ZHANG, PING. The Applications of Candidate Gene Approaches in the Characterization of Seed Quality Traits in Soybean. (Under the direction of Ralph Dewey.)

Over the past decade, candidate gene approaches have become increasing popular as a means to identify and characterize genes associated with the agronomically important traits in crop species. The studies outlined in this dissertation describe an attempt to elucidate the molecular basis of two important traits related to soybean seed quality through the implementation of this approach.

Soybean germplasm possessing a high stearic acid (18:0) content is favorable for meeting the demands for oxidatively stable vegetable oils that don’t require hydrogenation, a process that generates undesirable trans fatty acids. We isolated a novel \( \Delta^9 \)-stearoyl-ACP-desaturase gene, designated \textit{SACPD-C}, whose enzymatic function serves to convert stearic acid (18:0) to oleic acid (18:1) specifically during seed development. \textit{SACPD-C} was found to be mutated in two independent high stearic acid soybean lines, A6 and FAM94-41. A molecular marker was generated that can distinguish the mutant \textit{SACPD-C} gene in FAM94-41 and can thus serve as a useful tool in introducing the trait into elite germplasm using marker assisted selection.

Lowering the phytic acid levels in soybeans would not only enhance the nutritional value of the seed for its use as livestock feed, but also help reduce problems associated with phosphorus pollution of soils and groundwater. We applied a candidate gene approach to identify genes
responsible for one low phytic acid soybean germplasm G03PHY-443. Comparison of myo-inositol and myo-inositol phosphate intermediates between the wildtype and mutant lines suggested that the mutation occurs at an early step of the phytic acid biosynthetic pathway. Biochemical assays of phosphatidylinositol kinase and diacylglycerol kinase activities suggested that the lipid dependent branch of the phytic acid metabolic pathway may actually be up-regulated in the mutant G03PHY-443 germplasm. Even more surprising, analysis of what we considered the three best candidate genes failed to reveal any perturbation that could count for the reduced phytic acid phenotype. Cumulatively, the results from this study enhanced our understanding of phytic acid biosynthesis in soybean seed and suggested that nonconventional gene mutations are responsible for the low accumulations of the compound in line G03PHY-443.
The Applications of Candidate Gene Approaches in the Characterization of Seed Quality Traits in Soybean

by

Ping Zhang

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Crop Science
Raleigh, North Carolina
2008

APPROVED BY:

Dr. Wendy F. Boss
Dr. Rebecca S. Boston
Dr. Daniel W. Israel
Dr. Ralph E. Dewey
Chair of Advisory Committee
BIOGRAPHY

Born and raised in Anhui, China, Ping Zhang holds a Bachelor of Science from Anhui University where she developed her interest in biology. When her husband and she moved to Raleigh, North Carolina, she was fortunately accepted by Dr. Ralph Dewey’s lab to allow her passion for life science to grow. Upon completing graduate school, she would love to be a researcher either in academia or industry because she is remaining strong in faith that she would take a direct part in creating a new era, an era where the enormous successes brought by this field would eventually benefit humans the most.
ACKNOWLEDGEMENTS

First I would like to express my gratitude to Dr. Ralph Dewey for his knowledgeable guidance, consistent inspiration, and unyielding support throughout my graduate study. My gratitude is extended to my committee members, Dr. Boss, Dr. Boston, and Dr. Israel, for providing the critical direction and valuable advice of my research.

Next I would like to thank my friends for sharing their laughing and crying, discussing about things beyond science, and raising my spirits whenever I needed. I thank the members of Dr. Dewey’s lab and other labs for insightful views and experimental assistance toward my research. I thank Dr. Jeff Gillikin for his incredible aid in characterization of proteins. I thank Dr. Brian Phillippy for his great assistance in my phytic acid project. Special thanks to Ms. Carol Griffin who maximized the efficiency of my bench work in the lab due to her skillful management of the lab.

Finally I would like to acknowledge my family. My husband’s, my parents’ and my two sisters’ boundless love is always the major strength that gets me through all the difficulties. I extremely thank my husband for his caring, understanding, and tolerating during the course of my Ph. D study which makes this mission being completed much easier and less painful.
# TABLE OF CONTENTS

**LIST OF TABLES** ........................................................................................................ vii

**LIST OF FIGURES** ...................................................................................................... viii

**LIST OF ABBREVIATIONS** ........................................................................................ xi

The Application of Candidate Gene Approaches in the Characterization of Seed Quality Traits in Soybean

CHAPTER I
Candidate Gene Analysis of Mutant Soybean Germplasm ................................................. 1

I. Introduction .................................................................................................................. 2
II. Disease Resistance Genes ......................................................................................... 4
III. Lipid Metabolic Genes .............................................................................................. 6
IV. Pigmentation and Carbohydrate/Phytate Mutant Analysis ........................................ 13
V. Conclusions ................................................................................................................. 15
VI. References ................................................................................................................ 17
Original Research .............................................................................................................. 17

CHAPTER II
Mutations in a $\Delta^9$-Soyrtyl-ACP-Desaturase Gene Are Associated with Enhanced Stearic Acid Levels in Soybean Seeds ................................................................. 26

Abstract .......................................................................................................................... 26
Introduction ...................................................................................................................... 28
Materials and Methods ................................................................................................. 30
   Plant Materials .............................................................................................................. 31
   Gene Isolation and Analysis ......................................................................................... 31
   DNA and RNA Blot Analysis ...................................................................................... 32
   Molecular Marker Analysis ........................................................................................ 33
   Activity Analysis of Recombinant SACPD Enzymes .................................................. 34
Results .................................................................................................................................. 35
   Identification of a New Soybean $\Delta9$-Soyrtyl-ACP-Desaturase Gene ................... 35
   Spatial Distribution of SACPD-C Transcripts ............................................................ 37
   SACPD-C is Deleted in Mutant Line A6 ...................................................................... 37
   A Radical Amino Acid Substitution Is Found in the SACPD-C Gene of FAM94-41 .... 39
Identification of the low phyate conferring genetic loci··················  120
Refining the model of phytate biosynthesis in soybean·················  121

APPENDICS···································································································  123

Appendix I
Expression profiling on soybean leaves reveals integration of
ER- and osmotic-stress pathways ........................................  124

Appendix II
Analysis of Microarray Data .....................................................  126
A. Microarray experimental design .........................................  127
B. Date transformation, normalization and analysis .................  128

Appendix III
Additional Analysis of SACPD-C Gene and its Enzyme Product ······  132
A. Real-Time PCR of SACPD-C transcripts in
different soybean tissues......................................................  133
B. Western Blot Analysis of SACPD-C Using Antibody Generated
Against the Homologous Caster Enzyme...............................  135

References....................................................................................  137
LIST OF TABLES

CHAPTER III

Table 1
Genes among the seed-expressed ESTs (5,700) printed on the microarray chips that may be involved in the phytic acid biosynthesis

Table 2
Genes represented in the microarray chips that showed more than two-fold higher expression in the wildtype Prichard-RR comparing versus the mutant line G03PHY-443

105
106
LIST OF FIGURES

CHAPTER I

Figure 1
Primary storage lipid biosynthetic pathway in developing seeds of soybean .................................................. 9

Figure 2
A generalized scheme of stearic acid and phytic acid synthesis in developing soybean seeds ........................................... 25

CHAPTER II

Figure 1
Gene structure and predicted protein sequence of the SACPD-C gene of soybean .................................................. 58

Figure 2
RNA gel blot of SACPD-C transcripts in different soybean tissues ................................................................. 60

Figure 3
Analysis of SACPD-C in A6 and FAM94-41 .................................................. 61

Figure 4
Point mutation in the SACPD-C gene of FAM94-41 .......................... 62

Figure 5
CAPS marker analysis of representative F2 progeny from a FAM94-41 X N98-4445A-derived population .................................................. 63

Figure 6
In vitro synthesis of oleic acid using recombinant ∆9-S-ACP-Des proteins .................................................. 64

CHAPTER III

Figure 1
Phytic acid biosynthetic pathways proposed in plants .......................... 107
Figure 2
HPLC analysis of inositol phosphates in the wildtype Prichard-RR and the mutant G03PHY-443 soybean lines ................................. 108

Figure 3
Comparison of the levels of myo-inositol during three stages of soybean seed development between the wildtype Prichard-RR (Pri) and the low phytic acid mutant line G03PHY-443 (443) ................................................................. 109

Figure 4
Comparison of endogenous phosphatidylinositol (PI) kinase enzymatic activity between the wildtype line Prichard-RR (Pri) and the mutant line G03PHY-443 (443) .................................................................................. 110

Figure 5
A. Northern blot of the myo-inositol -3-phosphate synthase (MIPS) gene most abundantly expressed during seed development .......... 111
B. Western blot of the MIPS protein using the mid-maturation soybean seeds ........................................................................ 111

Figure 6
Alignment of 3 putative soybean ABC transporters with the maize ABC transporter involved in phytic acid biosynthesis (NP_00106060)...... 112

Figure 7
A. Southern blot analysis using a partial ABCT_63 transporter gene as a probe ................................................................. 114
B. Northern blot analysis of ABC transporters during 3 maturations of soybean seed development .................................................... 114

Figure 8
A. Comparison of putative soybean myo-inositol kinase (MIK1 and MIK2) and maize MIK (NP_001105790) ............................... 115
B. Southern blot analysis of Prichard-RR (Pri) and G03PHY-443 (443) probed with partial MIK1 cDNA ........................................... 115
C. Northern blot analysis of the same genotypes using RNAs from early- mid- and late- maturation soybean seeds ........................... 115
APPENDICS

Figure 1
Image of soybean cDNA chips used in the dissertation .......................... 125

Figure 2
Normalized data using the two-way ANOVA mixed modeling
statistical method ................................................................. 130

Figure 3
Real-time PCR analysis of SCAPD-C in developing seed, leaf, flower,
root and stem tissue ............................................................. 134

Figure 4
Western blot analysis of normal soybean cultivar Dare and enhanced 18:0
genotypes FAM94-41 and A6, using antibody generated against a casor
bean Δ9-stearoyl-ACP-desaturase enzyme .................................. 136
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCT</td>
<td>ATP-binding cassette transporter</td>
</tr>
<tr>
<td>ACCase</td>
<td>acetyl-CoA carboxylase</td>
</tr>
<tr>
<td>ACP</td>
<td>acyl carrier protein</td>
</tr>
<tr>
<td>ACS</td>
<td>acyl CoA synthetase;</td>
</tr>
<tr>
<td>Asp</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>CAPS marker</td>
<td>Cleaved Amplified Polymorphic Sequence Marker</td>
</tr>
<tr>
<td>DAG</td>
<td>diacylglycerol</td>
</tr>
<tr>
<td>DGAT</td>
<td>diacylglycerol acyltransferase</td>
</tr>
<tr>
<td>FAD2</td>
<td>ω-6 fatty acid desaturase</td>
</tr>
<tr>
<td>FAD3</td>
<td>ω-3 fatty acid desaturase</td>
</tr>
<tr>
<td>FAT-A</td>
<td>18:1-ACP thioesterase</td>
</tr>
<tr>
<td>FAT-B</td>
<td>16:0 (18:0)-ACP thioesterase</td>
</tr>
<tr>
<td>FAS</td>
<td>fatty acid synthase</td>
</tr>
<tr>
<td>G3P</td>
<td>glycerol 3-phosphate</td>
</tr>
<tr>
<td>IP₃</td>
<td>inositol trisphosphate</td>
</tr>
<tr>
<td>IP₄</td>
<td>inositol tetrakisphosphate</td>
</tr>
<tr>
<td>IP₅</td>
<td>inositol pentakisphosphate</td>
</tr>
<tr>
<td>IP₆</td>
<td>inositol hexakisphosphate</td>
</tr>
<tr>
<td>IPK1</td>
<td>myo-inositol phosphate kinase</td>
</tr>
<tr>
<td>KAS</td>
<td>3-keto-acyl-ACP synthase</td>
</tr>
<tr>
<td>lpa</td>
<td>low phytic acid</td>
</tr>
<tr>
<td>LPA</td>
<td>lysophosphatidic acid</td>
</tr>
<tr>
<td>LRR</td>
<td>leucine rich repeat</td>
</tr>
<tr>
<td>MIK</td>
<td>myo-inositol kinase</td>
</tr>
<tr>
<td>MIPS</td>
<td>myo-inositol 3-phosphate synthase</td>
</tr>
<tr>
<td>NBD</td>
<td>nucleotide-binding domain</td>
</tr>
<tr>
<td>PA</td>
<td>phosphatidic acid</td>
</tr>
<tr>
<td>PC</td>
<td>phosphatidylcholine</td>
</tr>
<tr>
<td>PI</td>
<td>phosphatidylinositol</td>
</tr>
<tr>
<td>PIP</td>
<td>phosphatidylinositol bisphosphate</td>
</tr>
<tr>
<td>PIP4, 5-kinas</td>
<td>phosphatidylinositol 4-phosphate 5-kinase</td>
</tr>
<tr>
<td>PLC</td>
<td>phospholipase C</td>
</tr>
<tr>
<td>RGA</td>
<td>resistance gene analogs</td>
</tr>
<tr>
<td>SACPD</td>
<td>Δ9-stearoyl-ACP-desaturase</td>
</tr>
<tr>
<td>16:0</td>
<td>palmitic acid</td>
</tr>
<tr>
<td>18:0</td>
<td>stearic acid</td>
</tr>
<tr>
<td>18:1</td>
<td>oleic acid</td>
</tr>
<tr>
<td>18:2</td>
<td>linoleic acid</td>
</tr>
</tbody>
</table>
18:3 – linolenic acid
$\Delta^9$- S-ACP-Des - $\Delta^9$-stearoyl-acyl carrier protein desaturase
CHAPTER I

(Literature Review)

Candidate Gene Analysis of Mutant Soybean Germplasm

R.E. Dewey and P. Zhang
Department of Crop Science
North Carolina State University

This manuscript was submitted as a chapter in the book <<Soybean Genomics>>, K.D. Bilyeu, M.B. Ratnaparkhe, eds., Wiley-Blackwell.
I. Introduction

A major goal for molecular genetic studies in soybean involves the identification and characterization of genes associated with important agronomic traits. In many cases, a trait of interest is manifest as a simple heritable mutation, either natural or induced, that gives rise to a desirable phenotype. Techniques such as map-based cloning have proven to be effective in identifying the specific genes responsible for such traits (Bhattacharyya 2008), yet are often time consuming and labor intensive. In cases where sufficient information is available from model systems concerning the molecular genetics or biochemistry of a specific metabolic pathway, and observed phenotypic outcomes that result from disruptions at a given step, reasonable predictions can be made regarding the gene or genes that may be involved when a similar phenotype is observed in soybean. This strategy, commonly referred to as a candidate gene approach, can serve as a relatively rapid and efficient means for characterizing gene mutations associated with traits of agronomic interest. In this chapter we will focus on how candidate gene approaches have been used to effectively define the molecular basis of important traits in soybean.

Candidate gene analysis for defining the molecular genetic basis of simply inherited phenotypic traits typically consists of four steps. First, one or more candidate genes are proposed according to cause-effect relationships that have been previously observed in the investigation of similar phenotypes in model species (such as *Arabidopsis thaliana*),
or alternatively, based on rational predictions deduced from an understanding of metabolite flow through well characterized biochemical pathways. Second, molecular polymorphisms having the potential of altering gene function are identified in the candidate gene between genotypes exhibiting measurable variability at the investigated trait. Third, an association is established between the molecular polymorphism and phenotype in question in a population segregating for the trait. Fourth, the candidate gene is validated. In cases where a mutation in a gene is believed to be responsible for the trait in question as a consequence of gene dysfunction, the ultimate means of verification is to demonstrate functional complementation through transgenic expression of the wild type gene.

Certain characteristics of soybean present challenges for candidate gene analyses. Molecular evolution studies support a model predicting the modern soybean genome to be the product of both an ancient and more recent duplication event (Shoemaker et al. 1996; Schlueter et al. 2004). Therefore, for any given gene product, there are typically multiple independent functional isoforms (paralogs) residing in the genome. This can complicate the candidate gene approach, in some cases requiring knowledge of all functional isoforms of a gene in order to identify the causal gene (or alternatively conclude that the particular gene family is not involved in the trait in question). Furthermore, soybean transformation remains a relatively labor intensive, time consuming process that is not routinely conducted in most soybean research labs.
Therefore, confirmation of candidate genes through genetic complementation is typically not used as a means of validation in soybean. However, when a candidate gene encodes an enzyme activity that can be assayed \textit{in vitro}, comparisons of mutant versus wild type activities of the recombinant enzymes can provide strong supportive evidence that an observed polymorphism in a candidate gene is responsible for the observed trait. In some cases, mutations resulting in an in-frame stop codon or frame-shift mutation have been observed upstream of an essential functional domain of the encoded product (see below). Under these circumstances, it is usually assumed that the mutation has rendered the gene nonfunctional.

Despite the above mentioned limitations, candidate gene analysis has proven to be a powerful tool in defining the molecular genetics of several important agronomic traits in soybean. Specific examples are highlighted below.

\textbf{II. Disease Resistance Genes}

One of the most important advancements in the field of plant molecular genetics has been the identification and characterization of disease resistance (R) genes (reviewed by Martin et al. 2003). Remarkably, from a wide diversity of plant species a small number of conserved structural motifs have been observed in R genes targeting a broad spectrum of pathogens (e.g. viruses, fungi, bacteria, insects and nematodes). The most abundant of the
R gene classes identified to date is defined by the presence of an N-terminal nucleotide-binding domain (NBD) and a C-terminal leucine rich repeat (LRR) motif. The NBD has been proposed to function in pathogen-dependent signal transduction cascades through a classical phosphorylation/dephospholation mechanism, while the LRR is believed to be involved in ligand binding and pathogen recognition via protein-protein interactions (McHale et al. 2006). The conservation of these motifs, particularly the NBD has provided a means for retrieving candidate R genes from essentially any plant genome.

Using degenerate primers corresponding to highly conserved regions of the NBD of the tobacco N and Arabidopsis RPS2 disease resistance genes, Kanazin et al. (1996) and Yu et al. (1996) identified several resistance gene analogs (RGAs) from soybean. In both studies, individual RGAs were genetically mapped to the vicinity of known resistance loci. Although this technique provided a rapid, powerful tool for identifying candidate R genes, subsequent analyses were hampered by the fact that RGAs reside throughout the genome as clustered arrays. For example, 16 RGAs were identified on a single 119 kb BAC clone corresponding to a region on soybean linkage group J where resistance to powdery mildew and Phytophthora stem and root rot had been mapped (Graham et al. 2002). Furthermore, 12 independent RGAs were found to map to a region of linkage group F where the Rsv1 gene that confers resistance to soybean mosaic virus resides (Jeong et al. 2001).
Defining the specific RGA that represents a genuine R gene for a given pathogen requires either the transformation of each candidate RGA into a susceptible background followed by pathogen screening, or high-resolution mapping to eliminate all RGAs within a cluster that do not perfectly co-segregate with the resistance phenotype. Owing to the recalcitrance of soybean transformation, the latter strategy has been the preferred option for soybean researchers. Through fine mapping, Hayes et al. (2004) were able to identify an RGA designated 3gG2 as the only candidate within the gene cluster that perfectly co-segregated with the major *Rsv1* soybean mosaic virus resistance allele. A PCR-based marker specific for the 3gG2 gene has been developed to facilitate marker-assisted breeding of *RsvI* (Shi et al. 2008). Although difficulties with regard to the clustering of RGA genes, together with the non-facile nature of soybean transformation remain an impediment to progress in defining specific R genes through a candidate gene approach, it is likely that continued advances in soybean genome information (Jackson 2008) and high-resolution mapping will help alleviate some of these obstacles.

III. Lipid Metabolic Genes

The myriad uses for soybean oil in both food and industrial applications have led to an intense investigation of the genetics and physiology that underlie the regulation of lipid storage reserves within the soybean seed. Oil quality is primarily determined by the fatty acid composition of the storage triacylglycerols, and what defines the optimal
combination and ratio of fatty acids in soybean oil differs depending on the specific end application. For example, oils containing minimal amounts of palmitic acid (16:0) are desired when used for human consumption because of the cholesterogenic nature of palmitic acid and its ability to promote arterial thrombosis formation (Hu et al. 1997; Hornstra and Kester 1997). Also, high levels of polyunsaturated fatty acids, particularly linolenic acid (18:3) reduce the oxidative stability of the oil and can lead to poor flavor quality. To meet the needs of the vegetable oil industry, plant breeders have developed an array of mutant germplasm lines (either natural or induced by mutagenesis) displaying a wide range of altered oil phenotypes. The availability of these lines, coupled with the wealth of information that has accrued regarding the biochemistry and molecular genetics of lipid metabolism in model plant systems, provides an ideal scenario for utilizing a candidate gene approach for mutant germplasm analysis.

A schematic representing the primary biochemical pathway responsible for triacylglycerol formation in oilseed species such as soybean is shown in Figure 1. The early steps of fatty acid formation occur within the plastids of the developing seed, whereas the subsequent attachment of the fatty acids to a glycerol backbone and the introduction of double bonds giving rise to polyunsaturated fatty acids take place in the cytosol/ER (Ohrogge and Jaworski 1997; Lung and Weselake 2006). The individual steps in the lipid biosynthetic pathway where researchers have been successful in identifying
the specific genes responsible for a novel oil phenotype using a candidate gene approach are highlighted in Figure 1.

The fatty acid thioesterase (FAT) enzyme that catalyzes the cleavage of palmitic and stearic (18:0) acids from their respective acyl carrier protein (ACP) intermediates during fatty acid biosynthesis is designated FAT-B (in contrast to FAT-A enzymes that recognized oleic acid substrates). Because 16:0 fatty acids do not become further modified from the time they are released from the plastid to when they become incorporated into triacylglycerols, the FAT-B enzymatic reaction is pivotal in establishing the amount of 16:0 found in the storage oil. Therefore, reductions in FAT-B enzymatic activity would be predicted to lead to a decrease in the palmitic acid content of the storage triacylglycerols. In a candidate gene investigation of soybean germplasm possessing the *fap*<sub>nc</sub> locus that confers a reduced palmitic acid phenotype, Wilson et al. (2001) concluded from Southern blotting assays that a copy of a *FAT-B* gene was deleted in lines carrying *fap*<sub>nc</sub>. Subsequently, it was determined that the soybean genome contains four unique functional isoforms encoding FAT-B enzymes, and that the isoform whose transcripts were most highly represented in soybean EST databases, *GmFATB1a*, was the gene specifically deleted in these lines (Cardinal et al. 2007). Molecular markers specific to *GmFATB1a* proved to be effective in following the *fap*<sub>nc</sub>-mediated reduced 16:0 trait in segregating breeding populations (Cardinal et al., 2008).
Figure 1. Primary storage lipid biosynthetic pathway in developing seeds of soybean. Steps where candidate gene approaches have been used to identify causal gene mutations are highlighted in gray. ACP, acyl carrier protein; LPA, lysophosphatidic acid; PA, phosphatidic acid; DAG, diacylglycerol; PC, phosphatidylcholine; G3P, glycerol 3-phosphate; ACCase, acetyl-CoA carboxylase; FAS, fatty acid synthase; KAS, 3-keto-acyl-ACP synthase; SACPD, 9-stearoyl-ACP-desaturase; FAT-B, 16:0 (18:0)-ACP thioesterase; FAT-A, 18:1-ACP thioesterase; ACS, acyl CoA synthetase; FAD2, ω-6 fatty acid desaturase; FAD3, ω-3 fatty acid desaturase; DGAT, diacylglycerol acyltransferase.
Although the metabolic flow of fatty acid intermediates depicted in the biochemical pathway shown in Figure 1 would correctly lead to the prediction that enhancement of FAT-B enzyme activity would lead to an increase in 16:0 in the storage oil (as proven using transgenic modification of oil content), such "gain-of-function" mutations are rarely obtained through mutagenesis. Instead, recessive mutations yielding an elevated palmitic acid phenotype would more likely be caused by mutations in the 3-keto-acyl-ACP synthase activity that elongates 16:0-ACP to 18:0-ACP (KAS II, Fig. 1). During an investigation of the \textit{fap2} locus associated with an enhancement of the 16:0 of the seed oil, it was revealed that the soybean genome expresses two major \textit{KAS II} genes, designated \textit{GmKAS IIA} and \textit{GmKAS IIB} (Aghoram et al., 2006). Sequence analysis of lines containing \textit{fap2} uncovered a point mutation within \textit{GmKAS IIA} that introduced a premature stop codon within the reading frame, a mutation that would be expected to render the encoded product nonfunctional, thus providing a rational explanation for the enhanced 16:0 phenotype associated with the \textit{fap2} locus.

Stearic acid levels are typically low in soybean (2 - 4% total fatty acid); therefore little attention has been paid toward further reducing its content in the oil. Mutant germplasm has been characterized, however, in which 18:0 levels as high as 30% have been observed (Pantalone et al. 2002). The most straightforward means for obtaining an elevated 18:0 phenotype via a recessive mutation would involve altering the stearoyl-ACP-desaturase (SACP D) enzyme responsible for converting 18:0-ACP to 18:1-ACP (Fig. 1). An initial
study of \textit{SACP}D genes in soybean identified two specific isoforms, \textit{SACP}D-A and \textit{SACP}D-B, yet neither gene appeared to be altered in germplasm possessing the \textit{fas}^a locus that mediates a 30\% 18:0 phenotype (Byfield et al. 2006). More recent studies, however, revealed a third isoform (designated \textit{SACP}D-\textit{C}) that is specifically expressed during soybean seed development (Zhang et al. 2008). A candidate gene analysis of germplasm containing the \textit{fas}^a locus showed that \textit{SACP}D-\textit{C} was completely deleted in this line (Zhang et al. 2008). Furthermore, an amino acid substitution mutation was found in the \textit{SACP}D-\textit{C} gene of soybean lines possessing \textit{fas}^{nc}, an elevated 18:0 locus that is allelic to \textit{fas}^a, yet less severe (9 - 15\% 18:0). Interestingly, in enzyme assays conducted \textit{in vitro} using recombinant proteins, the \textit{fas}^{nc}-associated mutant \textit{SACP}D-\textit{C} enzyme displayed greater activity than the wild type enzyme. Given the strict genetic correlation that was observed between the mutant \textit{SACP}D-\textit{C} gene and the elevated 18:0 phenotype, however, it was speculated that the consequence of the observed amino acid substitution may be to reduce enzyme stability \textit{in vivo} as opposed to reducing the inherent activity of the enzyme (Zhang et al. 2008).

The enzymes primarily responsible for the synthesis of polyunsaturated fatty acids during soybean seed development are ER-localized proteins termed FAD2 (converting 18:1 to 18:2) and FAD3 (converting 18:2 to 18:3) (Fig. 1). The genes encoding these enzymes represent the most obvious targets in mutant germplasm displaying reduced levels of 18:2 and/or 18:3. The soybean genome contains two distinct classes of \textit{FAD}2 genes. Genes
designated *FAD2-1* are expressed specifically during seed development, while *FAD2-2* genes are constitutively expressed throughout the entire plant. Within these classes, two functional *FAD2-1* isoforms (*Fad2-1a* and *Fad2-1b*) have been characterized (Tang et al. 2005), and at least three *FAD2-2* paralogs have been identified (Schlueter et al. 2007). Candidate gene analysis of germplasm line M23, an elevated 18:1 (reduced 18:2 + 18:3) line developed by X-ray mutagenesis, revealed that the embryo-specific *Fad2-1a* isoform is deleted in both the M23 parental line and elevated 18:1 segregants in breeding populations derived from M23 (Alt et al. 2005; Sandhu et al. 2007).

Candidate gene strategies have been particularly effective in revealing the molecular basis of low 18:3 soybean germplasm. Similar to the other lipid biosynthetic enzymes described above, FAD3 desaturases in soybean are encoded by multiple functional isoforms. In a seminal report by Bilyeu et al. (2003), three unique *FAD3* paralogs were characterized: Gm*FAD3A*, Gm*FAD3B* and Gm*FAD3C*. In a series of studies, primarily by this same group, candidate gene strategies have lead to a thorough understanding of the molecular genetics of linolenic accumulation in soybean oil. Mutations in all three soybean *FAD3* genes have been shown to be associated with lowering the 18:3 content of the seed. Breeders have developed multiple independent lines displaying a reduced 18:3 phenotype at a locus designated *fan*. In each case, null mutations (e.g. deletions, frameshifts, premature stop codons and splice junction defects) were found in the Gm*FAD3A* gene in germplasm possessing a *fan* allele (Bilyeu et al. 2003, 2005; Anai et
al. 2005; Chappell and Bilyeu 2006, 2007). Potentially debilitating mutations were also found within the \textit{GmFAD3B} and \textit{GmFAD3C} genes in additional reduced 18:3 germplasm lines (Bilyeu et al. 2005, 2006; Anai et al. 2005). By pyramiding these individual mutations, soybean germplasm containing just 1% linolenic acid was developed (Bilyeu et al. 2006), a trait highly desired by the food industry because of its enhanced oxidative and flavor stability. The residual 18:3 that remains in the triple mutant lines is likely either the product of plastid-localized 18:2 desaturase activities, or the product of yet a fourth ER localized \textit{FAD3} gene that has been reported (Anai et al. 2005).

Along with clarifying the genetic mechanisms responsible for altering soybean oil composition, characterizing the specific molecular lesions underlying these traits also provides information advantageous for the development of soybean cultivars with desirable oil profiles. In most of the studies described above, molecular markers were developed corresponding to the gene mutations that were revealed. Molecular markers based on causal mutations are "perfect" in that there is no chance that the trait of interest will become separated from the marker through recombination.

\textbf{IV. Pigmentation and Carbohydrate/Phytate Mutant Analysis}

Similar to the lipid biosynthetic pathway discussed above, the biochemistry and molecular genetics of pigment forming flavonoid pathways in plants have been
characterized to a level that makes candidate gene approaches for mutant analysis feasible. Previously published chromatographic data lead Toda et al. (2002) to predict that the gray pubescence coloration phenotype conferred by the recessive t locus in soybean is a consequence of deficient flavonoid 3'-hydroxylase activity. This supposition was supported by the discovery of a frameshift mutation within a gene encoding this activity that faithfully co-segregated with the phenotype. Using a similar rationale, it was proposed that the magenta flower color phenotype of soybeans possessing the recessive wm allele (versus the purple color mediated by Wm) could be explained by a mutation in a flavonol synthase gene (Takahashi et al. 2007). Candidate gene analysis of flavonol synthase genes from isogenic lines of the cultivar Harsoy revealed a debilitating frameshift mutation in wm-associated allele. The inability of the truncated protein to perform its normal function was further validated through in vitro enzyme assays of recombinant mutant versus wild type enzymes (Takahashi et al. 2007). Finally, prior knowledge of the role and function of chalcone synthase genes in pigment formation was the basis for the discovery that various duplications and deletions of members of this multigene family defined the I locus that controls seed coat color in soybean (Todd and Vodkin 1996; Senda et al. 2002; Kasai et al. 2007).

A final example that well exemplifies the utility of the candidate gene approach involves the characterization of a mutant soybean line that displayed both a decreased raffinosaccharide and phytic acid seed phenotype. This line, designated LR33, was
uncovered through a large scale screen of a mutagenized soybean population, followed by heritability studies that attributed the dual phenotype to a single, recessive locus (Sebastian et al. 2000). Although raffinosaccharides and phytate are products of distinct biochemical pathways, both can originate from myo-inositol 1-phosphate. This fact, combined with supportive in vivo labeling results, suggested a defect in a myo-inositol 1-phosphate synthase (MIPS) gene as the best candidate for mediating both phenotypes (Hitz et al. 2002). DNA sequence analysis revealed an amino acid substitution mutation in a highly conserved region of the enzyme. Subsequently, in vitro enzyme assays showed the specific activity of the recombinant MIPS protein originating from LR33 to be only ~10% of that observed with the wild type enzyme.

V. Conclusions

Candidate gene approaches have been used in soybean genetics for more than a decade and have become an increasingly popular means of determining the molecular genetic basis of agronomically useful traits. Although success in this area is typically reliant on previously established relationships between known metabolic pathways and their cognate genes, the continual progress that is being made in understanding these relationships in model systems will only further expand the opportunities for applying this information toward gene discovery in soybean. Furthermore, advances in the elucidation of the complete soybean genome (Jackson 2008) will serve to greatly
alleviate the problem of having to identify the complete set of paralogs of a given gene that may serve as viable targets. An obstacle that is likely to remain for at least the near future, however, is the inability to readily transform the soybean plant. Without this ultimate means of validation through complementation, it is possible that situations will arise where a benign polymorphism in a proposed gene candidate will be mistakenly assigned to a phenotype dictated by the true casual gene when the two are closely linked. Despite this limitation, it is likely that candidate gene approaches will continue to play an increasingly important role in the molecular genetic analysis of important agronomic traits in soybean.
VI. References


Cardinal AJ, Dewey RE, Burton JW (2008) Estimating the individual effects of the reduced palmitic acid fap\textsubscript{nc} and fap\textsubscript{l} alleles on agronomic traits in two soybean populations. Crop Sci (in press)


Sandhu D, Alt JL, Scherder CW, Fehr WR, Bhattacharyya MK (2007) Enhanced oleic acid content in the soybean mutant M23 associated with the deletion in the Fad2-1a gene encoding a fatty acid desaturase. J Amer Oil Chem Soc 84:229-235


Chapters II and III of this dissertation describe my efforts in applying a candidate gene approach toward understanding two important seed quality traits in soybean, enhanced stearic acid content in the storage oil, and a reduced phytic acid trait.

A generalized scheme of the biosynthetic pathways relevant to this project is depicted in Fig. 2. Chapter II details an analysis of a newly discovered stearoyl-ACP desaturase gene and mutations in this gene that result in elevating the stearic acid content in soybean seed. Chapter III describes my attempts to identify gene mutations responsible for lowering the phytic acid content of soybean seeds.
Fig. 2 A generalized scheme of stearic acid and phytic acid synthesis in developing soybean seeds. ACCase: acetyl-CoA carboxylase; FAS: fatty acid synthase; KAS: 3-ketoacyl-ACP synthase; SACPD: stearoyl-acyl-carrier-protein-desaturase; FAT: fatty acid thioesterase; ACS: acyl-CoA synthase; PA, phosphatidic acid; DAG, diacylglycerol; PI: Phosphatidylinositol; PIK: phosphatidylinositol kinase; PI-4-P: phosphatidylinositol 4-monophosphate; PI-4,5-P2: Phosphatidylinositol 4,5-bisphosphate; MIPS: myo-inositol 3-phosphate synthase; MIK: myo-inositol kinase; InP: myo-inositol phosphate; IPK1: myo-inositol phosphate kinase I; IPK2: myo-inositol phosphate kinase II. Stearic acid (18.0) and phytic acid (InP_{5}(1,2,3,4,5,6)) are highlighted in grey.
CHAPTER II

Mutations in a Δ⁹-Stearoyl-ACP-Desaturase Gene Are Associated with Enhanced Stearic Acid Levels in Soybean Seeds

Ping Zhang, Joseph W. Burton, Robert G. Upchurch, Edward Whittle, John Shanklin, and Ralph E. Dewey

P. Zhang and R.E. Dewey, Crop Science Department, North Carolina State University, Raleigh, NC, 27695; J.W. Burton and R.G. Upchurch, USDA-ARS, North Carolina State University, Raleigh, NC, 27695; J. Shanklin and E. Whittle, Department of Biology, Brookhaven National Laboratory, Upton, NY, 11973.

This manuscript has been accepted for publication in Crop Science.

I was responsible for all experimental aspects of this work except the in vitro SACPD enzyme assays conducted in the lab of Dr. John Shanklin (using purified recombinant protein that I provided them). Dr. Joseph Burton provided the soybean population segregating for the fas_{nc} locus.
ABSTRACT

Stearic acid (18:0) is typically a minor component of soybean [Glycine max (L.) Merr.] oil, accounting for only 2 to 4% of the total fatty acid content. Increasing stearic acid levels of soybean oil would lead to enhanced oxidative stability, potentially reducing the need for hydrogenation, a process leading to the formation of undesirable trans fatty acids. Although mutagenesis strategies have been successful in developing soybean germplasm with elevated 18:0 levels in the seed oil, the specific gene mutations responsible for this phenotype were not known. We report a newly identified soybean gene, designated SACPD-C, that encodes a unique isoform of Δ⁹-stearoyl-ACP-desaturase, the enzyme responsible for converting stearic acid to oleic acid (18:1). High levels of SACPD-C transcript were only detected in developing seed tissue, suggesting that the encoded desaturase functions to enhance oleic acid biosynthetic capacity as the immature seed is actively engaged in triacylglycerol production and storage. The participation of SACPD-C in storage triacylglycerol synthesis is further supported by the observation of mutations in this gene in two independent sources of elevated 18:0 soybean germplasm, A6 (30% 18:0) and FAM94-41 (9% 18:0). A molecular marker diagnostic for the FAM94-41 SACPD-C gene mutation strictly associates with the elevated 18:0 phenotype in a segregating population, and could thus serve as a useful tool in the development of cultivars with oils possessing enhanced oxidative stability.
INTRODUCTION

For decades the food industry has relied on hydrogenation to increase the oxidative stability of soybean oil. Because this process produces \textit{trans} fatty acids that have been associated with enhanced risk of coronary heart disease (Katan, 1998), developing an oxidatively stable, nonhydrogenated oil is highly desired. The production of solid-fat products, such as margarines, from soybean oil is particularly problematic since extensive hydrogenation is required for these applications. One avenue for achieving the goal of \textit{trans}-free solid fat products is to utilize soybean germplasm that possesses enhanced saturated fatty acid content. Palmitic acid (16:0), the major saturated fatty acid of commercial soybean oil, has desirable oxidative properties and a high melting point, yet is also associated with negative health consequences as a result of its cholestrogenic nature and its ability to promote arterial thrombosis formation (Hu et al., 1997; Hornstra and Kester, 1997). These negative effects, however, are generally not observed with stearic acid (KrisEtherton and Yu, 1997; Thijssen et al., 2005). In addition to enhancing oxidative stability, a high stearate-containing soybean oil may also be welcomed by the confectionery industry since it could be used as an economical cocoa butter equivalent (Chang et al., 1990).

The elevated stearic acid seed trait in soybean is controlled by homozygous recessive genetic loci designated either \textit{fas} or \textit{st} (Pantalone et al., 2004). Five elevated stearate lines
were developed using chemical or X-ray mutagenesis. Soybean germplasm A81-606085 (fas), A6 (fas\textsuperscript{a}), and FA41545 (fas\textsuperscript{b}) display seed oil 18:0 phenotypes of 19%, 30% and 15%, respectively, in contrast to normal soybeans which contain 2 to 4% stearic acid. Genetic studies demonstrated that these three fas loci are allelic, and therefore likely represent independent mutations within the same gene (Graef et al., 1985). Germplasm KK-2 (st\textsubscript{1}) and M25 (st\textsubscript{2}) possess non-allelic loci that when combined (st\textsubscript{1} st\textsubscript{1} st\textsubscript{2} st\textsubscript{2}) elevate 18:0 levels to over 35% of the fatty acid content of the seed oil (Rahman et al., 1997). It is not known whether st\textsubscript{1} or st\textsubscript{2} are allelic to the fas loci. In addition to the above mentioned soybean lines displaying elevated stearic acid phenotypes derived via mutagenesis, a naturally occurring source has been identified. Germplasm line FAM94-41 carries a natural mutation at a locus designated fas\textsubscript{nc} that results in a 9% 18:0 phenotype in the seed (Pantalone et al., 2002). Despite the considerable differences in the 18:0 accumulation profiles between FAM94-41 and A6 (9% versus 30%), genetic analysis demonstrated that fas\textsubscript{nc} and fas\textsuperscript{a} are allelic (Pantalone et al., 2002).

Although no experimental evidence has been presented that describes the molecular basis of the elevated stearic acid trait in any soybean line, studies using other plant systems have shown the Δ\textsuperscript{9}-stearyl-acyl carrier protein desaturase (Δ\textsuperscript{9} S-ACP-Des) enzyme to be pivotal in determining stearic acid accumulation (Knutzon et al., 1992; Zaborowska et al., 2002). Δ\textsuperscript{9} S-ACP-Des in plants is a plastid-localized soluble desaturase that catalyzes the conversion of stearic acid into the monounsaturated oleic acid (18:1), which in turn
can serve as a precursor in the synthesis of polyunsaturated fatty acids (18:2 and 18:3) (Shanklin and Cahoon, 1998). Therefore, this enzyme plays a key role in the processes that determine the physical properties of the majority of cellular glycerolipids; those utilized as membrane constituents as well as the storage triacylglycerols.

\[ \Delta^9 \text{- S-ACP-Des} \]

genes have been cloned and characterized from a variety of plant species (Shanklin et al., 1991; Thompson et al., 1991; Cahoon et al., 1998; Whittle et al., 2005; Tong et al., 2006; Kachroo et al., 2007). Recently two soybean \[ \Delta^9 \text{- S-ACP-Des} \] genes, designated \textit{SACPD-A} and \textit{SACPD-B}, have also been reported (Byfield et al., 2006; Byfield and Upchurch, 2007). Both \[ \Delta^9 \text{- S-ACP-Des} \] isoforms were detected in all soybean lines surveyed, including the high stearic acid line A6, but no correspondence was reported between these genes and the elevated stearic acid trait. Here we present a previously unidentified soybean \[ \Delta^9 \text{- S-ACP-Des} \] gene, designated \textit{SACPD-C}, that is specifically expressed during seed development. Defects in the \textit{SACPD-C} gene were observed in the elevated 18:0 soybean lines A6 and FAM94-41. These results define \textit{SACPD-C} as a key target in the development of soybean cultivars containing an elevated stearic acid oil trait.

**MATERIALS AND METHODS**
**Plant Materials**

Soybean lines 'Dare', A6, FA8077 and FAM94-41 were used in this study. Dare (FasFas) is a normal cultivar that accumulates typical levels of stearic acid in the mature seed (3 - 4%). The elevated stearate germplasm A6 (fas'fas') was generated by sodium azide treatment of breeding line FA8077 and contains greater than six-fold higher levels of 18:0 than normal lines (Hammond and Fehr, 1983). FAM94-41(fasncfasnc) possesses a natural mutation that confers 2 to 3-fold higher levels of stearic acid than normal lines (Pantalone et al., 2002). For the fasnc segregation analysis, F2 seeds from a cross between FAM94-41 and the mid-oleic acid breeding line N98-4445A (Burton et al., 2006) were used.

**Gene Isolation and Analysis**

Total cellular RNA was isolated from frozen developing seeds (mid-stage) of soybean cultivar Dare using the Trizol reagent according to the manufacturer's protocol (Invitrogen, Carlsbad, CA). First-strand cDNA was generated from total RNA using the SuperScript® First-Strand Synthesis System for RT-PCR (Invitrogen). Genomic DNA was extracted using young leaf tissue as previously described (Dewey et al., 1994). To clone both the full-length SACPD-C cDNA and its genomic version, primers 5’-ATCTCCACCTCTCCACAGTTC-3’ and 5’-AGTCACAAAGCCAAAAACCTG-3’, based on the tentative contig TC205834 (www.tigr.org), were utilized in amplification reactions using the Expand Hi-fidelity PCR system (Roche Applied Science, Indianapolis,
IN). Each 50-µl reaction contained 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 200 µM each dNTP, 25 pmol of each primer, 1 unit of Taq DNA polymerase, and 1 to 2 µl of the final first-strand product reaction for cDNA templates, and 50 – 100 ng when genomic DNA was used as template. Each reaction was conducted using the following thermocycling parameters: 94°C incubation for 2 min, followed by 30 cycles of 94°C for 30 s, 60°C for 45 s and 72°C for 1 min (for cDNA; 2 min for genomic DNA), and a final 72°C extension for 7 min. The polymerase chain reaction (PCR) products were cloned into the pCR 2.1-TOPO vector using the TOPO TA Cloning kit (Invitrogen). DNA sequence was obtained using the Iowa State University DNA Facility (www.dna.iastate.edu). Sequence alignments were conducted using bioinformatics programs available at http://align.bmr.kyushu-u.ac.jp/mafft/online/server/. Analysis of the translated open reading frame with the ChloroP program (http://www.cbs.dtu.dk/services/ChloroP/) lead to the prediction of an N-terminal chloroplast transit peptide as shown in Fig. 1A. During the course of this study it became apparent that the GenBank entries for the soybean SACPD-A and SACPD-B cDNA and predicted protein sequences (AY885234 and AY885233, respectively) contained some errors. These entries have been edited and corrected accordingly.

**DNA and RNA Blot Analysis**

Ten micrograms of genomic DNA from soybean genotypes Dare, A6 and FA8077 were digested with restriction endonuclease Hind III (Promega, Madison, WI) for 3 hours.
Digested DNA was separated on a 0.8% agarose gel and blotted to a nylon membrane as described (Sambrook and Russell, 2001). A $^{32}$P-labeled SACPD-C cDNA probe (4.4 x 10$^{10}$ cpm/E, Random Primed DNA Labeling Kit from Roche Applied Science, Indianapolis, IN) was hybridized to the DNA blot and washed according to standard protocols (Sambrook and Russell, 2001). Radioactive hybridization signals were detected by exposing the blot to BioMax X-ray film (Eastman Kodak, Rochester, NY). Total cellular RNA was isolated from leaf, flower, root, stem and 35 days after flowering (DAF) developing seed tissue using the Trizol reagent as outlined by the manufacturer (Invitrogen). RNA blots were generated by separating 10 $\mu$g of total cellular RNA on 1.5% agarose-formaldehyde denaturing gels (Sambrook and Russell, 2001). Blotting, hybridization and wash conditions were the same as described for DNA blots.

**Molecular Marker Analysis**

SACPD-C genomic DNAs were isolated from the progeny (seeds) of an F2 population derived from a FAM94-41 X N98-4445A cross using the FastDNA Kit (MP Biomedicals, Solon, OH). Primers 5’-CCACAAAAACTTCCCTCCCTGA-3’ (forward) and 5’-GAAATGCATACTCACCCAATCA-3’ (reverse) were designed to amplify a 432 bp fragment that encompasses the single nucleotide polymorphism found in the SACPD-C gene from FAM94-41 that alters codon 126. PCR components and conditions were the same as described above for cDNA amplifications except that 60°C was used as the annealing temperature. Eight microliters of amplified product was digested for 2 hrs at 37
°C with the restriction endonuclease Hga I (New England Biolabs, Beverly, MA) in 20 µl reaction volumes containing 4 units of enzyme and the recommended buffer. Digested DNA fragments were separated on 1.8% agarose gels stained with ethidium bromide and visualized using a Gel Logic 100 Imaging System (Eastman Kodak).

**Activity Analysis of Recombinant SACPD Enzymes**

To obtain mature SACPD-C enzyme suitable for expression in *E. coli* (without the predicted transit peptide), the following sense and antisense primers were used in PCR reactions with cloned full length cDNA as template: 5'-TTTCATATGGCGGCGCGCGTC-3' and 5'-TTTTGTCGACCCTGTTGACTGACGTTG-3'. Primers were designed to include NdeI and SalI restriction sites (no other amino acids were added between start codon and the target protein sequence) to facilitate cloning of the cleaved amplification product into the comparable restriction sites of the pET-24b expression vector (Novagen, San Diego, CA). Plasmids were transformed into host BL21 DE3 and expression was induced as directed by the manufacturer (Novagen). Cells were harvested by centrifugation and resuspended in 2.5 ml of extraction buffer (7 mM Hepes, 7 mM Mes, 7 mM sodium acetate, 6 units per ml of DNase I, pH 7.4) per gram of cells. Cells were lysed by sonication (10 bursts of 10 seconds each with 30 seconds intervals between each burst using a Fisher Sonic Dismembrator Model 100 at setting 2) and clarified by centrifugation at 48,000 X g for 30 min at 4°C. The supernatant was passaged onto a 1.8 ml column of CM-sephadex, C-
50 (Pharmacia LKB, Piscataway, NJ) that had been pre-equilibrated with 7 mM Hepes, 7 mM Mes, 7 mM sodium acetate, pH 7.4. The column was subsequently washed with 16 ml of equilibration buffer followed by elution with a 60 - 600 mM linear gradient of NaCl (in equilibration buffer). Fractions with the highest concentrations of recombinant protein (as judged by Coomassie Brilliant Blue stained SDS-PAGE gel) were concentrated using Centricon YM-30 columns according to the manufacturer's protocol (Millipore Corp. Bedford, MA). Purity of protein was estimated to be >90% by Coomassie Brilliant Blue stained SDS-PAGE. Desaturase preparations were assayed with [1-14C] 18:0-ACP substrate with the use of recombinant spinach ACP-I and *Anabaena* vegetative ferredoxin as described in a previous report (Cahoon et al. 1997). Methyl esters of fatty acids were analyzed by argentation TLC and radioactivity in products quantified (Cahoon et al. 1997); assays were performed with three or more replicates.

**RESULTS**

**Identification of a New Soybean Δ9-Stearoyl-ACP-Desaturase Gene**

In an attempt to characterize the number of independent a Δ⁹-S-ACP-Des genes that are actively expressed in soybean, we conducted TBLASTN alignments of the soybean ESTs deposited in GenBank using known plant Δ⁹-S-ACP-Des protein sequences as the query. The results of this analysis enabled us to define three discrete classes of a Δ⁹-S-ACP-Des-
like ESTs (data not shown). Two classes corresponded to the recently characterized, and
closely related, $SACPD-A$ and $SACPD-B$ genes (Byfield et al., 2006). The third class
formed a unique group that is represented in the TIGR Soybean Gene Index as tentative
contig TC205834 (www.tigr.org). Interestingly, 25 of the 27 ESTs that comprise
TC205834 originated from libraries generated from developing seed tissue or somatic
embryos (cultured cells that mimic soybean seed development). In contrast, the ESTs
corresponding to $SACPD-A$ or $SACPD-B$ were obtained from cDNA libraries
representing a diversity of plant tissues and/or stages of development (data not shown).

To obtain a full-length cDNA clone of this new $\Delta^9$-S-ACP-Des gene, which we
designated $SACPD-C$, PCR amplifications were conducting using primers based on the
putative 5’- and 3’- untranslated regions of TC205834. As shown in Fig. 1, $SACPD-C$ is
predicted to encode a 386-amino acid protein that shares approximately 63% sequence
identity with both of the 391-amino acid long SACPD-A and SACPD-B proteins.
Similar to other plant $\Delta^9$-S-ACP-Des enzymes, SACPD-C possesses a predicted N-
terminal transit peptide to facilitate transport into the plastid (Fig. 1A). The SACPD-C
protein sequence also contains the structural motifs typical of soluble acyl-ACP-
desaturases, including a di-iron center buried within two pairs of anti-parallel helices
(data not shown).
The same PCR primers used to amplify the full-length *SACP-D-C* cDNA were also successful in recovering the corresponding genomic sequence of the gene. *SACP-D-C* is comprised of two exons separated by an 846 bp intron (Fig. 1B). Structurally, *SACP-D-C* differs from the *SACP-D-A* and *SACP-D-B* genes in that it lacks the large intron located immediately after the putative transit peptide-encoding region (Fig. 1B). The sole *SACP-D-C* intron is located at the same position as *SACP-D-A/B* intron 2 (Fig. 1A), but is twice the size.

**Spatial Distribution of SACP-D-C Transcripts**

RNA blot analysis was conducted to determine the relative abundance of *SACP-D-C* transcripts in developing seeds (35 DAF), young leaves, flowers, roots, and stems from soybean cultivar Dare. As shown in Fig. 2, a strong hybridization signal was observed with RNAs isolated from developing seeds. In contrast, negligible signal was observed using RNA preparations from the other tissues tested. These results suggest that *SACP-D-C* is under the regulatory control of a highly seed-specific promoter. The RNA blotting results are also consistent with the above mentioned observation that the preponderance of the *SACP-D-C* ESTs found in GenBank were derived from developing seed (and somatic embryo) cDNA libraries.

*SACP-D-C* is Deleted in Mutant Line A6
Given the seed specificity of its expression profile, we speculated that \(SACPD-C\) functions to enhance the 18:0-ACP desaturation capacity of the immature seed to accommodate the great increase in fatty acid biosynthesis that occurs with the production of storage triacylglycerol reserves. As such, \(SACPD-C\) would represent a viable candidate gene for mutant soybean germplasm displaying an elevated 18:0 phenotype. The \(fas^a\) locus found in the mutant line A6 confers the highest 18:0 seed phenotype (30% total fatty acid content) of any single soybean locus characterized to date. To determine whether mutations in the \(SACPD-C\) gene may be responsible for this phenotype, we initially attempted to amplify \(SACPD-C\) from A6 using developing seed cDNA as template. Using the same primers and conditions that were successful in recovering the \(SACPD-C\) cDNA from cultivar Dare, we were unable to detect the expected product from A6 (Fig. 3A). Occasionally, a faint band of much smaller size is observed when amplifying A6 cDNA, the nature of which is unknown. Consistent with the PCR results, RNA blot analysis showed no detectable hybridization to 35 DAF developing seed RNA isolated from A6 when radiolabeled \(SACPD-C\) was used as a probe (Fig. 3B). To confirm that the absence of \(SACPD-C\) gene expression was truly the result of mutagenesis treatment, as opposed to natural cultivar variation, RNAs from breeding line FA8077, the parental line that was mutated to produce A6 (Hammond and Fehr, 1983), were included in this study. As shown in Fig. 3B, similar levels of \(SACPD-C\) transcripts as in Dare and FAM94-41 are observed in genotype FA8077.
To test whether our failure to detect evidence of \textit{SACPD-C} gene expression in A6 could be explained by a gross change in \textit{SACPD-C} gene structure within the A6 genome (such as a deletion, insertion or rearrangement), Southern blotting experiments were conducted using genomic DNAs isolated from Dare, A6 and FA8077. Regardless of the restriction enzyme used in these analyses, the most intensely hybridizing band (or bands) was always missing from genomic DNA preparations of A6 in comparison with Dare or FA8077. A typical example is shown in Fig. 3C using the restriction enzyme \textit{Hind} III. Although additional, cross-hybridizing bands were shared among all three genotypes, they appeared not to be contributing measurably to the transcript pool in A6 (Fig 3B). Cumulatively, the RNA and DNA blotting results present compelling evidence that the \textit{SACPD-C} gene has been deleted in the A6 mutant germplasm.

\textbf{A Radical Amino Acid Substitution Is Found in the \textit{SACPD-C} Gene of FAM94-41.}

Although the 18:0 phenotypes of the A6 and FAM94-41 germplasm lines differ substantially (30% versus 9% 18:0), allelism studies led to the prediction that the elevated stearate content in both lines was due to independent mutations within the same gene (Pantalone et al., 2002). Because we observed no difficulties in amplifying the expected \textit{SACPD-C} product from FAM94-41, and transcript accumulation levels appeared normal (Fig. 3A and B), DNA sequence analysis was conducted to determine whether more subtle differences were apparent in the FAM94-41 \textit{SACPĐ-C} gene. Sequence comparisons revealed two polymorphisms between the \textit{SACPD-C} genes from normal
cultivar Dare versus FAM94-41. One polymorphism, a G → A substitution at position 1079 (with respect to the SACPD-C start codon) does not alter the predicted amino acid sequence. The other polymorphism, however, results in a nonconserved amino acid substitution in a critical region of the enzyme. As shown in Fig. 4, a G → A polymorphism at position 376 of the cDNA changes an Asp residue to an Asn at position 126 of the predicted amino acid sequence. Alignment of the FAM94-41 protein sequence to the nonredundant protein database of GenBank revealed that this position is highly conserved among plant Δ⁹-S-ACP-Des proteins. The acidic residues Asp or Glu are found at the comparable location in every other Δ⁹-S-ACP-Des GenBank entry (data not shown). Furthermore, when the SACPD-C polypeptide sequence is superimposed on the three-dimensional structure that has been determined for the castor bean Δ⁹-S-ACP-Des (Lindqvist et al., 1996), Asp¹²⁶ is located within one of the two pairs of anti-parallel helices that comprise the catalytic di-iron center of the enzyme, lying immediately adjacent to an invariant Glu residue that serves as an iron binding ligand (Glu¹⁰⁵ of the castor enzyme; Glu¹²⁵ of SACPD-C).

**Association Analysis of the FAM94-41 SACPD-C Variant**

If the Asp¹²⁶ to Asn¹²⁶ polymorphism is indeed responsible for the enhanced 18:0 phenotype of FAM94-41, a perfect correlation should be observed between the elevated stearic acid trait defined by fasₘₐ and the nucleotide polymorphism responsible for the amino acid substitution in SACPD-C. The single nucleotide change at position 376
observed between the \textit{SACPD-C} gene from FAM94-41 versus normal soybean results in
the loss of a \textit{Hga} I restriction endonuclease site in the former (Fig. 4A). This
polymorphism allowed us to develop a cleavage amplified polymorphic sequence (CAPS)
marker (Konieczny and Ausubel, 1993) that can readily distinguish the two alleles. PCR
primers were designed that amplified a 432 bp region encompassing the polymorphic
restriction site. \textit{Hga} I digestion of the 432 bp \textit{SACPD-C} amplification product from
normal soybeans yields fragments of 295 and 137 bp, in contrast to the product derived
from FAM94-41 which remains uncut.

Segregation analysis was conducted on an F\textsubscript{2} population of a cross between FAM94-41
and N98-4445A, a mid-oleic acid germplasm line (Burton et al., 2006). Interestingly,
introducing the \textit{fasnc} locus into the mid-oleic background of N98-4445A resulted in
higher stearic acid levels (13 to 15\% total fatty acid content) than those observed in its
original background (9\%; data not shown). Three discrete classes of stearic acid
phenotypes were observed when fatty acid analysis was conducted on bulked seed from
each F\textsubscript{2} plant: wild type (2 to 4\% 18:0), high (12 to 15\%) and midpoint (7 to 9\%). Seed
from 40 F\textsubscript{2} plants displaying wild type 18:0 levels, and 47 plants with a high 18:0 content
were genotyped using the \textit{SACPD-C} CAPS marker. All lines classified as having a
normal 18:0 seed phenotype were homozygous for the wild type \textit{SACPD-C} allele, and
100\% of the F\textsubscript{2} lines displaying an elevated 18:0 phenotype were homozygous for the
\textit{SACPD-C} allele derived from FAM94-41. Representative results are shown in Fig. 5. We
predicted that the two *SACPD-C* alleles would be segregating in seed bulks from F$_2$ plants possessing a midpoint 18:0 phenotype. To test this, ten seed from each of three independent midpoint 18:0 lines were individually genotyped with the *SACPD-C* CAPS marker. As expected, representatives of all three possible marker genotypes were observed in each of the midpoint lines tested. Examples from one line are shown in Fig. 5. The precise co-segregation of the FAM94-41-derived *SACPD-C* marker genotype with the elevated 18:0 trait among the FAM94-41 X N98-4445A F$_2$ lines tested strongly supports the hypothesis that the aberrant *SACPD-C* allele defines the *fas$_{nc}$* locus.

**In Vitro Analysis of SACPD-C Enzyme Activity**

Given the occurrence of the Asn1266j6 substitution in the SACPD-C gene in FAM94-41 at a location where an acidic residue normally resides, we speculated that the resultant enzyme (SACPD-C$_{FAM}$) may be less active than the wild type enzyme (SACPD-C$_{WT}$) and thus provide a rational explanation of the elevated 18:0 phenotype observed in this line. To test the relative activities of SACPD-C$_{FAM}$ versus SACPD-C$_{WT}$, each enzyme was expressed in E. coli, purified, and assayed in the presence of 14C-labeled 18:0-ACP substrate. The recombinant Δ9-S-ACP-Des enzyme from castor bean was used as a positive control. Surprisingly, both soybean enzymes were far less active than the castor enzyme in catalyzing 18:0-ACP desaturation in vitro (Fig. 6). Equally unexpected, the mutant SACPD-C$_{FAM}$ enzyme consistently displayed ~10-fold higher activity than the wild type soybean enzyme in these assays.
DISCUSSION

The discovery and characterization of the seed-specific SACPD-C gene provides new insights regarding the mechanisms by which the stearic acid content of soybean seeds is controlled. The finding that SACPD-C is deleted in the A6 germplasm, and a mutant version of the gene from FAM94-41 faithfully segregates with the high 18:0 phenotype, provides compelling evidence that the encoded enzyme plays a major role in the biosynthesis of oleic acid during soybean seed development. Therefore, the observation that the specific activity of the purified, recombinant wild type SACPD-C enzyme was over 500-fold less than the prototypical castor enzyme (Fig. 6) was unexpected. Several independent replications of the in vitro assay-based experiments invariably yielded the same results, causing us to conclude that SACPD-C is inherently less active than the castor enzyme or FAB2, the predominant Δ9-S-ACP-Des enzyme of Arabidopsis (Kachroo et al., 2007) under the conditions of this assay. Two studies from the laboratory of J. B. Ohlrogge (Michigan State University) have shown that in vitro Δ9-S-ACP-Des enzyme activity can vary depending on the specific isoforms of ferredoxin (Schultz et al., 2000) or ACP (Suh et al., 1999) that are used as the electron-donating cofactor and in the 18:0-ACP substrate, respectively. It is possible that SACPD-C does not interact well with the Anabaena-derived ferredoxin and/or spinach ACP that were used in this study. Alternatively, SACPD-C may require a post-translational modification that is not
recapitulated during synthesis in the E. coli-based expression system, such as phosphorylation, in order to attain full activity. The fact that SACPD-C shares less than 65% amino acid sequence homology with SACPD-A or -B, the Arabidopsis FAB2 or the castor enzyme, whereas the latter four enzymes all share over 82% identity with each other (Fig. 1) lends plausibility to the speculation that SACPD-C is regulated differently than typical Δ9-S-ACP-Des enzymes. Finally, the "mature" recombinant SACPD-C enzyme used in the in vitro assays was engineered based on removal of the predicted chloroplast transit peptide shown in Fig. 1A. If this predicted cleavage site does not truly reflect the in vivo N-terminal maturation of the protein, failure to perfectly reconstitute the correct mature protein could provide another explanation for reduced overall activity.

Equally paradoxical is the observation that SACPD-C_{FAM} showed higher activity than SACPD-C_{WT} in the in vitro enzyme assays. The proximity of the Asp126 to Asn126 substitution found in SACPD-C_{FAM} to the predicted iron-binding Glu125 residue, together with the observation that only acidic residues (Glu or Asp) are found at the comparable position in every other soluble Δ9-S-ACP-Des enzyme sequence deposited in GenBank caused us to predict that SACPD-C_{FAM} enzyme activity would be reduced relative to the wild type. Instead, SACPD-C_{FAM} proved to be ~10-fold more active than SACPD-C_{WT} in the in vitro assay (Fig. 6). One plausible explanation for this phenomenon is that although the Asp126 to Asn126 substitution may result in greater inherent enzyme activity in vitro, it may also lead to greatly reduced stability in vivo. Regardless, the unexpected findings
revealed in this study highlight the fact that there remains much to be learned concerning the structure:function relationship of Δ9-S-ACP-Des enzymes in plants and the mechanisms by which they are regulated.

An interesting pattern is beginning to emerge concerning the organ specificity of genes associated with glycerolipid biosynthesis in soybean. Previously, both seed-specific and non-seed specific isoforms have been identified for genes encoding the endoplasmic reticulum-localized FAD2 (18:1 desaturase) and FAD3 (18:2 desaturase) enzymes in soybean (Heppard et al., 1996; Bilyeu et al., 2003; Tang et al., 2005). With our characterization of SACPD-C, it is apparent that the soybean genome possesses seed-specific isoforms for all three of the desaturase enzymes expected to be involved in synthesis of storage triacylglycerols. A reasonable explanation for the occurrence of seed-specific isoforms of these genes would be to assist in accommodating the great increase in lipid biosynthesis that occurs as the developing soybean seed is producing storage oil reserves. However, this phenomenon does not appear to be universal for all steps of the glycerolipid pathway that are required for triacylglycerol production, as no seed-specific isoforms were found among the four 16:0-ACP thioesterase genes or the two 3-keto-acyl-ACP synthase II genes that we have previously characterized from soybean (Aghoram et al., 2006; Cardinal et al., 2007).
The $fas^a$ locus of germplasm A6 can confer a stearic acid seed phenotype of ~30% total fatty acid; however, this locus is also associated with a severe depression in yield (Hartmann et al., 1997). In addition to the possibility that the high 18:0 seed phenotype per se is responsible for the observed yield drag, it is plausible that the deletion event that gave rise to A6 may have also encompassed other genes on the same chromosomal fragment that could impact yield. Although the gains in 18:0 content mediated by $fas_{nc}$ are more modest (13 to 15% in this study) than those attributed to $fas^a$, no yield penalty has been associated with $fas_{nc}$ (Pantalone et al., 2002). Introducing the elevated 18:0 phenotype of the $fas_{nc}$ locus into an elevated oleic acid, low linolenic acid germplasm, such as N98-4445A, represents a promising strategy for the development of soybean lines with greatly enhanced oxidative stability. The $fas_{nc}$-specific CAPS marker developed in this study could serve as a useful tool in achieving this goal by accelerating the introgression of the elevated 18:0 trait into desired backgrounds through marker-assisted selection.
ACKNOWLEDGEMENTS

We thank Jeff Gillikin for assistance with SACPD-C protein purification. This work was supported by a grant from the United Soybean Board. E. Whittle and J. Shanklin were supported through the Office of Basic Energy Sciences of the United States Department of Energy.


Cahoon, E. B., Y. Lindqvist, G. Schneider, and J. Shanklin. 1997. Redesign of soluble...


FIGURE LEGENDS

**Figure 1.** Gene structure and predicted protein sequence of the \textit{SACPD-C} gene of soybean. (A) Alignment of the predicted SACPD-C protein (GenBank \#EF113911) with $\Delta^9$-S-ACP-Des proteins from soybean (SACPD-A and -B, GenBank entries AY885234 and AY885233, respectively), Castor (GenBank \# M59857) and Arabidopsis (GenBank \# NP_850400). Residues identical among at least three of the five entries are box shaded. The putative SACPD-C transit peptide that was predicted using the program ChloroP (www.cbs.dtu.dk/services/ChloroP) is underlined. Arrow indicates the location of \textit{SACPD-C} intron (intron 2 of \textit{SACPD-A/B}). (B) Comparison of \textit{SACPD-C} gene structure with the \textit{SACPD-A} and -B isoforms.

**Figure 2.** RNA gel blot of \textit{SACPD-C} transcripts in different soybean tissues. Lower panel shows ethidium bromide staining of the gel prior to blotting to show the relative equivalency of RNA loading.

**Figure 3.** Analysis of \textit{SACPD-C} in A6 and FAM94-41. (A) Reverse transcriptase-PCR amplification of \textit{SACPD-C} using RNAs isolated from 35 DAF developing seeds of Dare, A6 and FAM94-41. Molecular weight markers are shown in the far left lane. (B) RNA blot analysis of \textit{SACPD-C} transcripts from developing seeds of Dare, FA8077, A6 and FAM94-41. Ethidium bromide stained gel is shown in lower panel as loading control. (C)
DNA blot analysis of Hind III digested genomic DNAs from A6, Dare and FA8077 using radiolabeled SACP-D-C cDNA as the hybridization probe.

**Figure 4.** Point mutation in the SACP-D-C gene of FAM94-41. (A) Sequence comparison of the SACP-D-C genes from Dare (GenBank #EF113911) and FAM94-41 (GenBank #EF113912) in the region surrounding the G→A polymorphism at position 376 (with respect to initiation codon of SACP-D-C cDNA). A rectangle denotes the recognition motif of the restriction enzyme Hga I that is absent in the FAM94-41 derived sequence. (B) Localized amino acid sequence comparison showing the Asp\textsuperscript{126} to Asn\textsuperscript{126} substitution in the predicted SACPD-C enzyme from FAM94-41. Amino acid positions are in reference to the initiator start Met residue. The predicted iron-binding Glu residue at position 125 is underlined.

**Figure 5.** CAPS marker analysis of representative F\textsubscript{2} progeny from a FAM94-41 X N98-4445A-derived population. SACP-D-C-specific amplification products from select individuals of F\textsubscript{2} families displaying wild type (WT, lanes 1 and 2), high (lanes 3 and 4) and midpoint (Mid, lanes 5 to 7) 18:0 phenotypes were digested with Hga I and separated on an agarose gel. M, 100 bp molecular weight ladder.

**Figure 6.** In vitro synthesis of oleic acid using recombinant Δ9-S-ACP-Des proteins. Activities are shown as nmol product/min/mg recombinant protein (from preparations
>90% pure). Rc, Ricinum communis (castor). Error bars represent standard error (n=3 for castor enzyme; n=5 for SACPD-CWT; n=4 for SACPD-C_FAM).
Figure 1. Gene structure and predicted protein sequence of the SACP-D-C gene of soybean. (A) Alignment of the predicted SACP-D-C protein (GenBank #EF113911) with Δ⁹-S-ACP-Des proteins from soybean (SACP-D-A and -B, GenBank entries AY885234 and AY885233, respectively), Castor (GenBank # M59857) and Arabidopsis (GenBank # NP_850400). Residues identical among at least three of the five entries are box shaded. The putative SACP-D-C transit peptide that was predicted using the program ChloroP (www.cbs.dtu.dk/services/ChloroP) is underlined. Arrow indicates the location of SACP-D-C intron (intron 2 of SACP-D-A/B). (B) Comparison of SACP-D-C gene structure with the SACP-D-A and -B isoforms.
A

SACPD-A : 
SACPD-B : 
Castor : 
At FAB2 : 
SALFD-C : 

B

SACPD-A/B

SACPD-C

Exon 1

Intron 1

111bp

1763 bp

504 bp

423 bp

561 bp

592 bp

846 bp

566 bp
**Figure 2.** RNA gel blot of *SACPD-C* transcripts in different soybean tissues. Lower panel shows ethidium bromide staining of the gel prior to blotting to show the relative equivalency of RNA loading.
Figure 3. Analysis of *SACPD-C* in A6 and FAM94-41. (A) Reverse transcriptase-PCR amplification of *SACPD-C* using RNAs isolated from 35 DAF developing seeds of Dare, A6 and FAM94-41. Molecular weight markers are shown in the far left lane. (B) RNA blot analysis of *SACPD-C* transcripts from developing seeds of Dare, FA8077, A6 and FAM94-41. Ethidium bromide stained gel is shown in lower panel as loading control. (C) DNA blot analysis of *Hind* III digested genomic DNAs from A6, Dare and FA8077 using radiolabeled *SACPD-C* cDNA as the hybridization probe.
**Figure 4.** Point mutation in the *SACPD-C* gene of FAM94-41. (A) Sequence comparison of the *SACPD-C* genes from Dare (GenBank #EF113911) and FAM94-41 (GenBank #EF113912) in the region surrounding the G→A polymorphism at position 376 (with respect to initiation codon of *SACPD-C* cDNA). A rectangle denotes the recognition motif of the restriction enzyme *Hga* I that is absent in the FAM94-41 derived sequence. (B) Localized amino acid sequence comparison showing the Asp$^{126}$ to Asn$^{126}$ substitution in the predicted SACPD-C enzyme from FAM94-41. Amino acid positions are in reference to the initiator start Met residue. The predicted iron-binding Glu residue at position 125 is underlined.
Figure 5. CAPS marker analysis of representative F₂ progeny from a FAM94-41 X N98-4445A-derived population. **SACPD-C**-specific amplification products from select individuals of F₂ families displaying wild type (WT, lanes 1 and 2), high (lanes 3 and 4) and midpoint (Mid, lanes 5 to 7) 18:0 phenotypes were digested with **Hga** I and separated on an agarose gel. **M**, 100 bp molecular weight ladder.
**Figure 6.** In vitro synthesis of oleic acid using recombinant $\Delta^9$-ACP-Des proteins.

Activities are shown as nmol product/min/mg recombinant protein (from preparations >90% pure). Re, Ricinum communis (castor). Error bars represent standard error (n=3 for castor enzyme; n=5 for SACPD-CWT; n=4 for SACPD-CFAM).
CHAPTER III

Candidate Gene Analysis of a Low Phytic Acid Soybean Germplasm
INTRODUCTION

The structural and functional complexities of inositol phosphates in living organisms have fascinated biologists since the 1980s when one of the D-myo-inositol phosphates, inositol 1, 4, 5-trisphosphate, was first shown to serve as a signaling molecule in the regulation of cellular Ca^{2+} levels (Streb et al. 1983). Subsequently, the mechanisms of synthetic pathways of inositol phosphates have been gradually clarified and a wide range of cellular functions have been assigned to various inositol phosphates (Irvine and Schell 2001; Laxalt and Munnik 2002; York 2006; Mulugu et al. 2007; Lee et al. 2007; Huang et al. 2007; Alcázar-Román and Wente 2008). It is expected that studies of inositol phosphates will establish a platform leading to a diversity of practical applications, many related to human health.

In agriculture, myo-inositol 1, 2, 3, 4, 5, 6-hexakisphosphate (IP_{6}, phytic acid) is of particular interest to scientists who are focusing on the enhancement of crop quality as it relates to both animal nutrition and environmental concerns. In addition to its role as a signaling molecule in basic cellular processes, phytic acid serves as a storage reservoir of phosphorous (P_{i}) in the seeds to assist in germination and young seedling growth. However, high levels of phytic acid in grains have actually been considered as an anti-nutritive factor for non-ruminant animals (the majority of livestock) and a contributor
toward phosphorous pollution in soils and groundwater (Raboy 2001; Raboy 2003). Because non-ruminant animals don’t have the digestive enzyme phytase that can hydrolyze phytic acid, they not only lack access to the beneficial phosphorous stores of the seed, but can also suffer from deficiencies in mineral cations (such as iron and zinc) that become tightly chelated to phytic acid. From an environmental perspective, undigested phytic acid excreted from livestock can be leached into watersheds and become a major source of phosphorous pollution in aquatic ecosystems. Several approaches are currently being used to address these problems, including the addition of phytase and/or inorganic phosphorus to animal feeds, a process that increases extra labor and costs without completely solving the problem. Alternatively, developing low phytic acid cultivars of important seed crops has been promoted as an ideal strategy toward meeting both animal nutrition and environment concerns (Raboy 2001). And identifying the specific genes that are responsible for a low phytic acid phenotype in a mutant grain crops would serve two purposes: 1) enhance our understanding of the cellular mechanisms by which phytic acid is synthesized and stored; and 2) enable the development of molecular markers that could be used to introgress the trait into elite cultivars.

Two major branches of the phytic acid biosynthetic pathway have been proposed to exist in plants based on studies mainly from yeast and animals, as well as a number of limited studies specific to plants (Fig.1). Lipid-dependent (beginning with phosphatidylinositol)
and -independent (beginning with inositol-3-monophosphate) pathways all putatively involve multiple sequential phosphorylation steps catalyzed by their corresponding kinases, with ATP as the Pi donor. Key enzymes that have been characterized in plants include myo-inositol 3-phosphate synthase (MIPS, Larson and Raboy 1999; Yoshida et al. 1999; Hegeman et al. 2001; Hitz et al. 2002; Pilu et al. 2003; Shukla et al. 2004; Chappell et al. 2006; Abreu and Aragao 2007), myo-inositol kinase (MIK, Shi et al. 2005), inositol 1, 4, 5-trisphosphate 6-/3- kinase (IPK2, Stevenson-Paulik et al. 2002; Xia et al. 2003), inositol 1, 3, 4, 5, 6-pentakisphosphate 2-kinase (IPK1, Phillippy et al. 1994; Stevenson-Paulik et al. 2005; Sweetman et al. 2006; Sun et al. 2007), inositol 1, 3, 4-trisphosphate 5/6-kinase (Wilson and Majerus 1997; Phillippy 1998; Shi et al. 2003). Most recently mutations in an ATP-binding cassette (ABC) transporter gene were shown to confer a low phytic acid phenotype in maize even though the gene specific function within the pathway has not been elucidated (Shi et al. 2007).

Mutant lines displaying a low phytic acid trait have been generated in several important crops including barley (Larson et al. 1998; Rasmussen and Hatzack 1998; Dorsch et al. 2003), maize (Raboy et al. 2000; Shi et al. 2003; Pilu et al. 2003; Shi et al. 2007), rice (Larson et al. 2000; Liu et al. 2007), and wheat (Guttieri et al. 2004). In soybean, low phytic acid cultivars have been developed by independent research groups. A point-mutation in a MIPS gene leading to decreased catalytic ability of the corresponding enzyme was shown to be responsible for the low phytic acid trait in line LR33 (Hitz et al. 2002).
A more recent report also described a 2 bp deletion in the same MIPS gene from a low phytic acid soybean line designated *Gm-lpa*-TW-1 (Yuan et al. 2007). Heritability studies and genetic mapping of a low phytic acid trait originating from an EMS-generated event named M153 (Wilcox et al. 2000) revealed two distinct recessive loci governing the mutant phenotype (Oltmans et al. 2004; Walker et al. 2006). The gene mutations responsible for the low phytic acid trait in these mutant soybean lines have not been characterized.

In this study, we utilized a candidate gene approach to examine the molecular basis of the low phytic acid germplam GO3PHY-443, an improved line originating from the M153 event. Furthermore, we conducted expression profiling of 5,700 soybean cDNAs in an attempt to identify genes that are differentially expressed between the low phytic acid versus wildtype lines. The results of this research failed to find mutations in the genes considered as the most likely candidates for conferring a low phytic acid phenotype. Moreover, none of the genes of the phytic acid pathway represented in the gene profiling study were shown to be differentially expressed in the mutant line in a manner consistent with reducing phytic acid content. Cumulatively, the results of this study suggest that the mechanism of gene control for the low phytic acid trait originating from M153 may differ from that observed in other phytate crop species.
MATERIALS AND METHODS

Plant Materials
Soybean germplasm lines Prichard-RR and G03PHY-443 were used in this study. Prichard-RR is a wildtype soybean line with normal phytic acid content (4.3g/kg\textsuperscript{-1} dry wt.). G03PHY-443 (< 2g/kg\textsuperscript{-1} dry wt.) is a low phytic acid line derived from the cross of Prichard-RR x CX1834-1-2 followed by three backcrosses to Prichard-RR. CX1834-1-2 is a F\textsubscript{3.5} progeny of the cross ‘Athrow’ x M153-1-4-6-14 developed by USDA and Purdue University. M153-1-4-6-14 originated from M153, one of the two low phytic acid mutants isolated by Wilcox et al. (2000).

Measurement of Inositol Phosphate Intermediates and Myo-Inositol
To analyze the inositol phosphate content of mutant and wildtype soybeans, one half gram of ground mature seeds was stirred with 10ml 0.5N HCl for 60 minutes. The mixtures were centrifuged at 20,000 x g for 30 minutes at 4°C and 5 ml of the supernatants were diluted 10-fold to give a concentration of 0.05N HCl. Diluted samples were applied to AG 1-X8 200-400 mesh chloride columns (0.8 ml). The columns were washed with 2 aliquots of 5 ml 0.05N HCl and inositol phosphates were eluted with 10 aliquots of 1 ml of 1N HCl. The 10 ml eluates were dried at 37°C in a rotary evaporator. The residues were dissolved in 500 µl distilled water and Millipore filtered for HPLC. HPLC protocols were followed as previously described (Phillippy and Bland 1988).
Briefly, gradient ion chromatography was performed on the Dionex AG7/AS7 columns with a range of 5% (v/v) to 100% (v/v) of 0.25N HNO₃ and 95% (v/v) to 0% (v/v) of coumarin (100% =25 mg/L) for 30 minutes followed by 5 minutes of 100% 0.25N HNO₃ at a flow rate of 1 ml/min. Eluates were combined with 0.1% Fe (NO₃)₃ (w/v) in 2% HClO₄ (w/v) at a total of 1.5 ml/min and absorbance was monitored at 290 nm.

To quantitate free myo-inositol, four independent sets of seeds, each representing three developmental stages (early, mid, and late maturation) were submitted to the Metabolomics and Proteomics Laboratory (MPL) at North Carolina State University for analysis by gas chromatography-mass spectrometry (GC-MS). Seeds were weighed, ground in a mixture of methanol and water, and put into 20 ml polypropylene scintillation vials. A portion of the homogenate was centrifuged to remove particulates and an aliquot of the supernatant was reduced to dryness in a 2-mL glass instrument vial for derivatization. The residue was suspended in 250µl of N, N-dimethyltrimethylsilylamine (TMS-DMA) and 750 µl of solvent (0.1% v/v trifluoroacetic acid in 3:1:1, acetonitrile:benzene : tetrahydrofuran, by volume) and incubated at approximately 60°C for one hour to convert the free inositol to trimethylsilyl (TMS) derivatives. The resulting extracts were centrifuged and analyzed by GC-MS without further cleanup. GC-MS was performed with an Agilent 5890 gas chromatograph coupled to an Agilent 5972 mass-selective detector. Chromatographic separations were achieved with a Thermo TR-50MS (50% phenyl polysilphenylene-siloxane) column (30m length; 0.25mm inside
diameter; 0.25μm film thickness). Helium, the carrier gas, was set at a constant flow of 1.2mL per minute (linear velocity of 40cm/s). One microliter of extract was injected into a split-splitless injector operated with a 100:1 split ratio at a constant temperature of 250°C. The column oven temperature was programmed for a 50°C initial temperature with a 10°C/min ramp to a final temperature of 350°C. Quantification was conducted by comparing peak areas obtained for TMS-inositol in the samples with a series of inositol reference standards analyzed concurrently with the extracts. The chromatographic data was processed using Agilent’s ChemStation software.

**Investigation of Cellular DAG Kinase and PI Kinase Activities**

Microsomes were isolated from frozen soybean seeds (mid-maturation) as follows: One half gram of seed tissue was powdered in liquid N₂ and immediately homogenized in 15 ml ice-cold buffer containing 100mM KCl, 15mM HEPES pH 7.5, 0.085mM EGTA pH 7.7, 3 ml Protease inhibitor cocktail (Sigma-Aldrich Corp. St. Louis, MO), and 1.5 ml of 100% glycerol. The homogenate was centrifuged at 3,000 x g for 12 minutes at 4°C. The resulting supernatant was centrifuged again at 30,000 x g for 1 hour at 4°C. The pellets were resuspended in 20μl of 20mM HEPES pH 7.5, placed on ice and immediately assayed for lipid kinase activity. Protein concentration was determined with the Bio-Rad protein reagent using γ-globulin as a standard.
Phosphatidylinositol kinase and diacylglycerol kinase activities were assayed in duplicate as described (Stevenson-Paulik J. et al. 2003). Each assay sample contained 30µg of microsomal protein, mixed with cofactors as outlined (Stevenson-Paulik J. et al. 2003) and 10 µCi $[^{32}P]$ ATP (7,000 Ci mmol$^{-1}$) in a reaction volume of 50 µl. Vortex-mixed samples were incubated at room temperature for 30 or 60 min before addition of 1.5 ml of ice-cold CHCl$_3$/methanol (1:2 v/v) to stop the reactions. The samples were maintained at 4°C until the lipids were extracted. Lipids were extracted, dried, and separated by TLC on silica gel plates (Whatman, Clifton, NJ) using a CHCl$_3$:MeOH:NH$_4$OH:H$_2$O (90:90:7:22, v/v) solvent system as described (Stevenson-Paulik J. et al. 2003). The $^{32}$P-labeled phospholipids were quantified with a Bioscan System 500 imaging scanner and the results were calculated accordingly (Stevenson-Paulik J. et al. 2003).

**Characterization of MIPS, ABC Transporter and MIK Genes**

First strand cDNAs were generated from total RNA using the kit “SuperScript® First-Strand Synthesis System for RT-PCR” (Invitrogen, Carlsbad, CA). A full-length MIPS gene was successfully PCR amplified from Prichard-RR and G03PHY-443 based on soybean MIPS gene information deposited in Genbank (accession number AF293970) using developing seed cDNA as a template (primers 5’-AGGATTCTCTTCTTTATTTCTTTTG-3’ and 5’-CAGAATATGCCCCGCTCTAA-3’). PCR reactions were conducted follows: 50 µl final reaction volumes contained 10 mM Tris-Cl (pH 8.3, 25°C), 50 mM KCl, 1.5 mM MgCl$_2$, 200 µM each dNTP, 25 pmol of
each primer, 1 unit of Taq DNA polymerase (Expand High Fidelity PCR System, Roche Applied Science, Indianapolis, IN), and 1 µl template cDNA. Each reaction was carried out under the following conditions: 94°C for 2 min followed by 30 cycles at 94°C for 30 s, 57°C for 45 s and 72°C for 2 min; terminating with a final 7 min extension at 72°C. Amplification products were cloned using TA cloning protocols according to the manufacturer’s instruction (Invitrogen, Carlsbad, CA). Sequencing was performed at the Iowa State University Biotechnology Center, and the sequences were compared using the software provided at http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html. Northern blotting and immunoblotting were performed as previously described (Dewey et al. 1994, Tang et al. 2005).

Three putative ABC transporter genes were cloned. Since the size of two of these genes, ABCT_63 and ABCT_84, is approximately 4.5 kb, two pairs of primers were designed to cover the entire sequences. Candidate soybean ABC transporters were selected based on BLAST searches of soybean databases using the low phytic acid-related ABC transporter protein sequence in maize (accession number NP_001106060) as a query sequence. The following primer sequences and annealing temperatures were selected for amplifying the soybean ABC transporter genes: **ABCT_63**, 5’- ACCTCCAGATGGTGAAATGC - 3’(forward) and 5’- TCTTCTTCTCTTTAATTGCTTTTTGA -3’(reverse), 5’ -TCAAAAAAGCAATTTAAAGAGAAGAAGA -3’ (forward) and 5’-
GCATTACACCGTTCCATT TG -3’(reverse), annealing temperature = 58°C, extension
time = 2 min 20 sec; **ABCT_84**, 5’ - CAATCTGCTTCTTCTTCTTCATCA-3’ (forward) and 5’ -CCAAGGGTACATTTTCATCTGA - 3’ (reverse), 5’ - 
TCAGATGAAATGTACCCTTGG -3’ (forward) and 5’ - 
GGTTTTGATGCATTTTCATCC -3’ (reverse), annealing temperature = 58°C, extension time = 2 min 20 sec; **ABCT_165**, 5’ - TATCATCTATCATCATCATCACCATCACC - 
3’ (forward) and 5’ - CTATGAGGAGGAAACCTCACTTTG-3’ (reverse), annealing temperature = 59°C, extension time = 2 min.

Two putative MIK genes were PCR cloned using the primers 5’ -
TCGTATTTTCCCTCCCATGA-3’ (forward) and 5’-AAGGAAGGGCACTGGAGAGT 
-3’ (reverse); 5’ - CGTCCCCACAGTTACCCTT-3’ (forward) and 5’-
GACATAGCTAGCAAAAATCCCTCA-3’ (reverse), respectively, designed after 
selecting soybean MIK candidates using the maize MIK protein sequence (accession number NP_001105790) as a query to search soybean databases. PCR and sequence comparisons were performed as described above with an annealing temperature of 58°C. Southern and Northern blot analyses were conducted as described previously (Dewey et al. 1994). For Southern blotting, partial gene fragments were used as probes. For 
Northern blots, full-length gene probes were synthesized.

**Microarray Analysis of Seed-Expressed Genes**
A soybean seed-specific cDNA library containing a 5,700 unigene set derived from numerous developing seed cDNA libraries was purchased from the American Type Culture Collection (ATCC, Manassas, VA). DNA slides for microarray analysis were prepared according to the protocol of Eisen and Brown (1999). Basically, cDNA inserts were PCR-amplified using universal M13 forward and reverse primers (Sigma-Aldrich Corp. St. Louis, MO) and concentrated with the Millipore Montage PCRµ96 plate system (Billerica, MA). Concentrated cDNAs were printed onto amino silane-coated slides (Corning GAPS II) by an Affymetrix GMS 417 array printer (Santa Clara, CA). Printed cDNA slides were UV crosslinked, followed by 2 hours of baking at 75°C to further immobilize cDNAs to the slide surface.

Total RNA was isolated from frozen soybean seeds using the Trizol reagent according to the protocol provided by the manufacturer (Invitrogen, Carlsbad, CA). An indirect amino-allyl labeling method was applied to generate fluorescence-labeled cDNAs from wildtype Prichard-RR or mutant line G03PHY-443 following the instructions from the protocol reported at http://pga.tigr.org/protocols.shtml. Pre-hybridization, hybridization, and post-hybridization washes were conducted according to the procedure of Alba et al. (2000). Microarray slides were scanned with SCANARRAY 2.1 (GSI Lumonics, Billerica, MA) and acquired images were normalized and quantified for signal intensity using QUANTARRAY analysis software (PerkinElmer, Waltham, MA) by the histogram method. The experiments were designed as “dye-swap” to control for variation and noise of the microarray data (Kerr et al. 2001).
RESULTS

Inositol Phosphate Intermediates and Free Myo-Inositol Levels in Wildtype versus Mutant Lines

The specific genes of the phytic acid biosynthetic pathway that are most likely to be impacted to give a low phytic acid phenotype can be best predicted once one has knowledge concerning the intermediates that accumulate in the mutant lines. For example, if accumulation of any inositol phosphate intermediates is observed in the mutant line compared to the wildtype, it may indicate that the activity of the enzyme that further metabolizes the intermediate has been disrupted. Therefore, genes encoding such an enzyme would be good candidates for causing the low phytic acid phenotype.

Measurement of IP₃, IP₄ and IP₅ in mid-maturation soybean seeds was conducted through HPLC analysis. These assays detected a small amount of IP₅ in seeds of Prichard-RR, which might indicate that the synthesis of IP₅ to IP₆ (phytic acid) has not gone to completion (Fig. 2). For the mutant line G03PHY-443, there was no detection of any IP₃, IP₄ and IP₅ species, suggesting that none of these phosphorylated intermediates accumulated in the mutant. As previous observed (Wilcox et al. 2000), the content of phytic acid (IP₆) was about 45% lower (by integrating the each area of each peak) in line G03PHY-443 than in normal line Prichard-RR.
Myo-inositol content was measured at three developmental stages for Prichard-RR and G03PHY-443 soybean seeds. Early-maturation seeds were 120 mg in size, while mid-maturation and late-maturation seeds were approximately 240 mg and 400 mg, respectively. High levels of myo-inositol were found in young seeds which were reduced as the seeds matured in both lines. The abundance of myo-inostol in early-maturation seeds did not differ in both lines (Fig. 3). When assayed at mid- and late- maturation, however, we observed that the decrease of myo-inositol was much greater in Prichard-RR than in G03PHY-443.

The fact that G03PHY-443 has a reduced phytic acid phenotype but shows no accumulation of inositol phosphate intermediates and a slower metabolism of myo-inositol, phenotypically paralleled the low phytic acid maize lines *lpa*1 and *lpa*3. The *lpa*1 locus is defined by a debilitating nucleotide substitution in an ATP-binding cassette transporter gene, and *lpa*3 is associated with a deletion of a MIK gene (Pilu et al. 2003; Shukla et al. 2004; Shi et al. 2007; Shi et al. 2005).

**Enzymatic Activities of PI Kinase and DAG Kinase**

It has been proposed that phytic acid can be synthesized from a branch of the pathway involving lipid metabolites such as phosphatidylinositol and diacylglycerol (Stevenson-Paulik et al. 2002, Xia et al 2003). The enzymes that function on these two compounds are phosphatidylinositol (PI) kinase and diacylglycerol (DAG) kinase (Fig. 1). We
compared the endogenous PI kinase and DAG kinase activities between microsomal extracts of Prichard-RR and G03PHY-443 using an in vitro assay to investigate the early steps of the lipid-dependent branch of the phytic acid synthetic pathway. Both PI kinase and DAG kinase activities were more than 2 times greater in G03PHY-443 microsomes than in Prichard-RR microsomes (Fig. 4A and 4B). Increases in these activities, however, would not be expected to result in reduced phytic acid accumulation. Instead, elevated lipid kinase activities may be reflecting a compensatory effort in response to the low phytic acid levels in soybean seeds through enhancement of the lipid-dependent phytic acid biosynthetic pathway. Alternatively, it is possible that the higher levels of myo-inositol present in the mutant line may serve to provide more endogenous substrate to stimulate the activities of these two kinases. Given the positions of PI and DAG within the pathway, combined with the information that no IP₃, IP₄ or IP₅ accumulation was identified in the mutant line, we do not believe that direct perturbations in the lipid-mediated synthetic pathway or late steps of IP₆ pathway (those after IP₃ formation) are likely to be the causal basis of the low phytic acid phenotype of G03PHY-443.

**MIPS, ABC Transporter and MIK Genes Analysis**

Cumulatively, the results of our biochemical analyses of the wildtype versus low phytic acid mutant soybean suggest that genes encoding MIPS, an ABC transporter and/or MIK represent the best candidates for mutations leading to the low phytic acid phenotype diagnostic of G03PHY-443. Furthermore, previous studies from soybean and other crops
have confirmed the importance of the MIPS gene in controlling the phytic acid level in crops (Hitz et al. 2002; Nunes et al. 2006; Kuwano et al. 2006). We thus initiated our candidate gene approach by analyzing a seed-specific MIPS gene in our soybean lines. The pioneering research of Hegeman et al (2001) followed by the report from Hitz et al (2002) demonstrated that a single MIPS isoform designated GmMIPS1 (access number: AF293970 or AY038802) was predominantly expressed in developing soybean seeds. Therefore, if the MIPS step is related to the low phytic acid trait in our mutant line, this isoform likely be the only one that could mediate the level of phytic acid reduction observed in G03PHY-443. Accordingly, a 1,791bp full-length GmMIPS1 cDNA was amplified and cloned based on gene information from these studies. Comparisons of the GmMIPS1 cDNA sequences between Prichard-RR and G03PHY-443 revealed no differences between the two lines (data not shown). Northern blotting was then performed to compare the relative expression level of the GmMIPS1 gene between Prichard-RR and G03PHY-443 (Fig. 5A). No differences were identified in early-, mid- or late-maturation seeds between Prichard-RR and G03PHY-443. The observation of greater transcript accumulation of GmMIPS1 during early-maturation seeds in both lines was consistent with the results reported by Hegeman et al (2001). Finally, immunoblot analysis was conducted using mid-maturation soybean seeds to compare MIPS protein levels between the two lines (Fig. 5B). GmMIPS1 enzyme accumulation appeared to be the same between the two genotypes.
Even though two separate studies reported decreases in the MIPS transcripts in the maize *lpa1* low phytic acid line and its allelic line *lpa241* when compared with their wildtype counterparts (Pilu et al. 2003; Shukla et al. 2004), a recently published paper provided convincing evidence that a multidrug resistance-associated ABC transporter is actually responsible of the low phytic acid trait of the *lpa1* locus (Shi et al. 2007). Furthermore, this group demonstrated that low phytic acid soybeans could be generated through RNAi silencing of a homologous soybean ABC transporter. Using the maize ABC transporter and partial soybean ABC transporter sequences as queries, we conducted TBLASTN searches against the first draft of soybean genome as reported in http://www.phytozome.net/soybean.php. Three candidate ABC transporter genes were identified and cloned. Of the three ABC transporters, ones encoded by ABCT_63 and ABCT_84 have the most similarity to the maize deduced as sequence of the ABC transporter (68.4% and 68.6% at the protein level, respectively); while the third one, ABCT_165, is a smaller protein, nearly half the size of the other two proteins (Fig. 6), sharing 33.4% protein identity to the maize ABC transporter (alignment software: http://bioinfo.hku.hk/services/analyseq/cgi-bin/alignn_in.pl). ABCT_63 represents the full-length genomic version corresponding to the partial ABC transporter fragment used by the Shi et al. group to produce their transgenic soybean plants (data not shown). DNA sequence analysis of the three putative ABC transporters indicated no differences between the wildtype and mutant soybean lines. Southern blotting was also conducted using several restriction enzymes to determine whether gross changes may have occurred
at the genomic level for any of these genes. No polymorphic bands were identified (Fig. 7A). Northern blotting results showed that ABCT_84 accumulates at very low levels in soybean seeds (data not shown). In contrast, transcripts of ABCT_63 and ABCT_165 can be readily detected during all three stages of soybean seed development (Fig. 6B). ABCT_165 is most highly expressed during early-seed development and ABCT_63 shows greatest transcript accumulation in late-maturation seeds. However, no differential expression was found between the wildtype and low phytic acid line, suggesting that the gene mutations underlying the phytic acid trait do not affect the RNA levels of these genes.

A new gene in the phytic acid pathway was elucidated by Shi et al (2005) who discovered that a deletion in a myo-inositol kinase (MIK) gene was responsible for the low phytic acid phenotype of the maize line lpa3 (Shi et al. 2005). Since our mutant soybean line displays the same diagnostic features as maize lpa3, such as enhanced myo-inositol accumulation and lack of observable inositol phosphate intermediates, soybean MIK genes were believed to be reasonable candidates for the low phytic acid trait within G03PHY-443. Two MIK genes in soybean were identified, using the maize MIK protein as a query. Sequence analysis of the two putative soybean MIKs revealed that similar to the maize MIK protein, the soybean MIKs also belonged to the pfkB family associated with carbohydrate metabolism. The two predicted soybean proteins, designated MIK1 and MIK2, share 72.1% amino acid identity with each other. When compared with maize
MIK (Fig. 8A), MIK1 and MIK2 share 44.3% and 44.5% identity to maize homolog, respectively. DNA and protein sequence comparisons revealed no differences between the wildtype and mutant lines for either gene. Southern blotting also revealed no polymorphisms (Fig. 8B). Northern blot analysis showed that higher expression levels in young seeds for both MIK1 and MIK2 (Fig. 8C). Interestingly, elevated levels of MIK1 transcripts were observed in G03PHY-443 compared to Prichard-RR. Similar to the lipid kinase results, the most likely interpretation of this result is that elevated MIK1 expression in G03PHY-443 may be due to a compensatory response to low phytic acid production in the mutant line.

**Seed-Specific Transcript Profiling**

To obtain a more global assessment of the effects of the mutation leading to the reduced phytic acid phenotype, a microarray approach was applied to a profile transcripts of wildtype line Prichard-RR in comparison to the mutant line G03PHY-443. Five thousand seven hundred unique EST sequences specifically expressed in soybean seed tissues (derived from numerous developing seed cDNA libraries including cotyledons, pods, and seed coats) were obtained and corresponding DNA chips were synthesized. The 5,700 unigene set was originally isolated and characterized as part of the soybean genome initiative (Vodkin et al. 2004).
To contrast gene expression levels in soybean seeds between Prichard-RR and G03PHY-443, three ‘dye-swap’ hybridizations were performed with fluorescence-labeled cDNAs generated from normal line Prichard-RR and low phytic acid line G03PHY-443. Analysis of these independent hybridizations revealed very few genes that were up-regulated or down-regulated between the two genotypes, suggesting that DNA chip quality was good. Northern blotting assays with select representative genes also supported the microarray results (data not shown). Considering that known genes in the proposed phytic acid biosynthetic pathway beyond those specifically examined by the candidate gene screen might be involved in the mutated phenotype, we first examined the potential phytic acid-related ESTs expression profiles in our microarray data. Table 1 shows the fold-change for each of these ESTs between wildtype Prichard-RR and mutant line G03PHY-443. The data indicate no significantly differential expression between any of these genes in the wildtype vs. the mutant line using a cut-off threshold of greater than 2 or less than 0.5. However, some notable trends were observed. For example, the equal accumulation of putative inositol 1,3,4-trisphosphate 5/6 kinase transcripts supports the conclusion from the measurement of inositol phosphate intermediates that these kinases aren’t likely to be related to the mutant phenotype; Slightly higher levels of both phosphatidyl-4-phosphate 5-kinase and a PLC transcript are consistent with the results of the enzymatic activity assays of PI knase and DAG kinase, which suggests that the lipid-dependent pathway is stimulated in the mutant line. Interestingly, we observed the modestly elevated levels of transcripts encoding myo-inositol monophosphatases (IMP) in the mutant line, suggesting
that this gene family is also up-regulated in G03PHY-443 compared to Prichard-RR. In Table 2 genes that are up-regulated in the wildtype line with fold change ratio ≥ 2 are shown (up-regulated genes are highlighted because it was more conceivable that inhibition of gene expression would be most likely associated with the low phytic acid trait in the mutant line). The 7S globulin gene encodes one of the major soybean seed storage proteins. Higher expression levels of this gene were observed in wildtype soybean suggesting that phytic acid biosynthesis may interact with genes associated with storage protein production. We also detected a differentially expressed auxin-controlling gene, ADR12, in our microarray data. The involvement of phytic acid and its intermediates as signaling components may provide an explanation for this result (Datta et al. 1993; van Leeuwen et al. 2004; Abreu and Aragao 2007; Yang et al. 2008). Finally genes encoding glucose and ribitol dehydrogenase, enzymes related to carbohydrate metabolism, were also differentially regulated, suggestive of cross-talk between the pathways as reported by Hitz et al. (2002). Overall, none of the observed differentially expressed genes (including those down-regulated) could be directly linked to the phytic acid production pathway.

DISCUSSION
Despite the fact that a low phytic acid seed trait has been generated both in important crop species (barley, maize, rice, wheat, and soybean) and the model plant Arabidopsis, knowledge concerning the molecular basis for phytic acid biosynthesis in plants is still fragmentary and many aspects are poorly understood. This is exemplified by our inability to clearly define the molecular perturbation responsible for the low phytate trait in soybean germplasm G03PHY-443, despite utilizing all the available information regarding metabolite flow through the pathway and knowledge of gene mutations in other plant species. Biochemical characterization of developing seeds from G03PY-443 failed to show accumulation of any phosphorylated inositol intermediates; furthermore, greater accumulation (or retention) of free inositol pools was observed as the seeds matured. These results suggested that an early step of the phytic acid pathway was impacted. However, of the three known gene candidates that could be expected to yield such a phenotype (MIPS1, MIK and an ABC transporter), no gene mutations were observed in any soybean homolog examined in this study. Although the soybean genome encodes four MIPS isoforms (Chappell et al. 2006), we only characterized the specific gene that is most highly expressed during seed development (and has been previously implicated in storage phytic acid accumulation (Hitz et al. 2002)). Other isoforms were not included because we did not consider it likely that mutations in a "minor" MIPS gene would result in lowering the phytic acid content to the levels observed in G03PHY-443 (50% of that found in wildtype soybean). This assumption was recently validated as a study of this
same soybean germplasm by Chappell et al. (2006) failed to find mutations in any of the four unique MIPS isoforms examined.

Molecular analysis of soybean MIK and ABC transporter isoforms also failed to identify mutations within these genes in line G03PHY-443. The MIK genes were selected as candidates for the low phytic acid phenotype based on the recent discovery that mutations in the maize MIK defined the low phytic acid locus known as lpa3 (Shi et al. 2005). Database searches, including examination of the first draft of the complete soybean genome (http://www.phytozome.net) suggest that the MIK1 and MIK2 genes included in this study are the only viable soybean orthologs of the maize MIK gene (data not shown). Similar to the low phytate trait found in G03PHY-443, maize plants carrying the lpa3 locus display no accumulation of inositol phosphate intermediates and contain elevated levels of free inositol. Interestingly, Northern blotting results suggest that soybean MIK1 transcript levels are actually greater in the mutant than wildtype line during early seed development. One possible explanation is that the cellular perception of abnormally low phytic acid levels in the mutant is signaling the enhancement of elements of the pathway as a means of compensation. This concept is further supported by our biochemical analyses showing that key enzyme activities of the lipid-dependent phytic acid pathway were increased in G03PHY-443 (Fig. 4) and the results of the microarrays that indicated a possible increase in accumulation of phosphatidylinositol-4 P, 5-kinase and PLC transcripts (Table 1).
The observed enhancement of elements of the lipid-dependent phytic acid pathway is also interesting in that it provides support for the hypothesis that this branch of the pathway is actively involved in the production of storage phytic acid within the seed. Although it has long been speculated that inositol-1,4,5-trisphosphate derived from the PLC-dependent release of phosphatidylinositol 4,5-bisphosphate may represent a viable pathway for the production of seed storage IP₆, there has been little evidence reported to support this. Our observation of an approximately two-fold increase in PI kinase and DAG kinase activities in G03PHY-443 not only support the involvement of the lipid-dependent phytic acid pathway in the synthesis of seed phytate in soybean, they also suggest that blocks to this branch of the pathway are unlikely to be directly involved in the mutant phenotype. A notable exception to this conclusion, however, could be the possibility that mutations in either a specific PLC or phosphatidlyinositol 4 P 5-kinase isoform dedicated to the formation of IP₆ from IP₃ during seed development could be responsible for the trait. A recent report suggests that PLC-dependent phytic acid biosynthesis is not the major route in rice (Suzuki et al. 2007). Nevertheless our results suggest the involvement of this branch of the pathway given that low levels of phytic acid in soybean seeds can appear to stimulate this branch of the pathway.

Map-based cloning of the lpa1 locus of maize revealed that mutations in a gene that is a member of the ABC transporter family is responsible for conferring the low phytic acid phenotype (Shi et al. 2007). Although the mechanism by which phytic acid is reduced in
lipa1 mutants is not known, the failure to find increased accumulation of inositol phosphate intermediates led to the speculation that this gene is also involved in an early step of the pathway. Our examination of soybean EST databases and the first draft of the entire soybean genome revealed three potential orthologs of the maize ABC transporter, ABCT_63, ABCT_84, and ABCT_165. A combination of sequence analysis and both DNA and RNA blotting experiments revealed no nucleotide polymorphisms or alterations in transcript accumulation patterns that could implicate any of these genes in the low phytic acid trait of G03PHY-443.

Cumulatively, the results of this study suggest that perturbations in an uncharacterized early step(s) of the phytic acid pathway are responsible for the low phytate seed phenotype found in G03PHY-443. One potential target would be the genes encoding the enzyme responsible for the synthesis of inositol bisphosphates from inositol monophosphates (Fig. 1). This reaction is perhaps the least understood step of the entire phytic acid pathway even though a recent study indicated that an inositol 1,3,4-trisphosphate 5/6 kinase in barley also showed some enzymatic activity toward mono- and bis-phosphates (Josefsen et al. 2007). Another possibility, as discussed above, is that mutations in a seed-specific isoform of PLC could reduce flux through the lipid-dependent branch of phytate biosynthesis, possibly lowering the overall levels of seed IP6. Finally, the microarray profiling studies suggest an alternative mechanistic possibility that warrants consideration. Transcript accumulation of three genes encoding inositol
monophosphatases was moderately enhanced in the G03PHY-443 line (Table 1). Increases in this activity would have the potential of "short-circuiting" the phytate pathway by metabolizing inositol monophosphates back to myo-inositol and free phosphate. Given that the mutations in G03PHY-443 are recessive, it is conceivable that a repressor of inositol monophosphatase gene expression may have been mutated, leading to enhanced phosphatase activity yielding a low phytate phenotype with enhanced myo-inositol accumulation and the absence of increased accumulation of inositol phosphate intermediates.
REFERENCES


Raboy V. 2003. myo-Inositol-1,2,3,4,5,6-hexakisphosphate. Phytochemistry 64: 1033-1033


representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5:73


expressing AtIpk2 beta, an inositol polyphosphate 6-/3-kinase from Arabidopsis thaliana. Plant Mol Biol 66: 329-343

York J.D. 2006. Regulation of nuclear processes by inositol polyphosphates. Biochim Biophys Acta 1761: 552-559


FIGURE LEGENDS

**Figure 1.** Phytic acid biosynthetic pathways proposed in plants. Characterized enzymes in plants include: MIPS, myo-inositol 3-phosphate synthase; IMP, myo-inositol monophosphatase; MIK, myo-inositol kinase; Ipk2, inositol 1,4,5-trisphosphate 6/3-kinase; Ipk1, inositol 1,3,4,5,6-pentakisphosphate 2-kinase; inositol 1,3,4-trisphosphate 5/6-kinase; ABC transporter. PI kinase, phosphatidylinositol kinase; DAG, diacylglycerol kinase; PI4P 5-kinase, phosphatidylinositol 4,P-5-kinase; and PLC, phospholipase L are enzymes in PLC-mediated signaling transduction pathways(shown in rectangles).

**Figure 2.** HPLC analysis of inositol phosphates in the wildtype Prichard-RR and the mutant G03PHY-443 soybean lines. Numbers on the images represent elution times for each eluted compound. Standard control, hydrolysates of 50µg IP₆.

**Figure 3.** Comparison of the levels of myo-inositol during three stages of soybean seed development between the wildtype Prichard-RR (Pri) and the low phytic acid mutant line G03PHY-443 (443). Early = ~ 120 mg average seed fresh weight, mid = ~ 240 mg average seed fresh weight late = 400 mg average seed fresh weight. Bars representing standard error were calculated based on 4 independent replicates.
Figure 4. A. Comparison of endogenous phosphatidylinositol (PI) kinase activity between the wildtype line Prichard-RR (Pri) and the mutant line G03PHY-443 (443). Reaction times were 30 or 60 minutes. B. Comparison of endogenous diacylglycerol (DAG) kinase activity between the same genotypes. Error bars represent the standard error of 2 independent replicates.

Figure 5. A. Northern blot of the myo-inositol-3-phosphate synthase (MIPS1) gene most abundantly expressed during seed development. RNAs were isolated from soybean seeds harvested at early- (average seed fresh weight, 120 mg per seed), mid (average seed fresh weight, 220 mg per seed) and late- (average seed fresh weight, 400 mg) maturation stages of development. Pri, Prichard-RR; 443: G03PHY-443. B. Western blot of the MIPS1 protein from mid-maturation soybean seeds. Ten micrograms of protein from two independent samples of each genotype were probed using an anti-MIPS antibody.

Figure 6. Alignment of 3 putative soybean ABC transporters with the maize ABC transporter involved in phytic acid biosynthesis (NP_00106060). Residues that are identical between two or more amino acids are box shaded.

Figure 7. A. Southern blot analysis using a partial ABCT_63 transporter gene as a probe. M, molecular marker. Pri, Prichard-RR; 443, G03PHY-443. B. Northern blot analysis of
ABC transporters during 3 stages of soybean seed development. Upper blot was probed with ABCT_165; lower blot was probed with ABCT_63.

**Figure 8.** A. Comparison of putative soybean myo-inositol kinases (MIK1 and MIK2) and maize MIK (NP_001105790). B. Southern blot analysis of Prichard-RR (Pri) and G03PHY-443 (443) probed with partial MIK1 cDNA. C. Northern blot analysis of the same genotypes using RNAs from early- (average seed fresh weight =120mg), mid- (average seed fresh weight =240mg) and late- (average seed fresh weight, 120mg) maturation soybean seeds. Upper panel was probed with MIK1 cDNA; lower panel was probed with MIK2.
Table 1. Genes among the seed-expressed ESTs (5,700) printed on the microarray chips that may be involved in the phytic acid biosynthesis. WT/MU ratio represents the ratio of expression levels between the wildtype line Prichard-RR versus the low phytic acid mutant line G03PHY-443. Ratios represent the average of three independent experiments.

<table>
<thead>
<tr>
<th>GenBank Accession #</th>
<th>Predicted Gene Product</th>
<th>WT/MU Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI941146</td>
<td>myo-inositol-1-phosphate synthase (MIPS)</td>
<td>0.79</td>
</tr>
<tr>
<td>AI746269</td>
<td>myo-inositol-1-phosphate synthase (MIPS)</td>
<td>1.02</td>
</tr>
<tr>
<td>AW100674</td>
<td>myo-inositol-1-phosphate synthase isozyme-2 (MIPS)</td>
<td>1.56</td>
</tr>
<tr>
<td>AW607367</td>
<td>inositol 1,3,4-trisphosphate 5/6-kinase</td>
<td>1.10</td>
</tr>
<tr>
<td>AW433103</td>
<td>inositol 1,3,4-trisphosphate 5/6-kinase</td>
<td>1.26</td>
</tr>
<tr>
<td>AW278716</td>
<td>putative phosphatidylinositol-4-phosphate 5'-kinase</td>
<td>0.76</td>
</tr>
<tr>
<td>AW308016</td>
<td>probable phosphatidylinositol-4-phosphate 5'-kinase</td>
<td>0.59</td>
</tr>
<tr>
<td>AW318222</td>
<td>myo-inositol-1(or 4)-monophosphatase 3 (IMP)</td>
<td>0.79</td>
</tr>
<tr>
<td>AW472201</td>
<td>myo-inositol-1(or 4)-monophosphatase 3 (IMP)</td>
<td>0.71</td>
</tr>
<tr>
<td>AW432980</td>
<td>myo-inositol-1(or 4)-monophosphatase 3 (IMP)</td>
<td>0.68</td>
</tr>
<tr>
<td>AI794656</td>
<td>phosphoinositol-specific phospholipase C (PLC)</td>
<td>0.55</td>
</tr>
<tr>
<td>AW164281</td>
<td>phosphoinositol-specific phospholipase C (PLC)</td>
<td>0.90</td>
</tr>
<tr>
<td>AW569264</td>
<td>phosphatidyglycerol/phosphatidylinositol transfer protein</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Table 2. Genes represented in the microarray chips that showed more than two-fold higher expression in the wildtype Prichard-RR comparing versus the mutant line G03PHY-443. Ratios represent the average of three independent experiments.

<table>
<thead>
<tr>
<th>GenBank Accession #</th>
<th>Predicted Gene Product</th>
<th>WT/MU Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW509499</td>
<td>basic 7S globulin precursor</td>
<td>6.44</td>
</tr>
<tr>
<td>AW509346</td>
<td>ADR12 protein</td>
<td>5.04</td>
</tr>
<tr>
<td>AW508896</td>
<td>glucose and ribitol dehydrogenase homolog</td>
<td>4.81</td>
</tr>
<tr>
<td>AV459657</td>
<td>hypothetical 65.1 KD protein slr1919</td>
<td>3.82</td>
</tr>
<tr>
<td>AV471959</td>
<td>metallothionein-II protein</td>
<td>3.07</td>
</tr>
<tr>
<td>AI933376</td>
<td>unknown gene</td>
<td>3.06</td>
</tr>
<tr>
<td>AI940862</td>
<td>similar to PAPA</td>
<td>2.65</td>
</tr>
<tr>
<td>AV471686</td>
<td>similar to DC59 protein</td>
<td>2.42</td>
</tr>
<tr>
<td>AI735922</td>
<td>similar to Convicilin Precursor</td>
<td>2.15</td>
</tr>
</tbody>
</table>
Figure 1. Putative phytic acid synthetic pathways in plants. Characterized enzymes in plants include (in blue): MIPS, myo-inositol 3-phosphate synthase; MIK, myo-inositol kinase; IMP, myo-inositol monophosphatase; Ipk2; inositol 1,4,5-trisphosphate 6/3-kinase; Ipk1, inositol 1,3,4,5,6-pentakisphosphate 2-kinase; inositol 1,3,4-trisphosphate 5/6-kinase; ABC transporter. Phosphatidylinositol kinase (PI kinase); phosphatidylinositol 4-phosphate 5-kinase (PI4P 5-kinase); phospholipase C (PLC); and diacylglycerol kinase (DAG kinase)(all in rectangle) are enzymes in PLC-mediated signaling transduction pathway.
Figure 2. HPLC analysis of inositol phosphates in the wildtype Prichard-RR and the mutant G03PHY-443 soybean lines. Numbers on the images represent elution times for each eluted compound. Standard control, hydrolysates of 50µg IP₆.
Figure 3. Comparison of the levels of myo-inositol during three stages of soybean seed development between the wildtype Prichard-RR (Pri) and the low phytic acid mutant line G03PHY-443 (443). Early = ~ 120 mg average seed fresh weight, mid = ~ 240 mg average seed fresh weight late = ~ 400 mg average seed fresh weight. Bars representing standard error were calculated based on 4 independent replicates.
Figure 4. A. Comparison of endogenous phosphatidylinositol (PI) kinase activity between the wildtype line Prichard-RR (Pri) and the mutant line G03PHY-443 (443). Reaction times were 30 or 60 minutes. B. Comparison of endogenous diacylglycerol (DAG) kinase activity between the same genotypes. Error bars represent the standard error of 2 independent replicates.
Figure 5. A. Northern blot of the myo-inositol-3-phosphate synthase (MIPS1) gene most abundantly expressed during seed development. RNAs were isolated from soybean seeds harvested at early- (average seed fresh weight, 120 mg per seed), mid (average seed fresh weight, 220 mg per seed) and late- (average seed fresh weight, 400 mg) maturation stages of development. Pri, Prichard-RR; 443: G03PHY-443. B. Immunoblot analysis of the MIPS1 protein from mid-maturation soybean seeds. Ten micrograms of protein from two independent samples of each genotype were probed using an anti-MIPS antibody.
Figure 6. Alignment of 3 putative soybean ABC transporters with the maize ABC transporter involved in phytic acid biosynthesis (NP_00106060). Residues that are identical between two or more amino acids are box shaded.
Figure 7. A. Southern blot analysis using a partial ABCT_63 transporter gene as a probe. M, molecular marker. Pri, Prichard-RR; 443, G03PHY-443. B. Northern blot analysis of ABC transporters during 3 stages of soybean seed development. Upper blot was probed with ABCT_165; lower blot was probed with ABCT_63.
**Figure 8.** A. Comparison of putative soybean myo-inositol kinases (MIK1 and MIK2) and maize MIK (NP_001105790). B. Southern blot analysis of Prichard-RR (Pri) and G03PHY-443 (443) probed with partial MIK1 cDNA. C. Northern blot analysis of the same genotypes using RNAs from early- (average seed fresh weight = 120mg), mid- (average seed fresh weight = 240mg) and late- (average seed fresh weight, 120mg) maturation soybean seeds. Upper panel was probed with MIK1 cDNA; lower panel was probed with MIK2.
CHAPTER IV

Future Research Directions
Further Characterization of the \textit{SACPD-C} Gene Product from Normal Soybean and the High Stearic Acid Line FAM94-41

My investigations of the genes responsible for the elevated 18:0 seed phenotypes found in germplasm lines A6 and FAM94-41 lead to the identification of a novel \(\Delta^9\)-S-ACP-Des encoding gene, \textit{SACPD-C}, that is expressed specifically within the developing seed. Although the genetic and phenotypic data provided compelling evidence that mutations in the \textit{SACPD-C} gene were responsible for the high 18:0 traits found in A6 and FAM94-41, the following questions concerning the SACPD-C enzyme remain: (1) why is the specific activity of SACPD-C so much lower than its castor bean or Arabidopsis orthologs when assayed \textit{in vitro}; and (2) why does the SACPD-C variant found in FAM94-41 (SACPD-C\textsubscript{FAM}) display greater \textit{in vitro} activity than the wild type enzyme (SACPD-C\textsubscript{WT})? Specific strategies to address these questions are outlined below:

\textit{Why does the wild type SACPD-C enzyme display such low \textit{in vitro} activity?}

There are several reasons why SACPD-C could show an activity profile \textit{in vitro} that does not reflect its true activity \textit{in vivo}. The initial priority would be to make sure that the mature version of the recombinant enzyme that I expressed in \textit{E. coli} represents the mature enzyme found within the soybean seed. The N-terminal region is the least conserved portion of SACPD-C with respect to other SACP\textDs, and therefore we had to rely solely on predictive computer programs to establish the likely plastid transit peptide
cleavage site. Unpublished work from Dr. John Shanklin's lab showed that modest alterations at the N-terminus of the castor SACPD enzyme significantly reduced its activity (J. Shanklin, personal communication). To determine the true SACPD-C N-terminus, I would purify the enzyme from developing soybean seeds and conduct N-terminal protein sequencing (using the resources of the NCSU Genomics Sciences Laboratory). Purification would involve adopting the protocols used by Dr. Shanklin's lab where they purified stearol-ACP-desaturase protein from castor beans, together with a SACPD-C specific antibody (discussed below) to separate it from the SACPD-A and -B isoforms. Should the true N-terminus prove to be different from the one I used to make the original recombinant constructs, the in vitro assays would be repeated using the corrected enzyme to determine whether this would lead to activity levels closer to that observed with previously characterized SACPD enzymes.

In Chapter II, I also discussed the possibility that the ACP and/or ferredoxin co-factors used in the in vitro experiments may not interact and function optimally with SACPD-C. To test this, I propose to conduct these assays using ACP and ferredoxin proteins that are expressed in developing soybean seeds. This could be accomplished through direct purification from developing soybean seeds. However, a more efficient approach would be to utilize the extensive EST database information available for soybean to identify the ACP and ferredoxin cDNAs expressed most highly during seed development. Purified,
recombinant versions of these co-factors would be used individually and in combination to test whether SACPD-C activity is increased using its endogenous partners.

Finally, we speculated that post-translational modifications that are not present on the recombinant *E. coli*-produced protein may be needed for SACPD-C to be fully functional. Since SACPD-C is not routed through the endoplasmic reticulum, Asn-glycosylation is not a consideration. Instead, phosphorylation would be the most likely post-translational modification that could potentially regulate enzyme activity. The purified SACPD-C protein that we would obtain for the above mentioned N-terminal sequencing experiments can also be analyzed by LC/MS/MS to determined whether any amino acids on the endogenous protein are phosphorylated. If these results suggest that a particular residue (or residues) is phosphorylated, our constructs will be re-engineered to introduce phospho-amino acid mimic mutations (for example, Asp residues in place of a Ser site that is phosphorylated) in the recombinant enzymes. These modified enzymes would then be tested *in vitro* to see whether enzyme activity is enhanced.

*Why is SACPD-C<sub>FAM</sub> more active than SACPD-C<sub>WT</sub>?*

In addition to SACPD-C<sub>WT</sub> showing atypically low activity levels *in vitro*, the other puzzling observation reported in Chapter II was that the mutant SACPD-C<sub>FAM</sub> enzyme was about 10-fold more active than the wild type enzyme. One possibility is that this observation is merely an artifact that may somehow be associated with our not using the
correct SACPD-C\textsubscript{WT} enzyme in the \textit{in vitro} assays as described above. Should any of the experiments outlined above yield a SACPD-C\textsubscript{FAM} activity that is much more active than observed originally, the same experiments will be repeated making the same modifications to SACPD-C\textsubscript{FAM}. It is possible that the differences between the normal and FAM94-41 derived enzyme activities will appear more as expected once the corrected versions of the enzymes, or use of endogenous co-factors, is applied in these assays.

Another scenario was also proposed in Chapter II. It is possible that the Asp to Asn mutation at position 126 in SACPD-C\textsubscript{FAM} may be inherently more active, but much less stable than the SACPD-C\textsubscript{FAM} enzyme. This could be tested by producing a SACPD-C specific antibody and using it in Western blotting assays. Should much higher levels of SACPD-C be observed in protein preparations from normal than from FAM94-41 soybeans, this would provide strong support for the hypothesis that the SACPD-C\textsubscript{FAM} enzyme is less stable \textit{in vivo}.

\addcontentsline{toc}{section}{Characterization of the Molecular Basis for the Low Phytate Soybean Germplasm}

\textbf{G03PHY-443}

\textit{Identification of the low phytate conferring genetic loci}

The results from Chapter III indicated that a candidate gene approach, while useful in revealing novel aspects of the phytic acid biosynthetic pathway in soybean, did not lead
to the identification of the specific gene mutations responsible for the reduced phytate phenotype of line G03PHY-443. Therefore, I believe that future efforts for identifying these mutations should be focus on a map-based cloning approach. Now that a preliminary draft of the complete soybean genome has been assembled, annotated and made available to the public, map-based cloning has become increasingly feasible. My approach would be to utilize the abundance of simple sequence repeat (SSR) markers that span the genome to identify markers closely linked to the trait. If no marker closer than 1 - 2 cm can be found, fine mapping would be used to obtain markers within this approximate distance by developing new SSR markers based on genomic sequences closer to the region of interest. Once SSR markers can be positioned that flank the low phytate locus within a region of contiguous genome sequence, analysis of the gene sequence information in this region can be combined with additional fine mapping to identify the nature of the mutant locus.

Refining the model of phytate biosynthesis in soybean

The results presented in Chapter III revealed some potentially interesting information regarding metabolite flow and the mechanisms involved in the low phytic acid line G03PHY-443. Microarray studies suggested that one or more genes encoding MIK, PLC, PI4,5P-kinase, and IMP activities may be upregulated in G03PHY-443. Our speculation was that increases in the transcription of these genes may be reflecting a compensatory effect in response to the perception of low phytate accumulation. To verify these results, I
would conduct a Real Time PCR analysis for each of the implicated genes to determine whether they are truly upregulated in G03PHY-443, or whether our microarray results were merely an anomaly. The speculation that phytate levels are possibly being reduced as a result of increased IMP activities could also be directly tested. An IMP-encoding gene could be placed under control of a strong seed-specific promoter (such as the β conglycinin promoter) and expressed in transgenic soybean plants. Should phytate levels be reduced in such plants, the idea that this may represent a means by which phytic acid accumulation is reduced in G03PHY-443 would become even more plausible.

Finally, our results still leave open the possibility that genes encoding inositol monophosphate kinase may be the step of the pathway that is perturbed in G03PHY-443. This is a step of the pathway where the enzyme and relevant gene(s) have yet to be identified. Dr. Brian Phillippy (a researcher currently working in the lab of Dr. Wendy Boss) is in the process of purifying the enzyme in soybean seeds. Once it has been purified, its identity can be revealed through LC/MS/MS analysis. Should this approach be successful, the genes encoding this activity in soybeans could then be identified, then analyzed using a candidate gene approach.
APPENDICES
APPENDIX I

Expression profiling on soybean leaves reveals integration of ER- and osmotic-stress pathways

André ST Irsigler$^{1,2}$, Maximiller DL Costa$^1$, Ping Zhang$^3$, Pedro AB Reis$^1$, Ralph E Dewey$^3$, Rebecca S Boston$^4$ and Elizabeth PB Fontes$^1$

$^1$Departamento de Bioquímica e Biologia Molecular, BIOAGRO, Universidade Federal de Viçosa, 36571.000 Viçosa, Minas Gerais, Brazil

$^2$Molecular Core Facility, Department of Biology, Florida State University, Tallahassee, FL 32306-4370, USA

$^3$Department of Crop Science, North Carolina State University, Raleigh, NC 27695, USA

$^4$Department of Plant Biology, North Carolina State University, Raleigh, NC 27695, USA

This paper was published in *BMC Genomics* 2007, 8:431.

I was responsible for generating the soybean seed-expressed DNA chips used in this study. Figure 1 shows a picture of the types of DNA chips generated not only for this study but also the microarray experiments described in Chapter III.
Figure 1. Image of soybean cDNA chips used in the dissertation. Two groups of chips were synthesized, each containing four blocks representing 5,700 seed-expressed genes and 60 selected control genes.
APPENDIX II

Analysis of Microarray Data
A. Microarray Experimental Design

Experimental design is a critical factor for microarray assays because a good design can avoid systematic errors and maximize the amount of information that can be derived from the data. Basically there are 3 types of experimental designs commonly used for microarrays:

(1) Reference Design: One dye is used to label the reference sample, another dye to label all treatment samples. Pairwise comparisons are then conducted between each treatment sample and the reference

(2) Loop Design: One dye is used to label one treatment sample (treatment 1), and another dye to label another treatment sample (treatment 2). Sequential pairwise comparisons are conducted in a linear fashion; treatment 1 versus treatment 2, treatment 2 versus treatment 3, etc…
(3) Dye-Swap Design: One dye is used to label one treatment sample, another dye to label a second treatment sample, followed by pairwise comparison. Simultaneously the dyes are swapped and the samples compared again.

The dye-swap design is most efficient for experiments (such as ours) where only two samples are being compared. This design not only provides replication, but also helps eliminate the variation from the dye treatment (there is a tendency for cy3 to give higher intensity readings than cy5).

B. Data Transformation, Normalization and Analysis

Raw microarray data (intensity measurements directly from the microarray scanner) are frequently transformed (base 2 log transformation) to avoid skewing the distribution of data and to simplify graphic presentation. However, most of the raw or transformed data can not be directly utilized because the experimental procedure involves multiple steps, each of which serves as a source of variation. Variability may arise from chip printing, cy3 and cy5 labeling, hybridization, or fluorescence detection. These variations are defined as technical variability, in contrast to the true biological variability due to differential accumulation of specific mRNA species.
Therefore, a normalization process of the raw data needs to be applied to adjust for technical biases. In Chapter III, we used the method recommended by the manufacturer (PerkinElmer) to normalize our raw data, i.e. subtract the mean intensity measurement of the array from each individual gene measurement and then divide by the standard deviation. In the paper shown in Appendix I by Irsigler et al << Expression profiling on soybean leaves reveals integration of ER- and osmotic-stress pathways, *BMC Genomics* 8:431>>, LOWESS analysis was applied as described in Materials and Methods.

Besides these two methods of normalization, we also used another statistical method called two-way ANOVA mixed modeling, which integrates the normalization process into the data analysis to automatically correct for systematic biases (Tsai et al. 2004). The model is as following:

\[
\log(y_{ijkg}) = \mu + A_i + D_j + V_k + G_g + (AG)_{ig} + (VG)_{kg} + \epsilon_{ijkg}
\]

\(\log(y_{ijkg})\): measured log-ratio for gene \(g\) of variety \(k\) measured on array \(i\) using dye \(j\).

\(\mu\): the error term for array \(i\), dye \(j\), variety \(k\) and gene \(g\).

\(A_i\): the effect of the \(i\)th array.

\(D_j\): the effect of the \(j\)th dye.

\(V_k\): the effect of the \(k\)th variety.

\(G_g\): the variation of the \(g\)th gene.
AG_{ig}: the effect of a particular spot on a given array.

VG_{kg}: the interaction between the \( k \)th variety and the \( g \)th gene.

\( e_{ijkg} \): the error term for array \( I \), dye \( j \), variety \( k \) and gene \( g \); assumed to be independent of \( y_{ijkg} \) and have a mean of zero.

The computations for the data analysis were carried using R/maanova software (http://research.jax.org/faculty/churchill/software/Rmaanova/index.html). Figure 2 shows a typical output reading from analysis using this model.

![Volcano plot for mixed effect model](image)

**Figure 2.** Normalized data using the two-way ANOVA mixed modeling statistical method.
The results obtained from the two-way ANOVA analysis were the essentially the same as those observed using the first two methods described. Since the first two methods are more straightforward and gave similar results, this alternative means of analysis was not used for the array analysis in Chapter III or the paper by Irsigler et al. (2007).
APPENDIX III

Additional Analysis of the SACPD-C Gene and its Enzyme Product
A. Real-Time PCR of SACPD-C transcripts in different soybean tissues

Real-time PCR was conducted to confirm the relative spatial distribution of SACPD-C transcripts in developing seeds, leaf, flower, root and stem tissue that was observed using Northern blot analysis.

Seed, leaf, flower, root, and stem cDNAs from cultivar DARE were isolated as described in Chapter II. Primers (forward 5’-ATGGAGTGAAAGATGACAGCGGCA-3’, reverse 5’-TGCTCGCTCTTGGAAATGACGTGTA-3’) were designed to flank the SACPD-C gene intron to prevent amplification from contaminating genomic DNA. An endogenous gene designated Gmpcc16 that has been reported to be expressed in soybean at similar levels in numerous tissues (Kelley et al. 2004). was assigned as a reference gene (Primers: forward 5’-TTCTTTTATCAGAAATAACGAGTTTAGCT-3’, reverse 5’-TGCTCATTTTTGCGGCAGCAGC-3’) A two-step quantitative real time PCR assay was carried out on instrument ABI PRISM 7900HT with the“SYBR GreenER qPCR SuperMix for ABI PRISM” kit (Invitrogen, Carlsbad, CA) according to the manufacturer’s instructions. Every reaction for each tissue was repeated independently at least 3 times. Relative quantification for SACPD-C accumulation levels was calculated using the comparative threshold cycle (Ct) method (Livak 1997) after similar efficiencies of PCR were assured (Ramakers et al. 2003). PCR products were verified by gel electrophoresis to make sure only true transcripts were counted
in real-time PCR. *SACPD-C* transcript data were normalized against the expression of Gmpcc16. As revealed in Figure 3., expression levels of *SACPD-C* in developing seeds were nearly 2000-fold higher than in any other tissues, indicating that *SACPD-C* is highly seed-specific.

![Figure 3](image_url)

Figure 3. Real-time PCR analysis of SACPD-C in developing seed, leaf, flower, root and stem tissue.
B. Western Blot Analysis of SACPD-C Using Antibody Generated Against the Homologous Castor Enzyme

The unexpected results obtained from the *in vitro* enzyme assays of SACPD-C (Chapter II, Fig.6) could have several possible explanations. One possibility was that the single nucleotide polymorphic mutation identified in the *SACPD-C* gene from FAM94-41 might cause a reduction in the stability of the SACPD-C enzyme, which would result in the accumulation of less SACPD-C protein in the mutated germplasm. Western blotting assays could test this hypothesis.

One gram of frozen developing soybean seeds was powdered in liquid nitrogen using a mortar and pestle, then immediately homogenized in 5 ml of homogenization buffer containing 0.1M Tris-HCl pH 7.5, 1mM EDTA pH 8.0, 10mM DTT, 10mM KCl, 1g sucrose, with the addition of protease inhibitors (1 mM PMSF and 1 ug mL-1 each of Aprotinin, Leupeptin, and Chymostatin). The mixture was centrifuged at 34,000 g at 4°C for 15 minutes. The supernatant was transferred to a new tube and protein concentration was determined by the Bradford method using Coomassie Blue Staining (Bio-Rad, Hercules, CA, USA). Ten micrograms of protein was separated on a 12% SDS- polyacrylamide gel and transferred to a polyvinylidene difluoride membranes. Protein blots were incubated with antibodies generated against the castor Δ9-stearoyl-ACP-desaturase (gift of J. Shanklin) according to the protocols for the ECL Plus kit (Amersham Biosciences, Amersham, UK).
Two closely migrating bands were observed on the western blot images. No differences were observed between SACPD bands among genotypes Dare, FAM94-41, and A6 (Fig. 4), despite the fact the SACPD-C gene is deleted in A6 germplasm. We believe this result is due to the castor antibody having significantly greater affinity for the SACPD-A and –B isoforms than SACPD-C. Comparisons of castor Δ9-18:0-ACP-desaturase with three soybean Δ9-18:0-ACP-desaturase isoforms identified in soybean revealed that the castor enzyme displays similarity ~87% identity to SACPD-A/B and only ~61% identity to SACPD-C. Therefore, it is reasonable that the antibody bound to SACPD-A/B proteins much more readily than it did to soybean SACPD-C protein, masking our ability to detect it on the Western blots.

![Western blot analysis of normal soybean cultivar Dare and enhanced 18:0 genotypes FAM94-41 and A6, using antibody generated against a castor bean Δ⁹-stearoyl-ACP-desaturase enzyme.](image)

Figure 4. Western blot analysis of normal soybean cultivar Dare and enhanced 18:0 genotypes FAM94-41 and A6, using antibody generated against a castor bean Δ⁹-stearoyl-ACP-desaturase enzyme.
References


