. coli* as host for excision plating. This work was done by Dr. David Yoder of this laboratory. The pSKCPc115b plasmid DNA was isolated and purified by HPLC (Sambrook et al., 1989; Merion and Warren, 1989) by Dr. Irma Pirtle. Both strands of the insert of pSKCPC115b were sequenced predominantly by Mr. David Yoder and Dr. Irma Pirtle using a primer-based approach with synthetic oligonucleotide primers (Pirtle et al., 1999). The cDNA insert in the pBluescript SK(-) phagemid

Figure 3. The nucleotide and deduced amino acid sequences for the cotton *PATE* cDNA clone SKCPc115b (Pirtle et al., 1999). The nucleotide sequence has been assigned the GenBank Accession Number AF034266. The numbering on the right refers to nucleotide residues, whereas the numbering on the left denotes amino acid residues. The first nine nucleotides (shown in parentheses), corresponding to the three amino acids MVA at the N-terminus of the PATE preprotein were determined from the DNA sequence of the corresponding region of the *PATE* gene (Yoder et al., 1999). The sequence corresponding to the presumptive poly(A) polymerase near-upstream element (nt 1685 to 1690) is underlined. The stromal-targeting domains (Cline and Henry, 1998) of the PATE preprotein corresponding to the N-terminal region (MVATAVTSAFF, amino acids 1-11), the variable middle region (TSSPDSSDSKNKKLGSIKSKPSVSSGS, residues 14-40), and the carboxy-proximal region (VKANA, residues 43-47) are denoted by underlining the respective amino acid domains. The putative transient peptide cleavage site is probably L84, when compared with other thioesterase preprotein sequences (Gavel and von Heijne, 1990; Dörmann et al., 1995; Cline and Henry, 1996). The catalytic cysteine required for formation of the covalent thiol enzyme intermediate and the histidine involved in general base catalysis in the plant acyl-ACP thioesterases (Yuan et al., 1996) would most likely correspond to the homologous residues C347 and H312 in the cotton amino acid PATE sequence. The two other conserved histidine residues would correspond to the homologous H167 and H377 residues.

```
1 (M  V  A)  T  A  V  T  S  A  F  F  P  V  T  S  S  P  D  S  S
 (ATGGTTGCT)ACTGCTGTGACATCGGCGTTTTTCCCAGTCACTTCTTCACCTGACTCCTCT +60

21  D  S  K  N  K  K  L  G  S  I  K  S  K  P  S  V  S  S  G  S
  GACTCGAAAAACAAGAAGCTCGGAAGCATCAAGTCGAAGCCATCGGTTTCTTCTGGAAGT +120

41  L  Q  V  K  A  N  A  Q  A  P  P  K  I  N  G  T  V  A  S  T
  TTGCAAGTCAAGGCAAATGCTCAAGCACCTCCGAAAATAAACGGCACTGTGGCGTCGACG +180

61  T  P  V  E  G  S  K  N  D  D  G  A  S  S  P  P  P  R  T  F
  ACTCCCGTGGAAGGTTCCAAGAACGATGACGGTGCAAGTTCCCCTCCTCCTAGGACGTTT +240

81  I  N  Q  L* P  D  W  S  M  L  L  A  A  I  T  T  I  F  L  A
  ATCAACCAGTTACCTGATTGGAGCATGCTTCTTGCTGCTATCACAACCATTTTCTTGGCT +300

101 A  E  K  Q  W  M  M  L  D  W  K  P  R  R  P  D  M  V  I  D
  GCTGAGAAGCAGTGGATGATGCTTGATTGGAAGCCGAGGCGGCCTGACATGGTCATTGAT +360

121 P  F  G  I  G  K  I  V  Q  D  G  L  V  F  S  Q  N  F  S  I
  CCGTTTGGCATAGGGAAGATTGTTCAGGATGGTCTTGTTTTCAGTCAGAACTTCTCGATT +420

141 R  S  Y  E  I  G  A  D  Q  T  A  S  I  E  T  L  M  N  H  L
  AGATCATATGAGATAGGCGCTGATCAAACAGCATCCATAGAGACACTAATGAATCATTTA +480

161 Q  E  T  A  I  N  H  C  R  S  A  G  L  L  G  E  G  F  G  A
  CAGGAAACAGCTATAAATCATTGTCGAAGTGCTGGACTGCTTGGAGAAGGTTTTGGTGCA +540

181 T  P  E  M  C  K  K  N  L  I  W  V  V  T  R  M  Q  V  V  V
  ACACCTGAGATGTGCAAGAAGAACCTAATATGGGTTGTCACACGGATGCAAGTTGTGGTT +600

201 D  R  Y  P  T  W  G  D  V  V  Q  V  D  T  W  V  S  A  S  G
  GATCGCTATCCTACTTGGGGTGATGTTGTTCAAGTCGACACTTGGGTCAGTGCATCGGGG +660

221 K  N  G  M  R  R  D  W  L  V  S  N  S  E  T  G  E  I  L  T
  AAGAATGGCATGCGAAGAGATTGGCTTGTCAGCAATAGTGAAACTGGTGAAATTTTAACA +720

241 R  A  T  S  V  W  V  M  M  N  K  L  T  R  R  L  S  K  I  P
  CGAGCCACAAGTGTATGGGTGATGATGAATAAACTGACTAGAAGGTTATCTAAAATCCCA +780

261 E  E  V  R  G  E  I  E  P  F  F  M  N  S  D  P  V  L  A  E
  GAAGAGGTTCGAGGGGAAATAGAACCTTTTTTTATGAATTCAGATCCTGTTCTGGCTGAG +840

281 D  S  Q  K  L  V  K  L  D  D  S  T  A  E  H  V  C  K  G  L
  GATAGCCAGAAACTAGTGAAACTCGATGACAGCACAGCTGAACACGTGTGCAAAGGTTTA +900

301 T  P  K  W  S  D  L  D  V  N  Q  H  V  N  N  V  K  Y  I  G
  ACTCCTAAATGGAGCGACTTGGATGTCAACCAGCATGTCAATAATGTGAAGTACATTGGC +960

321 W  I  L  E  S  A  P  L  P  I  L  E  S  H  E  L  S  A  L  T
  TGGATCCTTGAGAGTGCTCCATTACCAATCTTGGAGAGTCACGAGCTTTCCGCCTTGACT +1020

341 L  E  Y  R  R  E  C  G  R  D  S  V  L  Q  S  L  T  T  V  S
  CTGGAATATAGGAGGGAGTGCGGGAGGGACAGCGTGCTGCAGTCACTGACCACTGTGTCT +1080

361 D  S  N  T  E  N  A  V  N  V  G  E  F  N  C  Q  H  L  L  R
  GATTCCAATACGGAAAATGCAGTAAATGTTGGTGAATTTAATTGCCAACATTTGCTCCGA +1140

381 L  D  D  G  A  E  I  V  R  G  R  T  R  W  R  P  K  H  A  K
  CTCGACGATGGAGCTGAGATTGTGAGAGGCAGGACCCGATGGAGGCCTAAACATGCCAAA +1200
```

```
401 S  S  A  N  M  D  Q  I  T  A  K  R  A TER
   AGTTCCGCTAACATGGATCAAATTACCGCAAAAAGGGCATAGAAATCCAAGTAATCTCAT +1260

   TGCTGTGTGTAGTATCTATCGTGCTCTTTTCGGATTTATATACATATATTCCTTATGATT +1320

   ATTAGTCTTCCTTTGAGAAAAAAAAAGGGGGTTGTAATTAGGCTTGTTTAGGAGTCGGGT +1380

   TTTCGTACATAGCCTTGTAAGGCTCAGCTCGTATGACCCGAGCCTCGGACACGGATTTTG +1440

   TGAAGTTGGGCCCGTGCCCTAACCAGCATAGGCTCTTTCCATGGAAAGGTGGGTCTGCTT +1500

   TTGAAAAATTGAATAGCCATGTGAGATGGCTCTCTCCCTACATTATGGGCTTTTAACCAG +1560

   TTAGAGACCGGGTAGTTTAGGATAAAATTTATCTTTAATTTGGGAGGATTTGTATATTTT +1620

   TTTTGCCTTTATTTTAACCTAAATTTGCTTATAATTATTTGGTTTTATATTTAGGTATTG +1680

   AATCAATGAAGTTTTTAAATTTT                                      +1703
```

derivative designated pSKCPc115b was determined to be 1,694 bp and the deduced amino acid sequence is shown above the nucleotide sequence in Figure 3, and the presumptive cotton PATE preprotein is 410 amino acids in length. The 3'-untranslated region of this cDNA clone is 461 bp long. This sequence has been assigned the GenBank accession number AF034266.

A cotton (*Gossypium hirsutum, cv. Acala SJ-5*) genomic library harbored in the lambda vector EMBL3 was screened by the plaque hybridization procedure of Benton and Davis (1977) with the heterologous *Arabidopsis PATE* cDNA probe as previously described. Twelve presumptive genomic clones that gave intense hybridization signals were isolated and purified by a minilysate phage preparation procedure. These twelve genomic clones encompassing putative *PATE* genes were digested with *Bam*HI and *Eco*RI, which cleave at restriction sites in the polylinker region of lambda EMBL3, as shown in the agarose gels in Figures 4 and 6. Alkaline blot hybridization of these twelve genomic DNA fragments was done using the ^{32}P-labeled probe generated from the 1.4-kb *Sal*I*/Eco*RV fragment of the *Arabidopsis* cDNA clone TE 3-2 template by random priming. Four of the twelve clones gave intense positive hybridization signals on the autoradiograms (Figures 5 and 7). The genomic clone designated PATEg59B.1b was selected as representative of the genomic clones encompassing the *PATE* gene. This cotton genomic clone was amplified and purified using a large scale phage isolation procedure. The optimum condition used for amplification was at a M.O.I of 0.3, with an optimum temperature for bacteriophage propagation of 39°C. The lambda clone PATEg59B.1b was subjected to fine physical mapping and the name was re-designated as LCPg59 for convenience.

Figure 4. Agarose gel electrophoresis of cotton genomic clones encompassing putative *PATE* genes harbored in the phage EMBL3. DNAs from the 12 genomic clones were digested with *Eco*RI and fractionated on a 0.8% agarose gel. The DNA fragments were stained with ethidium bromide prior to photography. Standard molecular weight markers are shown in the lambda *Hin*dIII and pGEM lanes. The sizes are shown on the left and right, respectively.

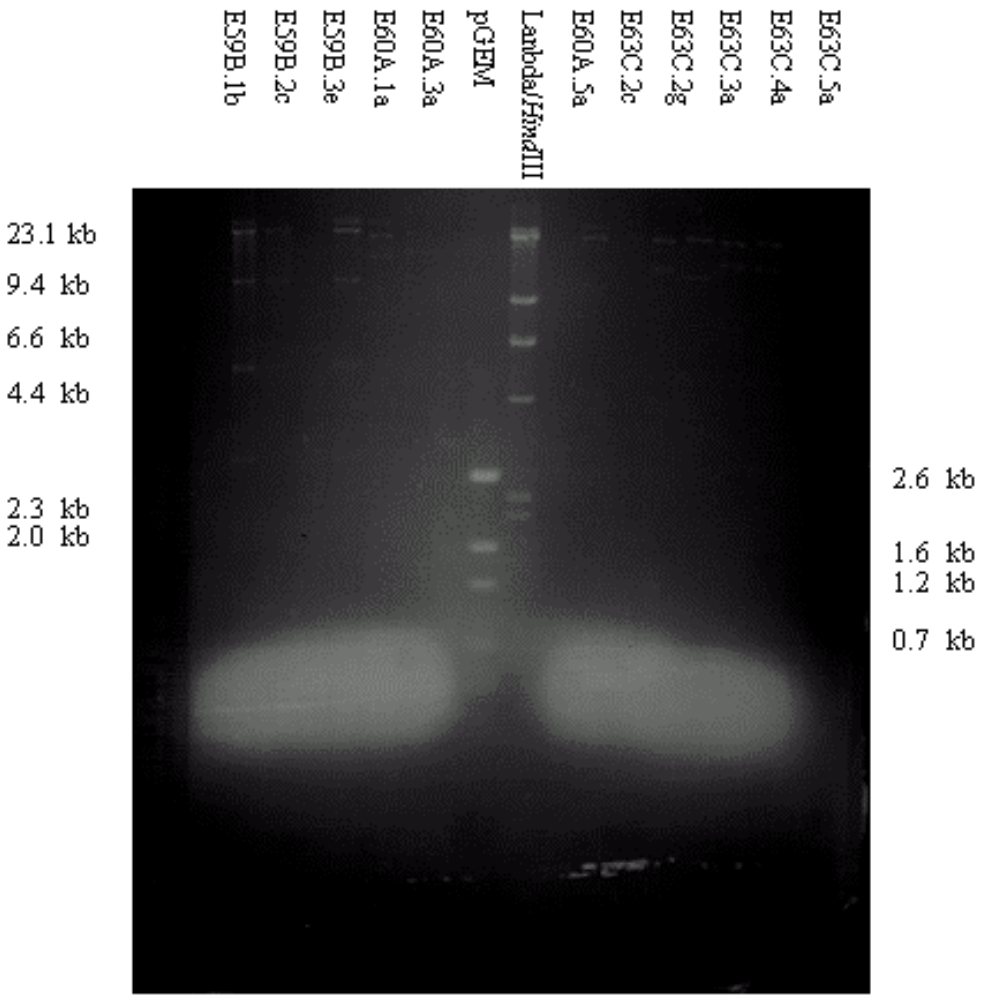

E59B.1b
E59B.2c
E59B.3e
E60A.1a
E60A.3a
pGEM
Lambda/HindIII
E60A.5a
E63C.2c
E63C.2g
E63C.3a
E63C.4a
E63C.5a
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 5. Alkaline blot hybridization analysis of cotton genomic clones containing putative *PATE* genes harbored in the phage EMBL3. The 12 genomic *PATE* clones (same as in Figure 4) were transferred to a positively-charged nylon membrane, and the DNA fragments immobilized on the membrane were hybridized with a ^{32}P-labeled probe generated from the 1.4-kb *Sal*I/*Eco*RV fragment of the *Arabidopsis* cDNA clone TE3-2.

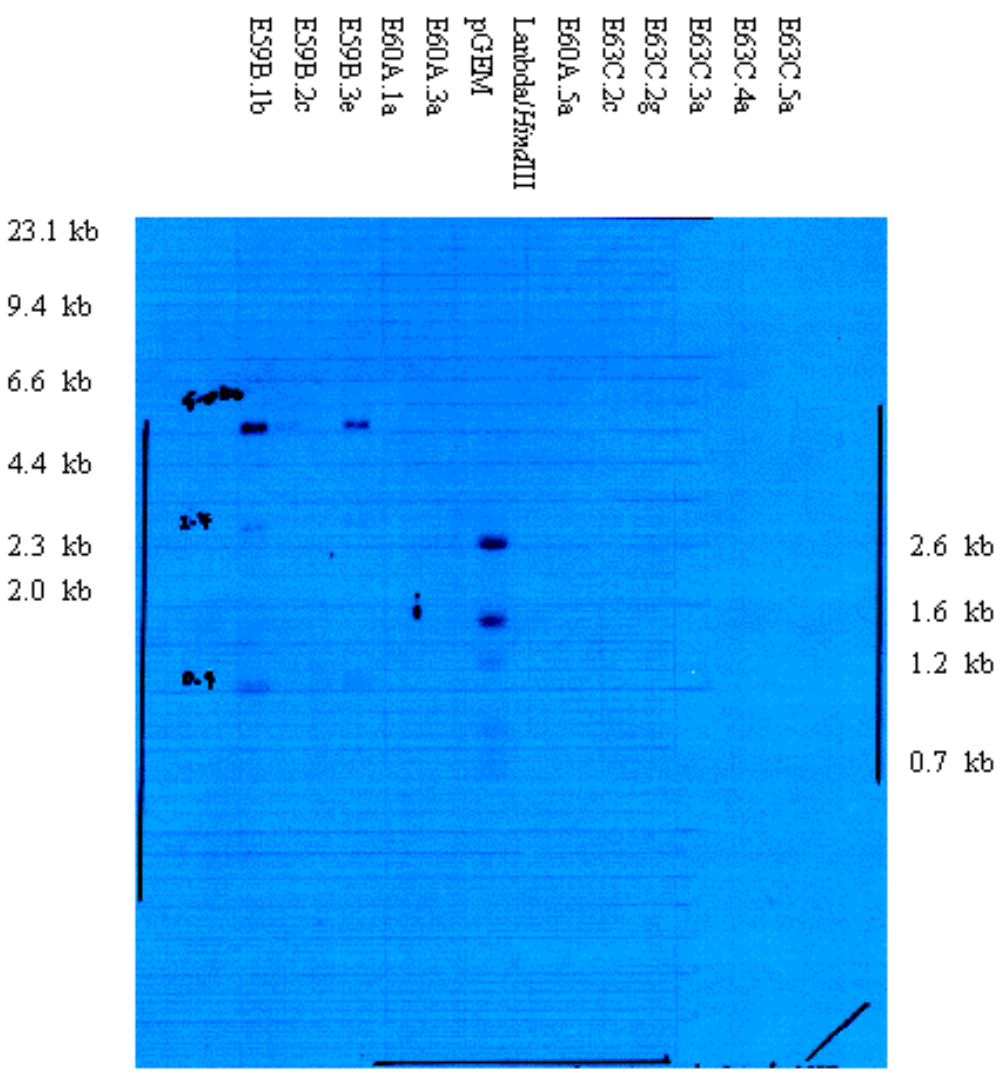

E63C.5a
E63C.4a
E63C.3a
E63C.2g
E63C.2c
E60A.5a
Lanbda/HindIII
pGEM
E60A.3a
E60A.1a
E59B.3e
E59B.2c
E59B.1b
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 6. Agarose gel electrophoresis of cotton genomic clones encompassing the putative *PATE* genes harbored in the phage EMBL3. DNAs from the 12 genomic clones were digested with *Bam*HI and fractionated on a 0.8% agarose gel. The DNA fragments were stained with ethidium bromide prior to photography. Standard molecular weight markers are shown in the lambda *Hin*dIII and pGEM lanes. The sizes are shown on the left and right, respectively.

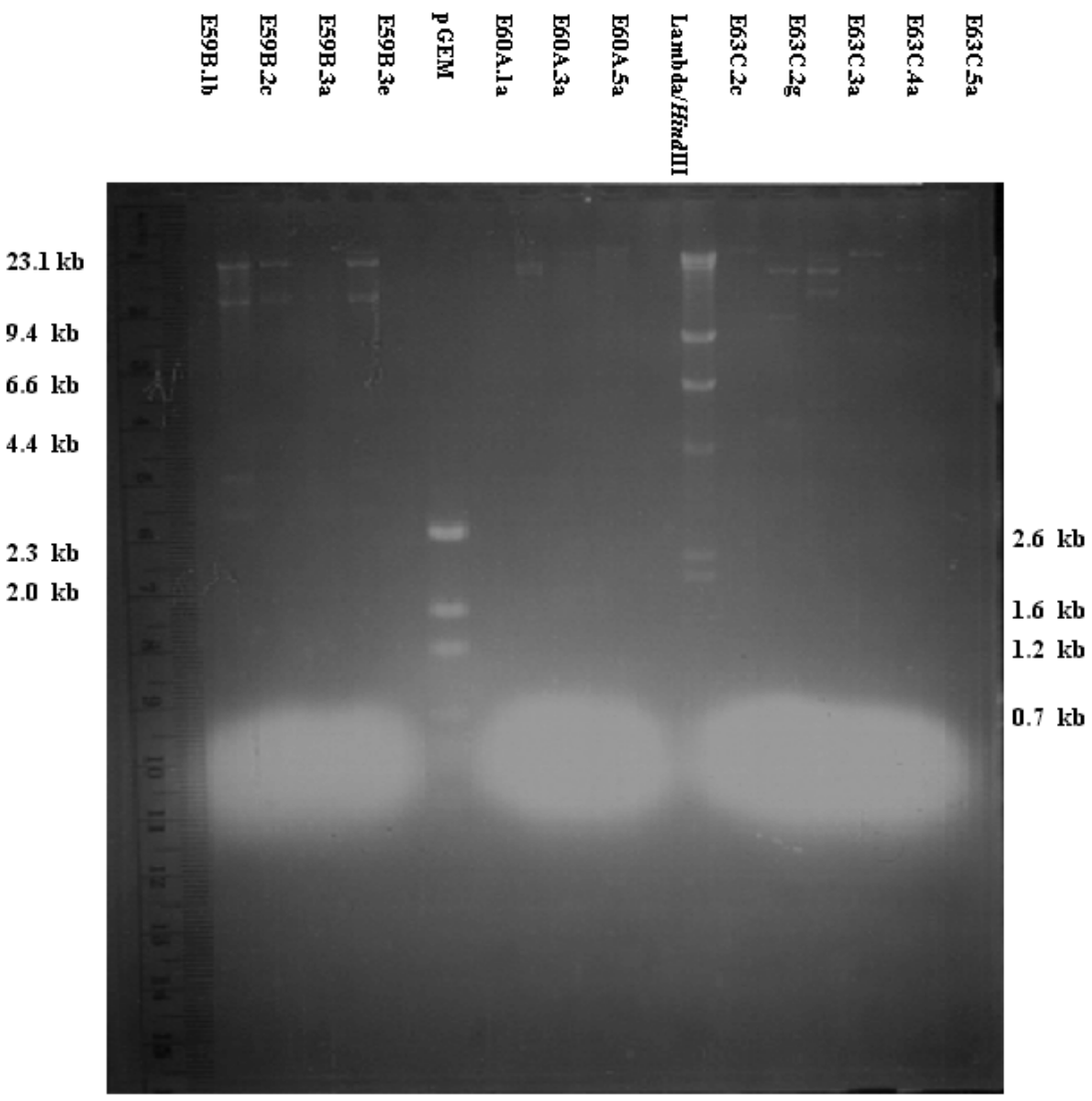
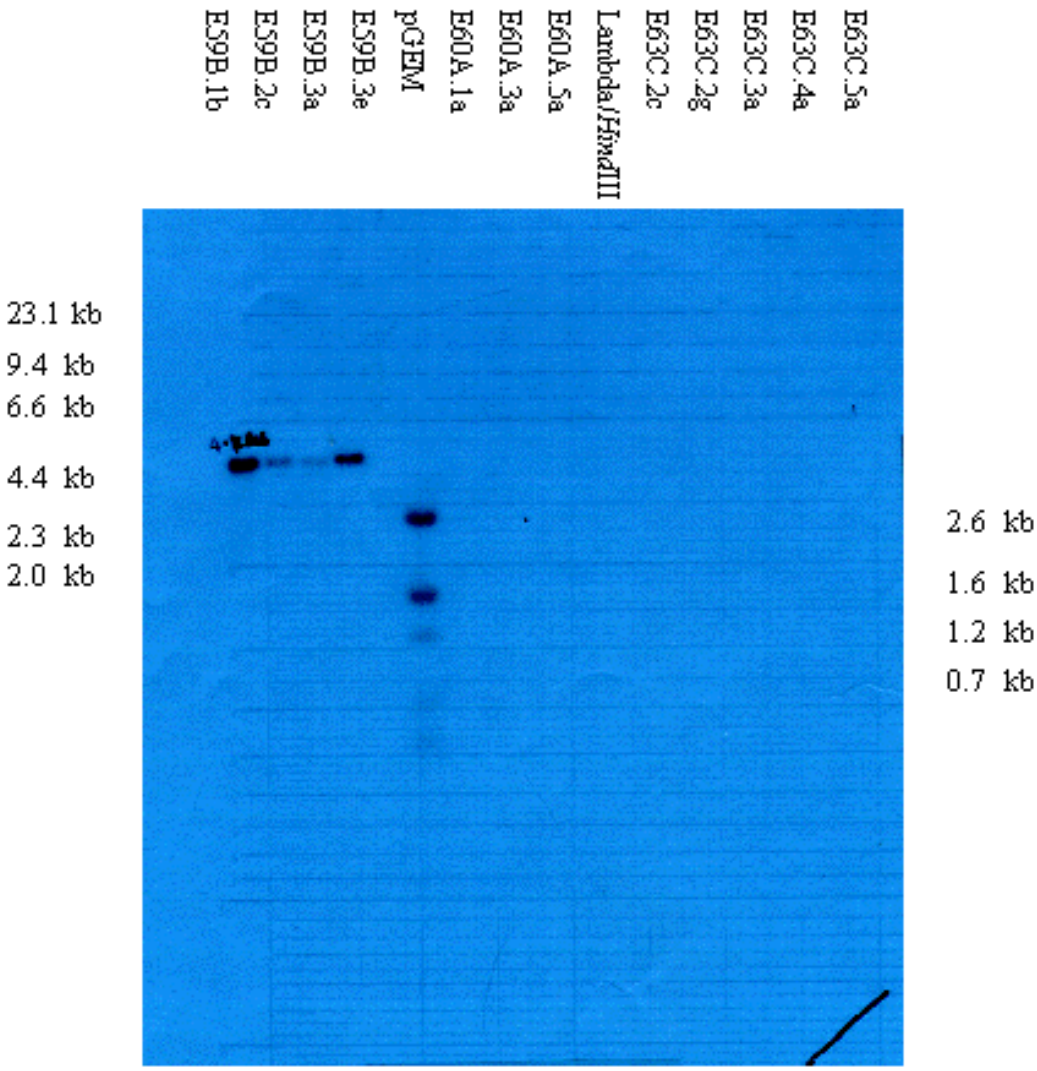

E59B.1b
E59B.2c
E59B.3a
E59B.3e
pGEM
E60A.1a
E60A.3a
E60A.5a
Lambda/HindIII
E63C.2c
E63C.2g
E63C.3a
E63C.4a
E63C.5a
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 7. Alkaline blot hybridization analysis of cotton genomic clones containing the putative *PATE* genes harbored in the phage EMBL3. The 12 genomic clones (same as in Figure 6) were transferred to a positively-charged nylon membrane and the DNA fragments immobilized on the membrane were hybridized with a ^{32}P-labeled probe generated from the 1.4-kb *Sal/Eco*RV fragment of the *Arabidopsis* cDNA clone TE3-2.


E59B.1b
E59B.2c
E59B.3a
E59B.3e
pGEM
E60A.1a
E60A.3a
E60A.5a
Lambda/HindIII
E63C.2c
E63C.2g
E63C.3a
E63C.4a
E63C.5a
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

The LCPg59 DNA was subjected to digestion with a variety of restriction endonucleases and characterized by agarose gel electrophoresis and alkaline blot hybridization. Figures 8 and 9 show examples of the agarose gel electrophoresis patterns with single and double restriction endonuclease digestions of LCPg59 DNA. Numerous agarose gel patterns were used to determine the sizes of the fragments and to construct the physical map of LCPg59 DNA spanning a cotton genomic insert of 17.4 kb. As shown in the physical map of LCPg59 in Figure 10, the 17.4-kb cotton DNA segment containing the putative *PATE* gene is between the 21-kb lambda EMBL3 left arm and the 9.6-kb lambda EMBL3 right arm.

The 5.2-kb *Xho*I fragment was found to completely encompass the *PATE* gene and was subcloned into the *Xho*I site of pGEM7Zf(+) and the recombinant plasmid designated pCPg59. The fine physical map of the 5.2-kb *Xho*I fragment in the recombinant plasmid pCPg59 encompassing the *PATE* gene, including the positions of exons and introns, is shown in Figure 11. The nucleotide sequence of 5,201 bp of the cotton insert in pCPg59 is shown in Figure 12, with the *Xho*I restriction sites at the vector/insert boundaries (Yoder et al., 1999). This sequence has been assigned the GenBank accession number AF076535. The gene has six exons (shown in capital lettering) and five introns (denoted by lower case lettering) discerned from comparison with the cotton *PATE* cDNA sequence (Pirtle et al., 1999) (GenBank Accession number AF034266). As shown above the nucleotide sequence in Figure 12, this cotton *PATE* gene would encode a presumptive thioesterase preprotein of 413 amino acids.

The 5'-flanking region of the *PATE* gene has several potential promoter/enhancer elements (Mitchell and Tjian, 1989; Young, 1991; Thomas, 1993; Guilfoyle, 1997),

Figure 8. Representative agarose gel of restriction endonuclease digestions of lambda genomic clone LCPg59 DNA encompassing the cotton *PATE* gene. The cloned LCPg59 DNA (2 μg per reaction) was digested with the restriction endonucleases as indicated. The resulting restriction fragments were resolved on a 0.8% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) of standard DNA fragments obtained by digesting lambda DNA with *Hin*dIII are shown on the left. The sizes of standard DNA markers derived from the plasmid vector pGEM-3 (Promega) are shown on the right.

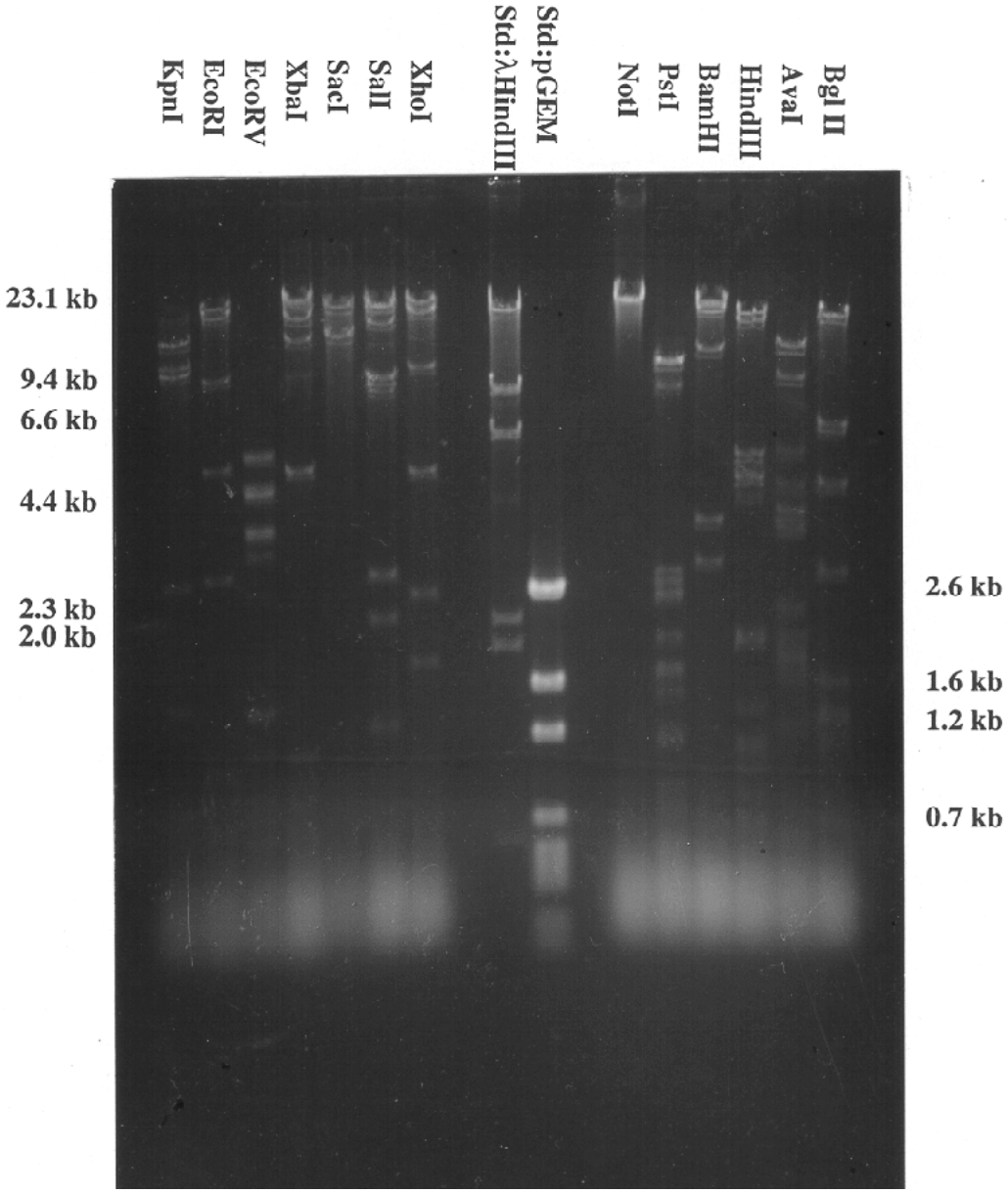

KpnI
EcoRI
EcoRV
XbaI
SacI
SalI
XhoI
Std:λHindIII
Std:pGEM
NotI
PstI
BamHI
HindIII
AvaI
Bgl II
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 9. Autoradiogram of alkaline blot of a cotton genomic clone LCPg59 encompassing a *PATE* gene. The LCPg59 DNA fragments fractionated on a 0.8% agarose gel were transferred to a positively-charged nylon membrane (Amersham Hybond N^+) in 0.4N NaOH and hybridized to a ^{32}P-labeled DNA probe generated from a heterologous *Arabidopsis* cDNA fragment by the random-priming procedure of Feinberg and Vogelstein (1983). The heterologous template probe was a 1.4-kb *Sal*I/*Eco*RV fragment isolated from the *Arabidopsis thaliana* cDNA clone designated TE3-2 encoding a long-chain acyl-ACP thioesterase (Dörmann et al., 1995). The size (in kb) of standard DNA fragments obtained by digesting lambda DNA with *Hin*dIII are shown on the left. The sizes of standard DNA markers derived from the plasmid vector pGEM-3 (Promega) are shown on the right.

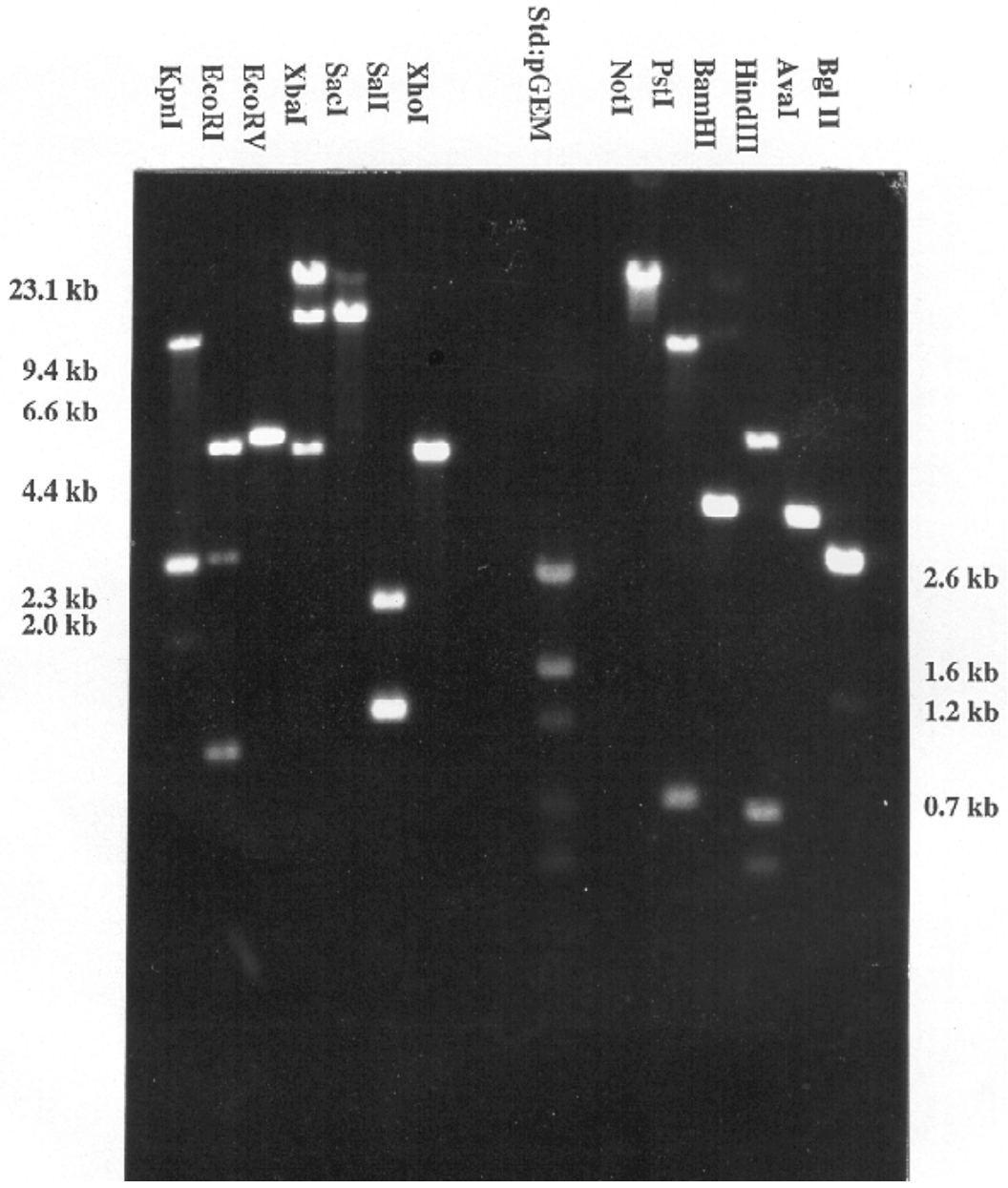

KpnI
EcoRI
EcoRV
XbaI
SacI
SalI
XhoI
Std:pGEM
NotI
PstI
BamHI
HindIII
AvaI
Bgl II
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 10. Physical map of the cotton genomic clone designated LCPg59 encompassing a *PATE* gene. The cotton genomic DNA segment is 17.4 kb in the EMBL3 lambda vector. The cotton DNA insert is represented by the horizontal line, the left and right arms of the vector are shown by the hatched areas labeled λ_L and λ_R, respectively. The stippled rectangle indicates the location of the *PATE* gene, with the arrow denoting the relative polarity of the transcription unit from 5' to 3'.

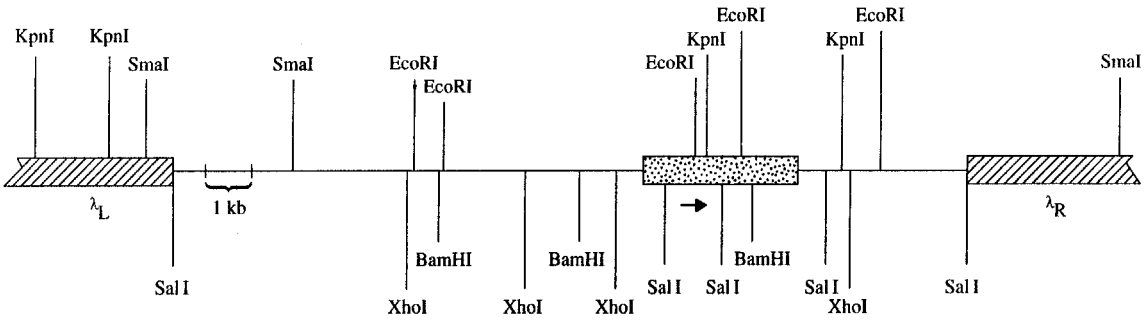
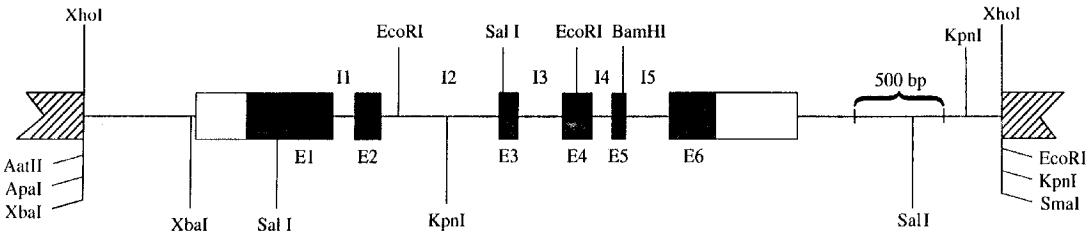

KpnI
KpnI
SmaI
SmaI
EcoRI
EcoRI
EcoRI
KpnI
EcoRI
KpnI
SmaI
λL
1 kb
λR
BamHI
BamHI
BamHI
Sal I
Sal I
Sal I
Sal I
Sal I
XhoI
XhoI
XhoI
XhoI

Figure 11. Physical map of a 5.2-kb *Xho*I fragment of the genomic clone LCPg59 subcloned into the plasmid vector pGEM7Zf(+) and the recombinant plasmid designated pCPg59. Both strands of the *Xho*I fragment were analyzed by DNA sequencing (Yoder et al., 1999). The black rectangles represent the coding regions of the six exons (E1-E6) deduced from a comparison with the cotton *PATE* cDNA sequence (Pirtle et al., 1999). The horizontal lines between the exons indicate the locations of the five introns (I1-I5). The open rectangles represent the extent of the 5'- and 3'-untranslated regions predicted by computer analysis of the DNA sequence. The vertically lined areas represent the pGEM7Zf(+) plasmid vector from Promega.


XhoI
EcoRI
SalI
EcoRI
BamHI
KpnI
XhoI
I1
I2
I3
I4
I5
500 bp
E1
E2
E3
E4
E5
E6
AatII
ApaI
XbaI
XbaI
SalI
KpnI
SalI
EcoRI
KpnI
SmaI

shown underlined in Figure 12 that could function as positive regulatory elements in gene expression. Two basic region helix-loop-helix (bHLH) or E box motif (Guilfoyle, 1997; Williams et al., 1992; Kawagoe et al., 1994) with the consensus sequence CANNTG occur at nucleotides (nt) 19 and 461 or 891 and 454 residues upstream from the ATG initiation codon. Potential CAAT and TATA basal promoter elements (Mitchell and Tjian, 1989; Young, 1991) occur at nt 537 and nt 591 or 379 and 325 residues, respectively, upstream from the ATG initiation codon. A putative cap site for formation of the 5' end of the mature cotton *PATE* mRNA occurs at nt 618 or 297 residues upstream from ATG initiation codon.

The 3' flanking region of the cotton gene has a likely near-upstream poly(A) polymerase CAA-TAAA like element (Hunt, 1994; Hunt and Messing, 1998) with the sequence 5'-AATGAA-3' (at nt 4034 in Figure 12) or 20 nt upstream from the poly(A) cleavage/polyadenylation sites at nt 4052, inferred from comparison with the *PATE* cDNA (Pirtle et al., 1999). The locations and sizes of the exons and introns of the cotton gene are shown in Table 1. All five of the introns in the cotton gene have the classical GT-AG exon-intron junction sequences, and are AT-rich.

Amino acid analysis of the *PATE* coding region

Comparison of alignments of the deduced amino acid sequences of cotton PATE polypeptide with other plant fatty acyl-ACP thioesterases was done using DNASIS software (Hitachi) and is shown in Figure 13. The alignment of the deduced amino acid sequences for the plant acyl-ACP thioesterases in Figure 13 indicates that there is a 63% identity between the cotton and *Arabidopsis* preproteins, a 54% identity between the

Figure12. Nucleotide sequence of the noncoding strand of the 5,201-bp *Xho*I fragment the plasmid subclone designated pCPg59 encompassing the cotton *PATE* gene. The sequence has been assigned GenBank accession number AF076535 (Yoder et al., 1999). The sequenced segment of the 5'-untranslated region is 914 nt, relative to the ATG translation initiation codon, starting at nt 915. The presumptive promoter/enhancer elements are indicated by underlining. A presumptive CAP site corresponding to the 5'-end of the putative *PATE* mRNA occurs at nt 618. The gene has five introns in the coding sequence numbered 1-5, indicated by lower case lettering. The putative near-upstream element (29-30) for polyadenylation by poly(A) polymerase in the 3'-untranslated region at nt 4,034 is underlined. The deduced amino acid sequence of 413 amino acids of the PATE preprotein is indicated above the exon sequences by capital lettering

```
     XhoI                bHLH
   1 CTCGAGAAAATTTCTCATCATTTGATTCGGTTTTTTCTTCGTTAAGGTACGGTCGTAGTT 60

  61 TTTCGTTTCTCAGTTTTAGTTTGATCTGAGCTTTTTTGGTTGAAATGATGCTGATCTCTT 120

 121 TAGTTACCGAGAAAATAAATGAGAAGGATGTATTGTTATATTGTTTTTTTTTCTCGAAAG 180

 181 GTTTCGAGTGATTGTGGCGGCCATAGGAGATCTAATGTTTTCTCGCATTTTCTCGGCAAT 240

 241 CAAAGGCAGCTTAAATTTATAATTTATATGAAAAAGTTCAGCTTATTGTTTATTTTATCG 300

 301 TATTTTTTTTCTCATGAAACACAGAAAGACCAAAGCTCCATTCTGTATTGTTTGCTATTT 360

 361 TCTCTTTCATTTTTTAAGTTTATTTTCCGGTTTTTCGTCTTGTAAACTTTTTAAGCGTAT 420
                                                  bHLH
 421 TACTGTTTTTGTTTTTAGCGATTCTACATCCATTTTGCAACATATGTTCTTCAACGCCGA 480
                                                          CAAT Box
 481 TTTAAGATTCTTTATGCGACGCTCTTTGTTTAAGTACTGTGATTTTTAGTACATAACAAT 540
                                                       TATA Box
 541 TTGCTTTGAGCTACTGGTTAATCCTATCTTCCTTTTTGTAAATAGATCCCTATAAATCTA 600
                     CAP Site
 601 GAACTTTGAAATATAAACAAACACGTGTATATTATGTTCTTTAGATTATGGCTGCTTATT 660

 661 AAATAAAGTAGAAAGAAATTATTTACCTATTAGATTGGAACCTGCTTTTAAACATATAGA 720

 721 ATATGTAATCTTCTATATAGTCAAAATCTGAAGATTTATGTGATTATATTTATTGTTAGT 780

 781 GCGTCTTTTGGATATGTTCTGATTCTTTAATACATGTGATACTGTTAGCTCATGCACCAG 840

 841 CTGCTTGTATAAAAAGCTTTAGATTTTGCAAAAGAAGGGATTTCAGCACGAAATTGAAGT 900
                   M  V  A  T  A  V  T  S  A  F  F  P  V  T  S
 901 TGTTTTTAAAAACCATGGTTGCTACTGCTGTGACATCGGCGTTTTTCCCAGTCACTTCTT 960
     S  P  D  S  S  D  S  K  N  K  K  L  G  S  I  K  S  K  P  S
 961 CACCTGACTCCTCTGACTCGAAAAACAAGAAGCTCGGAAGCATCAAGTCGAAGCCATCGG 1020
     V  S  S  G  S  L  Q  V  K  A  N  A  Q  A  P  P  K  I  N  G
1021 TTTCTTCTGGAAGTTTGCAAGTCAAGGCAAATGCTCAAGCACCTCCGAAAATAAACGGCA 1080
     T  V  A  S  T  T  P  V  E  G  S  K  N  D  D  G  A  S  S  P
1081 CTGTGGCGTCGACGACTCCCGTGGAAGGTTCCAAGAACGATGACGGTGCAAGTTCCCCTC 1140
     P  P  R  T  F  I  N  Q  L  P  D  W  S  M  L  L  A  A  I  T
1141 CTCCTAGGACGTTTATCAACCAGTTACCTGATTGGAGCATGCTTCTTGCTGCTATCACAA 1200
     T  I  F  L  A  A  E  K  Q  W  M  M  L  D  W  K  P  R  R  P
1201 CCATTTTCTTGGCTGCTGAGAAGCAGTGGATGATGCTTGATTGGAAGCCGAGGCGGCCTG 1260
     D  M  V  I  D  P  F  G  I  G  K  I  V  Q  D  G  L  V  F  S
1261 ACATGGTCATTGATCCGTTTGGCATAGGGAAGATTGTTCAGGATGGTCTTGTTTTCAGTC 1320
     Q  N  F  S  I  R  S  Y  E  I  G  A  D  Q  T  A  S  I  E  T
1321 AGAACTTCTCGATTAGATCATATGAGATAGGCGCTGATCAAACAGCATCCATAGAGACAC 1380
     L  M  N  H  L  Q
1381 TAATGAATCATTTACAGgtagagttacagttatttggctagtatgtttgaacaatgaaca 1440
                                Intron 1
1441 ttgggaaacaggatttacatttattggtttcttttgtagagatatggcataagcttgagt 1500
                                   E  T  A  I  N  H  C  R  S  A
1501 tttagttaatacatctcacttttttggcagGAAACAGCTATAAATCATTGTCGAAGTGCTG 1560
     G  L  L  G  E  G  F  G  A  T  P  E  M  C  K  K  N  L  I  W
1561 GACTGCTTGGAGAAGGTTTTGGTGCAACACCTGAGATGTGCAAGAAGAACCTAATATGGG 1620
```

```
     V  V  T  R  M  Q  V  V  V  D  R  Y  P  T  W
1621 TTGTCACACGGATGCAAGTTGTGGTTGATCGCTATCCTACTTGgtaagacatgctttttt 1680

1681 gctcatgattatagcaacaattcatgataagccactttgctctacagtatggctgtggc 1740

1741 atatcttttgatactaactagttcagttcttgaattccagcaatattctgtattatacaa 1800

1801 atgatctgtatcacatctggcggacttgtgtttgtttcattaaaacttggattgatgtta 1860

1861 ttgtttaagcttttaaaggttaagatatgaagtcgaagacaattaaggctagcccccagc 1920

1921 aatgaataacataagaaagataaacctgatacgcttctttgtttaatgagattccctgtt 1980
                                Intron 2
1981 taatactaagaacgggaccttaacttgtcatttttgttaatgttaatacttcccctctaa 2040

2041 ttgtgatttaggtacctatgcactatattgttttaagcacttggcaagttattggtgttg 2100

2101 aaaataacattcacttaactgatattagcttggcattagggcttgtttcaaaaataataa 2160

2161 taaaaacgaagaagttggcctaaaatagttacttttagcaaggtatgtacttgttggagc 2220

2221 accgttctttttgtatgtcacaaaattagtagttctggaggtattaagagtatatagaa 2280

2281 atattattattattttcgttaggtgtatgtctatattgatgaacattgaacttatttttg 2340
                   G  D  V  V  Q  V  D  T  W  V  S  A  S  G  K  N
2341 ttgtcttgcagGGGTGATGTTGTTCAAGTCGACACTTGGGTCAGTGCATCGGGGAAGAAT 2400
      G  M  R  R  D  W  L  V  S  N  S  E  T  G  E  I  L  T  R  A
2401 GGCATGCGAAGAGATTGGCTTGTCAGCAATAGTGAAACTGGTGAAATTTTAACACGAGCC 2460
      T  S
2461 ACAAGgtcggtcattgtttatggaaggatcatgaccatatgtttttttcgtgatgtaac 2520

2521 ctgttgaactctaaaatatttcaaaatttgttttgctagcatttaatatgttttcaatcg 2580
                                Intron 3
2581 atacacgaatcgatattttggtttcccaggccacatctaatgacttttttcctgaccttgt 2640

2641 gtttgcactttaatgaacagtgtttcatgagtgactaatcccagtctcctctgtttttgt 2700
                        V  W  V  M  M  N  K  L  T  R  R  L  S  K  I
2701 tttgttgatctgcagTGTATGGGTGATGATGAATAAACTGACTAGAAGGTTATCTAAAAT 2760
       P  E  E  V  R  G  E  I  E  P  F  F  M  N  S  D  P  V  L  A
2761 CCCAGAAGAGGTTCGAGGGGAAATAGAACCTTTTTTTATGAATTCAGATCCTGTTCTGGC 2820
       E  D  S  Q  K  L  V  K  L  D  D  S  T  A  E  H  V  C  K  G
2821 TGAGGATAGCCAGAAACTAGTGAAACTCGATGACAGCACAGCTGAACACGTGTGCAAAGG 2880
       L  T                    Intron 4
2881 TTTAACTgtaagtccccgcttcccctgtttctctttcatatacttacagcgtctgtcact 2940

2941 tgtaattgctgttatgttcattcgttgcaatttgtaataaagtttatatctaatttgcag 3000
      P  K  W  S  D  L  D  V  N  Q  H  V  N  N  V  K  Y  I  G  W
3001 CCTAAATGGAGCGACTTGGATGTCAACCAGCATGTCAATAATGTGAAGTACATTGGCTGG 3060
      I  L  E
3061 ATCCTTGAGgtagactcactccggttgtatttcaaggattttcttttgaacattctcacc 3120

3121 attacctcttcgtatccgaagataacaattaaatggaaatcataactgatttttatttcg 3180
                                Intron 5
3181 tataccttaatttttacctatgatgtcaagctaatatgaaattggaatttggtagaagt 3240
```

```
3241 ctgcttcagatcttcacgagttaagtcatttatagtctgttggttacatgcattttaacc 3300
                          S  A  P  L  P  I  L  E  S  H  E  L  S
3301 ggatggtagtacttgttgcagAGTGCTCCATTACCAATCTTGGAGAGTCACGAGCTTTCC 3360
      A  L  T  L  E  Y  R  R  E  C  G  R  D  S  V  L  Q  S  L  T
3361 GCCTTGACTCTGGAATATAGGAGGGAGTGCGGGAGGGACAGCGTGCTGCAGTCACTGACC 3420
      T  V  S  D  S  N  T  E  N  A  V  N  V  G  E  F  N  C  Q  H
3421 ACTGTGTCTGATTCCAATACGGAAAATGCAGTAAATGTTGGTGAATTTAATTGCCAACAT 3480
      L  L  R  L  D  D  G  A  E  I  V  R  G  R  T  R  W  R  P  K
3481 TTGCTCCGACTCGACGATGGAGCTGAGATTGTGAGAGGCAGGACCCGATGGAGGCCTAAA 3540
      H  A  K  S  S  A  N  M  D  Q  I  T  A  K  R  A Ter
3541 CATGCCAAAAGTTCCGCTAACATGGATCAAATTACCGCAAAAAGGGCATAGAAATCCAAG 3600

3601 TAATCTCATTGCTGTGTGTAGTATCTATCGTGCTCTTTTCGGATTTATATACATATATTC 3660

3661 CTTATGATTATTAGTCTTCCTTTGAGAAAAAAAAAGGGGGTTGTAATTAGGCTTGTTTAG 3720

3721 GAGTCGGGTTTTCGTACATAGCCTTGTAAGGCTCAGCTCGTATGACCCGAGCCTCGGACA 3780

3781 CGGATTTTGTGAAGTTGGGCCCGTGCCCTAACCAGCATAGGCTCTTTCCATGGAAAGGTG 3840

3841 GGTCTGCTTTTGAAAAATTGAATAGCCATGTGAGATGGCTCTCTCCCTACATTATGGGCT 3900

3901 TTTAACCAGTTAGAGACCGGGTAGTTTAGGATAAAATTTATCTTTAATTTGGGAGGATTT 3960

3961 GTATATTTTTTTTGCCTTTATTTTAACCTAAATTTGCTTATAATTATTTGGTTTTATATT 4020
                Poly A Signal    Poly A Site
4021 TAGGTATTGAATCAATGAAGTTTTTAAATTTTAAAATGTTTTGATTGGCTTACATTGACA 4080

4081 AAGCATTGTATCAACCAATTTTTTTAGATCAAATGTAAGCCCTTGTGTAACTATTGCATG 4140

4141 CATGTTAGGAAAAAAAAAACAAAAATTCACTTAAATTTAGGTGAGGTGTTTATTTTTTTA 4200

4201 AGGTATATATGCGTTTATAAATTGATTTATGTACGTAAATGTATTTCATGTGATCGGTTG 4260

4261 CATTCGAAATCAAATAAAGGGGTTAAATTAATTTTTAAGTTCATCGTTTCGAATTGGTAA 4320

4321 AAAAATAGAAAATTGAGATAAAACAGTGATTGAATTTGTTTTTATTATTTTTAACATTTT 4380

4381 ATTAATATTTTAATTGTTTATTTAATGATTGTTGGATTGATGATCGAACTAATTAAATTA 4440

4441 AGAATTAATGGTTTGATAAGTTTGATCATGACGATGAACTCATAATATCTGTATCTTGTC 4500

4501 TGAAACACAGTCTTCTGTACATCTTGCTCTTTCACTCTTAATTGGTATCCAGATCTCAAG 4560

4561 TCAATCTTAGAAAATACAATGGCTCCCTTTAACTAATCAAACAGGTCATCAAAGACGATA 4620

4621 TTACTACAGGCAGACTCATTTGGTCCAACTTGAAGAACATTCCCGTCTTTACATTTCAAC 4680

4681 TCAATTATCTTTTTCCCACAGTCGACTATAACACTATGACCAGTCAACCAATCTATCCCA 4740

4741 AGGATTATATCGAACTCATTAAATGACAATAACATCAAGTTAGTCGAAAAACAATGACCT 4800

4801 CTAATTGTCAAAGGGCAATTTCTACATACTTGATCCACTAGTGCATGTTTGCCTAAGGGA 4860

4861 TTGGACACTTTTATTACAAACTCAGTGGACTCTGCTAGCATATCCATATGAGGTATCAAT 4920
```

```
4921 TCCATACAAATATAAGAGTGGGTAGATCTGGGGTCAACTAAAGCAATGACAGATATTTCA 4980

4981 TGGATAGAAAAGGTACCCGTGATCACGTCAGGAGATTCTGCCTCTTCCCGAGCTTGAATA 5040

5041 GCATAAGTCCTTGCAGTTGTTCTACCCTCGGACCTTACTGCAGCATCTCTTGGTATGCCT 5100

5101 CTACTGCTAGCCCCACTTCTGGGGTTCTTTTGTGATCTACCCCTCAAGGGAGCACTGCTT 5160
                                                XhoI
5161 GCTTTCACATCTTGTTTTCTCTCTCTCTCATTCAACTCGAG                    5201
```

Table 1. Comparison of exon/intron junction organization between the cotton palmitoyl-ACP thioesterase (*Fat B*) gene and the *Brassica napus* acyl-ACP thioesterase (*Fat A*) gene

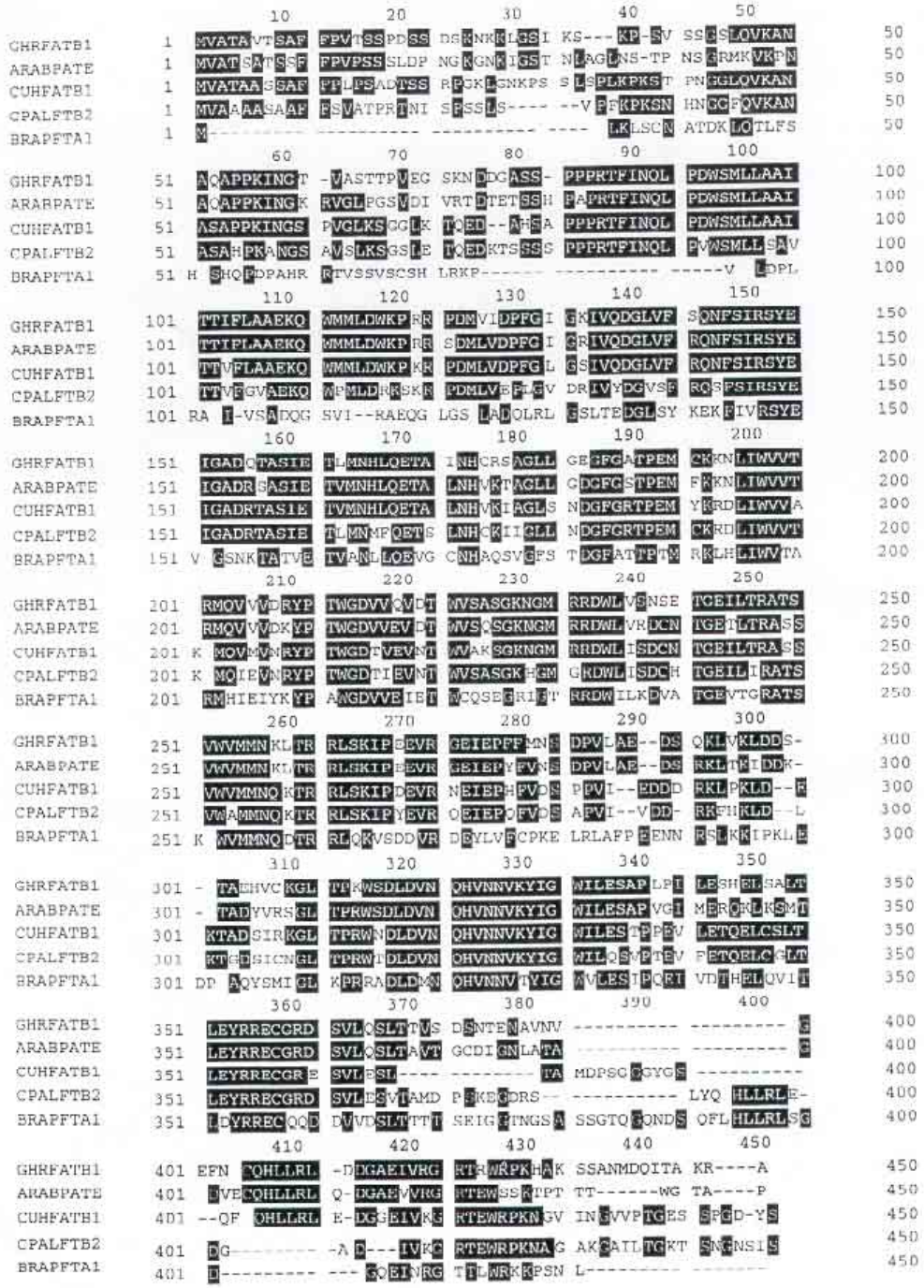
[a]Boundary class 0 refers to splice site after codon position 3; class 1 denotes splice site after codon position 1; class 2 denotes splice site after codon position 2

Table 1
Comparison of exon/intron organization between cotton palmitoyl-ACP thioesterase (*Fat B*) gene and *Brassica napus* acyl-ACP thioesterase (*Fat A*) gene[b]

Exon number	Length (bp)	Splice donor	Intron number	Length (bp)	Splice acceptor	Boundary class[a]
Cotton palmitoyl-ACP thioesterase gene						
1	503	TACAGgtaga	1	132	ggcagGAAAC	0
2	134	ACTTGgtaag	2	688	tgcagGGGTG	2
3	114	ACAAGgtcgg	3	250	tgcagTGTAT	2
4	172	TAACTgtaag	4	113	tgcagCCTAA	0
5	69	TTGAGgtaga	5	252	tgcagAGTGC	0
6	270					
Brassica napus acyl-ACP thioresterase gene						
1	327	TACAGgttcg	1	193	tatagGAGGT	0
2	134	GCTTGgtatt	2	75	aacagGAGTG	2
3	114	ACAAGgtttg	3	195	ctcagCAAGT	2
4	96	CCCAGgtgaa	4	100	cgtagACTAG	2
5	157	TTGAGgtgag	5	89	ggcagAGCAT	0
6	273					

[a]N.M. Loader and R. Safford, GenBank accession number X87842; [18].
[b]Boundary class 0 refers to splice site after codon position 3; class 1 denotes splice site after codon position 1; class 2 denotes splice site after codon position 2.

Figure 13. Alignment of the predicted amino acid sequences of plant acyl-ACP thioesterase preproteins. The amino acid sequences were aligned using the DNASIS software from Hitachi. The thioesterase sequences compared are: *Gossypium hirsutum FatB1* thioesterase (*GHRFATB1*), *Arabidopsis thaliana FatB1* thioesterase (*ARABPATE*), *Cuphea hookeriana FatB1* thioesterase (*CUHFATB1*), *Cuphea palustris FatB2* thioesterase (*CPALFTB2*), and *Brassica rapa FatA1* thioesterase (*BRAPFTA1*). Completely conserved amino acid residues are indicated by reverse contrast.



cotton and *Cuphea hookeriana* preproteins, whereas there is only a 22% identity between the cotton and *Brassica* preproteins.

Phylogenetic relationships between the cotton *PATE* open reading frame with the open reading frames of other acyl-ACP thioesterases

The phylogenetic relationships among the DNA sequences of the open reading frames of the plant fatty acyl-ACP thioesterases is shown by the dendrogram plot based on their DNA sequence similarities. As shown in Figure 14, the corresponding open reading frames/coding regions of a number of *FatA* and *FatB* thioesterase genes/ cDNAs were compared at the nucleotide sequence level. A high degree of identity of about 60% or higher occurs among the various members of the *FatA* or *FatB* DNA sequences of plant species. However, there is only about a 35% identity between the *FatA* and *FatB* cDNAs/genes sequences of the same or different species.

Genomic organization of the cotton *PATE* genes

The number of *PATE* genes in plant genomic DNA was determined by genomic blot analysis. The coding regions of the *Arabidopsis* and cotton *PATE* genes were used as a heterologous probe and a homologous probe, respectively, for genomic blot hybridization. The 5.2-kb *Xho*I and three *Eco*RI fragments (of sizes 5.5-, 3.0- and 1.0-kb) indicated in the physical map of the genomic clone LCPg59 in Figure 15A are derived from the *PATE* gene, since they hybridized to either the heterologous *Arabidopsis* probe or to the homologous cotton probe derived from random priming of the corresponding cDNA fragments. This is depicted in the representative alkaline blot shown in Figure 15A, with the hybridizing fragments (in lanes 2 and 4) clearly matching up with the corresponding fragments in ethidium bromide-staining pattern (in lanes 1 and

3). The relative locations of these fragments are shown in the physical maps of LCPg59 and pCPg59 in Figures 10 and 11 and the DNA sequence in Figure 12. The autoradiogram of an alkaline blot of *Xho*I and *Eco*RI digests of cotton genomic DNAs from the cultivars *TM1* and *Pima* is compared with that of the cloned LCPg59 DNA fragments in Figure 15B. The digests of both cultivar DNA samples have hybridizing fragments that basically correspond to the sizes of the fragments from the cloned DNA, but appear to have slightly lower mobilities than the corresponding cloned DNA fragments from a comparison of the cloned *Xho*I fragment and with the genomic *Xho*I fragments in Figure 15B. It also can be seen from Figure 15 that the *Xho*I and *Eco*RI digests of the genomic DNA have additional hybridizing fragments.

Discussion

Cotton (*Gossypium hirsutum*) is primarily grown as a textile fiber. In addition, secondarily, the cottonseed has important byproducts such as oil, meal, and hulls (National Cottonseed Products Association, http://www.cottonseed.com/facts.html). Cottonseed oil has been commercially used in many food industries as a source of frying /cooking oil. Therefore, success in designing a new cottonseed oil profile can provide higher value-added cottonseed products for the seed oil industry. Due to the high percentage of sequence identity among the plant acyl-ACP thioesterases and the difficulty in purifying to homogeneity acyl-ACP thioesterase specific for only long-chain acyl-ACP substrates by standard biochemical purification procedures (Voelker, 1996), the methods of molecular biology and genetic engineering have been employed in this research to clone and characterize the *PATE* genes in the cotton genome. Expression of

fatty acyl-ACP thioesterase genes may be one of the control points in the fatty acid metabolism required to alter fatty acid composition in storage lipids and membrane phospholipids (Ohlrogge, 1994). Palmitoyl-ACP thioesterases have been shown to be responsible for formation of palmitic acids in 16:0-rich oil (Dörmann et al., 1995; Jones et al., 1995). Thus, thioesterase may be a key enzyme controlling carbon flux through palmitic acids. By controlling expression of this enzyme, it may be possible to shunt the carbon flux through the formation of oleic acids, resulting in accumulation of increased levels of oleic acids for oil biosynthesis. To better understand the mechanism regulating fatty acid composition in cottonseed oil, a cotton palmitoyl-ACP thioesterase (*PATE*) gene was isolated and characterized (Yoder et al., 1999). In the future, we may introduce the *PATE* gene into transgenic cotton plants in order to improve the nutritional and economic values of cottonseed oils. The analysis of cis- and trans-acting promoter/enhancer regulatory elements of *PATE* may provide an understanding of the regulation of gene expression in oil accumulation. Thus, this research involved the isolation of the *PATE* gene by screening a cotton genomic DNA library. The gene structure was characterized by physical mapping, DNA sequencing, and genomic blot hybridization.

A cotton genomic clone encompassing the cotton palmitoyl-ACP thioesterase *PATE* gene and a cotton *PATE* cDNA clone were successfully isolated (Pirtle et al., 1999; Yoder et al., 1999). The cotton *PATE* cDNA clone has a 1,694-bp insert sequence, (Pirtle et al., 1999) which is almost full-length, lacking the first nine nucleotides at the 5'-terminus of the presumptive open reading frame. The open reading frame of this cDNA

Figure 14. Phylogenetic comparison of the gene or cDNA coding regions for plant acyl-ACP thioesterases. The DNA sequences obtained from GenBank were aligned using the DNASIS software from Hitachi. The sequence identities are shown above as percentages.

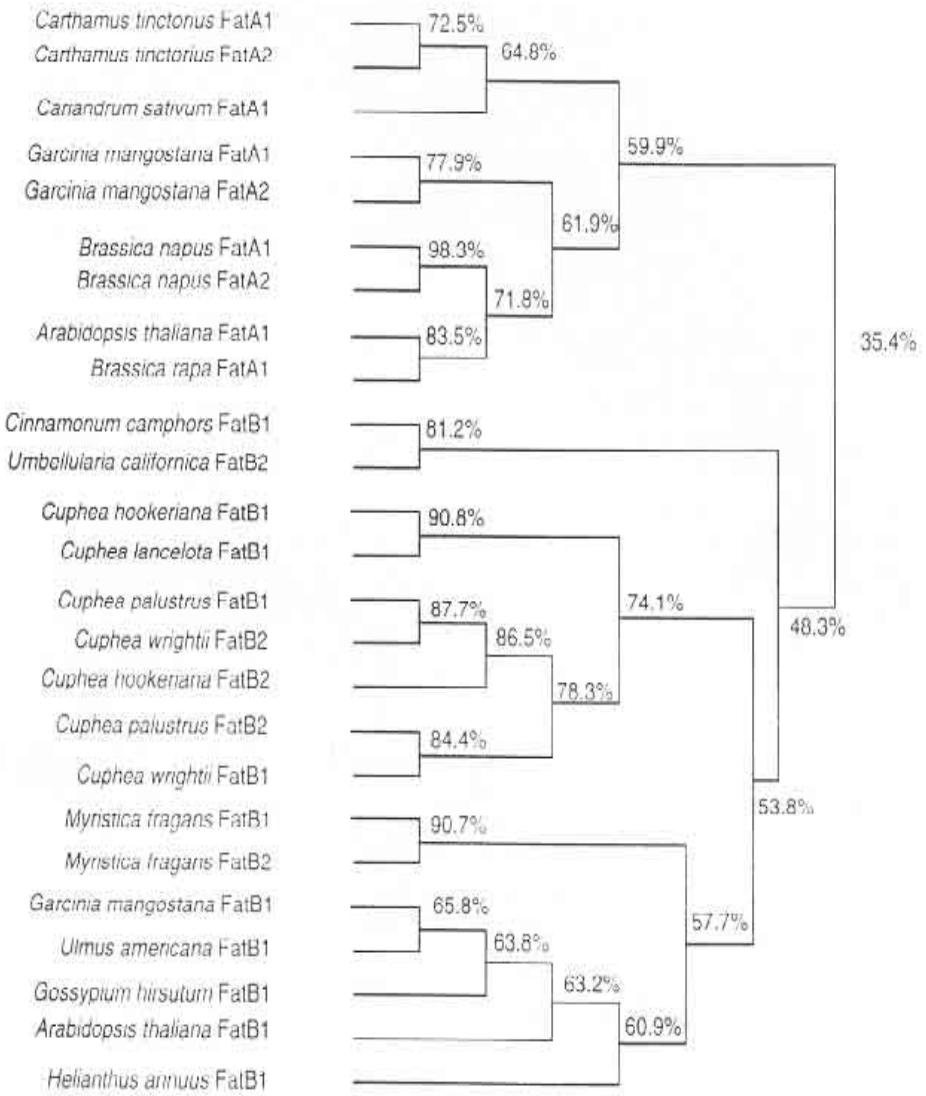
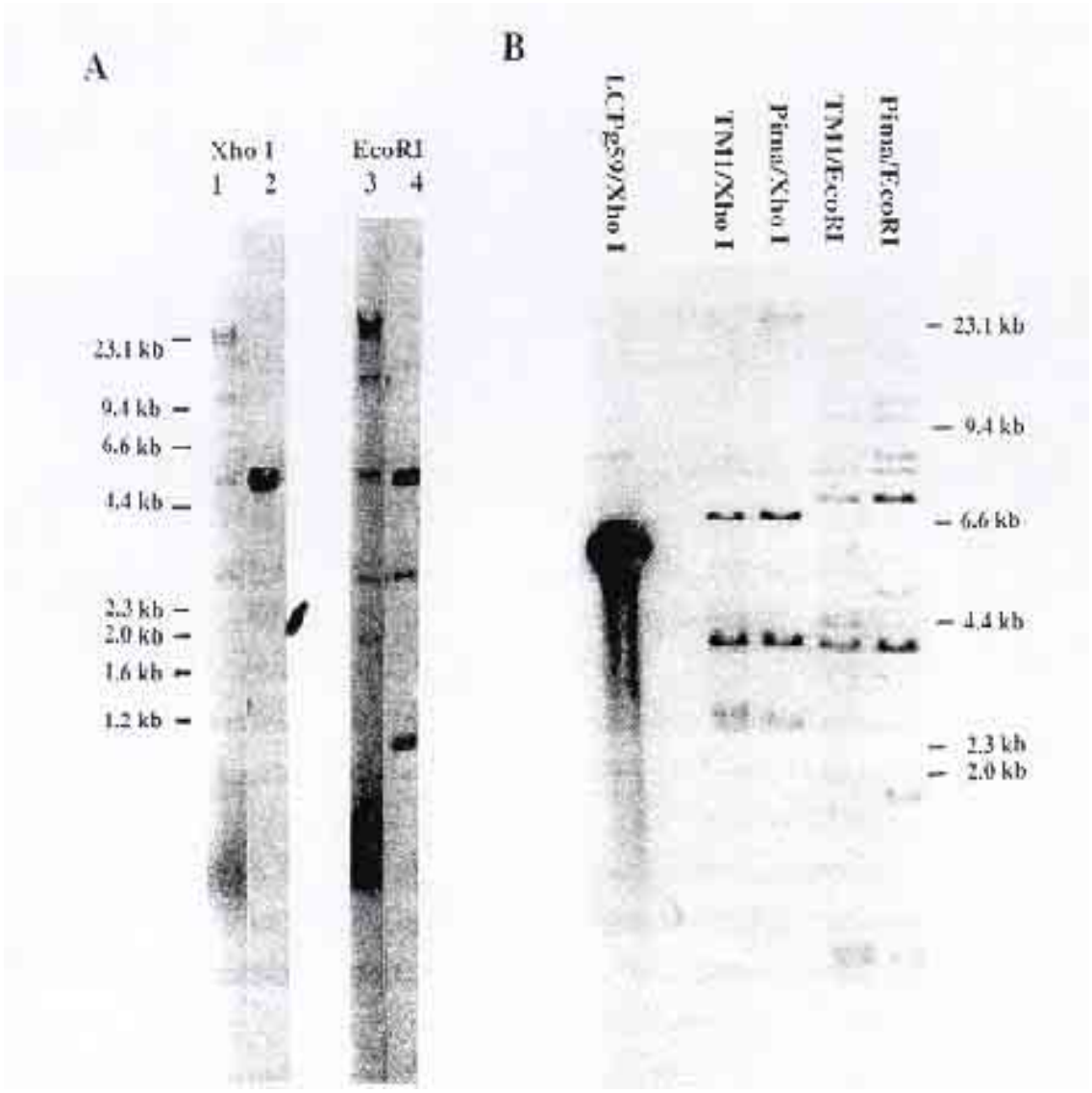

Carthamus tinctonus FatA1
Carthamus tinctorius FatA2
Coriandrum sativum FatA1
Garcinia mangostana FatA1
Garcinia mangostana FatA2
Brassica napus FatA1
Brassica napus FatA2
Arabidopsis thaliana FatA1
Brassica rapa FatA1
Cinnamonum camphors FatB1
Umbellularia californica FatB2
Cuphea hookeriana FatB1
Cuphea lancelota FatB1
Cuphea palustrus FatB1
Cuphea wrightii FatB2
Cuphea hookeriana FatB2
Cuphea palustrus FatB2
Cuphea wrightii FatB1
Myristica fragans FatB1
Myristica fragans FatB2
Garcinia mangostana FatB1
Ulmus americana FatB1
Gossypium hirsutum FatB1
Arabidopsis thaliana FatB1
Helianthus annuus FatB1
72.5%
64.8%
59.9%
77.9%
61.9%
98.3%
71.8%
83.5%
35.4%
81.2%
90.8%
74.1%
87.7%
86.5%
48.3%
78.3%
84.4%
53.8%
90.7%
65.8%
57.7%
63.8%
63.2%
60.9%

Figure 15. Comparison of the cloned cotton *PATE* gene with cotton genomic DNAs by alkaline blot hybridization. (A) LCPg59 DNA was digested with *Xho*I and *Eco*RI and the fragments resolved on a 0.8% agarose gel, alkaline-blotted, and hybridized with a heterologous *Arabidopsis* probe generated by random priming. The left lanes (1 and 3) are the photographs of the ethidium bromide-stained DNA fragments. The right lanes (2 and 4) are autoradiograms indicating the DNA fragments that hybridized to the probe. (B) Cotton genomic DNAs from the cultivars *TM1* and *Pima* were digested with *Xho*I and *Eco*RI, the fragments resolved on an 0.8% agarose gel, alkaline-blotted, and hybridized with a homologous cotton DNA probe derived from random priming. The hybridizing 5.2-kb *Xho*I fragment derived from LCpg59 DNA is shown on the left side. The sizes (in kb) of standard DNA fragments obtained by digesting lambda DNA with *Hin*dIII are shown on the left in A and on the right in B. Also, the size standards on the left in A are from the mobilities of pGEM markers from Promega.

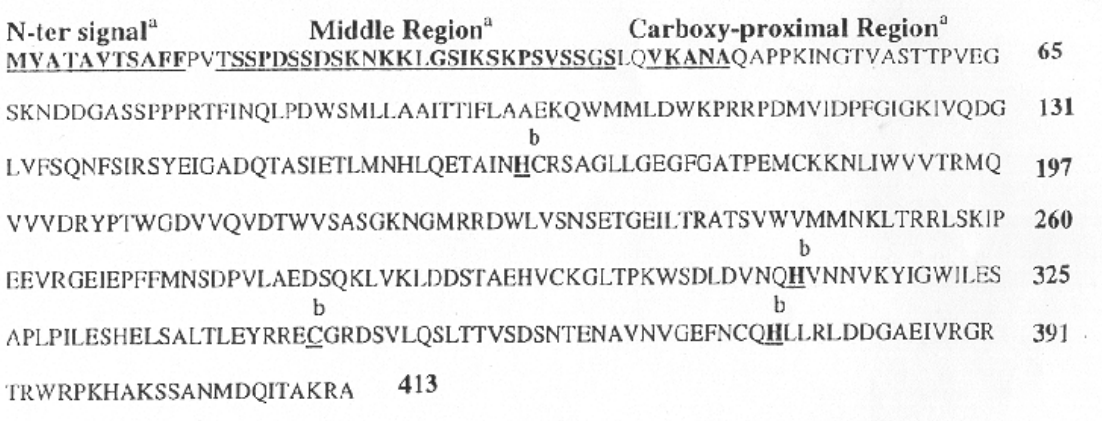

A
Xho I
1 2
EcoR1
3 4
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
1.6 kb
1.2 kb
B
LCPg59/Xho I
TM1/Xho I
Pima/Xho I
TM1/EcoRI
Pima/EcoRI
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb

insert would encode a polypeptide of 413 amino acids, lacking codons for MVA (shown in parenthesis in Figure 3), the three amino acids at the N-terminal. This was confirmed from the DNA sequence of the corresponding region of a cotton genomic clone encompassing this *PATE* gene (Yoder et al., 1999), and is consistent with the N-terminal amino acid sequence of other acyl-ACP thioesterases (shown in Figure 13). The predicted amino acid sequence of the cotton PATE preprotein is 413 amino acids in length and has a characteristic stroma-targeting domain as shown in Figure 3 and 16 (Cline and Henry, 1996). The putative transient peptide cleavage site for production of the mature cotton PATE is probably L84, when compared to other thioesterase preprotein sequences (Gavel and von Heijne, 1990; Dörmann et al., 1995; Cline and Henry, 1996). The catalytic cysteine required for formation of the covalent thiol enzyme intermediate and the histidine involved in general base catalysis in the plant acyl-ACP thioesterases (Yuan et al., 1996) would most likely correspond to the homologous residues C347 and H312 in the cotton amino acid PATE sequence. Two other conserved histidine residues also occur in the plant acyl-ACP thioesterases (Yuan et al., 1996), and would correspond to the homologous H167 and H377 residues. The 3'-untranslated region is 461 bp long with the poly(A) polymerase near-upstream (AATAAA-like) element (Hunt, 1994), 5'-AATGAA-3' (nt 1,685-1690), being 19 bp upstream from the poly(A) cleavage/polyadenylation site at 1,703.

A cotton genomic clone containing a 17.4 kb DNA segment was found to encompass a cotton palmitoyl-acyl carrier protein (ACP) thioesterase (*FatB1*) gene (Figure 10). The gene spans 3.6 kb with six exons and five introns (Figure 11 and 12). The six exons are identical in nucleotide sequence to the open reading frame of the

Figure 16. The stroma-targeting domains (Cline and Henry, 1996) of the PATE preprotein corresponding to the N-terminal region (MVATAVTSAFF, amino acids 1-11), the variable middle region (TSSPDSSDSKNKKLGSIKSKPSVSSGS, residue 14-40), and the carboxy-proximal region (VKANA, residues 43-47) are denoted by underlining the respective amino acid domains. The N-terminal signal is about 10-15 residues, lacking Gly, Pro, or charged residues. The middle region is a variable length region and rich in Ser, Thr, Lys, and Arg. The carboxy-proximal region is loosely conserved region (Ile/Val-X-Ala/CysAla) for proteolytic activity, and this clone demonstrates Val-X-Ala-X-Ala. The putative transient peptide cleavage site is probably L84, when compared with other thioesterase preprotein sequences (Gavel and von Hejine, 1990; Dörmann et al., 1995; Cline and Henry, 1996). The catalytic cysteine required for formation of the covalent thiol enzyme intermediate and the histidine involved in general base catalysis in the plant acyl-ACP thioesterase (Yuan et al., 1996) would most likely correspond to the homologous residues C347 and H312 in the cotton amino acid PATE sequence. The two other conserved histidine residues (Yuan et al., 1996) would correspond to the homologous H167 and H377 residue above.

Stromal Targeting Domain and Conserved Amino Acid Residues of the PATE Preprotein



corresponding cDNA and would encode a preprotien of 413 amino acids and have a characteristic stromal-targeting domain (Cline and Henry, 1996). The locations and sizes of the exons and introns of the cotton gene are shown in Table 1, along with a comparison of the sizes and locations of the exons and introns of *B.napus FatA* gene, the only other acyl-ACP thioesterase gene sequenced to date (GenBank Accession Number X87842; Loader et al., 1993). As shown in Table 1, the relative size of the six exons are about the same between the *FatA* and *FatB* genes, but intron 2 of the cotton gene is 688 bp while its counterpart in the *Brassica* gene is only 75 bp. Four of the five introns (1, 2, 3 and 5) have homologous splice site locations conserved between the corresponding nucleotide and amino acid loci of the cotton and *Brassica* DNA coding regions and deduced amino acid sequences, with the same boundary class types for these four introns. However the splice site location for intron 4 and the boundary class differ between the two classes of thioesterase genes, corresponding to about 60 nucleotide residues upstream from the splice site of intron 4 of the *Brassica* precursor mRNA. The amino acid sequences of the *FatA* and *FatB* thioesterases do not share a great degree of identity in this particular region (Voelker, 1996). This variation in the position of intron 4 may be a distinguishing difference between the nucleotide sequences of the Plant *FatA* and *FatB* genes. All five of the introns in the cotton gene have the classical AT-CG exon-intron junction sequences. The intron 5'-splice sites, 3'-splice sites, and branch sites have sequence similarities to the consensus sequences of these sites in other plant genes. The *FatB* gene introns are AT-rich as are introns of other genes of higher plants (Simpson and Filipowicz, 1996; Brendel et al., 1998).

In the 5'-flanking region of the *PATE* gene, the sequence of 914 bp upstream from the ATG initiation codon was determined. The 5'-flanking region of 914 bp was sequenced and has several potential promoter/enhancer elements (Mitchell and Tjian, 1989; Young, 1991; Thomas, 1993; Guilfoyle, 1997; William et al., 1992; Kawagoe et al., 1994), shown underlined in Figure 12. A putative cap site for formation of the 5' end of the mature cotton mRNA occurs at nt 618, or 297 residues upstream from the ATG initiation codon (CAAACA at nt 618 in Figure 12.) Many promoter element motifs which are identified as being facultative gene promoters for RNA polymerase II-directed transcription (Mitchell and Tjian, 1989; Young, 1991) occur within 618 bp upstream from the presumptive transcription start site in the cotton *PATE* gene. The promoter elements potentially designated as being facultative gene promoters include a TATA box (TATAA at nt 591, or 345 residues upstream from the ATG initiation codon) and a CCAAT box (CAAACA at nt 537, or 379 residues upstream from the ATG initiation codon). Other than the classical facultative gene promoter motifs (TATA box and CCAAT box), two presumptive basic region helix-loop-helix or E-box (Thomas, 1993; Guilfoyle, 1997; William et al., 1992; Kawagoe et al., 1994) with the consensus sequence CANNTG occur at nt 19 and nt 461 or 891 and 484 residues upstream from the ATG initiation codon. The E-box motif has been shown to be a positive regulatory element in seed-specific gene expression in the French bean β-phaseolin gene (Kawagoe et al., 1994). Therefore, this *PATE* gene might have tissue-specific gene expression characteristics. The *Brassica FatA* gene, the only other fatty acyl-ACP thioesterase identified to date, also has these three promoter/enhancer motifs. The transcription start point of the *Brassica* gene occurs at nt 312 and 227 residues, respectively, upstream from

the ATG initiation codon (GenBank Accession number X87842). Two bHLH or E-box motifs occur at nt 367 and 391 residues upstream from the initiation codon. Thus both cotton *PATE* and *Brassica FatA* genes have similar potential promoter/enhancer motifs for regulation of expression of their gene products.

In the 3'-flanking region of the *PATE* gene shown in Figure 12, the sequence of 1,609 bp downstream from the TAG termination codon was determined. The 3'-flanking region of the cotton *PATE* gene has a likely near upstream poly(A)polymerase CAATAAA-like element (Hunt, 1994; Hunt and Messing, 1998) with the sequence ATTGAA at nt 4,034 or 442 residues downstream from the termination codon with the poly(A) cleavage/polyadenylation site located 20 residues downstream from the polyadenylation signal or at nt 4,052, inferred from comparison with the *PATE* cDNA (Pirtle et al., 1999).

Alignment of deduced amino acid sequences for the plant acyl-ACP thioesterases in Figure 13 indicates that there is a 63% identity between the cotton and the *Arabidopsis* preproteins, a 54% identity between cotton and *Cuphea hookeriana* preproteins, whereas there is only a 22% identity between the cotton and *Brassica* preproteins. Similar identities of about 60% occur between the predicted sequences among members of either *FatA* or *FatB* types, whereas identities between members of the two groups are typically much less, being around 30% (Voelker 1996). As shown in Figure 14, the corresponding open reading frames/coding regions of a number of *FatA* and *FatB* thioesterase cDNAs and genes were compared at the nucleotide sequence level. A high degree of identity of about 60% or higher occurs among members of *FatA* and *FatB* sequences of plant species. However, there is only about a 35% identity between the *FatA and FatB* cDNAs

and genes sequences of the same or different species. *FatA* genes encode thioesterases specific for unsaturated fatty acids whereas *FatB* genes encode thioesterases specific for saturated fatty acids (Voelker, 1996; Jone et al., 1995). Based upon theses results, we conclude that this cotton cDNA and gene encodes a member of the *FatB1* class of thioesterase. The preprotein can clearly be identified as a *FatB* acyl-ACP thioesterase from its similarity to the deduced amino acid sequences of other *FatB* thioesterase preproteins.

The digests of both cultivar DNA samples have hybridizing fragments that basically correspond to the sizes of the fragments from the cloned DNA, but appear to have slightly lower mobilities than the corresponding cloned DNA fragments from a comparison of the cloned *Xho*I fragment and the genomic *Xho*I fragments in Figure 15B. Cotton and some other plant genomic DNA samples are notorious for having contaminating polysaccharides, pectins, and other polymeric components that interfere with the activities of restriction enzymes, various polymerases, and other enzymes (Paterson et al., 1993; Michaels et al., 1994; Ausubel et al., 1987). These two genomic DNA samples were subjected to rigorous extraction procedures widely used in the preparation of plant genomic DNAs, but it seem that they still contain some polymeric contaminants (probably polysaccharide) that can interferes with entry of the samples into agarose gels. In fact, the mobilities of the lambda *Hin*dIII fragments used as size standard mixed with the cotton genomic DNA samples were retard relative to the mobilities of the same standard fragments (data not shown). As can be seen from the genomic blot in Figure 15B, the *Xho*I and *Eco*RI digests of the genomic DNAs have additional hybridizing fragments that must be derived from other *PATE* genes. Therefore

it can be definitely concluded that there are at least two or more actual *PATE* genes in the cotton genome, which would likely be the case, since cotton is allotetraploid.

CHAPTER 2

THE DESIGN AND CONSTRUCTION OF SENSE AND ANTI-SENSE PALMITOYL-ACP THIOESTERASE (*PATE*) PLASMID VECTORS FOR ALTERING OILSEED COMPOSITION OF TRANSGENIC COTTON PLANTS

Introduction

Cotton (*Gossypium hirsutum*) is primarily grown as a textile fiber. In addition, cotton plants produce seeds, which in turn have oil, meal and hulls. Cottonseed oil has been commercially used in many food industries. Seed represents almost two-thirds of cotton by weight and oil is one of the products that can be extracted from cottonseed (National Cottonseed Products Association, http://www.cottonseed.com/facts.html). The cottonseed domestic edible oil market represents 4% of the total edible oil market compared with soybean oil, commanding 77% of the market. Interestingly, cottonseed oil is ranked the highest vegetable oil market, representing nearly 56% of the total oil supply, behind soybean and corn oil in the edible food market of baking/frying, margarine/shortening, and salad dressing (Oil Crop Yearbook; July 1995).

Other than safflower, corn, soybean, canola, and sunflower seed oils, cottonseed oil is among the most unsaturated plant oils. It generally consists of 70% unsaturated fatty acids including 18% monounsaturated (oleic), 52% polyunsaturated (linoleic), and 26% saturated (mainly palmitic and stearic) (National Cottonseed Products Association, http://www.cottonseed.com/facts.html). In cooking/frying and food processing industries of the United States, cottonseed oil is mainly used as salad oil and cooking oil because of

its flavor-enhancing property and stability factor. However, recent health concerns over consuming saturated and polyunsaturated fatty acids have raised some questions over the use of cottonseed oil in many of food processing industries. Because of increased public concerns for their health, consumers tend to be more careful when purchasing food products. The National Research Council recommendation on diet and health stated that consumers should cut down on fat, especially saturated fat (The U.S. Department of Agriculture, http://www.nal.usda.gov). The recommended level was 30% or less of total calories from fat consumption instead of the current 37% level. This group also recommended that the limitation of polyunsaturated be less than 10% of overall calories (The U.S. Department of Agriculture, http://www.nal.usda.gov). A study by Parthasrathy et al. (1990) showed a correlation between the amounts of saturated fats, polyunsaturated fats and reproductive (breast and prostate) cancers. Consumption of a diet rich in saturated and polyunsaturated fats is associated with a high cancer (breast and prostate) incidence rate. In addition, a diet rich in oleic acid (monounsaturated) was shown to exhibit a very low cancer incidence rate, when compared with diets rich in saturated and polyunsaturated fats (Parthasrathy et al., 1990).

A study by Dr. S. Grundy, Director of the Center of Human Nutrition at the University of Texas Southwestern Medical Center at Dallas (Grundy, 1989) showed that high-oleic safflower oil has benefits to personal health. Based on clinical studies, this fatty acid was correlated proportionally with the removal of more "bad cholesterol" associated with low density lipoproteins (LDLs) and had no correlation or no effect on the "good cholesterol" associated with high density lipoproteins (HDLs). Also, oils with low polyunsaturated fatty acid levels also have other benefits such as higher stability

when used as frying oil. There is also no need to hydrogenate oils, which have a high monounsaturated fatty acid component. Omitting this step is can reduce the cost-to-price ratio of the oil because it costs an additional 2-3 cents per pound to hydrogenate oils with highly unsaturated fatty acids. Therefore, success in manipulating fatty acid profiles to increase monounsaturated fatty acid (oleic) levels in cottonseed oil should provide a higher value cottonseed oil in the edible vegetable oil market.

Recent research indicated that it is possible to engineer changes in seed oil composition by genetic engineering without adverse effects on crop performance and seed viability (Ohlrogge, 1994; Kinney et al., 1997). In fact, manipulation of fatty acid composition in seed oils has become routine for several crop plants. To date, several transgenic oilseed crops have been produced with altered seed oil profiles without interfering with oil quantity and seed viability and without adverse effects in crop performances (Ohlrogge, 1994). Thus far, little or no adverse agronomic effects have been noted even with drastic changes in seed oils. For example, transgenic *Brassica* plants (which normally produce predominantly C18 unsaturated fatty acids) when genetically engineered to produce mostly C12 saturated fatty acids had little change in oil yield or overall crop performance (Delvecchio et al., 1996). An effort in altering seed oil profiles has been successful by altering expression of acyl-ACP thioesterases and has led to the commercial production of several modified seed oils, high laurate canola oil and high oleate soybean oil (Voelker, 1996; Delvecchio et al., 1996; Kinney et al., 1997).

Cottonseed oil is comprised mostly of palmitic acid (16:0, 22%), linoleic acid (18:2, 58%) and oleic acid (18:1, 15%) (Jones and King, 1996). The synthesis of these fatty acids and their incorporation into triacylglycerols are related metabolically. Their

relative proportions are determined by regulation of carbon flux through complex metabolic pathways and the capability to specifically alter them. To be capable of manipulating fatty acid profiles of cottonseed oil toward a higher level of monounsaturated fatty acid (oleic), understanding *de novo* fatty acid biosynthesis pathways and the subsequent assembly of those fatty acids into seed triacylglycerols is required. Understanding the metabolic links between the biosynthesis of palmitic, oleic, and linoleic acids is also necessary.

As can be seen in Figures 1 and 2, palmitoyl-ACP thioesterase and oleoyl-PC desaturase may be two key enzymes controlling cottonseed fatty acid composition. Recent studies suggested that several thioesterases of different chain-length specificities exist in plants and fatty acid profiles seen in storage oil is in large part controlled by the relative activities of different thioesterases (Ohlrogge, 1994; Dörmann et al., 1995; Ohlrogge and Browse, 1995; Jones et al., 1995; Voelker, 1996). Palmitoyl-ACP thioesterases have been shown to be responsible for formation of palmitic acid in 16:0-rich oil (Dörmann et al., 1995; Jones et al., 1996). Thus, this thioesterase may be a key enzyme controlling carbon flux through palmitic acid. Our preliminary evidence suggests that a *FatB1* acyl-ACP thioesterase activity is present in cotton and is probably responsible for the high level of palmitic acid incorporated in cottonseed oil (Pirtle et al., 1999). In this study, enzyme activity assays and Southern blot analyses indicated that a cotton homologue of the *Arabidopsis thaliana* palmitoyl-ACP thioesterase exists and is most active in embryos of developing cottonseed in the time course of the accumulation of seed storage proteins.

Altering the expression of the acyl-ACP thioesterases has been associated with changes in seed oil profiles (Voelker, 1996), and this has been accomplished by Calgene and other commercial companies (Delvecchio, 1996; Voelker, 1996; Dehesh et al., 1996). For example, transgenic rapeseed and *Arabidopsis* expressing a California bay thioesterase specific for lauroyl-ACP substrate produced over 40% lauric acid in the seed triacylglycerol (Voelker, 1996). This study clearly illustrates the ability to modify fatty acid composition in seed oil by controlling expression of the thioesterase enzyme in developing plant seeds. The first commercial transgenic oil seed crop (Laura Cal, a high lauric acid producing rapeseed) was harvested in 1996 (Delvecchio, 1996). Therefore, by controlling expression of palmitoyl-ACP thioesterase, it may be possible to shunt the carbon flux through formation of oleic acid, resulting in accumulation of an increased level of oleic acid for oil biosynthesis.

Recent evidence indicates that it is possible to engineer changes in seed oil composition by introducing new enzymes, over-expressing existing enzymes, or using anti-sense-RNAs to reduce expression of endogenous enzymes (Ohlrogge, 1994; Kinney et al., 1997). By over-expressing transgenes or expressing anti-sense transgenes corresponding to endogenous genes, expression of the endogenous genes should dramatically be reduced. The reduction in expression is called transgene co-suppression (Flavell et al., 1994; Matzke and Matzke, 1995) and anti-sense RNA suppression (Ohlrogge, 1994). For example, Seymour et al. (1993) reported that transgenic tomatoes with a gene construct having the 5'-end of a polygalacturonase cDNA and a pectin-esterase cDNA in the sense orientation showed no expression of the endogenous polygalacturonase and pectin-esterase genes in fruits. Also, no transgene mRNAs or

proteins were observed in transgenic fruits. Ju et al. (1994) reported that transgenic tomatoes with a gene construct having the polygalacturonase cDNA in an inverted orientation after a constitutive promoter had expression of the anti- polygalacturonase gene, but no polygalacturonase activity was detected. This observation indicated that the expression of the anti-polygalacturonase gene had effectively inhibited expression of the endogenous polygalacturonase gene. Therefore, an attempt to alter cottonseed oil profiles has been done by using the cotton palmitoyl-ACP thioesterase cDNA to generate gene-transgene and anti-sense-transgene constructs for further use in the manipulation of oilseed composition of transgenic cotton plants.

The focus of this chapter is primarily on the design and construction of sense and anti-sense vectors for altering oilseed composition of transgenic cotton plants. This research was done in collaboration with the laboratory of Dr. Kent Chapman. A sense transgene construct was generated by inserting a full-length palmitoyl-ACP thioesterase coding region amplified by PCR under control of a constitutive CaMV 35S promoter region in pBI121, a binary vector used in plant transformation (Jefferson, 1989). A anti-sense transgene construct was done by inserting a partial-length palmitoyl-ACP thioesterase coding region generated by PCR in an inverted orientation downstream from a CaMV 35S region in pBI121. The ultimate goal is to reduce seed-specific expression of palmitoyl-ACP thioesterase (*PATE*) genes only in developing seed during the maximum period of oil biosynthesis (30-45 DAA). To accomplish this goal, both sense and anti-sense seed-specific constructs were generated by replacing a CaMV 35S promoter region in both sense transgene and anti-sense transgene *PATE* constructs with a seed-specific promoter region.

Like all eukaryotes, plants use RNA polymerase II in protein-coding gene transcription (Lewin, 2000). Transcription can be constitutive, which occurs in expression of housekeeping genes encoding proteins in ubiquitous metabolic pathways, such as the tricarboxylic acid cycle, or tissue- and temporal-specific expression because many proteins are required by eukaryotic organisms in specific tissues and at specific stages of development. RNA polymerases use basic promoter elements like TATA-boxes and CCAAT-boxes to transcribe housekeeping genes at a nominal rates in all tissues and at all developmental periods for maintenance of the eukaryotic organism. RNA polymerase II requires tissue- or temporal-specific promoter and enhancer elements for the appropriate expression of a particular gene at a specific time and place (Lewin, 2000). For example, Thomas (1993) showed that a number of specific proximal and distal DNA sequence elements are involved in regulating gene expression in developing seeds. Nunberg et al. (1994) elucidated the promoter/enhancer elements upstream from sunflower helianthinin genes for three sequence motifs in a proximal promoter region that bind to seed nuclear proteins. Mutations in these structural motifs reduced the level of factor binding (Nunberg et al., 1994). AGATGT motifs are found at -111 and –58, TGATCT motifs are found at -83 and –41, and a CCAAT motif is found at -91. The AGATGT and TGATCT motifs have the consensus sequence WGATST (WS motif, where W = A or T and S = G or C).

Other upstream regulatory elements in the 5'-flanking region of the helianthinin gene were found to be responsive only to abscisic acid in the developing seed (Nunberg et al., 1994). The abscisic acid response elements and AT-rich sequences occur further upstream in the helianthinin gene regulatory region (Nunberg et al., 1994). These

elements are structurally similar to promoter/enhancer elements called ACGT motifs, upstream from plant genes like the French bean phaseolin gene that bind bZIP or trans-acting factors involved in abscisic acid-regulated seed expression (William et al., 1992; Thomas, 1993). A number of maturation-specific genes have been identified in cotton, including the genes encoding Lea (Late embryogenesis abundant) proteins (Baker et al., 1988; Roberts et al., 1993), several storage proteins (vicillis, leguminus and albumin (Chlan et al., 1987; Galau et al., 1991; Galau et al., 1992a), and lipid body proteins (oleosin) (Hughes et al., 1993). All these proteins are expressed at high levels in cotyledons of developing cotton embryos during reserve oil and protein accumulation. The cotton oleosin gene has an ACGTG-containing element involved in abscisic acid induction of other oleosin genes (Hughes et al., 1993) and cotton *Lea* genes (Galau et al., 1992b). Therefore, for the seed-specific expression of both palmitoyl-ACP thioesterase sense and anti-sense constructs in transgenic cottons plants, the cotton oleosin promoter region (Hughes et al., 1993) was inserted into both palmitoyl-ACP thioesterase sense and anti-sense constructs, and the CaMV 35S promoter cassette was deleted.

Materials and Methods

A partial cotton palmitoyl-ACP thioesterase (*PATE*) cDNA clone designated SKCPc115b was obtained from Mr. David W. Yoder of our laboratory (Pirtle et al., 1999). A full-length cotton *PATE* coding region generated from the coding region of pSKCPc115b by reverse transcription (RT) PCR amplification was obtained from Dr. Tu T. Huynh of the laboratory of Dr. Kent D. Chapman (Huynh, 2001). A cotton oleosin genomic clone was kindly provided by Dr. Glen Galau of the University of Georgia

(Hughes et al., 1993). Plasmid pBI121 and pGEM7Zf(+) vector DNAs were purchased from Clontech Laboratories, Inc. (Palo Alto, CA) and Promega Corp. (Madison, WI), respectively. ElectroMax *Agrobacterium tumefaciens LBA 4404* cells and *E.coli DH5α* cells were purchased from Invitrogen, Inc. Restriction endonucleases, T4 DNA ligase and calf intestinal alkaline phosphatase were purchased from Invitrogen Inc. Ampli*Taq* DNA polymerase was purchased from Perkin Elmer. Platinum *Taq* DNA polymerase and Platinum *Pfx* DNA polymerase were purchased from Invitrogen, Inc. Various agarose powders (NuSieve, LE and Sea Plaque GTG) were purchased from FMC Corp. Synthetic oligodeoxynucleotide for PCR and sequencing primers were purchased from Biosynthesis, Inc. (Lewisville, TX). Nitrocellulose and Hybond-N^+ nylon membranes for DNA transfer were purchased from Schleicher and Schuell (Keene, NH) and Amersham Life Sciences (Piscataway, NJ), respectively. Wizard *Plus* Miniprep DNA Purification kits were purchased from Promega Corp. QIAquick Gel Extraction and PCR Purification kits were purchased from QIAGEN, Inc (Valencia, CA). Sequenase Version 2.0 DNA Sequencing kits and ThermoSequenase Cycle Sequencing kits were from Amersham Life Sciences. Radioactive materials, including [α-^{33}P]-ddNTPs and [α-^{32}P]-dCTP, were from Amersham Life Sciences and NEN Life Science Products, Inc. Various other reagents, chemicals and materials were purchased from Kodak, Fisher Scientific, Sigma Chemical Co., and Millipore Corp.

Preparing the pBI121 Binary Vector for Design and Construction

A plasmid pBI121 binary vector from Clontech, Inc. (Palo Alto, CA) was used as the parental vector in the design and construction of all plasmid constructs. This binary vector is derived from the plasmid pBIN19 as shown in Figure 17 (Bevan, 1984). The

original pBIN19 vector contains the *Agrobacterium* Ti plasmid right (RB) and left border (LB) sequences, a nopaline synthase (Nos) gene promoter and terminator, a neomycin phosphotransferase II (NPT II-Kan^R) gene and a multiple cloning site (Bevan, 1984). pBI101 (as shown in Figure 18) is a plasmid derived from the plasmid pBIN19. It contains a promoterless 1.87-kb β-glucuronidase (GUS) cassette in the binary vector pBIN19. A 260-bp fragment containing a polyadenylation signal from the nopaline synthase gene (Nos-ter) of an *Agrobacterium* Ti plasmid was inserted downstream from the *GUS* gene. pBI101 also contains a low-copy-number RK2 origin of replication and a kanamycin resistance gene. pBI121 (as shown in Figure 19) is a plasmid derived from pBI101. It has an 800-bp *Hin*dIII/*Bam*HI fragment containing a cauliflower mosaic virus (CaMV) 35S promoter cloned upstream of the *GUS* gene. A CaMV 35S promoter is a constitutive promoter and expresses a high level of GUS activity after Ti-mediated transformation.

HPLC-purified pBI121 DNA (100 μg) was digested with *Sst*I and *Xba*I in 400 μl. solution at 37°C overnight. The digested and undigested DNA samples were resolved on a 1% agarose gel to check completion of digestion. The linearized pBI121 vector was then dephosphorylated by treatment with calf intestinal alkaline phosphatase (CIAP) in order to minimize the occurrence of vector self-ligation. The 40 μl (5 μg) of linear pBI121 vector DNA was mixed with 5 μl of 10x CIAP buffer (500 mM Tris-HCl (pH 8.5) and 1 μl (1 unit/μl) of CIAP and 4 μl of water in a 50 μl total reaction volume. The reaction mixture was incubated at 37°C for one hour. The reaction was stopped by heating at 75°C for 10 minutes and followed by extraction with phenol and chloroform.

The DNA sample was precipitated with 100% ethanol in a -90°C freezer for one hour. Then the pellet was redissolved in water.

Construction of *Agrobacterium tumefaciens* binary plasmid vector for constitutive anti-sense RNA suppression of cotton palmitoyl-ACP thioesterase gene expression.

To construct a constitutive *Agrobacterium tumefaciens* binary plasmid vector for anti-sense RNA suppression of palmitoyl-ACP thioesterase, a partial cDNA coding region for cotton palmitoyl-ACP thioesterase in a plasmid designated SKCPc115b was used as a template. The plasmid DNA (1 μg in 4 μl) was amplified by PCR using the forward primer, 5'-CCGGAGCTCAGACTGCTGTGACATCG-3' with a *Sst*I site (underlined) and the reverse primer, 5'-CCG TCTAGAAGAGCCATCTCACATGG-3'with a *Xba*I site (underlined). A solution containing 10 μl of 10xPCR buffer (100 mM Tris-HCl pH 8.3, 500 mM KCl, 15 mM $MgCl_2$, 0.01% (w/v) gelatin), 200 μM of deoxynucleotide triphosphates (dNTPs), 1 μM of both primers, 2.5 units of ampliTaq DNA polymerase, and 4 mM of $MgCl_2$ was added into a 500 μl PCR tube and gently mixed by tapping the tube. Then, water was added to make a final reaction volume of 100 μl. PCR amplification was conducted in a Perkin Elmer 2400 thermal cycler. PCR conditions were as follows. The template was denatured at 94°C for 5 minutes. Subsequent amplification of the targeted region within the *PATE* cDNA coding region was achieved through a two-step annealing cycle. The first step was achieved through five cycles at 94°C for 30 seconds, 53°C for 30 seconds, and 72°C for 30 seconds. The second step was achieved through 35 cycles at 94°C for 30 seconds, 69°C for 30 seconds, and 72°C for 30 seconds. The final polymerization step or extension step was then done

at 72°C for seven minutes. The PCR products were fractionated on a 1% agarose gel to check if there was nonspecific amplification during the PCR.

The 1.6-kb PCR-amplified fragment was purified from primers, free dNTPs, DNA polymerase, and salts using a QIAquick PCR Purification kit (QIAGEN). 500 µl of buffer PB (QIAGEN) was added to 100 µl of the PCR mixture and mixed by vortexing. The mixture was then applied to a QIAquick spin column and centrifuged for one minute in order to bind DNA to the membrane. The flow-through was discarded and 750 µl of buffer PE (QIAGEN) was added to the column to wash out all contaminants and centrifuged for one minute. The flow-through was discarded and the column was centrifuged for an additional one minute. Water (50 µl) was added to the center of the column and centrifuged for one minute in order to elute the DNA. Then, the purified 1.6-kb PCR fragment was digested with *Xba*I and *Sst*I restriction endonucleases and purified by electrophoresis on a 1% agarose gel. The fragment was extracted from the gel using a QIAquick Gel Extraction kit.

The amounts of both the purified 1.6-kb *Xba*I/*Sst*I PCR fragment and the CIAP-treated linear pBI121 vector were determined by comparison with standard bands of known sizes and amounts of DNA. The 1.6-kb *Xba*I/*Sst*I PCR fragment was directionally subcloned into the *Xba*I/*Sst*I site of pBI121 in the anti-sense orientation, downstream from the constitutive CaMV 35S promoter. The 1.6-kb *Xba*I/*Sst*I (100 ng/µl) PCR fragment (3 µl) and the dephosphorylated linear pBI121 (3 µl, 48 ng/µl) DNA were mixed in a 9:1 ratio (insert:vector). T_4 DNA ligase buffer (5 µl of 5x buffer (250 mM Tris-HCl (pH. 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethene

glycol-8000)), 1μl of T_4 DNA ligase (0.1 unit/μl), and water in a final volume of 20 μl were added to the reaction mixture. The ligation mixture was then incubated at room temperature for about one hour. The resulting recombinant 14.6-kb plasmid (as shown in Figure 20) was used for transforming electrocompetent *E. coli DH5α* cells.

Preparation and Transformation of Electrocompetent *E. coli DH5α* Cells

For preparation of electrocompetent cells, 200 μl of a frozen stock of *E. coli DH5α* cells was inoculated into 5 ml of LB media (Luria-Bertani media; 10 gm bactotryptone, 5 g bactoyeast extract, and 5 g NaCl per liter), and a culture was grown overnight at 37°C with shaking in a New Brunwick shaker/incubator at 250 rpm. A large scale culture of 1.0 liter in LB media was then done in two 2 L flasks containing 500 ml each by inoculation with 15 ml of the overnight culture, and the large scale culture was grown at 37°C with aeration. The A_{600} of the culture was monitored until it was between 0.5-1.0, and then the cells were chilled on ice. The bacterial cells were then harvested by centrifugation in a Sorvall GS-3 rotor at 5,500 rpm (4,300xg) for 15 minutes at 4°C. Supernatants were removed and the cells were resuspended in one volume of sterile cold water. The cells were then subjected to several washes and resuspensions in sterile cold water with centrifugation in a Sorvall GS-3 rotor at 5,500 rpm (4,300xg) for 15 minutes at 4°C (Resuspension was done twice in 0.5 volumes of sterile ice-cold water and once in 0.02 volumes of sterile ice-cold water). The final resuspension was in a 0.003 final volume of filter-steriled cold 10% glycerol. The electrocompetent cells were then aliquoted and frozen at -90°C.

For transformation, approximately 100 ng of the recombinant pBI121 construct plasmid DNA was mixed with 35 µl of electrocompetent *E. coli DH5α* cells. A BTX disposable cuvette (P/N 610, 1mm gap) was then filled with the transformation mixture. The filled cuvette was chilled on ice for one minute prior to electroporation, and then a single 1.5 kV electrical pulse was applied. This resulted in a field strength of 15.0 kV/cm with an exponential decay constant of approximately 5-6 milliseconds. Immediately following the electrical pulse, 960 µl of LB media was added to the cuvette and gently mixed. The cell suspension was then transferred to a fresh polypropylene tube and incubated at 37°C for one hour with shaking at 225 rpm. Portions of the mixture (50, 100 and 200 µl) were plated on LB-containing kanamycin agar plates and incubated overnight at 37°C. After overnight incubation, several prospective white colonies were observed. The colonies were then picked and streaked onto fresh LB-containing kanamycin agar plates to ensure that they were really recombinant clones and also to serve as master plates. The master plates, which serve as a library of the cloned DNA constructs, were incubated at 37°C overnight and then stored at 4°C.

Transformed bacterial colonies may be screened by several methods to determine if they contain the desired recombinant plasmid constructs (Sambrook et al., 1989). One popular method is to isolate small-scale amounts of plasmid DNA ("minipreps") and identify the desired clone by restriction digestion and gel analysis. In this method, a single colony streaked on a master plate was picked and inoculated in 5 ml of LB media containing kanamycin. The cultures were incubated overnight at 37°C with shaking. A 1.5 ml volume of the overnight culture was placed into a microfuge tube and centrifuged

at 12,000 xg for one minute. The supernatant was removed and the bacterial pellets were air-dried. The pellets were then resuspended by vortexing in 100 µl of ice-cold cell suspension solution (25 mM Tris-HCl, 10 mM EDTA (pH 8.0), and 50 mM glucose) and incubated at room temperature for five minutes. A volume of 200 µl of freshly prepared cell lysis solution (0.2 N NaOH, 1% SDS) was added and the solution was mixed by inversion and then incubated on ice for five minutes. This step was for the denaturation of bacterial chromosomal DNA. Ice-cold potassium acetate solution, pH 4.8 (150 µl) was then added to neutralize the lysates, mixed by inversion for 10 seconds, and then chilled on ice for five minutes. The bacterial DNA was pelleted by centrifugation at 12,000xg for five minutes, leaving the plasmid DNA in the supernatant.

The supernatant with the plasmid DNA construct was transferred to a fresh tube and 0.5 µl of DNase-free RNase A (100 µg/ul) was added and incubated at room temperature for five minutes. An equal volume of phenol/chloroform:isoamyl alcohol (24:24:1) was added, and the mixture was vortexed for 30 seconds, and then centrifuged at 12,000xg for two minutes. The upper aqueous phase was transferred to a fresh tube and 0.1 volume of 2.5 M sodium acetate (pH 5.2) and 2.5 volumes of ice-cold 100% ethanol were added. The supernatant was mixed and allowed to precipitate for one hour at -90°C. Pellets were collected by centrifugation at 12,000xg for five minutes, rinsed with ice-cold 70% ethanol, and dried under vacuum. The isolated plasmid DNA was digested with *Sst*I, *Xba*I and *Xba*I/*Sst*I restriction endonucleases. The reaction volume (20 µl) contained 10 µl of plasmid DNA, 1 µl of restriction endonuclease, and 2 µl of the appropriate 10x buffer. The digested plasmid DNAs were then fractionated by 1%

agarose gel electrophoresis. This approach was used to select subclones containing the appropriate 1.6-kb *Xba*I/*Sst*I PCR fragment containing the *PATE* anti-sense construct. One subclone, designated anti-sense PATE/pBI121, was selected for a large scale plasmid preparation.

Large Scale Isolation and Purification of Anti-sense PATE/pBI121 Plasmid DNA

A small 50 ml culture in LB media containing kanamycin (50 μg/ml) was prepared by inoculating with three loops of *E. coli DH5α* transformed with anti-sense PATE/pBI121 plasmid construct and incubated with aeration at 37°C overnight. Two ml of this overnight culture was then added to 100 ml of fresh LB media containing kanamycin. This overday culture was grown until the A_{600nm} reached 0.6 (late log phase). Then, 80 ml of this overday culture was added to 2L of LB media containing kanamycin and incubated with aeration at 37°c overnight. This was done in five 2L flasks containing 400 ml of culture each. Bacterial cells were harvested by centrifugation at 4,000 rpm (2,300xg) for 15 minutes at 4°C in the Sorvall GS-3 rotor. Supernatants were removed and the cell pellets were weighed. The bacterial cells were then subjected to several washes in ice-cold STE buffer (0.1 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The plasmid DNAs were isolated from the bacterial cells by an alkaline lysis procedure (Sambrook et al., 1989). The bacterial cells were resuspended in Solution I (50 mM glucose, 25 mM Tris-HCl, and 10 mM EDTA, pH 8.0), and lysed in a freshly prepared solution of lysozyme in 10 mM Tris-HCl (pH 8.0) and Solution II (0.2 N NaOH and 1% SDS). Then, ice-cold Solution III (5M potassium acetate, pH 4.8) was added to the lysed cells in order to precipitate chromosomal bacterial DNAs and proteins. The cellular debris was pelleted at 4,000 rpm (2,300xg) for 30 minutes at 4°C. Supernatants

were collected and precipitated by addition of a 0.6 volume of isopropanol at room temperature for 10 minutes.

The precipitated DNAs were then pelleted by centrifugation in a Sorvall GS-3 rotor at 6,000 rpm (5,200xg) for 15 minutes at room temperature. The pellets were then resuspended in 10 ml of TE buffer (10 mM Tris-HCl, 1 mM Na_2EDTA (pH 7.5)). RNase A was added to make a final concentration of 10 μg/ml and incubated at room temperature for one hour. A phenol/chloroform extraction was done by the addition of one volume of phenol/chloroform/isoamyl alcohol (24:24:1) with vigorous vortexing. Then samples were centrifuged in a Sorvall SS-34 rotor at 6,000 rpm (3,320xg) for 15 minutes. The aqueous phase was transferred to a new tube and one volume of chloroform:isoamyl alcohol (24:1) was added to the aqueous phase. The samples were vigorously vortexed for two minutes and centrifuged in the SS-34 rotor at 6,000 rpm (3,320xg) at room temperature for 15 minutes. The aqueous phase was again transferred to a clean tube. A 0.1 volume of 3 M sodium acetate (pH 5.2) and two volumes of 100% ethanol were added and mixed with the aqueous phase for overnight precipitation in a -90°C freezer. The ethanol-precipitated sample was warmed to 4°C on ice and then centrifuged in the Sorvall SS-34 rotor at 12,500 rpm (14,400xg) for 30 minutes at 4°C.

The pellets were washed with 10 ml of 70% ethanol, dissolved in an appropriate volume of HPLC-starting buffer (25 mM Tris-HCl and 1 mM EDTA, pH 8.0) and then purified by HPLC by Dr. Irma Pirtle using the general conditions of Merion and Warren (1989). The absorbance of the HPLC-purified plasmid DNA was measured at 260 nm and 280 nm to estimate the amounts and the purity of the plasmid DNA sample. The recombinant plasmid DNA was analyzed by cleavages with restriction endonucleases and

fractionated on a 1% agarose gel to determine the fragment sizes and to verify the identity of the desired recombinant plasmid. To further characterize this clone, the anti-sense PATE/pBI121 was used as a template for a PCR reaction using the same primers as when generating the 1.6-kb PCR fragment for the original construction. The PCR product then was fractionated on a 1% agarose gel. Both strands of the boundaries between the pBI121 vector and insert were sequenced by Dr. Irma Pirtle using a primer-based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, TX). The intact anti-sense PATE/pBI121 plasmid DNA was used as the template DNA for sequencing with either [α-^{35}S]dCTP (Tabor and Richardson, 1987) and T7 Sequenase 2.0 (Amersham) or terminator cycle sequencing (Fan et al., 1996) with ThermoSequenase (Amersham) and [α-^{33}P]-labeled dideoxynucleoside triphosphates (Amersham). Analysis of the DNA sequence was done with DNASIS software from Hitachi. The 14.6-kb construct designated anti-sense PATE/pBI121 was electroporated into electrocompetent *Agrobacterium tumefacien LBA4404* cells.

Preparation and Transformation of Electrocompetent *Agrobacterium tumefacien LBA4404* Cells

Agrobacterium LBA4404 cells contain the disarmed Ti plasmid pAL4404 which has only the *vir* and *ori* regions of the Ti plasmid (Wen-jun and Forde, 1989). Historically, the recombinant T-DNA has been introduced into *A. tumefacien* cells by electroporation. This recombinant DNA is then able to migrate from *A. tumefacien* cells into plant cells using components provided by the plasmid pAL4404 (Wen-jun and Forde, 1989).

For preparation of electrocompetent cells, 5 µl of a frozen stock of Electro MAX *Agrobacterium tumefaciens LBA4404* cells (Invitrogen, Inc.) was inoculated into five ml of LB broth (Luria-Bertani media; 10 grams of Bactotryptone, 5 grams of Bacto Yeast Extract, and 5 grams of NaCl per liter). The culture was grown overnight at 30°C with shaking in a New Brunswick shaker/incubator at 250 rpm. A large scale culture of 1.0 liter of LB media was then done in two 2L flasks containing 500 ml each by inoculating with 15 ml of the overnight culture. The large scale cultures were grown at 30°C with shaking at 250 rpm. The absorbance (at 600 nm) of the culture was monitored until it reached an absorbance between 0.5 and 2.0 and then the cells were chilled on ice. The bacterial cells were harvested by centrifugation in a Sorvall GS-3 rotor at 5,500 rpm (4,300xg) for 15 minutes at 4°C. The supernatant was removed and the cells resuspended in one volume of sterile ice-cold water. Then the cells were subjected to several washes and resuspension in sterile cold water (at least six times). The final resuspension was in five ml of cold filter-sterilized 10% glycerol. The electrocompetent cells were aliquoted and frozen at -90°C.

To transform electrocompetent *Agrobacterium tumefaciens LBA4404* cells, approximately 100 ng of recombinant pBI121 plasmid DNA construct was mixed with 35 µl of the electrocompetent cells. A BTX disposable cuvette (P/N 610, 1 mm gap) was then filled with the transformation mixture. The filled cuvette was chilled on ice for one minute prior to electroporation. The electroporator ECM apparatus was set on the HV Mode /3 kV. A single 1.8 kV electrical pulse was applied. This resulted in a field strength 18.0 kV/cm with an exponential decay constant of approximately 5-6

milliseconds. Immediately following the electrical pulse 960 µl of LB media was added to the cuvette and pipetted up and down gently. Then the cell suspension was transferred to a polypropylene tube and incubated at 30°C for three hours with aeration. Aliquots of cell suspension (50, 100 and 200 µl) were plated on LB agar plates containing kanamycin and streptomycin and incubated 2-3 days at 30°C. After 2-3 days incubation, numerous colonies were formed. The isolated colonies were picked and streaked on fresh LB plates containing kanamycin and streptomycin to ensure recombinant clones and to serve as master plates. Recombinant plasmid DNA was isolated from transformed cells by a mini-scale plasmid DNA preparation procedure (Sambrook et al., 1989). The desired recombinant plasmids were identified by restriction digestion, gel analysis, and alkaline blot hybridization as previously described.

Construction of *Agrobacterium tumefaciens* binary plasmid vector for constitutive gene-transgene co-suppression of palmitoyl-ACP thioesterase gene expression

To construct a *Agrobacterium tumefaciens* binary plasmid vector for a constitutive gene-transgene co-suppression of palmitoyl-ACP thioesterase gene expression, the 1.5-kb *Hin*dIII/*Xho*I fragment of the cotton *PATE* coding region in a pZero-2.1 sense plasmid construct obtained from Dr.Tu T. Huynh of the laboratory of Dr. Kent Chapman at the University of North Texas (Huynh, 2001) was used as template and amplified by PCR. The forward primer was 5'-CGTCTAGAATGGTTGCTACTGCTG TGACATCG-3' with a *Xba*I site (underlined), and the reverse primer was 5'-CCGAG CTCCTTGGATT TCTATGCCCTTTTTGCG-3' with a *Sst*I site (underlined). The *PATE* coding region in the *Hin*dIII/*Xho*I fragment (1 µg in 1µl) pZero-2.1 was used as a

template. High fidelity buffer (10µl of 10x buffer (600 mM Tris-SO_4 (pH. 8.9), 180 mM ammonium sulfate), 200 µM of dNTPs, 1 µM of both forward and reverse primers, 2.5 units of Platinum *Taq* DNA polymerase (Invitrogen), and 4 mM $MgCl_2$ were added into a 500 µl PCR tube and gently mixed by tapping the tube. Water was added to make the final reaction volume of 100 µl. The PCR amplification was conducted in a Perkin Elmer 2400 thermal cycler under the following conditions. The template DNA was denatured at 94°C for 5 minutes. Subsequent amplification of the targeted region was achieved through a single step annealing cycle. This was done for 35 cycles of 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. Then the final polymerization step or extension step was done at 72°C for seven minutes. The PCR products were fractionated on a 1% agarose gel to determine if there were nonspecific amplifications during the PCR.

The 1.3-kb PCR product was purified from primers, free dNTPs, *Taq* DNA polymerase, and salts using the QIAquick PCR Purification Kit (QIAGEN) as previously described. The 1.3-kb PCR fragment was digested with *Xba*I and *Sst*I restriction endonucleases and purified by electrophoresis on a 1% agarose gel and extracted from the gel using a QIAquick Gel Extraction Kit. The plasmid vector pGEM7Zf(+) was digested with *Xba*I and *Sst*I restriction enzymes, and agarose gel (0.8%) electrophoresis was done to check completeness of the double digest. The linearized vector was then extracted with phenol and chloroform and precipitated in 100% ethanol. The pellet was then re-dissolved in water. The linearized pGEM7Zf(+) vector was dephosphorylated through treatment with calf intestinal alkaline phosphatase (CIAP) in order to prevent

vector self-ligation. The 20 µl solution (containing 4.5 µg) of phenol/chloroform-extracted linear pGEM7Zf(+) plasmid DNA was mixed with 2.5 µl of 10xCIAP buffer (500 mM Tris-HCl (pH 8.5), 1 mM EDTA) and 1 µl (0.1 unit) of CIAP and 1.5 µl of water in a reaction volume of 25 µl. The reaction mixture was incubated for one hour at 37°C, followed by phenol/chloroform extraction, and precipitation with 100% ethanol. The CIAP-treated linear pGEM7Zf(+) vector was purified by electrophoresis on a 0.8% agarose gel and extracted from the gel using a QIAquick Gel Extraction Kit (QIAGEN). The amounts of both the purified 1.3-kb *Xba*I/*Sst*I PCR fragment and the CIAP-treated linear pGEM7Zf(+) vector were estimated by comparison of the desired bands with standard bands of known sizes and amounts of DNA.

The 1.3-kb *Xba*I/*Sst*I PCR fragment was directionally subcloned into the *Xba*I/*Sst*I polylinker site of the pGEM7Zf(+) vector in the sense orientation. The 1.3-kb *Xba*I/*Sst*I PCR fragment (2 µl of 100 ng/µl) and 3 µl of dephosphorylated linear pGEM7Zf(+) DNA (30 ng/µl) were mixed in a insert:vector ratio of 5:1. T_4 DNA ligase buffer (4 µl of 5x buffer (250 mM Tris-HCl (pH. 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethene glycol-8000)), 1 µl (0.1 unit) of T_4 DNA ligase, and water were added in a final reaction volume of 20 µl. The ligation mixture then was incubated at room temperature for about one hour. The recombinant plasmids were used for transforming electrocompetent *E. coli DH5α* cells as previously described. Portions of the transformed cell suspensions (50, 100 and 200 µl) were plated on LB agar plates containing amplicillin and incubated overnight at 37°C. After overnight incubation, a few transformants were observed. The isolated white colonies were picked and streaked

on fresh LB agar plates containing amplicillin to assess that they were recombinant clones and also to serve as master plates. A small amount of the plasmid construct DNAs was isolated using a Wizard *Plus* Minipreps DNA Purification Kit (Promega Corp.). In this procedure, a single colony streaked on a master plate was picked and inoculated into five ml of LB media containing amplicillin. Each culture was incubated overnight at 37°C with shaking at 250 rpm. The next day, cultures (1.5 ml) were placed into microfuge tubes, centrifuged at 12,000xg for one minute, and the supernatants removed. The bacterial pellets were then resuspended in 200 µl of Cell Resuspension Solution (50 mM Tris-HCl, 10 mM EDTA (pH 7.5), and 100 µg/ml of RNaseA). Cell Lysis Solution (200 µl of 0.2N NaOH and 1%SDS) was added and mixed by inverting the tube. Then, 200 µl of Neutralization Solution (3 M potassium acetate, pH 5.2) was added and mixed by inverting the tube. The lysates were centrifuged at 12,000xg in a microcentrifuge for five minutes.

In the next phase of the procedure, Wizard Miniprep DNA Resin (1 ml) was resuspended and loaded into each barrel of the minicolumn/syringe assembly attached to a vacuum manifold. The cleared lysate from each miniprep was removed and transferred to the barrel of the minicolumn/ syringe assembly containing resin. The stopcocks were opened and a vacuum was applied to pull the resin/lysate mixture into a minicolumn. The vacuum was then broken when all of mixtures had completely passed through the column. Then, 2 ml of Column Wash Solution (20 mM potassium acetate, 8.3 mM Tris-HCl (pH 7.5), 40 mM EDTA, and 55% ethanol) was added to the barrel and the vacuum was reapplied to draw the solution through the minicolumn. The resin was dried by continuing to draw a vacuum for 30 seconds after the solution had been pulled through

the column. The minicolumn was then transferred to a 1.5 ml microfuge tube. The minicolumn was centrifuged at 12,000xg for two minutes to remove any residual column wash solution. Water (50 μl) was applied to the minicolumn, incubated for one minute, and centrifuged at 12,000xg in a microcentrifuge for one minute to elute the DNA. The isolated plasmid DNA was digested with *Xba*I and *Sst*I restriction endonucleases. The reaction volume was 20 μl, containing 10 μl of plasmid DNA, 1 μl of each restriction enzyme, and 2 μl of the appropriate 10x buffer. The digested plasmid DNAs were then fractionated by 1% agarose gel electrophoresis. This approach was used to determine if the subclones contained the appropriate 1.3-kb *Xba*I/*Sst*I PCR fragment.

One subclone, designated PATE/pGEM, was selected and grown in 50 ml of LB media containing amplicillin in order to prepare sufficient amounts of plasmid DNA for sequencing and subcloning into the binary vector pBI121. Both strands of the boundaries between the pGEM7Zf(+) vector and insert were sequenced using a primer-based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, TX). The intact PATE/pGEM plasmid DNA was used as the template DNA for sequencing with either [α-^{35}S]dCTP (Tabor and Richardson 1987) and T7 Sequenase 2.0 (Amersham) or terminator cycle sequencing (Fan et al. 1996) with ThermoSequenase (Amersham) and [α-^{33}P]-labeled dideoxynucleotside triphosphates (Amersham). Analysis of the DNA sequence was done with DNASIS software from Hitachi. The isolated PATE/pGEM plasmid DNAwas digested with *Xba*I and *Sst*I restriction enzymes. A 1% preparative agarose gel was run to separate the 1.3-kb *PATE* coding region fragment from the pGEM7Zf(+) vector DNA, and the 1.3-kb band was cut from the gel.

The DNA was isolated and purified from the gel band using the QIAquick Gel Extraction Kit (QIAGEN).

For insertion of the *PATE* coding region in the pBI121 vector, the amounts of both the purified 1.3-kb *Xba*I/*Sst*I *PATE* coding region fragment and the CIAP-treated linearized pBI121 vector DNA were determined by comparison with standard bands of known sizes and amounts of DNA. The 1.3-kb *Xba*I/*Sst*I *PATE* coding region fragments were directionally subcloned into the *Xba*I/*Sst*I site of pBI121 in the sense orientation, downstream from a constitutive CaMV 35S promoter (Jefferson, 1989). The 1.3-kb *Xba*I/*Sst*I *PATE* coding region fragment and dephosphorylated linearized pBI121 DNA were mixed in a 5:1 insert:vector ratio. The resulting 14.3-kb recombinant plasmids (shown in Figure 21) were used for transforming electrocompetent *E. coli DH5α* cells as previously described. Portions of the transformed cell suspension (50, 100 and 200 µl) were plated on LB agar plates containing kanamycin and incubated overnight at 37°C. After overnight incubation, a few transformants were observed.

The recombinant plasmids were isolated from the transformed cells using the Wizard Plus Miniprep DNA Purification Kit (Promega). The isolated plasmid DNAs were then doubly digested with *Xba*I/*Sst*I, *Xba*I/*Eco*RI, *Xba*I/*Sal*I, and *Xba*I/*Bam*HI. The plasmid DNA fragments were then fractionated on a 1% agarose gel to determine if the subclones had the appropriate 1.3-kb *PATE* insert. One subclone, designated Sense PATE/pBI121 was selected and grown in 100 ml of LB media containing kanamycin in order to prepare sufficient amounts of plasmid DNAs for sequencing and transforming electrocompetent *Agrobacterium tumefaciens LBA 4404* cells. Both strands of the boundaries between the pBI121 vector and the insert were sequenced using a primer-

based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, TX). The intact sense PATE/pBI121 plasmid DNA was used as the template DNA for sequencing with either [α-^{35}S]dCTP (Tabor and Richardson 1987) and T7 Sequenase 2.0 (Amersham) or terminator cycle sequencing (Fan et al. 1996) with ThermoSequenase (Amersham) and [α-^{33}P]-labeled dideoxynucleotside triphosphates (Amersham). Analysis of the DNA sequence was done with DNASIS software from Hitachi. The 14.3-kb plasmid construct designated Sense PATE/pBI121 was then electroporated into electrocompetent *Agrobacterium tumefaciens LBA4404* cells as previously described.

Construction of *Agrobacterium tumefaciens* binary plasmid vectors for seed-specific anti-sense RNA suppression and gene-transgene co-suppression of palmitoyl-ACP thioesterase gene expression

To prepare the parental vector DNAs for the construction of seed-specific recombinant plasmid DNAs, 10 μg of pGEM7Zf(+) vector DNA was digested with *Xba*I and *Hin*dIII. The digested and undigested DNAs were fractionated on a 1% agarose test gel to check completeness of the digestion. The QIAquick PCR Purification Kit was used to purify the restriction fragments. Both anti-sense PATE/pBI121 and sense PATE/pBI121, which are *Agrobacterium tumefaciens* binary plasmid vectors for constitutive anti-sense RNA suppression and gene-transgene co-suppression of *PATE* gene expression, respectively, were used as parental vectors for generating *Agrobacterium tumefaciens* binary plasmid vectors for seed-specific anti-sense RNA suppression and gene-transgene co-suppression of *PATE* gene expression, respectively. Each plasmid DNA (10 μg) was digested with *Xba*I and *Hin*dIII. The samples were

fractionated on a 1% agarose gel to check completeness of digestion and also to isolate a *Xba*I/ *Hin*dIII parental vector fragment from the CaMV 35S promoter fragment. The gel band was cut from the gel and the DNA extracted from the gel band using the QIAquick Gel Extraction Kit (QIAGEN).

To construct *Agrobacterium tumefaciens* binary vectors for both seed-specific anti-sense RNA suppression and gene-transgene co-suppression of *PATE* gene expression, a 2.6-kb *Hin*dIII fragment of the cotton oleosin gene 5'-flanking region harbored in the Stratagene pBluescriptM13 vector obtained from Dr. Glenn Galau of the University of Georgia (Hughes et al., 1993) was used as template and amplified by PCR. The forward primer was 5'-GAAGCTTGATGGATGATAATGAAGAAAGCGGC-3' with a *Hin*dIII site (underlined) and the reverse primer was 5'GATCTAGAG CGAGAGAGAGAGAAACTTTGACAG-3'with a *Xba*I site (underlined). A *Hin*dIII fragment of the cotton oleosin gene in pBluescriptM13 was used as template (1μg in 10 μl). High Fidelity buffer (10 μl of 10xbuffer (10 μl of 10x buffer (600 mM Tris-SO_4 (pH. 8.9), 180 mM ammonium sulfate)), 2.5 units of Platinum *Taq* DNA polymerase, 4 mM of $MgSO_4$, 200 μM of dNTPs, and 1 μM of both the forward and reverse primers were gently mixed by tapping the PCR tube. Water was added to reach a final reaction volume of 100 μl.

The PCR conditions were as follows. The template was denatured at 94°C for five minutes. Subsequent amplification of a targeted region was achieved through a two step annealing cycle. The first step was achieved through five cycles of 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds. The second step was achieved

through 30 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds. Then, the final polymerization (or extension) step was done at 72°C for seven minutes. The PCR products were fractionated on a 1% agarose gel to check whether there was non-specific amplification occurring in the reaction. The 2.0-kb amplified fragment was then purified from primers, free nucleotides, *Taq* polymerase, and salts using the QIAquick PCR Purification Kit (QIAGEN). The purified fragment was digested to completion with *Hin*dIII and *Xba*I restriction enzymes. The 2.0-kb *Hin*dIII/*Xba*I PCR fragment was subsequently purified by electrophoresis on a 1% agarose gel and extracted from the gel using the QIAquick Gel Extraction Kit (QIAGEN).

The 2.0-kb *Hin*dIII/*Xba*I oleosin gene fragment was then directionally subcloned into the *Hin*dIII/*Xba*I polylinker site of pGEM7Zf(+). The 2.0-kb *Hin*dIII/*Xba*I fragment (10 µl of 80 ng/µl) and the linearized pGEM7Zf(+) (4 µl of 100 ng/µl) were mixed in a 3:1(insert:vector) ratio. Ligase buffer (4 µl of 5xbuffer (250 mM Tris-HCl (pH. 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethene glycol-8000)), 1.0 µl (0.1unit) of T4 DNA ligase, and water to a final volume of 20 µl were added and the ligation mixture was incubated for one hour at room temperature. The resulting recombinant plasmids were used for transforming *E. coli DH5* electrocompetent cells. Portions of the transformed cell suspension (50, 100 and 200 µl) were plated on LB/amplicillin agar plates containing IPTG and X-gal and incubated overnight at 37°C. After overnight incubation, several white transformants were observed . The colonies were picked and streaked onto fresh LB/amplicillin agar plate containing 10 µl of 10 mM IPTG and 50 µl of 2% X-gal to test the recombinant clones and to serve as master

plates. A small amount of plasmid DNA was isolated using the Wizard Plus Minipreps DNA Purification Kit. The plasmid DNAs were then digested with *Hin*dIII and *Xba*I and the DNA fragments resolved on a 1% aagrose gel. This approach was used to determine whether the subclones contained the appropriate 2.0-kb *Hin*dIII/*Xba*I amplified oleosin gene fragment. A subclone, designated 5'flanking-oleosin/pGEM7Z(+) was selected and grown in 50 ml of LB media containing amplicillin to isolate more plasmid DNA for sequencing and for subcloning the 5'-flanking region of the oleosin gene upstream from the *PATE* gene in the binary vector constructs designated anti-sense PATE/pBI121 and sense PATE/pBI121. Both strands of the boundaries between the pGEM7Zf(+) vector and oleosin gene 5'-flanking region insert were sequenced using a primer-based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, TX).

The 5'-oleosin/pGEM7Zf(+) plasmid DNA was digested with *Hin*dIII and *Xba*I restriction enzymes and a 1% preparative agarose gel was run to purify the 2.0-kb 5'-flanking region of the oleosin gene from the pGEM7Zf(+) vector. The band containing the fragment was cut from the gel and purified from the gel band using the QIAquick Gel Extraction Kit (QIAGEN). The amounts of the purified 2.0-kb *Hin*dIII/*Xba*I 5'-flanking oleosin fragment and the binary vector anti-sense PATE/pBI121 and sense pBI121/PATE DNA constructs were estimated visually by comparison with standard bands with known sizes and amounts of DNA on a 0.8% agarose test gel. The 2.0-kb *Hin*dIII/*Xba*I 5'-flanking oleosin gene fragment was directionally subcloned into the *Hin*dIII/*Xba*I sites of the binary vector anti-sense PATE/pBI121 and sense PATE/pBI121 constructs in place of the constitutive CaMV 35S promoter, as shown in Figures 22 and 23. The 2.0-kb

Figure 17. The pBIN19 vector (redrawn from Bevan, 1984) contains the *Agrobacterium* Ti plasmid right (RB) and left border (LB) sequences, a nopaline synthase (Nos) gene promoter and terminator, a neomycin phosphotransferase II (NPT II-Kan^{R}) gene and a multiple cloning site (Bevan, 1984).

Abbreviations are : RB and LB, *Agrobacterium* Ti plasmid right and left border sequences, respectively, essential for integration into the plant genome; NOS-pro, nopaline synthase promoter; NPTII, Tn5 neomycin phosphotransferase II; NOS-ter, nopaline synthase terminator

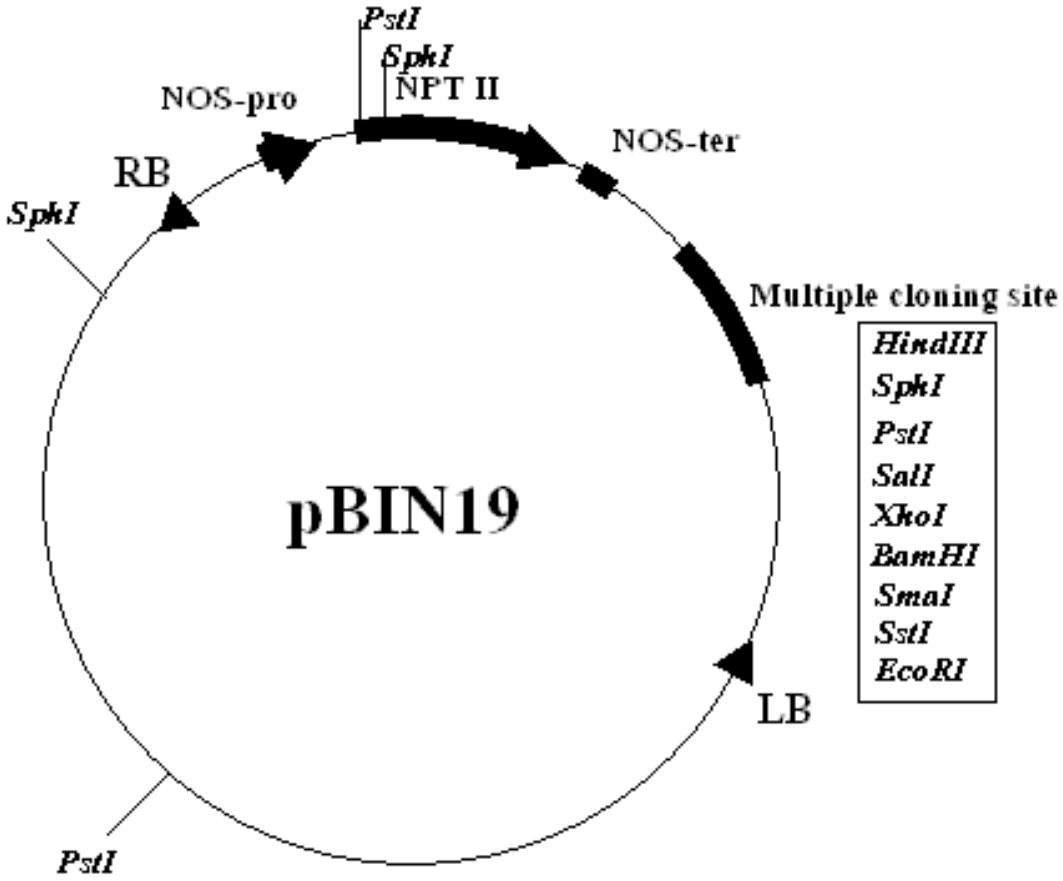

PstI
SphI
NPT II
NOS-pro
NOS-ter
RB
SphI
Multiple cloning site
HindIII
SphI
PstI
SalI
XhoI
BamHI
SmaI
SstI
EcoRI
pBIN19
LB
PstI

Figure 18. The pBI101 vector from Clontech (redrawn from Clontech manual, 1995), developed by Jefferson (1987,1989), encompasses the 1.9-kb bacterial beta-glucuronidase (*GUS*) reporter gene for detecting expression of genes from plant, yeast, and mammalian sources. The remainder of the parental binary vector pBIN19 (Bevan, 1984) is represented by the solid line. The pBI101 vector is designed for cloning and testing plant promoters using *GUS* expression. A 260-bp segment containing the polyadenylation signal from the nopaline synthase gene (NOS-ter) of the *Agrobacterium* Ti plasmid (Beavan et al., 1983) occurs downstream of the *GUS* gene. The NOS-promoter region from the nopaline synthase gene of the *Agrobacterium* Ti plasmid contains the TATA box and other essential promoter elements necessary for transcription of the bacterial NPT II-Kan^r gene in plant cells. pBI101 also contains a low-copy-number RK2 origin of replication.

Abbreviations are : RB and LB, *Agrobacterium* Ti plasmid right and left border sequences, respectively, essential for integration into the plant genome; NOS-pro, nopaline synthase promoter; NPTII, Tn5 neomycin phosphotransferase II; NOS-ter, nopaline synthase terminator; GUS, beta-glucuronidase.

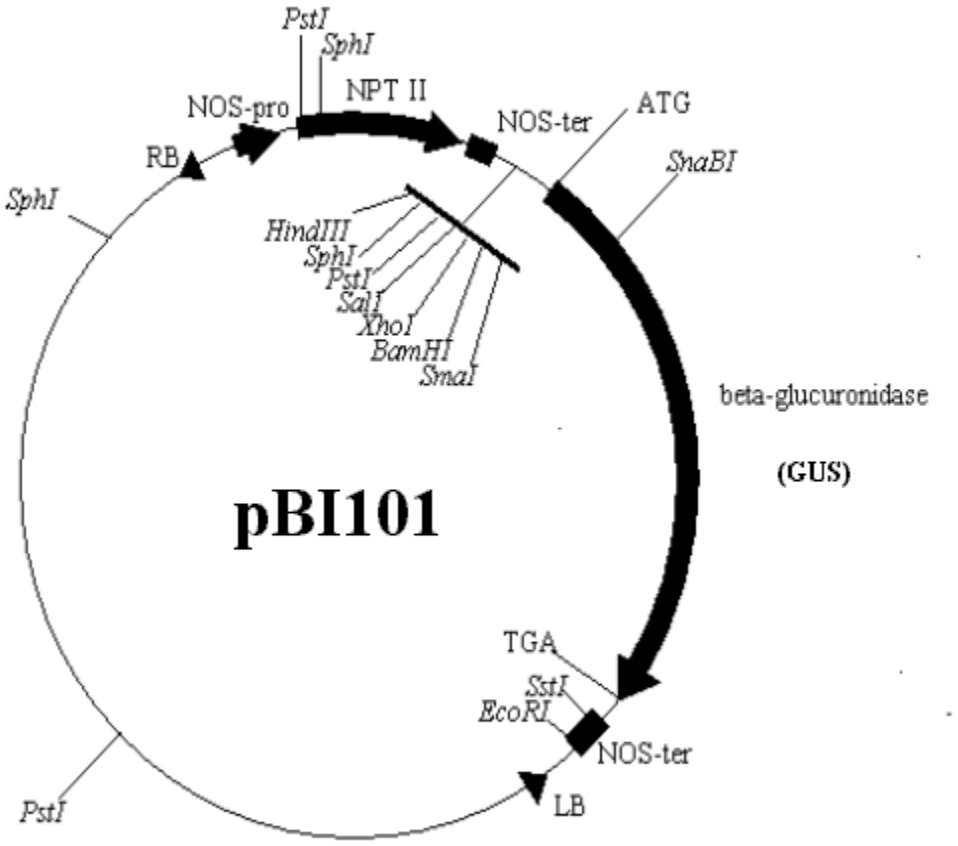

PstI
SphI
NPT II
NOS-pro
NOS-ter
ATG
RB
SnaBI
SphI
HindIII
SphI
PstI
SalI
XhoI
BamHI
SmaI
beta-glucuronidase
(GUS)
pBI101
TGA
SstI
EcoRI
NOS-ter
PstI
LB

Figure 19. pBI121 vector from Clontech (redrawn from Clontech manual, 1995) developed by Jefferson (1987,1989). pBI121 is the plasmid pBI101 with an 800-bp *Hin*dIII/*BamH*I fragment containing the cauliflower mosaic virus (CaMV) 35S promoter cloned upstream of the *GUS* gene. The CaMV 35S promoter is a constitutive promoter that expresses high levels of GUS activity.

Abbreviations are : RB and LB, *Agrobacterium* Ti plasmid right and left border sequences, respectively, essential for integration into the plant genome; NOS-pro, nopaline synthase promoter; NPTII, Tn5 neomycin phosphotransferase II; NOS-ter, nopaline synthase terminator; GUS, beta-glucuronidase; CaMV 35S promoter,cauliflower mosaic virus promoter.

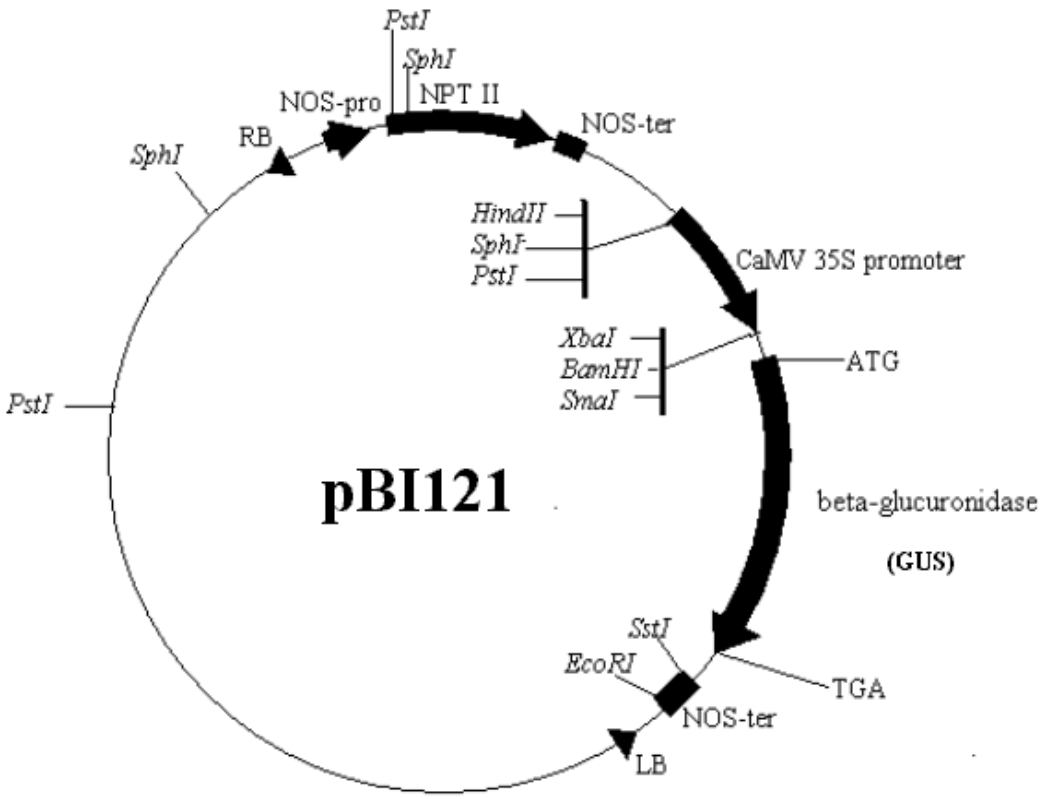

PstI
SphI
NPT II
NOS-pro
NOS-ter
RB
SphI
HindII
SphI
PstI
CaMV 35S promoter
XbaI
BamHI
SmaI
ATG
PstI
pBI121
beta-glucuronidase
(GUS)
SstI
EcoRI
TGA
NOS-ter
LB

Figure 20. Plasmid construct harboring the anti-sense *PATE* coding region. This recombinant plasmid was constructed from the pBI121 parental vector by replacing the *GUS* gene with the PCR-generated partial-length *PATE* coding region in an inverted orientation downstream from the CaMV 35S promoter.

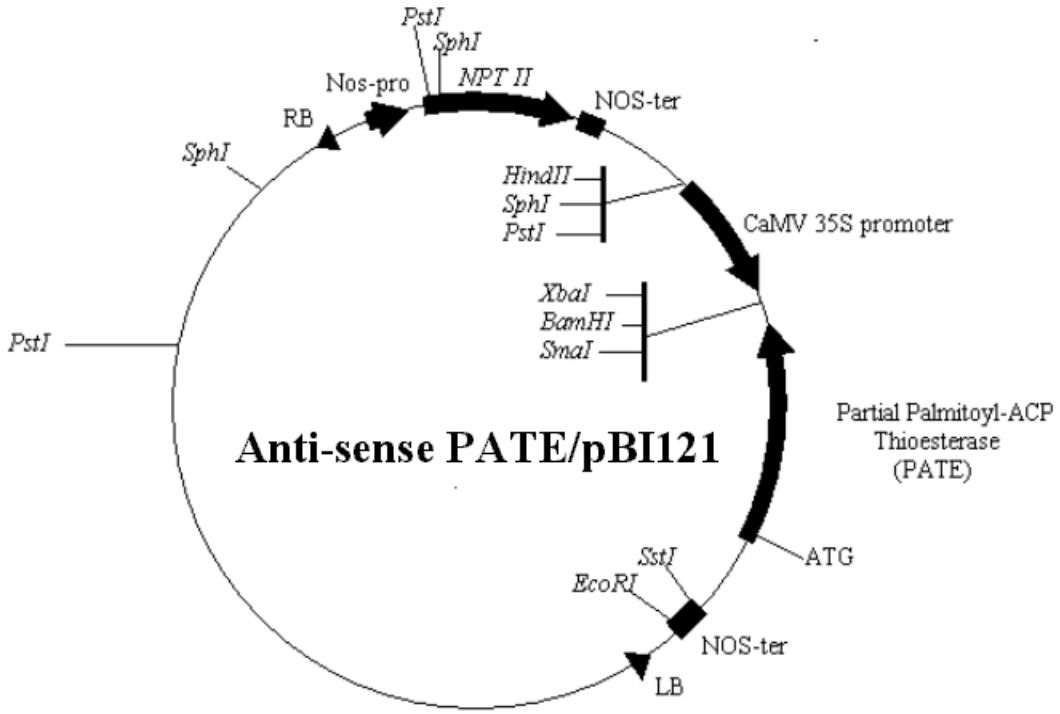

PstI
SphI
Nos-pro
NPT II
NOS-ter
RB
SphI
HindII
SphI
PstI
CaMV 35S promoter
XbaI
BamHI
SmaI
PstI
Anti-sense PATE/pBI121
Partial Palmitoyl-ACP
Thioesterase
(PATE)
ATG
SstI
EcoRI
NOS-ter
LB

Figure 21. Palmitoyl-ACP Thioesterase (*PATE)* sense construct in pBI121. This recombinant plasmid was constructed from the pBI121 parental vector by replacing the *GUS* gene with a PCR-generated full-length *PATE* coding region downstream from the CaMV 35S promoter cassette in the *Xba*I and *Sst*I sites of pBI121.

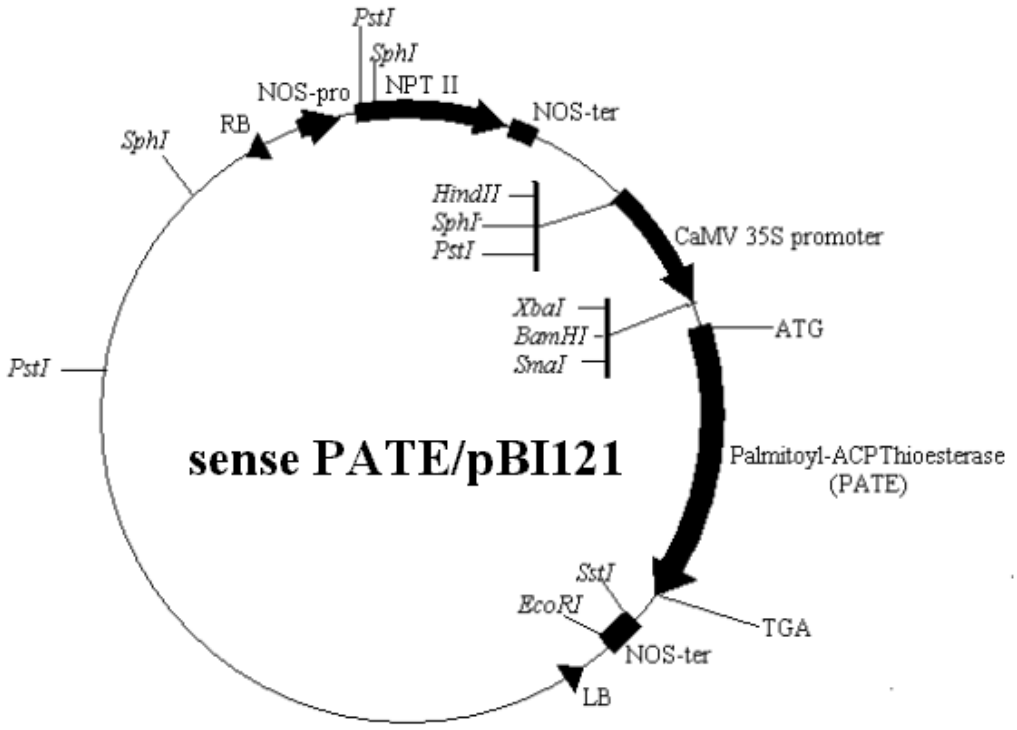
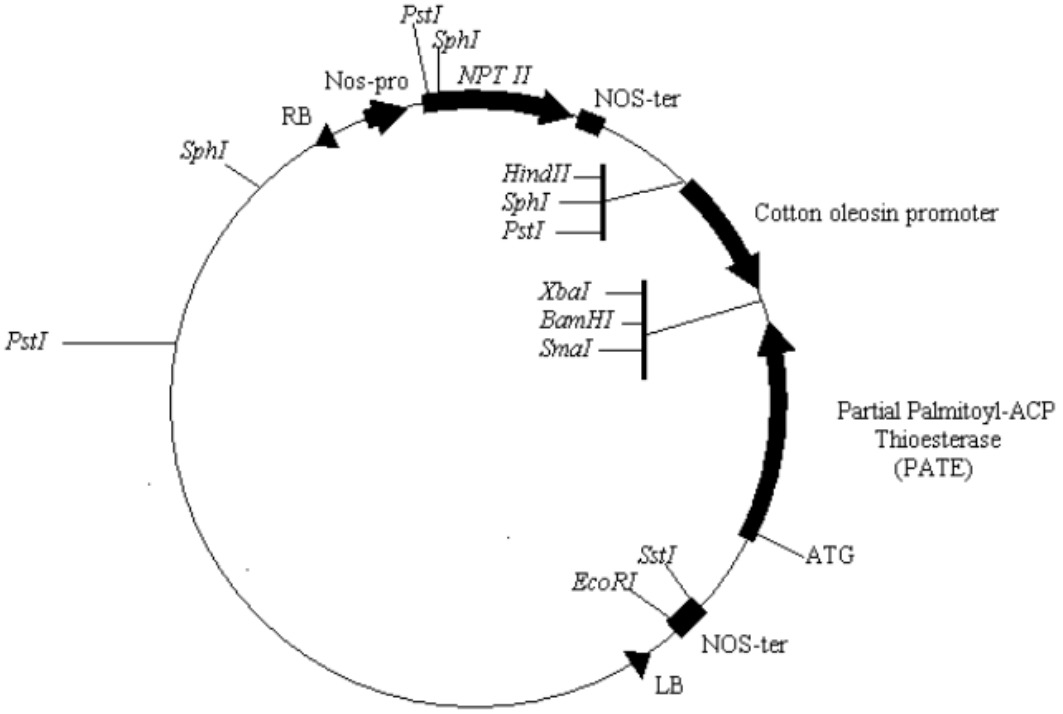
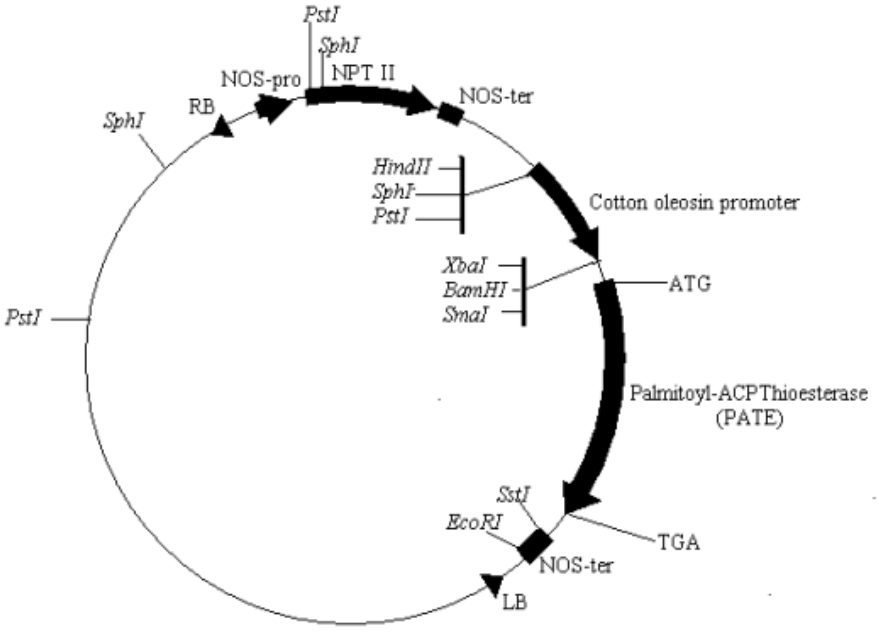

PstI
SphI
NPT II
NOS-pro
RB
NOS-ter
SphI
HindII
SphI
PstI
CaMV 35S promoter
XbaI
BamHI
SmaI
ATG
PstI
sense PATE/pBI121
Palmitoyl-ACPThioesterase
(PATE)
SstI
EcoRI
TGA
NOS-ter
LB

Figure 22. Diagram of the seed-specific *PATE* anti-sense construct in the binary vector pBI121 . This recombinant plasmid was constructed from the *PATE* anti-sense vector construct designated anti-sense 5'-flanking oleosin-PATE/pBI121 by replacing the CaMV 35S promoter with a 2.0-kb PCR-generated DNA fragment generated from the 5'-flanking region of the cotton oleosin gene (Hughes et al., 1993).



Anti-sense 5'-flanking oleosin-PATE/ pBI121

Figure 23. Diagram of the seed-specific *PATE* anti-sense construct in the binary vector pBI121. The plasmid construct designated sense 5'-flanking oleosin-PATE/pBI121 was constructed from the PATE sense construct vector designated by replacing the CaMV 35S promoter with a 2.0-kb PCR-generated fragment from 5'-flanking region of the cotton oleosin gene (Hughes et al., 1993).



sense 5'-flanking oleosin-PATE /pBI121

*Hin*dIII/*Xba*I 5'-flanking oleosin gene fragment and binary construct DNAs were mixed in a 5:1 ratio (insert:vector) for ligation. The resulting recombinant plasmids were used for transforming *E. coli DH5* electrocompetent cells, as previously described. Portions of the cell transformant suspensions (50, 100 and 200 μl) were plated on LB agar plates containing kanamycin and incubated overnight at 37°C.

The recombinant plasmids were then isolated from the transformed cells using the Wizard Plus Miniprep DNA Purification Kit (Promega) and double digested with *Hin*dIII/*Xba*I and *Xba*I/*Sst*I. The digested DNAs were then fractionated on a 1% agarose gel to determine which subclones contained the appropriate 2.0-kb 5'-flanking oleosin gene fragment. The new constructs, designated sense 5'-flanking oleosin-PATE /pBI121 and anti-sense 5'-flanking oleosin-PATE/pBI121, were selected and grown in 100 ml of LB media containing kanamycin in order to isolate sufficient plasmid DNA for sequencing and for transforming electrocompetent *Agrobacterium tumefaciens LBA4404* cells. Both strands of the boundaries between the pBI121 vector and insert were sequenced using a primer-based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, TX). Both seed-specific constructs were then electroporated into electrocompetent *Agrobacterium tumefaciens LBA4404* cells as previously described.

Results and Discussion

Construction of *Agrobacterium tumefaciens* binary plasmid vector for constitutive anti-sense RNA suppression of cotton palmitoyl-ACP thioesterase gene expression

Using a partial cotton palmitoyl-ACP thioesterase cDNA in the plasmid subclone SKCPc115b as the template with the forward primer, 5'-CCGAGCTCAGACTGCTGTG ACATCG-3' (with a *Sst*I site (underlined)) and the reverse primer 5'CCGCTAGAAGAG CCATCTCACATGG-3' (with a *Xba*I site (underlined)), a 1.5-kb partial cotton palmiotyl-ACP thioesterase PCR fragment was amplified by PCR, shown in Figure 24. This 1.5-kb PCR fragment was then directionally subcloned into the *Xba*I/*Sst* I site of the prepared pBI121 vector in the anti-sense orientation, downstream from the constitutive CaMV 35S promoter (as shown in Figure 20). The resulting recombinant 14.5-kb plasmid was used for transforming electrocompetent *E. coli DH5α* cells. At least 30 colonies were formed in the selective agar plates. Plasmid DNAs from these transformants were isolated and doubly digested with *Xba*I/*Sst*I restriction endonucleases. The digested plasmid DNAs were then fractionated on a 1% agarose gel. The clone #2 (as shown in Figure 25) generated the expected 1.5-kb band. This clone was then selected for large-scale plasmid preparation and its name was re-designated as anti-sense PATE /pBI121.

The anti-sense PATE /pBI121 DNA was subjected to digestion with *Sst*I, *Xba*I and *Xba*I/*Sst*I and the products were characterized by agarose gel electrophoresis (shown in Figure 26). To futher characterize this clone, the anti-sense PATE/pBI121 DNA was used as a template in the PCR amplification using the same primers as when generating 1.5-kb fragment insert for subcloning. The PCR product as shown in Figure 27 generated

a 1.5-kb fragment of the same size when pSKCPc115b DNA was used as a template. Both strands of the boundaries between pBI121 and insert and the 1.5-kb insert itself were sequenced using a primer-based approach. The DNA sequence of the boundaries between pBI121 and the 1.5-kb insert is shown in Figure 28. Analysis of the DNA sequence was done to compare between 1.5-kb insert and cotton *PATE* cDNA and also between the boundaries and pBI121 vector (data not shown). The sequences upstream before the *Xba*I site and downstream after the *Sst*I site are identical with the sequences of the pBI121 vector. The sequences between *Xba*I and *Sst*I sites are identical with the sequences of *PATE* cDNA. The anti-sense PATE/pBI121 was then used for transforming electrocompetent *Agrobacterium tumefaciens LBA4404* cells. Plasmid DNA was isolated from transformed *Agrobacterium* cells and subjected to digestion with *Sst*I, *Xba*I and *Xba*I/*Sst*I, and further characterized by agarose gel electrophoresis and alkaline blot hybridization (as shown in Figures 29 and 30). Transformed *Agrobacterium* cells indeed contain the anti-sense PATE/pBI121 construct, as can be seen by the presence of 1.5-kb hybridizing fragment. These transformed *Agrobacterium tumefaciens LBA4404* cells were then used in cotton transformation by Dr. Tu T. Huynh of Dr. Kent D. Chapman's laboratory (Huynh, 2001).

From Dr. Huynh's results, the expression of the anti-sense PATE/pBI121 construct in transgenic cotton plants did not have a reduction in the relative percentages of palmitic acid in somatic embryos compared to the pBI121 control transgenic plants and wild type plants (cv. Coker3.2). The fatty acid compositions of pooled T_1 seed cotyledons of 2 day-old cotton seedlings and mature leaves were similar to wild type (cv. Coker 3.2) and vector only (pBI121 control transgenic plants). The anti-sense transgenic

cotton plants had two different phenotypes. About 30% of the anti-sense transgenic plants displayed a dwarfed phenotype. The others had a phenotype similar to wild type plants. Both the dwarfed and the non-dwarfed anti-sense transgenic cotton plants were shown to have the transgene was incorporated into the cotton genome. The dwarfed anti-sense transgenic cotton plants had smaller bolls with a much reduced viable seeds and smaller leaves than the normal phenotype transgenic cotton plants and control plants (wild-type, Coker 312 cv. and vector-only control, pBI121). By comparing equivalent amounts of dwarf tissues with control tissues, the overall fatty acid composition in the T_1 seeds of dwarf plants or leaves of mature T_0 primary transformants were pretty much the same as in the normal plants. This suggests that plants tend to distribute a sufficient amounts of palmitic acid to only one or two viable seeds rather than to distribute an insufficient amounts of palmitic acid into many non-viable seeds when the level of palmitic acid was decreased. A reduction in palmitic acid levels may also affect cell expansion. Leaf expansion in developing leaves of dwarf plants was prematurely stopped as compared with the normal phenotype of transgenic cotton plants. This might be because the amounts of fatty acids available for incorporation into membrane lipids is insufficient and thus hinder cell expansion. Therefore, palmitic acid levels may be regulated in such a way that a certain range of palmitic acid levels must be maintained to avoid lethal effects (Huynh, 2001).

Figure 24. Agarose gel electrophoresis of the PCR product generated from the plasmid SKCPc115b as template to generate a 1.5-kb anti-sense *PATE* fragment, and pBI121 DNA digested with *Sst*I/*Xba*I restriction endonucleases. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively.

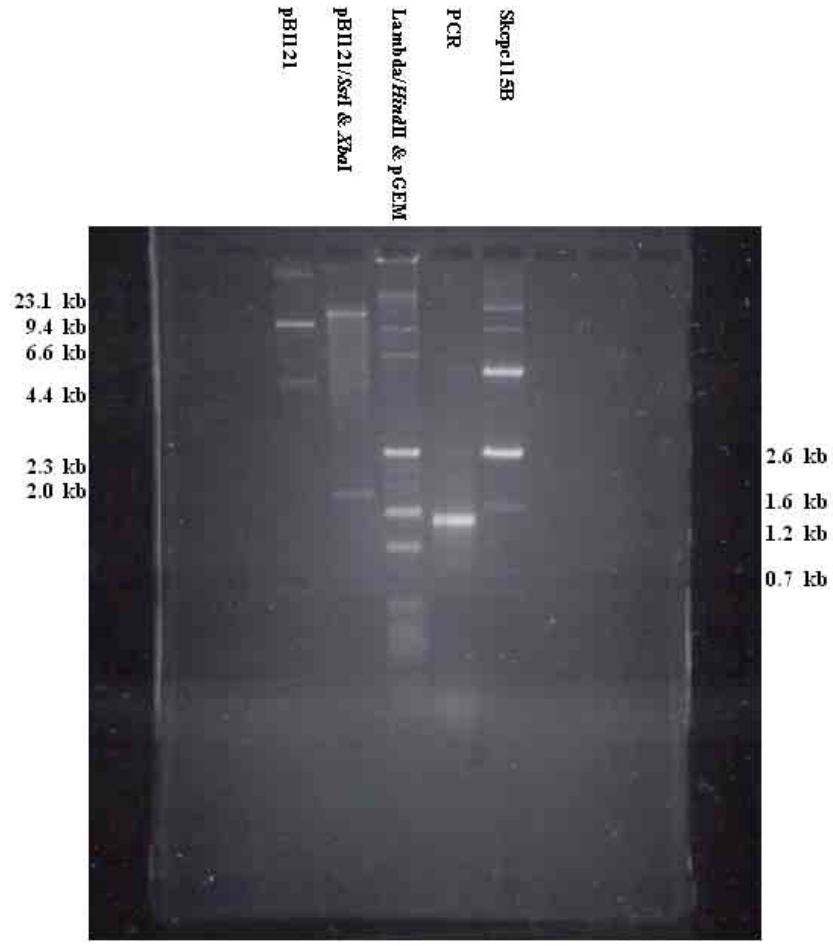

pBI121
pBI121/SstI & XbaI
Lambda/HindIII & pGEM
PCR
Skcpc115B
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 25. Agarose gel electrophoresis of putative anti-sense PATE/pBI121 clones. Plasmid DNAs isolated from these clones were digested with *Xba*I/*Sst*I restriction endonucleases and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively.

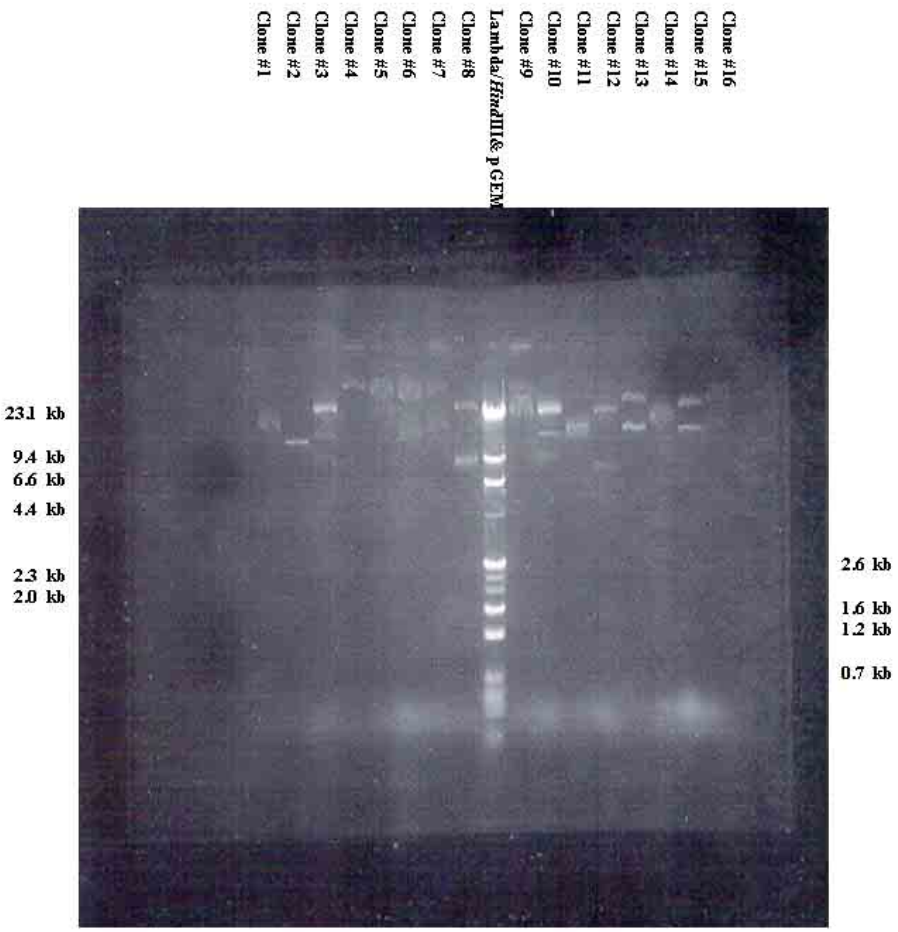

Clone #1
Clone #2
Clone #3
Clone #4
Clone #5
Clone #6
Clone #7
Clone #8
Lambda/HindIII& p GEM
Clone #9
Clone #10
Clone #11
Clone #12
Clone #13
Clone #14
Clone #15
Clone #16
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 26. Representative agarose gel of DNA fragments from restriction endonuclease digestion of Clone #2 (anti-sense PATE/pBI121). Plasmid DNA from this clone was digested with the restriction endonucleases as indicated. The resulting fragments were resolved on a 1% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) of λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively.

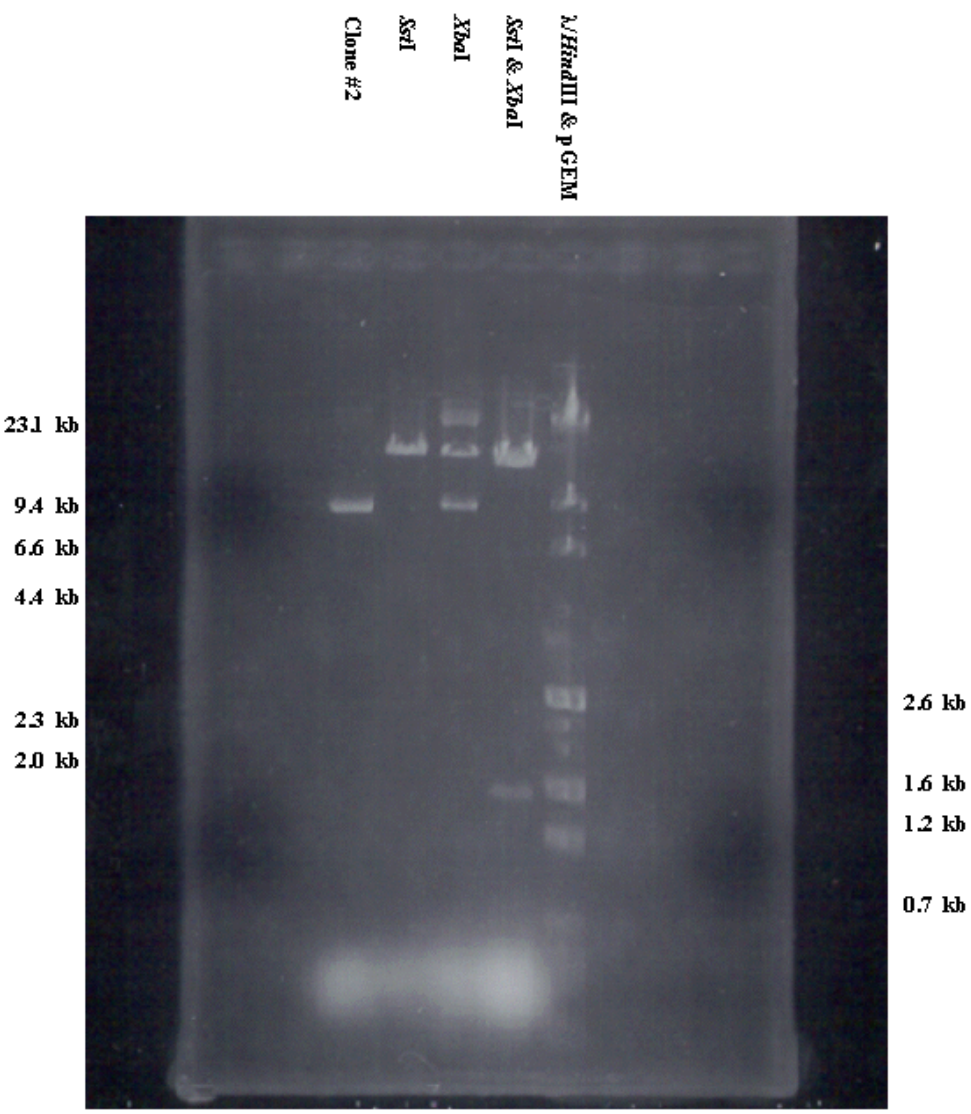

λ/HindIII & pGEM
SstI & XbaI
XbaI
SstI
Clone #2
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 27. Agarose gel electrophoresis of the PCR product using the anti-sense PATE/ pBI121 as a template compared with the PCR product using using plasmid SKCPc115b DNA as a template. The resulting fragments were resolved on a 1% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

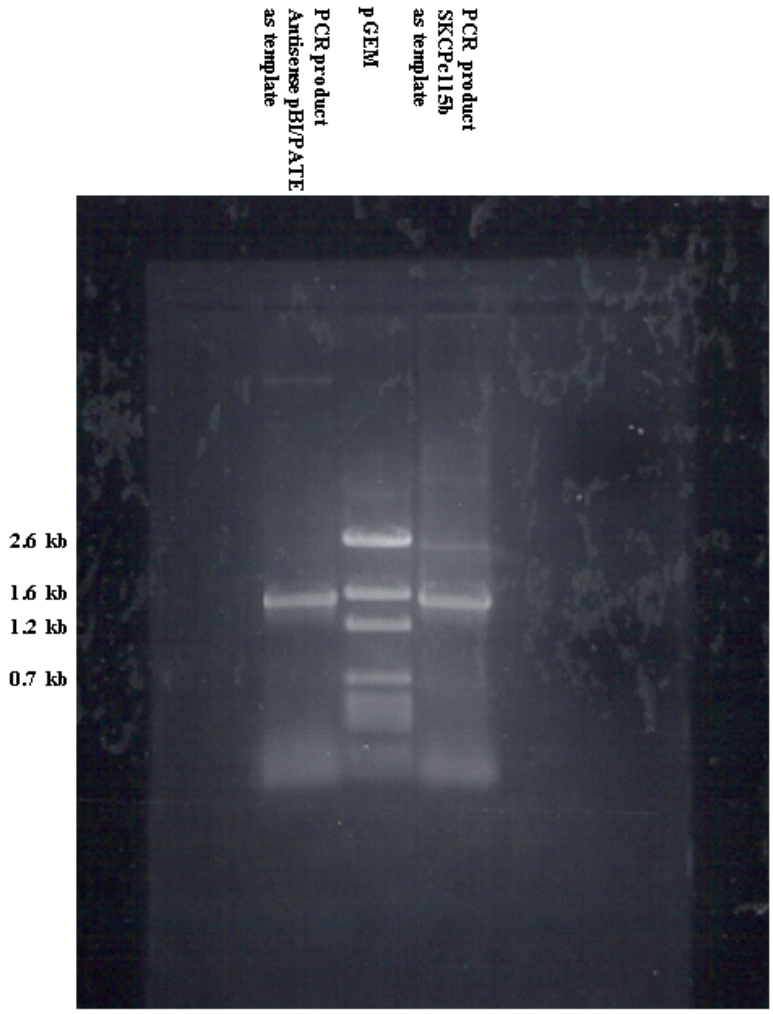

PCR product
SKCPc115b
as template
pGEM
PCR product
Antisense pBI/PATE
as template
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 28. The nucleotide sequence of the *Agrobacterium tumefaciens* binary plasmid vector construct designated anti-sense PATE/pBI121 for constitutive anti-sense RNA suppression of the cotton *PATE* gene. The number on the right refers to nucleotide residues. The sequences used for designing the forward and reverse primers for the anti-sense *PATE* gene are indicated by arrows and bold type. The sequences upstream before the *Xba*I site and downstream after the *Sst*I site are the sequences of the pBI121 vector DNA (Clontech, http://www.clontech.com). The sequences between *Xba*I and *Sst*I sites are matched with the sequences of *PATE* cDNA (Pirtle et al., 1999).

CCACGTCTTC AAAGCAAGTG GATTGATGTG ATATCTCCAC TGACGTAAGG GATGACGCAC 60

AATCCCACTA TCCTTCGCAA GACCCTTCCT CTATATAAGG AAGTTCATTT CATTTGGAGA 120

XbaI →

GAACACGGGG GAC**TCTAGAAG AGCCATCTCA CATGG**CTATT CAATTTTTCA AAAGCAGAC 180

CACCTTTCCA TGGAAAGAGC CTATGCTGGT TAGGGCACGG GCCCAACTTC ACAAAATCCG 240

TGTCCGAGGC TCGGGTCATA CGAGCTGAGC CTTACAAGGC TATGTACGAA AACCCGACTC 300

CTAAACAAGC CTAATTACAA CCCCCTTTTT TTTTCTCAAA GGAAGACTAA TAATCATAAG 360

GAATATATGT ATATAAATCC GAAAAGAGCA CGATAGATAC TACACACAGC AATGAGATTA 420

CTTGGATTTC TATGCCCTTT TTGCGGTAAT TTGATCCATG TTAGCGGAAC TTTTGGCATG 480

TTTAGGCCTC CATCGGGTCC TGCCTCTCAC AATCTCAGCT CCATCGTCGA GTCGGAGCAA 540

ATGTTGGCAA TTAAATTCAC CAACATTTAC TGCATTTTCC GTATTGGAAT CAGACACAGT 600

GGTCAGTGAC TGCAGCACGC TGTCCCTCCC GCACTCCCTC CTATATTCCA GAGTCAAGGC 660

GGAAAGCTCG TGACTCTCCA AGATTGGTAA TGGAGCACTC TCAAGGATCC AGCCAATGTA 720

CTTCACATTA TTGACATGCT GGTTGACATC CAAGTCGCTC CATTTAGGAG TTAAACCTTT 780

GCACACGTGT TCAGCTGTGC TGTCATCGAG TTTCACTAGT TTCTGGCTAT CCTCAGCCAG 840

AACAGGATCT GAATTCATAA AAAAAGGTTC TATTTCCCCT CGAACCTCTT CTGGGATTTT 900

AGATAACCTT CTAGTCAGTT TATTCATCAT CACCCATACA CTTGTGGCTC GTGTTAAAAT 960

TTCACCAGTT TCACTATTGC TGACAAGCCA ATCTCTTCGC ATGCCATTCT TCCCCGATGC 1020

ACTGACCCAA GTGTCGACTT GAACAACATC ACCCCAAGTA GGATAGCGAT CAACCACAAC 1080

TTGCATCCGT GTGACAACCC ATATTAGGTT CTTCTTGCAC ATCTCAGGTG TTGCACCAAA 1140

ACCTTCTCCA AGCAGTCCAG CACTTCGACA ATGATTTATA GCTGTTTCCT GTAAATGATT 1200

CATTAGTGTC TCTATGGATG CTGTTTGATC AGCGCCTATC TCATATGATC TAATCGAGAA 1260

GTTCTGACTG AAAACAAGAC CATCCTGAAC AATCTTCCCT ATGCCAAACG GATCAATGAC 1320

CATGTCAGGC CGCCTCGGCT TCCAATCAAG CATCATCCAC TGCTTCTCAG CAGCCAAGAA 1380

AATGGTTGTG ATAGCAGCAA GAAGCATGCT CCAATCAGGT AACTGGTTGA TAAACGTCCT 1440

AGGAGGAGGG GAACTTGCAC CGTCATCGTT CTTGGAACCT TCCACGGGAG TCGTCGACGC 1500

CACAGTGCCG TTTATTTTCG GAGGTGCTTG AGCATTTGCC TTGACTTGCA AACTTCCAGA 1560

AGAAACCGAT GGCTTCGACT TGATGCTTCC GAGCTTCTTG TTTTTCGAGT CAGAGGAGTC 1620

← ***Sst*I**

AGGTGAAGAA GTGACTGGGA AAAACGC**CGA TGTCACAGCA GTCTGAGCTC** GAATTTCCCC 1680

GATCGTTCAA ACATTTGGCA ATAAAGTTTC TTAAGATTGA ATCCTGTTGC CGGTCTTGCG 1740

ATGATTATCA TATAATTTCT GTTGAATTAC GTTAAGCATG TAATAATTAA CATGTAATGC 1800

ATGACGTTAT TTATGAGATG GGTTTTTATG ATTAGAGTCC CGCAATTATA CATTTAATAC 1860

GCGATAGAAA ACAAAATATA GCGCGCAAAC TAGGATAAAT TATCGCGCGC GGTGTCATCT 1920

ATGTTACTAG ATC 1980

Figure 29. Representative agarose gel of DNA fragments produced by restriction endonuclease digestion of the plasmid construct anti-sense PATE/pBI121 isolated from transformed *Agrobacteriun tumefaciens LBA4404* cells. Plasmid DNA from the transformants was digested with the restriction endonucleases as indicated. The resulting fragments were resolved on a 1% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively.

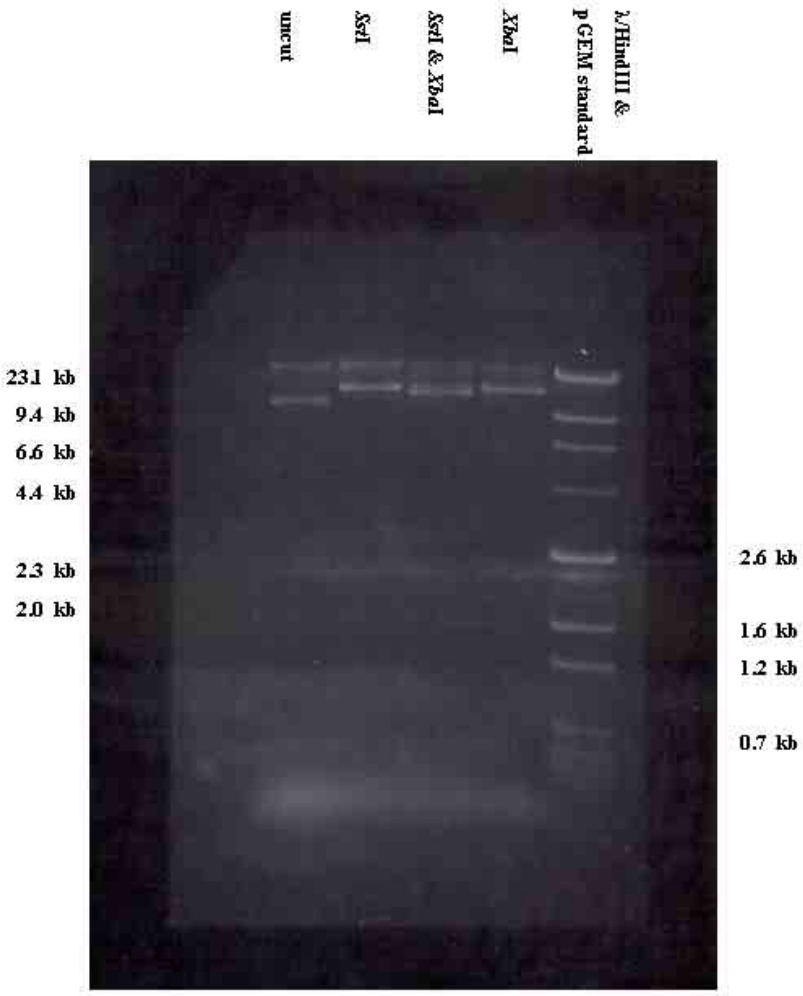
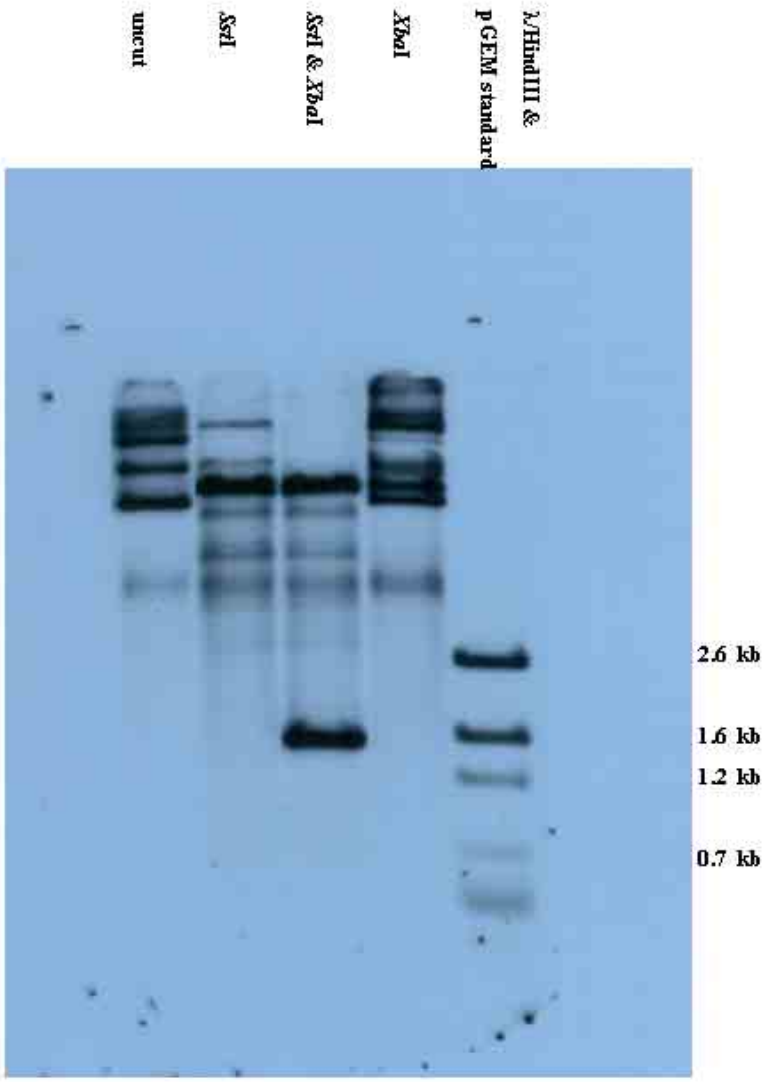

λ/HindIII &
pGEM standard
XbaI
SstI & XbaI
SstI
uncut
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 30. Alkaline blot of the DNA fragments from digested anti-sense PATE/pBI121 isolated from transformed *Agrobacterium tumefaciens LBA4404* cells. The anti-sense PATE/pBI121 fragments fractionated on 1% agarose gel were transferred to a positively-charged nylon membrane in a 0.4 N NaOH solution and hybridized to ^{32}P-labeled DNA fragments generated from a heterologous *Arabidopsis* DNA fragment by the random priming procedure of Feinberg and Vogelstein (1983). The heterologous probe was derived from 1.4-kb *Sal*I/*Eco*RV fragment isolated from the 5'-region of the *Arabdopsis thaliana* cDNA clone designated TE 3-2. The sizes (in kb) pGEM standard DNA fragments are shown on the right.


uncut
SstI
SstI & XbaI
XbaI
λ/HindIII &
pGEM standard
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Construction of *Agrobacterium tumefaciens* binary plasmid vector for constitutive gene-transgene co-suppression of palmitoyl-ACP thioesterase gene expression

A full-length cotton *PATE* coding region in pZero-2.1 obtained from Dr. Tu Huynh of Dr. Chapman's laboratory (Huynh, 2001) was used as a template. For PCR amplification of a 1.35-kb full-length cotton *PATE* coding region fragment (shown in Figure 31), the forward primer was 5'-CGTCTAGAATGGTTGCTACTGCTGTGA CATCG-3' with the *Xba*I site (underlined) and the reverse primer was 5'-CCGAGCTC TCAAAGGAAGACTAATAAT CATAAGG-3' with the *Sst*I site (underlined). This 1.35-kb PCR fragment was then directionally subcloned into the *Xba*I/*Sst*I site of the plasmid vector pGEM7Zf(+) (Promega). The resulting recombinant plasmid designated was used for transforming electrocompetent *E. coli DH5α* cells. At least 24 colonies formed on the selective agar plates. Plasmid DNAs were isolated from these transformants and digested with *Xba*I/*Sst*I restriction endonucleases, and the digested plasmid DNAs were then fractionated on a 1% agarose gel. The clone #17 DNA as shown in Figure 32 gave a 1.35-kb band upon digestion with *Xba*I and *Sst*I. This clone was then chosen for a large scale plasmid preparation and its name was re-designated to PATE/pGEM. The 1.35-kb DNA fragment from PATE/pGEM was isolated and purified away from the pGEM7Zf(+) vector DNA use for subcloning the fragment into *Xba*I/*Sst*I site of pBI121 in the sense orientation, downstream from the constitutive CaMV 35S (as shown in Figure 21).

The resulting recombinant 14.35-kb plasmid designated sense PATE/pBI121 was used for transforming electrocompetent *E. coli DH5α* cells. At least 15 colonies were

formed in the selective agar plates. Plasmid DNAs were isolated from these transformants and digested with *Xba*I/*Sst*I restriction endonucleases. The digested plasmid DNAs were then fractionated on a 1% agarose gel. Clone #5 (as shown in Figure 33) generated the expected a 1.35-kb band upon digestion with *Xba*I and *Sst*I. This clone was then chosen for a large scale plasmid preparation and its name was redesignated to sense PATE/pBI121. The sense PATE/pBI121 plasmid was subjected to digest with *Xba*I/*Sst*I, *Bam*HI/*Xba*I, *Xba*I/*Sal*I and *Xba*I/*Eco*RI restriction endonucleases and characterized by agarose gel electrophoresis as shown in Figure 34. Both strands of the boundaries between pBI121 and insert and the 1.35-kb insert itself were sequenced using a primer- based approach. The nucleotide sequence of the boundaries between pBI121 and insert and the 1.35-kb insert is shown in Figure 35. Analysis of the DNA sequence was done to compare between 1.35-kb insert and cotton *PATE* cDNA and also between the boundaries and pBI121 vector (data not shown). The sequences upstream before the *Xba*I site and downstream after the *Sst*I site are the sequences of the pBI121 vector. The sequences between *Xba*I and *Sst*I sites are identical with the sequences of *PATE* cDNA. The sense PATE/pBI121 was then used for transforming electrocompetent *Agrobacterium tumefaciens LBA4404* cells. Plasmid DNA was isolated from transformed *Agrobacterium* and subjected to digestion with *Sst*I, *Xba*I and *Xba*I/*Sst*I and further characterized by agarose gel electrophoresis and alkaline blot hybridization (data not shown). The transformants were shown to indeed contain the *PATE* coding region in the sense direction. The transformed *Agrobacterium tumefaciens LBA4404* cells were then used in cotton transformation by Dr. Tu T. Huynh of Dr. Kent D. Chapman's laboratory (Huynh, 2001).

The following results with the plasmid construct sense PATE/pBI121 in cotton plants were obtained from Dr. Tu T. Huynh of Dr. Kent D. Chapman's laboratory (Huynh, 2001). The fatty acid composition profile of somatic embryos overexpressing the sense PATE/pBI121 construct was different from the binary vector only (pBI121) control and wild type control (cv. Coker 3.2). Palmitic acid levels in the somatic embryos overexpressing the sense PATE/pBI121 construct was elevated as compared to the vector only (pBI121) control embryos and wild-type control (cv. Coker 3.2) embryos. Two transgenic lines had up to 70% of the total palmitic acid while the vector only control lines had only about 25% palmitic acid similar levels from wild-type lines. There was about three times more plamitic acid in these transgenic somatic embryos than is normally found in zygotic embryos (wild-type). Stearic acid levels (18:0) were also elevated up to 20% of the total in these transgenic lines. A corresponding drop in oleic acid (18:1) and linoleic acid (18:2) levels was also anticipated. Cell-free extracts of somatic embryos in these transgenic lines were shown to contain significantly higher amounts of the immuno-reactive 37kDa FatB protein compared to the vector only and wild type controls (Huynh, 2001). This suggests that indeed the *FatB* gene is over-expressed in these transgenic lines, resulting in the elevated levels of palmitic acid in these tissues. Therefore, overexpression of the*FatB* gene directly influences fatty acid composition profile of the transgenic somatic embryos.

Figure 31. Agarose gel electrophoresis of the PCR product using the full-length cotton *PATE* coding region in pZero 2.1 (Huynh, 2001) as a template to generate a 1.35-kb sense full-length cotton *PATE* coding region fragment. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively.

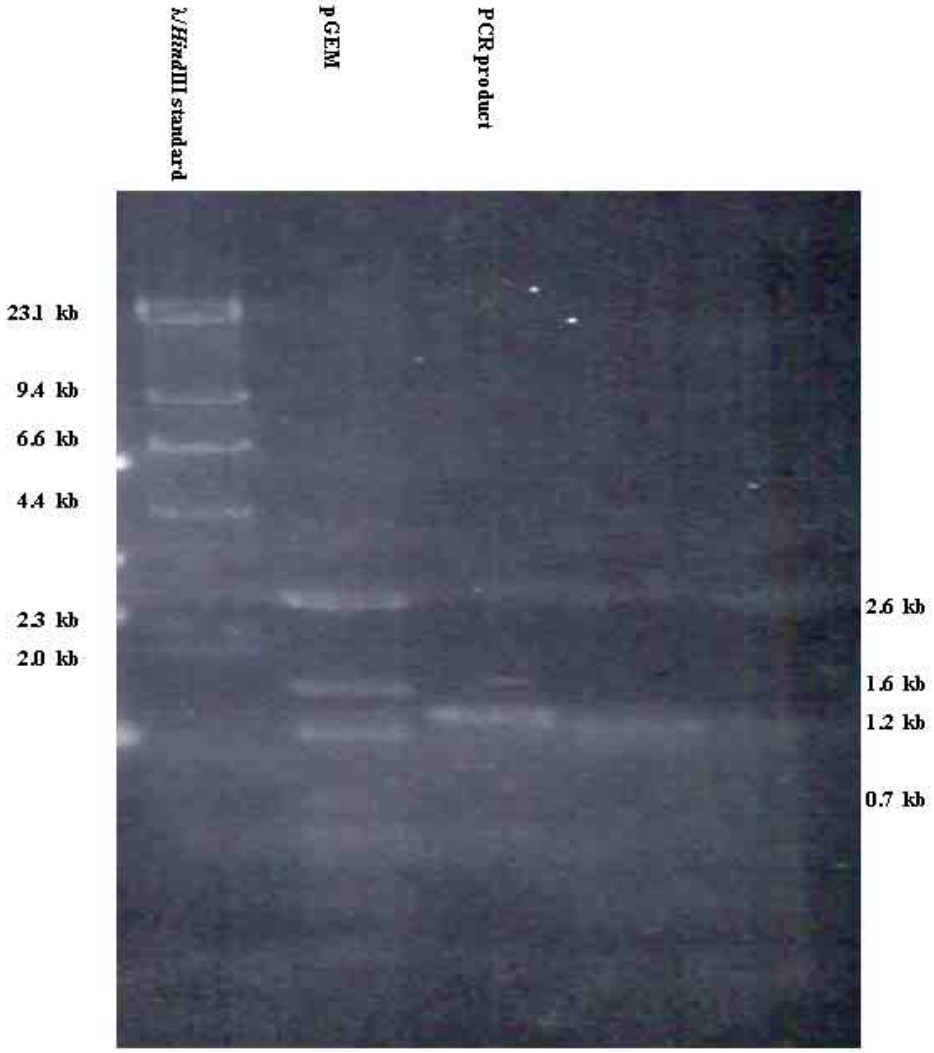

λ/HindIII standard
pGEM
PCR product
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 32. Agarose gel electrophoresis of DNA fragments derived from the putative PATE /pGEM constructs. Plasmid DNAs isolated from these clones were doubly digested with the *Xba*I/*Sst*I restriction endonucleases and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of the pGEM standard DNA fragments (Promega) are shown on the right.

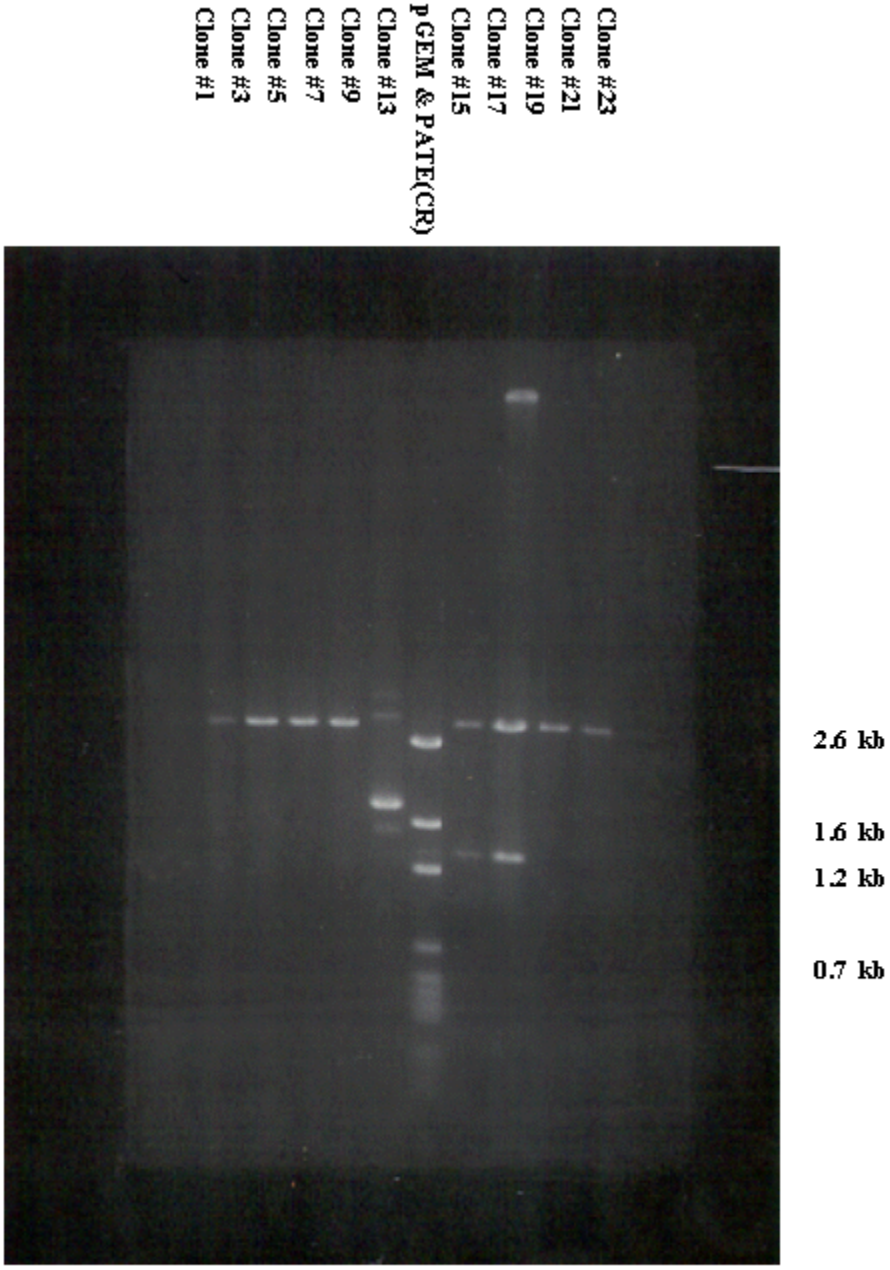

Clone #23
Clone #21
Clone #19
Clone #17
Clone #15
pGEM & PATE(CR)
Clone #13
Clone #9
Clone #7
Clone #5
Clone #3
Clone #1
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 33. Agarose gel electrophoresis of DNA fragments derived from the putative sense PATE/pBI121 constructs. Plasmid DNAs isolated from these clones were digested with the *Xba*I/*Sst*I restriction endonucleases and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively.

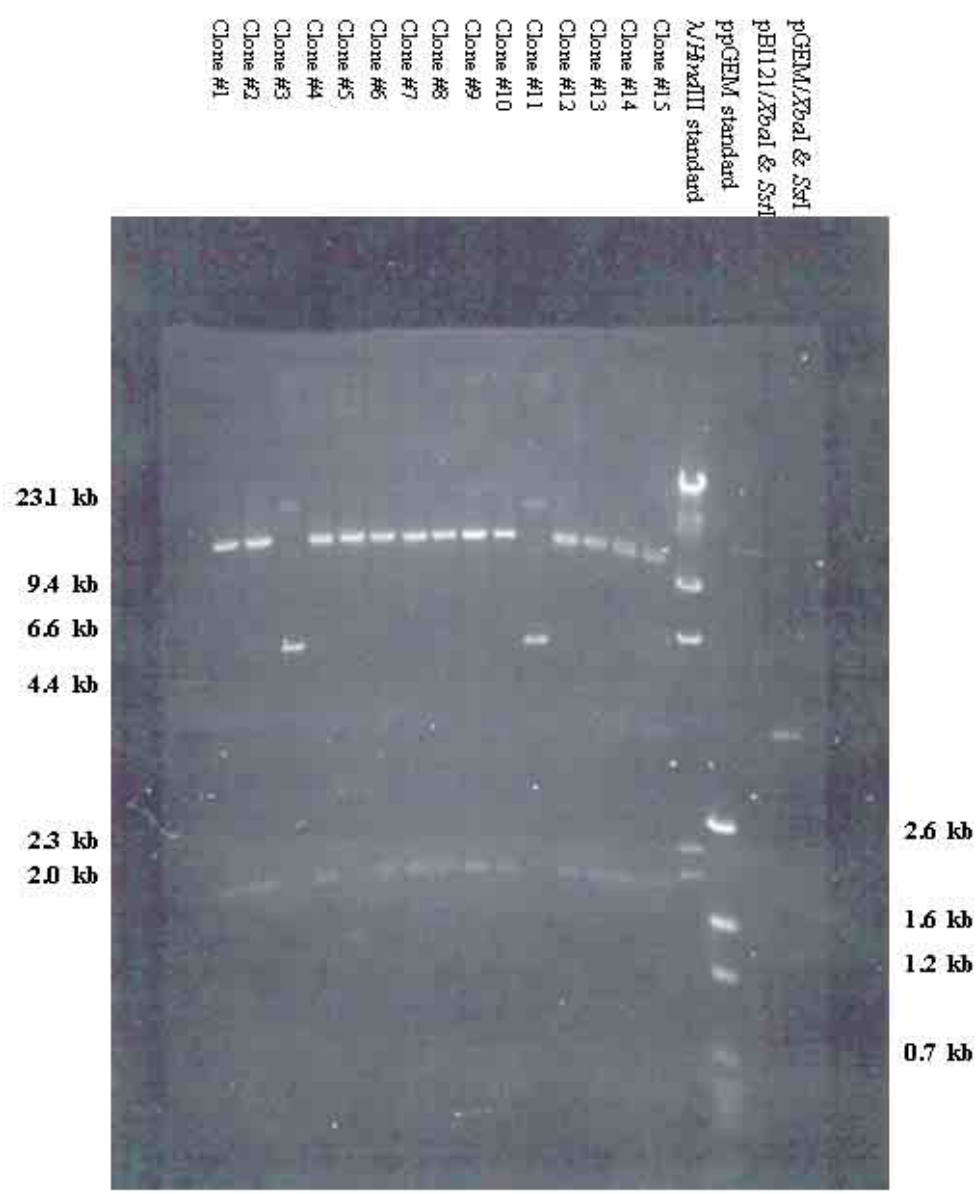

pGEM/XbaI & SstI
pBI121/XbaI & SstI
ppGEM standard
λ/HindIII standard
Clone #15
Clone #14
Clone #13
Clone #12
Clone #11
Clone #10
Clone #9
Clone #8
Clone #7
Clone #6
Clone #5
Clone #4
Clone #3
Clone #2
Clone #1
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 34. Representative agarose gel of DNA fragments generated by restriction endonuclease digestion of Clone #5 (sense PATE/pBI121). Plasmid DNA isolated from this clone was digested with the restriction endonucleases as indicated. The resulting fragments were resolved on a 1% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

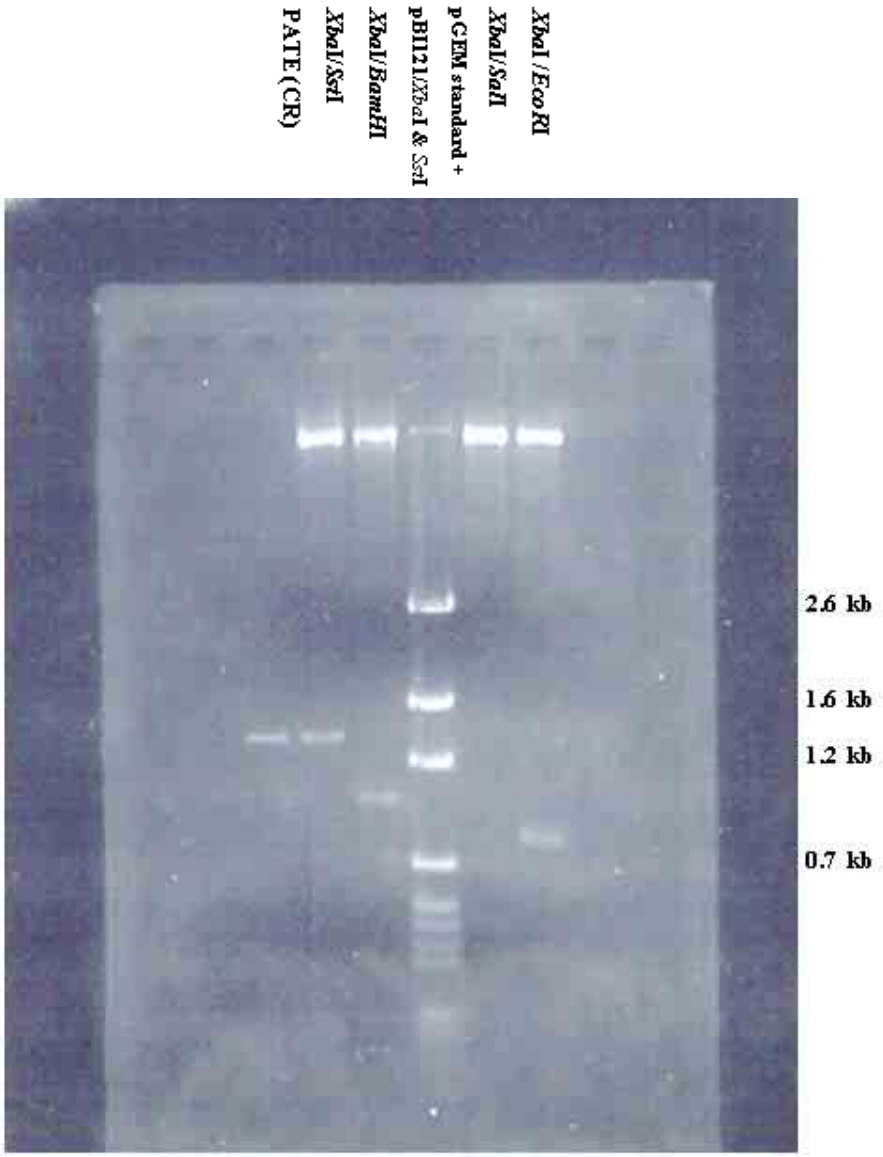

XbaI /EcoRI
XbaI/SalI
pGEM standard +
pBI121/XbaI & SstI
XbaI/BamHI
XbaI/SstI
PATE (CR)
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 35. The nucleotide sequence of the *Agrobacterium tumefaciens* binary plasmid vector sense PATE/pBI121 for constitutive gene-transgene co-suppression of *PATE* gene expression. The numbers on the right refers to nucleotide residues. The sequences used for designing the forward and reverse primers for the sense *PATE* coding region as are indicated by arrows and bold letters. The sequences upstream before the *Xba*I site and downstream after the *Sst*I site are identical with the sequences of the pBI121 vector. The sequences between *Xba*I and *Sst*I sites are identical with the sequences of cotton *PATE* cDNA (Pirtle et al., 1999).

CCACGTCTTC AAAGCAAGTG GATTGATGTG ATATCTCCAC TGACGTAAGG GATGACGCAC 60

AATCCCACTA TCCTTCGCAA GACCCTTCCT CTATATAAGG AAGTTCATTT CATTTGGAGA 120

XbaI →

GAACACGGGG GA**CTCTAGAA TGGTTGCTAC TGCTGTGACA TCG**GCGTTTT TCCCAGTCAC 180

TTCTTCACCT GACTCCTCTG ACTCGAAAAA CAAGAAGCTC GGAAGCATCA AGTCGAAGCC 240

ATCGGTTTCT TCTGGAAGTT TGCAAGTCAA GGCAAATGCT CAAGCACCTC CGAAAATAAA 300

CGGCACTGTG GCGTCGACGA CTCCCGTGGA AGGTTCCAAG AACGATGACG GTGCAAGTTC 360

CCCTCCTCCT AGGACGTTTA TCAACCAGTT ACCTGATTGG AGCATGCTTC TTGCTGCTAT 420

CACAACCATT TTCTTGGCTG CTGAGAAGCA GTGGATGATG CTTGATTGGA AGCCGAGGCG 480

GCCTGACATG GTCATTGATC CGTTTGGCAT AGGGAAGATT GTTCAGGATG GTCTTGTTTT 540

CAGTCAGAAC TTCTCGATTA GATCATATGA GATAGGCGCT GATCAAACAG CATCCATAGA 600

GACACTAATG AATCATTTAC AGGAAACAGC TATAAATCAT TGTCGAAGTG CTGGACTGCT 660

TGGAGAAGGT TTTGGTGCAA CACCTGAGAT GTGCAAGAAG AACCTAATAT GGGTTGTCAC 720

ACGGATGCAA GTTGTGGTTG ATCGCTATCC TACTTGGGGT GATGTTGTTC AAGTCGACAC 780

TTGGGTCAGT GCATCGGGGA AGAATGGCAT GCGAAGAGAT TGGCTTGTCA GCAATAGTGA 840

AACTGGTGAA ATTTTAACAC GAGCCACAAG TGTATGGGTG ATGATGAATA AACTGACTAG 900

AAGGTTATCT AAAATCCCAG AAGAGGTTCG AGGGGAAATA GAACCTTTTT TTATGAATTC 960

AGATCCTGTT CTGGCTGAGG ATAGCCAGAA ACTAGTGAAA CTCGATGACA GCACAGCTGA 1020

ACACGTGTGC AAAGGTTTAA CTCCTAAATG GAGCGACTTG GATGTCAACC AGCATGTCAA 1080

TAATGTGAAG TACATTGGCT GGATCCTTGA GAGTGCTCCA TTACCAATCT TGGAGAGTCA 1140

CGAGCTTTCC GCCTTGACTC TGGAATATAG GAGGGAGTGC GGGAGGGACA GCGTGCTGCA 1200

GTCACTGACC ACTGTGTCTG ATTCCAATAC GGAAAATGCA GTAAATGTTG GTGAATTTAA 1260

TTGCCAACAT TTGCTCCGAC TCGACGATGG AGCTGAGATT GTGAGAGGCA GGACCCGATG 1320

GAGGCCTAAA CATGCCAAAA GTTCCGCTAA CATGGATCAA ATTACCGCAA AAAGGGCATA 1380

GAAATCCAAG TAATCTCATT GCTGTGTGTA GTATCTATCG TGCTCTTTTC GGATTTATAT 1440

← *SstI*

ACATATATT**C CTTATGATTA TTAGTCTTCC TTTGAGGAGC TC**GAATTTCC CCGATCGTTC 1500

AAACATTTGG CAATAAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG CGATGATTAT 1560

CATATAATTT CTGTTGAATT ACGTTAAGCA TGTAATAATT AACATGTAAT GCATGACGTT 1620

ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCCGCAATTA TACATTTAAT ACGCGATAGA 1680

AAACAAAATA TAGCGCGCAA ACTAGGATAA ATTATCGCGC GCGGTGTCAT CTATGTTACT 1740

AGATC 1800

Plantlets developed from transgenic somatic embryos which have high levels of palmitic acid and very low levels of polyunsaturated fatty acids with less than 20% 18:1 and 18:2 (the major components of lipids in the thylakoid membranes) could not survive autotrophically after removal from sucrose-rich medium and placement into soil due to the less efficient photosynthetic machinery. The photosynthetic efficiency is greatly reduced (Somerville et al., 2000), due to leakiness of chloroplast membranes. Also, excess free palmitic acid, acting as a detergent to solubilize membranes in chloroplasts may have reduced photosynthetic efficiency. The leaves of the transgenic plants quickly lost their chlorophyll and senesced after the removal from the sucrose medium. Palmitic acid distributed to these embryos may be tightly regulated and hence, there is a direct correlation between the fatty acid compositions in cells and the normal development of plants. Normal plant development and growth possibly requires the right proportion of palmitic acid in these growing tissues (Huynh, 2001).

Construction of *Agrobacterium tumefaciens* binary plasmid vectors for seed-specific anti-sense RNA suppression and gene-transgene co-suppression of cotton palmitoyl-ACP thioesterase gene expression

Using a 2.6-kb *Hin*dIII fragment from the 5'-flanking region of the of cotton oleosin gene (Hughes et al., 1993) in pBluescriptM13 as a template, a 2.0-kb 5'-flanking region oleosin gene fragment was amplified by PCR (shown in Figure 36). The forward primer was 5'-GGAAGCTTGATGGATGATAATGAAGAAAGCG GC-3' with a *Hin*dIII site (underlined) and the reverse primer was 5'-GATCTAGAGCGAGAGGA GAGAAACTTTGACAG-3' with a *Xba*I site (underlined). This 2.0-kb *Hin*dIII/*Xba*I PCR fragment was then directionally subcloned into the *Hin*dIII/*Xba*I polylinker site of

the pGEM7Zf(+) plasmid vector (Promega). The resulting recombinant plasmid construct designated 5'- flanking oleosin/pGEM7Zf(+) was used for transforming electrocompetent *E. coli DH5α* cells. At least 10 colonies were formed in the selective agar plates. Plasmid DNAs were isolated from these transformants and digested with *Hin*dIII/*Xba*I restriction endonucleases to confirm their identity. Plasmid DNAs were isolated from these transformants and digested and fractionated on a 1% agarose gel. The DNA fragments from clone #4 and other clones shown in Figure 37 gave the expected 2.0-kb bands. Clone #4 was then chosen for a large scale plasmid preparation and further characterized by restriction digestion and agarose gel electrophoresis, as shown in Figure 38.

The 5'-flanking oleosin fragment was then isolated and purified from pGEM7Zf(+) in order to use for subcloning into the *Hin*dIII/*Xba*I sites of both anti-sense PATE/pBI121 and sense PATE/pBI121 constructs in place of the CaMV 35S promoter, as shown in Figures 22 and 23. The resulting recombinant plasmids were used for transforming electrocompetent *E. coli DH5α* cells. At least 3 colonies of each were formed in the selective agar plates. Plasmid DNAs were isolated from these transformants and digested with *Xba*I/*Sst*I and *Xba*I/*Hin*dIII restriction endonucleases. The digested plasmid DNAs fragments were then fractionated on a 1% agarose gel. Clone #S1 and clone #A2 (as shown in Figures 39 and 40) gave the expected 2.0-kb bands. These clones were then chosen for a large scale plasmid preparations, and further characterized by restriction digestion and agarose gel electrophoresis, as shown in Figure 41. The names were re-designated to sense 5'-flanking oleosin-PATE/ pBI121 and anti-sense 5'-flanking oleosin-PATE/ pBI121, respectively.

Both strands of the boundaries between pBI121 and insert and the 2.0-kb insert itself were sequenced using a primer-based approach. The nucleotide sequence of the boundaries between pBI121 and insert and the 2.0-kb inserts is shown in Figures 42 and 43. Analysis of the DNA sequence was done to compare between the 2.0-kb insert and the cotton oleosin gene and also between the boundaries and pBI121 vector (data not shown). The sequences upstream before the *Hin*dIII site and downstream after the *Sst*I site are the sequences of the pBI121 vector. The sequences between *Xba*I and *Sst*I sites are identical with the sequences of *PATE* cDNA. The sequences between *Hin*dIII and *Xba*I sites correspond with the 5'-flanking region of the cotton oleosin gene. The anti-sense and sense 5'-flanking oleosin-PATE/ pBI121 constructs were then used for transforming electrocompetent *Agrobacterium tumefaciens LBA4404* cells. Plasmid DNAs were isolated from these transformants and subjected to digestion with restriction enzymes and further characterized by agarose gel electrophoresis and alkaline blot hybridization (data not shown). The transformants indeed contained the seed-specific sense and anti-sense orientations of the *PATE* coding region constructs.

As previously discussed, the sense *PATE* and anti-sense *PATE* transgenic plants provided the correlation between palmitic acid levels and plant growth and development (Huynh, 2001). Overexpression of the *FatB* gene directly changed the fatty acid composition profile of transgenic plants in the way that higher amounts of saturated fatty acids (16:0 and 18:0) and lower amounts of mono (18:1) and polyunsaturated fatty acids (18:2) were produced. These changes directly affected the integrity of the chloroplast membranes and photosynthesis efficiency as previously discussed. Suppression of the *FatB* gene expression in anti-sense *PATE* transgenic plants did not deteted any changes in

the fatty acid composition profile of transgenic plants. It may be possible that a reduction in palmitic acid levels affected embryo development and limited leaf expansion. Thus, the dwarf transgenic plants had fewer viable seeds and smaller leaves (Huynh, 2001).

A future experiment is to generate transgenic cotton plants harboring the seed-specific sense and anti-sense *PATE* constructs. Dörmann et al. (2000) showed that higher amounts of palmitic acid accumulated in *Arabidopsis* seeds when the *Arabidopsis thaliana FatB1* thioesterase cDNA was overexpressed under the transcriptional control of the seed-specific napin promoter. These transgenic plants should provide knowledge on how the *FatB* gene is specifically expressed in seeds and influences fatty acid composition profile of seeds. The data presented in this chapter collectively indicates that recombinant plasmid constructs generated in this research altered *FatB* thioesterase expression, resulting in alteration of palmitic acid levles in some cotton tissues. Therefore, the cotton FatB thioesterase enzyme directly participates in regulating palmitic acid levels.

Figure 36. Agarose gel electrophoresis of the PCR products using the 2.6-kb *Hin*dIII fragment of the cotton oleosin gene as a template to generate the 2.0-kb 5'-flanking region fragment of the cotton oleosin gene (Hughes et al., 1993). The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

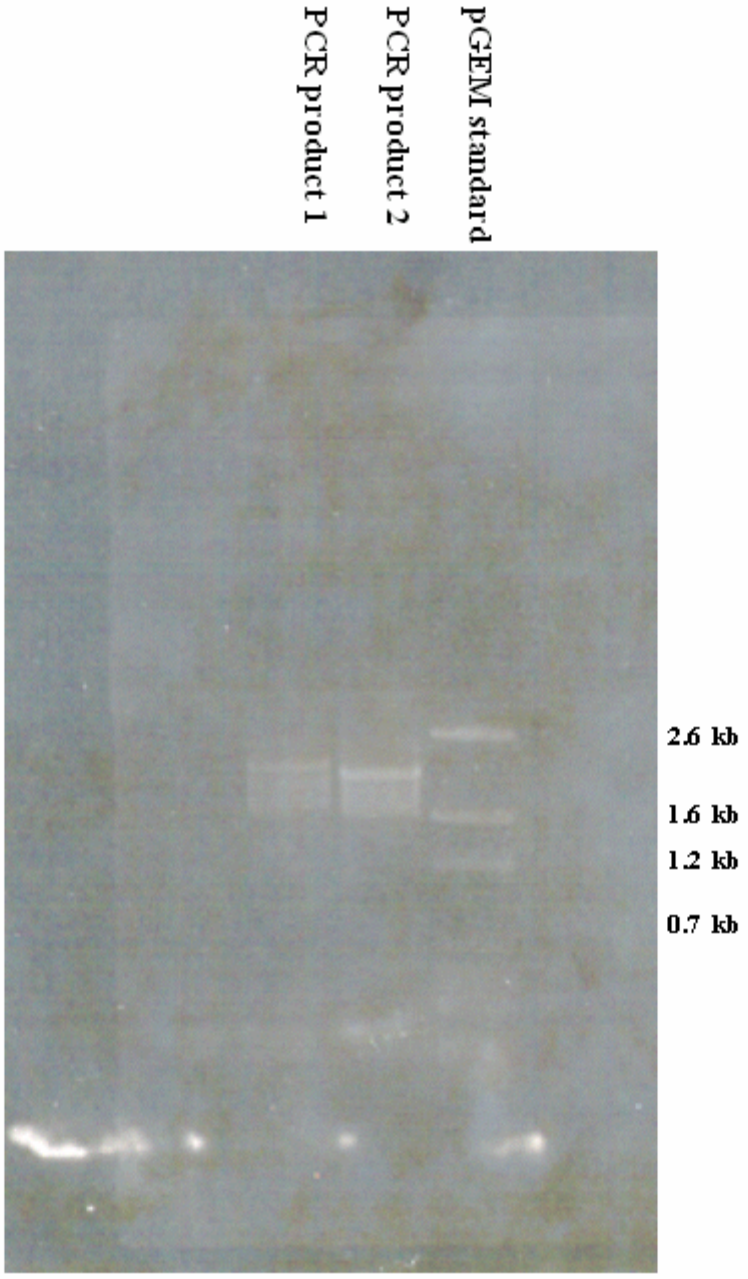

PCR product 1
PCR product 2
pGEM standard
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 37. Agarose gel electrophoresis of DNA fragment derived from the putative 5'-flanking oleosin/pGEM7Zf(+) constructs. Plasmid DNAs isolated from these clones were digested with the *Hin*dIII/*XbaI* restriction endonucleases and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

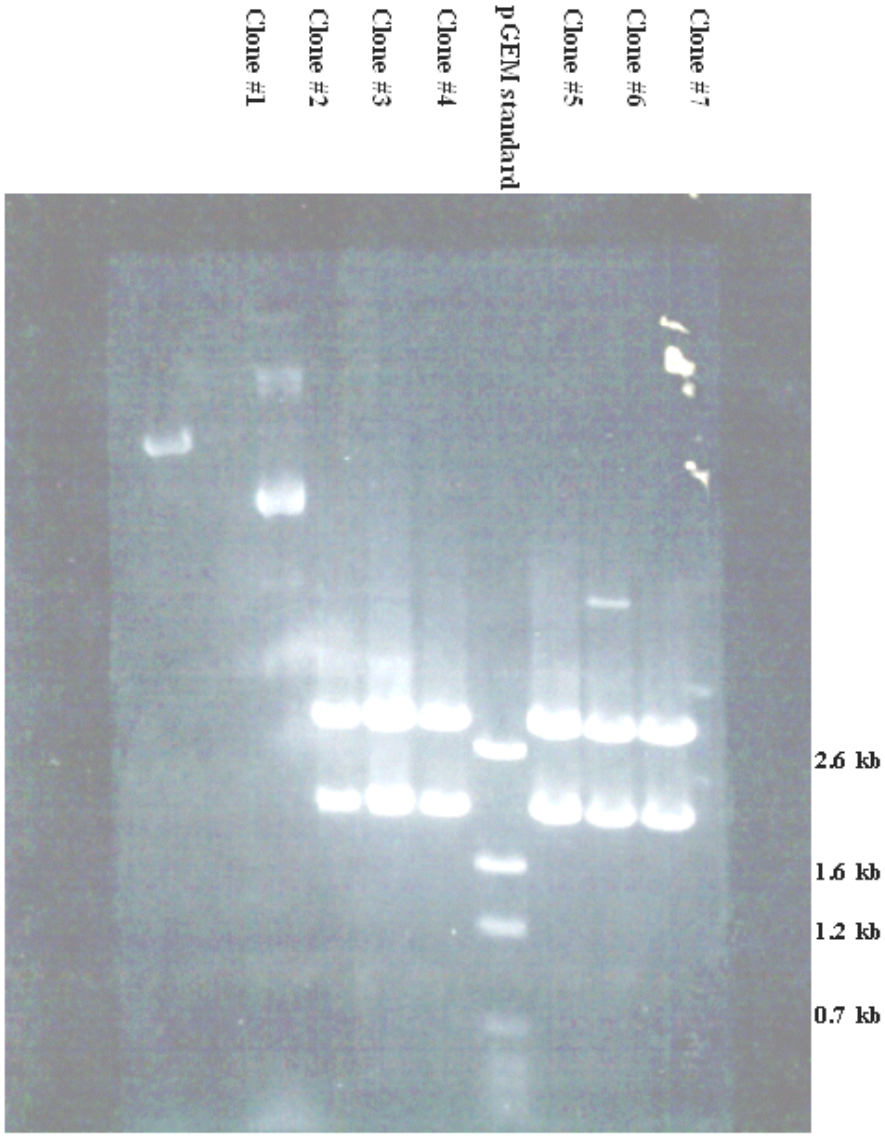

Clone #1
Clone #2
Clone #3
Clone #4
pGEM standard
Clone #5
Clone #6
Clone #7
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 38. Representative agarose gel of DNA fragments produced by restriction endonuclease digestion of Clone #4 (5'-flanking oleosin/pGEM7Zf(+)). The plasmid DNA isolated from this clone was digested with the restriction endonucleases as indicated. The resulting fragments were resolved on a 1% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

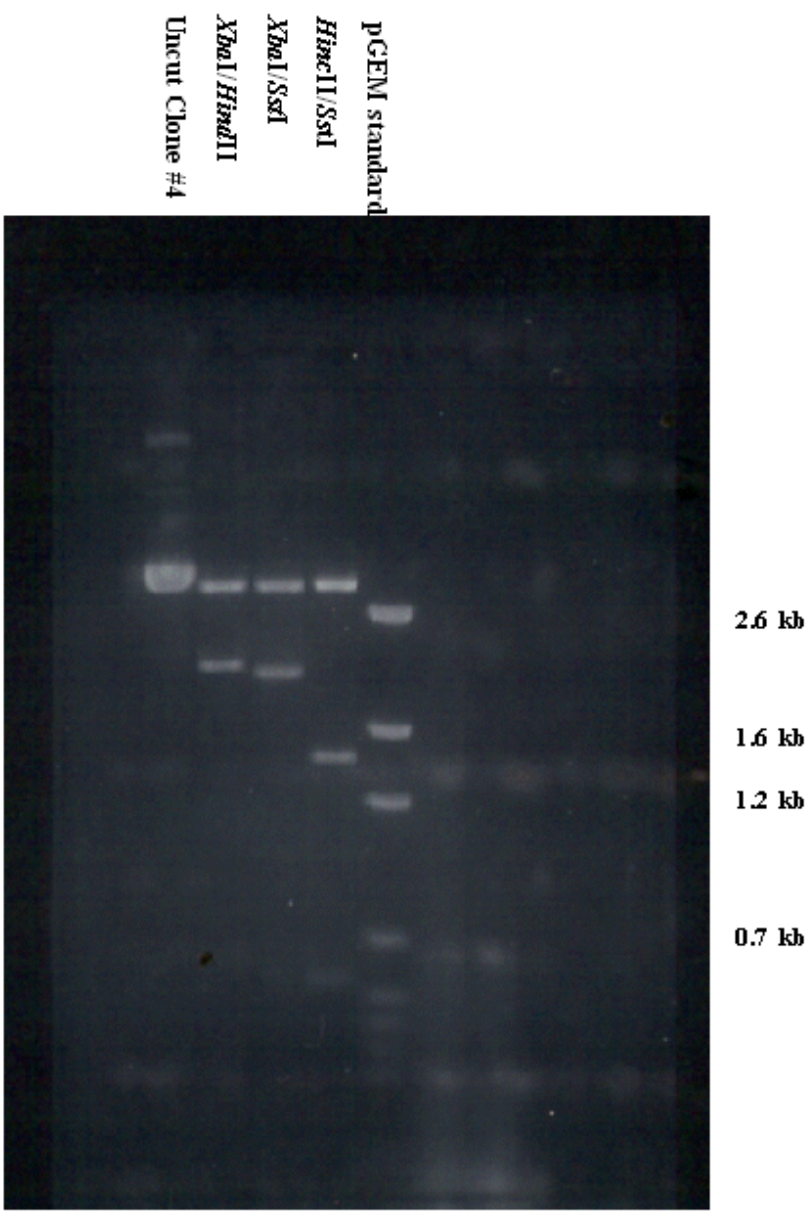

Uncut Clone #4
XbaI/HindIII
XbaI/SstI
HincII/SstI
pGEM standard
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 39. Agarose gel electrophoresis of DNA fragments from the putative sense 5'-flanking oleosin-PATE/ pBI121 constructs. Plasmid DNAs isolated from these clones were digested with the *Hin*dIII/*Xba*I and *Sst*I/*Xba*I restriction endonucleases and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

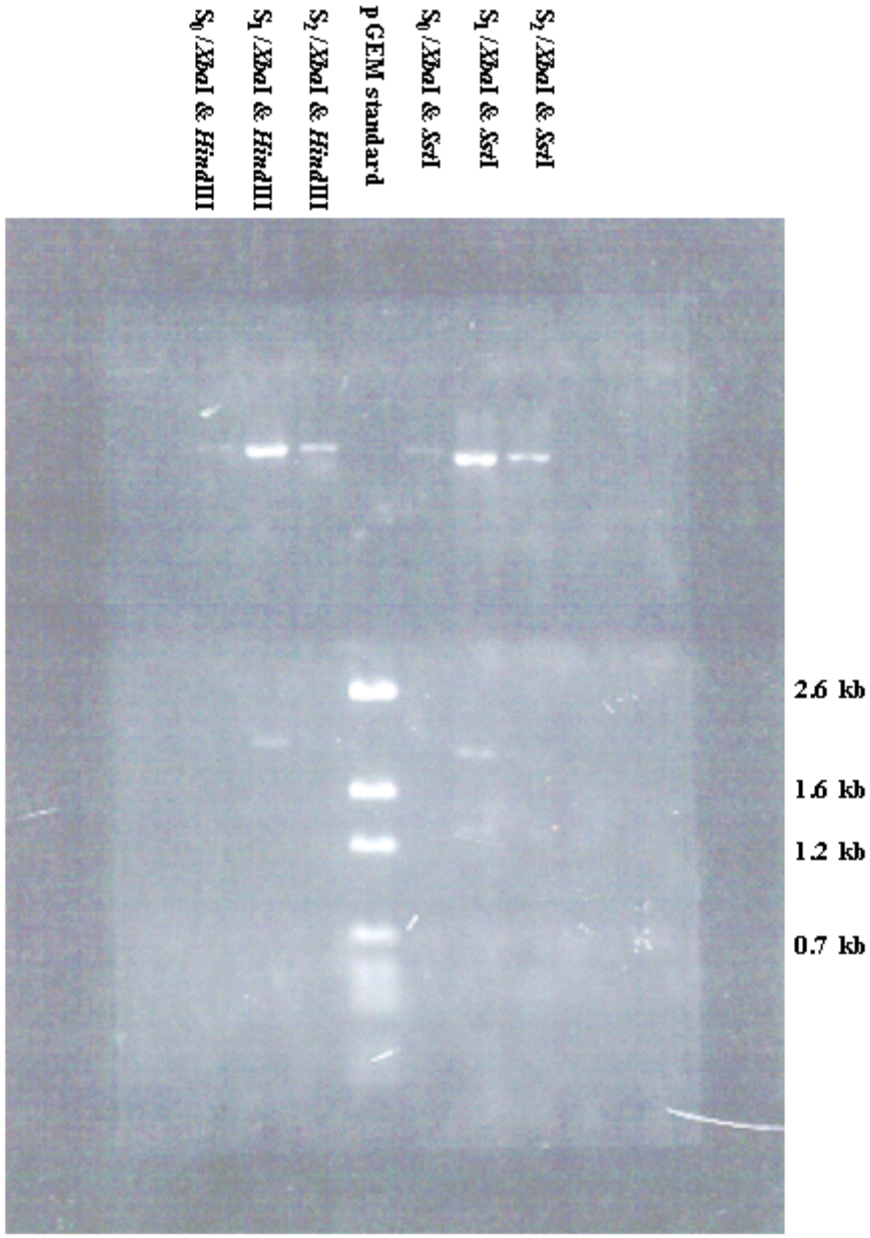

S0 /XbaI & HindIII
S1 /XbaI & HindIII
S2 /XbaI & HindIII
pGEM standard
S0 /XbaI & SstI
S1 /XbaI & SstI
S2 /XbaI & SstI
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 40. Agarose gel electrophoresis of DNA fragments from the putative anti-sense 5'-flanking oleosin-PATE/pBI121 constructs. Plasmid DNAs isolated from these clones were digested with the *Hin*dIII/*XbaI* and *Sst*I/*Xba*I restriction endonucleases and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

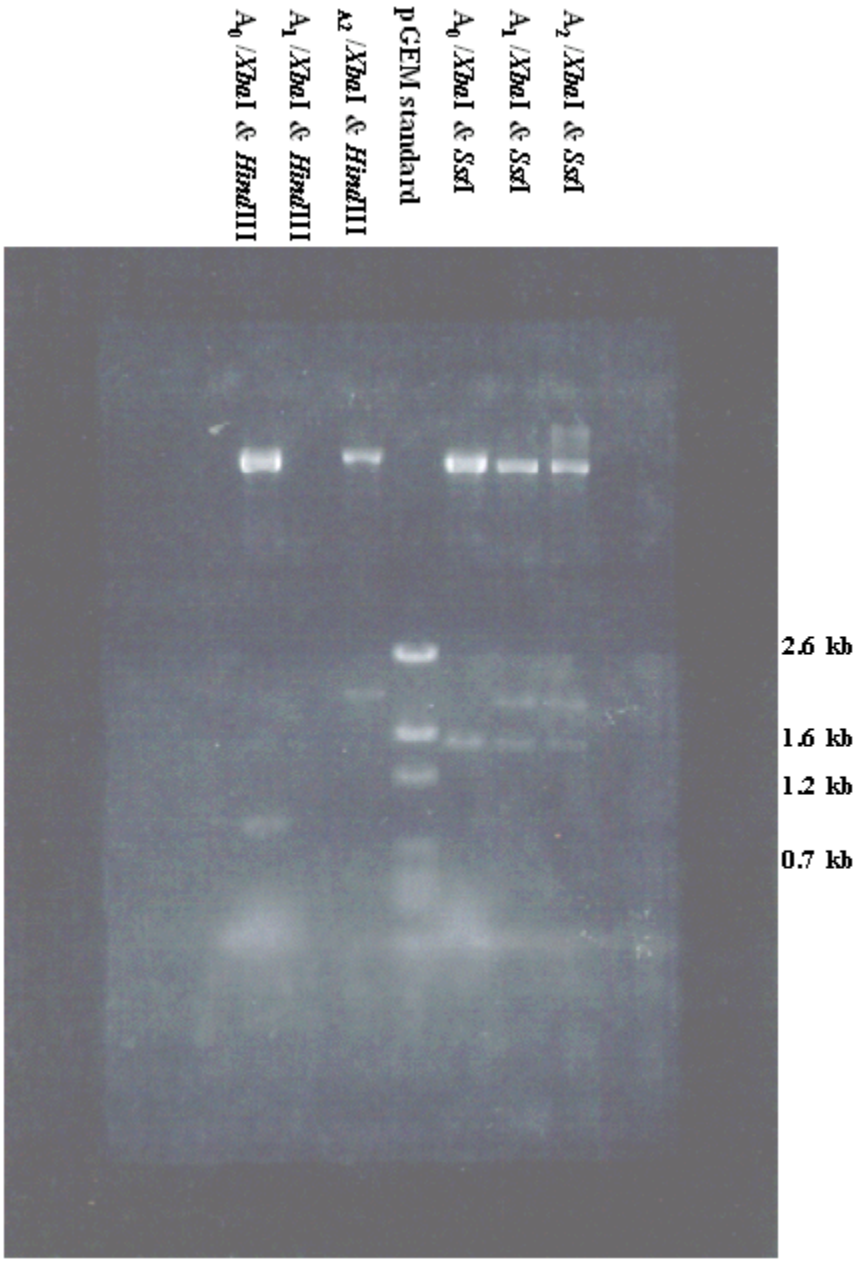

A0 /XbaI & HindIIII
A1 /XbaI & HindIIII
A2 /XbaI & HindIIII
pGEM standard
A0 /XbaI & SstI
A1 /XbaI & SstI
A2 /XbaI & SstI
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 41. Representative agarose gel of DNA fragments generated by restriction endonuclease digestion of clone #S1 and clone#A2. Plasmid DNA isolated from these clones were digested with the restriction endonucleases as indicated. The resulting fragments were resolved on a 1% agarose gel and stained with ethidium bromide prior to photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the left and lambda/*Hin*dIII standards are shown on the left.

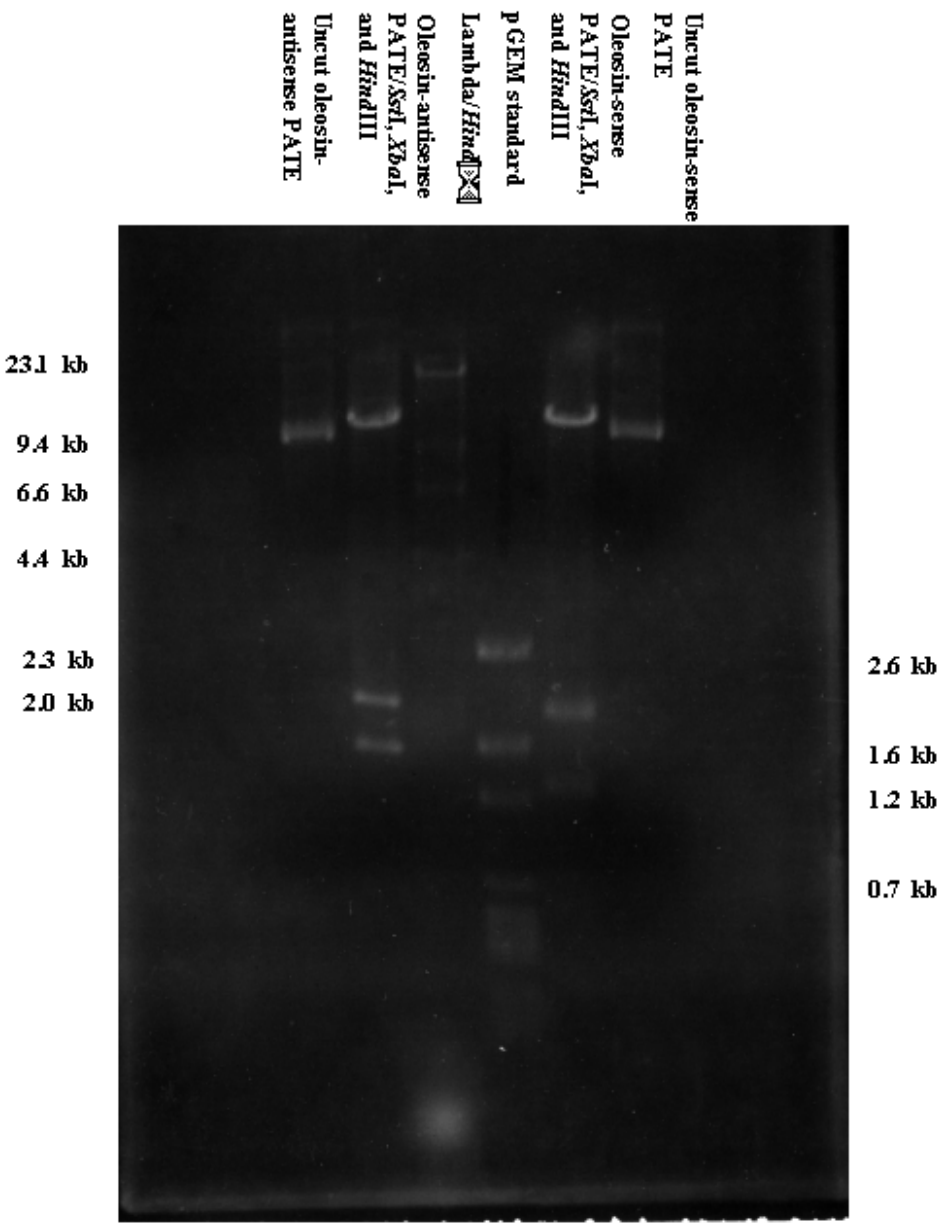

Uncut oleosin-sense PATE
Oleosin-sense PATE/SstI, XbaI, and HindIII
pGEM standard
Lambda/HindIII
Oleosin-antisense PATE/SstI, XbaI, and HindIII
Uncut oleosin-antisense PATE
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 42. The nucleotide sequence of the *Agrobacterium tumefaciens* binary plasmid vector anti-sense 5'-flanking oleosin-PATE/ pBI121 for seed-specific anti-sense RNA suppression of palmitoyl-ACP thioesterase gene expression. The numbers on the right refer to nucleotide residues. The sequences used for designing the forward and reverse primers for 5'-flanking oleosin and sense PATE as indicated by arrows and bold letters. The sequences upstream before the *Hin*dIII site and downstream after the *Sst*I site are the sequences of the pBI121 vector. The sequences between *Xba*I and *Sst*I sites are matched with the sequences of *PATE* cDNA (Pirtle et al., 1999). The sequences between *Hin*dIII and *Xba*I sites are matched with 5'flanking region of the cotton oleosin gene (Hughes et al., 1993).

AGGTTTCCCG ACTGGAAAGC GGGCAGTGAG CGCAACGCAA TTAATGTGAG TTAGCTCACT 60

CATTAGGCAC CCCAGGCTTT ACACTTTATG CTTCCGGCTC GTATGTTGTG TGGAATTGTG 120

***Hin*dIII**

AGCGGATAAC AATTTCACAC AGGAAACAGC TATGACCATG ATTACGCC**AA GCTTGATGGA** 180

→

TGATAATGAA GAAAGCGGCA GCTCAGATTC ACCAAATGAA CCATTACACC AAATCCAAGT 240

GAGCTCACCA TCACACCCAA TAACCACTAC GTTTTTGCTT AGTACACCAG AAGGGCATAA 300

GAAAAAGTTA TTAGCATCAA TTTCAGACAC CAATGAACCA AGCAGCTTTG AACAAACACC 360

TGAACCTGCA ATGAAAAGCC CATCTTCAGT TTTAAACCCA CCTTCTTCCT TGAAAAGAAA 420

AACCATTAAA GATTACTTTG TAGCAGCTCC ATGAAACTTT GGCGGTAAAT TTCTCATTGA 480

GGAAAAAGGC TTCGTATAAA AAATAGTAGT GTTGAGTAAA GTTCTTCAAA CTCCTCCTTT 540

CACTTTAAAC ACCAATTACT TGGTTAATGG GTATAGATTC TGGTTTTGTA AATTGGTTAG 600

CGAATTTAAA CCTAGCTACG ATTGTCTACG CCCTATGCCG GGGAATTTTA CTTAATTAAA 660

CTAAAGATAA TTTTTTATAT AAATTATTTA ATTTTTTAAT TATATAAATT GGATATGAAA 720

ATTTTTTAAT ATTAATTTTA TTATAAGTTT TAAAAAATTT AGACTCATTT ATTAATATTG 780

TAAGAATTTT TTTGTTGACG TAATATTTTG AAAAAAATAA AATAAATTTT GTATTCAAAT 840

AATTAAAAAA TAATATTGTA ATATATTTAA ATTTAATAAA GAATTTAGTG ATGTTAATTT 900

TTAAAAGTTT ACATAATTAT TTTAATTTGA GTAAAGTATT TAGTTTAATT AATGGTATTA 960

TAAAAATTGA GGTATTAAAT TATGTTAAAA TTTAAGTGAG ATATAGGAAT CAAATTGAAA 1020

TTTAACTATT CATATTGTTT AAAAAGTAGA GGGATTGGAA TTAATATTTA CCCATAATAT 1080

TCTCATGTAA AAGGTAAAAG GTAAAAGATG GACAGGTGTT CATTGTGGGC CGTCCTTAGT 1140

TGTATAGATT AAGAACACAT TTTTTATAAG GTTAAATTTT GCTATTTAGT AAGAGTAAAA 1200

AATTTAATTT TTGTATTTTA ATTTGATTAA TTTTAATTTT TTATATTTTT TTAATTTTAA 1260

AATTCTAGTC TTGATCTAAA TAAAAATAGT TAAATTCATT TGGTTAAGTT AAATTACTTG 1320

TCTTATAATA TATGTATAGT TGTAAATTTG ATTTATATAT TTCAATTGGA TCATTTTAAG 1380

TTTTTATACT TTTAAATTTT GAAATTTTGA TCTTAGTGCA AATGACAATT AATCTGTTAA 1440

TTAAATTTTT GGTGAGTAAT ATATGAAATA ACAAATTGAC ATAACATTAC ACATATGATA 1500

ATGTGTTTGT TGCATCAAAT TTTAAAAATA ACATGATTGA ACTTAATAAA TTTAACAGTT 1560

ATTGTTCAAT AAAATTAAAA TTTTATAATT TAAAATGTAC ATAAATTATA AATAACCAAA 1620

CTGTATAGTA TATGGACTGA ATCTATAATT TTCATAAACT ACAATACTAA TACCAAAAAA 1680

ATTAACCTTG TTTATAATTT GGAAAGGACA ACAGTGAGTA TTATTAAGAT TATTATTCAT 1740

TATTTCTAAT GAAGGAATTA AGTGTTAATT GGTGCAAGAG ATTAAGAAGA CCTGGAAAGG 1800

ACAACATAGA AATGACTTTT AGCAAAGACA TCATAAGGGG GCATATGGCC ACTTCAAAGT 1860

TCAAACCCAT ATGTCCCAAC ACAAAAACCA CAACGCAACA ATTTTGCACA ACCATGTCAA 1920

TTCCAGCAAT GTTTTCCCCA GATTTACTTT CCATATTTTT AATTTTAGAA ATTCAATTTT 1980

TATTTTTTAA AAATTGATAA AAAATACAAG AAATTATTAC ACATTTTAAA ACTTTCTTTA 2040

AATCTTGTCT GTTAAGATTC TTTTTTAATT TTTATTGCTA AGGAATCACA ATGAAACTTG 2100

TTCCCCATGC AAAGCCTCGA CTCACCATGT CATCCCCTAC GTGTCTTCCC ATTAAACCTC 2160

TTTATTCCTT TAAAAAACCC TCCCCTTTTC CCCCTATTCT CTCCATCTCT TTCTTTTACT 2220

← *Xba*I →

GTCAA**CTGTC AAAGTTTCTC TCTCTCTCGC TCTAGAAGAGC CATCTCACAT GGC**TATTCAA 2280

TTTTTCAAAA GCAGACCCAC CTTTCCATGG AAAGAGCCTA TGCTGGTTAG GGCACGGGCC 2340

CAACTTCACA AAATCCGTGT CCGAGGCTCG GGTCATACGA GCTGAGCCTT ACAAGGCTAT 2400

GTACGAAAAC CCGACTCCTA AACAAGCCTA ATTACAACCC CCTTTTTTTT TCTCAAAGGA 2460

AGACTAATAA TCATAAGGAA TATATGTATA TAAATCCGAA AAGAGCACGA TAGATACTAC 2520

ACACAGCAAT GAGATTACTT GGATTTCTAT GCCCTTTTTG CGGTAATTTG ATCCATGTTA 2580

GCGGAACTTT TGGCATGTTT AGGCCTCCAT CGGGTCCTGC CTCTCACAAT CTCAGCTCCA 2640

TCGTCGAGTC GGAGCAAATG TTGGCAATTA AATTCACCAA CATTTACTGC ATTTTCCGTA 2700

TTGGAATCAG ACACAGTGGT CAGTGACTGC AGCACGCTGT CCCTCCCGCA CTCCCTCCTA 2760

TATTCCAGAG TCAAGGCGGA AAGCTCGTGA CTCTCCAAGA TTGGTAATGG AGCACTCTCA 2820

AGGATCCAGC CAATGTACTT CACATTATTG ACATGCTGGT TGACATCCAA GTCGCTCCAT 2880

TTAGGAGTTA AACCTTTGCA CACGTGTTCA GCTGTGCTGT CATCGAGTTT CACTAGTTTC 2940

TGGCTATCCT CAGCCAGAAC AGGATCTGAA TTCATAAAAA AAGGTTCTAT TTCCCCTCGA 3000

ACCTCTTCTG GGATTTTAGA TAACCTTCTA GTCAGTTTAT TCATCATCAC CCATACACTT 3060

GTGGCTCGTG TTAAAATTTC ACCAGTTTCA CTATTGCTGA CAAGCCAATC TCTTCGCATG 3120

CCATTCTTCC CCGATGCACT GACCCAAGTG TCGACTTGAA CAACATCACC CCAAGTAGGA 3180

TAGCGATCAA CCACAACTTG CATCCGTGTG ACAACCCATA TTAGGTTCTT CTTGCACATC 3240

TCAGGTGTTG CACCAAAACC TTCTCCAAGC AGTCCAGCAC TTCGACAATG ATTTATAGCT 3300

GTTTCCTGTA AATGATTCAT TAGTGTCTCT ATGGATGCTG TTTGATCAGC GCCTATCTCA 3360

TATGATCTAA TCGAGAAGTT CTGACTGAAA ACAAGACCAT CCTGAACAAT CTTCCCTATG 3420

CCAAACGGAT CAATGACCAT GTCAGGCCGC CTCGGCTTCC AATCAAGCAT CATCCACTGC 3480

TTCTCAGCAG CCAAGAAAAT GGTTGTGATA GCAGCAAGAA GCATGCTCCA ATCAGGTAAC 3540

TGGTTGATAA ACGTCCTAGG AGGAGGGGAA CTTGCACCGT CATCGTTCTT GGAACCTTCC 3600

ACGGGAGTCG TCGACGCCAC AGTGCCGTTT ATTTTCGGAG GTGCTTGAGC ATTTGCCTTG 3660

ACTTGCAAAC TTCCAGAAGA AACCGATGGC TTCGACTTGA TGCTTCCGAG CTTCTTGTTT 3720

TTCGAGTCAG AGGAGTCAGG TGAAGAAGTG ACTGGGAAAA AC**GCCGATGT CACAGCAGTC** 3780

*Sst*I

T<u>**GAGCTC**</u>GAA TTTCCCCGAT CGTTCAAACA TTTGGCAATA AAGTTTCTTA AGATTGAATC 3840

CTGTTGCCGG TCTTGCGATG ATTATCATAT AATTTCTGTT GAATTACGTT AAGCATGTAA 3900

TAATTAACAT GTAATGCATG ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCCGC 3960

AATTATACAT TTAATACGCG ATAGAAAACA AAATATAGCG CGCAAACTAG GATAAATTAT 4020

CGCGCGCGGT GTCATCTATG TTACTAGATC 4080

Figure 43. The nucleotide sequence of the *Agrobacterium tumefaciens* binary plasmid vector sense 5'-flanking oleosin-PATE /pBI121 for seed-specific gene-transgene co-suppression of palmitoyl-ACP thioesterase. The numbers on the right refer to nucleotide residues. The sequences used for designing the forward and reverse primers for 5'-flanking oleosin and sense PATE are as indicated by arrows and bold letters. The sequences upstream before the *Hin*dIII site and downstream after the *Sst*I site are the sequences of the pBI121 vector. The sequences between *Xba*I and *Sst*I sites are matched with the sequences of *PATE* cDNA (Pirtle et al., 1999). The sequences between *Hin*dIII and *Xba*I sites are matched with 5'-flanking region of the cotton oleosin gene (Hughes et al., 1993)

AGGTTTCCCG ACTGGAAAGC GGGCAGTGAG CGCAACGCAA TTAATGTGAG TTAGCTCACT 60

CATTAGGCAC CCCAGGCTTT ACACTTTATG CTTCCGGCTC GTATGTTGTG TGGAATTGTG 120

***Hin*dIII**

AGCGGATAAC AATTTCACAC AGGAAACAGC TATGACCATG ATTACGCC**AA GCTTGATGGA** 180

→

TGATAATGAA GAAAGCGGCA GCTCAGATTC ACCAAATGAA CCATTACACC AAATCCAAGT 240

GAGCTCACCA TCACACCCAA TAACCACTAC GTTTTTGCTT AGTACACCAG AAGGGCATAA 300

GAAAAAGTTA TTAGCATCAA TTTCAGACAC CAATGAACCA AGCAGCTTTG AACAAACACC 360

TGAACCTGCA ATGAAAAGCC CATCTTCAGT TTTAAACCCA CCTTCTTCCT TGAAAAGAAA 420

AACCATTAAA GATTACTTTG TAGCAGCTCC ATGAAACTTT GGCGGTAAAT TTCTCATTGA 480

GGAAAAAGGC TTCGTATAAA AAATAGTAGT GTTGAGTAAA GTTCTTCAAA CTCCTCCTTT 540

CACTTTAAAC ACCAATTACT TGGTTAATGG GTATAGATTC TGGTTTTGTA AATTGGTTAG 600

CGAATTTAAA CCTAGCTACG ATTGTCTACG CCCTATGCCG GGGAATTTTA CTTAATTAAA 660

CTAAAGATAA TTTTTTATAT AAATTATTTA ATTTTTTAAT TATATAAATT GGATATGAAA 720

ATTTTTTAAT ATTAATTTTA TTATAAGTTT TAAAAAATTT AGACTCATTT ATTAATATTG 780

TAAGAATTTT TTTGTTGACG TAATATTTTG AAAAAAATAA AATAAATTTT GTATTCAAAT 840

AATTAAAAAA TAATATTGTA ATATATTTAA ATTTAATAAA GAATTTAGTG ATGTTAATTT 900

TTAAAAGTTT ACATAATTAT TTTAATTTGA GTAAAGTATT TAGTTTAATT AATGGTATTA 960

TAAAAATTGA GGTATTAAAT TATGTTAAAA TTTAAGTGAG ATATAGGAAT CAAATTGAAA 1020

TTTAACTATT CATATTGTTT AAAAAGTAGA GGGATTGGAA TTAATATTTA CCCATAATAT 1080

TCTCATGTAA AAGGTAAAAG GTAAAAGATG GACAGGTGTT CATTGTGGGC CGTCCTTAGT 1140

TGTATAGATT AAGAACACAT TTTTTATAAG GTTAAATTTT GCTATTTAGT AAGAGTAAAA 1200

AATTTAATTT TTGTATTTTA ATTTGATTAA TTTTAATTTT TTATATTTTT TTAATTTTAA 1260

AATTCTAGTC TTGATCTAAA TAAAAATAGT TAAATTCATT TGGTTAAGTT AAATTACTTG 1320

TCTTATAATA TATGTATAGT TGTAAATTTG ATTTATATAT TTCAATTGGA TCATTTTAAG 1380

TTTTTATACT TTTAAATTTT GAAATTTTGA TCTTAGTGCA AATGACAATT AATCTGTTAA 1440

TTAAATTTTT GGTGAGTAAT ATATGAAATA ACAAATTGAC ATAACATTAC ACATATGATA 1500

ATGTGTTTGT TGCATCAAAT TTTAAAAATA ACATGATTGA ACTTAATAAA TTTAACAGTT 1560

ATTGTTCAAT AAAATTAAAA TTTTATAATT TAAAATGTAC ATAAATTATA AATAACCAAA 1620

CTGTATAGTA TATGGACTGA ATCTATAATT TTCATAAACT ACAATACTAA TACCAAAAAA 1680

ATTAACCTTG TTTATAATTT GGAAAGGACA ACAGTGAGTA TTATTAAGAT TATTATTCAT 1740

TATTTCTAAT GAAGGAATTA AGTGTTAATT GGTGCAAGAG ATTAAGAAGA CCTGGAAAGG 1800

ACAACATAGA AATGACTTTT AGCAAAGACA TCATAAGGGG GCATATGGCC ACTTCAAAGT 1860

TCAAACCCAT ATGTCCCAAC ACAAAAACCA CAACGCAACA ATTTTGCACA ACCATGTCAA 1920

TTCCAGCAAT GTTTTCCCCA GATTTACTTT CCATATTTTT AATTTTAGAA ATTCAATTTT 1980

TATTTTTTAA AAATTGATAA AAAATACAAG AAATTATTAC ACATTTTAAA ACTTTCTTTA 2040

AATCTTGTCT GTTAAGATTC TTTTTTAATT TTTATTGCTA AGGAATCACA ATGAAACTTG 2100

TTCCCCATGC AAAGCCTCGA CTCACCATGT CATCCCCTAC GTGTCTTCCC ATTAAACCTC 2160

TTTATTCCTT TAAAAAACCC TCCCCTTTTC CCCCTATTCT CTCCATCTCT TTCTTTTACT 2220

← *XbaI* →

GTCAA**CTGTC AAAGTTTCTC TCTCTCTCGC TCTAGAATGG TTGCTACTGC TGTGACATCG** 2280

GCGTTTTTCC CAGTCACTTC TTCACCTGAC TCCTCTGACT CGAAAAACAA GAAGCTCGGA 2340

AGCATCAAGT CGAAGCCATC GGTTTCTTCT GGAAGTTTGC AAGTCAAGGC AAATGCTCAA 2400

GCACCTCCGA AAATAAACGG CACTGTGGCG TCGACGACTC CCGTGGAAGG TTCCAAGAAC 2460

GATGACGGTG CAAGTTCCCC TCCTCCTAGG ACGTTTATCA ACCAGTTACC TGATTGGAGC 2520

ATGCTTCTTG CTGCTATCAC AACCATTTTC TTGGCTGCTG AGAAGCAGTG GATGATGCTT 2580

GATTGGAAGC CGAGGCGGCC TGACATGGTC ATTGATCCGT TTGGCATAGG GAAGATTGTT 2640

CAGGATGGTC TTGTTTTCAG TCAGAACTTC TCGATTAGAT CATATGAGAT AGGCGCTGAT 2700

CAAACAGCAT CCATAGAGAC ACTAATGAAT CATTTACAGG AAACAGCTAT AAATCATTGT 2760

CGAAGTGCTG GACTGCTTGG AGAAGGTTTT GGTGCAACAC CTGAGATGTG CAAGAAGAAC 2820

CTAATATGGG TTGTCACACG GATGCAAGTT GTGGTTGATC GCTATCCTAC TTGGGGTGAT 2880

GTTGTTCAAG TCGACACTTG GGTCAGTGCA TCGGGGAAGA ATGGCATGCG AAGAGATTGG 2940

CTTGTCAGCA ATAGTGAAAC TGGTGAAATT TTAACACGAG CCACAAGTGT ATGGGTGATG 3000

ATGAATAAAC TGACTAGAAG GTTATCTAAA ATCCCAGAAG AGGTTCGAGG GGAAATAGAA 3060

CCTTTTTTTA TGAATTCAGA TCCTGTTCTG GCTGAGGATA GCCAGAAACT AGTGAAACTC 3120

GATGACAGCA CAGCTGAACA CGTGTGCAAA GGTTTAACTC CTAAATGGAG CGACTTGGAT 3180

GTCAACCAGC ATGTCAATAA TGTGAAGTAC ATTGGCTGGA TCCTTGAGAG TGCTCCATTA 3240

CCAATCTTGG AGAGTCACGA GCTTTCCGCC TTGACTCTGG AATATAGGAG GGAGTGCGGG 3300

AGGGACAGCG TGCTGCAGTC ACTGACCACT GTGTCTGATT CCAATACGGA AAATGCAGTA 3360

AATGTTGGTG AATTTAATTG CCAACATTTG CTCCGACTCG ACGATGGAGC TGAGATTGTG 3420

AGAGGCAGGA CCCGATGGAG GCCTAAACAT GCCAAAAGTT CCGCTAACAT GGATCAAATT 3480

ACCGCAAAAA GGGCATAGAA ATCCAAGTAA TCTCATTGCT GTGTGTAGTA TCTATCGTGC 3540

← ***Sst*I**

TCTTTTCGGA TTTATATACA TATATT**CCTT ATGATTATTA GTCTTCCTTT GAGAGCTC**GA 3600

ATTTCCCCGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT CCTGTTGCCG 3660

GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA ATAATTAACA 3720

TGTAATGCAT GACGTTATTT ATGAGATGGG TTTTTATGAT TAGAGTCCCG CAATTATACA 3780

TTTAATACGC GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG 3840

TGTCATCTAT GTTACTAGAT C 3900

CHAPTER 3

MOLECULAR CLONING AND ANALYSIS OF THE GENE FOR A COTTON Δ-12 FATTY ACID DESATURASE (FAD2-3) AND FUNCTIONAL EXPRESSION OF THE GENE

Introduction

Lipids in plants, as in all other organisms are the major structural components of biological membrane phospholipids and triacylglycerol storage oils which are essential for normal growth and development (Ohlrogge and Browse, 1995; Harwood, 1996; Ohlrogge and Jarworski, 1997; Somerville et al., 2000). Higher plants synthesize fatty acids de novo in the stroma compartment of plastids. A type II fully dissociable fatty acid synthase complex is utilized in this complex pathway (Ohlrogge et al., 1993; Ohlrogge and Browse, 1995; Ohlrogge and Jaworski, 1997; Harwood , 1996; Kinney, 1997). At least 30 enzymatic reactions are required to produce the common 16- or 18-carbon fatty acids with over 75% of fatty acid being unsaturated (Ohlrogge and Browse, 1995; Somerville et al., 2000) from acetyl-CoA and malonyl-CoA. Malonyl-CoA is a carbon donor for fatty acid biosynthesis and is produced from acetyl-CoA and carbon dioxide by acetyl CoA carboxylase in plastids. The malonyl group is then transferred from a malonyl-CoA to an acyl carrier protein (ACP) form malonyl-ACP.

Once malonyl-ACP is formed, it enters into a series of reactions including acylation, condensation and reduction reactions. In each round of the synthesis, a two-

carbon unit from malonyl-ACP is added to the growing acyl-ACP chain. Repeated cycles of the chain elongation of fatty acyl-ACP continue until the carbon length is 16 or 18, which is normal for a fatty acid found in phospholipid membranes and in triacylglycerols of seed oils. Hydrolysis of the acyl-ACP thioester bond by an acyl-ACP thioesterase enzyme terminates acyl chain elongation (Ohlrogge et al., 1993; Harwood, 1996; Somerville et al., 2000; Browse and Somerville, 1991). The introduction of the first double bond occurs while the fatty acid (18:0) is attached to the ACP by a soluble desaturase in plastids. Further desaturation of the fatty acid to introduce more double bonds occurs outside plastids by membrane-bound desaturases of either chloroplast or endoplasmic reticulum membranes (Ohlrogge and Browse, 1995). The possible role of polyunsaturated fatty acids in membrane structure and function in plants such as changes in the ultrastructure of the chloroplast membrane, the effect on thermal tolerance, the acting of signaling molecules, and the pathogen response have been studied and reported (Hamada et al., 1996b; Farmer, 1994; Browse and Somerville, 1991; Kirsch, et al., 1997).

Fatty acid desaturases are enzymes that catalyze the electron transfer from electron donors such as ferredoxin or cytochrome b_5 reductase to saturated fatty acids (the electron acceptors), and therefore form double bonds in fatty acids. There are three types of fatty acid desaturase enzymes depending upon the substrates: acyl-ACP, acyl-CoA and acyl-lipid desaturases (Shanklin and Cahoon, 1998; Los and Murata, 1998; Tocher et al., 1998). However, only two major types of desaturases are found in plants which are the soluble chloroplast Δ9-desaturases (an acyl-ACP desaturase), and the membrane-bound desaturases, (an acyl-lipid desaturase). The acyl-ACP desaturases, such as the chloroplast Δ9-desaturases found in the stroma of chloroplast are the

desaturases that introduce double bonds into fatty acyl-ACP (acyl carrier protein). The soluble chloroplast Δ9-desaturase, responsible for the conversion of saturated to unsaturated fatty acids introduces a cis-double bond at the Δ9 position of stearoyl-ACP (18:0) to produce oleoyl-ACP (18:1) in the stroma of chloroplasts. This soluble desaturase needs ferredoxin as the electron donor. The first Δ9 desaturase gene was isolated from safflower, and the polypeptide gene product is a soluble dimer of a 38-kDa subunit which contains a 33-amino acid transit peptide (Slabas and Fawcett, 1992).

The acyl-CoA desaturases and the acyl-lipid desaturases both are membrane-bound enzymes. The membrane-bound desaturases are integral membrane proteins found in the chloroplast and endoplsmic reticulum (ER) membranes. The acyl-CoA desaturases are found associated with the ER in animals, yeast and fungi (Los and Murata, 1998), while the acyl-lipid desaturases are found associated with the ER, chloroplast membrane, and cyanobacterial thylakoid membranes (Tocher et al., 1998; Los and Murata, 1998). Both acyl-CoA desaturases and acyl-lipid desaturases have quite similar structure, which are transmembrane proteins spanning the lipid bilayer membrane four times and consist of 300-500 amino acids (Fox et al., 1993; Shanklin et al., 1994; Murata and Wada, 1995). The acyl-CoA desaturases introduce double bonds into the fatty-acyl-CoA (Los and Murata, 1998). The acyl-lipid desaturases introduce double bonds into fatty acyl chain esterified in glycerolipids (Murata and Wada, 1995). Both desaturases act as the electron acceptors of the electron-transport system using the same cytochrome b_5 as the intermediates and the different electron donors. The acyl-CoA desaturses use the NADH-dependent cytochrome b_5 reductase as the eletron donors (Mitchell and Martin, 1995; Jeffcoat et al., 1977). The acyl-lipid desaturases use the NADH-dependent

cytochrome b_5 oxidoreductase as the electron donors (Kearns et al., 1991). The Δ5, Δ6, and Δ9 acyl-CoA desaturases are the examples of the acyl-CoA desaturses found in the ER of animal, yeast, and fungi (Mitchell and Martin, 1995; Jeffcoat et al., 1977). The fatty acid desaturases 4, 5, 6, and 7 (designated FAD4, FAD5, FAD6, and FAD7, respectively) are examples of the acyl-lipid desaturases found in chloroplast membranes, and the fatty acid desaturases 2 and 3 (designated as FAD2 and FAD3) are the examples of the acyl-lipid desaturases found in endoplasmic reticulum (Ohlrogge and Browse, 1995).

FAD2 and FAD3, the integral membrane proteins in the endoplasmic reticulum, primarily desaturase extrachloroplast oleate lipids (18:1) into linoleate (18:2) and linolenate (18:3) (Los and Murata, 1998; Shanklin and Cahoon, 1998). The enzymes are phosphatidylcholine desaturases, acting on fatty acids at both sn-1 and sn-2 positions phosphatidylcholine (Ohlrogge and Browse, 1995; Somerville et al., 2000; Harwood, 1996). The *FAD2* gene appears to be important in the chilling sensitivity of plants (Ohlrogge and Browse, 1995; Miquel and Browse, 1992; Okuley et al., 1994) because polyunsaturated membrane phospholipids are essential for maintaining cellular function and plant viability at lowered temperature (Browse and Xin, 2001). Studies with *Arabidopsis* mutants have shown that polyunsaturated lipids are significant components of membranes in order for membrane to function properly (Browse et al., 1986; Hugly and Somerville, 1992; Miquel et al., 1993; Wu et al., 1997). For example, *Arabidopsis fad2fad6* mutants, lacking the diunsaturated fatty acids $18:2^{\Delta 9,12}$ and $16:2^{\Delta 7,10}$ significantly lose photosynthetic efficiency (Somerville et al., 2000). *Arabidopsis* mutants lacking the triunsaturated fatty acids, $18:3^{\Delta 9,12,15}$ and $16:3^{\Delta 7,10,13}$, have diminished

the integrity of the thylakoid membranes which consist normal of about 70% polyunsaturated fatty acids (Routaboul et al., 2000). An *Arabidopsis fad3-2 fad7-2 fad8* mutant, lacking 16:3 and 18:3 fatty acids, loses chloroplast function at low temperatures (Routaboul et al., 2000). It has also been shown that the freezing response in plants depends upon lipid compositions in membranes (Havaux, 1987; Hugly and Somerville, 1992; Wu et al., 1997; Somerville et al., 2000).

Understanding the mechanisms underlying heat and cold tolerance, desiccation, salt tolerance and disease resistance in higher plants requires knowledge of the tissue-specific and development-specific regulation of fatty acid compositions of membranes. A major control point may be at the level of gene expression, such as how the genes expressing the enzymes of lipid biosynthesis are regulated in plants, and whether there are trans-acting transcription factors simultaneously coordinating the expression of many genes involved in lipid synthesis. A clearer understanding of the expression patterns of the *FAD2* and *FAD3* genes may ultimately permit the manipulation of the polyunsaturated fatty acid compositions of plant membranes predictably to improve the vigor and viability of crop plants. To date, the structure of only two plant *FAD2* genes have been determined, that of the single-copy *Arabidopsis FAD2* gene (Okuley et al., 1994) and the partial structure of a cotton *FAD2* gene designated the *FAD2-1* gene (Liu et al., 1997; Liu et al., 1999; Liu et al., 2001). The types of potential promoter and other regulatory elements in the 5'-flanking regions of *FAD2* genes have not been characterized, and the regulation of expression of these genes is not well understood.

Drs. Robert and Irma Pirtle and Ms. Wisatre Kongcharoensuntorn of our laboratory initially characterized the cotton genomic clone LCFg55 encompassing a

FAD2 gene (Pirtle et al., 2001). The coding region of this cotton FAD2 gene is 1,155 bp, and is continuous without introns, and would encode a polypeptide of 384 amino acids. This open reading frame has a 5'-flanking region of 568 bp and a 3'-flanking region of 242 bp. The deduced acid sequence of the cotton FAD2 polypeptide has significant identity with the amino acid sequence of the *Arabidopsis* FAD2 polypeptide (76%).

The *Arabidopsis FAD2* gene (Okuley ey al., 1994) has a 1.6-kb 5'-flanking region with the promoter region and a 1,134-bp intron, only four basepairs upstream from the ATG start codon. Thus, it was not surprising that the cotton gene might also have an intron in its 5'-flanking region. However, the 5'-flanking region of the *FAD2* gene in LCFg55 has only 568 basepairs upstream from the ATG start codon. Therefore, the 5' untranslated region of the *FAD2* gene was determined to contain only a portion of a potential intron in the 5'-flanking region by comparison with the 5'- flanking intron in the *Arabidopsis FAD2* gene. This cotton genomic clone LCFg55 also lacks the promoter/enhancer elements flanking the *FAD2* gene. Therefore, screening for an overlapping clone LCFg24 encompassing a potential full-length *FAD2* gene was necessary so that the presumptive upstream promoter/enhancer elements can be identified for analysis by mobility shift assays and DNA footprinting studies in the future.

A number of plant *FAD2* and *FAD3* cDNAs and genes that encode enzymes localized in the endoplasmic reticulum have been isolated and characterized, but the role and biochemical functions of the expressed polypeptide products is still largely uncharacterized. Yeast cells have been used successfully for functionally expression of several animal and plant fatty acid desaturases localized in the endoplasmic reticulum, such as the rat Δ-6 desaturase (Aki et al., 1999), the *C. elegans* Δ-5 desaturase (Watts and

Browse, 1999) and the *Arabidopsis* FAD2 enzyme (Covello and Reed, 1996; Kajiwara et al., 1996). The role of these enzymes can be studied in vivo in a yeast expression system because yeast cells produce only the monounsaturated fatty acid palmitoleic (16:1) and oleic (18:1) acids with the *OLE1* gene (Mitchell and Martin, 1995). The synthesis of the polyunsaturated fatty acid species linoleic (18:2) and linolenic (18:3) does not occur in yeast cells since they do not have the FAD2 and FAD3 enzymes found in plants (Tocher et al., 1995). The *Arabidopsis FAD2* gene was expressed in yeast cells using the *Gal1* promoter of *Saccharomyces cerevisiae* (Kajiwara et al., 1996; Covello and Reed, 1996). Yeast cells form the hexadecadienoyl (16:2) and linoleoyl (18:2) residues in their membrane phospholipids when the *Arabidopsis FAD2* coding region is induced by galactose.

In this chapter, the isolation and characterization of an overlapping cotton genomic clone encompassing the *FAD2* gene is described, because the cotton genomic fragment in LCFg55 lacks most of the 5'-flanking region of the *FAD2-3* gene, which is necessary for identifying the potential regulatory elements for expression of the gene. Several potential regulatory elements in its 5'-flanking region, particularly the large intron, have been identified that could be involved in tissue-specific and development specific regulation of the gene. In particular, our laboratory is interested in the regulatory mechanisms controlling the expression of the gene and its influence on the polyunsaturated fatty acid composition of membrane phospholipids for plant vigor and viability.

Materials and Methods

A cotton genomic library (constructed from *Gossypium hirsutum*, cv. Acala SJ2) harbored in the Lambda FIXII vector (Stratagene) was obtained from Dr. Thea Wilkins of the University of California at Davis. A partial cotton *FAD2-3* cDNA in the plasmid subclone pSKCF106A and *FAD2-3* gene in the plasmid subclone pCFg55 were obtained from Dr. Wisatre Kongcharoensuntorn, formerly of our laboratory. A cotton cDNA library in the vector pBluescript II SK (+) (Stratagene) was generously provided by Dr. Ed Cahoon of Dupont Ag Products, Experimental station, Wilmington, DE. The yeast/bacteria shuttle vector pYES2 and the plasmid pGEM7Zf(+) vector were purchased from Invitrogen (San Diego, CA) and Promega Corp. (Madison, WI), respectively. Restriction endonucleases, T4 DNA ligase and calf intestinal alkaline phosphatase were purchased from Invitrogen Life Technologies, Inc. Platinum *Taq* DNA polymerase and Platinum *Pfx* DNA polymerase were also purchased from Invitrogen Life Technologies, Inc. Various agarose powders (NuSieve, LE and Sea Plaque GTG) were purchased from FMC Corp. Synthetic oligonucleotides for PCR and sequencing primers were purchased from Biosynthesis, Inc. (Lewisville, TX). Nitrocellulose and Hybond N^+ nylon membranes for DNA transfer were purchased from Schleicher and Schuell (Keene, NH) and Amersham Life Sciences (Piscataway, NJ), respectively. Wizard *Plus* Miniprep DNA Purification kits were purchased from Promega Corp. QIAquick Gel Extraction and PCR Purification kits were purchased from QIAGEN, Inc (Valencia, CA). Sequenase Version 2.0 DNA Sequencing kits and ThermoSequenase Cycle Sequencing kits were purchased from Amersham Life Sciences. Radioactive materials, including [α-^{33}P]-ddNTPs and [α-^{32}P]-dCTP, were purchased from Amersham Life Sciences and NEN

Life Science Products, Inc. Various other reagents, chemicals and materials were purchased from Kodak, Fisher Scientific, Sigma Chemical Co., and Millipore.

Isolation and characterization of a cotton genomic clone overlapping the genomic clone LCFg55

The cotton genomic fragment in LCFg55 lacks most of the 5'-flanking region of the *FAD2-3* gene. Therefore it became necessary to isolate an overlapping cotton genomic clone. A unique, *FAD2-3* gene-specific 5'-flanking region fragment was generated by PCR amplification of nucleotides (nt) 2824-3213 of the partial 5'-flanking region of the *FAD2-3* coding region in LCFg55, using forward and reverse amplimers (5'CCGAGCTCACTAATC TTCAAGTGTATCC-3' and 5'-CGTCTAGATGGTT AAAGTGGTACTCGGG-3', with *Sst*I and *Xba*I sites (underlined), respectively. A solution containing 10 µl of 10X High fidelity PCR buffer (600 mM Tris-sulfate (pH. 8.9), 180 mM ammonium sulfate), 200 µM of deoxynucleotide triphosphates (dNTPs), 100 ng of the plasmid subclone pCFg55 encompassing the *FAD2-3* gene as template DNA, 1 µM of both primers, 2.5 units of Platinum Taq DNA polymerase, and 4 mM of $MgCl_2$ were added into a PCR tube and gently mixed by tapping the tube. Water was added to make a final reaction volume of 100 µl.

PCR amplification was conducted in a Perkin Elmer 2400 thermal cycler. The PCR conditions were as follows. The template was denatured at 94°C for 5 minutes. Subsequent amplification of a targeted region was achieved through 30 cycles at 94°C for 30 seconds, 53°C for 30 seconds, and 72°C for 30 seconds. Then, the final polymerization step or extension step was done at 72°C for seven minutes. The PCR

product was fractionated on a 1% agarose gel to check if there was nonspecific amplification occurring in the PCR. The 390-bp PCR fragment was purified from primers, free deoxynucleotides, polymerases, and salts using a QIAquick PCR purification kit (QIAGEN). The purified 390-bp fragment was then digested with *Xba*I and *Sst*I and purified by electrophoresis on a 1% agarose gel and extracted from the gel using a QIAquick Gel Extraction kit. The 390-bp *Xba*I/*Sst*I PCR fragment was then directionally subcloned into the *Xba*I and *Sst*I sites of the pGEM7Zf(+) vector and the plasmid construct designated FAD2pGEM. For ligation, the 390-bp fragment (2 μl of 100 ng/μl) and pGEM7Zf(+) DNA (3.4 μl of 30 ng/μl) were mixed in a 3:1 ratio (insert:vector). Ligase buffer (4 μl of 5X buffer (250 mM Tris-HCl (pH. 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethylene glycol-8000)), 1 μl of T_4 DNA ligase (0.1 unit/μl), and water in a final volume of 20 μl were added to the reaction mixture. The ligation mixture was then incubated at room temperature for one hour. The resulting recombinant FAD2pGEM plasmid was used for transforming electrocompetent *E. coli DH5α* cells as described in the previous chapters. The FAD2pGEM was then isolated from the transformants using a Wizard *Plus* Minipreps DNA Purification Kit. The FAD2pGEM was sequenced using a primer-based approach (Fan et al., 1996) with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, Tx) by Dr. Irma Pirtle of our laboratory. The nucleotide sequence was compared to the nucleotide sequence of pCFg55 DNA.

The 390-bp plasmid insert was used as the template to generate the hybridization probe by random priming (Feinberg and Vogelstein, 1983) to screen a second cotton

(*Gossypium hirsutum*, cv. Acala SJ2) genomic library harbored in the lambda FIXII vector (Stratagene), generously provided by Dr. Thea Wilkins of the University of California at Davis. This cotton genomic library was screened by the plaque hybridization method of Benton and Davis (1977). Conditions used to isolate the overlapping clone were as described in Chapter 1, except the stringency was increased by raising the temperature to 60°C. Two overlapping fragments from a genomic clone designated LCFg24, a 4.1-kb *Sal*I fragment and a 6.8-kb *Hin*dIII fragment, were subcloned into the corresponding sites of the pUC19 vector and designated pCFg24S and pCFg24H, respectively. The Epicentre EZ::TN Plasmid Based Deletion Machine System was used to generate nested random deletions of the LCFg24 subclones for sequencing. The screening and subcloning work was done by Drs. Robert and Irma Pirtle and Ms. Wisatre Kongcharoensuntorn of our laboratory. Both DNA strands of all subclones were sequenced using a combination of primer-based (Fan et al., 1996) and nested-deletion subcloning approaches (Wang et al., 1993; York et al., 1998). These sequencing efforts were done by Dr. Irma L. Pirtle of our laboratory and Dr. John E. Knesek of the Biology Department, Texas Woman's University (Pirtle et al., 2001, in press).

A cotton (*Gossypium hirsutum* cv. Deltapine 62) cDNA library, kindly provided by Dr. R. N. Trelease of Arizona State University harbored in the Stratagene Unizap lambda vector (Ni & Trelease, 1991) was screened and isolated by Drs. Robert and Irma Pirtle and Ms. Wisatre Kongcharoensuntorn using the *Arabidopsis* cDNA probe at 55°C. A partial length 791-bp *FAD2-3* cDNA clone designated pSKCF106A (GenBank AF 329635) was analyzed by DNA sequence analysis and found to lack the 5'-untranslated region (5'-UTR) and N-terminal coding region. Thus, the size of the 5'-UTR and

potential cap site of the *FAD2-3* gene could not be determined from the partial cDNA sequence.

The size of the 5'-UTR and potential cap site of the *FAD2-3* mRNA were deduced by comparison with the sequence of the 5'-end of a cotton *FAD2-3* cDNA fragment generated by PCR amplification of a *FAD2* cDNA sequence from a cotton cDNA library in the vector pBluescript II SK (+), generously provided by Dr. Ed Caoon of Dupont Ag Products Experimental Station, Wilmington, DE. For this PCR, the forward amplimer was 5'-CGTCATGTCTCTAGAACTAGT GGATCC-3' (complementary to the pBluescript II SK (+) multiple cloning site) and the reverse amplimer was 5'-GCAAGCTTTGTTTCC TTTGACTTGGAGG-3' (complementary to nt 3277-3297 in the 5'-region of the open reading frame of *FAD2-3* gene in LCFg 24 genomic clone). A solution containing 10 μl of 10X *Pfx* PCR buffer (Invitrogen), 200 μM of deoxynucleotide triphosphates (dNTPs), 100 ng of the total cotton cDNA (in the vector pBluescript II SK(+)) as template DNA, 1 μM of both primers, 2.5 units of Platinum *Pfx* Taq DNA polymerase, and 4 mM of $MgCl_2$ was added into a PCR tube and gently mixed. Water was added to make a final reaction volume of 100 μl. PCR conditions were as follows. The template was denatured at 94°C for 5 minutes. Subsequent amplification of a targeted region was achieved through a two-step annealing cycle. The first step was achieved through five cycles at 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 3 minutes. The second step was achieved through 30 cycles at 94°C for 30 seconds, 63°C for 30 seconds, and 72°C for 3 minutes. The final polymerization step or extension step was then done at 72°C for seven minutes.

The PCR product was fractionated on a 1% agarose gel to check if there was nonspecific amplification occurring in the PCR. The 216-bp amplified fragment was purified from primers, free deoxynucleotides, polymerases, and salts using a QIAquick PCR purification kit (QIAGEN). The purified 216-bp PCR fragment was then digested with *Xba*I and *Hin*dIII. The 216-bp *Xba*I/*Hin*dIII PCR fragment was purified by electrophoresis on a 1% agarose gel and extracted from the gel using a QIAquick Gel Extraction kit and then directionally subcloned into the *Xba*I and *Hin*dIII sites of the pGEM7Zf(+) vector and the plasmid construct designated 5'UTR-FAD2-3pGEM. The 216-bp *Xba*I/*Hin*dIII PCR fragment (10 µl of 3 ng/µl)) and pGEM7Zf(+) (1 µl of 100 ng/µl) DNA were mixed in a 3:1 ratio (insert:vector). The ligation mixture was then incubated at room temperature for one hour. The resulting recombinant 5'UTR-FAD2-3pGEM plasmid was used for transforming electrocompetent *E. coli DH5α* cells. The 5'UTR-FAD2-3pGEM was then isolated from the transformants using a Wizard *Plus* Minipreps DNA Purification kit and sequenced using a primer-based approach (Fan et al., 1996). The nucleotide sequence was then compared to the nucleotide sequence of pCFg24 DNA.

Expression of the cotton *FAD2-3* coding region in yeast cells

For functional expression of the *FAD2-3* open reading frame in yeast, the 1.2-kb open reading frame was amplified by PCR using Platinum *Taq* DNA polymerase (Invitrogen Gibco) and subcloned into the *Sst*I and *EcoR*I sites of the yeast-bacterial shuttle vector pYES2 (Invitrogen), which has the *GAL1* promoter for inducible expression of genes. The forward amplimer, 5'-GGGAGCTCATGGGTGCAGGTG GCAGAATGTCGG-3' was used to create a *Sst*I site (underlined) adjacent to the *FAD2-3*

initiation codon in the open reading frame. The reverse amplimer, 5'-GCGAATTC TTAGATCTTATTTCTAAACCAAAATACACC-3', was designed to provide an *Eco*RI (underlined) site in the 3'-flanking region. A solution containing 10 µl of 10X High fidelity PCR buffer (600 mM Tris-sulfate (pH. 8.9), 180 mM ammonium sulfate), 200 µM of deoxynucleotide triphosphates (dNTPs), 100 ng of pCFg55 template DNA, 1 µM of both primers, 2.5 units of Platinum Taq DNA polymerase, and 4 mM of $MgCl_2$ was added to a PCR tube and water was added to make a final reaction volume of 100 µl. The PCR conditions were as follows. The template was denatured at 94°C for 5 minutes. Subsequent amplification of a targeted region was achieved through 30 cycles at 94°C for 30 seconds, 53°C for 30 seconds, and 72°C for 30 seconds. Then, the final polymerization step or extension step was done at 72°C for seven minutes.

The PCR product was fractionated on a 1% agarose gel to check if there was nonspecific amplification occurring in the PCR. The 1.2-kb amplified fragment was purified from primers, free deoxynucleotides, polymerases, and salts using a QIAquick PCR purification kit (QIAGEN). Then the purified 1.2-kb PCR fragment was digested with *EcoR*I and *Sst*I and purified by electrophoresis on a 1% agarose gel, extracted from the gel slice using a QIAquick Gel Extraction kit, and then directionally subcloned into the *EcoR*I and *Sst*I sites of the pYES2 vector. The plasmid construct was designated pYES/FAD2-3. In the ligation reaction, the 1.2-kb *EcoR*I/*Sst*I (1.5 µl of 4 ng/µl) PCR fragment and pYES2 (2 µl of 50 ng/µl) DNA were mixed in a 3:1 ratio (insert:vector) for ligation with 0.1 unit of T_4 DNA ligase for 1 hour at room temperature. The pYES/FAD2-3 plasmid was then used for transforming electrocompetent *E. coli DH5α*

cells. The pYES/FAD2-3 was isolated from the transformants using a Wizard *Plus* Minipreps DNA Purification Kit as previously described. Both strands of this recombinant pYES/FAD2-3 plasmid were sequenced using a primer-based approach by Dr. Irma L. Pirtle of our laboratory to confirm that the coding region was truly identical to the *FAD2-3* gene in LCFg24 and that the construct was in frame relative to the yeast *GAL1* promoter.

The *S. cerevisiae* strain INVSc1 (Invitrogen) was transformed with both the pYES2/FAD2-3 DNA and the pYES2 vector DNA as a control by electroporation (Becker and Guarente, 1991) and the lithium lactate method (Gietz et al., 1992). Both yeast transformants were grown in SC-U medium (synthetic complete minus uracil) (Adams et al., 1998) at 30°C overnight, washed, and suspended in galactose induction medium (SC-U medium containing 2% galactose and 2% raffinose) and grown for three generations. The cells were pelleted and washes four times with water to remove any media or metabolites which could interfere with the lipid analyse. The yeast transformation and preparation was done by Ms. Wisatre Kongcharoensuntorn of our laboratory. The fatty acids were extracted and transmethylated with 5% HCl in methanol at 85°C for three hours (Christie, 1982). The fatty acid methyl esters were analyzed by gas chromatography and quantified by frame ionization detection (FID) (Chapman and Trelease, 1991) in comparison to an internal heptadecanoic acid (17:0) standard. The lipid analyses were done by Dr. Kent D. Chapman of our Biological Sciences Department.

Results

Isolation and characterization of a cotton genomic clone encompassing the *FAD2-3* gene overlapping the clone LCFg55

Using the plasmid pCFg55 encompassing the *FAD2-3* gene as a template with forward and reverse amplimers (5'CCGAGCTCACTAATCTTCAAGTGTATCC-3' and 5'-CGTCTAGATGGTTAAAGTGGTACTCGGG-3', with *Sst*I and *Xba*I sites (underlined), respectively), a 390-bp unique, *FAD2-3* gene-specific 5'-flanking region PCR fragment was amplified and shown in Figure 44. This 390-bp *Xba*I/*Sst*I PCR segment was directionally subcloned into the *Xba*I and *Sst*I sites of the pGEM7Zf(+) vector and the plasmid construct designated FAD2pGEM. The resulting recombinant FAD2pGEM plasmid DNA was used for transforming electrocompetent *E. coli DH5α* cells. The FAD2pGEM then was isolated from transformants using a Wizard *Plus* Minipreps DNA Purification Kit and digested with *Xba*I/*Sst*I enzymes. The digested plasmid DNAs were then fractionated on a 1% agarose gel, and Clone #1 (as shown in Figure 45) gave the expected 390-bp band. Both strands of the FAD2pGEM vector were sequenced by Dr. Irma Pirtle of our laboratory, and the DNA sequence is shown in Figure 46. The nucleotide sequence between the *Xba* I and *Sst* I sites is identical to that of the 5'-flanking region of the *FAD2-3* coding region in pCFg55 DNA. This 390-bp plasmid insert was used as template to generate hybridization probe in order to screen a second cotton (*Gossypium hirsutum*, cv. Acala SJ2) genomic library harbored in the lambda FIXII vector (Stratagene).

The cotton genomic DNA clone designated LCFg24 contains a 12.3-kb cotton genomic segment encompassing the *FAD2-3* gene (7,914 bp) including the coding

region, the 3'-flanking and the 5'-flanking regions containing the promoter elements and 5'-flanking intron (Pirtle et al., 2001, in press) The nucleotide sequence (GenBank AF331163) of a 7,914-bp segment of cotton DNA encompassing the FAD3-2 gene is depicted in Figure 47. The FAD2-3 open reading frame of 1,155 bp is continuous without introns, and encodes a presumptive polypeptide of 384 amino acid (Figure 47).

A partial-length 791-bp *FAD2-3* cDNA clone designated pSKCF106A (GenBank AF 329635) was isolated and analyzed by DNA sequence analysis. The *FAD2-3* gene in LCFg24 has a 3'-untranslated region (3'-UTR) of 242 bp deduced by comparison with 3'-flanking region of the corresponding partial-length cDNA clone pSKCF106A isolated by Drs. Robert and Irma Pirtle and Ms. Wisatre Kongcharoensuntorn. The near upstream polyadenylation signal (AATCAA, nt 4612-4617 in Figure 47) (Hunt and Messing, 1998) occurs 34 bp upstream from the 3'-polyadenylation site (nt 4,646 in Figure 47) in the putative *FAD2-3* mRNA transcript.

Unfortunately, this partial length FAD2-3 cDNA clone pSKCF106A lacks the 5'-untranslated region (5'-UTR) and N-terminal coding region. Therefore, the information is insufficient for comparison with the cotton *FAD2-3* gene to determine the size of the 5'-UTR and potential cap site of the *FAD2-3* mRNA. However, the size of the 5'-UTR and potential cap site of the cotton *FAD2-3* mRNA were deduced by comparison with the sequence of the 5'-end of a cotton *FAD2-3* cDNA fragment generated by PCR amplification of a cotton cDNA library in the vector pBluescript II SK (+). Using the cDNA isolated from a cotton cDNA library from Dr. Ed Cahoon of Dupont Ag Biotech, as a template with a forward amplimer 5'-CGTCATGTCTCTAGA ACTAGTGGATCC-3' (complementary to the pBluescript II SK (+) multiple cloning site) and the reverse

amplimer 5'-GCAAGCTTTGTTTCCTTTGACTT GGAGG-3' (complemenary to nt 3277-3297 in the 5'-region of the open reading frame of *FAD2-3* gene in LCFg 24 genomic clone), a 216-bp PCR fragment was amplified, as shown in Figure 48, and the 216-bp *Xba*I/*Hin*dIII PCR segment was directionally subcloned into the *Xba*I and *Hin*dIII sites of the pGEM7Zf(+) vector and the plasmid construct called 5'UTR-FAD2-3pGEM.

The resulting recombinant 5'UTR-FAD2-3pGEM plasmid was used for transforming electrocompetent *E. coli DH5*α cells. The 5'UTR-FAD2-3pGEM then was isolated from transformants using a Wizard *Plus* Minipreps DNA Purification Kit, digested with *Xba*I/*Hin*dIII enzymes, and the products fractionated on a 1% agarose gel. Clone #1A and clone#2B (as shown in Figure 49) both generated the expected 216-bp fragment. Both strands of the 5'UTR-FAD2-3pGEM. vector were sequenced by Dr. Irma Pirtle of our laboratory, and the DNA sequence is shown in Figure 50. The nucleotide sequence between the *Xba*I and *Hin*dIII is identical with the 5'-flanking region of the *FAD2-3* gene in the pCFg24 sequence, excluding the 5'-flanking intron. The locations of the putative cap site and the 5'- and 3'-intron splice junctions were deduced from a comparison of the gene sequence with that of a 216-bp PCR fragment derived from PCR amplification of the 5'-UTR of the FAD2-3 cDNA from (nt) 157 to (nt) 3,297, excluding intron.

A single, large intron is depicted in the 5'-UTR of the gene and is followed by a continuous open reading frame (as shown in Figure 47). The gene has 2,967-bp intron in its 5'-flanking region, only 12 bp upstream from the ATG initiation codon at nt 3,250. The *FAD2-3* gene in LCFg24 has the entire intron (2,967 bp) as compared to the *FAD2-3* gene in LCFg55, which has only 568 bp of the intron. Since the intron would be spliced

Figure 44. Agarose gel electrophoresis of the PCR products derived from amplification of the 5'-flanking region of the *FAD2-3* gene and the *FAD2-3* coding region using pCFg55 plasmid DNA as template to generate a 390-bp and 1.2-kb fragments, respectively. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments (Promega) are shown on the right.

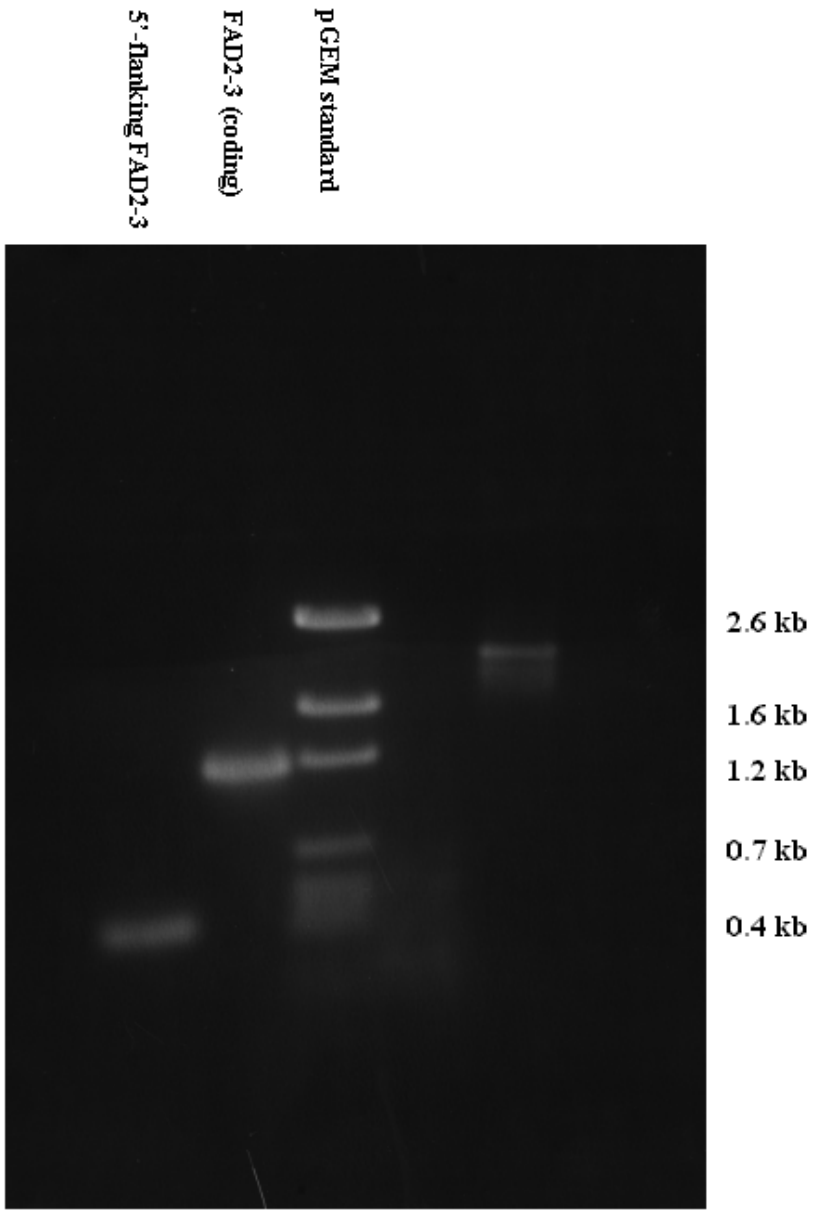

5'-flanking FAD2-3
FAD2-3 (coding)
pGEM standard
2.6 kb
1.6 kb
1.2 kb
0.7 kb
0.4 kb

Figure 45. Agarose gel electrophoresis of DNA fragments from restriction enzyme digestion of the putative FAD2pGEM clones and pYES/FAD2-3 clones. The plasmid DNAs isolated from these clones were digested with restriction endonucleases as indicated and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of λ/*Hin*dIII and pGEM standard DNA fragments are shown on the left and right, respectively**.**

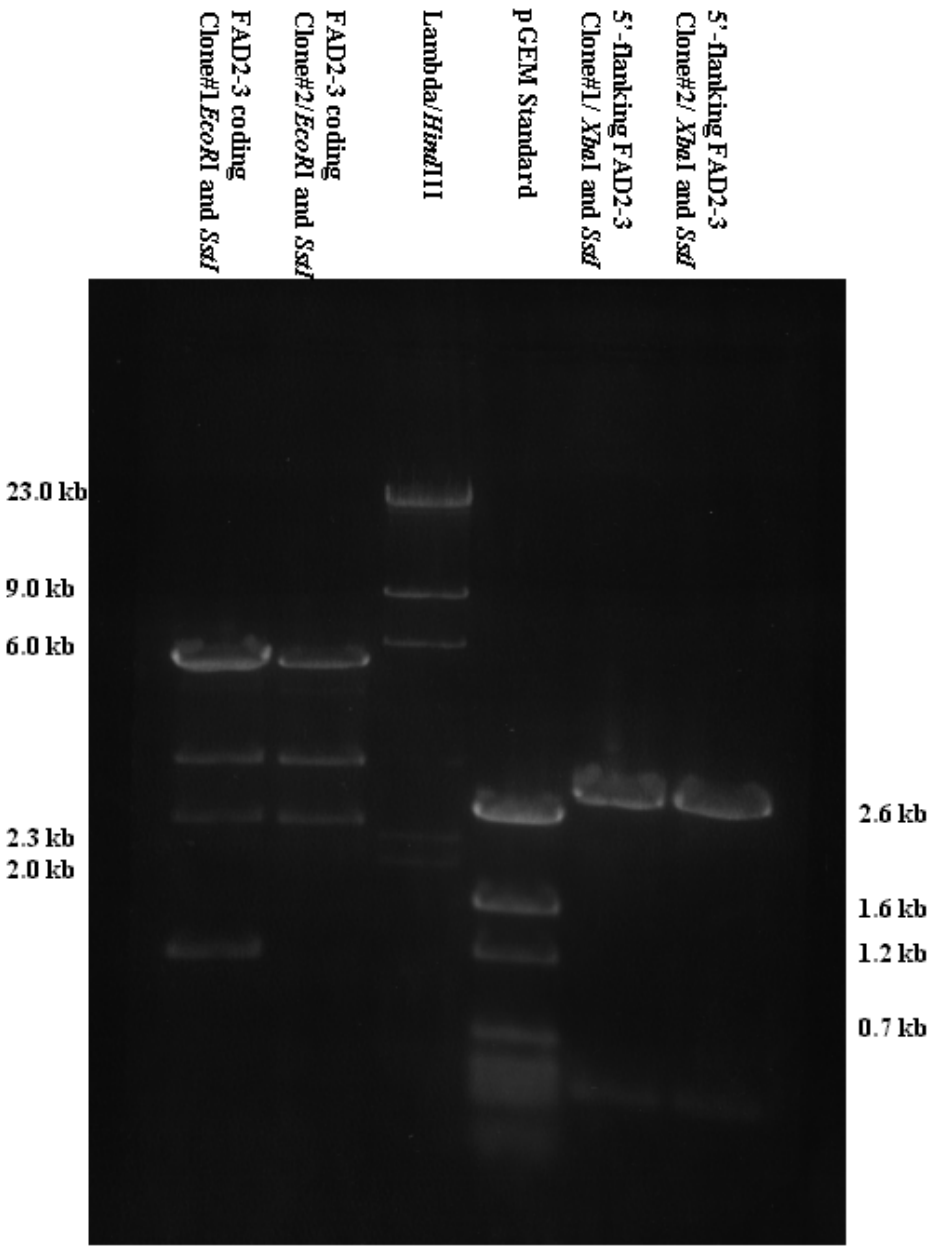

5'-flanking FAD2-3
Clone#2/ XbaI and SstI
5'-flanking FAD2-3
Clone#1/ XbaI and SstI
pGEM Standard
Lambda/HindIII
FAD2-3 coding
Clone#2/EcoRI and SstI
FAD2-3 coding
Clone#1EcoRI and SstI
23.0 kb
9.0 kb
6.0 kb
2.3 kb
2.0 kb
2.6 kb
1.6 kb
1.2 kb
0.7 kb

Figure 46. The nucleotide sequence of the 5'-flanking region of the PCR fragment derived from the cotton *FAD2-3* gene in the pCFg55 plasmid vector. The numbers on the right refer to nucleotide residues. The sequences used for designing the forward and reverse primers for the 5'-flanking of the cotton *FAD2-3* gene fragment are underlined. The sequences upstream before the *Sst*I site and downstream after the *XbaI* site are part of the sequences of the pGEM7Zf (+) vector. The sequences between *Xba*I and *Sst*I sites are matched with the 5'-flanking sequences of the *FAD2-3* gene in the pCFg55 plasmid vector shown in Figure 47.

*Sst*I

ATTTAGGTGACATATAGAATACTCAAGCTATGCATCCAACGCGTTGGGAGCTCACTAATC 60

TTCAAGTGTATCCTTATTGCTTTCATATTAGTTAGTTGAGATTGAATTATATAAACTTTC 120

AGATGTCAATTTGCAATTTTTACCAATATTTTACCCAAATAGTTTGTTTTTTTCAAAAAT 180

GACTGCTGTTGACCCCTTAGATGTAAATTAGCTGTCACCATATAGCATTGAGTTTATCTC 240

ATTTCAAATAGGAAGATGGGTCATTTGTGTTCATCTTATTGTATTTATATGTTACTTCCA 300

TTTGGACGTGAGTATTAAATATTATTATATATCTGACACGAATATAATTAAAATATTTCA 360

TATTTCTCGTATTATTTATAAAAAATCATATCTCTACATTCGTGTATGATAATGAGTATC 420

*Xba*I

GGACCCGAGTACCACTTTAACCATCTAGAGGAGCATGCGACGTCGGGCCCAATTCGCCCT 480

Figure 47. Nucleotide sequence of the noncoding (nontemplate) strand of a 7,914-bp cotton DNA segment in the lambda clones designated LCFg24 and LCFg55 encompassing the *FAD2-3* gene. The sequence has been assigned GenBank Accession Number AF331163. The numbering on the right refers to nucleotide (nt) residues, the numbering on the left refers to amino acid residues in the deduced sequence of the putative FAD2-3 polypeptide. The *FAD2-3* open reading frame is continuous and has 1,155 bp encoding 384 amino acids from nt 3250 to nt 4404, with the amino acid above the coding region in one letter abbreviations. The sequence of the 2,967-bp intron in the 5'-untranslated region (5'-UTR) is demarcated by brackets with 5'- and 3'- splice sites (GT…..AG) at nt 272 and 3238, respectively. The putative cap site of the mature mRNA is located at nt 153, with a 5'-UTR of the mRNA of about 130 nt. The location of the cap site and intron-exon junctions were deduced from the DNA sequence of a 174-bp PCR fragment derived from amplification of the 5'-UTR of a *FAD2-3* cDNA clone, corresponding to nt 157-3297, excluding the intron. The potential TATA box, a basic helix-loop-helix (bHLH) or E-box motifs, and a G-box element are shown underlined at nt 113, 44, and 17, respectively. Two potential GT-1 motifs are shown underlined at 32 and 65. Three putative *Dof* core elements occur at nt 92, 107, and 141. The 3'-polyadenylation site at nt 4,646 was deduced by comparison with the 3'-flanking region of a partial 791-bp *FAD2-3* cDNA clone designated pSKCF106A (GenBank AF329635), with a 3' untranslated region (3'-UTR) of 242 residues. The near-upstream polyadenylation signal is underlined at nt 4,612. The 3'-flanking region from nt 4,247 to 7,860 isomitted for brevity.

*Sal*I G-box GT-1 bHLH

GTCGACTCGATCACGGCACGTGGATGAGAGAGAAAATGAGAAACAAGTGGTGGAGTAAAA 60

GT-1 TATA Box

TGACGAAAATAGGTCCCTATTCCAAGGAGGGAAAGCTTAAAACAAAAAAGCTTAAATACA 120

CAP site

GGCGCCCCCCTTGAACACAGAAAGCACGGCCACCATAAAATAAGAAAATTAAGAGGCCGG 180

5'-UTR

ATTTCAAAACCCTTTCTCTTTAAAATATAGAGAAAAGAGGGACCAAAGTGAAAATCGAAA 240

TATAGATTTGATTTTCAATCTGCATTTTCAG|GT . . . . . .

. . . . . 2,967-bp INTRON, bp 272-3238 . . . . . .AG|GG 3240

M G A G G R M S V P P S Q R K Q E

TGTGGAACAATGGGTGCAGGTGGCAGAATGTCGGTTCCTCCAAGTCAAAGGAAACAAGAA 3300

18 S G S M K R V P I S K P P F T L S E I K

TCGGGCTCAATGAAAAGAGTCCCTATATCTAAACCACCATT TACTCTCAGTGAAATAAAA 3360

38 K A I P P H C F Q R S L I R S F S Y L V

AAAGCCATCCCACCACACTGTTTCCAACGCTCACT TATCCG TTCATTT TCCTATCTCGTT 3420

58 Y D F I L V S I F Y Y V A T T Y F H N L

TACGACTTCATT TTAGTCTCTATCTTTTACTACGTAGCCACCACTTACTTCCACAACCTC 3480

78 P Q P L S F V A W P I Y W T L Q G S V L

CCTCAGCCACTATCTTTCGTCGCCTGGCCAATTTATTGGACTCT TCAAGGTTCAGTCCTC 3540

98 T G V W V I A H E C G H H A F S D Y Q W

ACTGGCGTTTGGGTTATCGCCCATGAATGCGGTCACCATGCTTTTAGCGATTACCAATGG 3600

118 I D D T V G L I L H S S L L V P Y F S W

ATTGATGACACTGTCGGTCTCATCCTCCATTCATCCCTTCTTGTCCCGTACTT TTCGTGG 3660

138 K Y S H R R H H S N T G S L E R D E V F

AAATATAGTCACCGACGTCACCATTCCAACACTGGT TCCCTTGAACGCGACGAAGTATTT 3720

158 V P K K R S S I R W W A K Y L N N P P G

GTTCCGAAGAAACGGAGCAGCATTAGATGGTGGGCTAAATACCTCAACAATCCACCAGGT 3780

178 R F V T V T I Q L T L G W P L Y L A F N

CGTTTCGTCACAGTCACCATTCAGCTCACTCTCGGATGGCCTCTTTACTTAGCATTCAAT 3840

198 V A G R P Y E G L A C H Y N P Y G P I Y

GTAGCAGGTAGACCTTACGAAGGACTCGCTTGTCACTACAACCCATACGGTCCTATCTAC 3900

218 N D R E R L Q I Y I S D V G V L A V T Y

AACGACCGTGAACGACTTCAAATCTACATATCCGACGTCGGTGTCCTTGCTGTCACCTAT 3960

238 G L Y R L V L A K G L A W V I C V Y G V

GGGCTGTACCGTCTCGTGTTAGCCAAAGGTCTAGCTTGGGTCATTTGCGTTTACGGTGTC 4020

258 P L L I V N A F L V M I T Y L Q H T H P

CCATTGCTCATCGTTAATGCATTCCTCGTCATGATCACATACTTGCAACACACTCACCCC 4080

278 A L P H Y D S S E W D W L R G A L A T V

GCATTACCACACTACGACTCATCCGAATGGGACTGGTTACGTGGAGCCCTCGCGACGGTC 4140

298 D R D Y G I L N K V F H N I T D T H V A

GACCGAGATTATGGGATATTAAACAAGGTTTTCCATAACATAACTGATACTCATGTCGCT 4200

318 H H L F S T M P H Y H A M E A T K A I K

CATCATTTGTTTTCGACGATGCCGCATTACCACGCAATGGAAGCAACTAAGGCAATAAAA 4260

338 P I L G E Y Y S F D G T P V Y K A I F R

CCAATATTGGGAGAGTATTATTCATTTGATGGTACACCAGTTTATAAAGCGATATTTAGA 4320

358 E A K E C I Y V E P D E G E Q S S K G V

GAGGCAAAGGAGTGTATTTACGTTGAACCAGACGAAGGTGAGCAGAGCAGCAAAGGTGTA 4380

378 F W F R N K I TER

TTTTGGTTTAGAAATAAGATCTAACTTTGCCGATAGCGTTGCGGTTGCCGATGGTGATGC 4440

CTTTAGGAATGTGTTAAATTTGT TACATTAT TGTTAAGGATTTGGGGTTTTGGCGTTTGG 4500

3'-UTR

GTTACATCAATTTCAGATGCTTTCGAATTTGGACTTTGTATGGT TCTCATCGACTTTGTT 4560

Poly A Signal

GATCCCTGCAAAATTGGT TCGAGCTTTCAACTATCAAGTAGTTTTTTTTTTAATCAAATT 4620

Poly A site

TATTATTGGTGCCGAGT TATAAAAAA 4680

. . . . . .

GACAACCCGGTCTTGGAAACAAGGATCAGGGGATTCAAGCAAGGCTTGAAGCTT 7914

*Hind*III

Figure 48. Agarose gel electrophoresis of the PCR product derived from amplification of the *FAD2-3* cDNA in the cotton cDNA library as a template to generate the 216-bp fragment.. The PCR products were then doubly digested with *Xba*I and *Hind*III. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the right.

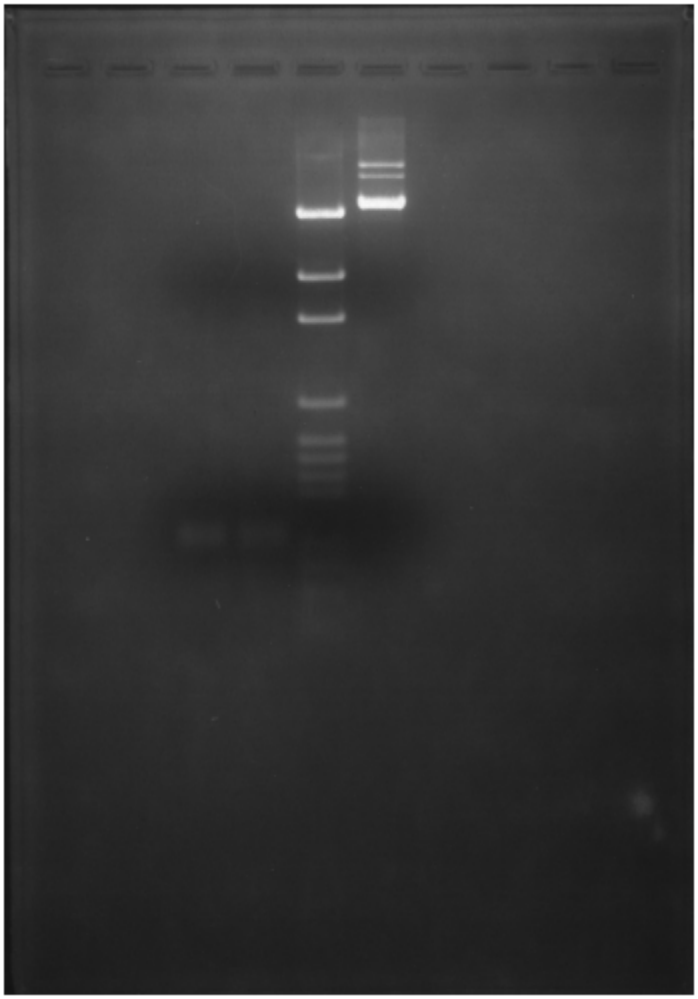

pGEM/XbaI & HindIII
pGEM standard
Library2/XbaI & HindIII
Library1/XbaI & HindIII
2.6 kb
1.6 kb
1.2 kb
0.7 kb
0.35 kb
0.2 kb

Figure 49. Agarose gel electrophoresis of the DNA fragments derived from prospective 5'-UTR-FAD2-3pGEM clones. The plasmid DNAs isolated from these clones were digested with *Xba*I and *Hin*dIII and fractionated on a 1% agarose gel. The DNA fragments were stained with ethidium bromide before photography. The sizes (in kb) of pGEM standard DNA fragments are shown on the righ**.**

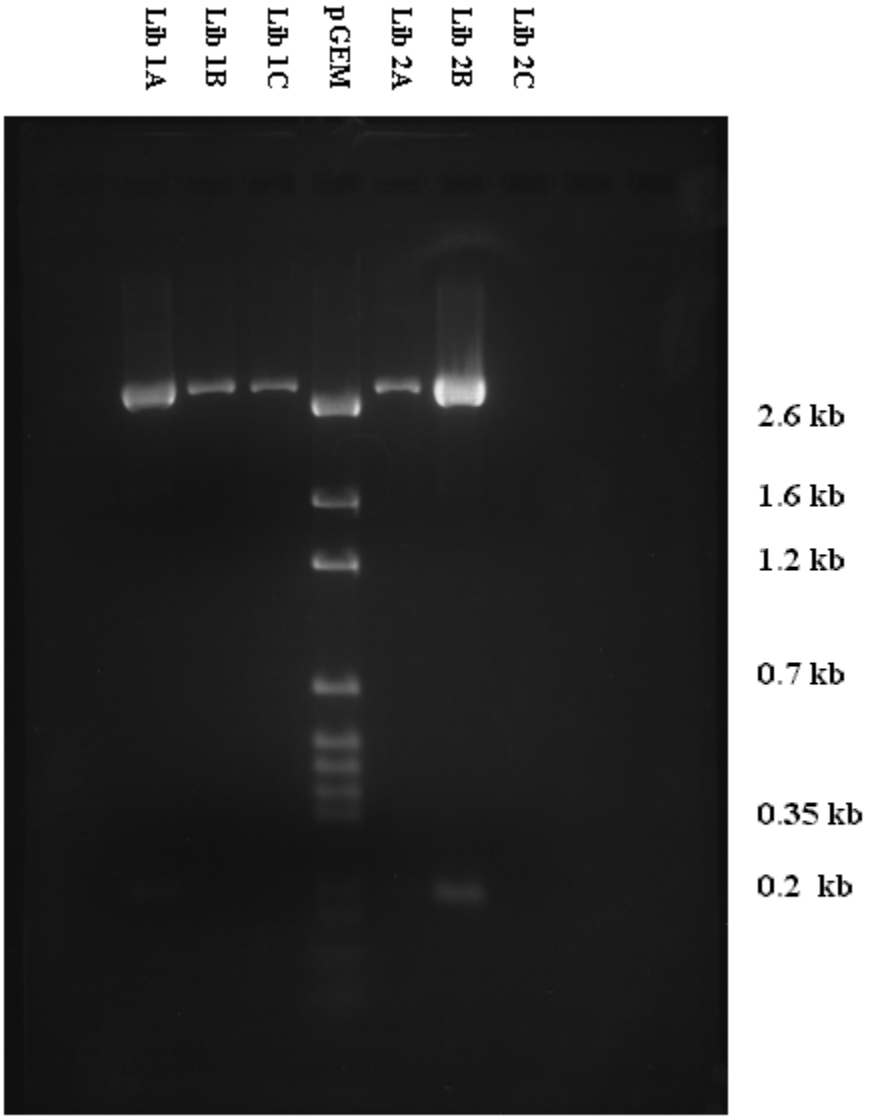

Lib 1A
Lib 1B
Lib 1C
pGEM
Lib 2A
Lib 2B
Lib 2C
2.6 kb
1.6 kb
1.2 kb
0.7 kb
0.35 kb
0.2 kb

Figure 50. The nucleotide sequence of the 5'-flanking region of a cotton *FAD2-3* cDNA fragment generated by PCR amplification of a cotton cDNA library in the vector pBluescript II SK (+) provided by Dr. Ed Cahoon of Dupont Ag Products, Experimental Station, Wilmington, DE. The numbers on the right refers to nucleotide residues. The sequences used for designing the forward and reverse primers for the 5'-end of a cotton *FAD2-3* cDNA fragment are underlined. The sequences upstream before the *Xba*I site and downstream after the *Hin*dIII site are part of the sequences of the pBluescript II SK (+) vector. The sequences between *Xba*I and *Hin*dIII sites are identical with the 5'-flanking sequence of the *FAD2-3* gene shown in Figure 47.

TTAAGTTGGTAACACGGTTTCCCAGTCACGCTTGTNAAAAGCACGGCCAGTGAATTGTAA 60

*Xba*I

TACGACTCACTATAGGGCGAATTGGGCCCGACGTCGCAGTCCTCTAGAACTAGTGGATCC 120

CCCGGGCTGCAGGAATTCGGCACGAGGAAAATAAGAAAATTAAGAGGCCGGATTTCAAAA 180

CCCTTTCTCTTTAAAATATAGAGAAAAGAGGGACCAAAGTGAAAGAAAATCGAAGTATAG 240

ATTTGATTTTCAATCTGCATTTCAGGGTGTGGAACAATGGGTGCAGGTGGCAGAATGTC 300

*Hind*III

GGTTCCTCCAAGTCAAAGGAAACAAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCC 360

CGGTACCCAATTCGCCCTATAGTGAGTCGTATTA

out to generate the mature *FAD2-3* mRNA, the 5'-UTR of the mRNA is about 130 nucleotides in length. The 5'-flanking region of the *FAD2-3* gene has several potential promoter elements (indicated in Figure 47) that could function as positive regulatory elements in gene expression. A TATA-like basal promoter element occurs 40 bp from the putative cap site. A basic region helix-loop-helix (bHLH) or E motif with the consensus sequence CANNTG occurs at 109 bp upstream from the potential cap site. There are three possible *Dof* core recognition sequences at 12 bp, 46 bp, and 61 bp upstream from the the tentative cap site. Several potential light-responsive promoter elements occur in the 5'-flanking region. A G-box (CACGTG) (Terzaghi et al., 1995; Guilfoyle, 1997) occur 136 bp upstream from the tentative cap site. Two consensus GT-1 motifs (GRWAAW), general features of light-responsive promoters (Terzaghi et al., 1995; Guilfoyle, 1997), occur 88 bp and 121bp prior to the putative cap site.

Expression of the cotton *FAD2* coding region in yeast cells

Using the *FAD2-3* gene in the subclone pCFg55 as a template with the forward amplimer, 5'-GGGAGCTCATGGGTGCAGGTGGCAGAATGTCGG-3' and the reverse amplimer, 5'-GCGAATTCTTAGATCTTATTTCT AAACCAAAATACACC-3', a 1.2-kb *FAD2-3* coding region fragment was amplified by PCR, as shown in Figure 44. The 1.2-kb *EcoR*I/*Sst*I PCR segment was directionally subcloned into the *EcoR*I and *Sst*I sites of the pYES2 vector and designated pYES/FAD2-3. The resulting recombinant pYES/FAD2-3 plasmid was used for transforming electrocompetent *E. coli DH5α* cells. The pYES/FAD2-3 then was isolated from transformants using a Wizard *Plus* Minipreps DNA Purification Kit and digested with *EcoR*I/*Sst*I enzymes. The digested plasmid DNAs were then fractionated on a 1% agarose gel. Clone # 1 (as shown in Figure 45)

gave a 1.2-kb band. Both strands of the pYES/FAD2-3 vector were sequenced using a primer-based approach with synthetic oligonucleotide primers from Biosynthesis, Inc. (Lewisville, Tx) by Dr. Irma Pirtle of our laboratory and the DNA sequence was shown in Figure 51. The nucleotide sequence was matched with the *FAD2-3* coding sequence of pCFg24 DNA sequence.

As shown in Figure 52, yeast cells transformed with the plasmid construct pYES2/FAD2-3 containing the cottonFAD2-3 open reading frame were found to have an appreciable accumulation of linoleic acid (18:2), not normally present in wild-type yeast cells. A slight amount of 9,12- hexadecadienoic acid (16:2) was also observed. As shown in Figure 52, the 18:2 and 16:2 fatty acids were not detected in the control yeast cells transformed by the shuttle vector pYES2. Also, the oleate (18:1) peak in the transformed cells was noticeably smaller than the corresponding oleate peak in the control cells (Pirtle et al., 2001, in press).

Discussion

One genomic clone designated LCFg55 encompassing the *FAD2-3* gene was successful isolated by Drs. Robert and Irma Pirtle and Ms. Wisatre Kongcharoensuntorn. It has a 13.2-kb cotton genomic fragment encompassing a putative *FAD2-3* gene. Unfortunately, prior to its ligation into the lambda vector EMBL3, the genomic DNA insert in this clone was cleaved in the 5'-flanking region of the *FAD2-3* gene. Therefore, this genomic clone contained an incomplete *FAD2-3* gene, consisting of the FAD2 protein-coding region (1,155 bp), 242 bp of the 3'-untranslated region, and only 568 bp of the 5'-untranslated region. Since the genomic clone LCFg55 contained an incomplete

Figure 51. The nucleotide sequence of the cotton *FAD2-3* coding region generated by PCR amplification of the *FAD2-3* gene in the pCFg55 plasmid vector. The numbers on the right refers to nucleotide residues. The sequences used for designing the forward and reverse primers for the cotton FAD2-3 coding region PCR fragment are underlined. The sequences upstream before the *Sst*I site and after the *EcoR*I site are part of the sequences of the pYES2 shuttle vector. The sequences between *Sst*I and *EcoR*I sites are identical with the open reading frame of the *FAD2-3* gene in Figure 47.

CCCCGGATCGGACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAAGCTTGGT 60

*Sst*I
ACCGAGCTCATGGGTGCAGGTGGCAGAATGTCGGTTCCTCCAAGTCAAAGGAAACAAGAA 120

TCGGGCTCAATGAAAAGAGTCCCTATATCTAAACCACCATT TACTCTCAGTGAAATAAAA 180

AAAGCCATCCCACCACACTGTTTCCAACGCTCACT TATCCG TTCATTT TCCTATCTCGTT 240

TACGACTTCATT TTAGTCTCTATCTTTTACTACGTAGCCACCACTTACTTCCACAACCTC 300

CCTCAGCCACTATCTTTCGTCGCCTGGCCAATTTATTGGACTCT TCAAGGTTCAGTCCTC 360

ACTGGCGTTTGGGTTATCGCCCATGAATGCGGTCACCATGCTTTTAGCGATTACCAATGG 420

ATTGATGACACTGTCGGTCTCATCCTCCATTCATCCCTTCTTGTCCCGTACTT TTCGTGG 480

AAATATAGTCACCGACGTCACCATTCCAACACTGGT TCCCTTGAACGCGACGAAGTATTT 540

GTTCCGAAGAAACGGAGCAGCATTAGATGGTGGGCTAAATACCTCAACAATCCACCAGGT 600

CGTTTCGTCACAGTCACCATTCAGCTCACTCTCGGATGGCCTCTTTACTTAGCATTCAAT 660

GTAGCAGGTAGACCTTACGAAGGACTCGCTTGTCACTACAACCCATACGGTCCTATCTAC 720

AACGACCGTGAACGACTTCAAATCTACATATCCGACGTCGGTGTCCTTGCTGTCACCTAT 780

GGGCTGTACCGTCTCGTGTTAGCCAAAGGTCTAGCTTGGGTCATTTGCGTTTACGGTGTC 840

CCATTGCTCATCGTTAATGCATTCCTCGTCATGATCACATACTTGCAACACACTCACCCC 900

GCAT TACCACACTACGACTCATCCGAATGGGACTGGTTACGTGGAGCCCTCGCGACGGTC 960

GACCGAGATTATGGGATATTAAACAAGGTTTTCCATAACATAACTGATACTCATGTCGCT 1020

CATCATTTGTTTTCGACGATGCCGCATTACCACGCAATGGAAGCAACTAAGGCAATAAAA 1080

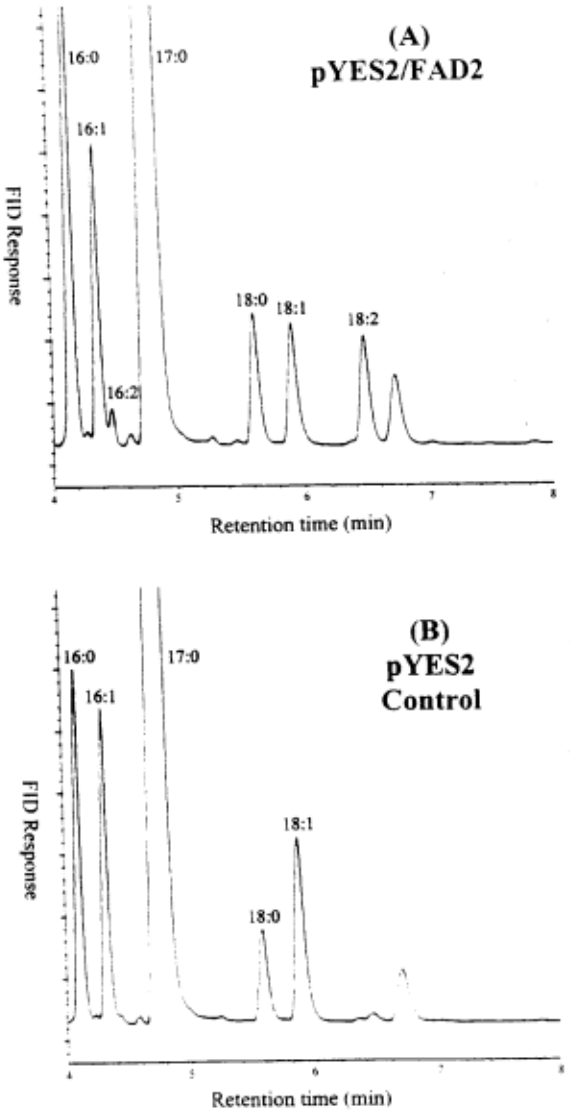
CCAATATTGGGAGAGTATTATTCATTTGATGGTACACCAGTTTATAAAGCGATATTTAGA 1140

GAGGCAAAGGAGTGTATTTACGTTGAACCAGACGAAGGTGAGCAGAGCAGCAAAGGTGTA 1200

*Eco*RI
TTTTGGTTTAGAAATAAGATCTAAGAATTCTGCAGATATCCATCACACTGGCGGCCGCTC 1260

GAGCATGCATCTAGAGGGCCGCATCATGTAATTAGTTATGTCACGCTTACATTCACGCC 1320

Figure 52. Gas chromatographic analysis of fatty acid methyl esters extracted from yeast transformants using frame ionization detection (FID). (A) *S. cerevisiae* strain INVSc1 cells (Invitrogen) transformed with the recombinant plasmid pYES/FAD2-3 were grown in SC-U (synthetic complete minus uracil) medium containing 2% galactose and 2% raffinose at 30°C for 3 generations. (B) As a control, *S. cerevisiae* strain INVSc1 cells (Invitrogen) transformed with the pYES2 shutle vector (Invitrogen) were grown in SC-U medium containing 2% galactose and 2% raffinose at 30°C for 3 generations. An additional peak with a retention time corresponding to that of linoleic acid (18:2) and a minor peak corresponding to 9,12-hexaadecadienoic acid (16:2) are seen in (A). Heptadecanoic acid (17:0) was the internal standard used for quantification and comparison of fatty acid content. In this analysis, the transformants expressing cotton *FAD2-3* had a linoleic acid content of 9.8% (by weight) of total fatty acid content.


(A)
pYES2/FAD2
16:0
16:1
17:0
18:0
18:1
18:2
16:2
FID Response
Retention time (min)
(B)
pYES2
Control
16:0
16:1
17:0
18:1
18:0
FID Response
Retention time (min)

FAD2-3 gene, it became necessary to isolate the overlapping genomic clone containing the complete *FAD2-3* gene. Using a unique, *FAD2-3* gene-specific 5'-flanking region hybridization probe, the overlapping genomic clone encompassing the *FAD2-3* gene designated LCFg24 was isolated. It was found to have a 12.3-kb cotton genomic DNA fragment. The nucleotide sequence encompassing the *FAD2-3* gene in this clone is 7,914-bp in length. The *FAD2-3* open reading frame is continuous without any introns and has 1,155 bp. The cotton *FAD2-3* gene encodes a putative FAD2 polypeptide of 384 amino acid residues. This *FAD2-3* gene has a 3'-UTR of 242 bp downstream from the UAA termination codon, with the near upstream polyadenylation signal (AATCAA, nt 4,612-4,617, in Figure 47) 34 bp upstream from the 3'-polyadenylation site (nt 4,646 in Figure 47).

The 5'-flanking region has a 2,967-bp intron located only 12 bp upstream from the ATG initiation codon. The 5'-UTR of the putative mRNA would be about 130 nucleotides in length. The presumptive promoter/enhancer elements that could function as positive regulatory elements in gene expression are located in the 5'-flanking region. For example, the basal promoter element (TATA box) is found 40 nucleotide upstream of the putative cap site. Potential seed-specific regulatory elements (E box or bHLH) with the consensus sequence CANNTG are found 109 bp and 135 bp upstream from the potential cap site. The E box motif has been shown to be a seed-specific regulatory element in the French bean β-phaseolin gene (Kawagoe et al., 1994). Three possible *Dof* transcription factors are found at 12 bp, 46 bp, and 61 bp upstream from the tentative cap site. *Dof* transcription factors are associated with genes involved in carbon metabolism in maize (Yanagisawa, 2000). Several potential light responsive promoter elements occur

in the 5'-flanking region. A G-box (CACGTG) (Terzaghi et al., 1995; Guilfoyle, 1997) occur 136 bp upstream from the tentative cap site. Two consensus GT-1 motifs (GRWAAW), general features of light-responsive promoters (Terzaghi et al., 1995; Guilfoyle, 1997), occur 88 bp and 121 bp prior to the cap site.

Alignment of the deduced amino acid sequences among the plant FAD2 desaturases showed that this cotton FAD2-3 polypeptide has the highest amino acid identity (85%) with the cotton FAD2-2 polypeptide (Liu et al., 1999) and also a high identity (75%) with the cotton FAD2-1 polypeptide (Liu et al., 1997). Moreover, this FAD2-3 polypeptide has high identity (72-85%) with the amino acid sequence of other plant such as *Arabidopsis*, soybean, and *Borago* FAD2 polypeptides. Comparison between the DNA sequence of this cotton *FAD2-3* gene and several other *FAD2* cDNA and gene sequences showed that this cotton *FAD2-3* DNA sequence has high identity to the cotton *FAD2-1* and *FAD2-2* genes (62% and 69%, respectively). The cotton *FAD2-3* gene is also highly similar to the *Arabidopsis FAD2* gene (64%) and the *Glycine ma*x *FAD2-1* gene (60%). However, the cotton *FAD2-3* gene sequence has only low to moderate identities (41-48%) with other fatty acid desaturase gene such as the *Arabidopsis FAD3*, *FAD6*, *FAD7*, *and FAD8* genes. Thus, this cotton *FAD2-3* gene has definitely shown to be a new member of the plant *FAD2* gene family.

Significant identities between both DNA and amino acid sequences of this cotton *FAD2-3* gene and other plant *FAD2* genes such as the cotton *FAD2-1*, *FAD2-2* genes (Liu et al., 1997; Liu et al., 1999), the *G. max FAD2-1* gene, and the *Arabidopsis FAD2* gene might help to provide an understanding how this cotton *FAD2-3* is expressed. The *Arabidopsis FAD2* gene has been shown to be equally expressed in both vegetative

tissues and embryos (Okuley et al., 1994). The soybean *FAD2-1* gene is expressed in developing seed while the soybean *FAD2-2* gene is expressed in both vegetative tissues and developing seeds (Heppard et al., 1996). The cotton *FAD2-1* gene is specifically induced in embryo development and is not detected in vegetative tissues like leaves, most likely having a major role in converting oleate into linoleate during storage lipid biosynthesis in cottonseed. In contrast, the cotton *FAD2-2* gene apparently has a low level constitutive expression in leaf tissue and throughout seed development, similar to the constitutive soybean *FAD2-2* gene. The cotton *FAD2-3* gene may be expressed in both vegetative tissues and developing seed because its high identity to the cotton *FAD2-2* gene both deduced amino acid sequence (85%) and nucleotide sequence (69%) (Liu et al., 1999).

The cotton *FAD2-3* mRNA transcripts has 130 nt in its 5'-UTR and 242 nt in its 3'-UTR (Pirtle et al., 2001, in press) while the cotton *FAD2-1* mRNA transcripts of Liu et al. (1999) has 78 nt in its 5'-UTR and 175 nt in its 3'-UTR. This eliminates any chance of significant homology between the *FAD2-3* mRNA 5'- and 3'-UTR sequences of the *FAD2-1* mRNA. Liu et al. (2001) analyzed the *FAD2-1* gene 5'-UTR intron structure across 31 speices of *Gossypium*, including the *Gossypium hirsutum* A and D genome orthologs. The cotton *FAD2-1* gene has a similar intron size to the *Arabidopsis FAD2* gene, which is 1,133-bp intron and 1,134-bp intron, respectively (Liu et al., 2001; Okuley et al., 1996). And, the location of intron is also quite the same. The intron of the *Gossypium FAD2-1* gene is located 9 bp upstream from the intiation codon (Liu et al., 2001) while the intron of the *Arabidopsis FAD2* gene is located merely 5 bp upstream from its ATG initiation codon (Okuley et al., 1996). In contrast, the cotton *FAD2-3* gene

has a much larger which is 2,967 bp and the locaton of intron is in a slightly different position, which is located 12 bp upstream from the translation start codon. Like introns of other higher plant genes, intron in the *FAD2-3* gene is AT rich with a T content of 44% and an A content of 30%. The differences in both the relative position and the sizes of the 5'-UTR introns are obvious structural differences between the *FAD2-3* and *FAD2-1* genes. This might result in the differential expression of the two genes as discussed in the above. The presence of the large introns in the 5'-flanking regions of plant *FAD2* genes might have the effects on their gene expressions. The positive effects of intron in 5'-UTRs on gene expression has been shown in a number of plant genes. The 5'-UTR intron having an enhancer-like element in *Arabidopsis* elongation factor 1β gene is required for its high levels of gene expression (Gidekel et al.,). Also, the 5'-UTR intron of a soybean phosphoenolpyruvate carboxylase gene dramatically increases gene expression in plant cells (Kato et al., 1998).

A number of plant microsomal *FAD2* and *FAD3* cDNAs and genes have been isolated and characterized, but the roles and biochemical functions of the expressed polypeptide products are still largely uncharacterized. Yeast cells have been used successfully for functionally expression of several animal and plant fatty acid desaturases localized in the endoplasmic reticulum, such as the rat Δ-6 desaturase (Aki et al., 1999), the *C. elegans* Δ-5 desaturase (Watts and Browse, 1999) and the *Arabidopsis* FAD2 enzyme (Covello and Reed, 1996; Kajiwara et al., 1996). Yeast cells can only produce the monounsaturated fatty acid palmitoleic (16:1) and oleic (18:1) acid with the *OLE1* gene (Mitchell and Martin, 1995). They cannot produce the polyunsaturated fatty acid linoleic (18:2) and linolenic (18:3) acids or in other words cannot convert oleic acid

(18:1) into linoleic acid (18:2) and linolenic (18:3) since they do not have *FAD2* and *FAD3* genes found in plants. Yeast cells transformed with the plasmid construct pYES2/FAD2-3 containing the cotton *FAD2-3* open reading frame were found to have an appreciable accumulation of linoleic acid (18:2), not normally present in wild-type yeast cells. A slight amount of 9,12- hexadecadienoic acid (16:2) was also observed. The 18:2 and 16:2 fatty acids were not detected in the control yeast cells transformed by the shuttle vector pYES2. Also, the oleate (18:1) peak in the transformed cells was noticeably smaller than the corresponding oleate peak in the control cells, clearly indicating the conversion of oleate into linoleate in the yeast cells containing the cotton FAD2-3 coding region. This conversion indicates that the *FAD2-3* gene is truly expressed in the yeast *S. cerevisiae*, and that the cotton *FAD2-3* gene is indeed is a functional gene. Thus, the cotton *FAD2-3* gene has been functionally identified, since it encodes an enzyme that catalyzes the desaturation of oleate into linoleate.

SUMMARY AND SIGNIFICANCE

The synthesis of palmitic, oleic, and linoleic acids and their incorporation into triacylglycerols are related metabolically, as shown in Figure 2. Their relative proportions are determined by regulation of carbon flux through these complex metabolic pathways. The metabolic relationship between palmitic acid, oleic acid and linoleic acid biosynthesis make it possible to manipulate the relative portion of these fatty acids in cottonseed oil profiles. As can be seen in Figures 1 and 2, palmitoyl-ACP thioesterase (PATE) and oleoyl-PC desaturase (FAD2) may be two key enzymes regulating cottonseed fatty acid composition toward higher oleic acid levels. The key to manipulating the proportions of these fatty acids in seed oils is an understanding of the regulatory points in these pathways and the ability to specifically alter them. It is necessary to determine the structure and function of the *Gossypium hirsutum PATE* and *FAD2* genes, since understanding the structure and function of these two genes will provide the necessary information to alter fatty acid composition of cottonseed oil.

In Chapter 1, the isolation and analysis of the cotton *PATE* cDNA and *PATE* gene were described (Pirtle et al., 1999; Yoder et al., 1999). The architecture of the cotton *PATE* gene and the potential regulatory (promoter/enhancer) elements of this cotton gene were identified. The cotton *PATE* cDNA clone has a 1.7-kb insert, with a coding region of 410 amino acids, lacking codons for the three N-terminal amino acids (Pirtle et al., 1999). The predicted amino acid sequence of the PATE preprotein has a characteristic stromal-targeting domain and a 63% identity to the *Arabidopsis* FatB1 thioesterase

sequence (Dörmann et al., 1994). A cotton genomic clone containing a 17.4-kb DNA segment was found to encompass a palmitoyl-ACP thioesterase (*PATE* or *FatB1*) gene (Yoder et al., 1999). The gene spans 3.6 kb with six exons and five introns. The six exons are identical in nucleotide sequence to the open reading frame of the corresponding cDNA, and would encode a preprotein of 413 amino acids. The preprotein was identified as a FatB thioesterase from its deduced amino acid sequence similarity to those of other FatB thioesterase preproteins. A 5'-flanking region of 914 bp was sequenced, with the potential promoter/enhancer elements, including basic helix-loop-helix elements (E boxes). Alkaline blot hybridization of cotton genomic DNA suggested the presence at least two *FatB1* thioesterase genes in the allotetraploid cotton genome.

In Chapter 2, the construction of *PATE* sense and antisense vectors for altering oilseed composition is described. Four plasmid constructs for both constitutive and seed-specific antisense RNA suppression and gene-transgene co-suppression of *PATE* gene expression were successfully generated. Dr. Tu T. Huynh (Huynh, 2001) of the Chapman laboratory at UNT, used these constructs to examine the *in planta* role of the cotton FatB thioesterase in transgenic cotton plants harboring the *PATE* sense and antisense constructs. The expression of the antisense PATE/pBI121 construct in transgenic cotton plants did not show a reduction in the relative percentages of palmitic acid in somatic embryos compared to the pBI121 control transgenic plants and wild type plants (cv. Coker 3.2). The antisense transgenic cotton plants had two different phenotypes, a dwarfed phenotype (30%) and a normal phenotype. The dwarf antisense transgenic cotton plants did not develop normally, having smaller bolls and smaller leaves. By comparing equivalent amounts of dwarf tissues with control tissues, the

overall fatty acid composition of dwarf plants was much the same as in the normal plants. This suggested that plants tend to distribute sufficient amounts of palmitic acid to only one or two viable seeds, rather than to distribute insufficient amounts of palmitic acid into many non-viable seeds, when the level of palmitic acid was decreased. Therefore, palmitic acid levels may be regulated in such a way that a certain range of palmitic acid levels must be maintained to avoid lethal effects (Huynh, 2001).

The fatty acid composition profile of somatic embryos overexpressing the sense PATE/pBI121 construct was elevated (up to 70% of the total) as compared to the vector-only (pBI121) control embryos and wild-type control (cv. Coker 3.2) embryos (25%). Stearic acid levels (18:0) were also elevated up to 20% of the total in these transgenic lines. A corresponding drop in oleic acid (18:1) and linoleic acid (18:2) levels was also anticipated. This suggested that indeed the *FatB* gene was over-expressed in these transgenic lines, resulting in the elevated levels of palmitic acid in these tissues. Therefore, overexpression of the *FatB* gene directly influences the fatty acid composition profile of the transgenic somatic embryos. These transgenic plants also showed the influence of palmitic acid content on plant membrane function. Plantlets developed from transgenic somatic embryos which have high levels of palmitic acid and very low levels of polyunsaturated fatty acids with less than 20% 18:1 and 18:2 (the major components of lipids in the thylakoid membranes) did not develop normally in soil due to the less efficient photosynthetic machinery (Huynh, 2001). The data presented in this chapter collectively indicates that recombinant plasmid constructs generated in this research altered *FatB* thioesterase expression, resulting in alteration of palmitic acid levels in

some cotton tissues. Therefore, the cotton FatB thioesterase enzyme directly participates in regulating palmitic acid levels.

In Chapter 3, the isolation and analysis of the cotton *FAD2-3* gene is described (Pirtle et al., 2001). The architecture of the cotton *FAD2-3* gene and the potential regulatory (promoter/enhancer) elements of this cotton gene were also determined. The functional expression of the cotton *FAD2-3* gene in transformed *Saccharomyces cerevisiae* cells was done. Two overlapping cotton genomic clones were found to encompass a Δ-12 fatty acid desaturase (*FAD2-3*) gene. The continuous *FAD2-3* coding region is 1,155 bp and would encode a protein of 384 amino acids. The *FAD2-3* gene has one large intron of 2,967 bp entirely within its 5'-untranslated region. Several potential promoter/enhancer elements, including several light responsive motifs, occur in the 5'-flanking region. Yeast cells transformed with a plasmid construct containing the cotton *FAD2-3* coding region accumulate an appreciable amount of linoleic acid (18:2), not normally present in wild-type yeast cells, indicating that the gene encodes a functional FAD2 enzyme (Pirtle et al., 2001).

The results obtained from this dissertation research should provide a starting point to delineate the role of the *FatB1* gene and *FAD2-3* gene expression in lipid biosynthesis and the information necessary to alter fatty acid composition of cottonseed oil. The data also provided the basis for future research to investigate the influence of phospholipids with palmitic acid and linoleic acid on plant membrane function.

As previously described, the sense *PATE* and antisense *PATE* transgenic plants did not develop normally when the *PATE* gene expression was changed throughout the plants. The transgenic plants showed the influence of palmitic acid content on plant

membrane function. Also a certain range of palmitic acid levels must be maintained to avoid lethal effects (Huynh, 2001). All of the data provide a correlation between palmitic acid levels and plant growth and development (Huynh, 2001). Future research will involve generating transgenic cotton plants harboring the seed-specific sense and antisense *PATE* constructs. Dörmann et al. (2000) showed that higher amounts of palmitic acid accumulated in *Arabidopsis* seeds when the *Arabidopsis thaliana FatB1* thioesterase cDNA was overexpressed under the transcriptional control of the seed-specific napin promoter. These transgenic plants should provide knowledge on how the *PATE* gene is specifically expressed in seeds and influences the fatty acid composition profile of seeds.

REFERENCES

Adams, A., Gottschling, D.E., Kaiser, C.A., Stearns, T. (1998) *Methods in Yeast Genetic*s. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 148.

Aki, T., Shimada, Y., Inagaki, K., Higashimoto, H., Kawamoto, S., Shigeta, S., Ono, K., and Suzuki, O. (1999) Molecular cloning and functional characterization of rat Δ-6 fatty acid desaturase. *Biochem. Biophys. Res. Com.* 255: 575-579.

Anderson, M.L.M., and Young, B.D. (1985) Quantitative Filter Hybridization. *In nucleic Acid Hybridization: a Practical Approach.* Edited by Hames, B.D. and Higgins, S.J. IRL press, Washington, DC pp. 73-111.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*. John Wiley & Sons, New York pp. 2.5.5, 6.0.1-6.3.6.

Baker, J., Steele, C., and Dure, L.,III (1988) Sequence and characterization of *6Lea* proteins and their genes from cotton. *Plant Mol. Biol.* 11: 277-291.

Beavan, M.V., and Flavell, R.B. (1983) A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature* 304: 184-187.

Bevan, M.V. (1984) Binary *Agrobacterium* Vectors for Plant Transformation. *Nucleic Acid Res.* 12: 8711-8721.

Becker, D.M., and Guarente, L. (1991) High-efficiency transformation of yeast by electroporation. *Methods in Enzymology*. Vol. 194, C.Guthrie and G.R. Fink, eds., Academic Press, New York, NY, pp.182-187.

Benton, W.D., and Davis, R.W. (1977) Screening λgt recombinant clones by hybridization to single plaques *in Situ*. *Science* 196: 180-182.

Blattner, F.R., Blechl, A.E., Denniston-Thompson, K. and Faber, H.E. (1978) Cloning human fetal-globin and mouse type-globin DNA: preparation and screening of shotgun Collections. *Science* 202: 1279-1284.

Brendel, V., Carle-Urioste, J.C., and Walbot, V. (1998) Intron recognition in plants, in: J. Bailey-Serres, D.R. Gallie (Eds.), *A Look Beyond Transription : Mechanism Determining mRNA Stability and Translation in Plants*. American Society of Plant Physiologists, Rockville, MD pp. 20-28.

Browse, J.A., McCourt, P.J., and Somerville, C.R. (1986) A mutant of *Arabidopsis* deficient in C18:3 and 16:3 leaf lipids. *Plant Physiol*. 81: 859-964.

Browse, J.A., and Somerville, C. (1991) Glycerolipid synthesis: biochemistry and regulation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42: 467-506.

Browse, J., Wu, J., James, Jr., D.W., and Dooner, H.K. (1994) A mutant of *Arabidopsis* deficient in the elongation of palmitic acid. *Plant Physiol.* 106: 143-150.

Browse, J., and Xin, Z. (2001) Temperature sensitive and cold acclimation. *Curr. Opin Plant Biol.* 4: 241-246.

Chapman, K.D., and Trelease, R.N. (1991) Acquisition of membrane lipids by differentiating glyoxysomes: role of lipid bodies. *J. Cell Biol.* 115: 995-1007.

Chasan, R. (1995) Engineering fatty acids-the long and short of it. *Plant Cell* 7: 235-237

Chlan, C.A., Borroto, K., Kamalay, J.A., and Dure, L., III (1987) Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequences and genomic organization of the alpha-globulin (vicillin) genes of cottonseed. *Plant Mol. Biol.* 9: 533-546.

Christie, W.W. (1982) Lipid Analysis, 2nd ed., Pegamon Press, New York, NY, pp. 52-54

Cline, K., and Henry, R. (1996) Import and routing of nucleus-encoded chloroplast proteins. *Annu. Rev. Cell dev. Biol.* 12: 1-26.

Covello, P.S., and Reed, D.W. (1996) Functional expression of the extraplastidial *Arabidopsis thaliana* oleate desaturase gene (*FAD2*) in *Saccharomyces cerevisiae. Plant Physiol. 111: 223-226.*

Davies, H.M., Anderson, L., Fan, C. and Hawkins, D.J. (1991) Developmental induction, purification and futher characterization of 12:0-ACP thioesterase from immature cotyledons of *Umbelllularia californica*. *Arch. Biochem. Biophys.* 290: 37-45.

Dehesh, K., Edwards, P., Hayes, T., Cranmer, A.M., and Fillatti, J. (1996) Two novel thioesterases are key determinants of the bimodal distribution of acyl chain length of *Cuphea palustris* seed oil. *Plant Physiol.* 110: 203-210.

Del Vecchio, A.J. (1996) High-laurate canola. *Inform* 7: 230-243.

Dörmann, P., Voelker, T.A., and Ohlrogge, J.B. (1995) Cloning and expression in *Escherichia coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long-chain acyl-acyl carrier proteins. *Arch.Biochem.Biophys.* 316: 612-618.

Dörmann, P., Voelker, T.A., and Ohlrogge, J.B. (2000) Accmulation of palmitate in *Arabidopsis* mediated by the acyl-acyl carrier protein thioesterase *FatB1*. *Plant Physiol.* 12: 637-643.

Fan, J., Ranu, R.S., Smith, C., Ruan, C., and Fuller, C.W. (1996) DNA sequencing with [α-^{33}P]-labeled ddNTP terminator: a new approach to DNA sequencing with Thermo Sequenase DNA polymerase. *BioTechniques* 21: 1132-1137.

Farmer, E.E. (1994) Fatty acid signaling in plants and their associated microorganisms. *Plant Mol. Biol.* 26: 1423-1437.

Flavell, R.B. (1994) Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl. Acad. Sci. USA* 91: 3490-3494.

Feinberg, A.P., and Vogelstein (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132: 6-13.

Feinberg, A.P., and Vogelstein (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. (Addendum) *Anal. Biochem*. 137: 266-267.

Fox, B.G. Shanklin, J., Somerville, C., and Munck, E. (1993) Stearoyl-acyl carrier protein Δ9 desaturase from *Ricinus communis* is a diiron-oxo-protein. *Proc. Natl. Acad. Sci. USA* 90: 2486-2490.

Galau, G.A., Wang, H.Y.-C., and Hughes, D.W. (1991) Sequences of the *Gossypium hirsutum* D-genome alloallele of Legumin A and its mRNA. *Plant Physiol.* 97: 1268-1270.

Galau, G.A., Wang, H.Y.-C., and Hughes, D.W. (1992a) Cotton *Mat5-A* (C164) gene and *Mat5-D* cDNAs encoding methionine-rich 2S albumin storage proteins. *Plant Physiol.* 99: 779-782.

Galau, G.A., Wang, H.Y.-C., and Hughes, D.W. (1992b) Cotton *Lea4* (D19) and *LeaA2* (D132) Group 1 *Lea* genes encoding water stress-related proteins containing a 20-amino acid motif. *Plant Physiol.* 99: 783-788.

Gavel, Y., and von Heijne, G. (1990) A conserved cleavage-site motif in chloroplast transit peptides. *FEBS Lett*. 261: 455-458.

Gibson, S., Falcone, D.L., Browse, J., and Somerville, C.R. (1994) Use of transgenic plants and mutants to study the regulation and function of lipid composition. *Plant cell environ.* 17: 627-637.

Gidekel, M., Jimenez, B., and Herrera-Estrella, L. (1996) The first intron of the *Arabidopsis thaliana* gene coding for elongation factor 1 beta contains an enhancer-like element. *Gene* 170: 201-206.

Gietz, D., St.Jean, A., Woods, R.A., and Schiest, R.H. (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* 20: 1425.

Grula, J.W., Hudspeth., R.L., Hobbs, S.L., and Anderson, D.M. (1995) Organization inheritance and expression of acetohydroxyacid synthase genes in the cotton allotetraploid *Gossypium hirsutum*. *Plant Mol. Biol.* 28: 837-846.

Grundy, S.M. (1989) Monounsaturated fatty acids and cholesterol metabolism: implication for dietary recommendations. *J. Nutrition* 119: 529-533.

Guilfoyle, T.J. (1997) The structure of plant gene promoters. *Gene. Engin.* 19: 15-47.

Hamada, T., Nichiuchi, U., Komada, H., Nishimura, M., and Iba, K. (1996a) Modification of fatty acid composition by over-expression and antisense of a microsomal omega-3-fatty acid desaturase gene in transgenic tobacco. *Transgenic Research.* 5: 115-121.

Hamada, T., Nichiuchi, U., Komada, H., Nishimura, M., and Iba, K. (1996b) cDNA cloning of a wounding-inducible gene encoding a plastid ω-3 fatty acid desaturase from tobacco. *Plant Cell Physiol.* 37: 606-611.

Harwood, J.L. (1996) Recent advances in the biosynthesis of plant fatty acids. *Biochim. Biophys. Acta* 1301: 7-56.

Havaux, M. (1987) Effects of chilling on the redox state of the primary electron acceptor Q_A of photosystem II in the chilling-sensitive and resistant plant species. *Plant Physiol. Biocchem.* 25: 735-743.

Helleyer, A., Leadlay, P.F., and Slabas, A.R. (1992) Induction, purification and characterization of acyl-ACP thioesterase from developing seeds of oil seed rape (*Brassica napus*). *Plant Mol. Biol.* 20: 763-780.

Heppard, E.P., Kinney, A.J., Stecca, K.L., and Miao, G. (1996) Developmental and growth temperature regulation of two different microsomal ω-6 desaturase genes in soybeans. *Plant Physiol.* 110: 311-319.

Hilditch, T.P., and Williams, P.N. (1964) The Chemical Constitution of Natural Fats. Chapman and Hall, London.

Hughes, D.W., Wang, H.Y.-C., and Galau, G.A. (1993) Cotton (*Gossypium hirsutum*) *MatP6* and *MatP7* oleosin genes. *Plant Physiol.* 101: 697-698.

Hugly, S., and Somerville, C. (1992) A role for membrane lipid polyunsaturation in chloroplast biogenesis at low temperature. *Plant Physiol.* 99: 197-202.

Hunt, G. A. (1994) Messenger RNA 3'-end formation in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 45: 47-60.

Hunt, G. A. and Messing, J. (1998) mRNA polyadenylation in plants, in: J. Bailey-Serres, D.R. Gallie (Eds.), *A Look Beyond Transcription: Mechanism Determining mRNA Stability and Translation in Plants*. American Society of Plant Physiologists, Rockville, MD, pp. 29-39.

Huynh, T.T. (2001) Palmitoyl-acyl carrier protein thioesterase in cotton (*Gossypium hirsutum* L.): biochemical and molecular characterization of a major mechanism for the regulation of palmitic acid content. *a dissertation prepared for the degree of Doctoral of Philosophy*, University of North Texas.

Jaworski, J.G., Tai, H., Ohlrogge, J.B., and Beittenmiller, D.P. (1994) The initial reactions of fatty acid biosynthesis in plants. *Prog. Lipid Res.* 33: 47-54.

Jeffcoat, R., Brawn, P.R., Safford, R., and James, A.T. (1977) Properties of rat liver microsomal stearoyl-coenzyme A desaturase. *Biochem. J.* 161: 431-437.

Jefferson, R.A. (1987) *GUS* fusion:β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6: 3901-3907.

Jefferson, R.A. (1989) The *GUS* reporter gene system. *Nature* 342: 837-838.

Jones, A., Davies, H.M., and Voelker, T.A. (1995) Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterase. *Plant Cell* 7: 359-371.

Jones, L.A., and King, C.C. (1996) Cottonseed Oil. *In Bailey's Industrial Oil & Fat Products*, 5th Ed., Vol. 2. Edible Oil and Fat Product: Oils and Oil Seeds, Y.H. Hui Ed. (John Wiley), pp. 159-240.

Ju, R., Tian, Y., Shen, Q., Liu, C., and Mang, K. (1994) Cloning of polygalacturonase (*PG*) cDNA and inhibition effects of its antisense RNA on the expression of *PG* gene intransgenic tomato plants. *Chin J. Biotechnol.* 10: (2) 67-74.

Kato, T., Itoh, E., Whittier, R.F., and Shibata, D. (1998) Increase of foreign gene expression in monocot and dicot cells by an intron in the 5'untranslated region of a soybean phosphoenolpyruvate carboxylase gene. *Biosci. Biotechnol. Biochem.* 62: 151-153.

Kawagoe, Y., Campbell, B.R., and Murai, N. (1994) Synergism between CACGTG (G-box) and CACCTG cis elements is required for activation of the bean seed storage protein beta-phaseolin gene. *Plant J.* 5: 885-890.

Kajiwara, S., Shirai, A., Fujii, T., Toguri, T., Nakamura, K., and Ohtaguchi, K. (1996) Polyunsaturated fatty acid biosynthesis in *Saccharomyces cerevisiae* expression of ethanol tolerance and the FAD2 gene from *Arabidopsis thaliana*. *Appl. Env. Microbiol.* 62: 4309-4313.

Kearns, E.V., Hugly, S., and Somerville, C.R. (1991) The role of cytochrome b_5 in Δ-12 desaturation of oleic acid by microsomes of safflower (*Cathanus tinctorias L.*). *Arch. Biochem. Biophys.* 284: 431-436.

Kinney, A.J. (1997) Genetic engineering of oilseeds for desired traits. *In Genetic Engineering*, vol 19, J.K. Setlow, ed., Plenum Press, New York, pp. 149-166.

Kirsch, C., Hahlbrock, K., and Somssich, I.E. (1997) Rapid and transient induction of a parsley microsomal Δ12 fatty acid desaturase mRNA by fungal elicitor. *Plant Physiol*. 115: 283-289.

Knutzon, D.S., Bleibaum, J.L., Nelsen, J., Kridl, J.C., and Thompson, G.A. (1992) Isolation and characterization of two safflower oleoyl-acyl carrier protein thioesterase cDNA clones. *Plant Physiol.* 100: 1751-1758.

Knutzon, D.S., Thompson, G.A., Radke, S.E., Johnson, W.B., Knauf, V.C., and Kridl, J.C. (1992) Modification of *Brassica* seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. *Proc. Natl. Acad. Sci. USA* 89: 2624-2628.

Lewin, B. (2000) *Gene VII*. Oxford University Press, Oxford, England.

Loader, N.M., Woolner, E.M., Helleyer, A., Salbas, A.R., and Safford, R. (1993) Isolation and characterization of two *Brassica napus* embryo acyl-ACP thioesterase cDNA clones. *Plant Mol. Biol*. 23: 769-778.

Los, D.A., and Murata, N. (1998) Structure and expression of fatty acid desaturases. *Biochem. Biophys. Acta.* 1394: 3-15.

Liu, Q., Singh, S.P., Brubaker, C.L., Sharp, P.J., Green, A.G., and Marshall, D.R. (1997) *In Physiology, Biochemistry and Molecular Biology of Plant Lipids,* William, J.P., Khan, M.U., and Lem, N.A.W., eds. Kluwer Academic Publishers, Boston, MA, pp.383-385.

Liu, Q., Singh, S.P., Brubaker, C.L., Sharp, P.J., Green, A.G., and Marshall, D.R. (1999) Molecular cloning and expression of a cDNA encoding a microsomal ω-6 fatty acid desaturase from cotton *(Gossypiun hirsutum)*. *Aust. J. Plant Physiol.* 26: 101-106.

Liu, Q., Brubaker, C.L., Green, A.G., Marshall, D.R., Sharp, P.J., and Singh, S.P. (2001) Evolution of the *FAD2-1* fatty acid desaturase 5'UTR intron and the molecular systematics of *Gossypium (Malvaceae)*. *Amer. J. Botany* 88: 91-102.

Matzke, M.A., and Matzke, A.J.M. (1995) How and why plants inactivate homologous (trans) Genes. *Plant Physiol.* 107: 679-685.

Merion, M. and Warren, W. (1989) Purification of supercoiled plasmids from crude cell lysates using high performance anion exchange chromatography. *Biotech*. 7 (1): 60-67.

Michaels, S.D., John, M.C., and Amasino, R.M. (1994) Removal of polysaccharides from plant DNA by ethanol precipitation. *BioTechniques* 17: 274-276.

Miquel, M., and Browse, J. (1992) *Arabidopsis* mutants deficient in polyunsaturated fatty acid synthesis. *J. Biol. Chem.* 267: 1502-1509.

Miquel, M., James, D Jr, Donner, H., and Browse, J. (1993) *Arabidopsis* requires polyunsaturated lipids for low-temperature survival. *Proc. Natl. Acad. Sci. USA* 90(6): 208-212.

Mitchell, A.G., and Martin, C.E. (1995) A novel cytochrome b_5-like domain is linked to the carboxyl terminus of the *Saccharomyces cerevisiae* Δ9 fatty acid desaturase. *J. Biol. Chem.* 270: 29766-29772.

Mitchell, P., and Tjian, R. (1989) Transcriptional Regulation in mammalian cells by sequence-specific DNA-binding proteins. *Science* 245: 371-378.

Murata, N., and Wada, H. (1995) Acyl-lipid desaturases and their importance in the tolerance and acclimatization to cold of cyanobacteria. *Biochem. J.* 308: 1-8.

Ni, W., and Trelease, R.N. (1991) Two genes encode the two subunits of cottonseed catalase. *Arch. Biochem. Biophys.* 289: 237-243.

Nunberg, A.N., Li, Z., Bogue, M.A., Vivekananda, J., Reddy, A.S., and Thomas, T.L. (1994) Developmental and hormonal regulation of sunflower helianthinin genes: proximal promoter sequences confer regionalized seed expression. *Plant Cell* 6: 473-486.

Ohlrogge, J.B., Shine, W.E., and Stumpf, P.K. (1978) Fat metabolism in higher plants: Characterization of plant acyl-ACP and acyl-CoA hydrolases. *Arch. Biochem. Biophys*. 189: 382-391.

Ohlrogge, J.B., Jarworski, J.G., and Post-Beittenmiller, D. (1993) *De novo* fatty acid Biosynthesis. In *Lipid Metabolism in Plants*, T.S. Moore, Jr., Ed. (CRC Press, Boca Raton, FL), pp. 3-32.

Ohlrogge, J.B. (1994) Design of new plant products: engineering of fatty acid metabolism. *Plant Physiol.* 104: 821-826.

Ohlrogge, J., and Browse, J. (1995) Lipid biosynthesis. *Plant Cell* 7: 957-970.

Ohlrogge, J.B., and Jaworski, J.G. (1997) Regulation of fatty acid synthesis. *Annu. Rev. Plant Mol. Biol.* 48: 109-136.

Oil Crops Yearbook. USDA-ERS. July, 1995.

Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., and Browse, J. (1994) *Arabidopsis FAD2* gene encodes the enzyme that essential for polyunsaturated lipid synthesis. *Plant Cell* 6: 147-158.

Parthasarathy, S., Khoo, J.C., Miller, E., Barnett, J., Witztum, J.L., and Steinberg, D. (1990) Low density lipoprotein rich in oleic acid is protected against oxidative modification: Implication for dietary prevention of atherosclerosis. *Proc. Natl. Acad. Sci. USA* 87: 3894-3898.

Paterson, A.H., Brubaker, C.L., and Wendel, J.F. (1993) A rapid method for extraction of cotton (*Gossypium spp*.) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol. Biol. Rep.* 11: 122-127.

Pirtle, I.L., Kongcharoensuntorn, W., Nampaisansuk, M., Knesek, J.E., Chapman, K.D., and Pirtle, R.M. (2001) Molecular cloning and functional expression of the gene for a cotton Δ-12 fatty acid desaturase (FAD2). *Biochim. Biophys. Acta* (in press).

Pirtle, R.M., Yoder, D.W., Huynh, T.T., Nampaisansuk, M., Pirtle, I.L., and Chapman, K.D. (1999) Characterization of a Palmitoyl-Acyl Carrier Protein Thioesterase (*FatB1*)in Cotton. *Plant Cell Physiol.* 40: 155-163.

Pollard, M.R., Anderson, L., Fan, C., Hawkins, D.J., and Davies, H.M. (1991) A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of *Umbelllularia californica*. *Arch. Biochem. Biophys.* 284: 306-312.

Reed, K.C., and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon Membranes. *Nucleic Acids Res*. 13: 7207-7221.

Roberts, J.K., DeSimone, N.A., Lingle, W.L., and Dure, L., III (1993) Cellular concentrations and uniformity of cell-type accumulation of two *Lea* proteins in cotton embryos. *Plant Cell* 5: 769-780.

Routaboul, J., Fisher, S.F., and Browse, J. (2000) Trienoic fatty acids are required to maintain chloroplast function at low temperatures. *Plant Physiol.* 124: 1697-1705.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning : A Laboratory Manual*, 2nded." Cold Spring Habor Laboratory Press, Cold Spring Habor, NY.

Seymour, G.B., Fray, R.G., Hill, P., and Tucker, G.A. (1993) Down-regulation of two non-Homologous endogenous tomato genes with a single chimeric sense gene construct. *Plant Mol Biol.* 23(1): 1-9

Shanklin, J., and Cahoon, E.B. (1998) Desaturation and related modifications of fatty acids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49: 611-641.

Shanklin, J., Whittle, E., and Fox, B.G. (1994) Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry* 33: 12787-12794.

Shine, W.E., Mancha, M., and Stumpf, P.K. (1976) Fat metabolism in higher plants: The function of acyl thioesterases in the metabolism of acyl-Coenzymes A and acyl-acyl carrier proteins. *Arch. Biochem. Biophys.* 172: 110-116.

Simpson, G.G., and Filipowicz, W. (1996) Splicing of precursors to mRNA in higher plants: Mechanism, regulation and subnuclear organization of the spliceosomal machinery. *Plant Mol. Biol.* 32: 1-41.

Slabas, A.R., and Fawcett, T. (1992) The biochemistry and molecular biology of plant lipid biosynthesis. *Plant Mol. Biol.* 19: 169-191.

Somerville, C., Browse, J., Jaworski, J. and Ohlrogge, J.B. (2000) Lipids. *In Biochemistry and Molecular Biology of Plants*, B. Buchanan, W. Gruissem, R. Jones, eds., American Society of Plant Physiologists, Rockville, MD, pp 456-527.

Tabor, S., and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 84: 4767-4771.

Terzaghi, W.B., and Cashmore, A.R. (1995) *Annu. Rev. Plant Physiol. Mol. Biol.* 46: 445-474.

Thomas, T.L. (1993) Gene expression during plant embryogenesis and germination: an Overview. *Plant Cell* 5: 1401-1410.

Tiemeier, D.C., Tilghman, A., and Leder, P. (1977) Purification and cloning of a mouse ribosomal gene fragment in coliphage lambda. *Gene* 2: 173-191.

Tocher, D.R., Leaver, M.J., and Hodgson, P.A. (1998) Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. *Prog. Lipid Res.* 37: 73-117.

Topfer, R., and Martini, N. (1994) Molecular cloning of cDNAs or genes encoding proteins involved in de novo fatty acid biosynthesis in plants. *J. Plant Physiol.* 143: 416-425.

Voelker, T.A., Worrell, A.C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D.J., Radke, S.E., and Davies, H.M. (1992) Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science* 257: 72-74.

Voelker, T. (1996) Plant acyl-ACP thioesterases: chain-length determining enzymes in plant fatty acid biosynthesis. *In Genetic Engineering*, Vol.18. Edited by Setlow, J.K. pp. 111-133. Plenum Press, New York.

Voelker, T.A., Jones, A., Cranmer, A.M., Davies, H.M., and Knutzon, D.S. (1997) Broad-range and binary-range acyl-acyl-carrier-protein thioesterases suggest an alternative mechanism for medium-chain production in seeds. *Plant physiol.* 114: 669-677.

Watts, J.L., and Browse, J. (1999) Isolation and characterization of a Delta 5-fatty acid desaturase from *Caenor habditis* elegans. *Arch. Biochem. Biophys.* 362: 175-182.

Wang, G., Blakesley, R.W., Berg, D.E., and Berg, C.M. (1993) pDUAL: A transposon-based cosmid cloning vector for generating nested deletions and DNA sequencing templates *in vivo. Proc. Natl. Acad. Sci. USA* 90: 7874-7878.

Wen-jun, S., and Forde, B.G. (1989) Efficient Transformation of *Agrobacterium spp*. by High Voltage Electroporation. *Nucleic Acids Res.* 17: 8385

Williams, M.E., Foster, R. and Chua, N. -H. (1992) Sequence flanking the hexameric G-box core CACGTG affect the specificity of protein binding. *Plant Cell* 4: 485-496.

Wu, J., Lightner, J., Warwick, N., and Browse, J. (1997) Low temperature damage and subsequent recovery of *fab1* mutant *Arabidopsis* exposed to 2°C. *Plant Physiol.* 113: 347-356.

Yanagisawa, S. (2000) *Dof1* and *Dof2* transcription factors are associated with expression of mutiple genes involved incarbon metabolism in maize. *Plant J.* 21: 281-288.

Yoder, D.W., Nampaisansuk, M., Pirtle, I.L., Chapman, K.D., and Pirtle, R.M. (1999) Molecular cloning and nucleotide sequence of a gene encoding a cotton palmitoyl-acyl carrier protein thioesterase. *Biochim. Biophys. Acta* 1446: 403-413.

York, D., Welch, K., Goryshin, I.Y., and Reznikoff, W.S. (1998) Simple and efficient generation in vitro of nested deletions and inversions: Tn5 intramolecular transposition. *Nucleic Acids Res.* 26: 1927-1933.

Young, R.A. (1991) RNA polymerase II. *Annu. Rev. Biochem.* 60: 689-715.

Yuan, L., Nelson, B. A., and Caryl, G. (1996) The catalytic cysteine and histidine in the plant acyl-acyl carrier protein thioesterases. *J. Biol. Chem.* 271: 3417-3419.