MODIFICATION OF FATTY ACID COMPOSITION IN SOYBEAN SEEDS TO IMPROVE SOYBEAN OIL QUALITY AND FUNCTIONALITY

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri

In Partial Fulfillment

Of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

By

Anh TUNG PHAM
Dr. Kristin Bilyeu, Dissertation Supervisor

July 2011

The undersigned, appointed by the dean of the Graduate School,

have examined the Dissertation entitled

MODIFICATION OF FATTY ACID COMPOSITION IN SOYBEAN SEEDS TO IMPROVE SOYBEAN OIL QUALITY AND FUNCTIONALITY

Presented by Anh Tung Pham

A candidate for the degree of

DOCTOR OF PHILOSOPHY OF PLANT BIOLOGY AND GENETICS

And hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Kristin D. Bilyeu

Dr. David A. Sleper

Dr. J. Grover Shannon

Dr. Emmanual Liscum

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Kristin Bilyeu not only for giving me the opportunity to join her research group and conduct exciting research in plant biology and genetics, but also for her continuous advising, support, and encouragement during my doctoral studies. The quality of this study was greatly influenced by her help in avoiding adversities and in solving problems which arose during the course of the research project, and also by her research ethics and her perseverance for a deeper understanding of this project.

I would also like to thank Dr. Grover Shannon for his collaboration and contribution to my dissertation projects. I would like to thank Dr. David Sleper for your guidance and support. I would like to thank Dr. David Sleper, Dr. Grover Shannon, Dr. Mannie Liscum for serving on my doctoral committee. Also, I would like to thank all the professors from different educational levels who have encouraged me or inspired me, one way or another, to appreciate academia and to pursue an advanced degree.

I greatly acknowledge the help of the current and past members of Dr.Shannon‟s group and Dr. Bilyeu‟s group: Dr. Jeong-Dong Lee for his helpful insights and for supplying the germplasm and segregating populations used in this study, Dr. Jason Gillman who share with me your deep knowledge and your willingness to help with all the difficulties that I encountered during the course of the study. My thanks also are for Kerry Clark, Christi Shannon, Stewart Selves and Paul Little for helping me with field works, protein, oil and fatty acid analysis. Thanks and good luck to my colleagues Ashley

Tetlow, Katherine Hagely, and Julian Lenis. You have created a family-like asmosphere that makes me enjoy and appreciate our friendship.

At last, many thanks to all friends that I was fortunate to meet in Columbia MO and had so much fun with. I am so happy that we shared a part of our “journeys” and I really hope to stay in touch. I am especially thankful to all you that our friendship has already lasted so long and your support has not changed, although so far away. I am also grateful to my family, especially my parents, Binh Nguyen and Minh Pham, and my brother, Linh Pham, for their love and understanding. I deeply wish that in the near future I will be able to share more time with you. Finally, I would like to thank my husband, Giang Nguyen for being there to share my excitement and disappointment, encourage me and keep my motivation at all times, and make me smile even those days that “everything went wrong”

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

CHAPTER 1 Literature review 1

Figure 49

Literature cited 51

CHAPTER 2 Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait 69

Abtract 71

Introduction 72

Materials and methods 77

Population development 77

DNA isolation and PCR for sequencing of *FAD2-1A* and *FAD2-1B* 79

Sequence analysis 79

FAD2-1B allele specific molecular marker assay 80

FAD2-1A allele specific molecular marker assay for 17D 81

FAD2-1A allele specific molecular marker assay for M23 81

Fatty acid, protein and oil determination 82

Population genotyping 83

Results 84

Identification of mutant alleles of *FAD2-1B* in soybean lines PI 283327 and PI 567189A 84

The PI 283327 *FAD2-1B* allele is associated with an increase in seed oleic acid content 86

Combinations of mutations in *FAD2-1A* and *FAD2-1B* produce high oleic acid levels in the seed oil 87

Excess desaturase activity: A single wild-type *FAD2-1* allele prevents high oleic acid accumulation 90

The high oleic acid phenotype is stable in plants grown in alternate environments.. 91

Full fatty acid profiles and total oil and protein content 92

Discussion 94

Figures 100

Tables 108

Literature cited 111

CHAPTER 3 A novel *FAD2-1A* allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content 115

Summary 116

Introduction 117

Materials and method 121

DNA isolation and sequencing of *FAD2-1A* and *FAD2-1B* 121

Population development.. 121

FAD2-1A allele specific molecular marker assay 122

Fatty acid, protein and oil determination 123

Population genotyping using SimpleProbe assay 123

Design of field experiments 124

Statistical analysis 125

Results 126

Identification of novel alleles of *FAD2-1A* and *FAD2-1B* in soybean plant introductions 126

Combinations of novel mutant alleles of *FAD2-1A* with mutant *FAD2-1B* alleles produce high oleic acid levels in the seed oil 128

Full fatty acid profiles and total oil and protein content 130

Discussion 133

Figures 137

Tables 139

Liturature cited 145

CHAPTER 4 Combinations of mutant *FAD2* and *FAD3* genes to produce high oleic acid and low linolenic acid soybean oil 148

Summary 149

Introduction 150

Materials and methods 155

Population development 155

FAD2-1A allele specific molecular marker assay 157

Genotyping for the gene combination of interest 158
Fatty acid determination 158
Stability of high oleic/low linolenic soybean..................................... 159
Statistical analysis 160
Results 161
Maturity of the HOLL lines 162
For the HOLL soybean lines with null *FAD2-1A*(Δ) alleles derived from M23 162
For the HOLL soybean lines with S117N missense *FAD2-1A* alleles from 17D 165
Agronomic characteristics of HOLL lines 167
Oil and protein content... 167
Discussion 169
Figures 175
Tables 178
Liturature cited 182
CHAPTER 5 Identification of candidate genes encoding for three acyltransferase enzymes controlling the triacylglycerol biosynthesis in soybean seed.......................... 186
Summary 187
Introduction 188
Materials and methods 193

Database search for *Arabidopsis* sequences of *GPAT, LPAAT* and *DGAT* genes 193
Homologous sequence identification and transcriptional expression data mining 193
RNA isolation, reverse transcriptase reactions, and quantitative Real-Time PCR 194
Results 195
Discussion 199
Figures 203
Tables 208
Liturature cited 211
VITA 216

LIST OF ILLUSTRATION

LIST OF FIGURES

Figure Page

1.1 Generalized scheme for triacylglycerol (TAG) assembly in developing seeds of oleaginous plants 50

2.1 Characterization of mutations in the *FAD2-1B* alleles from soybean lines PI 283327 and PI 567189A 101

2.2 Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 1 103

2.3 Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 2 104

2.4 Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 3 105

2.5 Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 4 106

2.6 Seed oleic acid phenotype and *FAD2-1* genotype association analysis for F_2 seeds of population 4 107

3.1 Weblogo output of the amino acid conservation *FAD2* enzyme as part of the BLINK feature at NCBI using GI number 197111722 138

4.1 Seed linoleic (18:2) and linolenic acid (18:3) contents of high oleic low linolenic acid (HOLL) soybean lines developed with null allele ofM23 *FAD2*-1A in Portageville Missouri summer 2010 176

4.2 Seed linoleic (18:2) and linolenic acid (18:3) content of HOLL soybean lines developed with either M23 or 17D *FAD2-1A* alleles in Columbia summer 2010 177

5.1 Generalized scheme for triacylglycerol (TAG) assembly in developing seeds of oleaginous plants 204

5.2 Relative expression of putative glycerol-3-P acyltransferase (*GPAT*) genes in leaf and seeds at different sizes of cultivar Williams 82 205

5.3 Relative expression of putative lysophosphatidic acid acyltransferase (*LPAAT*) genes in leaf and seeds at different sizes of cultivar Williams 82.. 206

5.4 Relative expression of putative diacylglycerol acyltransferase (*DGAT*) genes in leaf and seeds at different sizes of cultivar Williams 82 207

LIST OF TABLE

Table Page

2.1 Oleic acid content, standard deviation, and seed generation of soybean lines with different combinations of mutant *FAD2-1A* and mutant *FAD2-1B* produced in three environments.. 108

2.2 Fatty acid profiles (means and standard deviations) for different homozygous *FAD2*-1 genotypes in four segregating populations developed by crosing soybean lines carrying different sources of mutant *FAD2-1A* alleles with different sources of mutant *FAD2*-1B alleles 109

3.1 Twenty two soybean lines selected for cloning and sequencing of *FAD2*-1A and *FAD2*-1B genes.. 139

3.2 Variants in DNA sequences of *FAD2-1B* of 24 tested soybean lines 141

3.3 Fatty acid profiles and protein and oil contents for homozygous mutant *FAD2-1A and FAD2-1B* genotypes of population KB09-13 (population 1) in field trials in Portageville and Columbia in summer 2010................................ 142

3.4 Fatty acid profiles for different homozygous *FAD2-1* genotypes in population KB09-35 (population 2), the two parents, two soybean lines with high oleic acid content from population 1 and control in Columbia MO summer 2010 144

4.1 Gene combination of HOLL lines, parental lines and control lines used in the study 178

4.2 Fatty acid profiles and protein and oil contents for high oleic low linolenic soybeans with M23 *FAD2-1A* allleles in field trials in Portageville and Columbia in summer 2010 180

4.3 Fatty acid profiles for different homozygous *FAD2-1* genotypes in population KB09-35 (population 2), the two parents, two soybean lines with high oleic acid content from population 1 and control in Columbia MO summer 2010 181

5.1 Information of homologous gene of *GPAT*, *LPAAT* and *DGAT* genes used in the RT-PCR analysis 209

5.2 Expression levels of all soybean homologous genes of *GPAT*, *LPAAT* and *DGAT* genes in leaf and different stages of seed development obtained from microarray data in soybase database 210

CHAPTER 1

Literature review

Soybean domestic and world production: Soybean [*Glycine max* (L). Merrill] is one of the most important oilseed crops in the world. Soybean originated from China with the most ancient archaeological evidence dating back to the 11th BCE (Hymowitz and Shurtleff, 2005). Soybeans were brought to Savannah, Georgia, United States from China in 1765 by Samuel Bowen, a seaman employee of the East India Company. However, the crop was not grown in the field in the United States until 1851 and was mostly used as a forage for livestock for nearly 100 years afterward (Hymowitz, 1990). In 1917, Osborne and Mendel claimed that properly cooked soybean meal has superior nutritional values compared to unheated soybean meal. This finding established the value of soybean seed meal for potential human consumption and the development of a soybean processing industry (Hymowitz, 1990). To date, soybean represents the largest oilseed crop in the world with 260 million tons recorded in April 2011, contributing 59% of the total oilseed production (Oilseeds: World Markets and Trade, Foreign Agricultural Service, USDA, http://www.fas.usda.gov/oilseeds/circular). The majority producer is the U.S., with 100 million tons. For more than 10 years, the U.S has maintained its position as the world"s single largest soybean producing country, providing 46% of world production in 2000 and 38% in 2010. In the U.S., soybean has also dominated other oilseed crops, with 74 million acres planted in 2009 (30% of the total U.S. crop area). Soybean production totaled 91.48 million tons, ranking it the second most valuable agricultural export in the United States behind corn and first among oilseed crops (soystats.com).

Soybean products: Soybean seeds constitute 40% protein, 20% oil, 35% carbohydrate and 5% ash (Lui, 1997). In addition, soybeans are also rich in calcium,

which benefits bone health, and isoflavones, which are found to play a role in cancer prevention and relief of menopausal symptoms (Messina, 1999). To date, it is estimated that more than 200 categories of both edible and industrial products are produced from soybean (soystats.com). Soybean has been the largest provider of vegetable oil and vegetable protein for human food and livestock feed. Furthermore, soybean is the leading source of biodiesel, accounting for 80% of the domestic biodiesel production of the U.S. Soybean is also present in a wide range of industrial products like cosmetics, plastics, inks, pesticides, clothing, spraying oils and lubricant, among others.

One of the most valuable products from soybean seed is the oil. To remove oil, soybean seeds were first cleaned, and subsequently cracked, rolled into flakes and solvent-extracted with hexane (soystats.com). The materials remaining after extracting the oil is processed further to produce soybean meal for swine and poultry feed, or soy concentrate and soy protein isolate (Lui, 1997).

Soy foods: Soymilk is the beverage made by soaking dry soybean seeds and then grinding them with water until they become liquid. An eight ounce glass of soy milk has 30% more protein and 50% less fat compared to whole cow milk (soyfoods.com, USDA nutrient database at http://www.nal.usda.gov/fnic/foodcomp/search/). Other well known products in liquid form made from soybean are miso and soy sauce (Hymowitz and Newell, 1981). Miso is a rich, salty paste that characterizes the essence of Japanese cooking. It is made by fermenting a mixture of soybeans, another grain such as rice, salt and a mold culture that may contain fungi such as Aspergillus oryzae (Ahlburg) Cohn for one to three years (Hesseltine, 1983; Hymowitz, 1970). The Japanese make miso soup and use it to flavor a variety of foods in their daily meals. Similar to miso, soy sauce is a

condiment produced by fermenting a mixture of soybeans with another grain and Aspergillus oryzae or Aspergillus sojae molds, plus water and salt (Hesseltine, 1983). The fermentation first yields moromi, and soy sauce is the liquid product obtained after the moromi is pressed. All varieties of soy sauce are salty, brownish liquids used for seasoning food while cooking or at the table. Soy sauce is a traditional ingredient in East and Southeast Asian cuisines, and more recently it is also being used in Western cuisine and prepared foods. The high protein concentration in miso and soy sauce is one of the reasons that made them become the most popular ingredient used in cooking for populations in East and Southeast Asia.

Another popular food made from soybean is tofu, which is one of the most important sources of protein in daily meals for populations in East and Southeast Asia (Hymowitz and Newell, 1981). Tofu, also known as soybean curd, is produced by curdling hot soymilk using a coagulant. Because it is rich in high-quality protein, B-vitamins and low in sodium, tofu has been a favorite meat substitute for Asian vegetarians. Tofu can be found in three different forms: firm, soft, or silken, which can subsequently be processed in a variety of ways for different dishes.

Soy meal is the high-protein portion that remains after the oil is removed from the seeds (Lui, 1997). This fiber like material is often toasted and prepared into animal feed for poultry, pork, cattle, other farm animals and pets. Soy protein is increasingly found in fish food, both for home aquariums and for the fish grown for eating due to the scarcity and increasing cost of fish meal. Around the world, soy protein may be found in feed for most animals due to its competitive low price and high quality protein content.

Soybean oil uses and trends: In the 2009/10 growing season, nearly 38.7 million tons of soybean oil were produced worldwide, constituting about a quarter of worldwide edible vegetable oil production (Oilseeds: World Markets and Trade, Foreign Agricultural Service, USDA). The U.S. is the single biggest soybean oil producer in the world with approximately 9.3 million tons in 2010, representing approximately a quarter of the total soybean oil produced worldwide. In the 1950s, the majority of soybean oil (95%) was used for human consumption in the forms of margarine (12%), cooking and salad oils (46%), or baking and frying fats (37%). In addition to vegetable oil, soybean oil can be found in bakery and manufactured foods such as baked breads, crackers, cakes, cookies and pies etc. More recently, with the advance of technology, more soybean oil is being utilized for industrial purposes (18%), at the expense of decreased percent of soybean oil used for baking and frying fats (25%). Some of the newly developed products in the industry using soybean oil as a component include biodiesel, polyols (for the production of polyurethane products such as foams and resins), renewable bio-lubricants, renewable plasticizers and soy-based toner (http://www.associationdatabase.com/aws/OHSOY/pt/sp/osc_newuses). Utilization of soybean oil in these products is expected to create environmentally friendly alternatives to the conventional products and is also very cost effective.

Vegetable oil properties: The term “vegetable oil” is dedicated to oil extracted from seeds or fruits of oilseed plants in commercial practice, although different parts of plants may also contain oil. This term also indicates oils that remain in liquid form at room temperature, contrasting to fats (either originally from vegetable or animal) which are solid. Some vegetable oils like rapeseed and castor bean oils would not be used for

human consumption without further processing to get rid of harmful components. Most vegetable oils are composed of triacylglycerols, which constitutes a glycerol backbone attached to three fatty acids. Therefore, the fatty acid composition determines the chemical and physical quality of vegetable oil (Ensminger and Ensminger, 1993). The five major fatty acids commonly found in vegetable oils contain 16-18 carbons in their molecules including palmitic (C16), stearic, oleic, linoleic and linolenic (C18) (Kim et al., 2010a), though short chain fatty acids (C8-caprylic acid, C10-capric acid, C12-lauric acid, and C14-myristic acid) can be found in coconut and palm oils and longer chain fatty acids (C20- eicosanoic acid or C22-erucic acid) can be found in natural rapeseed oils or peanut oils (White, 2007). Depending on the number of double bonds in the molecules, fatty acids can be classified into saturated fatty acid (no double bond), monounsaturated fatty acid (one double bond), and polyunsarturated fatty acid (two or more double bonds) (Scrimgeour, 2005). Oils from palm, coconut and palm kernel have a high percentage of saturated fatty acids; therefore, they are more resistant to rancidicity, can last up to two years, and are stable at high temperatures. However, one of the disadvantages of oils high in saturated fatty acids is that the consumption of a diet high in saturated fatty acids has been linked to higher risks of cardiovascular diseases (Artaud-Wild et al., 1993). Oils high in monounsaturated fatty acid (oleic acid) like olive oil and canola oil have been demonstrated to be stable to oxidation and have various health benefits (Covas et al., 2006; Oomah and Mazza, 1999; Waterman, 2007). Oils high in polyunsaturated fatty acids include cotton seed oil, sunflower oil, soybean oil and safflower oil, etc. These oils are nutritionally healthy because linoleic and linolenic acids are essential fatty acids that the human body cannot synthesize on its own so they must be obtained from food or

supplements. However, because double bonds in the molecules of these fatty acids are more susceptible to oxidation (White, 2007), oils high in polyunsaturated fatty acid have low oxidative stability and frying stability resulting in quick rancidity, a rapid decrease in optimum flavor, and shortened storage time of manufactured food products (Warner and Fehr, 2008)

For biodiesel production, vegetable oils are characterized by numerous parameters including viscosity, density, cetane number, cloud and pour points, distillation range, flash point, ash content, sulfur content, carbon residue, acid value, and copper corrosion according to ISO norms (Ali et al., 1995; Harrington, 1986). Some of the properties of vegetable oils such as high viscosity, low volatility and poor cold flow properties can cause severe engine deposits, injector coking and piston ring sticking if it is directly used as biodiesel. However, several practices have been suggested to improve the fuel properties of vegetable oils such as pyrolysis, microemulsification, dilution, and transesterification (Srivastava and Prasad, 2000). Biodiesel can be blended with hydrocarbon-based diesel fuels at any ratio. Biodiesel not only has appropriate viscosity and boiling point, but also is simple to use, nontoxic, biodegradable, and essentially free of sulfur, halogen and aromatics (Ryan III et al., 1984; Srivastava and Prasad, 2000)

Soybean fatty acid profile: The average fatty acid composition of soybean oil is 11% palmitic acid (16:0), 4% stearic acid (18:0), 23% oleic acid (18:1), 55% linoleic acid (18:2), and 8% linolenic acid (18:3) (Wilson, 2004). This profile of soybean oil has been modified depending on the end products of soybean oil to meet different demands of the markets (Fehr, 2007).

Soybean oils with more than 50% linolenic acid of the total fatty acid were developed by overexpression of the fatty acid desaturase 3 (*FAD3*) gene, which was suggested to be good for production of drying oils for paint or salad oils (Cahoon, 2003). Accessions of wild soybean, *G. soja*, contain up to 16% linolenic acid (Wilson, 2004). Soybean oils that are high in linoleic and linolenic acid content offer health benefits because these two fatty acids are essential fatty acids that humans must incorporate from their diets. Moreover, they are important for proper structure and function of multiple body systems such as brain and eye (Crawford et al., 1981; Uauy et al., 2001). On the other hand, the high concentration of polyunsaturated fatty acids is responsible for low oxidative stability of soybean oil and quick rancidity and off flavor in soybean oil-processed foods (Warner et al., 1997). Therefore, for vegetable oil and food production purposes, soybean oil that is high in oleic acid content or low linolenic acid content is preferred. Soybeans with more than 80% oleic acid content or 1-3% linolenic acid content or the combination of the two traits were produced (Bilyeu et al., 2011; Brace et al., 2011; Buhr et al., 2002; Pham et al., 2010). In addition, conventional breeding was successful in the generation of soybeans that are high in saturated fatty acid (palmitic and stearic acids totaled) up to 40% and low in saturated fatty acid content down to 7% (Fehr et al., 1991; Stoltzfus et al., 2000) for different utilization purposes. The low saturated fatty acid soybean is suitable for production of low saturated fat containing foods because diets high in saturated fatty acid have been linked to higher risk of cardiovascular diseases (Artaud-Wild et al., 1993). The high saturated fatty acid soybean is desired for some special food production because saturated fatty acids have no double bonds, hence they resist oxidation and make the oil and foods last longer without any off-flavors. The

combination of two oil traits in a single line has been achieved with promising results: 23% palmitic acid and 20% stearic acid (Kok et al., 1999); 25% palmitic acid and 1% linolenic acid (Bravo et al., 1999); 4% palmitic acid and 50% oleic acid (Rahman et al., 2004); 26% stearic acid and 61% oleic acid (Booth et al., 2005). Oils with elevated oleic acid and stearic acid could be used to produce oil for use in confectionary applications and other products requiring a high level of stability such as margarine and other spread products (Booth et al., 2005).

Soybeans with elevated oleic acid content: Soybean with elevated levels of oleic acid content was identified by screening the soybean germplasm collection, treating seeds with X-ray or Ethane methyl sulfonate (EMS) and conventional breeding. Plant introductions (PI) with oleic acid content ranging from 27% to 50% were selected; however, the oleic acid levels in these PIs are not as stable as the mutation induced sources (Lee et al., 2009). X-radiation and chemical mutagenesis was successful in creation of soybean lines with mid oleic acid content ranging from 30-50% of the total fatty acid content such as M23, KK21, and 17D (Dierking and Bilyeu, 2009; Rahman et al., 1996a). Finally, using conventional breeding, soybean lines with oleic acid content from 50-70% were also released such as N00-3350, N98-4445A or F22, although the oleic acid contents of these lines varied greatly across environments (Alt et al., 2005b; Fehr, 2007; Oliva et al., 2006). The molecular genetics and mechanism underlying the elevated oleic acid content is largely unknown except for a few cases. Some PIs with mid oleic acid contents were reported to have down-regulated expression levels of *FAD2-1A*, *FAD2-1B*, and oleate-ACP thioesterase (*GmFATB1a*) genes and/or up-regulated expression levels of delta-nine stearoyl acyl carrier protein desaturase A, B and C

(G*mSACPD*) genes (Upchurch and Ramirez, 2010). M23 had a deletion of a 164 kilobase region containing the *FAD2-1A* gene while KK21 and 17D have missense mutations in *FAD2-1A* gene occurring at conserved regions which probably affect the *FAD2-1A* enzyme"s function (Anai et al., 2008; Bolon et al., 2011; Dierking and Bilyeu, 2009). In other cases, mapping studies revealed that the oleic acid content in soybean lines N00-3350 or N98-4445A or was controlled by at least 6 QTLs (Bachlava et al., 2009; Monteros et al., 2008). The usage of these mid oleic acid soybean cultivars and plant introductions for production in reality has been very limited as they carry several disadvantages such as unstable phenotype, yield drag (as shown for M23), unacceptable agronomical traits (due to the PIs" background) and the genetic complexity underlying the trait which would be challenging for soybean breeders (in case of N00-3350 or N98-4445A or the PIs) (Oliva et al., 2006; Scherder and Fehr, 2008)

Transgenic high oleic (HO) soybeans: Transgenic high oleic acid soybeans with more than 80% oleic acid content were first created by inserting a copy of the *FAD2-1* gene into the soybean genome to induce downregulation of the native *FAD2-1* transcripts (Knowlton, 1999). After that, using ribozyme-terminated antisense and standard antisense constructs to post-transcriptionally suppress the expression level of *FAD2-1* genes, Buhr et al. generated transgenic soybean lines with 57-82% oleic acid contents. Recently, transgenic HO soybean were produced by introducing a hairpin suppression cassette containing a region of 100-400 bp of the intron of the *FAD2-1A* gene to trigger the RNAi process (Mroczka et al., 2010; Wagner et al., 2010). It was shown by the study that the suppression by this RNAi-like mechanism is not an on/off process but a gradual process and an intermediate state can be achieved. However, the intron region used for triggering

the RNAi suppression must be larger than 100 bp. Compared to soybeans with elevated or mid oleic acid content, HO transgenic soybeans have a higher and more stable oleic acid content. In addition, their agronomic traits and yield are not adversely impacted by the high oleic acid content (Graef et al., 2009).

High oleic acid and yield: The affect of elevated oleic acid content to yield has been reported but the results are controversial. Kinney and Knowlton (1998) claimed that the high oleate (HO) trait did not have any negative effects on yield or other agronomic traits. Similarly, evaluation of transgenic high oleic acid and low palmitic acid soybeans (HOLP) in both irrigated and non- irrigated fields showed that either yield or seed composition of the HOLP event was not compromised (Graef et al., 2009). On the other hand, M23-derived mid oleic soybean lines have been shown to have reductions in yield from 9-15% compared to soybean with conventional oleic acid content, and this affect was predicted to be linked to the deleted portion of chromosome 10 in M23 (Scherder and Fehr, 2008). Recently, Brace et al. 2011 reported the same adverse affect of transgenic HO soybean (80%) to yield and oil content but at smaller magnitude (Brace et al., 2011). Yield reduction for high oleic soybean was in a range of 3.6-5% and reduction in oil was about 3%. Other traits such as maturity, lodging, height, and seed weights in transgenic HO were not significantly different from those of conventional soybeans. Because the transgenic HO soybeans were previously found to yield 4-5% less than the untranformed parent “Jack” in six environments with three replications although the statistics indicated the differences were not significant; Brace et al. 2011 suggested further testing these HO soybean lines to the determine if yield reduction resulted from undesirable pleiotropic effects of the HO transgene or undesirable linkage between the

transgene and other genomic components. The difference in performance of transgenic HO soybean in the Graef et al. study and Brace et al. study may also be due to the difference in testing locations, temperature and maturity of the plants.

Transgenic high oleic low linolenic acid (HOLL) soybean: HOLL soybeans were generated by incorporating two loci for low linolenic acid including mutant Gm*FAD3A*(C1640) (Wilcox et al., 1984) and Gm*FAD3C* (RG10) (Reinprecht et al., 2009) into an HO transgenic line developed by Kinney et al. 2008. The HOLL lines had oleic acid content of 78.5% and linolenic acid content of 2.4% in average across four growing regions in the Midwest U.S (Brace et al., 2011). However, the transgenic HOLL soybeans in that study were shown to have unwanted properties such as lower levels of oil content, seed weight and yield compared to a control cultivar with normal fatty acid composition. Because lower yield in HOLL lines compared to conventional oil lines, the results contrast with previous reports, which showed that the high oleic low linolenic acid did not affect yield. The authors recommended that further studies needed to be done to confirm the difference of HOLL lines compared to normal oil soybean.

High oleic trait in other crops: High oleic acid oils were produced in *Arabidopsis* and oilseed crops with great economic importance by different approaches: traditional breeding using germplasm lines with elevated oleic acid content or mutagenized lines and genetic engineering.

In *Arabidopsis*, high oleic acid level was first found to be caused by a T-DNA insertional mutation in the fatty acid desaturase 2 (*FAD2*) gene (Okuley et al., 1994). After that, **gene-**silencing of the *FAD2* gene of *Arabidopsis* using RNAi (intron-spliced hairpin construct) caused the 18:1 substrate accumulation from 17% to 53% (Stoutjesdijk

et al., 2002), while using artificial trans-acting small interfering RNAs also succeeded in increasing the oleic acid content in seed (Gutiérrez-Nava et al., 2008)

In canola (rapeseed), the value of a high oleic acid trait was recognized and appreciated early; hence, breeding efforts have been spent on developing mutant lines to produce high oleic acid content in canola up to 86% (Schierholt et al., 2001). Using EMS, several mutant lines with elevated oleic acid content were also developed with up to 76% oleic acid (Spasibionek, 2006). It was demonstrated in a high oleic acid *Brassica rapa* line that a single nucleotide polymorphism (SNP) in the *FAD2* DNA sequence resulting in an amino acid (aa) change from leucine to proline in protein sequence was responsible for the high oleic acid content up to 90% (Tanhuanpää et al., 1998). In another study, the high oleic acid content in the seed oil of a *Brassica napus* line was found to associate with a mutation in a *FAD2* gene that created a stop codon resulting in a truncated protein of *FAD2* (Hu et al., 2006a). By down-regulating the expression level of *FAD2* gene using RNAi constructs or antisense *FAD2* gene constructs, high oleic acid canola lines with 83- 88 % oleic acid content were successfully generated (Peng et al., 2010; Töpfer et al., 1995).

In sunflower (*Helianthus annuus*), treating seeds with dimethyl sulphate (DMS) resulted in sunflower mutant plants with 75% oleic acid content in seed oil compared to 10-14 % in wild type seeds. The mutation in a high oleic acid sunflower line Pervenets was identified to be the duplication of the *FAD2-1* gene, leading to silencing of the gene in trans (Hongtrakul et al., 1998; Lacombe et al., 2009).

In *peanut (Arachis hypogaea L.),* high oleic peanut lines were mainly generated by seed treatment with chemicals like EMS or DMS. In a mutant high oleic acid peanut

line, F435, the change in amino acid sequence of the *FAD2*A gene from aspartate^{150} to asparagine, together with a reduction in the transcription level of the high activity gene *FAD2*B are causative for the high oleic acid (Bruner et al., 2001; Jung et al., 2000). Furthermore, in two chemical induced high oleic acid peanut lines, miniature inverted-repeat transposable element (MITE) insertions were identified to create stop codons in the coding sequence of the peanut *FAD2B* genes. This mutation when combined with the point mutation in *FAD2A* described by Jung et al. 2000 created peanuts with oleic acid content in seed oil up to 80% (Patel et al., 2004).

In cotton (*Gossypium hirsutum*), utilization of a homologous promoter to drive the expression of an antisense *FAD2* gene construct to down regulate this gene in seed resulted in a doubled oleic acid content (Sunilkumar et al., 2005). Another effort to silence *FAD2* gene expression by hairpin RNA-mediated post-transcriptional approach resulted in cotton with 77% oleic acid content in the seed oil (Liu et al., 2002).

In maize (*Zea mays* L.), a substitution at a highly conserved position leucine^{71}-to-threonine was found to negatively affect the activity of the *FAD2* enzyme, which caused the oleic acid content in seed oil to increase from 30% in wild type to 60% in the variant accessions (Beló et al., 2008)

Crops with naturally high in oleic acid content: In nature, vegetable oils extracted from nuts (except walnut), olive and canola naturally often have high oleic acid contents (more than 50% of the total fatty acid) (Beuchat and Worthington, 1978; White, 2007). Among those, olive and canola oils are the most common vegetable oils, while oils originating from nuts represent a small portion in the market and are mostly used in skin therapy, aromatherapy and beauty products. Olive (*Olea europaea*) oil has a healthy

fatty acid composition with 15% saturated fatty acids, 75% oleic, 9% linoleic and 1% linolenic and an abundance of phenolic antioxidants (Owen et al., 2000). Archeological evidence suggested that olive oil was cultured about 6000 years ago in Mediterranean Coast regions (Galili et al., 1997). Not only being a healthy and popular cooking ingredient for the Mediterranean population, it was also used for religious rituals, medicinal and magical purposes, and it was used a symbol for wealth and power in ancient times (Galili et al., 1997). Currently, Spain is the largest olive oil production country, and together with Italy and Greece, it produced approximately 75% of the total olive oil in the world (Vossen, 2007). In the U.S., the largest production area is in California with 400,000 gallons in 2004-05, though this only accounts for 0.6% of the domestic consumption (Vossen, 2007). Consumption of olive oil has been demonstrated to bring enormous health benefits including lower total cholesterol and low-density lipoprotein levels in the blood, lower blood sugar levels and blood pressure (Waterman, 2007). Additionally, several preliminary studies reported the preventive capacity of olive oil against peptic ulcer, gastric cancer, DNA and RNA damage, and tumor and skin cancer (Budiyanto et al., 2000; Ferrara et al., 2000; Machowetz et al., 2007). In addition to the usage as a food, olive oil has become one of the most common ingredients used in skin therapy, cosmetics, pharmaceuticals and medicinal uses (Waterman, 2007).

Canola oil is another alternative source of high oleic acid oil with 61% oleic acid content, 7% saturated fatty acid and 32% polyunsaturated fatty acids. Canola oil is extracted from seeds of two rapeseed species (*Brassica napus* L. and *B. rapa* L.) (Raymer, 2002). Natural rapeseed oil contains 50% erucic acid, and is also high in glucosinolates, which is not healthy for animal feeds. Through the breeding process, the

content of these acids was lowered to an acceptable level. The name "canola" was derived from the term "**Can**adian **o**il, **l**ow **a**cid" used to label the seeds by the Manibota government in 1978 (Raymer, 2002). Consumption of canola oil was shown to improve cardiovascular health and lower LDL and cholesterol level, to reduce blood pressure, plasma glucose and ventricular fibrillation (Simopoulos, 1999; Valsta et al., 1992).

Although these oils are nutritiously healthy, the expensive price of olive oil and canola oil compared to other oils such as soybean oil has made the usage of these oils in the food production industry rather limited in the U.S. In 2006, canola oil‟s price was $700/ton (http://www.canolacouncil.org/canolaprices.aspx), while olive oil „s was $5500-6000/ton (http://www.mongabay.com/commodities/price-charts/olive-oil-price.html) compared to the price of $480/ton of soybean (National Monthly Feedstuff Prices, Agricultural Marketing Service, USDA and Global Agricultural Trade System, Foreign Agricultural Service, USDA and Oilseed Crushings, U.S. Census Bureau). In 2009-10, average olive oil‟s price was nearly $3029/ton, while soybean‟s was $354/ton, and canola‟s was $941/ton.

Trans fat labeling 2006: Trans fats were generated as an unexpected outcome of the hydrogenation process of crude soybean oil in order to decrease the amount of polyunsaturated fatty acids and improve the oxidative stability of soybean oil. The patent of the fat hardening technology by hydrogenation was filed in Britian in 1903 by a pharmacist Wilhelm Norman. Proctor & Gamble Company saw the potential and bought the patent from Mr. Norman. They used the protocol to produce Crisco in America, a cooking fat that has a long lifetime and is good for baking. During World War II and the Great Depression, the economic hardship pushed Americans to use the economic

products like partially hydrogenated Crisco vegetable oil to supplement or substitute rationed butter. Since 1948 when the first McDonald"s restaurant was opened, the diet of American people has been deeply involved with trans fat, without knowing about the impact of this fat. However, during the 1970s-1990s decades, scientific evidence has accumulated the influence of trans fat on human health risks for coronary, heart diseases and high levels of cholesterol (Ascherio and Willett, 1997; Hu et al., 1997; Mozaffarian et al., 2006). Discerning the health impacts of trans fat, in July 2003, the FDA adopted a rule to make trans fat to be labeled on food products beginning January 2006. The appearance of the amount of trans fat on food labeling was hoped to raise the awareness of consumers about the harmful fat intake, and be a strong incentive to reduce or eliminate the trans fat in the diet of American people.

Fatty acid synthesis and desaturase pathway: Fatty acid synthesis takes place only in the chloroplast, while the desaturation of the fatty acids occurs in both the chloroplast and Endoplasmic Reticulum (ER) (Ohlrogge and Browse, 1995). The central carbon donor for the fatty acid synthesis is malonyl-CoA. Malonyl-CoA is synthesized using acetyl-CoA and CO_2 as substrates for the enzyme acetyl-CoA carboxylase. After that, malonyl-CoA:ACP transacylase transfers malonyl from CoA to ACP to form malonyl-ACP. ACP is a small (9 kD) acidic protein that contains a phosphopantethein prosthetic group to which the growing acyl chain is attached as a thioester. Malonyl-ACP is the carbon donor for later condensation steps to produce fatty acids with two carbons longer than before the condensation steps. Malonyl-ACP is condensed with Acetyl-CoA to produce a 3-ketoacyl-ACP. After each condensation, the 3-ketoacyl-ACP is reduced to remove the carbonyl group, dehydrated, and reduced again using three enzymes: 3-

ketoacyl-ACP reductase,3-hydroxyacyl-ACP dehydratase, and enoyl-ACP reductase, respectively. These three enzymes are the easily dissociable multisubunits which constitute the fatty acid synthase enzyme complex and use NADPH as the electron donor (Mroczka et al., 2010). The condensation steps continue until a 16 or 18 carbon fatty acid is produced. There are three condensing: KAS III (3-ketoacyl-ACP synthases) catalyze the first condensation of acetyl-coA and malonyl-ACP to form a four carbon product. KAS I is responsible for catalyzing condensation steps to produce chain lengths from six to 16 carbons. The generation of stearoyl –ACP from palmitoyl-ACP was catalyzed by a different enzyme, KAS II. To produce the mono- and polyunsaturated fatty acids, a double bond is introduced into stearoyl-ACP by stearoyl-ACP desaturase. This enzyme is the only desaturase that is soluble in the plant kingdom. The synthesis of fatty acids in the plastids is finished when the acyl group is removed from ACP. This reaction happens either when the long chain acyl groups are hydrolysed by an acyl- ACP thioesterase and release free fatty acids to be imported outside of the plastid or the fatty acids are transferred from ACP to glycerol-3-phosphate or to monoacylglycerol-3-phosphate by two acyltransferases in the plastid (Kim et al., 2007). On the outer membrane of the chloroplast envelope, free fatty acids are finally incorporated to coenzyme A (CoA) esters by an acyl-CoA synthetase and exported to the endoplasmic reticulum. Polyunsaturated fatty acids are subsequently produced through the subsequent desaturation reactions catalyzed by the membrane-bound desaturase enzymes of either the plastids or the ER. In the eukaryote biosynthesis pathway, the FAD2 and FAD3 desaturase enzymes which bind to the endoplasmic reticulum‟s membrane act on the fatty

acids esterified to the first and second position of phosphatidyl choline (PC) (Kim et al., 2007; Spasibionek, 2006)

TAG assembly pathway: Oil is stored in the form of triacylglycerol (TAG) in the developing oilseeds. The TAG assembly stage occurs in the endoplasmic reticulum after the fatty acid synthesis process finishes and creates the Acyl-CoA pool in the ER. Fatty acyl groups are incorporated into glycerol-3 phosphate backbone via the Kennedy pathway, which is the sequential acylations of glycerol-3-phosphate to produce TAG (Figure 1.1). In the first acylation reaction, glycerol-3-P acyltransferase (*GPAT*) catalyzes the esterification of sn-glycerol-3-phosphate by a fatty acid coenzyme A ester at position sn-1 to form lysophosphatidic acid (Figure 1.1). The second acylation reaction is catalyzed by lysophosphatidic acid acyltransferase (*LPAAT*) enzyme to the position sn-2 to produce phosphatidic acid (PA). Subsequently, a phosphatidic acid phosphatase participates in the dephosphorylation of PA to form sn-1,2 diacylglycerol (DAG), an immediate precursor of TAG production. Finally, diacylglycerol acyltransferase (DGAT) catalyses the acyl-coenzyme A (acyl-CoA)-dependent acylation of DAG to generate TAG. However, TAGs could also be produced via an acyl-CoA independent pathway, in which the acyl groups can be transfered to lysophosphatidylcholine via the enzyme lysophosphatidylcholine acyltransferase. The enzyme phospholipid:diacylglycerol acyltransferase (PDAT) will then transfer the acyl groups to DAG to generate TAG (Raneses et al., 1999; Stahl et al., 2004).

FAD2/6 in Arabidopsis and other species: *FAD2* and *FAD6* genes are often called omega-6 fatty acid desaturase, and the enzymes they encode are called delta12-fatty acid dehydrogenase. This is because *FAD2*/*FAD6* genes catalyze the addition of a

double bond to the sixth carbon of the oleic acid molecule from the end without the carboxyl group (or the twelfth carbon counting from the carboxyl group end) to make linoleic acid. *Arabidopsis* only habors one *FAD2* gene targeting the enzyme to the ER and one plastidial targeted FAD6 gene. The *FAD2* gene was first characterized by Okuley et al. 1994 using T-DNA insertion. The *FAD2* gene contains a single large intron (1134bp) just four basepairs downstream from the start codon and lacks the lysine rich carboxyl terminal motifs that are characteristic of integral membrane proteins in the ER. *FAD2* uses 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine as a substrate and it requires oxygen, NADH, NADH:Cytochrome b5 oxidoreductase and Cytochrome b5 to complete the reaction (Smith et al., 1990; Stymne and Appelqvist, 1978; Vijay and Stumpf, 1972).

The *Arabidopsis FAD6* gene was found to contain distinctive properties compared to the *FAD2* gene (Burton et al., 2004). The *FAD6* cDNA encodes a 418 amino acid peptide and shares a higher similarity with the desA 12 desaturase enzyme of cyanobacterium Synechocystis (54%) compared to the *Arabidopsis FAD2* gene (21%). Instead of having a signal recognizing peptide which is a characteristic of chloroplast-localized proteins, this enzyme has a long extension at the amino terminal which resembles those in the omega 3 desaturases FAD7 and FAD8 protein sequences. However, the FAD6 enzyme retains three characteristic histidine-rich motifs that are thought to be important for possible iron coordination in the enzyme active site of desaturase enzymes. Expression of this gene in mutant *FAD6 Arabidopsis* lines rescued the mutant oleic acid content in leaf, which indicates that this enzyme is capable of catalyzing the desaturase reaction.

FAD6 genes from soybean and *B.napus* were characterized by Hitz et al., 1994. Using a polypeptide sequence from a desaturase gene from *Arabidopsis*, one ω-6 desaturase gene was identified for either soybean or rapeseed. Because both of the enzymes encoded by these two genes possess a characteristic transit peptide sequence for the chloroplast, they were suggested to be FAD6 enzymes. The soybean ω-6 desaturase contains 424 aa while the rapeseed"s contain 443 aa. Expression of the identified rapeseed ω-6 desaturase gene in a yeast strain incapable of producing polyunsaturated fatty acids yielded polyunsaturated fatty acids in the yeast cell, confirming the function of this gene in vivo.

The number of *FAD2* genes in canola was estimated to be four to six, and four of these genes were mapped on four different linkage groups, two of A- (*Brassica rapa*) and two of C-genome (*Brassica oleracea*) origin (Scheffler et al., 1997). One actively transcribed copy of these *FAD2* genes was identified to be in the linkage group N5 (Schierholt et al., 2000). This gene encodes a 384 aa protein and a single nucleotide deletion in this gene was found to be responsible for the increased oleic acid content in the seed oil of a chemical induced mutant canola line (Hu et al., 2006).

Cotton was reported to harbor at least five *FAD2* genes in its genome, although only four homologous *FAD2* genes were thoroughly characterized (Zhang et al., 2009). The *FAD2-1* was partially analyzed and reported by an Australian group in 1999 and *FAD2*-3 was characterized by using cloning and physical mapping in 2001(Liu et al., 1999; Liu et al., 2001; Pirtle et al., 2001). *FAD2-1* gene has one intron with 1133 basepairs (bp) located 9 bp downstream from the initiation codon. The *FAD2-3* has an intron of 2967 bp in the 5"-untranslated region, just 11 bp upstream from the initial codon

ATG, and one single exon encoding 384 amino acids. This enzyme also has the three histidine-rich motifs which are conserved among plant membrane desaturases. When the *FAD2*-3 homologous gene was expressed in yeast, the transformed yeast cells were capable of catalyzing the reaction from 18:1 to 18:2. Recently, another *FAD2* homologous gene, *FAD2-4*, was identified by Zhang et al., 2009. This novel gene encodes a protein that shares a similarity of 98% with the *FAD2*-3 protein. Similar to *FAD2*-3, the 5′-UTR intron of the cotton *FAD2*-4 has a large 5′-UTR intron (2780 bp) located 12 bp upstream of the start codon. This large intronic region is one of the characteristics that distinguish *FAD2*-3 and *FAD2*-4 genes to *FAD2*-1 gene. The *FAD2*-4 green fluorescent protein (GFP) fusion polypeptide appeared to be localized in the ER of transgenic *Arabidopsis* plants. Transformation of this gene in a mutant yeast strain incapable of producing polyunsaturated fatty acids rescued the phenotype suggesting that this gene should play a role in controlling the unsaturated fatty acid composition in cotton seeds. Gene expression data indicated that there is a differential expression pattern among the four cotton *FAD2* genes: *FAD2-1* was seed and flower specific, *FAD2-2* was expressed in all other tissues except for roots, hypocotyls, stems, and fibers, and *FAD2*-3 and *FAD2-4* were ubiquitously expressed in all types of tissues (Zhang et al., 2009).

The number and location of *FAD2*/6 genes in maize was reported in 2003 by Mikkilineni and Rocheford, 2003. Using Southern blot, the authors suggested that there are at least three homologous *FAD2* genes in the maize genome located on chromosome 4, 5 and 10, and two of them were only expressed during embryogenesis. There are also two *FAD6* cDNAs sharing 81% similarity with the *Arabidopsis FAD6* gene that mapped to chromosome 1. The temporal transcript level of the *FAD2* genes that were expressed in

both embryo and endosperm was found to be steady at 14, 21, 28 and 35 days after pollination, with the highest peak at 14 days. One of the homologs of the maize *FAD2* genes was cloned and sequenced (Tao et al., 2006). This homolog contains 387 amino acids encoded by one single exon. RT-PCR analysis showed that this gene is highly expressed in maize immature embryos, but its mRNA also can be found in leaves, stems and roots with less abundance.

In peanut, two homologous genes of *FAD2* were reported and designated as ah*FAD2A* and ah*FAD2B*. Although these two genes encode for two 379 aa protein sequences that are different at only four amino acid residues, ah*FAD2B* was found to be more active than ah*FAD2B* (Bruner et al., 2001). It was demonstrated that in order to obtain a high oleic acid content up to 80%, both of the genes must be silenced or mutated (Jung et al., 2000; Patel et al., 2004).

In sesame (*Sesamum indicum*), there are at least two homologous genes of *FAD2* but only one of the homologs was characterized (Jin et al., 2001; Kim et al., 2006). Jin et al. reported the seed specific expression as well as some variations in the histidine-rich motifs of sesame *FAD2* compared to *FAD2* genes of other species. Later, Kim et al. suggested that the negative cis -regulatory elements in the sesame *FAD2* promoter and enhancers in the 5′-UTR intron in combination control the expression level of the seed specific homologous FAD gene.

Studies on FAD2 in soybean prior to this work: In 1996, Heppard et al. reported the presence of two genes in soybean: *FAD2-1* and *FAD2-2*. These two genes share amino acid sequence identity of 73% and different expression patterns as *FAD2-1* is more expressed in seed while *FAD2-2* is mostly expressed in vegetative tissues.

However, the transcript levels of both genes were not influenced by cold temperature. Subsequently, Kinoshita et al. (1998) found that the mid oleic acid content in soybean line M23 controlled by the ol locus is possibly due to the deletion of the *FAD2-1* gene. In 2007, Schlueter et al. reported that there are actually two homologous genes of *FAD2-1* and three homologous genes of *FAD2-2* due to two ancient genome duplication events. The expression levels of the *FAD2-1* and *FAD2-2* genes were similar to those reported by Heppard et al. The exception was *FAD2*-2*A*, for which expression was not detected; *FAD2-2A* was predicted to be nonfunctional as it has a deletion of 100 bp in the coding region (Schlueter et al., 2007). *FAD2*-1*A* and *FAD2*-1*B* are considered to play an important role in controlling the oleic acid level in developing soybean seeds and induced mutations in *FAD2-1A* were reported to associate with the elevated oleic acid content in soybean (Anai et al., 2008; Dierking and Bilyeu, 2008). As the sequencing project of the soybean genome was completed (Schmutz et al., 2010), sequences and location of these genes became more precise: the *FAD2*-2 desaturases consisting of *FAD2-2A* (Glyma19g32930), *FAD2-2B* (Glyma19g32940), and *FAD2-2C* (Glyma03g30070); the two microsomal *FAD2*-1 desaturases include *FAD2-1A* (Glyma10g42470) and *FAD2-1B* (Glyma20g24530).

FAD2-1A and *FAD2-1B* are most closely related to one another, with a shared genomic organization containing a single intron and 99% identity in encoded amino acid sequence, and are present on homologous chromosome regions mapped to linkage group O (chromosome 10) and I (chromosome 20), respectively (Li et al., 2008). Characterization of the expression of the individual soybean *FAD2* genes confirmed the importance of *FAD2*-1A and *FAD2*-1B with expression of these genes during peak oil

synthesis; a possible role was also revealed for *FAD2*-2C under cool temperature conditions (Schlueter et al., 2007). Transcript levels of *FAD2-1B* are more abundant than that of *FAD2-1A*, but the response to temperatures of the two genes was similar in fashion and magnitude, with transcript levels reduced at higher temperatures (Byfield and Upchurch, 2007). At the enzyme level, *FAD2-1B* was found to be significantly more stable than *FAD2-1A* at 30 ^{o}C due to the possession of some unique amino acid residues in the region 241–334 in the C terminus, and the two enzymes were found to be regulated by phosphorylation (Tang et al. 2005).

Further desatuation by FAD3: *Arabidopsis* FAD3 (At2g29980), or omega-3 desaturase, was isolated both by map-based chromosome walking (Arondel et al., 1992) and T-DNA tagging (Yadav et al., 1993). In maturing seeds, the transcript level of this gene increases significantly. The corresponding desaturase is inserted co-translationally into the ER where it acts on fatty acids esterified to phosphatidylcholine (PC) (Somerville and Browse, 1991). The desaturase reaction requires NADP-cytochrome b5 reductase as an electron donor and cytochrome b5 as an intermediate electron donor (Kearns et al., 1991; Taton and Rahier, 1996). Mutations in *FAD3* genes were created by seed treatment with EMS and *Arabidopsis* lines with overexpression of *FAD3* genes had higher linolenic acid content compared to the wild type (Shah et al., 1997) while those with mutant *FAD3* alleles had no or decreased linolenic acid content (James and Dooner, 1990).

***FAD3* in soybean and low linolenic acid soybeans:** The interest in lowering the linolenic acid content in soybean has arisen since 1952 to find an alternative method for hydrogenation. Efforts to screen the soybean germplasm collection were initially employed; however, none of them were found to contain less than 4% 18:3 (Fehr, 2007).

The second strategy was to use X-ray irradiation to stimulate genetic changes (Hammond et al., 1972; Rahman and Takagi, 1997; Rahman et al., 1996b). Three lines with reduced linolenic acid content, M-5 (4.8%), IL-8 (4.5%), and KL-8 (6.5%), were developed by X-ray irradiation of „Bay" (Rahman and Takagi, 1997; Rahman et al., 1996b). M-5 and IL-8 were determined genetically to have the fan1 locus that are designated as *fan1*(M-5) and fan1(IL-8). The allele in KL-8 is not at the same locus as *fan1*(M-5) and was designated fanx(KL-8) (Rahman and Takagi, 1997).

In addition to X-ray irradiation, seed treatment with chemical mutagens such as EMS achieved some interesting results. Nearly at the same time, two soybean lines with lower linolenic acid content than their parents were released including A5 (4.1%) at the Iowa State University (Hammond and Fehr, 1983) and C1640 (3.6%) at the USDA/ARS in conjunction with Purdue University (Wilcox et al., 1984). The reduced linolenic acid contents in these two soybean lines were uncovered to be under the control of fan1 alleles (Rennie and Tanner, 1991; Wilcox et al., 1984). Other sources of low linolenic acid conditioned by mutations at the *fan1* locus include RG10 (this line was actually C1640 treated with EMS)(Rennie and Tanner, 1989b; Stojsin et al., 1998), PI 123440 and PI 361088B (Rennie et al., 1988). Later, a soybean line with lower palmitic acid content-A23 was crossed with A5 and in the progeny two soybean lines with 2.5% 18:3 were identified (Fehr, 2007). A23 was then identified to carry alleles different from fan1, which was called *fan2*. Subsequently, the third locus responsible for linolenic acid content was identified and designated as *fan3* in M24 (X-ray mutation of Bay) (Anai et al., 2005) and A26 (Fehr, 2007).

The molecular basis of all of the three fan loci and their alleles were determined by two groups, resulting in name contradiction about these three genes. (Bilyeu et al., 2003) reported that there are at least three versions of the *FAD3* omega-3 fatty acid desaturase gene in the soybean genome that are responsible for the conversion of linoleate to linolenate designated as Gm*FAD3A*, Gm*FAD3B*, and Gm*FAD3C*. Anai et al. 2005 reported four isoforms of *FAD3* genes and name them *GmFAD3-1a* (equivalent to Gm*FAD3B*), Gm*FAD3-1b* (Gm*FAD3A*), Gm*FAD3*-2a (Gm*FAD3C*), and Gm*FAD3-2b*. Up to date, it is now known that Glyma14 g37350 is the gene coding for Gm*FAD3A*, which is the Fan1 locus present on linkage group B2/chromosome 14; Glyma02 g39230 is the gene encoding Gm*FAD3B*/Fan3 present on LG D1b/chromosome 2; and Glyma18 g06950/ is the gene encoding Gm*FAD3C*/Fan2 present on LG G chromosome 18 (Bilyeu et al., 2011).

By sequencing the three candidate genes in mutant soybean lines, it was demonstrated that the reduced linolenic acid content of most of the mutagenized lines was caused by mutations in one or more *FAD3* genes, although modifying genes were also reported to influence the linolenic acid content in addition to three major genes (Graef et al., 1988). The lower linolenic acid content of A5, C1640, or J18 and M5 from Japan was associated to deletion or nonsense mutations in Gm*FAD3A* gene coding sequence (Anai et al., 2005; Bilyeu et al., 2011).The molecular basis of the fan1 allele in RG10 was determined to be a mutation in the Gm*FAD3A* gene resulting in a stop codon in the sixth exon inherited from the induced mutation in C1640 (Chappell and Bilyeu, 2006). In addition, RG10 was found to contain a second mutation in the 5′ splice site of the fifth intron of the Gm*FAD3B* gene which may result in abnormal mRNA splicing

products (Reinprecht et al., 2009). Similar to RG10, the linolenic acid content in A16 is also conditioned by two mutant Gm*FAD3* genes. Beside the mutant Gm*FAD3A* gene from A5, A16 was identified to have a single nucleotide mutation in Gm*FAD3C* that resulted in the substitution of a tyrosine for a histidine (Bilyeu et al., 2006b). In the 3% linolenic acid soybean line CX1512-44, a single nucleotide mutation in Gm*FAD3C* that resulted in the substitution of a glutamic acid for a glycine in addition to a splice site mutation in Gm*FAD3A* was found to control the low linolenic acid phenotype (Bilyeu et al., 2005). By combining the fan1 (A5), fan2 (A23) , and fan3 (A89–144003) alleles, the line A29 was developed that had only ≈1.0% linolenate (Fehr and Hammond, 2000; Ross et al., 2000). It was later known that *GmFAD3B* in A29 contains a mutation at a 3‟ splice site consensus sequence between intron 3 and exon 4, while the Gm*FAD3C* contains a missense mutation of H304Y which occurs at a very conserved region (Bilyeu et al., 2006). These mutations in three *FAD3* genes obviously knock out *GmFAD3A* and *GmFAD3B* and severely affect *GmFAD3C*‟s gene activity, leading to significantly lower linolenic acid content of 1%. Based on the DNA sequences of the three *GmFAD3* genes, molecular markers were designed and used as an effective tool to select for reduced linolenic acid content in breeding populations (Beuselinck et al., 2006).

The impact of reduced linolenic acid content on agronomic traits and seed traits has been evaluated in several studies. These studies indicated that the reduced linolenic acid content did not adversely affect other traits such as protein, oil content or yield, but it lowered tocopherol (vitamin E) content (McCord et al., 2004; Ross et al., 2000; Walker et al., 1998). Moreover, it was demonstrated by three different studies that oil with reduced linolenic acid content has higher oxidative stability than conventional soybean oil and

foods fried with this oil have better flavor scores compared to those prepared with conventional oil (Liu and White, 1992; Mounts et al., 1988; Warner and Gupta, 2003).

Currently, soybean oils with two levels of linolenate were released to the market: a low linolenate oil with 2.5 to <3.0% of the fatty acid and an ultra-low linolenate oil with ≈1.0% (Fehr, 2007). In 2006, approximately 280,000 ha (700,000 acres) of the low- and ultra-low linolenate soybeans in the United States were grown in the field. Production is expected to expand in the future as the food industry decreases its use of hydrogenated oils.

Improvement of oil content by overexpression of acyltransferase genes: In the group of acyltranferase genes in the Kennedy pathway, *GPAT*, *LPAAT* and *DGAT* have been shown to be effective targets to be used for the modification of fatty acid composition and seed oil content (Courchesne et al., 2009; Raneses et al., 1999). Though less information is available for *GPAT* and *LPAAT* compared to *DGAT*, their role in promoting oil content in seeds is strongly supported. Expression of plastidal safflower *GPAT* or E. coli *GPAT* gene increased seed oil content in *Arabidopsis* from 15 to 22% (Jain et al., 2000). In the same fashion, overexpression of a yeast *LPAAT* gene in soybean, rapeseed and *Arabidopsis* resulted in oil levels increased up to 1.5%, 22% and 45% in average, respectively (Rao and Hildebrand, 2009; Scherder and Fehr, 2008a; Zou et al., 1997). Consistent with this result, Maisonneuve et al, reported that expressing individually two *B. napus LPAAT* isozymes in *Arabidopsis* produced seeds with 14% greater lipid content and 7% greater seed mass (Maisonneuve et al.).

For *DGAT* gene, considerable evidence exists that it holds an important role in TAG accumulation in plants. There are two distinct gene families of *DGAT* that share no

homology with each other: *DGAT1* and *DGAT2* (Shockey et al., 2006; Yen et al., 2008). DGAT1 was first cloned from mouse while *DGAT2* was first identified in the fungus Morteriella ramanniana with two homologs *DGAT2A* and *DGAT2B* (Cases et al., 1998; Lardizabal et al., 2001). In both plants and animals, DGAT1 proteins often have about 500 amino acids (aa) with six to nine transmembrane domains, while DGAT2"s protein sequences are much shorter with about 300 aa and two transmembrane domains (Shockey et al., 2006; Yen et al., 2008) These two enzymes were also reported to localize in the different subdomains of the ER and have different expression level profiles in seeds of six oilseed crops (Kim et al., 2010b; Kroon et al., 2006; Shockey et al., 2006).

Despite many differences in protein sequence and structure, cellular location, and expression level, both of them were demonstrated to play an important role in TAG production. In plants, the significant role of *DGAT1* gene in accumulation of seed oil was reported in olive, tobacco, *Arabidopsis*, canola, castor bean, burning bush, soybean, tung tree, and maize (Bouvier-Navé et al., 2000; He et al., 2004; Jako et al., 2001; Milcamps et al., 2005; Mozaffarian et al., 2006; Nykiforuk et al., 2002; Shockey et al., 2006; Wang et al., 2006; Weselake et al., 2008; Zheng et al., 2008). Overexpression of a *DGAT1* gene in *Arabidopsis* and four different crops led to increases in seed oil content from 11-47% (Bouvier-Navé et al., 2000; Jako et al., 2001; Lardizabal et al., 2008; Zheng et al., 2008). Meanwhile, DGAT2 was suggested to be the most important enzyme in TAG production in yeast (Reinprecht et al., 2009). Notably, expression of M. ramanniana DGAT2A in insect cells and soybean showed a 3.1 fold increase in the amount of TAG and 1.5% increase in seed oil content per dry weight, respectively (Lardizabal et al., 2008). In developing seeds of tung tree and castor bean, the expression level of *DGAT2*

was higher than that of *DGAT1* (approximately 18 fold in castor bean), and was suggested to be responsible for the formation of seed oil (Kroon et al., 2006; Shockey et al., 2006). *DGAT* is also a good candidate gene for improvement of oleic acid content in soybean seed by altering the acyl channel in to TAG. An insertion mutation of 81 bp in the exon 2 of *Arabidopsis DGAT1* gene resulted in half the amount of oleic acid, compensated by a doubled amount of linolenic acid in seed oil compared to those of wild-type seed (Jako et al., 2001). In addition, overexpression of a *DGAT* gene with an insertion of three base pairs encoding for a phenylalanine at position 469 in maize resulted in increases in seed oil content up to 41% and oleic content up to 107% Recently, Oakes et al. reported that expression of a fungal *DGAT2* in maize kernels led to a change in fatty acid composition in which the oleic acid content increased up to 18% and palmitic, linoleic and linolenic content all decreased (Oakes et al., 2011), suggesting that *DGAT* can influence both oleic acid and seed oil content in plants.

Soybean *GPAT, LPAAT* and *DGAT* genes: Though the three acyltransferase genes mentioned above are good targets that can be used to increase the soybean seed oil content, up to date, there have been no published data for *GPAT* and *LPAAT* homologous genes and their roles in TAG production in soybean. For *DGAT*, sequence and expression level of a *DGAT1* gene was characterized in cultivated and wild soybean (Wang et al., 2006). However, the expression level of this gene was similar in different types of tissues including seed, leaf and flower, which prompted the authors to propose that this gene may not be the most important gene controlling the oil content in soybean seeds. In 2008, two homologous genes of *DGAT1* were reported: *DGAT1a*, which consists of 7575 bp in the genomic sequence (AB257589) and 99 % similarity to the *DGAT1* gene reported by

Wang et. al. (2006); and *DGAT1b*, which consists of 8164 bp in the genomic sequence (AB257590) (Hildebrand et al., 2008). Though *DGAT1b* has a greater activity compared to *DGAT1a*, activities of *DGAT1*s from soybean were five-fold less than the activity of a *DGAT1* gene from *Vernonia galamensis* (Hildebrand et al., 2008). Expression of *DGAT2* gene in soybean was reported to be significantly lower compared to that of *DGAT1* gene (Heppard et al., 1996a). However, it was not clear from the study whether the expression level of *DGAT1* and *DGAT2* each was obtained from a homologous gene with the highest expression in each family or was averaged from expression levels of all homologous genes in each gene family.

Soybean fatty acid profile and the environment: Temperature has been demonstrated to be the most essential environmental factor affecting the fatty acid composition in soybean seed oil. Among five fatty acid species, stearic acid was shown to be more dependent on soybean genotype rather than on temperature (Byfield and Upchurch, 2007; Kumar et al., 2005). Byfield and Urchurch elaborated that this phenomenon may be due to the regulation of stearic acid content in seeds that involves more genes other than desaturase enzymes such as acyl-ACP thioesterase genes *FATB* and *FATA*. The other four fatty acid species were more susceptible to temperature though the affect of temperature to each fatty acid in soybean is not in the same fashion. It was reported in several studies that linolenic acid content of soybean often decreases when temperatures are raised, although the correlation cofactor of linolenic acid content and temperature was small (Burton et al., 1983; Chapman et al., 1976; Howell and Collins, 1957; Kumar et al., 2005; Werteker et al., 2010; Wilcox et al., 1993; Wolf et al., 1982). This adverse relationship was shown in both field conditions (Chapman et al.,

1976), greenhouse (Howell and Collins, 1957) and growth chamber (Wolf et al., 1982). In contrast, palmitic acid content of soybean varieties often increase as the temperature goes up (Kumar et al., 2005). In the same fashion, a higher percent of oleic acid content is observed with warmer environment (Byfield and Upchurch, 2007; Carver et al., 1986; Dornbos and Mullen, 1992; Howell and Collins, 1957). The influence of temperature on the fatty acid composition in soybean has been linked to the inhibitory affect of temperature on the desaturase enzymatic activities when it increases. Because of this effect, the amount of the substrate of the desaturation reactions (i.e. oleic acid) will be increased and the amount of product will be reduced (i.e. linolenic acid). Cheesbrough (1989) provided evidence to support this hypothesis in which *FAD2* and *FAD3* enzymatic activities in seeds cultured in vitro greatly declined when the temperature increased from 20 ^{o}C to 25 ^{o}C (100 fold for *FAD2* and 60 fold for *FAD3*, and almost inactive at 35 ^{o}C for both of the enzymes). In addition, Tang et al. (2005) showed that the FAD2-1A degraded much faster than FAD2-1B at high temperature (30 ^{o}C), and this difference was attributed to the difference in amino acid sequence of two specific domains between two enzymes. Furthermore, increasing temperature is also found to inhibit the transcript levels of two desaturase enzymes including stearoyl-ACP desaturase and *FAD2* (Byfield and Upchurch 2007).

It is noted that soybeans with modified fatty acid profile showed less variation compared to wild type phenotype. However, this is true only for soybean lines that were mutated or were created by combining mutant genes encoding for desaturase enzymes to eliminate the enzyme activity (Bilyeu et al. 2006). The exception is stearic acid, which was shown by Byfield and Upchurch (2007) that soybean line A6 with high stearic acid

content up to 40% has standard deviation of 10% as the temperature changes compared to little or almost no change in the stearic acid content of the control cultivar „Dare".

Among other non-biological factors, latitudes and rainfall was shown to only influence stearic acid content but not other fatty acid species (Kumar et al., 2005). Other factors like irrigation, photoperiod, light intensity and quality, N, P, K, and S nutrition, or the addition of manure or plant residues had little or no effect on the levels of these five fatty acids (Howell and Collins, 1957; Lee et al., 2008).

Besides temperature, other factors documented to influence fatty acid composition in soybean are node position and maternal effect. In the study of Bennett et al. 2003, the oleic and linoleic acid contents of seeds harvested in the plant apex are lower than those of seeds in the basal nodes (Bennett et al. 2003). Using grafting technique, Carver et al. 1987 showed that the oleic acid content of soybean seeds of the grafted branches was influenced by the mother plant (Carver et al., 1987). Recently, using reciprocal populations, Gilginger et al. also detected maternal effects to not only oleic acid but also other fatty acids in soybean seeds in field grown soybeans (Warner and Fehr, 2008). Therefore, the authors suggested that the selection for changes in fatty acid composition may not be effective at early generations (i.e. F_2), except for those with molecular markers associated with the trait.

Soybeans and environment, maturity groups: Maturity is an important factor to be considered in breeding and selection for modification in fatty acid profile. Soybean is a short day plant and it is the length of the night that controls the flowering time of soybean (Holdsworth, 1964). Comparison of the fatty acid composition of two soybean lines planted at one location with different maturity can lead to misleading results and

conclusions because the pod filling periods of these two soybean lines may occur at different temperatures. Soybean maturity reflects soybean adaptation to certain latitudes in which daylength and warm temperature are the main factors that control soybean maturity via triggering the onset of soybean flowering (Green et al., 1965). Maturity groups range from 000 in the northern U.S. to VIII in the southern Gulf Coast states and most of Florida. It is important that soybeans flower after they reach the first trifoliate stage in growth so that the pods will develop and mature properly. Soybean lines with maturity group 000, I and II in the northern U.S. have longer minimum daylength requirements for the onset of flowering (often greater than 14 hours of daylight). Planting a specific variety in a latitude further south than its normal growing zone will shorten the vegetative period, triggering early flowering and early maturity of the variety due to warmer temperature and longer daylength, while planting a variety further north of its adaptive region will extend the vegetative stage and delay flowering and maturity due to shortened daylength and cooler temperature.

Use of winter nursery, temperature and daylength differences: The U.S. has a long and harshly cold winter; therefore, advancement of soybean breeding populations to two generations in a year is restricted. Therefore, it is now a commonly used practice to advance soybean breeding populations for one or two more generations in a winter nursery. Areas with semitropical or temperate climates like Costa Rica and Puerto Rico are ideal for winter nursery because in these places the temperature only drops to nearly 20 ^{O}C as the coldest in the winter. Determinate soybean lines only can be advanced for one more generation whereas indeterminate lines with early maturity can be grown for

two generations. In winter nursery, cultivation of soybean of group I-V is often improved due to longer day length, higher light intensity and higher temperature (Anai et al., 2005).

Soybean genetic diversity: Genetic diversity of commercial soybean lines used in the U.S. was shown to be very limited. Gizlice et al. (1994) provided the first insight of this issue by showing that the genetic basis of 258 public cultivars in North America released between 1947 and 1988 was contributed from 80 ancestors. However, six of the ancestors including Mandarin, CNS, Richland, S-100 and Lincoln‟s presumed two unknown parents contributed to nearly half of the genetic base of those soybean cultivars. Subsequently, Bilyeu and Beuselinck (2005) found that five out of six North American ancestral lines including Noir 1, Williams 82, Minsoy, Harosoy and Forrest have the same chloroplast haplotype #49. The exception was Peking which has haplotype #25. However, by sequencing 111 fragments in 102 genes from four populations, Hyten et al. 2006 demonstrated that the assumption that the limited genome diversity of current elite soybean cultivars is due to the utilization of only a small number of ancestral lines to develop new cultivar in modern breeding practices should not be valid for soybean. They claimed that the low nucleotide diversity in modern elite soybean cultivars is mainly due to an unusually low level of genetic variability in the wild progenitor, G. soja, and loss of diversity during the domestication bottleneck is about 50%. However, in this 50% loss, there was a significant loss of rare alleles present in *G. soja* and the Asian origin lines, which are important sources for soybean improvement. The authors hence suggested a solution to overcome this loss by screening the Asian soybean germplasm collection for variants of interest and then incorporating the desired variants in the elite germplasm pool.

Soybean germplasm collection: The headquarters of the USDA soybean germplasm collection resides at the University of Illinois. This center is in charge of storage, maintenance, management and distribution of soybean seeds classified into five categories: North American varieties (700 mostly publicly developed varieties), plant introductions (16,600 from 92 countries), wild soybeans (1,100 from four countries), perennial Glycine species (900 lines in 16 species), and the genetic collection (900 special research lines) (http://agronomyday.cropsci.illinois.edu/2003/exhibits/soybean_germplasm.html). It is estimated that more than 20,000 seed samples each year were distributed to approximately 400 soybean scientists in 35 states and 20 foreign countries. This collection provides excellent materials for studies aimed at screening and identification of new sources of resistance to various economically important diseases and abiotic stresses, of soybean seed composition, and recently of yield (Dorrance and Schmitthenner, 2000; Hartman et al., 1997; Hildebrand and Hymowitz, 1981; Lee et al., 2009; Mohamed and Rangappa, 1992). Less than 1% of the collection have been used for soybean variety development by commercial seed companies, therefore, this reservoir contains a vast untapped genetic diversity to be discovered and used to improve the soybean quality and production.

Soybean mutant collections: Three different soybean mutant collections were generated to provide a resource available to the soybean research community for identification of function of novel genes using reverse genetics approach. The first one was created by transposon-based mutagenesis and maintained by Mathieu et al. 2009. This collection contains 900 soybean events and analysis of the insertions revealed

approximately 70% disrupted known genes. In addition, four TILLING populations were produced using chemical mutagens and this is another valuable resource for studying gene function in soybean (Cooper et al., 2008). More importantly, the authors designed a strategy to facilitate the fast identification of mutant genes and avoid amplification of homologous genes in soybean. Recently, Bolon et al. (2011) reported on a release of a fast neutron mutant population with over 23,000 independent mutant lines as a public resource for future genetic screens and functional genomics research. All the information including genome-wide coverage of deletion events, phenotypic traits and photographs are now available and free to public access in a website along with parallel data from the unmutagenized wild-type M92-220.

Sources of mutations when project initiated: Soybean lines with elevated or mid oleic acid content used in this study come from different sources, reflecting different strategies used to increase oleic acid content in soybean. M23 was a mutant soybean line created by X-ray treatment on the cultivar Bay (Rahman et al., 1996a). 17D is a soybean line produced in a TILLING project using EMS to induce genetic changes in soybean (Dierking and Bilyeu, 2009). It was identified to have a change from $G^{350}A$ in the coding sequence resulting in a missense mutation from S117N in amino acid sequence. The missense mutation was found to be associated with the elevated oleic acid content up to 37% in segregating populations developed using 17D as the elevated oleic acid parent. PI 2833327, PI 567189A and PI 603452 were identified by screening the national soybean germplasm collection for changes in fatty acid composition (G. Shannon, personal communication, 2005).

Forward genetics and reverse genetics: In the modern molecular biology realm, forward and reverse genetics are two fundamental approaches to identify the function of a specific gene or gene family. However, forward genetics starts with introducing a mutagen agent to create a mutant phenotype and subsequently identifying the mutated genes using different standard molecular techniques (Peters et al., 2003). Reverse genetics, on the other hand, starts with inducing genetic changes or disrupting a certain gene or a gene family and try to gain more insights about the gene function and the phenotypes controlled by the selected gene or gene family (Hongtrakul et al., 1998). So in short, in forward genetics, a phenotype is known, and it is the goal to uncover the gene(s) responsible for the phenotype, while in reverse genetics a gene is selected for the study and the goal is to find out the phenotype that gene is controlling. Forward and reverse genetics approaches when combined offer the opportunity to hasten the process of identification of function of a specific gene/gene family.

In soybean, forward genetics-based studies have used both chemical mutagens and radiation for the creation of genetic modifications in the soybean genome. The most common chemical mutagens are EMS and N-nitroso-N-methylurea (NMU) because they often cause point mutations to only a single nucleotide, usually from G-C or A-T (Bilyeu, 2008; Till et al., 2007; Till et al., 2004). In contrast, radiation including X-ray, fast neutrons and gamma ray are very destructive as they caused deletions in the soybean genome either in a form of a small deletion within a gene or a large region containing many different genes (Anai et al., 2008; Bolon et al., 2011; Men et al., 2002; Rahman, 1994). Both methods have been successful in the generation and selection of very important traits such as: fatty acid composition, nodulation, oil and protein (Bolon et al.,

2011; Palmer et al., 2004; Takagi et al., 1989). Although map-based positional cloning method was successfully used to reveal the mutant genes in *Arabidopsis* and other species (Choe et al., 2002; Gonzaléz-Guzman et al., 2002; Lukowitz et al., 2000; Peters et al., 2003), this method when applied to soybean was shown to be difficult with only few successes and most of the time only molecular markers associated with the genes were identified (Gao et al., 2005). In addition to map-based cloning, the candidate gene approach is another method that is being effectively used in soybean, especially for traits such as fatty acid composition (Bilyeu et al., 2006b; Bilyeu et al., 2003; Cardinal et al., 2007; Pham et al., 2010). The candidate genes are often selected based on knowledge gained from studies on function of those gene(s) in a model species or other species that are comparable to the species of interest. The procedure of the candidate gene-based studies starts with sequencing candidate genes to find mutations in their DNA sequences and subsequently analyzing the association of a certain mutation with the phenotype in a segregating population to determine if the mutation is responsible for the mutant phenotype. Using this method, allele-specific markers were designed enabling the selection of genotype of interest at early generation.

In reverse genetics-based studies, several methods have been used to mutate or disrupt a gene of interest including: RNA-mediated interference, insertional mutagenesis, fast-neutron mutagenesis, virus-induced gene silencing (VIGS), TILLING (Targeting Induced Local Lesions IN Genomes), zinc-finger nuclease and transcription activator-like (TAL) effector nuclease (Christian et al., 2011; Curtin et al., 2011; Miller et al., 2011). Each method has its own advantages and disadvantages, therefore, the decision of which method to be employed depends on the resources, the species and the question to be

asked. Two repositories of mutations induced by Ds (disassociation) transposon elements and fast-neutron mutagenesis have just been established for the soybean research community (Bolon et al., 2011; Mathieu et al., 2009). VIGs was shown to effectively elucidate genes involving the pathway of soybean flavonoid synthesis (Nagamatsu et al., 2007), and soybean defense pathway (Pandey et al., 2011). A TILLING procedure was developed by Cooper et al. 2008 capable of being used in high-throughput pipeline and available to the soybean research community. New source of TILLING-induced mutations result in valuable traits being identified and characterized (Dierking and Bilyeu, 2009). Although the methods mentioned above have been used effectively in soybean to discover gene functions, they also have some disadvantages. Considering the soybean genome has been duplicated twice, resulting in multiple homologous genes controlling the same phenotype, it is always a challenge to gain a targeted phenotype when applying chemical mutagens or radiation because random mutations often produce no phenotype due to complementation of redundant genes. Usage of VIGs or RNAi interference often results in opposite results as all homologous genes in the family may be silenced at the same time.

To study the function of each member of a gene family of interest in soybean, a novel technique named zinc-finger nuclease (ZFN) has been established while a similar method named transcription activator-like effector nuclease (TALEN) may be used as an alternative (Curtin et al., 2011; Miller et al., 2011). ZFNs is a dimer complex, with each monomer consisting a DNA binding domain (a zinc finger array) fused to the catalytic domain of the FokI restriction enzyme 5 (Cathomen and Keith Joung, 2008). The zinc-finger array in each monomer will be designed to bind specifically to its target sequence

and the target sequences of two arrays from two monomers are separated by a spacer (5-7bp). Double-stranded break is generated in the spacer sequence between both ZFA binding sites and subsequently stimulates the cellular DNA repair pathways, which in turn generates small nucleotide insertions and deletions (indels) at the target DNA site that will disrupt the gene"s open reading frame. In the case of TALEN, zinc finger arrays are substituted with the DNA recognition domain of TAL effectors, which are proteins secreted by Xanthomonas bacteria via their type III secretion system when they infect various plant species, to create TALENs complex that recognize and cleave DNA targets. Because these techniques can distinguish homologous genes and create a specific mutation specifically in the individual gene of interest, they are becoming powerful tools for soybean molecular researchers for studying functional genomics in the highly duplicated soybean genome.

Population development: A soybean population is developed by crossing two soybean lines different for both DNA sequence and phenotype for the trait of interest. After crossing these two soybean lines together, true F_1 seeds are used to develop populations. Depending on the purpose of the research, populations can be developed into different types such as F_2, F_3, Recombinant inbred lines (RILs) or Near isogenic lines (NILs), backcross, etc. that are most commonly used for mapping and candidate gene approaches. Here in this document only the four types of populations mentioned above will be discussed.

F_2 populations produced by allowing self-fertilization (selfing) of F_1 individuals in a segregating population are generated by cross-fertilizing the selected parents; hence, it is produced from a single meiotic cycle (Allard, 1999; Schneider, 2005). Though it costs

less time and effort in generating this population, an F_2 population only can be used for preliminary mapping but not fine mapping because individuals in the population are genetically different. It is also hard to draw a precise estimation of the interaction between genetics and environment using F_2 populations. Moreover, it is impossible to be replicated or increased for seeds if the phenotypic analysis destroys the F_2 seeds.

F_3 population is obtained by selfing the F_2 individuals for one more generation (Allard, 1999). This population is better than F_2 population because it can be used for mapping quantitative traits or mapping recessive genes. However, it also has the same disadvantages of an F_2 population.

Recombinant inbred lines (RILs) population are produced by continuously selfing the individuals in F_2 population until complete homozygosity is achieved (Schneider, 2005). Single seed descent (SSD) method is the most appropriate technique in which for every initial F_2 plant, only one seed is kept to be propagated to yield the next generation. Usually 99% of homogenosity can be achieved by the F_8 generation, and each RIL represents a unique combination of genome segments from the two parents. RILs are the most common type of population used for mapping because an RIL population can be replicated over locations and years. Moreover, as RILs are produced after several meiosis cycles, the identified association of marker and the gene of interest of this population type is tight and it has a higher recombination frequency than F_2 population.

Near isogenic lines (NILs) are created by selfing or backcrossing F_1 plants to the recurrent parents (Maughan et al., 1996). In the selfing method, advanced selfed lines that are still segregating for a phenotype of interest is identified and allowed to self and produce progeny that can be phenotyped and genotyped. Genomes of NILs developed by

both methods are very similar except for the gene of interest. Consequently, NILs does not allow testing for genetic interactions and thereby the detection of QTL expressed in specific genetic backgrounds (epistasis). However, NILs have the power to detect small-effect QTL because it has one introgression per line. Although NILs are not ideal for linkage mapping and are time consuming to develop, they are very useful for tagging the trait (Keurentjes et al., 2007).

Difference of the two testing environments: Columbia, Missouri is at 38°58'' North with soil type of Mexico silt loam (fine, smectitic, mesic, Aeric Vertic Epiaqualfs)while Portageville, MO is at 36°44'' Dundee silt loam (fine-silty, mixed, thermic Typic Hapludalfs) (Lee et al. 2009). Both maximum and minimum average temperature of Columbia are often at least 2 ^{0}C lower than those at Portageville, MO during time course of the typical soybean production experiment.

Other candidate genes for high oleic acid content: In addition to *FAD2*, which directly involves the desaturation of oleic acid, various enzymes or proteins were shown to affect the oleic acid content in seed oil of oilseed crops. *DGAT* (diacylglycerol acyltransferase) was first reported to have an impact on oleic acid content via substrate specific preference. It was reported that DGAT prefers unsaturated fatty acids to saturated fatty acids. In an *Arabidopsis* line with a mutant *DGAT1* gene, oleic acid content was decreased by half and linolenic acid content increased two fold (Jako et al., 2001). Overexpression of a maize *DGAT* resulted in transgenic maize lines with oleic acid content increased by two fold (Zheng et al., 2008). Another candidate gene that was reported to have an influence on the oleic acid content in seed oil is the *ROD1* gene

(At3g15820), which is an abbreviation for reduced oleic desaturase, although the enzyme encoded by the gene is actually a phosphatidylcholine:diacylglycerol cholinephosphotransferase (*PDCT*) (Lu et al., 2009). *ROD1* mutant *Arabidopsis* plants have a marked increase in oleic acid content and reduction of 18:2 and 18:3 contents up to 40%. Using both map-based cloning and candidate gene methods in *Arabidopsis*, a gene mutation was identified as a G>A substitution predicted to change Tryptophan^{76} to a stop codon, which probably will produce a truncated and nonfunctional protein (Lu et al., 2009). PDCT is an important enzyme catalyzing the transfer of the phosphocholine headgroup from PC to diacylglycerol, although the increase of oleic acid content of mutant *ROD1* gene is less than the increase in mutant *FAD2* genes (Lu et al., 2009) (Figure 1.1).

At the transcriptional level, a transcription factor and one of its partners in the transcriptional activation apparatus of *FAD2* gene were found to be needed to activate the *FAD2* enzyme in sesame (Kim et al., 2007; Kim et al., 2010b). Several PIs with mid oleic acid contents were reported to have down-regulated expression levels of *FAD2-1A* and *FAD2-1B* (Reinprecht et al., 2009); therefore, these two genes, *DGAT* and PDCT, are good targets for a candidate gene-based approach to identify the cause of the reduced transcriptional level of *FAD2-1* genes in the PIs.

Soybean genome resource, potential SSR and SNP markers: Soybean has a palaeopolyploid genome with approximately 1,115-Mb. The size and complexity of the genome of soybean is the result of two duplication events which occurred at approximately 59 and 13 million years ago, resulting in a highly duplicated genome with nearly 75% of the genes present in multiple copies (Schmutz et al., 2010). Following two

rounds of duplication, most of the genes experienced diversification and limited loss, and numerous chromosomes were rearranged. Therefore, it is likely to have more than two homologous genes that control a same phenotype in soybean. The effort to construct a linkage map of soybean was completed by Song et al 2004. Using 420 SSR markers, the authors located all the markers into 20 linkage groups, providing a great resource for mapping studies. Subsequently, in 2010, Schmutz et al. reported the assemble of nearly 950megabases (Mb) of Glycine max var. Williams 82 into 20 chromosomes corresponding to 20 linkage groups, representing about 85% of the predicted genome. In addition, there is about 17.7 Mb is assembled into 1,148 unanchored sequence scaffolds that contain fewer than 450 predicted genes. This resource also contains information of 4,991 single nucleotide polymorphisms (SNPs), 874 simple sequence repeats (SSRs), ~20,000 low confidence protein encoded loci and 46,430 high-confidence loci that can be accessed at phytozome.net. This information will facilitate the mapping and fine mapping studies of economically important traits, as well as the functional genomics study in soybean.

Because soybean seeds are the most important product of soybean plants, the urge to know about all the biological process in seed formation has led to the foundation of a resource for transcriptional level of genes that are actively transcribed during embryogenesis, “Gene network in seed development” (http://seedgenenetwork.net/). The profiles of the mRNA sets present in different seed regions and compartments throughout seed development were identified using soybean and *Arabidopsis* Affymetrix GeneChips, Laser Capture Microdissection (LCM), and 454 high-throughput sequencing technologies. Up to date, mRNA profiles of all genes found to be present in 71 soybean

and *Arabidopsis* seed compartments from preglobular- to early maturation-stage seeds are identified and recorded. Additional resources for transcript levels of genes in soybean in different biological processes can be found at several online databases including Genevestigator (https://www.genevestigator.com/) (Hruz et al., 2008), soybase.org/soyseg, Soybean Knowledge Base (http://soykb.org/gene_card.php?gene=Glyma08g12270.1) (Joshi et al. 2010), and Transcriptome atlas of *Glycine max* (http://digbio.missouri.edu/soybean_atlas/) (Libault et al., 2010). The data present at these databases are obtained from microarray experiments, high through-put sequencing experiments as well as large scale qRT-PCR. Undoubtedly, the information will be greatly useful to identify genes involving a certain biological process in soybean‟s growth, development and reproduction.

The purpose of this project is to modify the fatty acid composition in soybean seeds to improve soybean oil quality and functionality. By sequencing the *FAD2-1A* and *FAD2-1B* genes in 24 plant introductions, we identified two novel mutant alleles: one for each gene that is responsible for the elevated oleic acid content in four plant introductions. The combination of the newly identified mutant *FAD2-1B* allele with existing or the novel mutant *FAD2-1A* alleles created soybean lines with more than 80% oleic acid content. Combination of two mutant *FAD2-1A* and *FAD2-1B* with mutant *FAD3A* or mutant *FAD3C* or both resulted in high oleic acid content of 80 – 85% and linolenic acid content in the range from 1.5 – 4%. Perfect molecular markers associated with these mutant alleles were designed to help select the soybean lines with genotypes of interest in early generations in breeding. The high oleic acid and high oleic acid low

linolenic soybeans produced have an improved stability across growing environments compared to existing sources.

FIGURE

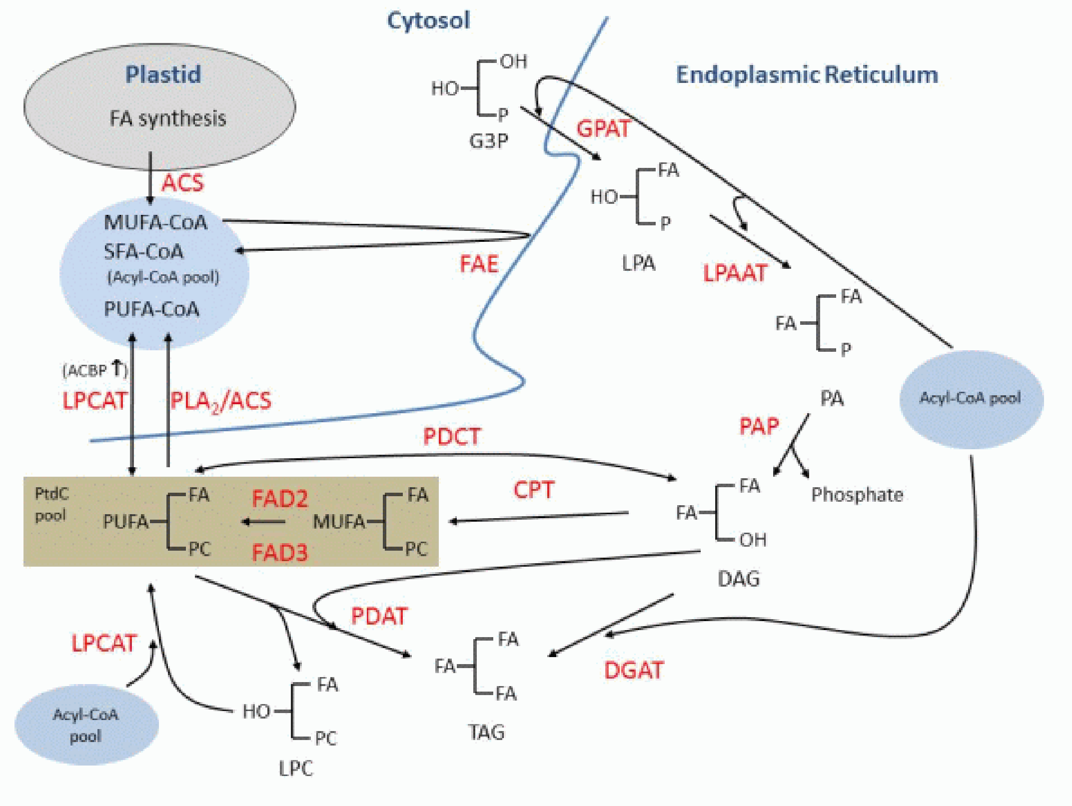


Figure 1.1: Generalized scheme for triacylglycerol (TAG) assembly in developing seeds of oleaginous plants (figure was taken from the website lipid library athttp://lipidlibrary.aocs.org/plantbio/tag_biosynth/index.htm with the permission granted from Dr. Randall Weselake, Dr. Jitou Zou and Dr. David Taylor). Abbreviations: CoA, coenzyme A; CPT,cholinephosphotransferase; DAG, sn-1,2-diacylglycerol; FA, fatty acid; FA-CoA, fatty acyl-coenzyme A; FAD2 and FAD3: fatty acid desaturases 2 and 3, G3P, sn-glycerol-3-phosphate; FAE, fatty acid elongase; GPAT, sn-glycerol-3-phosphate acyltransferase; LPA, lysophosphatidic acid; LPAAT,lysophosphatidic acid acyltransferase; LPC, lysophosphatidylcholine; LPCAT,lysophosphatidylcholine acyltransferase; MUFAs, Monounsaturated fatty acids; PA, phosphatidic acid; PAP, phosphatidic acid phosphatase; PC, phosphatidylcholine; PDAT,phospholipid:diacylglycerol acyltransferase; PLA2, phospholipase A2; PUFA, polyunsaturated fatty acids; SFA, Saturated fatty acids;

LITERATURE CITED

Ali, Y., M. Hanna, and S. Cuppett. 1995. Fuel properties of tallow and soybean oil esters. Journal of the American Oil Chemists' Society 72:1557-1564.

Allard, R.W. 1999. Principle of plant breeding. 2nd ed. John Wiley, New York.

Alt, J.L., W.R. Fehr, G.A. Welke, and J.G. Shannon. 2005. Transgressive segregation for oleate content in three soybean populations. Crop Science 45:2005-2007.

Anai, T., T. Yamada, T. Kinoshita, S.M. Rahman, and Y. Takagi. 2005. Identification of corresponding genes for three low-[alpha]-linolenic acid mutants and elucidation of their contribution to fatty acid biosynthesis in soybean seed. Plant Science 168:1615-1623.

Anai, T., T. Yamada, R. Hideshima, T. Kinoshita, S. Rahman, and Y. Takagi. 2008. Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by *GmFAD2*-1a. Breeding Sciences 58:447 - 452.

Arondel, V., B. Lemieux, I. Hwang, S. Gibson, H. Goodman, and C. Somerville. 1992. Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258:1353-1355.

Artaud-Wild, S., S. Connor, G. Sexton, and W. Connor. 1993. Differences in coronary mortality can be explained by differences in cholesterol and saturated fat intakes in 40 countries but not in France and Finland. A paradox. Circulation 88:2771-2779.

Ascherio, A., and W. Willett. 1997. Health effects of trans fatty acids. American Journal of Clinical Nutrition 66:1006 - 1010.

Bachlava, E., R. Dewey, J. Burton, and A. Cardinal. 2009. Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. Crop Science 49:433 - 442.

Baud, S., J.-P. Boutin, M. Miquel, L. Lepiniec, and C. Rochat. 2002. An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. Plant Physiology and Biochemistry 40:151-160.

Beló, A., P. Zheng, S. Luck, B. Shen, D.J. Meyer, B. Li, S. Tingey, and A. Rafalski. 2008. Whole genome scan detects an allelic variant of *FAD2* associated with increased oleic acid levels in maize Molecular Genetics and Genomics 279:1-10.

Beuchat, L.R., and R.E. Worthington. 1978. Technical note: Fatty acid composition of tree nut oils. International Journal of Food Science & Technology 13:355–358.

Beuselinck, P., D. Sleper, and K. Bilyeu. 2006. An assessment of phenotype selection for linolenic acid using genetic markers. Crop Science 46:747 - 750.

Bilyeu, K. 2008. Forward and Reverse Genetics in Soybean, p. 135-139, In G. Stacey, ed. Genetics and Genomics of Soybean. Springer Science Business Media, LLC, New York.

Bilyeu, K., L. Palavalli, D. Sleper, and P. Beuselinck. 2005. Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science 45:1830-1836.

Bilyeu, K., L. Palavalli, D. Sleper, and P. Beuselinck. 2006. Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science 46:1913 - 1918.

Bilyeu, K.D., and P.R. Beuselinck. 2005. Genetic Divergence between North American Ancestral Soybean Lines and Introductions with Resistance to Soybean Cyst Nematode Revealed by Chloroplast Haplotype. Journal of Heredity 96:593-599.

Bilyeu, K.D., J.D. Gillman, and A.R. LeRoy. 2011. Novel *FAD3* mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Science 51:259-264

Bilyeu, K.D., L. Palavalli, D.A. Sleper, and P.R. Beuselinck. 2003. Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science 43:1833-1838.

Bolon, Y.-T., W.J. Haun, W.W. Xu, D. Grant, M.G. Stacey, R.T. Nelson, D.J. Gerhardt, J.A. Jeddeloh, G. Stacey, G.J. Muehlbauer, J.H. Orf, S.L. Naeve, R.M. Stupar, and C.P. Vance. 2011. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiology.

Booth, J.R., R.M. Broglie, W.D. Hitz, A.J. Kinney, S. Knowlton, and S.A. Sebastian. 2005. Gene combinations that alter the quality and functionality of soybean oil 2005.

Bouvier-Navé, P., P. Benveniste, P. Oelkers, S.L. Sturley, and H. Schaller. 2000. Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase. European Journal of Biochemistry 267:85-96.

Brace, R.C., W.R. Fehr, and S.R. Schnebly. 2011. Agronomic and seed traits of soybean lines with high oleate concentration. Crop Sci. 51:534-541.

Bravo, J.J., W.R. Fehr, G.A. Welke, E.G. Hammond, and S.R. Cianzio. 1999. Family and Line Selection for Elevated Palmitate of Soybean. Crop Sci. 39:679-682.

Bruner, A., S. Jung, A. Abbott, and G. Powell. 2001a. The naturally occurring high oleate oil character in some peanut varieties results from reduced oleoyl-PC desaturase

activity from mutation of Aspartate 150 to Asparagine. Crop Science 41:522 - 526.

Budiyanto, A., N.U. Ahmed, A. Wu, T. Bito, O. Nikaido, T. Osawa, M. Ueda, and M. Ichihashi. 2000. Protective effect of topically applied olive oil against photocarcinogenesis following UVB exposure of mice. Carcinogenesis 21:2085-2090.

Buhr, T., S. Sato, F. Ebrahim, A. Xing, Y. Zhou, M. Mathiesen, B. Schweiger, A. Kinney, P. Staswick, and T. Clemente. 2002. Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. Plant Journal 30:155-163.

Burton, J.W., R.F. Wilson, and C.A. Brim. 1983. Recurrent Selection in Soybeans. IV. Selection for Increased Oleic Acid Percentage in Seed Oil1. Crop Sci. 23:744-747.

Byfield, G.E., and R.G. Upchurch. 2007. Effect of temperature on delta-9 stearoyl-ACP and microsomal omega-6 desaturase gene expression and fatty acid content in developing soybean seeds. Crop Science 47:1698 - 1704.

Cahoon, E.B. 2003. Genetic enhancement of soybean oil for industrial uses: Prospects and challenges. AgBioForm 6:11-13.

Cardinal, A.J., J.W. Burton, A.M. Camacho-Roger, J.H. Yang, R.F. Wilson, and R.E. Dewey. 2007. Molecular analysis of soybean lines with low palmitic acid content in the seed oil. Crop Science 47:304-310.

Carver, B.F., J.W. Burton, and R.F. Wilson. 1987. Graft-Transmissible Influence on Fatty Acid Composition of Soybean Seed1. Crop Sci. 27:53-56.

Carver, B.F., J.W. Burton, T.E. Carter, and R.F. Wilson. 1986. Response to Environmental Variation of Soybean Lines selected for Altered Unsaturated Fatty Acid Composition1. Crop Sci. 26:1176-1181.

Cases, S., S.J. Smith, Y.-W. Zheng, H.M. Myers, S.R. Lear, E. Sande, S. Novak, C. Collins, C.B. Welch, A.J. Lusis, S.K. Erickson, and R.V. Farese. 1998. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of the National Academy of Sciences of the United States of America 95:13018-13023.

Cathomen, T., and J. Keith Joung. 2008. Zinc-finger Nucleases: The Next Generation Emerges. Mol Ther 16:1200-1207.

Chapman, G., J. Robertson, D. Burdick, and M. Parker. 1976. Chemical composition and lipoxygenase activity in soybeans as affected by genotype and environment. Journal of the American Oil Chemists' Society 53:54-56.

Chappell, A.S., and K.D. Bilyeu. 2006. A Gm*FAD3*A mutation in the low linolenic acid soybean mutant C1640. Plant Breeding 125:535-536.

Choe, S., R.J. Schmitz, S. Fujioka, S. Takatsuto, M.-O. Lee, S. Yoshida, K.A. Feldmann, and F.E. Tax. 2002. *Arabidopsis* Brassinosteroid-Insensitive dwarf12Mutants Are Semidominant and Defective in a Glycogen Synthase Kinase 3Î²-Like Kinase. Plant Physiology 130:1506-1515.

Christian, M., T. Cermak, E.L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A.J. Bogdanove, and D.F. Voytas. 2011. TAL Effector Nucleases Create Targeted DNA Double-strand Breaks. Genetics:genetics.110.120717.

Cooper, J., B. Till, R. Laport, M. Darlow, J. Kleffner, A. Jamai, T. El-Mellouki, S. Liu, R. Ritchie, N. Nielsen, K. Bilyeu, K. Meksem, L. Comai, and S. Henikoff. 2008. TILLING to detect induced mutations in soybean. BMC Plant Biology 8:9.

Courchesne, N.M.D., A. Parisien, B. Wang, and C.Q. Lan. 2009. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. Journal of Biotechnology 141:31-41.

Covas, M.-I., V. Ruiz-Gutiérrez, R. de la Torre, A. Kafatos, R.M. Lamuela-Raventós, J. Osada, R.W. Owen, and F. Visioli. 2006. Minor Components of Olive Oil: Evidence to Date of Health Benefits in Humans. Nutrition Reviews 64:S20-S30.

Crawford, M.A., A.G. Hassam, and P.A. Stevens. 1981. Essential fatty acid requirements in pregnancy and lactation with special reference to brain development. Progress in Lipid Research 20:31-40.

Curtin, S.J., F. Zhang, J.D. Sander, W.J. Haun, C. Starker, N.J. Baltes, D. Reyon, E.J. Dahlborg, M.J. Goodwin, A.P. Coffman, D. Dobbs, J.K. Joung, D.F. Voytas, and R.M. Stupar. 2011. Targeted mutagenesis of duplicated genes in soybean with zinc finger nucleases. Plant Physiology.

Dahlqvist, A., U. StÃ¥hl, M. Lenman, A. Banas, M. Lee, L. Sandager, H. Ronne, and S. Stymne. 2000. Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proceedings of the National Academy of Sciences of the United States of America 97:6487-6492.

Dierking, E., and K. Bilyeu. 2008. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome 1:135 - 145.

Dierking, E., and K. Bilyeu. 2009. New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biology 9:89.

Dornbos, D., and R. Mullen. 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists' Society 69:228-231.

Dorrance, A.E., and A.F. Schmitthenner. 2000. New Sources of Resistance to Phytophthora sojae in the Soybean Plant Introductions. Plant Disease 84:1303-1308.

Ensminger, M.E., and A.H. Ensminger. 1993. Foods & Nutrition Encyclopedia. 2nd ed. CRC Press.

Falcone, D.L., S. Gibson, B. Lemieux, and C. Somerville. 1994. Identification of a Gene that Complements an *Arabidopsis* Mutant Deficient in Chloroplast [omega]6 Desaturase Activity. Plant Physiology 106:1453-1459.

Fehr, W.R. 2007. Breeding for modified fatty acid composition in soybean. Crop Science 47:S-72-87.

Fehr, W.R., and E.G. Hammond. 2000. Reduced linolenic acid production in soybeans. U.S. Patent No. 6 133 509. Issued October 17, 2000 Patent 6 133 509. Date Issued: October 17, 2000 2000.

Fehr, W.R., G.A. Welke, S.R. Cianzio, D.N. Duvick, and E.G. Hammond. 1991. Inheritance of Reduced Palmitic Acid Content in Seed Oil of Soybean. Crop Sci. 31:88-89.

Ferrara, L.A., A.S. Raimondi, L. d'Episcopo, L. Guida, A. Dello Russo, and T. Marotta. 2000. Olive Oil and Reduced Need for Antihypertensive Medications. Arch Intern Med 160:837-842.

Galili, E., Stanley. D.J., J. Sharvit, and M. Weinstein-Evron. 1997. Evidence for Earliest Olive-Oil Production in Submerged Settlements off the Carmel Coast, Israel. Journal of Archaeological Science 24:1141-1150.

Gao, H., N.N. Narayanan, L. Ellison, and M.K. Bhattacharyya. 2005. Two Classes of Highly Similar Coiled Coil-Nucleotide Binding-Leucine Rich Repeat Genes Isolated from the Rps1-k Locus Encode Phytophthora Resistance in Soybean. Molecular Plant-Microbe Interactions 18:1035-1045.

Gilchrist, E., and G. Haughn. 2010. Reverse genetics techniques: engineering loss and gain of gene function in plants. Briefings in Functional Genomics 9:103-110.

Gilsinger, J.J., J.W. Burton, and T.E. Carter. 2010. Maternal Effects on Fatty Acid Composition of Soybean Seed Oil. Crop Sci. 50:1874-1881.

Gonzaléz-Guzman, M., N. Apostolova, J.M. Bellés, J.M. Barrero, P. Piqueras, M.R. Ponce, J.L. Micol, R. Serrano, and P.L. Rodríguez. 2002. The Short-Chain

Alcohol Dehydrogenase ABA2 Catalyzes the Conversion of Xanthoxin to Abscisic Aldehyde. The Plant Cell Online 14:1833-1846.

Graef, G., B. LaVallee, P. Tenopir, M. Tat, B. Schweiger, A. Kinney, J. Gerpen, and T. Clemente. 2009. A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. Plant Biotechnology Journal 7:411 - 421.

Graef, G.L., W.R. Fehr, L.A. Miller, E.G. Hammond, and S.R. Cianzo. 1988. Inheritance of Fatty Acid Composition in a Soybean Mutant with Low Linolenic Acid. Crop Sci. 28:55-58.

Green, D.E., E.L. Pinnell, L.E. Cavanah, and L.F. Williams. 1965. Effect of Planting Date and Maturity Date on Soybean Seed Quality1. Agron. J. 57:165-168.

Gutiérrez-Nava, M.d.l.L., M.J. Aukerman, H. Sakai, S.V. Tingey, and R.W. Williams. 2008. Artificial trans-Acting siRNAs Confer Consistent and Effective Gene Silencing. Plant Physiology 147:543-551.

Hammond, E., W. Fehr, and H. Snyder. 1972. Improving soybean quality by plant breeding. Journal of the American Oil Chemists' Society 49:33-35.

Hammond, E.G., and W.R. Fehr. 1983. Registration of A5 germplasm line of soybean (Reg. No. GP44). Crop Sci 23:192.

Harrington, K.J. 1986. Chemical and physical properties of vegetable oil esters and their effect on diesel fuel performance. Biomass 9:1-17.

Hartman, G.L., Y.H. Huang, R.L. Nelson, and G.R. Noel. 1997. Germplasm Evaluation of Glycine max for Resistance to Fusarium solani, the Causal Organism of Sudden Death Syndrome. Plant Disease 81:515-518.

He, X., C. Turner, G.Q. Chen, J.-T. Lin, and T.A. McKeon. 2004. Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean Lipid 39:311-319.

Heppard, E., A. Kinney, K. Stecca, and G. Miao. 1996. Developmental and growth temperature regulation of two different microsomal [omega]-6 desaturase genes in soybeans. Plant Physiol 110:311 - 319.

Hesseltine, C.W. 1983. The Future of Fermented Foods. Nutrition Reviews 41:293-301.

Hesseltine, C.W., and K. Shibasaki. 1961. Miso: III. Pure Culture Fermentation with Saccharomyces rouxii. Appl. Environ. Microbiol. 9:515-518.

Hildebrand, D., and T. Hymowitz. 1981. Two soybean genotypes lacking lipoxygenase-1. Journal of the American Oil Chemists' Society 58:583-586.

Hildebrand, D.F., R. Li, and T. Hatanaka. 2008. Genomics of soybean oil traits, p. 185-209, In G. Stacey, ed. Genetics and genomics of soybean, Vol. 2. Springer New York, New York.

Hitz, W.D., T.J. Carlson, J.R. Booth Jr, A.J. Kinney, K.L. Stecca, and N.S. Yadav. 1994. Cloning of a Higher-Plant Plastid [omega]-6 Fatty Acid Desaturase cDNA and Its Expression in a Cyanobacterium. Plant Physiology 105:635-641.

Holdsworth, M. 1964. The Leaf Movements of Soybean, a Short-day Plant. Journal of Experimental Botany 15:391-398.

Hongtrakul, V., M.B. Slabaugh, and S.J. Knapp. 1998. A seed specific delta-12 oleate desaturase gene is duplicated, rearranged, and weakly expressed in high oleic acid sunflower lines. Crop Science v38:p1245(5).

Howell, R.W., and F.I. Collins. 1957. Factors Affecting Linolenic and Linoleic Acid Content of Soybean Oil1. Agron. J. 49:593-597.

Hruz, T., O. Laule, G. Szabo, F. Wessendorp, S. Bleuler, L. Oertle, P. Widmayer, W. Gruissem, and P. Zimmermann. 2008. Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Advances in Bioinformatics:420747.

Hu, F.B., M.J. Stampfer, J.E. Manson, E. Rimm, G.A. Colditz, B.A. Rosner, C.H. Hennekens, and W.C. Willett. 1997. Dietary fat intake and the risk of coronary heart disease in women. N Engl J Med 337:1491-1499.

Hu, X., M. Sullivan-Gilbert, M. Gupta, and S. Thompson. 2006b. Mapping of the loci controlling oleic and linolenic acid contents and development of *FAD2* and *FAD3* allele-specific markers in canola (*Brassica napus* L.). Theoretical Applied Genetics 113:497 - 507.

Hymowitz, T. 1990. Soybeans: The success story, p. 159-163, In J. Janick and J. E. Simon, eds. Advances in new crops. Timber Press Portland, OR.

Hymowitz, T., and C. Newell. 1981. Taxonomy of the genus Glycine, domestication and uses of soybeans. Economic Botany 35:272-288.

Hymowitz, T., and W.R. Shurtleff. 2005. Debunking Soybean Myths and Legends in the Historical and Popular Literature. Crop Sci 45:473-476.

Hyten, D.L., Q. Song, Y. Zhu, I.-Y. Choi, R.L. Nelson, J.M. Costa, J.E. Specht, R.C. Shoemaker, and P.B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences 103:16666-16671.

Jain, R.K., M. Coffey, K. Lai, A. Kumar, and S.L. MacKenzie. 2000. Enhancement of seed oil content by expression of glycerol-3-phosphate acyltransferase genes. Biochem. Soc. Trans. 28:958-961.

Jako, C., A. Kumar, Y. Wei, J. Zou, D.L. Barton, E.M. Giblin, P.S. Covello, and D.C. Taylor. 2001. Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol. 126:861-874.

James, D.W., and H.K. Dooner. 1990. Isolation of EMS-induced mutants in *Arabidopsis* altered in seed fatty acid composition. TAG Theoretical and Applied Genetics 80:241-245.

Jin, U.-H., J.-W. Lee, Y.-S. Chung, J.-H. Lee, Y.-B. Yi, Y.-K. Kim, N.-I. Hyung, J.-H. Pyee, and C.-H. Chung. 2001. Characterization and temporal expression of a [omega]-6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds. Plant Science 161:935-941.

Jung, S., G. Powell, K. Moore, and A. Abbott. 2000. The high oleate trait in the cultivated peanut [Arachis hypogaea L.]. II. Molecular basis and genetics of the trait. Molecular and General Genetics 263:806-811.

Kearns, E.V., S. Hugly, and C.R. Somerville. 1991. The role of cytochrome b5 in [Delta]12 desaturation of oleic acid by microsomes of safflower (*Carthamus tinctorius* L.). Archives of Biochemistry and Biophysics 284:431-436.

Keurentjes, J.J.B., L. Bentsink, C. Alonso-Blanco, C.J. Hanhart, H. Blankestijn-De Vries, S. Effgen, D. Vreugdenhil, and M. Koornneef. 2007. Development of a Near-Isogenic Line Population of *Arabidopsis* thaliana and Comparison of Mapping Power With a Recombinant Inbred Line Population. Genetics 175:891-905.

Kim, J., D.N. Kim, S.H. Lee, S.-H. Yoo, and S. Lee. 2010a. Correlation of fatty acid composition of vegetable oils with rheological behaviour and oil uptake. Food Chemistry 118:398-402.

Kim, M., J.-K. Kim, J. Shin, and M. Suh. 2007. The SebHLH transcription factor mediates trans-activation of the Se*FAD2* gene promoter through binding to E- and G-box elements. Plant Molecular Biology 64:453-466.

Kim, M., H. Kim, J. Shin, C.-H. Chung, J. Ohlrogge, and M. Suh. 2006. Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis -regulatory elements in the Se*FAD2* promoter and enhancers in the 5′-UTR intron. Molecular Genetics and Genomics 276:351-368.

Kim, M., Y. Go, S. Lee, Y. Kim, J. Shin, M. Min, I. Hwang, and M. Suh. 2010b. Seed-expressed casein kinase I acts as a positive regulator of the Se*FAD2* promoter via phosphorylation of the SebHLH transcription factor. Plant Molecular Biology 73:425-437.

Kinney, A.J., and S. Knowlton. 1998. Designer oils: The high oleic acid soybean, p. 193-213, In S. Roller and S. Harlander, eds. Genetic modification in the food industry. Blackie Academic, London.

Kinney, A.J., K.L. Stecca, and K. Meyer. 2008. Soybean event DP-305423-1 and compositions and methods for the identification and/or detection thereof 2008.

Kinoshita, T., S.M. Rahman, T. Anai, and Y. Takagi. 1998. Genetic analysis of restriction fragment length polymorphism on the fatty acid synthesis in soybean mutants and their progenies: II High oleic acid mutants with two microsomal x-6 fatty acid desaturase cDNAs as probes. Bull Fac Agric Saga Univ 83:37-42.

Knowlton, S. 1999. Soybean oil having high oxidative stability. U.S. Patent 5 981 781. Date issued: 9 November. Date Issued: 9 November 1999 1999.

Kok, L., W. Fehr, E. Hammond, and P. White. 1999. Trans-free margarine from highly saturated soybean oil. Journal of the American Oil Chemists' Society 76:1175-1181.

Kroon, J.T.M., W. Wei, W.J. Simon, and A.R. Slabas. 2006. Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. Phytochemistry 67:2541-2549.

Kumar, V., A. Rani, S. Solanki, and S.M. Hussain. 2005. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. Journal of Food Composition and Analysis 19:188-195.

Lacombe, S., I. Souyris, and A. Bervillé. 2009. An insertion of oleate desaturase homologous sequence silences via siRNA the functional gene leading to high oleic acid content in sunflower seed oil. Molecular Genetics and Genomics 281:43-54.

Lardizabal, K., R. Effertz, C. Levering, J. Mai, M.C. Pedroso, T. Jury, E. Aasen, K. Gruys, and K. Bennett. 2008. Expression of Umbelopsis ramanniana DGAT2A in seed increases oil in soybean. Plant Physiol. 148:89-96.

Lardizabal, K.D., J.T. Mai, N.W. Wagner, A. Wyrick, T. Voelker, and D.J. Hawkins. 2001. DGAT2 Is a new diacylglycerol acyltransferase gene family. Purification, cloning and expression in insect cells of two polypeptides from Mortierella

ramanniana with diacylglycerol acyltransferase activity. J. Biol. Chem. 276:38862-38869.

Lee, J.D., M.L. Oliva, D.A. Sleper, and J.G. Shannon. 2008. Irrigation has Little Effect on Unsaturated Fatty Acid Content in Soya Bean Seed Oil within Genotypes Differing in Fatty Acid Profile. Journal of Agronomy and Crop Science 194:320-324.

Lee, J.D., M. Woolard, D.A. Sleper, J.R. Smith, V.R. Pantalone, C.N. Nyinyi, A. Cardinal, and J.G. Shannon. 2009. Environmental effects on oleic acid in soybean seed oil of Plant Introductions with elevated oleic concentration. Crop Science 49:1762-1768.

Li, L., X. Wang, J. Gai, and D. Yu. 2008. Isolation and characterization of a seed-specific isoform of microsomal omega-6 fatty acid desaturase gene (*FAD2*-1B) from soybean.

Libault, M., A. Farmer, T. Joshi, K. Takahashi, R.J. Langley, L.D. Franklin, J. He, D. Xu, G. May, and G. Stacey. 2010. An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. The Plant Journal 63:86-99.

Liu, H.-R., and P. White. 1992. Oxidative stability of soybean oils with altered fatty acid compositions. Journal of the American Oil Chemists' Society 69:528-532.

Liu, Q., S. Singh, and A. Green. 2002. High-oleic and high-stearic cottonseed oils: nutritionally improved cooking oils developed using gene silencing. Journal of the American College of Nutrition 21:205S-211S.

Liu, Q., P. Singh Surinder, Brubaker Curt L., Sharp Peter J., Green Allan G., and M.D. R. 1999. Molecular cloning and expression of a cDNA encoding a microsomal w-6 fatty acid desaturase from cotton (*Gossypium hirsutum*). Australian Journal of Plant Physiology 26:101-106.

Liu, Q., C.L. Brubaker, A.G. Green, R.M. Don, P.J. Sharp, and S.P. Singh. 2001. Evolution of the *FAD2*-1 Fatty Acid Desaturase 5' UTR Intron and the Molecular Systematics of Gossypium (Malvaceae). American Journal of Botany 88:92-102.

Lu, C., Z. Xin, Z. Ren, M. Miquel, and J. Browse. 2009. An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of *Arabidopsis*. Proceedings of the National Academy of Sciences 106:18837-18842.

Lui, K. 1997. Soybeans: Chemistry, Technology, and Utilization Chapman & Hall, New York.

Lukowitz, W., C.S. Gillmor, and W.R. Scheible. 2000. Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. Plant Physiology 123:795-805.

Machowetz, A., H.E. Poulsen, S. Gruendel, A. Weimann, M. Fita, J. Marrugat, R. de la Torre, J.T. Salonen, K. NyyssÃnen, J. Mursu, S. Nascetti, A. Gaddi, H. Kiesewetter, H. BÃcumler, H. Selmi, J. Kaikkonen, H.-J.F. Zunft, M.-I. Covas, and C. Koebnick. 2007. Effect of olive oils on biomarkers of oxidative DNA stress in Northern and Southern Europeans. The FASEB Journal 21:45-52.

Maisonneuve, S., J.-J. Bessoule, R. Lessire, M. Delseny, and T.J. Roscoe. Expression of Rapeseed Microsomal Lysophosphatidic Acid Acyltransferase Isozymes Enhances Seed Oil Content in *Arabidopsis*. Plant Physiol. 152:670-684.

Mathieu, M., E. Winters, F. Kong, J. Wan, S. Wang, H. Eckert, D. Luth, M. Paz, C. Donovan, Z. Zhang, D. Somers, K. Wang, H. Nguyen, R. Shoemaker, G. Stacey, and T. Clemente. 2009. Establishment of a soybean (*Glycine max* Merr. L) transposon-based mutagenesis repository. Planta 229:279-289.

Maughan, P., M. Saghai Maroof, G. Buss, and G. Huestis. 1996. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. TAG Theoretical and Applied Genetics 93:392-401.

McCord, K.L., W.R. Fehr, T. Wang, G.A. Welke, S.R. Cianzio, and S.R. Schnebly. 2004. Tocopherol Content of Soybean Lines with Reduced Linolenate in the Seed Oil. Crop Sci. 44:772-776.

Men, A.E., T.S. Laniya, I.R. Searle, Iturbe-Ormaetxe, Inaki; Gresshoff, Irma; Jiang, Qunyi; Carroll, B. J.;, Gresshoff, and P. M. 2002. Fast Neutron Mutagenesis of Soybean (*Glycine soja* L.) Produces a Supernodulating Mutant Containing a Large Deletion in Linkage Group H. Genome Letters 1:147-155(9).

Messina, M.J. 1999. Legumes and soybeans: overview of their nutritional profiles and health effects. American Journal of Clinical Nutrition 70:439-450.

Mikkilineni, V., and T.R. Rocheford. 2003. Sequence variation and genomic organization of fatty acid desaturase-2 (*FAD2*) and fatty acid desaturase-6 (fad6) cDNAs in maize. TAG Theoretical and Applied Genetics 106:1326-1332.

Milcamps, A., A.W. Tumaney, T. Paddock, D.A. Pan, J. Ohlrogge, and M. Pollard. 2005. Isolation of a gene encoding a 1,2-Diacylglycerol-sn-acetyl-CoA Acetyltransferase fromd eveloping seeds of *Euonymus alatus*. J. Biol. Chem. 280:5370-5377.

Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, D.E. Paschon, E. Leung, S.J. Hinkley, G.P. Dulay, K.L. Hua, I. Ankoudinova, G.J. Cost, F.D.

Urnov, H.S. Zhang, M.C. Holmes, L. Zhang, P.D. Gregory, and E.J. Rebar. 2011. A TALE nuclease architecture for efficient genome editing. Nat Biotech 29:143-148.

Mohamed, A., and M. Rangappa. 1992. Screening soybean (grain and vegetable) genotypes for nutrients and anti-nutritional factors. Plant Foods for Human Nutrition (Formerly Qualitas Plantarum) 42:87-96.

Monteros, M., J. Burton, and H. Boerma. 2008. Molecular mapping and confirmation of QTLs associated with oleic acid content in N00-3350 soybean. Crop Science 48:2223 - 2234.

Mou, Z., Y. He, Y. Dai, X. Liu, and J. Li. 2000. Deficiency in Fatty Acid Synthase Leads to Premature Cell Death and Dramatic Alterations in Plant Morphology. The Plant Cell Online 12:405-418.

Mounts, T., K. Warner, G. List, R. Kleiman, W. Fehr, E. Hammond, and J. Wilcox. 1988. Effect of altered fatty acid composition on soybean oil stability. Journal of the American Oil Chemists' Society 65:624-628.

Mozaffarian, D., M.B. Katan, A. Ascherio, M.J. Stampfer, and W.C. Willett. 2006. Trans fatty acids and cardiovascular disease. New England Journal of Medicine 354:1601-1613.

Mroczka, A., P.D. Roberts, J.J. Fillatti, B.E. Wiggins, T. Ulmasov, and T. Voelker. 2010. An Intron Sense Suppression Construct Targeting Soybean *FAD2*-1 Requires a Double-Stranded RNA-Producing Inverted Repeat T-DNA Insert. Plant Physiology 153:882-891.

Nagamatsu, A., C. Masuta, M. Senda, H. Matsuura, A. Kasai, J.-S. Hong, K. Kitamura, J. Abe, and A. Kanazawa. 2007. Functional analysis of soybean genes involved in flavonoid biosynthesis by virus-induced gene silencing. Plant Biotechnology Journal 5:778-790.

Nykiforuk, C.L., T.L. Furukawa-Stoffer, P.W. Huff, M. Sarna, A. Laroche, M.M. Moloney, and R.J. Weselake. 2002. Characterization of cDNAs encoding diacylglycerol acyltransferase from cultures of *Brassica napus* and sucrose-mediated induction of enzyme biosynthesis. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1580:95-109.

Oakes, J., D. Brackenridge, R. Colletti, M. Daley, D.J. Hawkins, H. Xiong, J. Mai, S.E. Screen, D. Val, K. Lardizabal, K. Gruys, and J. Deikman. 2011. Expression of Fungal diacylglycerol acyltransferase2 Genes to Increase Kernel Oil in Maize. Plant Physiology 155:1146-1157.

Ohlrogge, J., and J. Browse. 1995. Lipid Biosynthesis. Plant Cell 7:957-970.

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

Oliva, M., J. Shannon, D. Sleper, M. Ellersieck, A. Cardinal, R. Paris, and J. Lee. 2006. Stability of fatty acid profile in soybean genotypes with modified seed oil composition. Crop Science 46:2069 - 2075.

Oomah, B.D., and G. Mazza. 1999. Health benefits of phytochemicals from selected Canadian crops. Trends in Food Science & Technology 10:193-198.

Owen, R.W., A. Giacosa, W.E. Hull, R. Haubner, B. Spiegelhalder, and H. Bartsch. 2000. The antioxidant/anticancer potential of phenolic compounds isolated from olive oil. European Journal of Cancer 36:1235-1247.

Palmer, R.G., T.W. Pfeiffer, G.R. Buss, and T.C. Kilen. 2004. Qualitative genetics, p. 137-234, In J. E. Specht and H. R. Boerma, eds. Soybeans: improvement, Production, and Uses. Crop Sci.Society of America Monograph, Madison, WI.

Pandey, A.K., C. Yang, C. Zhang, M.A. Graham, H.D. Horstman, Y. Lee, O.A. Zabotina, J.H. Hill, K.F. Pedley, and S.A. Whitham. 2011. Functional Analysis of the Asian Soybean Rust Resistance Pathway Mediated by Rpp2. Molecular Plant-Microbe Interactions 24:194-206.

Patel, M., S. Jung, K. Moore, G. Powell, C. Ainsworth, and A. Abbott. 2004. High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theoretical Applied Genetics 108:1492-1502.

Peng, Q., Y. Hu, R. Wei, Y. Zhang, C. Guan, Y. Ruan, and C. Liu. 2010. Simultaneous silencing of *FAD2* and FAE1 genes affects both oleic acid and erucic acid contents in Brassica napus seeds. Plant Cell Reports 29:317-325.

Peters, J.L., F. Cnudde, and T. Gerats. 2003. Forward genetics and map-based cloning approaches. Trends in Plant Science 8:484-491.

Pham, A.-T., J.-D. Lee, J.G. Shannon, and K. Bilyeu. 2010. Mutant alleles of *FAD2*-1A and *FAD2*-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology 10:195.

Pirtle, I.L., W. Kongcharoensuntorn, M. Nampaisansuk, J.E. Knesek, K.D. Chapman, and R.M. Pirtle. 2001. Molecular cloning and functional expression of the gene for a cotton [Delta]-12 fatty acid desaturase (*FAD2*). Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1522:122-129.

Rahman, S.M., and Y. Takagi. 1997. Inheritance of reduced linolenic acid content in soybean seed oil. TAG Theoretical and Applied Genetics 94:299-302.

Rahman, S.M., Y. Takagi, and T. Kinoshita. 1996a. Genetic control of high oleic acid content in the seed oil of two soybean mutants. Crop Science 36:1125-1128.

Rahman, S.M., Y. Takagi, and T. Kumamaru. 1996b. Low linolenate sources at the Fan locus in soybean lines M-5 and IL-8. Breeding Science 46:155-158.

Rahman, S.M., T. Anai, T. Kinoshita, S. Arima, and Y. Takagi. 2004. Three novel soybean germplasms with unique fatty acid composition using multiple mutant alleles. Breeding Science 54:225-229.

Rahman, S.M., Takagi, Y., Kubota, K., Miyamoto, K. and Kawakita, T. . 1994. High Oleic Acid Mutant in Soybean Induced by X-Ray Irradiation. Bioscience, biotechnology, and biochemistry 58:1070-1072.

Raneses, A., L. Glaser, J. Price, and J. Duffield. 1999. Potential biodiesel markets and their economic effects on the agricultural sector of the United States. Industrial Crops and Products 9:151 - 162.

Rao, S., and D. Hildebrand. 2009. Changes in Oil Content of Transgenic Soybeans Expressing the Yeast SLC1 Gene. Lipids 44:945-951.

Raymer, P.L. 2002. Canola: An Emerging Oilseed Crop, In J. Janick and A. Whipkey, eds. Trends in new crops and new uses. ASHS Press, Alexandria, VA.

Reinprecht, Y., S.Y. Luk-Labey, J. Larsen, V.W. Poysa, K. Yu, I. Rajcan, G.R. Ablett, and K.P. Pauls. 2009. Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. Plant Breeding 128:253-258.

Rennie, B.D., and J.W. Tanner. 1989. Genetic analysis of low linolenic acid levels in the soybean line PI 123440. Soybean Genet. Newsl. 16:25-26.

Rennie, B.D., and J.W. Tanner. 1991. New Allele at the Fan Locus in the Soybean Line A5. Crop Sci. 31:297-301.

Rennie, B.D., J. Zilka, M.M. Cramer, and W.D. Beversdorf. 1988. Genetic analysis of low linolenic acid levels in the soybean line PI 361088B. Crop Sci 28:655–657.

Ross, A.J., W.R. Fehr, G.A. Welke, and S.R. Cianzio. 2000. Agronomic and Seed Traits of 1%-Linolenate Soybean Genotypes. Crop Sci. 40:383-386.

Ryan III T.W., Dodge L.G., and C. T.J. 1984. The effects of vegetable oil properties on injection and combustion in two different diesel engines. Journal of the American Oil Chemists Society 61:1610-1619.

Scheffler, J.A., A.G. Sharpe, H. Schmidt, P. Sperling, I.A.P. Parkin, W. Lühs, D.J. Lydiate, and E. Heinz. 1997. Desaturase multigene families of *Brassica napus*

arose through genome duplication. TAG Theoretical and Applied Genetics 94:583-591.

Scherder, C.W., and W.R. Fehr. 2008. Agronomic and seed characteristics of soybean lines with increased oleate content. Crop Sci. 48:1755-1758.

Schierholt, A., H.C. Becker, and W. Ecke. 2000. Mapping a high oleic acid mutation in winter oilseed rape (*Brassica napus* L.). TAG Theoretical and Applied Genetics 101:897-901.

Schierholt, A., B. Racker, and H.C. Becker. 2001. Inheritance of High Oleic Acid Mutations in Winter Oilseed Rape (L.). Crop Sci. 41:1444-1449.

Schlueter, J.A., I.F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. Siegfried, B.A. Roe, S.D. Schlueter, B.E. Scheffler, and R.C. Shoemaker. 2007. The *FAD2* gene family of soybean:Insights into the structural and functional divergence of a paleopolyploid genome. Crop Science 47:S-14 - 26.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.-C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, and S.A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Schneider, K. 2005. Mapping Populations and Principles of Genetic Mapping Wiley-VCH Verlag GmbH & Co. KGaA.

Scrimgeour, C. 2005. Chemistry of Fatty Acids, In F. Shahidi, ed. Bailey‟s Industrial Oil and Fat Products, Vol. Six Volume Set, Sixth ed. John Wiley & Sons, Inc.

Shah, S., Z. Xin, and J. Browse. 1997. Overexpression of the *FAD3* Desaturase Gene in a Mutant of *Arabidopsis*. Plant Physiology 114:1533-1539.

Shockey, J.M., S.K. Gidda, D.C. Chapital, J.-C. Kuan, P.K. Dhanoa, J.M. Bland, S.J. Rothstein, R.T. Mullen, and J.M. Dyer. 2006. Tung Tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 18:2294-2313.

Simopoulos, A.P. 1999. Essential fatty acids in health and chronic disease. The American Journal of Clinical Nutrition 70:560S-569S.

Sleper, D.A., and J.M. Poehlman. 2006. Breeding field crops. 5th ed. Blackwell Publishing Professional, 2121 State Ave., Ames, IA 50014-8300.

Smith, M.A., A.R. Cross, O.T. Jones, W.T. Griffiths, S. Stymne, and K. Stobart. 1990. Electron-transport components of the 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine delta 12-desaturase (delta 12-desaturase) in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons. Biochem Journal 272:23-29.

Somerville, C., and J. Browse. 1991. Plant Lipids: Metabolism, Mutants, and Membranes. Science 252:80-87.

Song, Q., L. Marek, R. Shoemaker, K. Lark, V. Concibido, X. Delannay, J. Specht, and P. Cregan. 2004. A new integrated genetic linkage map of the soybean. TAG Theoretical and Applied Genetics 109:122-128.

Spasibionek, S. 2006. New mutants of winter rapeseed (Brassica napus L.) with changed fatty acid composition. Plant Breeding 125:259-267.

Srivastava, A., and R. Prasad. 2000. Triglycerides-based diesel fuels. Renewable & sustainable energy reviews 4:111-133.

Stahl, U., A.S. Carlsson, M. Lenman, A. Dahlqvist, B. Huang, W. Banas, A. Banas, and S. Stymne. 2004. Cloning and Functional Characterization of a Phospholipid:Diacylglycerol Acyltransferase from *Arabidopsis*. Plant Physiol. 135:1324-1335.

Stojsin, D., B.M. Luzzi, G.R. Ablett, and J.W. Tanner. 1998. Inheritance of Low Linolenic Acid Level in the Soybean Line RG10. Crop Sci. 38:1441-1444.

Stoltzfus, D.L., W.R. Fehr, G.A. Welke, E.G. Hammond, and S.R. Cianzio. 2000. A Allele for Elevated Palmitate in Soybean. Crop Sci. 40:647-650.

Stoutjesdijk, P.A., S.P. Singh, Q. Liu, C.J. Hurlstone, P.A. Waterhouse, and A.G. Green. 2002. HpRNA-mediated targeting of the *Arabidopsis FAD2* gene gives highly efficient and stable silencing. Plant Physiol. 129:1723-1731.

Stymne, S., and L.-Å. Appelqvist. 1978. The Biosynthesis of Linoleate from Oleoyl-CoA via Oleoyl-Phosphatidylcholine in Microsomes of Developing Safflower Seeds. European Journal of Biochemistry 90:223-229.

Sunilkumar, G., L.M. Campbell, M. Hossen, J.P. Connell, E. Hernandez, A.S. Reddy, C.W. Smith, and K.S. Rathore. 2005. A comprehensive study of the use of a homologous promoter in antisense cotton lines exhibiting a high seed oleic acid phenotype. Plant Biotechnology Journal 3:319-330.

Takagi, Y., A.B.M. Mamun Hossain, T. Yanagita, and S. Kusaba. 1989. High linolenic acid mutant in soybean induced by X-ray irradiation. Japan J. Breed. 39:403-409.

Tang, G., W. Novitzky, H. Griffin, S. Huber, and R. Dewey. 2005. Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. Plant Journal 44:433 - 446.

Tanhuanpää, P., J. Vilkki, and M. Vihinen. 1998. Mapping and cloning of *FAD2* gene to develop allele-specific PCR for oleic acid in spring turnip rape (*Brassica rapa* ssp. oleifera). Molecular Breeding 4:543-550.

Tao, F., S.-W. Zhu, J. Fan, and B.-J. Cheng. 2006. Cloning and Sequence Analysis of Maize *FAD2* Gene. Journal of Plant Physiology and Molecular Biology 6.

Taton, M., and A. Rahier. 1996. Plant Sterol Biosynthesis: Identification and Characterization of Higher Plant [Delta]7-Sterol C5(6)-Desaturase. Archives of Biochemistry and Biophysics 325:279-288.

Till, B., J. Cooper, T. Tai, P. Colowit, E. Greene, S. Henikoff, and L. Comai. 2007. Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biology 7:19.

Till, B., S. Reynolds, C. Weil, N. Springer, C. Burtner, K. Young, E. Bowers, C. Codomo, L. Enns, A. Odden, E. Greene, L. Comai, and S. Henikoff. 2004. Discovery of induced point mutations in maize genes by TILLING. BMC Plant Biology 4:12.

Töpfer, R., N. Martini, and J. Schell. 1995. Modification of Plant Lipid Synthesis. Science 268:681-686.

Uauy, R., D. Hoffman, P. Peirano, D. Birch, and E. Birch. 2001. Essential fatty acids in visual and brain development. Lipids 36:885-895.

Upchurch, R., and M. Ramirez. 2010. Gene Expression Profiles of Soybeans with Mid-Oleic Acid Seed Phenotype. Journal of the American Oil Chemists' Society 87:857-864.

Valsta, L.M., M. Jauhiainen, A. Aro, M.B. Katan, and M. Mutanen. 1992. Effects of a monounsaturated rapeseed oil and a polyunsaturated sunflower oil diet on lipoprotein levels in humans. Arteriosclerosis, Thrombosis, and Vascular Biology 12:50-57.

Vijay, I.K., and P.K. Stumpf. 1972. Fat Metabolism in Higher Plants. Journal of Biological Chemistry 247:360-366.

Vossen, P. 2007. Olive oil: history, production, and characteristics of the World‟s classic oils. Horticultural science 42:1093-1100.

Wagner, N., A. Mroczka, P.D. Roberts, W. Schreckengost, and T. Voelker. 2010. RNAi trigger fragment truncation attenuates soybean *FAD2*-1 transcript suppression and yields intermediate oil phenotypes. Plant Biotechnology Journal:1-8.

Walker, J.B., W.R. Fehr, G.A. Welke, E.G. Hammond, D.N. Duvick, and S.R. Cianzio. 1998. Reduced-Linolenate Content Associations with Agronomic and Seed Traits of Soybean. Crop Sci. 38:352-355.

Wang, H.-W., J.-S. Zhang, J.-Y. Gai, and S.-Y. Chen. 2006. Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean. Theoretical Applied Genetics 112.

Warner, K., and M. Gupta. 2003. Frying quality and stability of low-and ultra-low-linolenic acid soybean oils. Journal of the American Oil Chemists' Society 80:275-280.

Warner, K., and W. Fehr. 2008. Mid-Oleic/Ultra Low Linolenic Acid Soybean Oil: A Healthful New Alternative to Hydrogenated Oil for Frying. Journal of the American Oil Chemists' Society 85:945-951.

Warner, K., P. Orr, and M. Glynn. 1997. Effect of fatty acid composition of oils on flavor and stability of fried foods. Journal of the American Oil Chemists' Society 74:347-356.

Waterman, E. 2007. Active Components and Clinical Applications of Olive Oil. Alternative Medicine Review 12:331-342.

Werteker, M., A. Lorenz, H. Johannes, E. Berghofer, and C.S. Findlay. 2010. Environmental and Varietal Influences on the Fatty Acid Composition of Rapeseed, Soybeans and Sunflowers. Journal of Agronomy and Crop Science 196:20-27.

Weselake, R.J., S. Shah, M. Tang, P.A. Quant, C.L. Snyder, T.L. Furukawa-Stoffer, W. Zhu, D.C. Taylor, J. Zou, A. Kumar, L. Hall, A. Laroche, G. Rakow, P. Raney, M.M. Moloney, and J.L. Harwood. 2008. Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. J. Exp. Bot.:ern206.

White, P.J. 2007. Fatty acid in oilseeds (Vegetable oils), p. 210-263, In C. K. Chow, ed. Fatty acids in foods and their health implications. CRC press, Marcel Dekker, Inc., New York.

Wilcox, J., J. Cavins, and N. Nielsen. 1984. Genetic alteration of soybean oil composition by a chemical mutagen. Journal of the American Oil Chemists' Society 61:97-100.

Wilcox, J.R., A.D. Nickell, and J.F. Cavins. 1993. Relationships between the fan Allele and Agronomic Traits in Soybean. Crop Sci. 33:87-89.

Wilson, R. 2004. Seed composition, In Boerma HR and S. JE, eds. Soybeans: improvement, production, and uses. ASA,Madison, pp 621-677.

Wolf, R., J. Cavins, R. Kleiman, and L. Black. 1982. Effect of temperature on soybean seed constituents: Oil, protein, moisture, fatty acids, amino acids and sugars. Journal of the American Oil Chemists' Society 59:230-232.

Yadav, N.S., A. Wierzbicki, M. Aegerter, C.S. Caster, L. Perez-Grau, A.J. Kinney, W.D. Hitz, J.R. Booth Jr., B. Schweiger, K.L. Stecca, S.M. Allen, M. Blackwell, R.S. Reiter, T.J. Carlson, S.H. Russell, K.A. Feldmann, J. Pierce, and J. Browse. 1993. Cloning of Higher Plant [omega]-3 Fatty Acid Desaturases. Plant Physiology 103:467-476.

Yen, C.-L.E., S.J. Stone, S. Koliwad, C. Harris, and R.V. Farese, Jr. 2008. Thematic Review Series: Glycerolipids. *DGAT* enzymes and triacylglycerol biosynthesis. J. Lipid Res. 49:2283-2301.

Zhang, D., I.L. Pirtle, S.J. Park, M. Nampaisansuk, P. Neogi, S.W. Wanjie, R.M. Pirtle, and K.D. Chapman. 2009. Identification and expression of a new delta-12 fatty acid desaturase (*FAD2*-4) gene in upland cotton and its functional expression in yeast and *Arabidopsis* thaliana plants. Plant Physiology and Biochemistry 47:462-471.

Zheng, P., W.B. Allen, K. Roesler, M.E. Williams, S. Zhang, J. Li, K. Glassman, J. Ranch, D. Nubel, W. Solawetz, D. Bhattramakki, V. Llaca, S. Deschamps, G.-Y. Zhong, M.C. Tarczynski, and B. Shen. 2008. A phenylalanine in *DGAT* is a key determinant of oil content and composition in maize. Nat Genet 40:367-372.

Zou, J., V. Katavic, E.M. Giblin, D.L. Barton, S.L. MacKenzie, W.A. Keller, X. Hu, and D.C. Taylor. 1997. Modification of Seed Oil Content and Acyl Composition in the Brassicaceae by Expression of a Yeast sn-2 Acyltransferase Gene. Plant Cell 9:909-923.

CHAPTER 2

Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait

ABSTRACT

The alteration of fatty acid profiles in soybean [*Glycine max* (L.) Merr.] to improve soybean oil quality is an important and evolving theme in soybean research to meet nutritional needs and industrial criteria by soybean oil processors and end-users. Soybean oil with elevated oleic acid is desirable because this monounsaturated fatty acid improves the nutrition, flavor and oxidative stability of the oil. The objective of this work was to create the high oleic acid trait in soybeans by identifying and combining mutations in two omega-6 fatty acid desaturase genes, *FAD2-1A* and *FAD2-1B*. Soybean plant introduction (PI) germplasm lines that contained elevated oleic acid content in the seed oil were characterized for the sequence and genetic association of their *FAD2-1A* and *FAD2-1B* alleles. Three polymorphisms found in the *FAD2-1B* alleles of two soybean lines resulted in missense mutations. For each of the two soybean PI lines, there was one unique amino acid change within a highly conserved region of the protein. The sequence polymorphisms were developed into highly efficient molecular markers for the mutant alleles. The mutant *FAD2-1B* alleles were associated with an increase in oleic acid levels, although the *FAD2-1B* mutant alleles alone were not capable of producing a high oleic acid phenotype. When existing *FAD2*-1A mutations were combined with the novel mutant *FAD2-1B* alleles, a high oleic acid phenotype was recovered only for those lines that were homozygous for both of the mutant alleles. The high oleic acid soybean germplasm developed contained a desirable fatty acid profile, and it was stable in multiple environments. The resources described here for the creation of high oleic acid soybeans provide a framework to efficiently develop soybean varieties to meet changing market demands.

INTRODUCTION

Soybean oil consumed in the U.S. accounted for approximately 70% of the total U. S. edible fat and oil consumption in 2008; three quarters of the soybean oil was used as cooking oil and baking and frying fat [http://www.soystats.com/2009/]. Soybean oil"s utilization is determined by its fatty acid composition, with commodity soybean oil typically containing 13% palmitic acid (16:0), 4% stearic acid (18:0), 20% oleic acid (18:1), 55% linoleic acid (18:2), and 8% linolenic acid (18:3). Consumption of oils with high oleic acid content is desirable because this monounsaturated fatty acid not only improves flavor and shelf life but also reduces the need for hydrogenation, a process adding to the cost of the oil and generating unwanted trans-fat that has been linked to many health problems in humans (Ascherio and Willett, 1997). Additionally, in biodiesel production, there is also a need for oils high in oleic acid and low in saturated fatty acids in order to improve the oxidative stability while augmenting cold flow (Raneses et al., 1999). Moreover, the enhanced oxidative stability of soybean oil with higher oleic acid content will also open up a variety of food uses and industrial applications like spraying oils or machine lubricants (Butzen and Schnebly, 2007).

In the lipid biosynthetic pathway, conversion of oleic acid (18:1) precursors to linoleic acid (18:2) precursors is catalysed by the omega-six fatty acid desaturase 2 enzyme (*FAD2*) (Okuley et al., 1994; Schlueter et al., 2007). While in *Arabidopsis* and maize only one copy of a *FAD2* gene was identified (Beló et al., 2008; Okuley et al., 1994), soybean appears to possess multiple copies of the gene due to the consequence of repeated rounds of genome duplication followed by limited sequence loss (Schlueter et al., 2007). The soybean *FAD2* gene family has been previously characterized at the

genome level for structure and expression (Schlueter et al., 2007). Among the *FAD2* genes identified in soybean, the *FAD2*-2 desaturases consisting of *FAD2-2A* (Glyma19g32930), *FAD2-2B* (Glyma19g32940), and *FAD2-2C* (Glyma03g30070) were found to be widely expressed in the vegetative tissues of the soybean plant (Schlueter et al., 2007). The exception was *FAD2-2A*, for which expression was not detected; *FAD2-2A* was predicted to be nonfunctional as it has a deletion of 100 bp in the coding region (Schlueter et al., 2007). The two microsomal FAD2-1 desaturases *FAD2-1A* (Glyma10g42470) and *FAD2-1B* (Glyma20g24530) were mainly expressed in developing seeds (Schlueter et al., 2007; Tang et al., 2005). Thus, *FAD2-1A* and *FAD2-1B* are considered to play an important role in controlling the oleic acid level in developing soybean seeds and were selected as candidate genes to elucidate the molecular genetic basis of soybean lines from the germplasm collection that contained elevated levels of oleic acid.

FAD2-1A and *FAD2-1B* are most closely related to one another, with a shared genomic organization containing a single intron and 99% identity in encoded amino acid sequence, and are present on homologous chromosome regions mapped to linkage group O (chromosome 10) and I (chromosome 20), respectively (Schlueter et al., 2007; Tang et al., 2005). Characterization of the expression of the individual soybean *FAD2* genes confirmed the importance of *FAD2-1A* and *FAD2-1B* with expression of these genes during peak oil synthesis; a possible role was also revealed for *FAD2-2C* under cool temperature conditions (Schlueter et al., 2007). Differential response to temperature was also demonstrated for the soybean FAD2-1A and FAD2-1B enzymes expressed in yeast (Tang et al., 2005). The temperature during soybean pod fill has been shown to influence

the fatty acid composition of soybean oil, with cooler temperatures leading to decreased oleic acid accumulation (Heppard et al., 1996; Lee et al., 2009; Oliva et al., 2006).

Selection for breeding and genetic engineering resulting in elevated oleic acid levels were reported in many oilseed crops: safflower (Knowles and Hill, 1964), sunflower (Liu et al., 2002), peanut (Bruner et al., 2001; Jung et al., 2000; Patel et al., 2004), canola (Hu et al., 2006; Stoutjesdijk et al., 2000), cotton (Liu et al., 2002) and maize (Beló et al., 2008). While the elevated oleic acid phenotype was often observed after a single *FAD2* gene was mutated, the very highest oleic acid phenotypes (> 80% of the total oil) were achieved most frequently by silencing all copies of *FAD2* genes that were expressed in developing seeds, or in particular for peanut, by combining the mutation in the active site of ah*FAD2A* with the loss of transcription of ah*FAD2B* (Jung et al., 2000; Patel et al., 2004). No soybean lines exist in the USDA National Plant Germplasm System collection with the high oleic acid trait (oleic acid content above 70% of the oil fraction), although multiple lines contain elevated oleic acid levels (Lee et al., 2009).

Several elevated oleic acid soybean lines have been characterized at the molecular level. The destruction of the *FAD2-1A* gene by X-ray mutagenesis yielded two soybean lines with oleic acid content of approximately 50% of the oil (Alt et al., 2005a; Anai et al., 2008; Sandhu et al., 2007). A reverse genetics approach was utilized to identify a soybean line containing a missense mutation in *FAD2-1A* that associated with an elevated oleic acid content of the oil (Dierking and Bilyeu, 2009). Many soybean lines were developed through recurrent selection that contained elevated oleic acid content such as N00-3350, N98-4445A, and N97-3363-3 (Burton et al., 2006), but the genetic basis for

the trait was extremely complex, with at least six QTLs conditioning the phenotype (Bachlava et al., 2009; Monteros et al., 2008). In addition, the level of oleic acid in the oil of these soybean lines was particularly susceptible to environmental effects when compared to the X-ray *FAD2-1A* deletion line (Lee et al., 2009; Oliva et al., 2006). No mutations in *FAD2*-1B that associate with elevated oleic acid content have been reported to date.

Suppression of *FAD2-1* gene expression by means of genetic engineering has been successful in creating soybean lines with the high oleic acid trait, with oleic acid content above 80% of the oil fraction reported and very little environmental impact on the trait (Buhr et al., 2002; Butzen and Schnebly, 2007; Graef et al., 2009). Transgenic expression of ribozyme terminated sense or antisense *FAD2-1* constructs was successful in eliminating the *FAD2-1* mRNA expression signal in developing embryos and producing siRNAs for the *FAD2-1* genes (Buhr et al., 2002). No attempt was made to distinguish between the *FAD2-1A* and *FAD2-1B* genes in the transgenic work, and it is also possible that the *FAD2*-2 genes were targeted in the experiments (Buhr et al., 2002).

The objective of this work was to create the high oleic acid trait in soybeans using conventional plant breeding technology. We hypothesized that combinations of mutant alleles of the soybean *FAD2-1A* and *FAD2-1B* genes would greatly reduce the FAD2 enzyme activity in developing seeds and thus result in an accumulation of oleic acid at the expense of linoleic and linolenic acid in the triacylglycerol fraction of the seed oil. We took a candidate gene approach with the *FAD2-1A* and *FAD2-1B* genes present in soybean germplasm accessions containing elevated levels of oleic acid in the oil. Two mutant alleles of *FAD2-1B* were identified that associated with elevated oleic acid

content. In addition, mutant alleles of *FAD2-1A* and *FAD2-1B* were combined to create soybeans with the high oleic acid trait.

MATERIALS AND METHODS

Population development

Recombinant inbred line from (RIL) population 1 (F_6 RIL of Jake x PI 283327), 2 ($F_{2:6}$ and $F_{2:7}$ RIL of M23 x PI283327) and 3 ($F_{2:5}$ and $F_{2:7}$ RIL of M23 x PI 567189 A) were created at the same time. Three crosses were made in summer 2005 at the Delta Research Center at Portageville, MO including Jake x PI 283327, M23 x PI 283327 and M23 x PI 567189A. PI 283327 and PI 567189A are two elevated oleic acid lines in maturity group V and IV, respectively (GRIN USDA), while Jake is a conventional high yielding soybean in group V that contains a typical oleic acid content (Shannon et al., 2007). M23 was selected for elevated oleic acid after mutagenesis of the cultivar Bay (Takagi and Rahman, 1996). In 2005 and early 2006, F_1 seeds were advanced to the F_2 generation in Costa Rica. Each RIL tracing to a single F_2 plant except population 1 was also advanced in Costa Rica from 2006 to 2007 for F_5 seeds. In 2007, a bulk of five seeds from each RIL in each population was analyzed to obtain fatty acid profile for the Costa Rica location. Population 1 was grown in Portageville, MO to produce F_7 seeds. Population 2 was grown in Portageville, MO to produce F_6 seeds, and then soybean RILs with more than 60% oleic acid were advanced to the F_7 generation. In population 3, only F_5 RILs producing more than 60% oleic acid were selected to generate F_7 seeds at Portageville, MO in subsequent generations.

Population 4 was developed from the cross 17D x S08-14788 which was created in the summer of 2008 at Portageville, MO. S08-14788 was selected from population 1 because it carried the *FAD2-1B* P137R mutant alleles derived from PI 283327. 17D is an

elevated oleic soybean line developed by mutagenesis with 35% oleic acid content (Dierking and Bilyeu, 2009). True F_1 seeds were sent to Costa Rica and F_2 seeds were produced in the winter 2009. $F_{2:3}$ seeds were produced in Columbia, MO and Portageville, MO during the summer 2009 growing season.

Population 5 was initiated in summer 2008 at Portageville, MO. Soybean line KB07-1#123 was crossed with soybean line #93 from population 2. Soybean line #93 (>80% oleic acid) was genotyped to contain the *FAD2-1A* Δ alleles from M23 and the *FAD2-1B* P137R alleles derived from PI 283327. KB07-1#123 is a soybean line with the pedigree [W82 x (M23 x 10-73)]. This soybean line was selected to contain three mutant alleles affecting the fatty acid profile, including *FAD2-1A* Δ alleles from M23, and mutant *FAD3A* and *FAD3C* alleles from soybean line 10-73 (Bilyeu et al., 2005; Dierking and Bilyeu, 2009). F_1 seeds were genotyped to confirm the heterozygosity and then advanced to obtain F_2 seeds in summer 2009 at Bradford Research and Extension Center, Columbia MO.

In 2008, populations 1 and 2 were grown in Portageville, MO to produce the seeds analyzed for fatty acids in figures 2.2 and 2.3. In 2009, population 4 was grown in Columbia, MO to produce the seeds analyzed for fatty acid analysis in figure 2.5. Data in figure 2.4 was from F_5 seeds of population 3 produced in Costa Rica. In addition, five lines with the highest oleic acid content from populations 2 and 3 were grown in Columbia, MO in 2009. Similarly, four to eleven lines from each of four combinations of homozygous *FAD2-1A* and *FAD2-1B* genes from population 4 were grown in Columbia MO and selected lines from population 4 were grown in Portageville, MO in 2009.

DNA isolation and PCR for sequencing of FAD2-1A and FAD2-1B

Genomic DNA was isolated from approximately 30mg using the DNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA) and used at 5 to 50 ng per PCR reaction. PCR was carried out using Ex Taq according to manufacturer‟s recommendation (Takara, Otsu, Shiga, Japan) in a PTC-200 thermocycler (MJ Research/Bio-Rad, Hercules, CA). Primers for *FAD2-1A*: forward was ACTGCATCGAATAATACAAGCC and reverse was TGATATTGTCCCGTGCAGC. Primers for *FAD2-1B*: forward was CCCGCTGTCCCTTTTAAACT and reverse was TTACATTATAGCCATGGATCGCTAC. PCR was programmed as the following: 95 ^{o}C for 5 minutes followed by 35 cycles of 95 ^{o}C for 30 seconds, 60 ^{o}C for 30 seconds, and 72 ^{o}C for 1 minute 30 seconds. PCR products were examined for size by running on Flashgel (Lonza Group Ltd., Switzerland) for 5 minutes. PCR products were then isolated with the Qiaprep Spin Miniprep kit (Qiagen, Inc.) and sequenced at the University of Missouri DNA Core facility.

Sequence analysis

Sequences were aligned using Multiple Sequence Alignment by CLUSTALW (http://align.genome.jp/), and evaluated for variant nucleotides between „Williams 82‟ reference (http://www.phytozome.net/soybean) and the PIs. Protein translation was conducted using ExPaSy (http://ca.expasy.org/tools/dna.html) and protein alignment was done using Multiple Sequence Alignment program.

***FAD2-1B* allele specific molecular marker assay**

SimpleProbe assays were based on the disassociation kinetics of SimpleProbe obligonucleotides (Roche Applied Sciences) to be exactly complimentary to the „Williams 82‟ reference sequence. SimpleProbe was purchased from Roche Applied Sciences. The *FAD2-1B* probe consists of 5'-Fuorescence-AGTCCCTTATTTCTCATGGAAAA**T**AAGC--Phosphate-3'. C>G mutation (P137R allele from PI 283327) is indicated by underline and T>C mutation (I143T allele from PI 567189A) is indicated by bold font. Primers used to generate template for Simpleprobe genotyping assay were designed by aligning the *FAD2-1A* and *FAD2-1B* region containing the SNPs. Primers were selected to be as close as possible to the SNPs while differing in at least 3 nucleotides between the two genes to specifically amplify the targeted region in *FAD2-1B*. Genotype reactions used asymmetric PCR to generate additional single stranded DNA to which the Simpleprobe could bind with less competition from the opposite amplification stand. Because the *FAD2-1B* SNPs found in the two PIs were only a few nucleotides apart, the SimpleProbe was designed to detect both of the SNPs. Genotyping reactions were performed with a 5:2 asymmetric mix of primers (5‟-ACTGCATCGAATAATACAAGCC-3‟ at 2 µM final concentration, and 5‟-TGATATTGTCCCGTCCAGC-3‟ at 5 µM final concentration). Reactions were carried out in 20 µl; containing template, primers, 0.2 µM final concentration of SimpleProbe, buffer (40 mM Tricine-KOH [pH 8.0] 16mM $MgCl_2$, 3.75 µg ml^{-1} BSA,), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Genotyping reactions were performed using a Lightcycler 480 II real time PCR instrument (Roche), using the following PCR parameters: 95 ^{o}C for 5 minutes followed

by 40 cycles of 95 ^{o}C for 20 seconds, 60 ^{o}C for 20 seconds, 72 ^{o}C for 20 seconds, and then a melting curve from 50 ^{o}C to 70 ^{o}C. When DNA from PI 283327 and PI 567189A is amplified with gene specific primers and used in melting curve analysis with the SimpleProbe, a mismatch between the Simpleprobe and the amplicon results in altered disassociation kinetics. Each genotype produced a characteristic melting profile, as measured by Tm of the negative first derivative of the disappearance of fluorescent signal. PI 283327 and all soybean lines with an identical *FAD2-1B* allele genotype have a characteristic peak at 56.7 ^{o}C, while the PI 567189A *FAD2-1B* allele genotype yielded a characteristic peak at 60.2 ^{o}C. M23 and Jake (wild-type *FAD2-1B*) have a peak at 62.5 ^{o}C. Heterozygous individuals‟s genotype showed two peaks at either 56.7 ^{o}C or 60.2 ^{o}C and 62.5 ^{o}C.

***FAD2-1A* allele specific molecular marker assay for 17D** was conducted as described by Dierking 2009 (Dierking and Bilyeu, 2009).

***FAD2-1A* allele specific molecular marker assay for M23**

An allele specific molecular marker assay was developed to distinguish soybean lines with deletion of *FAD2-1A* (*FAD2-1A* Δ alleles from M23) and the soybean lines with the presence of one (heterozygous) or two *FAD2-1A* alleles.

The reactions contain two primer pairs: one pair specific for *FAD2-1A* gene amplification and one pair specific for PEPC16 gene amplification (The *FAD2-1A* primer pair has forward primer T2AF: (5‟-ATCTTTAGATTTTTCACTACCTGGTTTAAAATTGAGGGATTG-3‟) and reverse primer HOLL1 (5‟- CTTTGCTAGACCCTGTGTCAAAGTATAAAC-3‟). The PEPC16

primer pair has forward primer PEPC16fwd (5‟-TTCCTTTATCAGAAATAACGAGTTTAGCT-3‟) and reverse primer PEPC16rev (5-TGTCTCATTTTGCGGCAGC-3‟).

Reactions were carried out in 15 µl; each primer was at 1.3 µM final concentration in reactions containing template, buffer (40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM $MgCl_2$, 3.75 µg ml^{-1} BSA, 200 µM dNTPs), 5% DMSO, 1.25 µM EvaGreen (Biotium Inc., Hayward, CA) and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). PCR parameters on a DNA Engine Opticon 2 (MJ Research/Bio-Rad) were as follows: 95 ºC for 5 minutes followed by 35 cycles of 95 ºC for 20 seconds, 64 ºC for 20 seconds, 72 ºC for 20 seconds, and then a melting curve from 70 ºC to 85 ºC. The fluorescence was read after each cycle and every 0.2 ºC with a one second hold during the melt with excitation at 470-505 nm and detection at 523-543 nm. Each genotype produced a product with a characteristic melting profile, as measured by Tm of the negative first derivative of the disappearance of fluorescent signal. Homozygous wild-type *FAD2*-1A alleles and heterozygous samples produced a peak at 76 ^{o}C and possibly another peak at 78 ^{o}C; homozygous mutant alleles (*FAD2-1A* Δ) only produced a peak at 78 ºC. Templates for PCR were either genomic DNA samples isolated using the DNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA) or 1.2 mm washed FTA (Whatman) card punches prepared from leaves according to the manufacturer‟s instructions.

Fatty acid, protein and oil determination

The method of gas chromatography of total fatty acid methyl esters of extracted oil was used to examine the fatty acid profiles of all samples (Beuselinck et al., 2006).

For the RIL populations, individual whole crushed seeds for each soybean line were used as samples for fatty acid determination, except the *FAD2-1aabb* lines from population 3 produced in Costa Rica, for which a bulk sample of 5 seeds was analyzed for each line. For the F_2 seeds from population 4 and population 5, seed chips were used by removing a "chipped" portion of the seed opposite the embryo with a scalpel for fatty acid analysis such that the remainder of the seed containing the embryo could be germinated.

Protein and oil contents were determined for seeds of $F_{2:3}$ lines for population 4 using NIR spectroscopy (Hartwig and Hurburgh, 1990).

Population genotyping

For every RIL population, a seed from each line was germinated in a germination package. Eight to ten days later, a small developing unifoliate leaf was excised and pressed onto an FTA card for DNA storage. Based on fatty acid profiles obtained from the other five seeds, to obtain a set of the contrasting homozygous *FAD2-1A* and *FAD2-1B* genotypes, selected lines of the three populations Jake x PI 283327, M23 x PI 283327, M23 x PI 567189A, were genotyped with all allele- specific assays as described. For the F_2 population of the population 4 17D x S08-14788 cross, 200 seeds were chipped, the small portion of each without hypocotyls was sent for fatty acid profiling, the remaining chipped seed was germinated to collect DNA for genotyping. Not all samples were gentoyped.

RESULTS

Identification of mutant alleles of *FAD2-1B* in soybean lines PI 283327 and PI 567189A

In an effort to identify novel alleles of the soybean *FAD2-1A* and *FAD2-1B* genes, genomic DNA was characterized for the sequence of both genes from plant introduction (PI) lines selected from the National Genetic Resources Program containing elevated oleic acid levels from 27 to approximately 50% percent of the oil, compared to commodity soybeans which produce 19-23% oleic acid (Lee et al., 2007). The *FAD2-1A* alleles from PI 283327 and PI 567189 A were identical to the reference „Williams 82"(Bernard and Cremeens, 1988) allele. In contrast, for the *FAD2-1B* genes from PI 283327 and PI 567189 A, seven common single nucelotide polymorphisms (SNPs) and one unique SNP for each line were identified when compared to the reference sequences from the cultivar „Williams 82" (Figure 2.1A). Other soybean lines were also characterized for their *FAD2*-1B alleles, and two independent lines, PI 210179 and PI 578451, contained exactly the same *FAD2-1B* alleles as PI 283327 and PI 567189 A, respectively.

For the PI 283327 *FAD2-1B* allele, there were three missense mutations resulting in S86F, M126V, and P137R. For the PI 567189 A *FAD2-1B* allele, two of the missense mutations were identical (S86F and M126V), and the third missense mutation was unique, I143T. Other soybean lines were identified that had *FAD2-1B* alleles which contained different combinations of the silent mutations and the S86F and M126V missense mutations (data not shown).

To predict the potential effect of the amino acid changes to soybean *FAD2-1B* enzyme function, we used the program PolyPhen to analyze the potential severity of each amino acid change (Ramensky et al., 2002). In addition, the relative amino acid conservation for each position in the enzyme was evaluated visually using Weblogo after alignment of 100 FAD2 protein sequences present in the National Center for Biotechnology Information database (Crooks et al., 2004). The two shared amino acid substitutions M126V and S86F each occur at a highly variable position according to the alignment in Weblogo and were classified by Polyphen as benign substitutions, indicating they are likely to have no phenotypic effect. For the M126V change, methionine and valine have similar chemical properties.

FAD2-1B alleles that encoded only the S86F misssense mutation were identified and analyzed for the functional consequence of that isolated mutation. Although serine and phenylalanine are amino acids with different chemical properties, several lines of evidence indicated that the S86F *FAD2-1B* allele was functional, including detection of omega-6 fatty acid desaturase activity in a yeast recombinant expression experiment and no association of the S86F *FAD2-1B* allele variant with an elevated oleic acid phenotype (data not shown).

The *FAD2-1B* P137R mutation present in PI 283327 represents a charge change for the substituted amino acid since proline is nonpolar while arginine is classified as charged (basic). At position 137 of wild-type *FAD2-1B*, the proline is perfectly conserved; a proline at that position is invariant for all of the tested *FAD2* sequences represented in the protein database (Figure 2.1B). In PolyPhen, the amino acid change

was classified as probably damaging for P137R, which means that this SNP is predicted to affect protein structure and/or function.

For the isoleucine present at position 143 of wild-type *FAD2-1B*, other amino acids were observed in this position in the FAD2 sequences present in the protein database, indicating that an isoleucine at position 143 is less conserved than the proline at the 137 position (Figure 2.1B). However, the substitution of threonine for isoleucine at this position was not observed in the database analysis. Moreover, isoleucine and threonine also have contrasting chemical properties since isoleucine is nonpolar while threonine is uncharged polar. In PolyPhen, the I143T amino acid change was classified as probably damaging. The conservation of amino acids in the general region of the *FAD2-1B* P137R and I143T mutations combined with the chemical nature of the changes is suggestive of the potential deleterious effects of these mutations to the *FAD2-1B* enzyme‟s structure and function.

The PI 283327 *FAD2-1B* allele is associated with an increase in seed oleic acid content

To test the hypothesis that the newly identified mutations in *FAD2-1B* are causative for the elevated oleic acid level in the plant introduction lines, an analysis of the oleic acid phenotype and *FAD2-1B* genotype association was examined for Population 1, an F_6 recombinant inbred line (RIL) population developed from the cross „Jake‟ (Shannon et al., 2007) x PI 283327. The commodity soybean line Jake typically produces approximately 22% oleic acid in the seed oil and contains functional *FAD2-1A* and *FAD2-1B* alleles, represented as genotype *FAD2-1* AABB. PI 283327 was selected from

the germplasm collection because of elevated levels of oleic acid in the seed oil and carries the mutant *FAD2-1B* P137R allele, represented as genotype *FAD2-1* AAbb. Of the 54 lines in the RIL population 1 that contained homozygous alleles of *FAD2-1B*, the 30 lines carrying the mutant *FAD2-1B* P137R alleles from PI 283327 had an average of 29.4% oleic acid, while 24 lines carrying wild type alleles had an average of 20.5% oleic acid (Figure 2.2). Although the variation in the data was large, the difference in oleic acid contents between the two contrasting *FAD2-1B* genotypes was confirmed significant using Student‟s t-test at the 0.05 probability level (P>0.05).

Combinations of mutations in *FAD2-1A* and *FAD2-1B* produce high oleic acid levels in the seed oil

We hypothesized that combining the mutant alleles of both *FAD2-1A* and *FAD2-1B* in one soybean line would eliminate most of the enzyme activity responsible for converting oleic acid precursors into linoleic acid precursors, and thus result in a higher oleic acid content compared to soybean lines containing mutations in either gene individually. Two mutations in *FAD2-1A* were available: the deletion of the *FAD2-1A* gene in soybean line M23 (designated herein as *FAD2-1A* Δ) (Alt et al., 2005a) and the missense mutation in *FAD2-1A* from line 17D (designated herein as *FAD2-1A* S117N) (Dierking and Bilyeu, 2009). The identification of the two missense mutant alleles in *FAD2-1B* from PI 283327 (*FAD2-1B* P137R) and PI 567189A (*FAD2-1B* I143T) created the opportunity to evaluate the oleic acid phenotype in soybean lines containing different combinations of mutant *FAD2-1A* and *FAD2-1B* alleles.

An association analysis of the oleic acid phenotype and the *FAD2-1A* and *FAD2-1B* genotypes was performed for Population 2, a RIL population consisting of $F_{2:6}$ and $F_{2:7}$ lines developed from the cross M23 x PI 283327 and grown in Portageville MO 2008. Since soybean line M23 contained *FAD2-1A* Δ and wild-type alleles of *FAD2-1B*, the genotype is herein represented as *FAD2-1* aaBB with the lowercase allele designation always specifying the mutant allele and the capital case specifying the wild-type allele; likewise, the *FAD2-1* P137R genotype of PI 283327 is represented here as *FAD2-1* AAbb. Individual seeds from each of 40 lines produced in an appropriate field environment and carrying the different homozygous combinations of *FAD2-1A* and *FAD2-1B* were analyzed along with the parental lines for the fatty acid phenotype of the seed oil (Figure 2.3). Transgressive segregation for oleic acid content was observed for the lines that inherited the *FAD2-1* AABB and aabb genotypes, while the lines that recovered the parental *FAD2-1* genotypes contained oleic acid contents similar to the phenotype of the parental lines. Lines with the genotype *FAD2-1* AABB had an average oleic acid content similar to that of a conventional soybean „Jake", which was 22.6% of total oil content. In contrast, individuals with the genotype *FAD2-1* aabb had an average of 82.2% oleic acid, with a very narrow standard deviation of 1.2%. In this population, lines with either homozygous mutant *FAD2-1A* or *FAD2-1B* alleles had an average of 39.4% and 30.6% oleic acid, respectively, reiterating the relatively minor increase in oleic acid level conditioned by the *FAD2-1B* alleles in the presence of functional *FAD2-1A* alleles.

A similar experiment investigated the impact on seed oleic acid levels for the genetic combination of the *FAD2-1A* Δ alleles from M23 with the alternate *FAD2-1B*

I143T alleles present in PI 567189 A. For Population 3, a M23 x PI 567189A RIL population, we genotyped lines based on the fatty acid profiles of F_5 seeds harvested in Costa Rica, and then selected 31 lines classified into four homozygous combinations of *FAD2-1A* and *FAD2-1B*. The average oleic acid contents of the four *FAD2-1* genotypes were significantly different from each other and again demonstrated transgressive segregation (Figure 2.4). Lines with *FAD2-1* genotype AABB had an average oleic acid content not statistically different from that of Jake. In contrast, soybean lines with the *FAD2-1*aabb genotype had an average of 80.3% oleic acid. The lines that recovered the parental *FAD2-1* genotypes also recovered the respective parental oleic acid phenotype.

Soybean line 17D was discovered in a reverse genetics screen for mutations in *FAD2-1A* (Dierking and Bilyeu, 2009). Line 17D contains elevated oleic acid in the seed oil due to a FAD2-1A missense mutation in a conserved amino acid referred to here as *FAD2-1A* S117N. Soybean lines which contained the *FAD2-1A* S117N alleles consistently accumulated lower oleic acid levels in the seed oil than lines containing *FAD2-1A* Δ alleles derived from M23, and the phenotype was not stable in different environments (Dierking and Bilyeu, 2009). We next examined the combining ability in Population 4, the F_2 and $F_{2:3}$ individuals derived from the cross of the *FAD2-1A* S117N alleles derived from line 17D with the *FAD2-1B* P137R alleles derived from PI 283327 (17D x S08-14788). Homozygous *FAD2-1A* and *FAD2-1B* allele combinations were selected from *FAD2-1* genotyped F_2 plants for field growth in an appropriate environment and subsequent F_3 seed oil fatty acid phenotype determination (Figure 2.5). Transgressive segregation was observed for the genotypes that inherited both the homozygous wild-type *FAD2-1* alleles and the homozygous mutant *FAD2-1* allele

combinations. The *FAD2-1* aabb combination demonstrated an average oleic acid content of 77.3%; the AABB combination displayed a typical commodity soybean oleic acid level. The parental oleic acid phenotype was recovered for the *FAD2-1* aaBB genotype but not for the *FAD2-1* AAbb genotype.

Excess desaturase activity: A single wild-type FAD2-1 allele prevents high oleic acid accumulation

Our initial investigation of both the *FAD2-1* genotype and fatty acid phenotype in F_2 seeds from Population 4 (*FAD2-1A* S117N x *FAD2-1B* P137 cross) demonstrated the epistatic nature of the mutant alleles working in combination, and the results revealed that only homozygous combinations of both mutant *FAD2-1A* and *FAD2-1B* were capable of producing the high oleic acid phenotype. Of the 200 F_2 seeds that were phenotyped, there were 12 individual F_2 seeds with genotype *FAD2-1* aabb, and they had an average oleic acid content of 81%, ranging from 75.2% to 83.9% oleic acid (Figure 2.6). The next highest oleic acid phenotype in the set was 48.8%, and that seed had the *FAD2-1* Aabb genotype. For a two recessive gene model, one sixteenth of the individuals should inherit the phenotype; recovery of 12 individuals with the high oleic acid phenotype satisfies this expectation by Chi-Square test at the 0.05 probability level.

Individuals with a single wild-type version of either *FAD2-1A* or *FAD2-1B* in combination with three mutant *FAD2-1* alleles (Aabb or aaBb) contained approximately 40% oleic acid. No seeds from any of the other *FAD2-1* genotypes contained oleic acid levels above 49% of the seed oil. Individuals with two or more wild-type *FAD2-1* alleles contained oleic acid content with a range of 18-47% of the seed oil.

The necessity of the homozygous *FAD2-1A* and *FAD2-1B* mutant combination requirement for the high oleic acid phenotype was confirmed in an independent analysis of *FAD2-1* genotype and fatty acid phenotype of field produced F_2 seeds that contained homozygous *FAD2-1A* Δ alleles but which were segregating for *FAD2-1B* P137R alleles (Population 5). While the average oleic acid level of those seeds with the *FAD2-1* aabb genotype was 82.5%, *FAD2-1* aaBb seeds averaged 55.4%; *FAD2-1* aaBB seeds averaged 43.4% oleic acid in the seed oil. The presence of a single wild-type version of the *FAD2-1B* allele also prevented a high oleic acid content in the seed oil, although the magnitude of the difference was greater for the F_2 seeds from Population 4.

The high oleic acid phenotype is stable in plants grown in alternate environments

Some of the high oleic acid soybean lines developed in this study demonstrated stability for the high oleic acid trait when grown in different environments (Table 2.1). Of the three environments, Costa Rica typically has the warmest temperatures during seed development, followed by the Portageville, MO environment; the Columbia, MO environment is the coolest of the three environments during seed development (Lee et al., 2009; Oliva et al., 2006). The differences in the oleic acid contents between environments when the *FAD2-1B* P137R alleles were present were minor. Soybean lines with genotype *FAD2-1*aabb of population 2 and 4 produced more than 80% oleic acid content in Costa Rica and Portageville, MO environments, and the oleic acid level was an average of 2-4% lower when grown in the Columbia, MO environment. It is notable that the variation in the phenotype was narrow in all of the environments. In contrast, the *FAD2-1*aabb soybean lines of population 3 containing the *FAD2-1B* I143T alleles had

lower and more variable oleic acid content in the cooler environments, and failed to produce a high oleic acid phenotype in either the Columbia, MO or Portageville, MO environments.

Full fatty acid profiles and total oil and protein content

The full fatty acid profiles of the seeds of contrasting *FAD2-1*genotypic classes produced from Populations 2, 3, and 4 in this study revealed additional alterations in palmitic acid, linoleic acid, and linolenic acid content (Table 2.2). As expected for a major decrease in seed expressed FAD2 enzyme activity that results in an accumulation of oleic acid, the FAD2 reaction products linoleic acid and linolenic acid were dramatically reduced in the high oleic *FAD2-1A* and *FAD2-1B* homozygous mutant lines when either of the *FAD2-1A* mutations were present along with the *FAD2-1B* P137R or I143T alleles.

By evaluating the proportions of oleic, linoleic, and linolenic acids present in the oil extracted from mature seeds, the relative FAD2 and FAD3 desaturase activities of the developing seeds were determined for the contrasting homozygous *FAD2-1* genotypes from each population. The *FAD2-1* AABB genotypes contained FAD2 desaturase activities (final oleic acid content divided by the sum of final oleic, linoleic, and linolenic acid contents) of 76%, 76%, and 74% for Population 2, Population 3, and Population 4, respectively. The *FAD2-1* aabb genotypes contained FAD2 desaturase activities of 7%, 10%, and 14%, for Population 2, Population 3, and Population 4, respectively. Also noted is that the accumulation of linolenic acid follows a different pattern for the *FAD2-1* aabb mutant lines compared to the *FAD2-1* AABB lines, with increased FAD3 desaturase

activity (final linolenic acid content divided by the sum of final linoleic and linolenic acid contents) for the *FAD2-1* mutant lines.

While no significant differences were observed for the stearic acid levels in the contrasting *FAD2-1* genotypes, the *FAD2-1* aabb mutant lines consistently produced lower palmitic acid levels than lines with the *FAD2-1* AABB genotype. The most dramatic change was for Population 2. In that case, the content of palmitic acid was 7.9% for the *FAD2-1* aabb mutant lines compared to 12.3% for the *FAD2-1* AABB lines.

Because of the concern that improvement in fatty acid profiles might have negative impacts on the total oil and protein profiles of the seeds, we also evaluated the protein and oil contents for the field produced $F_{2:3}$ seeds from Population 4. There were no significant differences in the protein or oil contents among the different homozygous *FAD2-1* genotypes, or with those lines compared to either Williams 82 or the 17D parental line. The *FAD2-1B* P137R allele donor parental line had a minor decrease in the average oil content and the highest mean protein content of all of the lines examined.

DISCUSSION

Increasing the oleic acid content in soybean seed oil is one of the most effective and efficient ways to enhance the nutritional value and practical utilization of soybean oil. However, the previously characterized sources of elevated oleic acid soybean involved mutation of the *FAD2-1A* gene alone, which failed to achieve oleic acid levels stable above 60% of the oil (Anai et al., 2008; Dierking and Bilyeu, 2009; Scherder and Fehr, 2008) or utilized approaches that have proven to have limited usage in soybean breeding due to the complexity of the trait (Bachlava et al., 2009). A transgenic approach was successful in downregulation of the *FAD2-1* genes leading to high oleic acid levels in the oil (Buhr et al., 2002; Graef et al., 2009). We have demonstrated here that an allele of the *FAD2-1B* gene containing a conserved amino acid substitution is responsible for the elevated oleic acid content in PI 283327 since soybean lines inheriting the homozygous mutant alleles have higher levels of oleic acid compared to lines inheriting wild type *FAD2-1B* alleles. The PI 567189A *FAD2-1B* allele, which contains a very rare amino acid substitution, was also predicted to have a negative impact on the enzyme activity and function.

The most significant finding of this research was that the mutant *FAD2-1B* alleles from either PI 283327 or PI 567189A not only contribute to the elevated oleic acid content in soybean seed oil of the two germplasm accessions but also enable the accumulation of oleic acid content nearly four-fold, to more than 80% of the oil, in soybean seeds when combined with different sources of a mutant *FAD2-1A* gene. Our results define the requirement of the two major contributors to the FAD2 enzyme activity present in developing soybean seeds, *FAD2-1A* and *FAD2-1B*. Other *FAD2* genes exist

in the soybean genome and some are expressed during seed development. However, when FAD2-1A and FAD2-1B are non-functional, very little FAD2 activity remains in developing seeds, as evidenced by the minor accumulation of linoleic and linolenic acid in the seed oil of the lines containing the *FAD2-1A* and *FAD2-1B* mutant combinations.

Our study demonstrated that different types of *FAD2-1A* mutant alleles bring about a similar high level of oleic acid content in the oil of soybean seeds when combined with the P137R *FAD2-1B* allele from PI 283327. M23 has lost its *FAD2-1A* alleles during the X-ray treatment while 17D carries the alleles with a mutagenesis-induced mutation that appeared to be less effective in accumulating oleic acid in the seed oil (Anai et al., 2008; Dierking and Bilyeu, 2009). Hence, it was assumed that 17D would be less useful than M23 in creating a high oleic acid trait. In contrast to what we expected, the *FAD2-1A* alleles from 17D were able to combine with mutant *FAD2-1B* to produce oleic acid levels similar to those when M23 was the source of the deleted *FAD2-1A* gene, and the phenotype was stable across different environments. We conclude that for the *FAD2-1A* gene from 17D the encoded enzyme was unable to function appropriately, so when in combination with the mutant P137R *FAD2-1B* allele, the conversion of oleic acid precursors to linoleic acid precursors was almost completely blocked. Presumably there is some compensatory difference in the activity of the functional FAD2-1B when the S117N allele of *FAD2-1A* was present compared to the situation from M23 when *FAD2-1A* is deleted.

The occurrence of soybean lines with the high oleic acid phenotype in population 4 fit a model of two independent recessive genes segregating with very little evidence for additional modifier genes, demonstrating that in this circumstance only *FAD2-1A* and

FAD2-1B are contributing to the high oleic acid content in soybean seed oil. Based on our results, we hypothesize that any soybean line carrying a null or mutated *FAD2-1A* allele that is crossed with PI 283327 is likely to produce individual soybean seeds in the F_2 progeny with at least 80% oleic acid content of the seed oil.

Although there is evidence of influence of temperature on the soybean seed oleic acid content(Lee et al., 2009; Oliva et al., 2006) , two of our three high oleic acid soybean genotypes proved to be capable of producing a high and stable oleic acid content in three environments. Moreover, there was no reduction in oil and protein content in the evaluated high oleic acid soybean lines. Soybean lines with the combination of *FAD2-1A* Δ and *FAD2-1B* I143T alleles from population 3 failed to produce the high oleic acid phenotype when grown in the non-tropical environments. A possible explanation is the mutation in the *FAD2-1B* allele of PI 567189 A encodes at least nominal enzyme function. This explanation is supported by the fact that the I143T substitution is in a less conserved amino acid of the FAD2 enzyme than the P137R substitution. Other than that, our high oleic acid soybean lines showed a reduction of 4% at most when they were grown in the cooler environment, with a small variation in the oleic acid content. It will be necessary to test the performance of these high oleic acid soybean lines in the main North American soybean growing locations in more northern latitudes. The mutant *FAD2-1A* and *FAD2-1B* alleles will have to be combined in soybean lines with the appropriate maturity for those experiments to be conducted. However, based on the stability of the trait that we have observed so far, any reduction of oleic acid content due to the environment is likely to be minor because very little FAD2 enzyme activity remains in developing seeds in the mutant *FAD2-1A* and *FAD2-1B* lines. An additional

factor is that the end use market has not matured sufficiently to define the exact oleic acid content desired for different oil uses. Another question that should be addressed is whether the trait will affect yield or other agronomic traits. It has been reported that the transgenic soybean lines with the *FAD2-1* genes being silenced did not show any yield drag or abnormal physiology characteristics (Graef et al., 2009).

The relative contribution of FAD2-1A and FAD2-1B in oleic acid accumulation could not be fully explored in this study due to the lack of a true null allele of *FAD2-1B*. Previous research has indicated that *FAD2-1B* is expressed at a higher level than *FAD2-1A*, and that FAD2-1B is more stable than FAD2-1A when expressed recombinantly in yeast (Anai et al., 2008a; Sandhu et al., 2007; Schlueter et al., 2007; Tang et al., 2005). Our sequencing results of *FAD2-1A* and *FAD2-1B* alleles from 24 PIs with elevated oleic acid content revealed that the *FAD2-1A* gene sequence is much more conserved than *FAD2-1B* (data not shown). If the assumption is made that the P137R allele of *FAD2-1B* is non-functional, then the contribution of FAD2-1A to the FAD2 desaturase activity in developing seeds appears greater than that of FAD2-1B, although the variability in the fatty acid profiles for the lines that contain functional versions of FAD2-1A or FAD2-1B obscures the contribution from each allele. If the P137R allele of *FAD2-1B* retains some activity, then it could account for only one tenth of the original FAD2 activity present (in population 2, for the aabb genotypes containing a null *FAD2-1A* compared to the AABB genotypes).

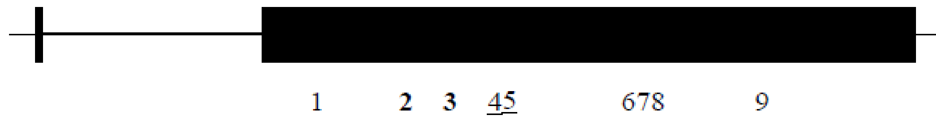
Traditional breeding has been used previously to produce soybean with up to 70% oleic acid content in the seed oil (Alt et al., 2005b). However, the phenotype is not consistent across environments, and the genetics of the trait is not very well understood,

which limits the usage of these soybean lines. Practices to boost oleic acid content in soybean to more than 80% have been achieved by means of suppression of the expression level of *FAD2* genes, and as a result, transgenic high oleic soybeans were produced (Buhr et al., 2002a). Commercial release of transgenic plants still has to overcome regulatory hurdles, and production and importation of transgenic plants remain unacceptable in various countries.

Our research results have demonstrated the capacity to develop soybeans containing more than 80% oleic acid in the oil based on very simple genetic manipulation, the combination of two recessive genes. As part of this research, we developed molecular marker assays that allow the selection of the desired mutant *FAD2-1A* and *FAD2-1B* alleles, even when they are present in the heterozygous state. Molecular marker selection thus eliminates the time necessary to produce an extra generation of plants that must be screened for the fatty acid phenotype. Simple genetics combined with perfect molecular marker assays will make it possible for soybean breeders to quickly incorporate the high oleic acid trait in their breeding programs. The resulting high oleic acid soybean lines can then be developed and released as cultivars to producers without extensive regulatory delays. Also, because only two genes control the high oleic acid phenotype, other genes may be added to enhance soybean oil quality such as low linolenic acid, low allergens, or a growing list of traits involved in soybean meal quality (Bilyeu et al., 2006). In the short term, examination of the stability of the high oleic acid soybean lines across different environments is of particular interest. Also, the soybean lines with the high oleic acid trait should be used for development of soybean varieties with favorable agronomic traits including high yield.

In conclusion, this research demonstrates that when mutant alleles of *FAD2-1A* and *FAD2-1B* are combined together by means of traditional plant breeding, they can significantly enhance the oleic acid content of the oil, up to 80%, providing a means for the development of soybean varieties with superior oil quality.

FIGURES



B.

PI 283327	61	VVYDLSLAFIFYIATTYFHLLPHPFFLIAWPIYWVLQGCILTGVWVIAHECGHHAFSKYP
PI 567189 A	61	VVYDLSLAFIFYIATTYFHLLPHPFFLIAWPIYWVLQGCILTGVWVIAHECGHHAFSKYP
Williams 82	61	VVYDLSLAFIFYIATTYFHLLPHPFSLIAWPIYWVLQGCILTGVWVIAHECGHHAFSKYP
PI 283327	121	WVDDVVGLTVHSALLVRYFSWKISHRRHHSNTGSLDRDEVFVPKPKSKVAWYTKYLNNPL
PI 567189 A	121	WVDDVVGLTVHSALLVPYFSWKTSHRRHHSNTGSLDRDEVFVPKPKSKVAWYTKYLNNPL
Williams 82	121	WVDDVMGLTVHSALLVPYFSWKISHRRHHSNTGSLDRDEVFVPKPKSKVAWYTKYLNNPL

C.

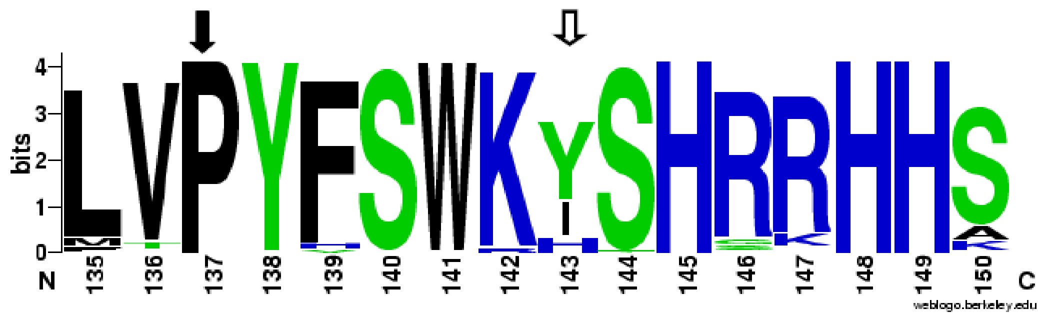


Figure 2.1: Characterization of mutations in the *FAD2-1B* alleles from soybean lines PI 283327 and PI 567189A. A. Schematic of the *FAD2-1B* gene and the polymorphisms identified in the alleles from PI 283327 and PI 567189A. The horizontal line represents the DNA for *FAD2*-1B, with the heavier line indicating the intron and lighter lines representing the 5' and 3' untranslated portions of the gene. The dark vertical line represents the portion of exon 1 that contains only the start codon and the darkened rectangle represents exon 2. Numbers beneath the schematic indicate the positions of polymorphisms compared to the Williams 82 reference *FAD2-1B* coding sequence, with shared missense mutations in bold and non-shared missense mutations underlined. 1 = a105g, silent; 2 = c257t, S86F; 3 = a376g, M126V; 4 = c410g, P137R unique to PI 283327 and PI 210179; 5 = t428c, I143T unique to PI 567189A and PI 578451; 6 = c657t, silent; 7 = t669c, silent; 8 = t682c, silent; 9 = a918g, silent. **B.** Fatty acid desaturase *FAD2*-1B amino acid sequence alignment in the region surrounding the mutations in PI 283327 and PI 567189 A. Amino acid positions are indicated at the beginning of each line of the alignment. Identical amino acid residues are highlighted in black, a similar amino acid substitution is highlighted in gray, and the S86F, P137R, and I143T mutations are not highlighted. C. Weblogo output of the amino acid conservation

FAD2 enzyme as part of the BLINK feature at NCBI using GI number 197111722. The top 100 best matched sequences were aligned and used as input for sequence LOGO http://weblogo.berkeley.edu/logo.cgi webcite. The logo consists of stacks of symbols, one stack for each position in the amino acid sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino acid at that position. Closed arrow indicates residue changed due to the P137R *FAD2-1B* mutation in PI 283327 and open arrow indicates residue changed due to the I143T *FAD2-1B* mutation in PI 567189A.

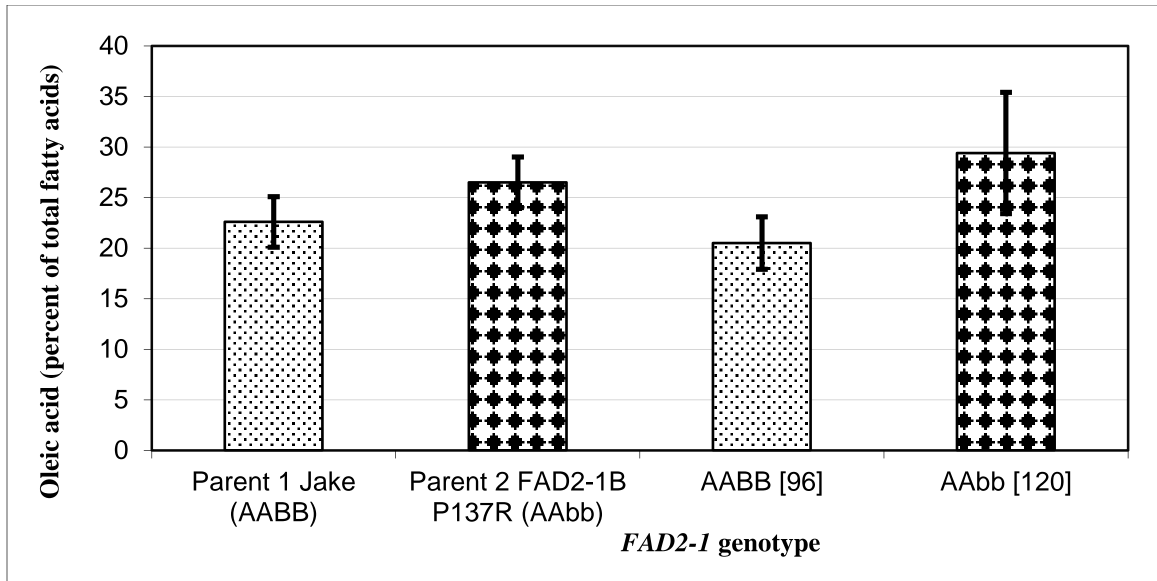


Figure 2.2: Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 1. Average oleic acid content of soybean seeds produced in Portageville, MO from the F_6 RILs developed from the cross Jake x PI 283327, population 1. Labels on the X-axis include: two parents with contrasting *FAD2-1B* genotypes and the RILs grouped by *FAD2-1B* genotype. BB indictates wild-type *FAD2-1B* alleles, and bb indicates mutant P137R *FAD2-1B* alleles derived from PI 283327; brackets surround the number of soybean seed samples represented in the genotype class. Error bars indicate plus and minus one standard deviation from the mean. Oleic acid phenotype data of each genotype class is the mean of oleic acid content as a percentage of the total fatty acid content of the oil of all lines carrying the *FAD2-1* genotype, four individual samples for each line.

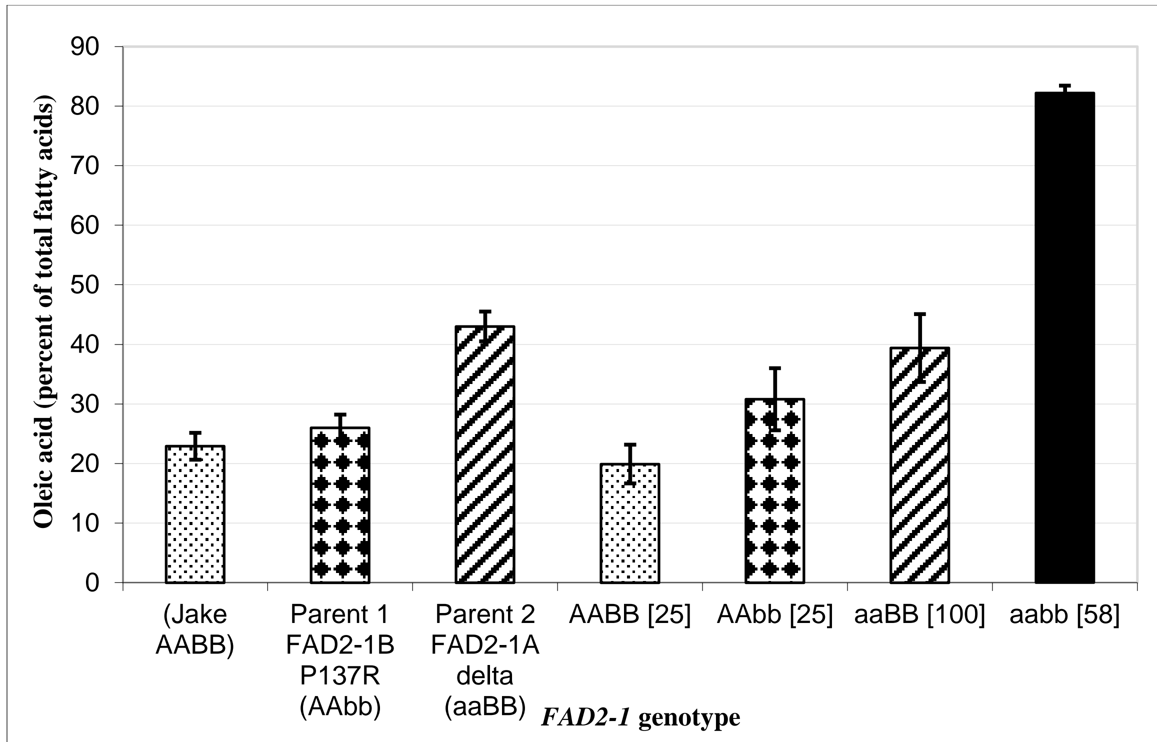


Figure 2.3: Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 2. Average oleic acid content of soybean seeds produced in Portageville, MO from the segregating F_6 and F_7 RILs developed from the cross M23 x PI 283327, population 2. X-axis labels indicate a typical commodity soybean line (Jake), the two RIL parents, and the RILs grouped by *FAD2-1A* and *FAD2-1B* genotype. AA = wild-type *FAD2-1A* alleles, aa = mutant *FAD2-1AΔ* alleles derived from M23, BB = wild-type *FAD2-1B* alleles, bb = mutant P137R *FAD2-1B* alleles derived from PI 283327; brackets surround the number of soybean seed samples represented in the genotype class. Error bars indicate plus and minus one standard deviation from the mean. Oleic acid phenotype data of each genotype class is the mean of oleic acid content as a percentage of the total fatty acid content of the oil of all lines carrying the *FAD2-1* genotype, five individual samples for each line, with two exceptions, one for class aaBB, where six lines had ten seed samples and eight lines had five seed samples. Additionally, two separate individual seed samples from the aabb lines were ommitted from the analysis because their genotypes could not be verified. Including these two samples would have led to a mean of 80.8 ± 8.0% oleic acid for the *FAD2-1* aabb genotype class. Subsequent sampling of ten additional seeds from these lines produced oleic acid levels with a range of 81.5% to 86.2 % oleic acid in the seed oil.

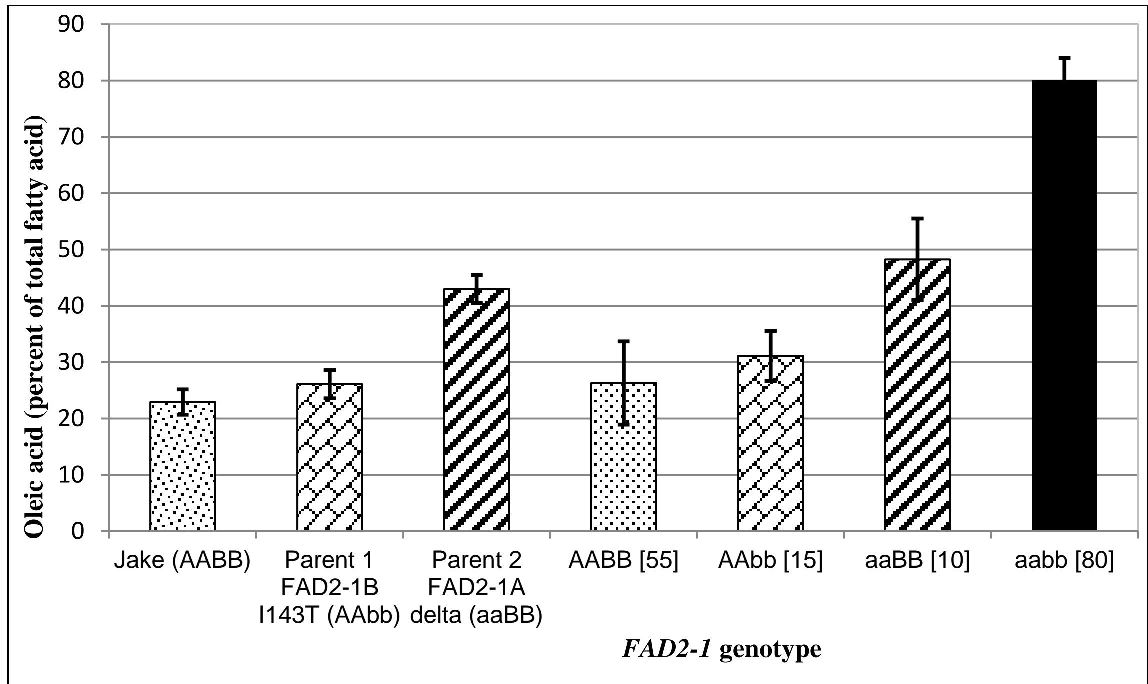


Figure 2.4: Seed oleic acid phenotype and *FAD2-1* genotype association analysis for population 3. Average oleic acid content of soybean seeds produced in Costa Rica from the segregating F_6 and F_7 RIL developed from the cross M23 x PI 567189 A population 3. X-axis labels indicate a typical commodity soybean line (Jake), the two RIL parents, and the RILs grouped by *FAD2-1A* and *FAD2-1B* genotype. AA = wild-type *FAD2-1A* alleles, aa = mutant *FAD2-1A*Δ alleles derived from M23, BB = wild-type *FAD2-1B* alleles, bb = mutant I143T *FAD2-1B* alleles derived from PI 567189 A; brackets surround the number of soybean seed samples represented in the genotype class. Error bars indicate plus and minus one standard deviation from the mean. Oleic acid phenotype data of each genotype class is the mean of oleic acid content as a percentage of the total fatty acid content of the oil of all lines carrying the *FAD2-1* genotype, five individual samples for each line, with the exception of the aaBB class, for which there was only one line and ten individual seeds were analyzed.

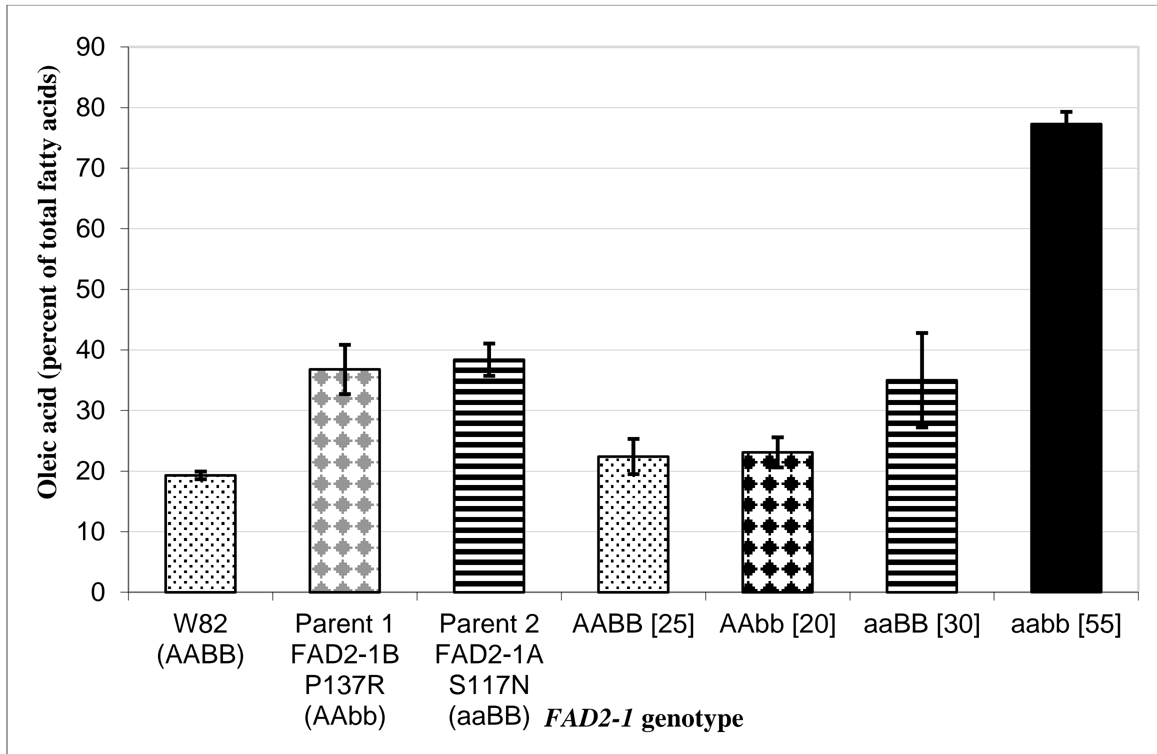


Figure 2.5: Seed oleic acid phenotype and *FAD2-1* genotype association analysis for F_3 seeds of population 4. Average oleic acid content of soybean seeds produced in Columbia, MO from the segregating F_2 population developed from the cross 17D x S08-14788, population 4. X-axis labels indicate a typical commodity soybean line (Williams 82, W82), the two RIL parents, and the RILs grouped by *FAD2-1A* and *FAD2-1B* genotype. AA = wild-type *FAD2-1A* alleles, aa = mutant S117N *FAD2-1A* alleles derived from 17D, BB = wild-type *FAD2-1B* alleles, bb = mutant P137R *FAD2-1B* alleles derived from PI 283327; brackets surround the number of soybean seed samples represented in the genotype class. Error bars indicate plus and minus one standard deviation from the mean. Oleic acid phenotype data of each genotype class is the mean of oleic acid content as a percentage of the total fatty acid content of the oil of all lines carrying the *FAD2-1* genotype, five individual samples for each line.

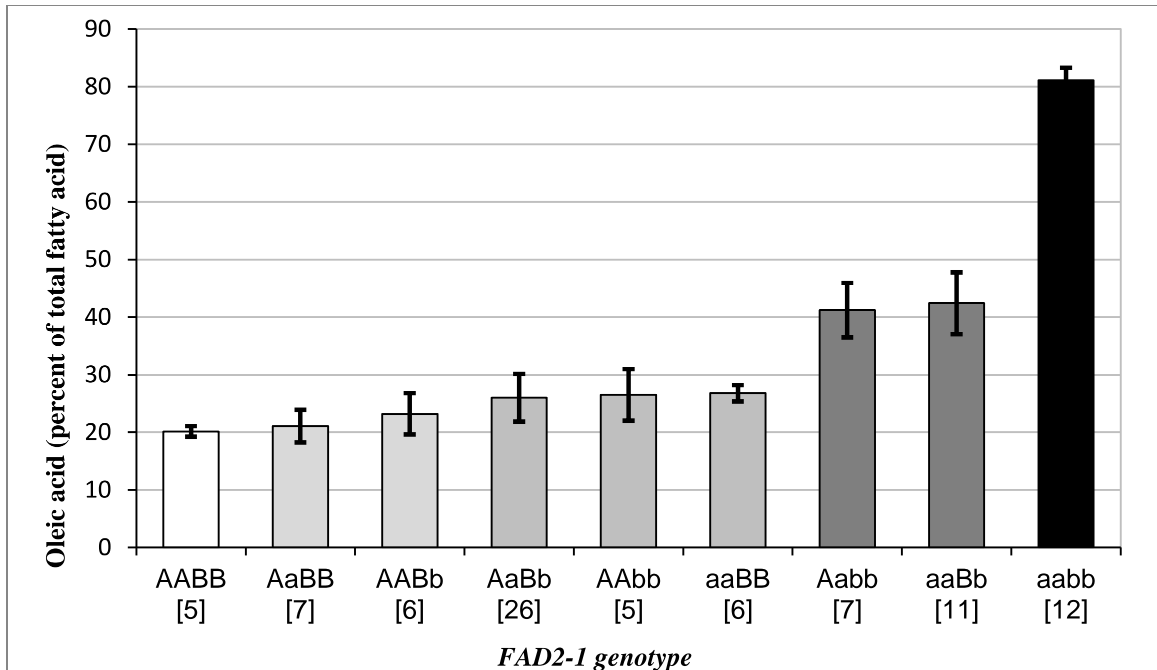


Figure 2.6: Seed oleic acid phenotype and *FAD2-1* genotype association analysis for F_2 seeds of population 4. Average oleic acid content of F_2 soybean seeds produced in Costa Rica from the segregating population developed from the cross 17D x S08-14788, population 4. X-axis labels indicate a the soybeans grouped by *FAD2-1A* and *FAD2-1B* genotype. AA = wild-type *FAD2-1A* alleles, aa = mutant S117N *FAD2-1A* alleles derived from 17D, BB = wild-type *FAD2-1B* alleles, bb = mutant P137R *FAD2-1B* alleles derived from PI 283327; brackets surround the number of soybean seed samples represented in the genotype class. Error bars indicate plus and minus one standard deviation from the mean. Oleic acid phenotype data of each genotype class is the mean of oleic acid content as a percentage of the total fatty acid content of the oil of all lines carrying the *FAD2-1* genotype.

TABLES

Table 2.1: Oleic acid content, standard deviation, and seed generation of soybean lines with different combinations of mutant *FAD2*-1A and mutant *FAD2*-1B produced in three environments.

Population	FAD2-1A	FAD2-1B	Oleic acid content (percent of total fatty acid)		
			Costa Rica[1]	Portageville, MO[2]	Columbia, MO[3]
2	Δ	P137R	81.4 ± 5.7 ^{F5}	82.2 ± 1.2 ^{F7}	79.1 ± 1.3 ^{F8}
3	Δ	I143T	80.0 ± 4.0 ^{F5}	65.0 ± 4.3 ^{F7}	58.7 ± 7.7 ^{F8}
4	S117N	P137R	81.1 ± 2.2 ^{F2}	81.7 ± 2.1 ^{F3}	77.3 ± 2.0 ^{F3}

[1]Research station in Costa Rica. Seeds of F_5 generation of population of 2 and 3 were produced in winter 2006-2007, while F_2 seeds of population 4 were produced in winter 2008-2009.

[2] Plants were grown in Delta Research Center, seeds of F_7 generation of the populations 2 and 3 were produced in summer 2008 and F_3 generation of population 4 was produced in summer 2009.

[3] All of the plants were grown summer 2009 at the Bradford Research & Extension Center, Columbia MO

Table 2.2: Fatty acid profiles (means and standard deviations) for different homozygous *FAD2*-1 genotypes in four segregating populations developed by crosing soybean lines carrying different sources of mutant *FAD2*-1A alleles with different sources of mutant *FAD2*-1B alleles.

	Fatty Acid				
	16:0	**18:0**	**18:1**	**18:2**	**18:3**
Population 1 (Jake[1] x PI 283327)					
BB[2] (n=24)	12.2 ± 0.9	3.9 ± 0.5	20.5 ± 2.6	53.4 ± 2.8	10.0 ± 0.3
bb (n=30)	11.2 ± 0.7	3.8 ± 0.6	29.4 ± 6.0	47.0 ± 5.1	8.7 ± 0.5
Population 2 (M23 x PI283327)					
AABB (n=5)	12.3 ± 0.5	3.7 ± 0.4	19.9 ± 3.3	55.4 ± 2.7	8.7 ± 1.0
AAbb (n=5)	11.0 ± 0.5	3.9 ± 0.4	30.8 ± 5.2	45.9 ± 4.6	8.5 ± 0.9
aaBB (n=14)	10.8 ± 0.8	3.8 ± 0.6	39.4 ± 5.7	37.1 ± 4.8	8.9 ± 1.2
aabb (n=16)	7.9 ± 0.7	3.7 ± 0.6	82.2 ± 1.2	2.3 ± 0.6	3.9 ± 0.5
Population 3 (M23 x PI 567189A)					
AABB (n=11)	12.5 ± 0.9	2.9 ± 0.4	26.3 ± 7.4	51.4 ± 6.4	6.1 ± 1.2
AAbb (n=3)	12.4 ± 0.8	2.8 ± 0.4	31.1 ± 4.5	47.5 ± 3.3	6.1 ± 1.0
aaBB (n=1)	10.3 ± 0.6	2.8 ± 0.3	48.2 ± 7.2	32.5 ± 6.1	6.2 ± 0.9
aabb (n=16)	8.4 ± 0.8	2.6 ± 0.4	80.0 ± 4.0	5.0 ± 3.0	3.8 ± 0.6
Population 4 F_2(17D x S08-14788)					
AABB (n=5)	12.3 ± 0.9	3.2 ± 0.3	20.1 ± 0.9	55.7 ± 1.0	8.7 ± 0.6
AAbb (n=5)	12.1 ± 1.0	3.4 ± 0.5	26.5 ± 4.5	47.8 ± 3.7	10.2 ± 0.9
aaBB (n=6)	11.7 ± 0.3	3.0 ± 0.2	26.8 ± 1.4	48.2 ± 0.7	9.9 ± 0.5
aabb (n=12)	7.8 ± 0.5	3.1 ± 0.2	81.1 ± 2.2	3.2 ± 1.4	4.9 ± 0.6
Population 4 $F_{2:3}$ (17D x S08-14788)					
AABB (n=5)	9.6 ± 0.6	3.9 ± 0.4	22.4 ± 2.9	56.0 ± 2.8	8.2 ± 0.9
AAbb (n=4)	10.5 ± 0.5	3.8 ± 0.3	23.1 ± 2.5	54.0 ± 2.6	8.6 ± 0.5
aaBB (n=6)	9.3 ± 0.6	3.2 ± 0.3	35.0 ± 7.8	42.9 ± 5.9	9.6 ± 2.2
aabb (n=11)	6.9 ± 0.4	3.2 ± 0.2	77.3 ± 2.0	6.3 ± 1.5	6.3 ± 0.6

[1] Jake is a soybean line with normal oleic acid content and wild-type FAD2-1A; M23 and 17D are soybean lines with mutant *FAD2-1A*Δ and *FAD2-1A* S117N alleles, respectively. PI 283327 and PI 567189A are soybean lines with mutant *FAD2-1B* P137R and *FAD2-1B* I143T alleles, respectively. S08-14788 is a soybean line selected from population 2 that inherited the mutant *FAD2-1B* P137R alleles from PI 283327.

[2] AA = wild-type *FAD2-1A* alleles, aa = mutant *FAD2-1A* alleles derived from M23 or17D, BB = wild-type *FAD2-1B* alleles, bb = mutant *FAD2-1B* alleles derived from PI 283327 or PI 567189A.

LITERATURE CITED

Alt, J., W. Fehr, G. Welke, and D. Sandhu. 2005a. Phenotypic and molecular analysis of oleate content in the mutant soybean line M23. Crop Science 45:1997 - 2000.

Alt, J.L., W.R. Fehr, G.A. Welke, and J.G. Shannon. 2005b. Transgressive segregation for oleate content in three soybean populations. Crop Science 45:2005-2007.

Anai, T., T. Yamada, R. Hideshima, T. Kinoshita, S.M. Rahman, and Y. Takagi. 2008. Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by Gm*FAD2*-1a. Breeding Sciences 58:447-452.

Ascherio, A., and W.C. Willett. 1997. Health effects of trans fatty acids. American Journal of Clinical Nutrition 66 (Suppl):1006-1010.

Bachlava, E., R.E. Dewey, J.W. Burton, and A.J. Cardinal. 2009. Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. Crop Sci 49:433-442.

Beló, A., P. Zheng, S. Luck, B. Shen, D.J. Meyer, B. Li, S. Tingey, and A. Rafalski. 2008. Whole genome scan detects an allelic variant of *FAD2* associated with increased oleic acid levels in maize Molecular Genetics and Genomics 279:1-10.

Bernard, R., and C. Cremeens. 1988. Registration of 'Williams 82' Soybean. Crop Science 28:1027 - 1028.

Beuselinck, P.R., D.A. Sleper, and K.D. Bilyeu. 2006. An assessment of phenotype selection for linolenic acid using genetic markers. Crop Science 46:747-750.

Bilyeu, K., L. Palavalli, D. Sleper, and P. Beuselinck. 2005. Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science 45:1830-1836.

Bilyeu, K., L. Palavalli, D.A. Sleper, and P. Beuselinck. 2006. Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science 46:1913-1918.

Bruner, A.C., S. Jung, A.G. Abbott, and G.L. Powell. 2001. The naturally occurring high oleate oil character in some peanut varieties results from reduced oleoyl-PC desaturase activity from mutation of Aspartate 150 to Asparagine. Crop Science 41:522-526.

Buhr, T., S. Sato, F. Ebrahim, A. Xing, Y. Zhou, M. Mathiesen, B. Schweiger, A. Kinney, P. Staswick, and T. Clemente. 2002b. Ribozyme termination of RNA

transcripts down-regulate seed fatty acid genes in transgenic soybean. Plant Journal 30:155-163.

Burton, J.W., R.F. Wilson, G.J. Rebetzke, and V.R. Pantalone. 2006. Registration of N98-4445A mid-oleic soybean germplasm line. Crop Science 46:1010-1012.

Butzen, S., and S. Schnebly. 2007. High oleic soybean. Crop Insights 17:3.

Crooks, G., G. Hon, J.-M. Chandonia, and S. Brenner. 2004. WebLogo: A sequence logo generator. Genome Research 14:1188 - 1190.

Dierking, E., and K. Bilyeu. 2009. New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biology 9:89.

Graef, G., B. LaVallee, P. Tenopir, M. Tat, B. Schweiger, A. Kinney, J. Gerpen, and T. Clemente. 2009. A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. Plant Biotechnology Journal 7:411 - 421.

Hartwig, R.A., and C.R. Hurburgh. 1990. Near-infrared reflectance measurement of moisture, protein and oil content of ground crambe seed. Journal of the American Oil Chemists' Society 67:435-437.

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

Hu, X., M. Sullivan-Gilbert, M. Gupta, and S.A. Thompson. 2006. Mapping of the loci controlling oleic and linolenic acid contents and development of *FAD2* and *FAD3* allele-specific markers in canola (*Brassica napus* L.). Theoretical Applied Genetics 113:497-507.

Jung, S., G. Powell, K. Moore, and A. Abbott. 2000. The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. II. Molecular basis and genetics of the trait. Molecular and General Genetics 263:806-811.

Knowles, P.F., and A.B. Hill. 1964. Inheritance of fatty acid content in the seed oil of a safflower introduction from Iran. Crop Science 4:406-409.

Lee, J.-D., K.D. Bilyeu, and J.G. Shannon. 2007. Genetics and breeding for modified fatty acid profile in soybean seed oil. Journal of Crop Science and Biotechnology 10:201-210.

Lee, J.D., M. Woolard, D.A. Sleper, J.R. Smith, V.R. Pantalone, C.N. Nyinyi, A. Cardinal, and J.G. Shannon. 2009. Environmental effects on oleic acid in soybean seed oil of Plant Introductions with elevated oleic concentration. Crop Science 49:1762-1768.

Liu, Q., S. Singh, and A. Green. 2002. High-oleic and high-stearic cottonseed oils: nutritionally improved cooking oils developed using gene silencing. Journal of the American College of Nutrition 21:205S-211S.

Monteros, M.J., J.W. Burton, and H.R. Boerma. 2008. Molecular mapping and confirmation of QTLs associated with oleic acid content in N00-3350 soybean. Crop Science 48:2223-2234.

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

Oliva, M.L., J.G. Shannon, D.A. Sleper, M.R. Ellersieck, A.J. Cardinal, R.L. Paris, and J.D. Lee. 2006. Stability of fatty acid profile in soybean genotypes with modified seed oil composition. Crop Science 46:2069-2075.

Patel, M., S. Jung, K. Moore, G. Powell, C. Ainsworth, and A. Abbott. 2004. High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theoretical Applied Genetics 108:1492-1502.

Ramensky, V., P. Bork, and S. Sunyaev. 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Research 30:3894-3900.

Raneses, A.R., L.K. Glaser, J.M. Price, and J.A. Duffield. 1999. Potential biodiesel markets and their economic effects on the agricultural sector of the United States. Industrial Crops and Products 9:151-162.

Sandhu, D., J. Alt, C. Scherder, W. Fehr, and M. Bhattacharyya. 2007. Enhanced oleic acid content in the soybean mutant M23 is associated with the deletion in the *FAD2*-1a gene encoding a fatty acid desaturase. Journal of the American Oil Chemists' Society 84:229-235.

Scherder, C.W., and W.R. Fehr. 2008. Agronomic and seed characteristics of soybean lines with increased oleate content. Crop Science 48:1755-1758.

Schlueter, J.A., I.F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. Siegfried, B.A. Roe, S.D. Schlueter, B.E. Scheffler, and R.C. Shoemaker. 2007. The *FAD2* gene family of soybean:Insights into the structural and functional divergence of a paleopolyploid genome. Crop Science 47:S-14 - 26.

Shannon, J.G., J.A. Wrather, D.A. Sleper, R.T. Robinson, H.T. Nguyen, and S.C. Anand. 2007. Registration of 'Jake' Soybean. Journal of Plant Registration 1:29-30.

Stoutjesdijk, P.A., C. Hurlestone, S.P. Singh, and A.G. Green. 2000. High-oleic acid Australian *Brassica napus* and *B. juncea* varieties produced by co-suppression of endogenous D12-desaturases. Biochemical Society Transactions 28:938-940.

Takagi, Y., and S.M. Rahman. 1996. Inheritance of high oleic acid content in the seed oil of soybean mutant M23. Theoretical Applied Genetics 92:179-182.

Tang, G.Q., W.P. Novitzky, H.C. Griffin, S.C. Huber, and R.E. Dewey. 2005. Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. Plant Journal 44:433-446.

CHAPTER 3

A novel *FAD2-1A* allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content

SUMMARY

The alteration of fatty acid profiles in soybean to improve soybean oil quality has been a long-time goal of soybean researchers. Soybean oil with elevated oleic acid is desirable because this monounsaturated fatty acid improves the nutrition and oxidative stability of soybean oil compared to other oils. In the lipid biosynthetic pathway, the enzyme fatty acid desaturase 2 (FAD2) is responsible for the conversion of oleic acid precursors to linoleic acid precursors in developing soybean seeds. Two genes encoding *FAD2-1A* and *FAD2-1B* were identified to be expressed specifically in seeds during embryogenesis and considered to hold an important role in controlling the seed oleic acid content. Twenty two soybean plant introduction (PI) lines identified to have an elevated oleic acid content were characterized for sequence mutations in the *FAD 2-1A* and *FAD2-1B* genes. PI 603452 was found to contain a deletion of a nucleotide in the second exon of *FAD2-1A*. These important SNPs were used in developing molecular marker genotyping assays, and the assays appear to be a reliable and accurate tool to identify the *FAD 2-1A* and *FAD2-1B* genotype of wild type and mutant plants. PI 603452 was subsequently crossed with PI 283327, a soybean line that has a mutation in *FAD2-1B*. Interestingly, soybean lines carrying both homozygous insertion/deletion mutation (indel) *FAD2-1A* alleles and mutant *FAD2-1B* alleles have an average of 82 to 86% oleic acid content, compared to 20% in conventional soybean, and low levels of linoleic and linolenic acids. The newly identified indel mutation in the *FAD2-1A* gene offers a simple method for the development of high oleic acid commercial soybean varieties.

INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) oil is one of the most economically important products of soybean. Due to its neutral flavor and a competitive price, soybean oil is used extensively in food industry and has been the most consumed vegetable oil in the world (http://www.soystats.com/). It is reported to account for more than 70% of the total edible fat and oil consumption in the U.S. from 2001-2009, and three quarters of this amount was used as cooking oil, margarine, and baking and frying fat (http://www.soystats.com/). Therefore, soybean oil is almost ubiquitously present in daily diets as a hidden ingredient in processed foods, snacks, and fast foods or directly consumed as vegetable oil. As a result, enhancement of soybean oil quality is desirable.

Soybean oil is composed mostly of triacylglycerol, an ester product of fatty acids and glycerol. Therefore, the chemical properties of soybean oil ultimately depend on the seed fatty acid composition (Ensminger and Ensminger, 1993). Commodity soybean oil typically contains 11% palmitic acid (16:0), 4% stearic acid (18:0), 25% oleic acid (18:1), 52% linoleic acid (18:2), and 8% linolenic acid (18:3) (Fehr, 2007). This high percentage of linoleic acid (ω-6) and linolenic acid (ω-3) in soybean oil is nutritionally positive but disadvantageous for food production due to the fact that the oil is oxidized easily and the foods go rancid quickly (Mounts et al., 1988; Warner and Gupta, 2005). Thus, to improve the quality of soybean oil and processed foods, chemical hydrogenation has been employed to increase oleic acid content to nearly 50% and reduce the amount of the polyunsaturated fatty acids (List et al., 2000). Oil hydrogenation not only adds to the cost of oil processing, but it also generates *trans* fat that has been linked to heart disease and stroke, high levels of "bad" cholesterol (low-density lipoproteins), a higher risk of

developing type 2 diabetes and others (Hu et al., 1997; Mozaffarian et al., 2006). Another effective approach to improve soybean oil functionality without hydrogenation is to genetically increase the oleic acid content in soybean seeds at the expense of linoleic and linolenic acids. As a result, a healthier and less expensive soybean oil would be available to improve consumers‟ health without making drastic changes in their diets.

In the lipid biosynthetic pathway, the delta-twelve fatty acid desaturase 2 enzyme (FAD2) is responsible for the conversion of oleic acid (18:1) precursors to linoleic acid (18:2) precursors (Okuley et al., 1994; Schlueter et al., 2007a). In developing soybean seeds, among FAD2 genes identified, the two microsomal FAD2-1 desaturases *FAD2-1A* (Glyma10g42470) and *FAD2-1B* (Glyma20g24530) have the highest expression levels (Schlueter et al., 2007b; Tang et al., 2005). Thus, *FAD2-1A* and *FAD2-1B* are considered to play an important role in controlling the oleic acid level in developing soybean seeds and were used as targets or candidate genes to produce high oleic acid soybean. Two approaches were used to generate soybeans with oleic acid content over 80% of the total oil content including genetic engineering and candidate gene-based molecular breeding (Buhr et al., 2002; Hoshino et al., 2010; Pham et al., 2010). Transgenic high oleic acid soybeans were produced using ribozyme-terminated antisense to down regulate *FAD 2-1* gene transcript levels (Buhr et al., 2002). Though the transgenic high oleic acid soybeans had the desirable fatty acid profile, production and importation of these transgenic soybeans may still face opposition in some important markets like Europe and Asia. On the other hand, we used a candidate gene approach to identify a mutant *FAD2-1B* gene in PI 283327 and combine this gene with existing mutant *FAD2-1A* genes from two soybean lines M23 and 17D to produce a stable, non-transgenic high oleic acid (> 80%) content in

soybean seed oil (Pham et al., 2010). The usefulness of this approach to produce high oleic soybeans was confirmed by another research group although different sources of mutant *FAD2-1A* (KK21) and *FAD2-1B* genes (from two EMS mutant soybean lines) were used (Hoshino et al., 2010).

It has been shown the more severe the mutations in either *FAD2-1A* or *FAD2-1B* are, the higher and more stable the oleic acid content is in the soybean oil, due to residual enzymatic activity of FAD2. For instance, high oleic acid soybeans with the 17D derived mutant *FAD2-1A* alleles containing a missense mutation showed a larger reduction of oleic acid content (more flux through the desaturase pathway) when grown in cooler environments, compared to those that had the M23-derived *FAD2-1A* null alleles, which are completely deleted (Anai et al., 2008; Pham et al., 2010). However, M23-derived lines have been shown to have reductions in yield that appear to be linked to the deleted portion of chromosome 10 (Scherder and Fehr, 2008). Therefore, identifying natural mutations involving single base pair deletions/mutations of these two candidate genes in soybean lines in the public germplasm collection and combining them by marker assisted breeding appears to be a promising strategy to produce high and more stable oleic acid content with less effect on yield. Fortunately, a set of soybean plant introductions (PIs) in the U.S. soybean germplasm collection was identified to have elevated oleic acid contents ranging from 30-50% of the total fatty acid in seed oil. Fifteen soybean lines with the highest oleic acid content among those were evaluated for the stability of the trait in multi-locations field trials (Lee et al. 2009). These PIs are promising sources of novel mutant *FAD2-1A* and *FAD2-1B* alleles. Therefore, we initiated a project to: 1) screen for mutations in candidate *FAD2-1* genes that may be responsible for the elevated

oleic acid content in these PIs; and 2) produce high oleic soybean lines using the identified mutant alleles of *FAD2-1A* and *FAD2-1B*.

MATERIALS AND METHOD

DNA isolation and sequencing of FAD2-1A and FAD2-1B

Primers specific for *FAD2-1A* and *B* soybean genes and PCR amplifications were described by Pham et al. 2010. PCR products were obtained and validated using agarose gel- electrophoresis to confirm that the products"authenticity. Sequencing reactions were performed at the DNA Core Center, University of Missouri. Sequence alignment and analysis were accomplished using Multiple Sequence Alignment by CLUSTALW (http://align.genome.jp/). Variant nucleotides between „Williams 82" reference (http://www.phytozome.net/soybean) and the PIs were identified. Protein translation was conducted using ExPaSy (http://ca.expasy.org/tools/dna.html) and protein alignment was done using Multiple Sequence Alignment.

Population development

PI 603452 was crossed with two different lines that contain the P137R *FAD2-1B* allele derived from PI 283327. In cross 1, generating population KB09-13, PI 603452 was crossed to a soybean (Jake x PI283327) -derived line that contained the mutant *FAD2-1B* allele P137R and for cross 2, generating population KB09-35, PI 603452 was crossed to a (17D x PI283327)-derived line that contained the same mutant allele of *FAD2-1B* and wild-type alleles of *FAD2-1A*. True F_1 seeds were confirmed using a molecular marker assay of *FAD2-1B* from PI 283327 (Pham et. al. 2010)

F_1 and F_2 seeds of the KB09-13 population were grown in Costa Rica in Fall 2009 and Spring 2010, and F_3 seeds were sown in field conditions in Columbia MO and Portageville MO in summer 2010. F_1 KB09-35 seed was planted in a growth chamber in

Spring 2010 and F_2 seedlings were transplanted to the field at the Bradford Research and Extension Center, Columbia MO in Summer 2010.

***FAD2-1A* allele specific molecular marker assay**

SimpleProbe assay for the deletion in the *FAD2-1A* of PI 603452 was developed based on the SimpleProbe protocol described by Pham et al. 2010. The Probe contained 5‟-Fluorescence-SPC-CCTCTAGG**A**AGGGCTGTTTCTCT-Phosphate-3‟ (the deleted nucleotide is indicated by bold font and underline). Primers used to generate template for Simpleprobe genotyping assay were designed by aligning the *FAD2-1A* and *FAD2-1B* region containing the SNPs. Primers were selected to be as close as possible to the SNPs while differing in at least 3 nucleotides between the two genes to specifically amplify the targeted region in *FAD2-1A*. Genotyping reactions were performed with a 5:2 asymmetric mix of primers (5‟-CCAAGGTTGCCTTCTCACTGGT-3‟ at 2 µM final concentration, and 5‟-TAGGCCACCCTATTGTGAGTGTGAC-3‟ at 5 µM final concentration). Reactions were carried out in 20 µl; containing template, primers, 0.2 µM final concentration of SimpleProbe, buffer (40 mM Tricine-KOH [pH 8.0] 16 mM $MgCl_2$, 3.75 µg ml^{-1} BSA,), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Genotyping reactions were performed using a Lightcycler 480 II real time PCR instrument (Roche), using the following PCR parameters: 95 °C for 5 minutes followed by 40 cycles of 95 °C for 20 seconds, 65 °C for 20 seconds, 72 °C for 30 seconds, and then a melting curve from 50 °C to 68 °C. PI 603452 and all soybean lines with an identical *FAD2-1A* allele genotype have a

characteristic peak at 56 °C, while Williams 82 (wild-type *FAD2-1A*) have a peak at 62 °C. Heterozygous individuals's genotype showed two peaks at 56 °C and 62 °C.

Fatty acid, protein and oil determination

Fatty acid profiles of all samples were obtained using the method of gas chromatography of total fatty acid methyl esters of extracted oil (Beuselinck et al., 2006). The individual fatty acid contents are reported as the relative percents of palmitic, stearic, oleic, linoleic, and linolenic acids in the extracted oil. For population KB09-13, five whole crushed individual seeds were used as samples to determine the fatty acid content of each soybean lines.

For population KB09-35, seed chips were used by removing a “chipped” portion of the seed opposite the embryo with a scalpel for fatty acid analysis such that the remainder of the seed containing the embryo could be germinated.

Protein and oil contents were determined for seeds of high oleic lines KB09-13-744, KB09-13-760, KB09-35-3-5, Williams 82 and a soybean line with genotype wild-type for both *FAD2-1A* and *FAD2-1B* genes using NIR spectroscopy. Fifty seeds from each line were used for the analysis (Hartwig and Hurburgh, 1990).

Population genotyping using SimpleProbe assay

Populations were genotyped using SimpleProbe assays designed specifically for each mutant allele (Pham et al., 2010). For the KB09-13 population, DNAs of F_1 and F_2 plants were collected for PCR reactions using Whatman FTA card protocol BD05 (whatman.com). For the KB09-35 population, seeds were chipped to get a small portion

for fatty acid determination. The remaining portion with hypocotyl was germinated in germination packages to collect DNA on Whatman FTA cards.

Design of field experiments

Seeds of KB09-13-744 and KB09-13-760 were grown in Columbia and Portageville in summer 2010 together with two parents (Jake x PI283327)/ (17D x PI 283327), PI 603452 (maturity group (MG) III) and various checks including Williams 82 (MG III), LG04-6863 (MG mid group IV), 5002T (MG IV late to V early), N98-4445A (MG IV), M23 (MG V), AP09INCR1, AP09INCR2 (80% oleic lines from M23 x PI 283327) . There were three replications for each location, and Randomized Complete Block Design was used for setting up the experiment. In both locations, plantings were made in rows spaced 76cm apart. Within the rows, ten seed of each soybean line was planted in a hill plot, spaced 51 cm apart. After seedling emergence, hills were thinned to five plants from which all data was collected for each soybean line. In Columbia, seeds were planted on May 28th and harvested on September 21st for KB09-13-744 and September 23rd for KB09-760, which were similar to Williams 82 with MG group III. In Portageville, seeds were planted June 2nd and harvested September 24th for KB09-13-744 and Sep 21st for KB09-760, which were also similar to Williams 82. Seeds from five plants were bulked and five seeds from each line were used for fatty acid analysis.
For population KB09-35, the germinated chipped seeds of soybean lines KB09-35-3-5 and 2 soybean lines for each of the genotype aaBB, AAbb and AABB were transplanted into the field. These seedlings were grown 10 cm apart from each other in a hill plot right

behind the population KB09-13. All the plants were transplanted on June 9th and KB09-35-3-5 was harvested on September 28th, which was considered to be in MG III.

Statistical analysis

The data and statistical analysis presented for population KB09-13 was generated using proc mixed procedure, SAS® 9.2 Enhanced Logging Facilities, Cary, NC: SAS Institute Inc., 2008 (Table 3.3). The data presented for population KB09-35 was generated using fatty acid values of all seeds used for phenotypic analysis (Table 3.4). Comparisons between two soybean lines presented in this table were done using Student‟s t test with $\alpha = 0.05$.

RESULTS

Identification of novel alleles of *FAD2-1A* and *FAD2-1B* in soybean plant introductions

A set of 22 soybean lines were selected for the sequencing of *FAD2-1A* and *FAD2-1B* genes (Table 3.1). These PIs have mean oleic acid contents in the range of 27- 49% of the total fatty acid content in seed oil. Lines with identifier number 1-15 were evaluated in three years, 2005-2007, in Portageville, MO (J. G. Shannon, unpublished data) while lines 16-22 were evaluated in 16 environments during 2005-2007 (Lee et al., 2009).

<u>*FAD2-1A*.</u>

Among 22 PIs, 17 lines have a wild-type *FAD2-1A* identical to the Williams 82 reference sequence. Four soybean lines (KLG11028, PI 404160B, PI 506885, PI 507420) contained a missense mutation (g64c in coding sequence) resulting in a change from G22R in protein sequence and a silent mutation of t990c. To predict the potential effect of the amino acid changes to soybean FAD2-1A enzyme function, the program PolyPhen was used to analyze the potential severity of each amino acid change (Ramensky et al., 2002). In addition, the relative conservation for each amino acid position in the enzyme was evaluated visually using Weblogo after alignment of 100 FAD2 protein sequences present in the National Center for Biotechnology Information database (Crooks et al., 2004). G22R was classified by Polyphen as a benign substitution (score 0.6<1), indicating it is likely to have no phenotypic effect. Additionally, alignment in Weblogo also showed that the position 22 in the FAD2-1A protein is extremely variable, indicating

that modification of the glycine residue at this position would not necessarily affect the enzyme‟s function (Figure 3.1A).

Notably, there was a deletion mutation in the *FAD2-1A* gene of PI 603452. A deletion of a single adenosine at position 544/545 in an exon region of this gene resulted in a frameshift of the translation and premature termination of the peptide after 191 amino acids.

FAD2-1B.

A total of 12 SNPs were identified in *FAD2-1B* gene for 22 PIs. Because some of the SNPs are shared among the PIs, and each PI contains a unique combination of these SNPs, the results of the SNP discovery in the 22 PIs are summarized and presented in Table 3.2. Among the 12 SNPs, five are missense mutations and seven are silent mutations in the *FAD2-1B* sequence of these 22 PIs.

Of the five missense mutations, four were reported previously and only one is new (Pham et al., 2010). Interestingly, P137R in PI 210179 and I143T in PI 578451 are identical to the mutations in PI 283327 and PI 567189A, respectively, and these pairs of lines have identical *FAD2-1B* alleles and origins (Pham et al., 2010). S86F and M126V were found in most of the PIs, and they were determined benign in Polyphen and not to be associated with the oleic acid phenotype (Pham et al., 2010).

The only novel missense mutation found in *FAD2-1B* gene is L242M in soybean line KLG10926. However, this mutation was classified by Polyphen as a benign substitution (score 0.2<1), indicating it likely does not affect the enzyme‟s functionality. This conclusion was supported by other evidence including the variability of the position 242 amino acid residue in the FAD2 protein sequence among 100 FAD2 aligned protein

sequences (Figure 3.1B), and the similarity of chemical properties between methionine and lysine. As a result, we predict that this missense mutation is not causative for the elevated oleic acid content of this soybean line.

Combinations of novel mutant alleles of *FAD2-1A* with mutant *FAD2-1B* alleles produce high oleic acid levels in the seed oil

Two independent studies report that the combination of mutant *FAD2-1A* and mutant *FAD2-1B* alleles resulted in soybean lines with more than 80% oleic acid (Pham et al., 2010; Hoshino et al., 2010). We hypothesized that the newly identified mutant *FAD2-1A* gene of PI 603452 when combined with a mutant *FAD2-1B* gene would also result in progeny with 80% oleic acid. To prove this, PI 603452 was crossed to two different lines that contain the P317R *FAD2-1B* gene derived from PI 283327. In cross 1, PI 603452 was crossed to a (Jake x PI283327) -derived soybean line to produce population KB09-13 and for cross 2, PI 603452 was crossed to a different (17D x PI283327)-derived soybean line to generate population KB09-35. The association of oleic acid phenotype and four combinations of the novel PI 603452 *FAD2-1A* alleles and P137R *FAD2-1B* alleles was analyzed using field-produced F_3 and F_4 soybean seeds developed from these two crosses. The genotype of PI 603452 herein is represented as *FAD2-1* aaBB with the lowercase allele designation always specifying the mutant allele and the capital case specifying the wild-type allele; likewise, the *FAD2-1* P137R genotype of PI 283327 is represented here as *FAD2-1* AAbb.

Two soybean lines, KB09-13-744 and KB09-13-760, from population KB09-13 and soybean line KB09-35-3-5 from population KB09-35 were identified to possess both

mutant *FAD2-1A* and *FAD2-1B* alleles (*FAD2-1* aabb) and have similar maturity. There were significant differences for average oleic, stearic, linolenic acid contents between KB09-13-744 and KB09-13-760 from population 1 (Table 3.3). Hence, the data for these two soybean lines will be presented separately. In addition, the contents of palmitic, oleic, linoleic and linolenic acids of these two lines were significantly different between the two production environments, but were similar in stearic acid contents (P = 0.12 for stearic acid, P < 0.001 for others). Statistics run on soybean lines with different combinations of genotypes for *FAD2-1*: aaBB, AAbb and AABB from population KB09-13 also gave the same results (data not shown). Therefore, data from the KB09-13 population and the two parental lines, a M23 derived high oleic acid soybean line, and Williams 82 (control) were presented by location (Table 3.3). Moreover, data for KB09-35 population were presented only for the Columbia location where all the plants were in the F_2 generation (Table 3.4).

In both of the populations, transgressive segregation for oleic acid content was observed for the lines that inherited the *FAD2-1* AABB and aabb genotypes. In population KB09-13, lines with the genotype *FAD2-1* AABB had an average oleic acid content significantly higher compared to that of the control Williams 82 in both locations (Table 3.3) (P <0.0001 for Portageville and P = 0.008 for Columbia MO). Likewise, lines with either homozygous mutant *FAD2-1A* or *FAD2-1B* alleles had average oleic acid contents significantly higher than the oleic acid content of either PI 603452 or (Jake x PI 283327) line in both locations (Table 3.3, P<0.0001 for all of the comparisons).

Two individual lines KB09-13-744 and KB09-13-760 had oleic acid contents not significantly different from each other (P = 0.06 in Columbia and 0.39 in Portageville),

but significantly higher than that of the soybean line with genotype aabb derived from M23 x PI 283327 (aabb_M23) in both locations ($P < 0.0001$ for both locations) and also higher than that of KB09-35-3-5 in Columbia location (Table 3.4).

In population KB09-35, lines with the genotype *FAD2-1* AABB had an average oleic acid content significantly lower than that of the control Williams 82 (Table 3.4) ($P = 0.003$). The soybean lines with *FAD2-1* genotypes resembling their parents contained oleic acid contents not significantly different from those of their parents ($P = 0.97$ for AAbb genotype and $P = 0.98$ for aaBB genotype). Soybean line KB09-35-3-5 had an oleic acid content of 82.4% and not significantly different from that of aabb_M23 ($P =1$), but significantly lower than those of KB09-13-744 and KB09-13-760 ($P =0.02$ and $P <0.001$, respectively).

Full fatty acid profiles and total oil and protein content

The full fatty acid profiles of the seeds of contrasting *FAD2-1*genotypic classes produced from two populations in this study revealed changes in palmitic acid, linoleic acid, and linolenic acid content (Table 3.3). As expected for a major increase in the accumulation of oleic acid, the linoleic acid and linolenic acid contents were dramatically reduced in the high oleic *FAD2-1A* and *FAD2-1B* homozygous mutant lines. The linoleic acid contents of KB09-13-744, KB09-13-760, KB09-35-3-5 and the high oleic M23-derived line in both locations were not significantly different from each other but significantly lower than those from other homozygous combinations and parental lines. Linoleic acid content was reduced to 1% for line KB09-13-760 in Columbia, which was approximately 50 times lower than soybean lines with wild type *FAD2-1* alleles (Table

3.3 and table 3.4). The linolenic acid contents of KB09-13-744, KB09-35-3-5 and the aabb_M23 line were significantly higher than that of KB09-13-760, but significantly lower than those of other homozygous combinations of *FAD2-1A* and *FAD2-1B* alleles and parental lines in both locations (Table 3.3). The greatest reduction in linolenic acid content was once again observed for line KB09-13-760 with less than 3% linolenic acid contents in both locations, compared to 8-11% in parental lines.

Interestingly, KB09-13-744 and KB09-13-760 had significantly lower stearic acid levels compared to those of the contrasting *FAD2-1* genotypes, the parental lines, aabb_M23 and KB09-35-3-5 in both populations (Table 3.3 and table 3.4). Consistent with previous studies in which mutant *FAD2-1A* and *FAD2-1B* genes were combined by Pham et al. 2010 and Hoshino et al. 2010, all of the three *FAD2-1* aabb mutant lines produced lower palmitic acid levels than lines with the *FAD2-1* AABB genotype with the most dramatic reduction recorded for KB09-13-744 and KB09-13-760. The content of palmitic acid was approximately 7.0-7.3% for the *FAD2-1* aabb mutant lines compared to 10.2-11.4% for the *FAD2-1* AABB lines.

To evaluate the impact of the enhanced oleic acid content on the total oil and protein profiles of the seeds, we analyzed protein and oil contents for the field produced F_4 seeds of KB09-13-744, KB09-13-760 harvested from Columbia and Portageville, KB09-35-3-5 from Columbia, their parental lines and Williams 82 (Table 3.3 and 3.4). The oil content of the high oleic acid soybean KB09-13-744 was not significantly different from those of their parents (P =1 for all of the comparisons) in both Columbia and Portageville. However, the protein content of this line was significantly higher in protein contents than those of the parental lines (P< 0.05 for all of the comparisons).

KB09-13-744 had significantly lower oil content and higher protein content compared to Williams 82 ($P < 0.001$ for both oil and protein content), but oil contents of both of its parental lines were also lower compared to Williams 82 (Table 3.3 and 3.4). In Portageville, the oil and protein content of KB09-13-760 were not statistically different from those of KB09-13-744. In Columbia, because only one data point was obtained for either oil or protein content of each of KB09-13-760 and KB09-35-3-5, the data were presented but not used for statistics analysis.

DISCUSSION

Plant introductions from the national germplasm collections are excellent sources of natural mutations and genetic variation that can be exploited to enrich the means to produce high oleic acid soybeans. Our study identified one novel null allele of *FAD2-1A* in PI 603452 and confirmed two PIs that have the same mutations in *FAD2-1B* which we previously reported (Pham et al. 2010). Using this novel null *FAD2-1A* allele, three soybean lines were developed to have more than 82 % oleic acid content in both testing environments. Among those, KB09-13-744 and KB09-13-760 had oleic acid content up to 86% and total linolenic and linoleic content less than 5 % in two production environments. These combinations produce the highest oleic acid content and lowest polyunsaturated acid and saturated acid contents that have been reported to date. For the first time, stearic acid contents were recorded to be lower than those of the parents and Williams 82, the control cultivar. Combinations of mutant *FAD2-1A* and mutant *FAD2-1B* created in the two studies by Pham et al. 2010 and Hoshino et al. 2010 did not have any reduction in stearic acid content compared to the parental lines or wild type soybean lines. The significantly low levels in the saturated fatty acids (palmitic + strearic) and the reduced levels in linoleic and linolenic acids of the high oleic acid genotypes offers the opportunity develop soybean oils with higher oxidative stability than previous ones. However, it is necessary that these mutant *FAD2-1* alleles are incorporated into various maturity groups and tested in various environments for stability of the high oleic acid trait, particularly in the northern US where temperatures at seed fill are cooler.

The significant contribution of this study is the identification of a natural null allele of *FAD2-1A* in PI 603452 that when combined with mutant alleles of *FAD2-1B*

produced the highest oleic acid content in soybean to date. Prior to this discovery, there were only three sources of mutant *FAD2-1A* genes in soybean including M23, KK21, and 17D to combine with four reported mutant *FAD2-1B* alleles for generation of high oleic acid soybean. The M23 *FAD2-1A* alleles, have been shown to contribute to higher and more stable oleic acid content compared to 17D and KK21 when combined with soybean lines carrying mutant *FAD2-1B* genes (Hoshino et al., 2010; Pham et al., 2010). In this study, we observed that using the null *FAD2-1A* alleles of PI 603452 resulted in significantly higher oleic acid content compared to those derived from M23, although the difference may be due to a relative maturity effect. Temperature during seed fill has long been correlated with differences in fatty acid profile, with cooler temperatures producing generally lower oleic acid contents (Heppard et al., 1996; Lee et al., 2009; Oliva et al., 2006). For both testing locations, the two PI 603452-derived lines were harvested one month earlier than the M23 derived lines, therefore they experienced warmer temperature during seed filling periods (the differences in minimum temperatures between 30 day-before-harvesting periods of two high oleic acid sources were approximately 7-8 ^{o}C in both locations), resulting in the record oleic acid contents observed for high oleic acid soybean lines grown in two testing locations. Because PI 603452 is available for public use, it is a valuable resource of the mutant *FAD2-1A* allele for the production of a stable, non-transgenic high oleic acid soybean.

Our sequencing data indicates the cause of the elevated oleic acid contents in the majority of the plant introductions used in this experiment is not due to changes in DNA coding for FAD2-1 enzymes. We hypothesize that the mechanisms or factors that cause the increased oleic acid content in these soybean lines can happen at either a regulatory

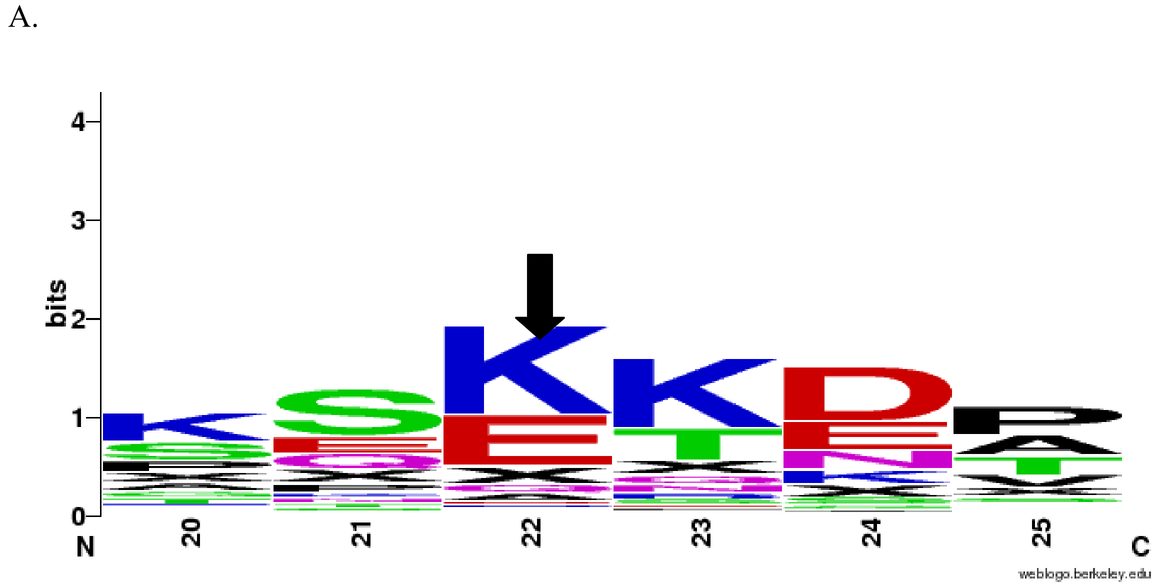
level, post transcriptional or post-translational level. Recently, a transcriptional factor and one of its partners in the transcriptional activation apparatus of *FAD2* gene were found to be needed to activate the FAD2 enzyme in sesame (Kim et al., 2007; Kim et al., 2010), whereas a mutation in a DGAT gene was identified to cause high oleic acid content in maize (Zheng et al., 2008). Also, some PIs with mid oleic acid contents were reported to have down-regulated expression levels of *FAD2-1A*, *FAD2-1B* and oleate-ACP thioesterase (*GmFATB1a*) genes and/or up-regulated expression levels of delta-nine stearoyl acyl carrier protein desaturase A, B and C (*GmSACPD*) genes (Upchurch and Ramirez, 2010). These genes are good targets for a candidate gene-based approach to identify the cause of the elevated oleic acid contents in the PIs. In addition, mapping approach can be used to identify additional genes that control the increased oleic acid content in these PIs.

It is interesting to observe that though PI 603452 and M23 carry the same type of null allele of *FAD2-1A* in which the former has a deletion of a single nucleotide in the coding region of the gene while the latter has a large section of the genome containing *FAD2-1A* gene completely deleted, the oleic acid contents in these two soybean lines are statistically different. In both locations, the oleic acid contents of PI 603452 were approximately 15% less than those of M23 (data not shown). In this case, the difference in oleic acid content between these two lines would not be explained by maturity because PI 603452 is in MG (maturity group) III while M23 is in MG V. Additionally, crossing PI 603452 to two soybean lines containing the PI 283327 *FAD2-1B* P137R allele in either Jake or the 17D background with the same maturity resulted in different levels of high oleic acid content, and it is the first time to observe that in one of the populations soybean

progeny with the parental genotype had higher oleic acid contents compared to those of the parents. This implies that there are modifying genes with small effects or other complicated mechanisms regulating of the oleic acid accumulation in soybean seed oil that have yet to be discovered. It is likely that these mechanisms may be not only determined by the genetics factor but also influenced by the environmental factor or the interaction of both.

FIGURE

A.



B.

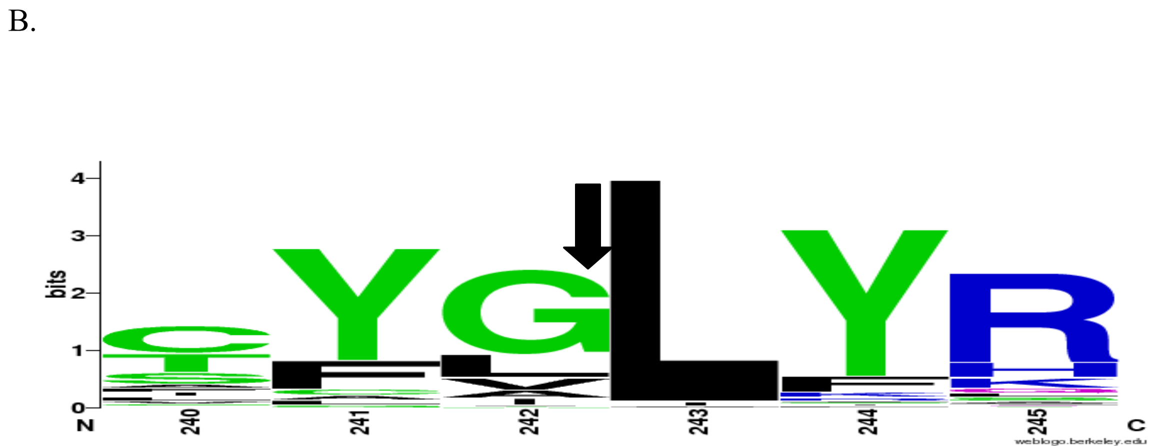


Figure 3.1: Weblogo output of the amino acid conservation *FAD2* enzyme as part of the BLINK feature at NCBI using GI number 197111722. The top 100 best matched sequences were aligned and used as input for sequence LOGO http://weblogo.berkeley.edu/logo.cgi. The logo consists of stacks of symbols, one stack for each position in the amino acid sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino acid at that position. A. Arrow indicates residue changed due to the Q22R *FAD2*-1A mutation in KLG11028, PI404160B, PI506885, and PI507420. B. Arrow indicates residue changed due to the L242M *FAD2*-1B mutation in KLG10926

TABLES

Table 3.1: Twenty two soybean lines selected for cloning and sequencing of *FAD2-1A* and *FAD2-1B* genes

Identifier number	Name	Maturity group	Origin	Oleic acid[1] (% of total oil content)
1	PI 561315	I	China	36.4
2	PI 437593 B	II	China	46.0
3	PI 567155 B	II	Japan	47.3
4	PI 603452	III	China	36.7
5	PI 592974	III	China	42.0
6	PI 578451	IV	Vietnam	28.7
7	KLG11028	ND[1]	Korea	30.9
8	KLG10926	ND	South Korea	35.0
9	Kwangankong	ND	South Korea	30.8
10	PI 416908	V	Japan	29.2
11	PI 210179	V	Taiwan	29.4
12	PI 458238	V	South Korea	31.3
13	PI 506885	VI	Japan	27.0
14	PI 467310	II	China	45.4
15	PI 561338A	II	China	48.3
16	PI 196165	III	Japan	37.6
17	PI 417054	III	Japan	49.1
18	PI 567205	IV	Germany	32.8
19	PI 458044	V	South Korea	30.0
20	PI 507307	V	Japan	30.1
21	PI 507420	V	Japan	31.0
22	PI 404160B	III	Russian Federation	47.0

[1] Oleic acid data were obtained from Dr. Grover Shannon (unpublished data, 2006-2007) and Dr. Jeong-Dong Lee (Lee et al., 2009)

Table 3.2: Variants in DNA sequences of *FAD2-1B* of 24 tested soybean lines. In the first row, numbers with amino acid abbreviation are single nucleotide polymorphisms (SNPs) positions that caused amino acid changes specified in the bracket, numbers without amino acid abbreviations are SNPs positions that have no effect to protein sequences.

Nucleotide position / Soybean lines	66	105	257 (S86F)	376 (M126V)	410 (P137R)	426 (I143)	636	657/669/682	724 (L242M)	918
Williams 82 (control) PI437593B, Kwangankong	G	A	C	A	C	T	C	CTT	T	A
PI 467310, PI 404160B, PI 561338A, PI 561315, PI 603452, PI 417054				G				TCC		G
KLG10926				G			T	TCC	A	G
PI 567155 B			T	G				TCC		G
KLG11028			T	G						G
PI 592974, PI 196165, PI 416908, PI 458044		G		G						G
PI 578451		G	T	G		C		TCC		G
PI 210179		G	T	G	G			TCC		G
PI 567205	A									
PI 458238	A	G		G						G
PI 506885, PI 507307	A		T	G				TCC		G
PI 507420	A	G		G				TCC		G

Table 3.3: Fatty acid profiles and protein and oil contents for homozygous mutant *FAD2-1A and FAD2-1B* genotypes of population KB09-13 (population 1) in field trials in Portageville and Columbia in summer 2010

		Fatty Acid					**Oil**	**Protein**
		16:0	**18:0**	**18:1**	**18:2**	**18:3**		
Portageville								
AABB_1[1]	IS[2]	11.4 ± 0.6	3.3 ± 0.6	30.2 ± 5.9	48.6 ± 3.7	7.5 ± 1.5	15.9[3]	39.5
aaBB_1	IS	10.4 ± 0.5	3.4 ± 0.2	50.2 ± 7.2	29.1 ± 6.3	6.8 ± 0.8	-[5]	-
AAbb_1	IS	10.7 ± 1.3	3.2 ± 0.5	45.4 ± 6.7	34.5 ± 4.6	6.3 ± 1.3	-	-
KB09-13-744	REP	7.3 ± 0.5a[4]	2.4 ± 0.1a	84.8 ± 0.1a	2.0 ± 0.4 a	3.2 ± 0.3b	17.8 ± 2.3 bc	42.3 ± 2.3 a
KB09-13-760	REP	7.4 ± 0.4a	2.8 ± 0.3a	85.3 ± 0.3a	1.6 ± 0.4 a	2.5 ± 0.1a	18.1 ± 0.1 b	41.1 ± 1.1 a
aabb_M23	REP	7.2 ± 0.2a	4.4 ± 0.3c	82.6 ± 0.4b	2.4 ± 0.1 a	3.2 ± 0.2b	19.4 ± 0.4 a	38.6 ± 0.3 b
PI 603452	REP	11.5 ± 0.4b	2.9 ± 0.1ab	34.5 ± 2.6c	41.5 ± 1.8b	6.5 ± 0.6c	17.1 ± 1.0 c	38.5 ± 1.0 b
PI 283327*	REP	11.0 ± 0.5b	4.7 ± 0.3e	28.6 ± 2.2d	41.2 ± 1.6b	9.2 ± 0.4d	17.6 ± 0.1 bc	38.0 ± 0.9 b
Williams 82	REP	11.0 ± 0.2b	3.7 ± 0.3d	21.8 ± 2.7e	56.1 ± 2.0c	6.9 ± 0.6c	19.4 ± 0.1 a	35.9 ± 0.1 c
Columbia								
AABB_1	IS	10.2 ± 0.4	3.5 ± 0.4	25.3 ± 4.0	53.3 ± 2.7	7.7 ± 1.4	16.6	38.3
aaBB_1	IS	8.8 ± 0.5	3.2 ± 0.2	34.5 ± 5.0	25.4 ± 4.3	7.0 ± 0.5	-	-
AAbb_1	IS	10.9 ± 0.8	3.5 ± 0.5	55.7 ± 6.0	44.0 ± 3.5	7.2 ± 2.0	-	-
KB09-13-744	REP	7.5 ± 0.3b	2.7 ± 0.1a	83.9 ± 0.5ab	1.8 ± 0.4a	4.0 ± 0.3 b	17.0 ± 0.3 b	41.1 ± 0.5a
KB09-13-760	REP	6.9 ± 0.6a	2.9 ± 0.4ab	86.4 ± 0.3a	1.0 ± 0.5a	2.9 ± 0.1 a	17.4	40.9
aabb_M23	REP	7.3 ± 0.1b	3.6 ± 0.2b	82.7 ± 1.4b	2.3 ± 1.0a	4.2 ± 0.5 b	18.9 ± 0.5 a	38.3 ± 0.6b
PI 603452	REP	11.1 ± 0.2c	3.2 ± 0.1b	31.9 ± 1.7c	46.2 ± 1.3b	7.6 ± 0.8 c	16.9 ± 0.8 b	36.9 ± 1.9b
PI 283327*	REP	10.8 ± 0.1c	4.3 ± 0.3d	23.6 ±2.3d	50.0 ±1.6b	11.3 ±1.0d	16.9 ± 0.1 b	37.4 ± 0.4b
Williams 82	REP	10.8 ± 0.1c	3.8 ± 0.1bc	21.0 ±0.5d	57.1 ±0.9b	7.4 ± 0.3 c	19.5 ± 0.2 a	33.8 ± 0.6c

[1] AA = wild-type *FAD2-1A* alleles, aa = mutant *FAD2-1A* alleles derived from PI 603452, BB = wild-type *FAD2-1B* alleles, bb = mutant *FAD2-1B* alleles derived from PI 283327. AABB_1 means that this genotype is from population 1 (KB09-13) KB09-13-744 and KB09-13-760 are soybean lines with mutant *FAD2-1A* alleles derived from PI 603452 and mutant *FAD2-1B* alleles derived from PI 283327. PI 603452 is a soybean line with mutant indel*FAD2-1A*. PI 283327* is a (Jake x PI283327) -derived line with mutant *FAD2-1B* P137R alleles and wild-type *FAD2-1A* alleles. aabb_M23 = soybean line with mutant *FAD2-1A* alleles derived from M23 and mutant *FAD2-1B* alleles derived from PI 283327. Williams 82 is the control line with normal oleic acid content and wild-type *FAD2-1.*

[2] IS = individual seeds. This means that plants that have this genotype were grown without replication in the field. Therefore, this genotype was excluded from statistical analysis in SAS. Mean value ± standard deviation was obtained by averaging fatty acid values of ten seeds used for fatty acid analysis, five seeds per line, two lines for each genotype.

REP=Replicated. This means that this soybean line was grown with replication in the field and was included in statistical analysis in SAS. Mean value ± standard deviation was obtained by averaging means of three replications which were averaged from fatty acid values of five individual seeds per replication. Standard deviation was calculated using mean values of three replications.

[3] The data have no standard deviation due to either the genotype was planted without replication or because the genotype was planted with replication but only one replication had enough seeds for oil and protein analysis.

[4] Letters of significance. Two values with same letter are not statistically different at $\alpha=0.05$. Lines with letters were analyzed using data of three replications.

[5] No data were collected for these soybean lines

Table 3.4: Fatty acid profiles for different homozygous *FAD2-1* genotypes in population KB09-35 (population 2), the two parents, two soybean lines with high oleic acid content from population 1 and control in Columbia MO summer 2010

	Fatty Acid					Oil	Protein
	16:0	18:0	18:1	18:2	18:3		
KB09-35 $F_{2:3}$ (PI 603452 x(17D xPI283327))							
AABB_2[1] (n=5[2])	11.4 ± 0.8[3] a[4]	3.7 ± 0.2 bc	17.3 ± 0.7 f	57.3 ± 1.0 a	10.4 ± 0.6 a	17.7[5]	36.8
AAbb_2 (n=10)	11.3 ± 0.4 a	3.6 ± 0.2 bc	26.8 ± 4.1 d	50.4 ± 3.3 b	8.0 ±1.0 b	-[6]	-
aaBB_2 (n=10)	10.6 ± 0.6 a	3.5± 0.2 cd	36.8± 4.7 c	41.6 ± 3.9 c	7.5± 0.9 b	-	-
KB09-35-3-5 (n=10)	7.8 ± 0.5 b	3.4 ± 0.2 cd	82.4 ± 1.9 b	2.3 ± 0.9 d	4.1 ± 0.9 c	18.3	39.3
aabb_M23 (n=15)	7.3 ± 0.2 cd	3.6 ± 0.3 bc	82.7 ± 1.8 b	2.3 ± 1.1 d	4.2 ± 0.9 c	18.9 ± 0.5 b	38.3 ± 0.6 b
KB09-13-744 (n=15)	7.5 ± 0.5 bc	2.7 ± 0.2 f	83.9 ± 1.2 ab	1.8 ± 0.7 d	4.0 ± 0.5 c	17.0 ± 0.3 c	41.1 ± 0.5 a
KB09-13-760 (n=15)	7.0 ± 0.6 d	3.0 ± 0.5 ef	86.4 ± 1.0 a	1.0 ± 0.5 d	2.9 ± 0.4 d	17.4	40.9
PI 603452 (n=15)	11.1 ± 0.3 a	3.2 ± 0.3 de	31.9 ± 5.3 c	46.2 ± 5.3 c	7.6 ±1.8 b	16.9 ± 0.8 c	36.9 ± 1.9bc
PI 283327** (n=15)	11.0 ± 0.2 a	4.0 ± 0.2 a	21.7 ±3.4 de	53.5 ±2.2 b	10.8 ±1.3 a	16.7 ± 0.2 c	37.6 ± 0.3 bc
Williams 82 (n=15)	10.8 ± 0.3 a	3.8 ± 0.3 b	21.0 ±1.6 de	57.1 ±1.4 a	7.4 ± 0.6 b	19.5 ± 0.2 a	33.8 ± 0.6 d

[1] AA = wild-type *FAD2-1A* alleles, aa = mutant *FAD2-1A* alleles derived from PI 603452, BB = wild-type *FAD2-1B* alleles, bb = mutant *FAD2-1B* alleles derived from PI 283327. AABB_2 means that this genotype is from population 2 (KB09-35) KB09-13-744 and KB09-13-760 are soybean lines with mutant *FAD2-1A* alleles derived from PI 603452 and mutant *FAD2-1B* alleles derived from PI 283327. PI 603452 is a soybean line with mutant indel*FAD2-1A*. PI 283327** is a (17D x PI283327) -derived line with mutant *FAD2-1B* P137R alleles and wild-type *FAD2-1A* alleles. aabb_M23 = soybean line with mutant *FAD2-1A* alleles derived from M23 and mutant *FAD2-1B* alleles derived from PI 283327. Williams 82 is the control line with normal oleic acid content and wild-type *FAD2-1*.

[2] Number of seeds used for fatty acid analysis

[3] Mean value ± standard deviation was obtained by averaging fatty acid values of all seeds used for fatty acid analysis

[4] Letters of significance based on P values of Student‟s t tests. Two values with same letter are not statistically different at α=0.05.

[5] The data have no standard deviation due to either the genotype was planted without replication or because the genotype was planted with replication but only one replication had enough seeds for oil and protein analysis.

[6] No data were collected for these soybean lines

LITERATURE CITED

Anai T, Yamada T, Hideshima R, Kinoshita T, Rahman S, Takagi Y (2008) Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by Gm*FAD2*-1a. Breeding Sciences 58:447 - 452.

Beuselinck P, Sleper D, Bilyeu K (2006) An assessment of phenotype selection for linolenic acid using genetic markers. Crop Science 46:747 - 750.

Buhr T, Sato S, Ebrahim F, Xing A, Zhou Y, Mathiesen M, Schweiger B, Kinney A, Staswick P, Clemente T (2002) Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. Plant Journal 30:155-163.

Crooks G, Hon G, Chandonia J-M, Brenner S (2004) WebLogo: A sequence logo generator. Genome Research 14:1188 - 1190.

Ensminger ME, Ensminger AH (1993) Foods & Nutrition Encyclopedia, 2nd edn. CRC Press.

Fehr WR (2007) Breeding for modified fatty acid composition in soybean. Crop Science 47:S-72-87.

Hartwig RA, Hurburgh CR (1990) Near-infrared reflectance measurement of moisture, protein and oil content of ground crambe seed. Journal of the American Oil Chemists' Society 67:435-437.

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

Hoshino T, Takagi Y, Anai T (2010) Novel Gm*FAD2*-1b mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with Gm*FAD2*-1a mutant alleles. Breeding Science 60:419-425.

Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, Hennekens CH, Willett WC (1997) Dietary fat intake and the risk of coronary heart disease in women. N Engl J Med 337:1491-1499.

Kim M, Go Y, Lee S, Kim Y, Shin J, Min M, Hwang I, Suh M (2010) Seed-expressed casein kinase I acts as a positive regulator of the Se*FAD2* promoter via phosphorylation of the SebHLH transcription factor. Plant Molecular Biology 73:425-437.

Kim M, Kim J-K, Shin J, Suh M (2007) The SebHLH transcription factor mediates trans-activation of the Se*FAD2* gene promoter through binding to E- and G-box elements. Plant Molecular Biology 64:453-466.

Lee JD, Woolard M, Sleper DA, Smith JR, Pantalone VR, Nyinyi CN, Cardinal A, Shannon JG (2009) Environmental effects on oleic acid in soybean seed oil of Plant Introductions with elevated oleic concentration. Crop Science 49:1762-1768.

List G, Neff W, Holliday R, King J, Holser R (2000) Hydrogenation of soybean oil triglycerides: Effect of pressure on selectivity. Journal of the American Oil Chemists' Society 77:311-314.

Mounts T, Warner K, List G, Kleiman R, Fehr W, Hammond E, Wilcox J (1988) Effect of altered fatty acid composition on soybean oil stability. Journal of the American Oil Chemists' Society 65:624-628.

Mozaffarian D, Katan MB, Ascherio A, Stampfer MJ, Willett WC (2006) Trans fatty acids and cardiovascular disease. New England Journal of Medicine 354:1601-1613.

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

Oliva M, Shannon J, Sleper D, Ellersieck M, Cardinal A, Paris R, Lee J (2006) Stability of fatty acid profile in soybean genotypes with modified seed oil composition. Crop Science 46:2069 - 2075.

Pham A-T, Lee J-D, Shannon JG, Bilyeu K (2010) Mutant alleles of *FAD2*-1A and *FAD2*-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology 10:195.

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Research 30:3894-3900.

Scherder C, Fehr W (2008) Agronomic and seed characteristics of soybean lines with increased oleate content. Crop Science 48:1755 - 1758.

Schlueter J, Lin J, Schlueter S, Vasylenko-Sanders I, Deshpande S, Yi J, O'Bleness M, Roe B, Nelson R, Scheffler B, Jackson S, Shoemaker R (2007a) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC genomics 8:330.

Schlueter J, Vasylenko-Sanders I, Deshpande S, Yi J, Siegfried M, Roe B, Schlueter S, Scheffler B, Shoemaker R (2007b) The *FAD2* gene family of soybean:Insights into the structural and functional divergence of a paleopolyploid genome. Crop Science 47.

Tang G, Novitzky W, Griffin H, Huber S, Dewey R (2005) Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. Plant Journal 44:433 - 446.

Upchurch R, Ramirez M (2010) Gene expression profiles of soybeans with mid-oleic acid seed phenotype. Journal of the American Oil Chemists' Society 87:857-864

Warner K, Gupta M (2005) Potato chip quality and frying oil stability of high oleic acid soybean oil. Food Science 70:395-400

Zheng P, Allen WB, Roesler K, Williams ME, Zhang S, Li J, Glassman K, Ranch J, Nubel D, Solawetz W, Bhattramakki D, Llaca V, Deschamps S, Zhong G-Y, Tarczynski MC, Shen B (2008) A phenylalanine in *DGAT* is a key determinant of oil content and composition in maize. Nat Genet 40:367-372

CHAPTER 4

Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil

SUMMARY

High oleic acid soybeans were produced by combining a mutant *FAD2-1A* and a mutant *FAD2-1B* gene. Despite having a high oleic acid content, the linolenic acid content of these high oleic acid soybeans was in the range of 4-6%. Therefore, a study was conducted to incorporate one or two mutant *FAD3* gene(s) into the high oleic background to further reduce the linolenic acid content. As a result, soybean lines with high oleic acid and low linolenic acid (HOLL) content were produced using different sources of mutant *FAD2-1A* genes. While oleic acid content of these HOLL lines was stable across two testing environments, the reduction of linolenic acid content varied depending on the number of mutant *FAD3* genes combined with mutant *FAD2-1* genes, on the severity of mutation in the *FAD2-1A* gene, and on the testing environment. Combination of two mutant *FAD2-1* genes and one mutant *FAD3* gene resulted in less than 2% linolenic acid content in Portageville, Missouri while four mutant genes were needed to achieve the same amount of linolenic acid content in Columbia, Missouri. This study generated non-transgenic soybeans with the highest oleic acid content and lowest linolenic acid content reported to date, offering a unique alternative to produce a fatty acid profile very similar to olive oil.

INTRODUCTION

Soybean is the largest oilseed crop worldwide with 59% of the total oilseed production, and 38% of those soybeans were produced in the U.S. (Oilseeds: World Markets and Trade, Foreign Agricultural Service, USDA at http://www.fas.usda.gov/oilseeds/circular). Accordingly, soybean is the foremost provider of oil globally and in the U.S, representing 24% of the total vegetable oil consumption in the world and ~70% total fat and oil consumption in the United States in 2010 (http://www.soystats.com/2010/). The majority of soybean oil (79%) was used for edible products, including salad and cooking oil (50%), baking and frying fat (25%) and margarine (4%) [http://www.soystats.com/]. Commercial oil made from commodity soybean has a nutritious fatty acid composition, which is low in saturated fatty acid (13%) and high in essential fatty acids including linolenic acid (ω-3, 6.2 %), linoleic acid (ω-6, 54.4%), and oleic acid (ω-9, 24.2%) compared to other fats and oils (Fehr, 2007; Mounts et al., 1994). Unfortunately, the high concentration of linoleic acid and linolenic acid in soybean oil is also responsible for the low oxidative stability and frying stability of soybean oil, resulting in a tendency towards rancidity, a rapid decrease in optimum flavor, and shortened storage time of manufactured food products (Warner and Fehr, 2008). To overcome this disadvantage, soybean oil used for food manufacturing purposes is usually partially hydrogenated to reduce the linolenic and linoleic acid contents and increase the oleic acid content. However, an unwanted outcome of this process is the creation of 10-40% *trans*-fat, an unhealthy chemical for which human consumption has been linked to a higher risk of obesity, decreased cardiovascular health, and other health problems (Hu et al., 1997; Mounts et al., 1994; Mozaffarian et al., 2006).

Hence, there have been numerous efforts using traditional breeding or genetic engineering approaches to increase the oxidative stability of soybean oil without hydrogenation by means of producing soybean seeds high in oleic acid concentration and low in polyunsaturated fatty acid concentration.

Because the soybean genome has experienced rounds of duplication followed by limited sequence loss, one trait is often controlled by multiple homologous genes; this feature of the soybean genome presents a challenge for soybean breeders to develop soybean lines with multiple traits (Chi et al., 2011; Schmutz et al., 2010). Traditional breeding was successful in identifying the loci that are responsible for oleic acid and linolenic acid content, but selection was primarily based on phenotype when knowledge about the genes and molecular markers associated with the trait was limited (Fehr, 2007). When the soybean genome was completely sequenced, the two oleate desaturase genes *FAD2-1A* (Glyma10g42470) and *FAD2-1B* (Glyma20g24530) and the three linoleate desatuase genes *FAD3A* (Glyma14g37350), *FAD3B* (Glyma02 g39230) and *FAD3C* (Glyma18g06950) were unambiguously assigned a chromosomal position, and were characterized to mainly control the oleic acid content and linolenic acid content in soybean seed oil, respectively, using candidate gene based approaches (Bilyeu et al., 2003; Heppard et al., 1996; Schlueter et al., 2007; Schmutz et al., 2010). Mutations identified in either *FAD2-1A* or *FAD2-1B* resulted in elevated oleic acid contents up to 50% (Anai et al., 2008; Dierking and Bilyeu, 2009; Pham et al., 2010). Mutation in one or more of the three homologous genes of *FAD3* also resulted in lower linolenic acid content from 7-10% in lines containing wild-type alleles to approximately 4% or 3% for lines containing mutations in one or two *FAD3* genes, respectively (Bilyeu et al., 2006;

Bilyeu et al., 2011; Reinprecht et al., 2009). When all of the mutant homologous genes of either *FAD2-1* or *FAD3* were combined, the catalytic activities of the FAD2 and FAD3 enzymes were severely reduced, resulting in more than 80% oleic acid content or 1% linolenic acid content (Bilyeu et al., 2006; Bilyeu et al., 2011; Hoshino et al., 2010; Pham et al., 2010).

Although utilization of the mutant alleles of the *FAD2-1* and *FAD3* genes was successful in creating either high oleic acid or low linolenic acid soybean, these soybean lines still have some disadvantages. Soybeans with low linolenic acid content often were influenced greatly by temperature with lower linolenic acid content often seen with higher temperature (Rennie and Tanner, 1989; Wilcox and Cavins, 1992) In addition, Warner et al. 2008 showed that tortilla chips fried in oils that have 1% linolenic acid content showed higher stability than those prepared with normal soybean oil, but significantly less stability than those fried in high oleic acid oil (80%), especially when they aged more than two months after 35 and 55 hours of frying. The authors reasoned that it is because oils with 1% linolenic acid content still have wild type oleic acid and linoleic acid content. On the other hand, soybean with more than 80% oleic acid content generated by the combination of two mutant *FAD2-1A* and *FAD2-1B* genes had the linolenic acid content reduced to half of the linolenic acid content in commercial soybean oil, which still requires hydrogenation (Hoshino et al., 2010; Pham et al., 2010). Moreover, the linoleic acid content (18:2) is usually equal or lower than the linolenic acid content (18:3) in the high oleic acid soybean lines (Pham et al., 2010). In a study by Warner et al. 2005, it was suggested that the linoleic acid content in the oil is partially responsible for the deep fried flavor of the foods, and tortillas fried with high oleic acid

oils had an unexpected off-flavor because of the exceptionally low linoleic acid content, which was even lower than the linolenic acid content (Warner and Gupta, 2005). Therefore, in order to have soybean oils with high stability but no adverse effect on the flavor of the food, it is rational to combine mutant *FAD3* genes into the high oleic acid background to reduce the linolenic acid content at the expense of the linoleic acid content to avoid both off-flavor effects and hydrogenation requirements.

One question addressed in this study was to find out how many mutant *FAD3* genes needed to be combined with two mutant *FAD2-1A* and *FAD2-1B* genes to produce less than 3% linolenic acid content, the current industry standard for low linolenic acid soybean oils. Because the amount of linoleic acid in 80% high oleic acid soybean oil is only 3-5% compared to 55% in commercial soybeans, we hyphothesized that less than three mutant *FAD3* genes would be needed to produce the target linolenic acid content in the high oleic acid background. Among the three *FAD3* genes, the *FAD3A* gene was shown to have a greater impact on linolenic acid level in soybean seed than *FAD3B* and *FAD3C*, consistent with higher expression of *FAD3A* in developing seeds (Bilyeu et al., 2005; Bilyeu et al., 2003). We hypothesized that crossing an 80% high oleic soybean line carrying two mutant *FAD2-1* genes to a low linolenic soybean line carrying two mutant *FAD3A* and *FAD3C* genes will produce in the progeny soybeans with oleic acid content more than 80% and linolenic acid content lower than 3%. Although it is expected that this may require incorporation of four genes into one soybean line and a larger effort to develop populations, it may be possible to achieve this goal using only a single *FAD3* mutation in combination with two *FAD2* mutations. The objectives of this project were: 1) to combine two mutant *FAD2-1* genes with one or two mutant *FAD3* genes to produce

high oleic acid (>75%) and low linolenic acid (<3%) soybean. 2) to examine the stability of the high oleic low linolenic soybeans in appropriate environments to select the most stable phenotype/genotype combination.

MATERIALS AND METHODS

Population development

Our strategy was to cross a soybean line with genotype *FAD2-1*(aabb) *FAD3*(AACC), designated herin as HO(Δ or $S^{117}N$) to a soybean line with genotype *FAD2-1*(aaBB) *FAD3*(aacc) or *FAD2-1*(AABB) *FAD3*(aacc), designated herin as MO(Δ)LLac or NOLL3c, respectively (Table 4.1). The genotype herein is represented with the lowercase allele designation always specifying the mutant allele and the capital case specifying the wild-type allele. Two crosses were made for this project: Cross 1=HO(Δ)LL, soybean line S08-1692, homozygous for M23 *FAD2-1A*Δ allele and PI 283327 *FAD2-1B* P137R allele, was crossed to a mid-oleic low linolenic acid soybean line KB 07-1 #123, homozygous for M23 *FAD2-1a*Δ allele, 10-73 *FAD3A* and *FAD3C* alleles; and cross 2= HO(S117N)LL, a high oleic parent homozygous for 17D *FAD2-1A* $S^{117}N$ alleles and PI 283327 *FAD2-1B* P137R alleles*,* was crossed to 10-73, homozygous for *FAD3A* and *FAD3C* alleles (Bilyeu et al., 2005). The mutant *FAD2-1B*, *FAD3A and FAD3C* genes used in this research are identical, while two mutant *FAD2-1A* alleles were evaluated: the *FAD2-1A* alleles donated from either M23 (Δ) or 17D ($S^{117}N$) (Dierking and Bilyeu, 2009; Pham et al., 2010; Sandhu et al., 2007)

F_1 seeds of the two crosses were collected and verified for true hybridization by SimpleProbe assay for *FAD2-1b* allele (Pham et al., 2010) True F_1 seeds were planted in the field and advanced to get F_2 seeds. F_2 seeds were germinated and genotyped for three genes *FAD2-1B*, *FAD3A* and *FAD3C* in cross 1‟s population (HO(Δ)LL) or all of four mutant genes in cross 2‟s population (HO($S^{117}N$)LL).

For cross 1 (HO(Δ)LL), there were nine F_2 soybean plants with the genotype *FAD2-1(aabb) FAD3(AaCc)* , two plants for each of the three genotypes *FAD2-1(aabb) FAD3(aacc)*= HO(Δ)**LL4ac**, *FAD2-1*(*aabb*) *FAD3*(*aaCC*) =HO(Δ)**LL3a** or *FAD2-1*(*aabb*) *FAD3*(*AAcc*)= HO(Δ)**LL3c** were identified. Only F_2 plants with the desirable gene combinations (*FAD2-1(aabb) FAD3(aaCC/AAcc/ aacc*) were selected and advanced to $F_{2:3}$ generation in Sears growth chamber, University of Missouri, Columbia . $F_{2:3}$ seeds were harvested and bulked for each of the three genotypes and analyzed for fatty acid profile. The bulked $F_{2:3}$ seeds for each genotype of HO(Δ)**LL4ac** , HO(Δ)**LL3a** or HO(Δ)**LL3c** were used for stability test in Columbia and Portageville Missouri described below and data from F_4 seeds were used for statistical analysis and presented in Table 4.1. For the cross 2 (HO(S117N)LL), ten true F_1 seeds were planted but only one F_2 population was used for subsequent experiment. The F_2 population was planted at a winter nursery center near Upala, Costa Rica in February 2010 and DNA samples from individual F2 plants were collected in FTA cards for genotyping assay conducted in our lab at the University of Missouri, Columbia in March 2010. After genotyping, one F_2 soybean plant with the genotype *FAD2-1*(aabb) *FAD3*(AaCc), and two soybean plants with genotype *FAD2-1*(aabb) *FAD3*(aaCC), were kept to advance for $F_{2:3}$ seeds in the nursery center. A total of 120 F_3 seeds of the plant with genotype *FAD2-1*(aabb) *FAD3*(AaCc) were harvested from Costa Rica in May 2010 and germinated for a second round of genotyping. Subsequenly, three soybean plants with genotype *FAD2-1*(aabb) *FAD3C*(aacc) =HO(S117N)**LL4ac**, four with genotype *FAD2-1(aabb)FAD3(aaCC)* =HO(S117N)**LL3a_3**, and one with genotype *FAD2-1*(aabb) *FAD3C*(AAcc) = HO(S117N)**LL3c** were identified and transplanted to the field in Columbia, MO for seed

production. In addition, ten F_3 seeds from each of the two F_2 plants with genotype *FAD2-1*(aabb) *FAD3*(aaCC), designated as HO(S117N)**LL3a_1** and HO(S117N)**LL3a_2,** were also planted at the same time with the transplants in the same field in Columbia, MO for seed increase. However, only seeds harvested from three individual plants for either HO(S117N)LL3a_1 or HO(S117N)LL3a_2 genotype were used for fatty acid analysis together with seeds from plants with genotypes HO(S117N)**LL4ac,** HO(S117N)**LL3a_3 and** HO(S117N)**LL3c .**

FAD2-1A allele specific molecular marker assay

SimpleProbe assay for the deletion in the *FAD2-1A* of 17D was developed based on the SimpleProbe protocol described by Pham et al. 2010. The Probe contained 5‟-Fluorescence-SPC- GTACTTGCTGAAGGCATGGTGA -Phosphate-3‟ (the underlined base is mutated to T in the *FAD2-1A* (S117N) allele). Primers used to generate template for Simpleprobe genotyping assay were designed by aligning the *FAD2-1A* and *FAD2-1B* region containing the SNPs. Primers were selected to be as close as possible to the SNPs while differing in at least 3 nucleotides between the two genes to specifically amplify the targeted region in *FAD2-1A*. Genotyping reactions were performed with a 5:1 asymmetric mix of primers (5‟- CCAAGGTTGCCTTCTCACTGGT -3‟ at 5 µM final concentration, and 5‟- TAGGCCACCCTATTGTGAGTGTGAC -3‟ at 1 µM final concentration). Reactions were carried out in 20 µl; containing template, primers, 0.2 µM final concentration of SimpleProbe, buffer (40 mM Tricine-KOH [pH 8.0] 16 mM $MgCl_2$, 3.75 µg ml^{-1} BSA,), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Genotyping reactions were performed

using a Lightcycler 480 II real time PCR instrument (Roche), using the following PCR parameters: 95°C for 5 minutes followed by 40 cycles of 95°C for 20 seconds, 65°C for 20 seconds, 72°C for 30 seconds, and then a melting curve from 50°C to 68°C. 17D and all soybean lines with an identical *FAD2-1A* allele genotype have a characteristic peak at 54°C, while Williams 82 (wild-type *FAD2-1A*) have a peak at 62°C. Heterozygous individuals's genotype showed two peaks at 54°C and 62°C.

Genotyping for the gene combination of interest

Except for F_1 seeds in which the whole seeds were germinated, for every genotyping assay conducted, seeds were first chipped to get a small portion for fatty acid determination. The remaining portion with hypocotyl was germinated in germination packages to collect unifoliate leaf tissue for DNA on Whatman FTA cards and the DNA was collected for PCR reactions using Whatman FTA card protocol BD05 (whatman.com). SimpleProbe assays of 17D *FAD2-1A*, PI 283327 *FAD2-1B*, *FAD3A* and *FAD3C* alleles were conducted as described (Bilyeu et al., 2011; Pham et al., 2010).

Fatty acid determination

Gas chromatography of total fatty acid methyl esters of extracted oil method was used to examine the fatty acid profiles of each individual seed in F_3 and F_4 generations (Beuselinck et al., 2006). A small portion of the seeds (without the hypocotyl) was chipped and used as samples for the fatty acid analysis. For the field produced seeds from the HO(Δ)LL-derived lines, seeds from five plants in one replication of a genotype were bulked and five seeds from each bulk were assayed individually for fatty acid analysis. For the field produced seeds from the HO(S117N) LL-derived lines, seeds from

individual F_3 plants of each genotype were bulked and five seeds from each bulk were assayed individually for fatty acid analysis.

Stability of high oleic/low linolenic soybean

F_3 seeds of three genotypes HO(Δ)**LL4ac**, HO(Δ)**LL3a** and HO(Δ)**LL3c** from cross 1 were grown in Columbia and Portageville, Missouri in May 28th 2010 together with the parents, KB07-1-123 and a soybean line with the same *FAD2* genotype as S08-1692, and various checks including Williams 82 (MG III), LG04-6863 (MG mid group IV), 5002T (MG IV late to V early), N98-4445A (MG IV), M23 (MG V). There were three replications for each location, and Randomized Complete Block Design was used for setting up the experiment. In both locations, plantings were made in rows spaced 76cm apart. Within the rows, ten seed of each soybean line was planted in a hill plot, spaced 5 cm apart. After seedling emergence, hills were thinned to five plants from which all data was collected for each soybean line.

For the cross 2 (HO(S117N)LL), F_3 seedlings with genotypes HO(S117N)**LL4ac,** HO(S117N)**LL3a_3 and** HO(S117N)**LL3c** were grown 10 cm apart from each other in a hill plot adjacent to the experiment for the population derived from cross 1 on June 9th. Ten F_3 seeds from each of two lines with either HO(S117N)**LL3a_1** or HO(S117N)**LL3a_2** genotype were grown in a hill plot in the same field, spaced 5 cm apart. Ten seeds for two parents, a 17D high oleic soybean line and 10-73were also planted in the field at the same time.

Statistical analysis

The data and statistical analysis presented for HO(Δ)LL lines with M23 derived *FAD2-1A* gene was generated using proc mixed procedure, SAS® 9.2 Enhanced Logging Facilities, Cary, NC: SAS Institute Inc., 2008 (Table 4.2). The data presented for HO(S117N) LL lines with 17D *FAD2-1A* alleles was generated using fatty acid values of all seeds used for gas chromatography analysis (Table 4.3). Comparisons of significant differences between two soybean lines presented in this table were done using Student"s t test with $\alpha = 0.05$.

RESULTS

The combination of four mutant genes: *FAD2*-1A, *FAD2*-1B, *FAD3*A and *FAD3*C herein is represented as either HO(Δ)LL4ac or HO(S117N)LL4ac, to distinguish the null *FAD2*-1A allele donated from either M23 (Δ) or 17D (S117N) (Dierking and Bilyeu, 2009; Sandhu et al., 2007). The mutation in *FAD2*-1B is the P137R missense mutation from PI 283327, and the *FAD3*A splice-site mutation and *FAD3*C $G^{128}E$ misense mutation are derived from CX1512-44 (Bilyeu et al., 2003; Pham et al., 2010).

The three mutant gene combination of *FAD2*-1A, *FAD2*-1B and *FAD3*A is designated as HO(Δ or S117N)**LL3a**, while the three mutant gene combination of *FAD2*-1A, *FAD2*-1B and *FAD3*C is designated as HO(Δ or S117N)**LL3c**. The comparison lines were: high oleic Parent:HO(Δ or S117N) indicating a parental genotype of mutant *FAD2*-1A (Δ or S117N) and *FAD2*-1B along with wild-type *FAD3* genes; mid-oleic acid low linolenic acid line “Parent:MO(Δ)**LLac**", which is line KB07-1-123 with a genotype of mutant *FAD2-1A*(Δ), but wild-type *FAD2-1B*, as well as mutant *FAD3A* and *FAD3C* (Bilyeu et al., 2003); and Parent: NOLLac, which was soybean line 10-73, with wild-type *FAD2-1* genes and mutant *FAD3*A and *FAD3*C genes (Bilyeu et al., 2003). Low linolenic acid lines B1-52abc with wild-type *FAD2*-1 genes and mutant *FAD3A*, *FAD3B*, and *FAD3C* genes (splice site mutant *FAD3B* alleles derived from A29 (Bilyeu et al., 2006). The reference line "Williams 82" with wild-type alleles of the *FAD2* and *FAD3* genes was also used as a check cultivar.

Maturity of the HOLL lines

In Columbia, average days from planting to harvesting across three replications of HO(Δ)**LL4ac** and HO(Δ)**LL3c** was 111 days, of HO**LL**(Δ)**3a** was 118 days, which were similar to Williams 82 with MG group III (114 days). In Portageville, average days from planting to harvesting across three replications of HO(Δ)**LL4ac** was 110, of HO(Δ)**LL3a** was 106, of HO(Δ)**LL3c** was 111 which were also similar to Williams 82 (112 days).

In Columbia, average days from planting to harvest across three replications of HO(S117N)**LL4ac** was 116 days, of HO(S117N)**LL3a** and HO(S117N)**LL3c** was around 120 days.

For the HOLL soybean lines with null FAD2-1A(Δ) alleles derived from M23

Combination of four mutant genes

At Portageville, except for linolenic acid, the contents of the other four fatty acids of HO(Δ)**LL4ac** were not different from those of HO(Δ)**LL3a** or HO(Δ)**LL3c** (Table 4.2). The oleic acid contents of HO(Δ)**LL4ac** in both locations were approximately 85% on average, and not significantly different from those of HO(Δ)**LL3a** or HO(Δ)**LL3c**. The linolenic acid content of HO(Δ)**LL4ac** was 1.5%, significantly higher than that of B1-52abc, the low linolenic acid control soybean line (1.2 %) and lower than that of HO(Δ)**LL3c** (2 %), but not significantly different from that of HO(Δ)**LL3a** (1.8 %) (Figure 4.1). In contrast, at Columbia, the linolenic acid content of HO(Δ)LL**4ac** was 1.9%, significantly higher than that of B1-52abc (1.3%) but significantly lower than those of HO(Δ)**LL3a** or HO(Δ)**LL3c**, which were 2.6% and 2.5%, respectively (Table 4.2 and Figure 4.2). In this location, stearic acid and linoleic acid contents of the four mutant

gene combination was also significantly different from those of the three mutant gene combinations while no difference was seen for oleic acid and palmitic acid contents. Stearic acid content of HO(Δ)**LL4ac** in Columbia was significantly lower than that of HO(Δ)**LL3a** but not different from that of HO(Δ)**LL3c**. Linoleic acid content of HO(Δ)**LL4ac** was significantly higher than that of both HO(Δ)**LL3a** and HO(Δ)**LL3c**, which reflects less enzymatic activity for two mutant *FAD3* genes combined compared to one mutant *FAD3* gene in HO(Δ)**LL3a** and HO(Δ)**LL3c**.

Compared to its parents, HO(Δ)**LL4ac** had significantly higher palmitic acid, oleic acid and linoleic acid contents and lower stearic acid and linolenic acid contents than those of the high oleic parent in both Columbia and Portageville. It also had significantly lower linolenic acid contents than those of the mid oleic low linolenic parent MO(Δ)**LLac** in both locations. Compared to the control cultivar „Williams 82", there was a significant change in the fatty acid profile of HO(Δ)LL**4ac** soybeans as it was observed a significantly higher oleic acid content and a significant reduction in contents of all of the other fatty acids (Table 4.2). It was reported that there is often a reduction in palmitic acid content but not stearic acid content in high oleic acid soybean (80%) (Hoshino et al., 2010; Pham et al., 2010); however, the HO(Δ)**LL4ac** soybean lines developed in this study showed a significant increase in palmitic acid content and a significant reduction in stearic acid content compared to the parental lines, and a significant reduction in both palmitic and stearic acid contents compared to those of Williams 82. The smallest significant change observed was for stearic acid, which was about a 25% reduction in the high oleic acid lines compared to the wild type stearic acid content of Williams 82 (from 3.6% to 2.7%). Connected to the dramatic change in oleic acid content in the high oleic

acid lines was a reduction in linoleic acid content, which was equivalent to a 97% reduction compared to the wild-type linoleic acid content of Williams 82 (from 56% to less than 3%).

Combination of three mutant genes

There were no significant differences in fatty acid contents between HO(Δ)**LL3a** and HO(Δ)**LL3c** in Portageville. On average, both HO(Δ)**LL3a** and HO(Δ)**LL3c** have about 86% oleic, 2 % linoleic and 2% linolenic acid. However, in Columbia, these two genotypes had statistically different contents of stearic, oleic, and linoleic acid but not the contents of palmitic and linolenic acid. Location-wise, only the linolenic acid content of the three mutant gene combination is influenced by the planting location with linolenic acid levels in Columbia significantly higher than those of Portageville.

Compared to its high oleic parents, HO(Δ)**LL3a** or HO(Δ)**LL3c** lines had higher oleic acid contents and lower linolenic acid content. In addition, compared to those of the mid oleic low linolenic parent MO(Δ)**LLac**, the linolenic acid content of HO(Δ)**LL3a** and HO(Δ)**LL3c** was significantly lower in Portageville but was not significantly different in Columbia.

Statistical analysis also indicated that, for the HO(Δ)**LL3a** and HO(Δ)**LL3c** lines developed with null alleles of *FAD2*-1A derived from M23, location is the factor that influenced the contents of palmitic and linolenic acids but not stearic, oleic, and linoleic acids.

For the HOLL soybean lines with S117N missense FAD2-1A alleles from 17D

For this population, data from F_4 seeds with genotypes HO(S117N)**LL4ac**, HO(S117N)**LL3a**, and HO(S117N)**LL3c** derived from one F_2 high oleic acid plant heterozygous for *FAD3*A and *FAD3*C and two additional HO(S117N)**LL3a** lines derived from two independent F_2 high oleic acid plants that were genotyped to be homozygous for mutant *FAD3*A and wild-type *FAD3*C. The three HO(S117N)LL3a lines are designated as HO(S117N)**LL3a_1**, HO(S117N)**LL3a_2**, HO(S117N)**LL3a_3** (Table 4.1).

Compared to HO**LL3a**_1, HO(S117N)**LL4ac** had significantly lower palmitic and linolenic acid contents, higher linoleic content and similar stearic and oleic acid contents (Table 4.3). Compared to HO**LL3a**_2, HO(S117N)**LL4ac** had significantly higher palmitic content and lower linolenic acid content, and similar stearic, oleic and linoleic acid contents. Fatty acid composition of HO(S117N)**LL4ac** was not significantly different from that of HO(S117N)**LL3a_3**, except for linolenic acid content Linolenic acid content of HO(S117N)**LL4ac** (2.5%) was significantly lower than that of HO(S117N)**LL3a_3** (3.5%).

Compared to HO(S117N)**LL3c**, HO(S117N)**LL4ac** had significantly difference in content of all fatty acids except for oleic acid. The palmitic and linolenic acid contents of HO(S117N)**LL4ac** were significantly lower than those of (S117N)**LL3c** while the stearic and linoleic acid contents were significantly higher than those of HO(S117N)**LL3c** (Table 4.3 and Figure 4.2).

The oleic acid content of HO(S117N)**LL4ac** (80.2%) was not significantly different from that of the high oleic parent but significantly higher than that of Williams

82. Its linolenic acid content was not significantly different from that of 10-73, the low linolenic acid parent, but significantly lower than that of Williams 82.

Combination of three mutant genes

There was a marked difference between fatty acid composition of seeds of HO(S117N)**LL3a** and HO(S117N)**LL3c**. HO(S117N)**LL3c** soybean lines had significantly higher palmitic and linolenic acid contents, lower stearic and linoleic content and similar oleic acid content compared to the three HO(S117N)**LL3a** lines. The only exception is that HO(S117N)**LL3c**"s linoleic acid content was not different from that of HO(S117N)**LL3a** _1. Three HO(S117N)**LL3a** lines had approximately 3.7% stearic, 80% oleic acid and 3-4% linolenic acid. Their palmitic and linoleic acid contents vary depending on individual line. HO(S117N)**LL3a**_2 had the lowest palmitic acid content (6.4%) and highest linoleic acid content (7%) among three HO(S117N)**LL3a** lines (Table 4.3). The linolenic acid content of HO(S117N)**LL3a_1**is significantly lower than that of HO(S117N)**LL3a_3**, and the linolenic acid content of HO(S117N)**LL3a _2** is in between of these two values and not different from that of HO(S117N)**LL3a _1** or HO(S117N)**LL3a_3**. The oleic acid contents of all the HO(S117N)**LL3a** or HO(S117N)**LL3c** were not significantly different from that of the high oleic acid parent but significantly higher than that of Williams 82. In addition, their linolenic acid contents were significantly lower than that of Williams 82 but higher than that of 10-73, probably because these two genotypes only contain one mutant *FAD3* gene compared to two mutant gene in 10-73.

Comparisons between HO(Δ)LL and HO(S117N)LL lines in Columbia MO

The total content of linoleic and linolenic acid in HO(Δ)LL lines was lower (~5%) than those in HO(S117N)LL lines (Figure 4.2). On the other hand, the oleic acid content of HO(Δ)LL lines were 5 % higher than those of HO(S117N)LL lines (Table 4.3). Unlike HO(Δ)LL lines, which showed an increase in palmitic acid content and a reduction in stearic acid content compared to those of their high oleic acid parental lines, HO(S117N)LL lines had an equal or significantly reduced content of palmitic acid and an equal or significantly increased content of stearic acid content compared to those of their high oleic acid parents.

Agronomic characteristics of HOLL lines

The soybean lines used in this study responded similarly to the daylength requirement for flowering time and maturity, with maturity equivalent to those of maturity group (MG) III. The plants from population 1 were short, with thick green leaves, and appeared to have inherited a dominant gene for determinate growth habit (Tian et al., 2010). Seed weight of these soybean lines with *FAD2*-1A derived from either M23 or 17D was not different from that of the parents (data not shown).

Oil and protein content

The oil contents of HO(Δ)**LL4ac**, HO(Δ)**LL3a** and HO(Δ)**LL3c** all showed a reduction of 1% in oil content compared to those of the parents and Williams 82, but this was not observed for HO(S117N)LL lines (Table 4.2 and 4.3). As a result, the protein contents of the HO(Δ)LL lines were approximately 1% greater than those of the parents and 3-4 % greater than that of Williams 82. While the oil contents of HO(S117N)LL lines in general were not different from those of the parents and Williams 82,Tthe protein

contents of these lines were 3-5% higher than those of the parents and Williams 82. The oil and protein contents of HO**LL4ac** were not significantly different from those of HO**LL3a** and HO**LL3c** independent of the source of the mutant *FAD2*-1A alleles.

DISCUSSION

Prior to this work, conventional breeding and genetic engineering have been employed to combine elevated oleic acid and low linolenic acid in one soybean line (Brace et al., 2011; Rahman et al., 2001). The mid oleic low linolenic acid soybean developed by Rahman et al. had approximately 55% oleic acid and 4.2% linolenic acid and possesses three mutant genes: *FAD2-1A*(Δ) from M23, mutant *FAD3B* from M5 and mutant *FAD3A* from M24 (Anai et al., 2005; Rahman et al., 2001). By combining a transgenic event which silences transcriptional activities of *FAD2-1A* and *FAD2-1B* genes and two mutant *FAD3A* (C1640) and *FAD3C* (RG10) genes, HOLL soybeans with oleic acid content in the range of 77-79 % and linolenic acid content in the range of 2.1-2.8 % have been produced by Brace et al. 2011. The novelty of our study is the production of non-transgenic HOLL lines conditioned by three or four mutant genes with similar fatty acid profiles to that of transgenic HOLL. In this study, the four mutant gene combination with null *FAD2-1A*Δ alleles from M23 produced seeds with higher oleic acid content and lower linolenic acid content compared to the transgenic HOLL reported by Brace et al. 2011, while the 17D derived missense *FAD2-1A* S117N HOLL lines produced seeds with similar oleic acid and linolenic acid contents compared to the transgenic HOLL. However, this difference may be due to the relative maturity difference between soybean lines and/or temperature difference between testing locations. Because our testing locations are at lower latitudes compared to those of Brace et al. 2011, we predict that our HOLL lines with four mutant genes may have a fatty acid profile similar to that of the transgenic HOLL soybeans reported by Brace et al. when the

genes are incorporated into the appropriate maturity group for planting in locations in more northern latitudes described by Brace et al. 2011.

Our study indicates that in two testing environments in Missouri, with M23 mutant *FAD2-1A* Δ alleles contributing to the high oleic acid phenotype, only one mutant gene of either *FAD3A* or *FAD3C* is needed to lower the linolenic acid content to less than 3%. For the HO(S117N)LL lines, all combinations of three mutant genes with either mutant *FAD3A* or *FAD3C* had 3-5% linolenic acid content in the Columbia, MO location. In this environment it requires four mutant genes to produce <3% linolenic acid content consistently, and combining four genes could be more challenging for soybean breeders. Recently, our group has generated another source of high oleic acid soybeans with nearly 85% oleic and less than 3% linolenic acid content with a combination of a mutant *FAD2*-1*A* allele containing a single base deletion resulting in a frameshift and premature translation termination from PI 603452 and the missense mutant *FAD2*-1*B* gene from PI 283327 (Bilyeu, unpublished). We anticipate that with this high oleic acid soybean background, similar to the situation with the null *FAD2-1* alleles from M23, only one mutant *FAD3* gene will be needed to lower the content of linolenic acid to below 3% in Missouri production environments and possibly two mutant *FAD3* genes will be necessary to further reduce linolenic acid content to 1% in cooler environments.

Oleic acid and linolenic acid content was demonstrated to be influenced greatly by temperature and modifier genes by several studies (Chapman et al., 1976; Dornbos and Mullen, 1992; Graef et al., 1988; Hyten et al., 2004; Wilcox et al., 1993). It is shown clearly in this study that when two mutant *FAD3A* and *FAD3C* genes were incorporated into a high oleic acid background to reduce the enzymatic activity of *FAD3*, the linolenic

acid content of HOLL soybeans was still affected by environmental factors while the oleic acid content showed a reduction that was statistically insignificant. Although the oleic acid content was not reduced significantly in a cooler environment, the stability of the fatty acid constituents of the oil of HOLL soybean needs to be further evaluated by conducting experiments in a wider scale, with more diverse environments, and across different years. The instability of linolenic acid content across environments was possibly due to the influence of temperature on the enzymatic activity of the wild type *FAD3B* enzyme in the HOLL lines, because mutant *FAD3A* and *FAD3C* genes from 10-73ac used in this study would not produce enzymes with proper function. A study demonstrated that FAD2 and FAD3 enzymatic activities in soybean seeds cultured in vitro greatly declined when the temperature increased from 20 ^{o}C to 25 ^{o}C (100 fold for *FAD2* and 60 fold for *FAD3*, and almost inactive at 35 ^{o}C for both of the enzymes) (Cheesbrough, 1989). In addition to the environmental effect, modifier genes may also play a role in controlling the linolenic content in the HOLL lines developed in this study. We identified three HO(S117N)LL3a lines with the same gene combination and had significantly different linolenic acid contents when planted in Columbia MO. Because they were planted close to each other in the same environment, the *FAD3* enzymes in each plant in each line should receive the same environmental signals and cues, and therefore should respond in the same fashion. However, the existing variation for linolenic acid content between these lines HO(S117N)LL3a was significant, indicating of a more complicated regulatory mechanisms that may involve other genes besides *FAD3* genes for linolenic acid content.

The changes of saturated acid contents, oil and protein contents of HO(Δ)LL compared to those of high oleic acid parents were different from HO(S117N)LL lines. HO(Δ)LL lines showed a small but significant increase in palmitic acid content and a small but significant decrease in stearic acid content compared to the high oleic acid content while it was reduction for palmitic acid and increase for stearic acid for HO(S117N)LL compared to its high oleic acid parent. The change in the saturated acid content of HO(Δ)LL is more valid because it was shown in both of the testing locations while data for HO(S117N)LL were obtained from individual plants grown only in Columbia. In addition, few but not all of the HO(S117N)LL lines showed the described changes. Nevertheless, because the reverse changes in the palmitic and stearic acid content of these HOLL lines resulted in no change the total content of saturated fatty acid compared to that of the high oleic acid content, but still resulted in lower saturated fatty acid content compared to Williams 82, therefore, this change should not affect the usage values of the HOLL soybean lines. The increase in protein content at the expense of a reduction in oil content of the HO(Δ)LL lines with M23 derived null *FAD2*-1A alleles is in agreement with the report by Brace et al. 2011. In contrast, the HO(S117N)LL lines with 17D mutant *FAD2*-1A alleles maintained the oil content equivalent to those of the high oleic parent (19-20%) and had higher protein contents compared to those of the two parents and Williams 82 in the Columbia location. The small reduction in oil content of the HO(Δ)LL lines, however, can be improved when they are incorporated into an elite background to enhance agronomic traits, including oil and protein content.

It is necessary that the new high oleic low linolenic oils produced from our HOLL soybeans be tested for effect on flavor of the foods. Although the oils extracted from

HOLL soybeans may probably have the highest oxidative stability of soybean oil to date, the influence to characteristics of food products including taste, flavor, and textures must be carefully evaluated considering that 75% of soybean oils are being used for food preparation and production. Warner and Gupta (2005) have shown that although having the highest oxidative and frying stability, HOLL oil with 85% oleic, 1.3% linoleic and 2% linolenic had the lowest sensory scores compared to those of low linolenic acid oil (2%) and the 1:1 mixture of the high oleic acid and low linolenic oils (50% oleic, 2% linolenic). They also suggested that as the linolenic acid contents were the same in the three tested types of oil, the low sensory scores of high oleic soybean oil may be due to the extremely low content of linoleic acid; the low linolenic acid oil which had the highest linoleic acid content also had the highest sensory scores (Warner and Gupta, 2005). However, in another study, only the potato chips that were freshly fried with high oleic low linolenic sunflower oil (8% saturated fatty acids, 78 % oleic acid, 12% linoleic, 0.1% linolenic) had lower flavor scores compared to those fried in other oils. During the storage time up to 6 months, the flavor scores of food prepared with HOLL sunflower oil were still lower than those with other oils but they were not significantly different (Warner et al., 1997). It is hoped that by combining mutant *FAD3* genes with two mutant *FAD2-1* genes, the linolenic acid content of the HO soybean would be reduced and become less than the linoleic acid content so that the flavor issue can be evaluated. It was reported in three studies in human infants, rat and chicken that the most beneficial ratio of linoleic: linolenic acid content to improve health is equal or higher than 4:1 (Clark et al., 1992; Puthpongsiriporn and Scheideler, 2005; Yehuda et al., 1996). Olive oil, which has 15% saturated fatty acid, 75% oleic, 9% linoleic and 1% linolenic acid has been long

claimed to be a one of the most healthy natural vegetable oils (White, 2007). Among the HOLL lines we created, seeds of HO(S117N)**LL4ac** soybean lines with 17D *FAD2-1A*(S117N) alleles had the fatty acid profile that is close to that of olive oil with 11% saturated fatty acids, 80% oleic, 7% linoleic and 2% linolenic and the ratio of linoleic: linolenic acid content was 3.5. It is believed that with this olive- like fatty acid composition, high oleic low linolenic soybean can offer more applications for industrial purposes and can be used as a less expensive but equally healthy alternative to olive oil.

FIGURES

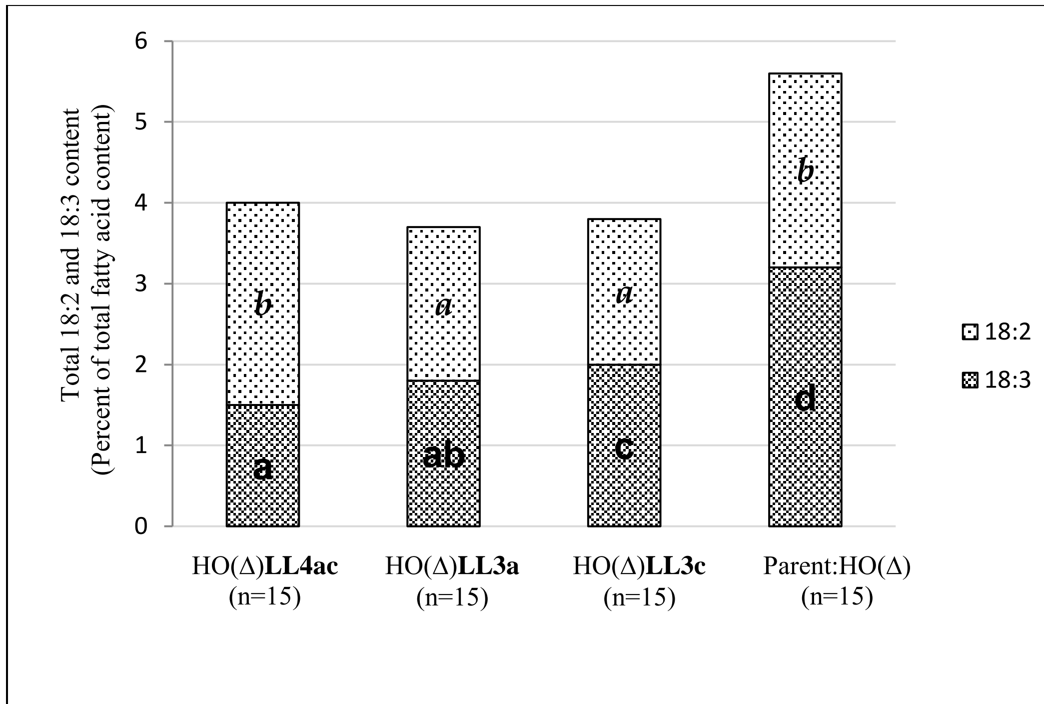


Figure 4.1: Seed linoleic (18:2) and linolenic acid (18:3) contents of high oleic low linolenic acid (HOLL) soybean lines developed with null allele ofM23 *FAD2-1A* in Portageville Missouri summer 2010. Linoleic and linolenic acid phenotypic data of each genotype are the mean of linoleic and linolenic acid contents as a percentage of the total fatty acid content of the oil from five individual samples for each line, three replications per line. Bars with the same letter represent two values that are not statistically significantly different. Italic letters indicate significance for linoleic acid data while bold bigger font letters are for linolenic acid data.

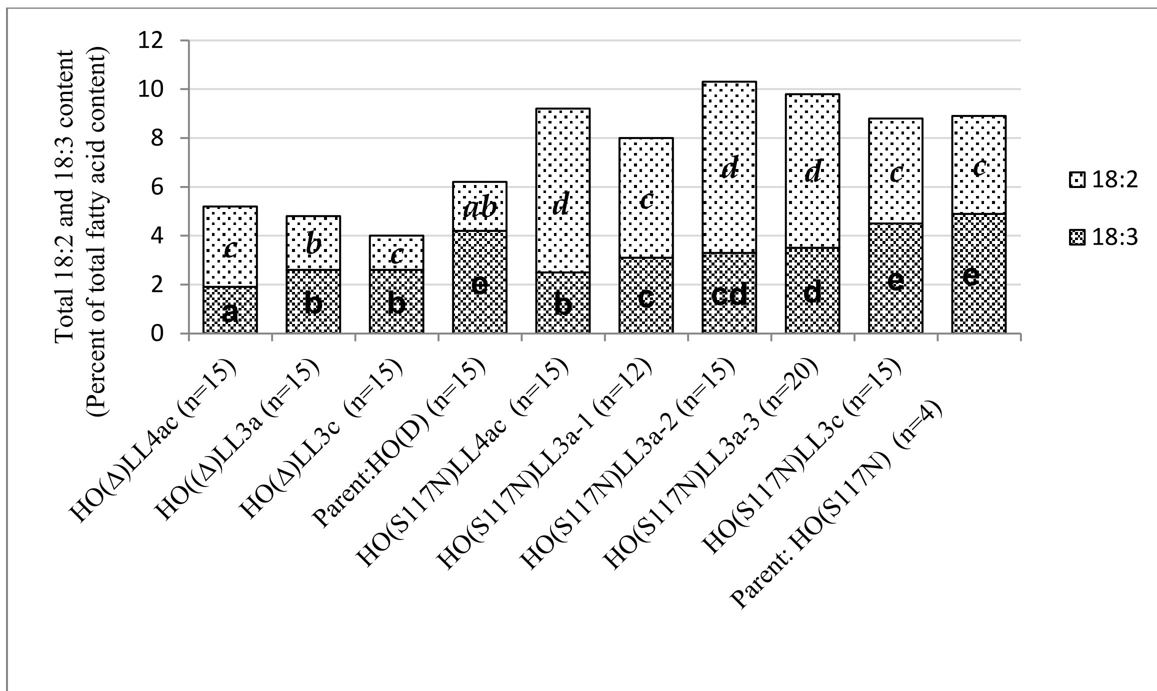


Figure 4.2: Seed linoleic (18:2) and linolenic acid (18:3) content of HOLL soybean lines developed with either M23 or 17D *FAD2-1A* alleles in Columbia summer 2010. Linoleic and linolenic acid phenotype data of each genotype are the mean of linolenic acid content as a percentage of the total fatty acid content of the oil from the a number of individual samples for each line indicated by n. Two bars with the same letter are not statistically significant different from each other using Student"s t test comparison. Bars with the same letter represent two values that are not statistically significantly different. Italic letters indicate significance for linoleic acid data while bold bigger font letters are for linolenic acid data.

TABLES

Table 4.1: Gene combination of HOLL lines, parental lines and control lines used in the study

Genotype	Gene				
	FAD2-1A	**FAD2-1B**	**FAD3A**	**FAD3B**	**FAD3C**
HO(Δ)**LL4ac**	Δ	P137R[3]	Splice site ($G^{810}A$)[4]	WT	G128E[6]
HO(Δ)**LL3a**	Δ	P137R	Splice site ($G^{810}A$)	WT	WT
HO(Δ)**LL3c**	Δ	P137R	WT	WT	G128E
Parent:HO(Δ)	Δ	P137R	WT	WT	WT
Parent:MO(Δ)LLac	Δ	WT	Splice site ($G^{810}A$)	WT	G128E
HO(S117N **LL4ac**	S117N[2]	P137R	Splice site ($G^{810}A$)	WT	G128E
HO(S117N) **LL3a-1**	S117N	P137R	Splice site ($G^{810}A$)	WT	WT
HO(S117N) **LL3a-2**	S117N	P137R	Splice site ($G^{810}A$)	WT	WT
HO(S117N) **LL3a-3**	S117N	P137R	Splice site ($G^{810}A$)	WT	WT
HO(S117N)LL**3c**	S117N	P137R	WT	WT	G128E
Parent: HO(S117N)	S117N	P137R	WT	WT	WT
Parent: NOLLac	WT	P137R	Splice site ($G^{810}A$)	WT	G128E
B1-52abc	WT	WT	Splice site ($G^{810}A$)	Splice site (G>A)[5]	G128E
Williams 82	WT[7]	WT	WT	WT	WT

[1] FAD2-1A allele derived from M23 (Sandhu et al., 2007)
[2] FAD2-1A allele derived from 17D (Dierking and Bilyeu, 2009)
[3] FAD2-1B allele derived from PI 283327 (Pham et al., 2010)
[4] FAD3C allele derived from CX1512-44 (Bilyeu et al., 2005)
[5] FAD3B allele derived from A29 (Bilyeu et al., 2006)
[6] FAD3C allele derived from CX1512-44 (Bilyeu et al., 2005)
[7] A wild type allele of the gene

Table 4.2: Fatty acid profiles and protein and oil contents for high oleic low linolenic soybeans with M23 *FAD2-1A* allleles in field trials in Portageville and Columbia in summer 2010

	Fatty Acid					**Oil**	**Protein**
	16:0	**18:0**	**18:1**	**18:2**	**18:3**		
Portageville							
HO(Δ)**LL4ac**	7.9 ± 0.2[2]d	2.7 ± 0.2 a	85.3 ± 0.9 a[2]	2.5 ± 0.6 b	1.5 ± 0.1 b	18.1 ± 0.1	39.3 ± 0.4
HO(Δ)**LL3a**	7.6 ± 0.2 c	2.9 ± 0.3 a	85.9 ± 1.0 a	1.9 ± 0.4 a	1.8 ± 0.4 bc	18.4 ± 0.3	39.8 ± 0.3
HO(Δ)**LL3c**	7.5 ± 0.01b	2.8 ± 0.1 a	85.9 ± 0.4 a	1.8 ± 0.2 a	2.0 ± 0.2 c	18.1 ± 0.3	39.8 ± 0.9
Parent:HO(Δ)	7.2 ± 0.2 a	4.4 ± 0.3 c	82.8 ± 0.4 b	2.4 ± 0.1 b	3.2 ±0.2 e	19.4 ± 0.4	38.6 ± 0.3
Parent:MO(Δ)LLac	10.8 ±0.4 d	3.8 ± 0.1 b	29.6 ± 2.0 c	53.1 ± 1.9c	2.6 ± 0.2 d	19.1 ± 0.4	36.8 ± 1.5
B1-52abc	10.7 ± 0.6 d	3.9 ± 0.3 b	28.1 ± 5.7 c	56.2 ± 5.3d	1.2 ± 0.1 a	20.0 ± 0.3	33.9 ± 0.6
Williams 82	11.2 ± 0.2 e	3.7 ± 0.2 b	22.5 ± 2.7 d	56.1 ± 2.0d	6.5 ± 0.6 f	19.4 ± 0.1	35.9 ± 0.1
Columbia							
HO(Δ)**LL4ac**	7.5 ± 0.1 b	2.9 ± 0.1 a	84.5 ± 0.7 b	3.3 ± 0.6 c	1.9 ± 0.2 b	17.5 ± 0.5	39.5 ± 0.5
HO(Δ)**LL3a**	7.5 ± 0.2 b	3.3 ± 0.2 b	84.4 ± 0.6 b	2.2 ± 0.3 b	2.6 ± 0.2 c	17.9 ± 0.6	39.1 ± 0.7
HO(Δ)**LL3c**	7.5 ± 0.1 b	2.8 ± 0.1 a	85.9 ± 0.5 a	1.4 ± 0.4 a	2.5 ± 0.1 c	17.6 ± 0.3	39.8 ± 0.3
Parent:HO(Δ)	7.3 ± 0.1 a	3.5 ± 0.2 b	82.7 ± 1.4 b	2.0 ± 1.0ab	4.2 ± 0.6 d	19.1 ± 0.2	37.8 ± 0.5
Parent:MO(Δ)LLac	10.1 ± 0.01c	3.8 ± 0.1 c	33.6 ± 1.9 c	49.8 ± 1.8d	2.6 ± 0.1 c	18.7 ± 0.2	36.1 ± 0.3
B1-52abc	10.2 ± 0.1 c	4.2 ± 0.1 d	24.9 ± 0.9 d	59.4± 0.8 f	1.3 ± 0.1 a	20.1 ± 0.1	35.2 ± 0.6
Williams 82	10.8 ± 0.1 d	3.8 ± 0.1 c	21.0 ± 0.5 e	57.1 ± 0.9 e	7.4 ± 0.3 e	19.5 ± 0.2	33.8 ± 0.6

[1] Mean value ± standard deviation was obtained by averaging means of three replications which were averaged from fatty acid values of five individual seeds per replication. Standard deviation was calculated using mean values of three replications.

[2] Letters of significance. Two values with same letter are not statistically different at α=0.05. Lines with letters were analyzed using data of three replications and only assigned for oleic and linolenic acid content only.

Table 4.3: Fatty acid profiles and protein and oil contents for high oleic low linolenic soybeans with 17D *FAD2-1A* allleles in field trials in Columbia in summer 2010

	Fatty Acid					Oil	Protein
	16:0	**18:0**	**18:1**	**18:2**	**18:3**		
HO(S117N)**LL4ac** (n=15)[1]	7.0 ± 0.2[2]b[3]	3.6 ± 0.5 ab	80.2± 1.5 ab[4]	6.7 ± 1.3 b	2.5 ± 0.7 b	19.4 ± 0.1	38.2 ± 0.0
HO(S117N)**LL3a-1** (n=12)	7.4 ± 0.3 c	3.7 ± 0.5 b	80.8 ± 0.8 a	4.9 ± 0.8 a	3.1 ± 0.5 c	20.2 ± 0.6	36.6 ± 0.2
HO(S117N)**LL3a-2** (n=15)	6.4 ± 0.2 a	3.6 ± 0.1 b	79.7 ± 1.2 b	7.0 ± 0.8 b	3.3 ± 0.3 cd	20.1 ± 0.1	36.4 ± 0.6
HO(S117N)**LL3a-3** (n=20)	7.0 ± 0.3 b	3.6 ± 0.4 ab	79.5 ± 2.1 ab	6.3 ± 1.4 b	3.5 ± 0.5 d	19.9 ± 0.1	37.3 ± 0.3
HO(S117N)LL**3c** (n=15)	7.7 ± 0.4 d	3.3 ± 0.1 a	80.1 ± 2.3 ab	4.3 ± 2.0 a	4.5 ± 0.5 e	19.4 ± 0.2	37.4 ± 0.5
HO(S117N) (n=10)	7.1 ± 0.2 b	3.4 ± 0.2 a	80.1 ± 1.7 ab	4.3 ± 0.8 a	5.3 ± 0.7 f	19.7 ± 0.5	36.7 ± 1.4
Parent: HO(S117N) (n=4)	7.6 ± 0.4 cd	3.3 ± 0.4 a	80.1 ± 2.2 ab	4.0 ± 1.0 a	4.9 ± 1.0 ef	19.4 ± 0.7	37.8 ± 1.5
Parent: NOLLac (n=10)	11.6 ±0.4 g	4.2 ± 0.3 c	21.8 ± 3.0 d	59.9 ± 2.8 d	2.5 ± 0.1 b	20.4 ± 0.3	33.6 ± 0.2
B1-52abc (n=15)	10.2 ± 0.2e	4.2 ± 0.3 c	24.9 ± 3.0 c	59.4 ± 2.9 d	1.3 ± 0.1 a	20.1 ± 0.1	35.2 ± 0.6
Williams 82 (n=15)	10.8 ± 0.3f	3.8 ± 0.3 b	21.0 ±1.6 d	57.1 ± 1.4 c	7.4 ± 0.6 g	19.5 ± 0.2	33.8 ± 0.6

[1] Number of seeds used for fatty acid analysis
[2] Mean value ± standard deviation was obtained by averaging fatty acid values of all seeds used for fatty acid analysis
[3] Letters of significance based on P values of Student‟s t tests. Two values with same letter are not statistically different at α=0.05. Letters were assigned for oleic and linolenic acid content only.

LITERATURE CITED

Anai T., Yamada T., Hideshima R., Kinoshita T., Rahman S., Takagi Y. (2008) Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by Gm*FAD2*-1a. Breeding Sciences 58:447 - 452.

Anai T., Yamada T., Kinoshita T., Rahman S.M., Takagi Y. (2005) Identification of corresponding genes for three low-[alpha]-linolenic acid mutants and elucidation of their contribution to fatty acid biosynthesis in soybean seed. Plant Science 168:1615-1623.

Beuselinck P.R., Sleper D.A., Bilyeu K.D. (2006) An assessment of phenotype selection for linolenic acid using genetic markers. Crop Science 46:747-750.

Bilyeu K., Palavalli L., Sleper D., Beuselinck P. (2005) Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science 45:1830-1836.

Bilyeu K., Palavalli L., Sleper D., Beuselinck P. (2006) Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science 46:1913 - 1918.

Bilyeu K.D., Gillman J.D., LeRoy A.R. (2011) Novel *FAD3* mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Science 51:259-264

Bilyeu K.D., Palavalli L., Sleper D.A., Beuselinck P.R. (2003) Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science 43:1833-1838.

Brace R.C., Fehr W.R., Schnebly S.R. (2011) Agronomic and seed traits of soybean lines with high oleate concentration. Crop Sci. 51:534-541.

Chapman G., Robertson J., Burdick D., Parker M. (1976) Chemical composition and lipoxygenase activity in soybeans as affected by genotype and environment. Journal of the American Oil Chemists' Society 53:54-56.

Cheesbrough T.M. (1989) Changes in the Enzymes for Fatty Acid Synthesis and Desaturation during Acclimation of Developing Soybean Seeds to Altered Growth Temperature. Plant Physiology 90:760-764.

Chi X., Yang Q., Lu Y., Wang J., Zhang Q., Pan L., Chen M., He Y., Yu S. (2011) Genome-Wide Analysis of Fatty Acid Desaturases in Soybean (*Glycine max*). Plant Molecular Biology Reporter:1-15.

Clark K.J., Makrides M., Neumann M.A., Gibson R.A. (1992) Determination of the optimal ratio of linoleic acid to linolenic acid in infant formulas. The Journal of pediatrics 120:S151-S158.

Dierking E., Bilyeu K. (2009) New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biology 9:89.

Dornbos D., Mullen R. (1992) Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists' Society 69:228-231.

Fehr W.R. (2007) Breeding for modified fatty acid composition in soybean. Crop Science 47:S-72-87.

Graef G.L., Fehr W.R., Miller L.A., Hammond E.G., Cianzo S.R. (1988) Inheritance of Fatty Acid Composition in a Soybean Mutant with Low Linolenic Acid. Crop Sci. 28:55-58.

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

Hoshino T., Takagi Y., Anai T. (2010) Novel Gm*FAD2*-1b mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with Gm*FAD2*-1a mutant alleles. Breeding Science 60:419-425.

Hu F.B., Stampfer M.J., Manson J.E., Rimm E., Colditz G.A., Rosner B.A., Hennekens C.H., Willett W.C. (1997) Dietary fat intake and the risk of coronary heart disease in women. N Engl J Med 337:1491-1499.

Hyten D., Pantalone V., Saxton A., Schmidt M., Sams C. (2004) Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. Journal of the American Oil Chemists' Society 81:1115-1118.

Mounts T.L., Warner K., Usta G.R., Neffa W.E., Wilson R.F. (1994) Low linolenic acid soybean oils-Alternatives to frying oils. Journal of the American Oil Chemists' Society 71:495-499.

Mozaffarian D., Katan M.B., Ascherio A., Stampfer M.J., Willett W.C. (2006) Trans fatty acids and cardiovascular disease. New England Journal of Medicine 354:1601-1613.

Pham A.-T., Lee J.-D., Shannon J.G., Bilyeu K. (2010) Mutant alleles of *FAD2*-1A and *FAD2*-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology 10:195.

Puthpongsiriporn U., Scheideler S. (2005) Effects of dietary ratio of linoleic to linolenic acid on performance, antibody production, and in vitro lymphocyte proliferation in two strains of leghorn pullet chicks. Poult Sci 84:846-857.

Rahman S.M., Kinoshita T., Anai T., Takagi Y. (2001) Combining ability in loci for high oleic and low linolenic acids in soybean. Crop Science 41:26-29.

Reinprecht Y., Luk-Labey S.Y., Larsen J., Poysa V.W., Yu K., Rajcan I., Ablett G.R., Pauls K.P. (2009) Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. Plant Breeding 128:253-258.

Rennie B., Tanner J. (1989) Fatty acid composition of oil from soybean seeds grown at extreme temperatures. Journal of the American Oil Chemists' Society 66:1622-1624.

Sandhu D., Alt J., Scherder C., Fehr W., Bhattacharyya M. (2007) Enhanced oleic acid content in the soybean mutant M23 is associated with the deletion in the *FAD2*-1a gene encoding a fatty acid desaturase. Journal of the American Oil Chemists' Society 84:229-235.

Schlueter J.A., Vasylenko-Sanders I.F., Deshpande S., Yi J., Siegfried M., Roe B.A., Schlueter S.D., Scheffler B.E., Shoemaker R.C. (2007) The *FAD2* gene family of soybean:Insights into the structural and functional divergence of a paleopolyploid genome. Crop Science 47:S-14 - 26.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.-C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Tian Z., Wang X., Lee R., Li Y., Specht J.E., Nelson R.L., McClean P.E., Qiu L., Ma J. (2010) Artificial selection for determinate growth habit in soybean. Proceedings of the National Academy of Sciences.

Warner K., Fehr W. (2008) Mid-Oleic/Ultra Low Linolenic Acid Soybean Oil: A Healthful New Alternative to Hydrogenated Oil for Frying. Journal of the American Oil Chemists' Society 85:945-951.

Warner K., Gupta M. (2005) Potato chip quality and frying oil stability of high oleic acid soybean oil. Food Science 70:395-400.

Warner K., Orr P., Glynn M. (1997) Effect of fatty acid composition of oils on flavor and stability of fried foods. Journal of the American Oil Chemists' Society 74:347-356.

White P.J. (2007) Fatty acid in oilseeds (Vegetable oils), in: C. K. Chow (Ed.), Fatty acids in foods and their health implications, CRC press, Marcel Dekker, Inc., New York. pp. 210-263.

Wilcox J.R., Cavins J.F. (1992) Normal and Low Linolenic Acid Soybean Strains: Response to Planting Date. Crop Sci. 32:1248-1251.

Wilcox J.R., Nickell A.D., Cavins J.F. (1993) Relationships between the fan Allele and Agronomic Traits in Soybean. Crop Sci. 33:87-89.

Yehuda S., Brandys Y., Blumenfeld A., Mostofsky D.I. (1996) Essential Fatty Acid Preparation Reduces Cholesterol and Fatty Acids in Rat Cortex. International Journal of Neuroscience 86:249-256.

CHAPTER 5

Identification of candidate genes encoding for three acyltransferase enzymes controlling the triacylglycerol biosynthesis in soybean seed

SUMMARY

Soybean oil is one of the most economically important products of soybean, which accounts for more than 25% of the vegetable oil consumption in the world and 70% of the total fat and edible consumption of the U.S. However, soybean oil represents only 20% of the soybean seed dry weight and negatively correlates to the protein content. To enhance the oil content in soybeans, it is essential to gain more knowledge about genes controlling key points in the biosynthesis pathway of triacylglycerol (TAG), the predominant component of soybean oil. Among the genes in the TAG biosynthesis pathway, three genes involved in the final acylation steps, *GPAT*, *LPAAT* and *DGAT* are good targets because overexpression experiments conducted on these three genes in *Arabidopsis* and some oilseed crops resulted in increases of up to 50% oil content. Unfortunately, in soybean, little is known about members within these three gene families and their roles in TAG production. Moreover, as the soybean genome is paleopolyploid, one gene in *Arabidopsis* often corresponds to two or four corresponding homologous genes in soybean, and some of them may be pseudogenes or may not be expressed in seeds. Therefore, a series of experiments were conducted to identify *GPAT*, *LPAAT* and *DGAT* genes that are highly expressed during soybean embryogenesis, and thus may hold an important role for oil accumulation in soybean seed. Characterization of the expression level of individual genes in each gene family will not only provide us with a better understanding of the biosynthesis of TAG in soybean, but also help to identify genes that can be used for candidate gene-based approaches to enhance soybean oil content.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is the largest oilseed crop worldwide with 58% of the total oilseed production, and 35 % of those soybeans were produced in the U.S. (http://www.soystats.com/2010/). Accordingly, soybean is also the foremost provider of oil globally and in the U.S, representing 29 % of the total vegetable oil consumption in the world and 70 % total fat and oil consumption in the United States (http://www.soystats.com/2010/). However, oil content in soybean seeds (20% seed dry weight) is two to three times lower than oil content of other oilseed crops such as rapeseed (40%), sunflower (40-50%) and castor been (60%), and is known to negatively correlate with the protein content (Chi et al., 2011; Kim et al., 2007; Yadav et al., 1993). In addition to traditional breeding methods, candidate gene-based approaches have proved to be an effective means to enhance soybean oil quality and quantity (Bilyeu et al., 2005; Courchesne et al., 2009; Frentzen, 1998; Xu et al., 2010). However, for this approach, it is essential to determine candidate genes that hold key points in the biosynthesis pathway of triacylglycerols (TAG), the predominant component of soybean oil (Chi et al., 2011).

The biosynthesis pathway of TAG involves three different stages: synthesis of fatty acid in plastids, triacylglycerol assembly in the endoplasmic reticulum, and storage of TAGs in the cytosol (Kim et al., 2007; Weselake, 2005). In the context of this study, only the second stage will be described as it contains three enzymes that will be the focus of this project. During triacylglycerol assembly in the endoplasmic reticulum, fatty acyl groups produced in the plastids are incorporated into glycerol-3-phosphate (glycerol-3-P) backbones via the Kennedy pathway. The Kennedy pathway involves the sequential

acylations of glycerol-3-phosphate to produce TAG. In the first acylation reaction, glycerol-3-P acyltransferase (*GPAT*) catalyzes the esterification of sn-glycerol-3-phosphate by a fatty acid coenzyme A ester at position sn-1 to form lysophosphatidic acid (Figure 5.1). The second acylation reaction is catalyzed by lysophosphatidic acid acyltransferase (*LPAAT*) enzyme to position sn-2 to produce phosphatidic acid (PA). Subsequently, a phosphatidic acid phosphatase participates in the dephosphorylation of PA to form sn-1,2 diacylglycerol (DAG), an immediate precursor of TAG production. Finally, diacylglycerol acyltransferase (DGAT) catalyses the acyl-coenzyme A (acyl-CoA)-dependent acylation of DAG to generate TAG. However, TAGs could also be produced via an acyl-CoA independent pathway, in which the acyl groups can be transfered to lysophosphatidylcholine via the enzyme lysophosphatidylcholine acyltransferase. The enzyme phospholipid:diacylglycerol acyltransferase (PDAT) will then transfer the acyl groups to DAG to generate TAG (Raneses et al., 1999a; Stahl et al., 2004).

In the group of acyltranferase genes in the Kennedy pathway, *GPAT*, *LPAAT* and *DGAT* have been shown to be effective targets to be used for the modification of fatty acid composition and seed oil content (Courchesne et al., 2009; Raneses et al., 1999). Though less information is available for *GPAT* and *LPAAT* compared to *DGAT*, their role in promoting oil content in seeds is strongly supported. Expression of plastidal safflower *GPAT* or E. coli *GPAT* gene increased seed oil content in *Arabidopsis* from 15 to 22% (Jain et al., 2000). In the same fashion, overexpression of a yeast *LPAAT* gene in soybean, rapeseed and *Arabidopsis* resulted in oil levels increased up to 1.5%, 22% and 45% in average, respectively (Rao and Hildebrand, 2009; Scherder and Fehr, 2008a; Zou

et al., 1997). Consistent with this result, Maisonneuve et al, reported that expressing individually two *B. napus LPAAT* isozymes in *Arabidopsis* produced seeds with 14% greater lipid content and 7% greater seed mass (Maisonneuve et al.).

For *DGAT* gene, considerable evidence exists that it holds an important role in TAG accumulation in plants. There are two distinct gene families of *DGAT* that share no homology with each other: *DGAT1* and *DGAT2* (Shockey et al., 2006; Yen et al., 2008). DGAT1 was first cloned from mouse while *DGAT2* was first identified in the fungus Morteriella ramanniana with two homologs *DGAT2A* and *DGAT2B* (Cases et al., 1998; Lardizabal et al., 2001). In both plants and animals, DGAT1 proteins often have about 500 amino acids (aa) with six to nine transmembrane domains, while DGAT2"s protein sequences are much shorter with about 300 aa and two transmembrane domains (Shockey et al., 2006; Yen et al., 2008) These two enzymes were also reported to localize in the different subdomains of the ER and have different expression level profiles in seeds of six oilseed crops (Kim et al., 2010b; Kroon et al., 2006; Shockey et al., 2006).

Despite many differences in protein sequence and structure, cellular location, and expression level, both of them were demonstrated to play an important role in TAG production. In plants, the significant role of *DGAT1* gene in accumulation of seed oil was reported in olive, tobacco, *Arabidopsis*, canola, castor bean, burning bush, soybean, tung tree, and maize (Bouvier-Navé et al., 2000; He et al., 2004; Jako et al., 2001; Milcamps et al., 2005; Mozaffarian et al., 2006; Nykiforuk et al., 2002; Shockey et al., 2006; Wang et al., 2006; Weselake et al., 2008; Zheng et al., 2008). Overexpression of a *DGAT1* gene in *Arabidopsis* and four different crops led to increases in seed oil content from 11-47% (Bouvier-Navé et al., 2000; Jako et al., 2001; Lardizabal et al., 2008; Zheng

et al., 2008). Meanwhile, DGAT2 was suggested to be the most important enzyme in TAG production in yeast (Reinprecht et al., 2009). Notably, expression of *M. ramanniana DGAT2A* in insect cells and soybean showed a 3.1 fold increase in the amount of TAG and 1.5% increase in seed oil content per dry weight, respectively (Lardizabal et al., 2008). In developing seeds of tung tree and castor bean, the expression level of *DGAT2* was higher than that of *DGAT1* (approximately 18 fold in castor bean), and was suggested to be responsible for the formation of seed oil (Kroon et al., 2006; Shockey et al., 2006). *DGAT* is also a good candidate gene for improvement of oleic acid content in soybean seed by altering the acyl channel in to TAG. An insertion mutation of 81 bp in the exon 2 of *Arabidopsis DGAT1* gene resulted in half the amount of oleic acid, compensated by a doubled amount of linolenic acid in seed oil compared to those of wild-type seed (Jako et al., 2001). In addition, overexpression of a *DGAT* gene with an insertion of three base pairs encoding for a phenylalanine at position 469 in maize resulted in increases in seed oil content up to 41% and oleic content up to 107% Recently, Oakes et al. reported that expression of a fungal *DGAT2* in maize kernels led to a change in fatty acid composition in which the oleic acid content increased up to 18% and palmitic, linoleic and linolenic content all decreased (Oakes et al., 2011), suggesting that *DGAT* can influence both oleic acid and seed oil content in plants.

Though the three acyltransferase genes mentioned above are good targets that can be used to increase the soybean seed oil content, up to date, there have been no published data for *GPAT* and *LPAAT* homologous genes and their roles in TAG production in soybean. For *DGAT*, sequence and expression level of a *DGAT1* gene was characterized in cultivated and wild soybean (Wang et al., 2006). However, the expression level of this

gene was similar in different types of tissues including seed, leaf and flower, which prompted the authors to propose that this gene may not be the most important gene controlling the oil content in soybean seeds. In 2008, two homologous genes of *DGAT1* were reported: *DGAT1a*, which consists of 7575 bp in the genomic sequence (AB257589) and 99 % similarity to the *DGAT1* gene reported by Wang et. al. (2006); and *DGAT1b*, which consists of 8164 bp in the genomic sequence (AB257590) (Hildebrand et al., 2008). Though *DGAT1b* has a greater activity compared to *DGAT1a*, activities of *DGAT1*s from soybean were five-fold less than the activity of a *DGAT1* gene from *Vernonia galamensis* (Hildebrand et al., 2008). Expression of *DGAT2* gene in soybean was reported to be significantly lower compared to that of *DGAT1* gene (Heppard et al., 1996a). However, it was not clear from the study whether the expression level of *DGAT1* and *DGAT2* each was obtained from a homologous gene with the highest expression in each family or was averaged from expression levels of all homologous genes in each gene family.

The objectives of this project were: 1) Identification of homologous genes within the soybean genome for three gene families (*GPAT*, *LPAAT* and *DGAT*) and their expression in public sources of microarray data; and 2) Identification of differential gene expression of all homologous genes of the three candidate genes to select for the ones with highest expression levels in seed.

MATERIALS AND METHODS

Database search for Arabidopsis sequences of *GPAT, LPAAT* and *DGAT* genes

Methods of identifying orthologous genes among plant species have been used extensively and shown to be powerful in studying genes that may control the same phenotype across various species (Bruner et al., 2001a; Frentzen, 1998). In general, orthologs in different species hold the same function as they preserve one or more protein motifs and/or 3-dimensional structures. Therefore, publicly available sources for nucleotide and protein sequence of the National Center for Biotechnology Information were used to search for sequences of *Arabidopsis GPAT*, *LPAAT* and *DGAT* genes. In addition, gene entries were taken from all the papers that reported on these genes in *Arabidopsis* (Bilyeu et al., 2011; Gidda et al., 2009; Jako et al., 2001; Salt, 1957).

Homologous sequence identification and transcriptional expression data mining

Protein sequences of *Arabidopsis GPAT*, *LPAAT*, and *DGAT* were used as a query to determine all putative homologs of soybean *GPAT*, *LPAAT* and *DGAT* genes. The query protein sequence of GPAT1 is At1g06520, GPAT2 is At3g11430, LPAAT protein sequence of *Arabidopsis* is At3g57650, DGAT1 is At2g19450, DGAT2 is At3g51520. The soybean Phytozome databases were used for BLAST searches of soybean sequences (http://www.phytozome.net/). Genes were selected based on similarity in sequence and annotated function. After that, expression data of these genes were exploited from Soybase (http://soybase.org/soyseg). The expression data from Soybase is obtained by microarray data using Affymetrix GeneChip technology. All of the homologous genes identified were selected for RT-PCR experiments.

RNA isolation, reverse transcriptase reactions, and quantitative Real-Time PCR

Based on DNA sequence of each gene, specific primer pairs for each of the putative *GPAT*, *LPAAT*, and *DGAT* genes were designed using Primer3Plus program (Mounts et al., 1994). PCR products‟ size was in the range from 100-150 basepairs. Each PCR-amplified product was sequenced and compared to the correlative sequence of Williams 82 to ensure that the primers amplified the right gene. Primer efficiency was tested by the construction of standard curves using C(T) value (Pfaffl, 2001). Total RNA was extracted from mature leaves and seeds of Williams 82 at three different seed diameter categories including 3-4 mm (harvested at growth stage R3), 6-7mm (harvested at growth stage R5), and 11mm (growth stage R6-R7) using TRIzol reagent (Invitrogen) following the extraction protocol of the producer. DNase-treated total RNA template was reverse transcribed into cDNA and subsequently amplified by PCR in one reaction tube. The RT-PCR reaction was performed using the QuantiTect SYBR Green RT-PCR Kit (Qiagen, Valencia, CA) in 20 µL reactions. The experiment was conducted with technical triplicates as described previously (Dierking and Bilyeu, 2008). The house-keeping gene used in this study was cons7, one of the four genes were characterized as genes with the most stable and consistent expression level across different soybean tissues including leaf and seed (Libault et al., 2008).

RESULTS

For *GPAT*, four putative soybean homologous genes were identified using *Arabidopsis GPAT1-2* sequences, which are the two genes that were shown to have a role to the TAG production in seed (Gidda et al., 2009). They are encoded as Glyma14g03210 (*GPAT1A*), Glyma02g45600 (*GPAT1B*), Glyma02g41660 (*GPAT2A*), and Glyma14g07290 (*GPAT2B*) (Table 5.1 and 5.2). Protein-wise, *GPAT* genes are the largest proteins among three acyltransferase enzymes tested in this study with protein sizes in the range of 461-540 amino acid (aa). However, these proteins are encoded by small genes with simple structures. All of the *GPAT* genes have the size of nearly 2 kilobases (kb) with two exons and one intron. For *GPAT* genes, only seeds with the size of 6-7 mm (corresponding to R5 growth stage) were used for RT-PCR. Although the data in soyseg showed that these genes were not expressed in either leaf or seeds, RT-PCR data showed that *GPAT1A* and *GPAT2B* were specifically expressed in leaf, *GPAT1B* was only expressed in seed and *GPAT2A* gene had a trace expression in both leaf and seed (Figure 5.2). The expression of *GPAT2B* is about nine fold higher than that of *GPAT1A* and about 15 fold higher than those of *GPAT1B* and *GPAT2A*.

For *LPAAT* genes, five putative *LPAAT* genes were identified: Glyma02g31320 (*LPAAT1*), Glyma03g29600 (*LPAAT2*), Glyma10g12560 (*LPAAT3*), Glyma15g03880 (*LPAAT4*), and Glyma19g32420 (*LPAAT5*) (Table 5.1). The LPAAT proteins have about 377-387 aa, encoded by genes with size in the range from 4kb to 13kb (Table 5.2). All genes contain 11 small exons separated by ten large introns. The expression scores in soyseg indicated that *LPAAT4* is only expressed in leaf while other four genes are equally

expressed in both seeds and leaf, and *LPAAT2* is the gene with the highest expression in seeds. The RT-PCR data, however, indicated that *LPAAT4* was expressed specifically in leaf while the other four genes had significant higher expression in seeds than in leaf (Figure 5.3). Among four genes that were expressed in seed, *LPAAT3* but not *LPAAT2* had the highest expression level, which was two-fold higher than those of the other three genes in all three seed diameter categories. Three genes *LPAAT1, LPAAT2,* and *LPAAT5* had the same expression pattern and expression levels. Except for *LPAAT4*, all the *LPAAT* genes had bell-shaped curves of expression level across three different seed sizes from small to big during embryogenesis, with the peaks of these curves were seen at seed sized 6-7mm.

Three homologs for *DGAT1* genes (Glyma13g16560-*DGAT1A*, Glyma17g06120-*DGAT1B*, and Glyma09g07520-*DGAT1C*) and four for *DGAT2* genes (Glyma01g36010-*DGAT2A*, Glyma11g09410-*DGAT2B*, Glyma16g21960-*DGAT2C*, and Glyma09g32790-*DGAT2D*) were found (Table 5.1). DGAT proteins are about 500 aa for DGAT1s and 300aa for DGAT2s (Table 5.2). They are encoded by large genes encompassing approximately 8-12kb. *DGAT1* genes often have 16 exons and 15 introns while *DGAT2* genes have nine exons and eight introns. Similar to *LPAAT* genes, the exons in these DGAT genes are small and separated by large introns.

Based on soyseg database, *DGAT1A* and *DGAT1B* genes have the highest expression scores in different seed development stages than those of *DGAT1C* and all *DGAT2*s. Among *DGAT2* candidate genes, *DGAT2A* and *DGAT2C* have some detectable expression level in seed, while *DGAT2D* is more expressed in leaf than in seed and

DGAT2B did not appear to be expressed in any of the tissues (Table 5.1). The RT-PCR data were in agreement with online database trends for a few *DGAT* genes but not all of them (Figure 5.4). The expression level of *DGAT1A* was the highest among seven *DGAT* genes across three different seed sizes. The expression level of *DGAT1B* was second to the *DGAT1A* and higher than the other genes only at seed size 6-7mm; otherwise, its expression levels was not different from those of the other genes at early and later stages. The expression level of *DGAT1C* was lowest among DGAT1 but equal to that of DGAT2C, the gene with the highest expression level among DGAT2 for seeds sized 6-7 mm. At the smaller or bigger seed size, *DGAT1C*"s expression level was lower than those of *DGAT1A* but similar to those of *DGAT1B*.

Among *DGAT2* genes, *DGAT2A* seems to be seed specific though its expression level is the lowest among the four *DGAT2* genes. The other three genes all had some detectable expression levels in leaf, although the expression levels in leaf were lower than the expression in seeds. *DGAT2C* steadily expressed in seed of all size; its expression level is highest among *DGAT2* genes and equivalent to the expression of *DGAT1C* in seeds at 6-7mm diameter. *DGAT2B* and *DGAT2D* had some expression levels mostly shown in seeds at 3-4mm or 6-7mm, which were higher than the expression of these genes in leaf. Except for *DGAT2B* and *DGAT2C*, the expression levels of all other genes across three seed size categories from small to big size corresponding to growth stages from R3-R7 follow bell-shaped curves similar to those of *LPAAT* genes.

In this study, the expression levels of *GPAT, LPAAT* and *DGAT* gene families were not comparable to each other because no attempt were made to include all

individuals genes in three gene families in one RT-PCR run for a specific seed stage. Moreover, the threshold values (C(t) values) for cons7 showed small variation for each run with each of the gene family. However, the expression levels of *GPAT* genes were estimated to be higher than those of *LPAAT* genes, and the expression level of *DGAT* is the least among three genes family. The expression of *GPAT1B*, the *GPAT* gene with the highest expression in seeds, is about three folds higher than that of *LPAAT3* (the *LPAAT* gene with the highest expression in seed). The expression level of *LPAAT3* is two fold higher than that of *DGAT1A* (the *DGAT* gene with the highest expression in seeds).

DISCUSSION

In this study, we identified multiple homologous genes for each of the three candidate acyltransferase genes, *GPAT*, *LPAT* and *DGAT*. This once again reflects the palaeopolyploid characteristics of soybean genome. The size and complexity of the genome of soybean is the result of two duplication events which occurred at approximately 59 and 13 million years ago, resulting in a highly duplicated genome with nearly 75% of the genes present in multiple copies (Schmutz et al., 2010). Following two rounds of duplication, most of the genes experienced diversification and limited loss, and numerous chromosomes were rearranged. Therefore, it is reasonable that for one gene from *Arabidopsis* there are often two to four corresponding genes of *GPAT*, *LPAT* and *DGAT* in soybean identified. The results of this experiment indicated that all of the homologous genes of *GPAT*, *LPAT* and *DGAT* showed some differential expression either in seed, leaf, or both, which means that they all are actively transcribed in all of the tissues tested. The mechanisms that caused the expression divergence of homologous genes occur are diverse and poorly understood (Chen and Ni, 2006) but epigenetic mechanisms were proposed to be plausible causes (Wolffe and Matzke, 1999). In order to increase the oil content for an oilseed crop, it is often required that the expression level of one candidate gene in the TAG biosynthesis pathway is up-regulated (Courchesne et al., 2009; Frentzen, 1998). Therefore, the findings of this study will provide multiple candidate genes for such studies in soybean. It is recommended that the gene(s) with highest expression level during seed oil accumulation period would be selected for the over-expression experiments. Although several studies on overexpression of genes in the TAG biosynthesis pathway showed that that there is a low correlation between the

composition of seeds and the expression level of transgenes or transgene enzymatic activities (Jako et al., 2001; Maisonneuve et al.; Xu et al., 2008; Zou et al., 1997), usage of genes with highest expression level for an overexpression experiment will more likely to get the phenotype of interest.

The results of this study demonstrated the sensitivity and power of RT-PCR technique which can detect expression level of genes with low expression or homologous genes sharing a high level of similarity. Several transcriptome databases exist for soybean genes; however, one of the disadvantages they have is that for genes that have more than two homologous genes, only data for one of the homologous genes can be found. In this experiment, we designed specific primers to amplify each homologous gene in each gene family of *GPAT*, *LPAAT* and *DGAT*. Therefore, it enables us to distinguish and compare the expression level of each independent homologous gene accurately. Overall, the data for genes that were showed to have high expression levels in public databases such as soyseg or Transcriptome atlas of *Glycine max* were in agreement with the data from our experiment. The difference of expression profile for the tested genes between our results and the public databases was mostly seen in genes that have low expression level in public resources, such as the *GPAT* genes (especially *GPAT2B)*, *LPAAT5* or *DGAT2B*. *GPAT2B* was showed to have a high expression level compared to other *GPAT* genes in leaf although in Soyseg data it is not expressed in either seed or leaf. *LPAAT5* and *DGAT2B* were found to have no detectable expression in seeds in Soyseg or to have higher expression in leaf than seed in Transcriptome atlas while in our experiment they were found to have expression levels in both tissues and the expression levels in seeds

was higher than that in leaf. Because the seed size categories we selected are relative correlated to the time scale of the online public data, with 3-4mm seeds were harvested at about 14 days after flowering (DAF), 6-7 mm seeds were harvested at 25-28 DAF, and 11mm seeds were harvested around 42 DAF, the expression pattern of *LPAAT* genes which formed the bell-shaped curve resembles the expression pattern of these genes in Soyseg. However, for *DGAT* genes, expression data on Soyseg showed that the expression levels of *DGAT1A* and *DGAT1B* genes are low from 10 DAF to 28 DAF and then increase at 35-42 DAF while our data showed a bell-shapred curve expression pattern for these two genes with the peaks of the curves observed at about 25-28 DAF. Although there was a discrepancy with the online database, the expression patterns of *DGAT1* genes in our study is in agreement with RT-PCR data for overall *DGAT1* gene from the study by Li et al. 2010. This demonstrates the sensitivity of RT-PCR for the detection of gene expression level compared to that of microarray data.

The expression level patterns of *DGAT1* and *DGAT2* homologs in our study in general support the expression data reported by Li et al 2010, in which *DGAT1* genes are more expressed in seeds than *DGAT2* genes. In addition, we provided more details of the expression profile of each individual gene within the *DGAT1* and *DGAT2* families at aspecific growth stage during seed development because not all of them are expressed in all development stages. Although the differential expression pattern of homologous genes of *GPAT, LPAAT*, and *DGAT* were identified in this study, it is not conclusive which homologous gene holds the most important role in the TAG production in soybean seed. It was demonstrated in other species such as in *Saccharomyces cerevisiae* yeast that

although the expression level of genes involving galactose utilization pathway was found to be moderately correlated with their protein abundance counterparts in (r = 0.5 to 0.61)(Griffin et al., 2002; Ideker et al., 2001), there were genes that were reported to have mRNA abundance up to five times higher than that of the control gene but their protein abundance was similar to that of the control (Ideker et al., 2001). Although the expression of *DGAT1A* gene was higher than *DGAT1B* in our study, Hilderbrand et al. 2009 showed that the amount of TAG formed when *DGAT1A* gene was expressed in yeast is significantly lower than the TAG level formed when expressing *DGAT1B* (Hildebrand et al., 2009). Therefore, until more evidence about protein abundance and enzyme activities of these acyltransferase genes in soybean seeds are found, we can not underestimate genes that have moderate or low expression levels. All of the genes that are expressed in seeds should be considered to contribute and influence the quantity of the product of the reaction that the corresponding enzymes participate in.

FIGURES

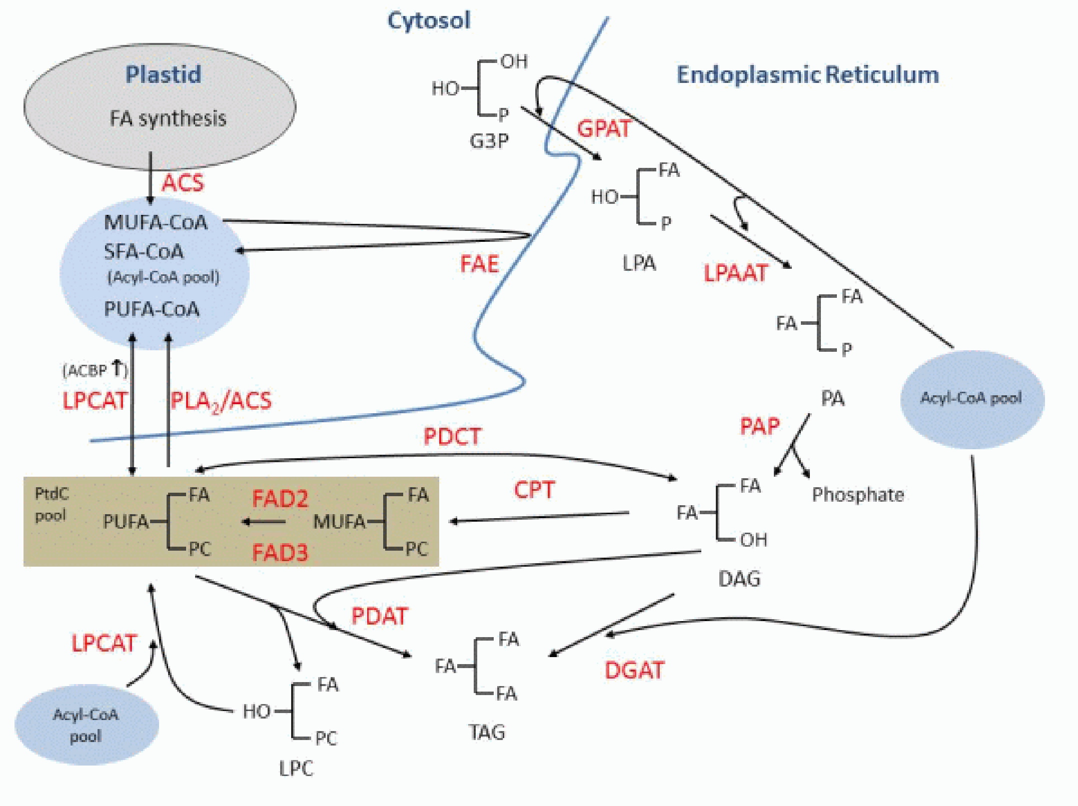


Figure 5.1: Generalized scheme for triacylglycerol (TAG) assembly in developing seeds of oleaginous plants (figure was taken from the website lipid library athttp://lipidlibrary.aocs.org/plantbio/tag_biosynth/index.htm with the permission granted from Dr. Randall Weselake, Dr. Jitou Zou and Dr. David Taylor). Abbreviations: CoA, coenzyme A; CPT,cholinephosphotransferase; DAG, sn-1,2-diacylglycerol; FA, fatty acid; FA-CoA, fatty acyl-coenzyme A; FAD2 and FAD3: fatty acid desaturases 2 and 3, G3P, sn-glycerol-3-phosphate; FAE, fatty acid elongase; GPAT, sn-glycerol-3-phosphate acyltransferase; LPA, lysophosphatidic acid; LPAAT,lysophosphatidic acid acyltransferase; LPC, lysophosphatidylcholine; LPCAT,lysophosphatidylcholine acyltransferase; MUFAs, Monounsaturated fatty acids; PA, phosphatidic acid; PAP, phosphatidic acid phosphatase; PC, phosphatidylcholine; PDAT,phospholipid:diacylglycerol acyltransferase; PLA2, phospholipase A2; PUFA, polyunsaturated fatty acids; SFA, Saturated fatty acids;

Figure 5.2: Relative expression of putative glycerol-3-P acyltransferase (*GPAT*) genes in leaf and seeds at different sizes of cultivar Williams 82. The four putative *GPAT* genes are along the x-axis. Bar heights represent the average values from three replicates expressed relative to the housekeeping gene (cons7) control. Error bars represent plus and minus one standard deviation from the mean.

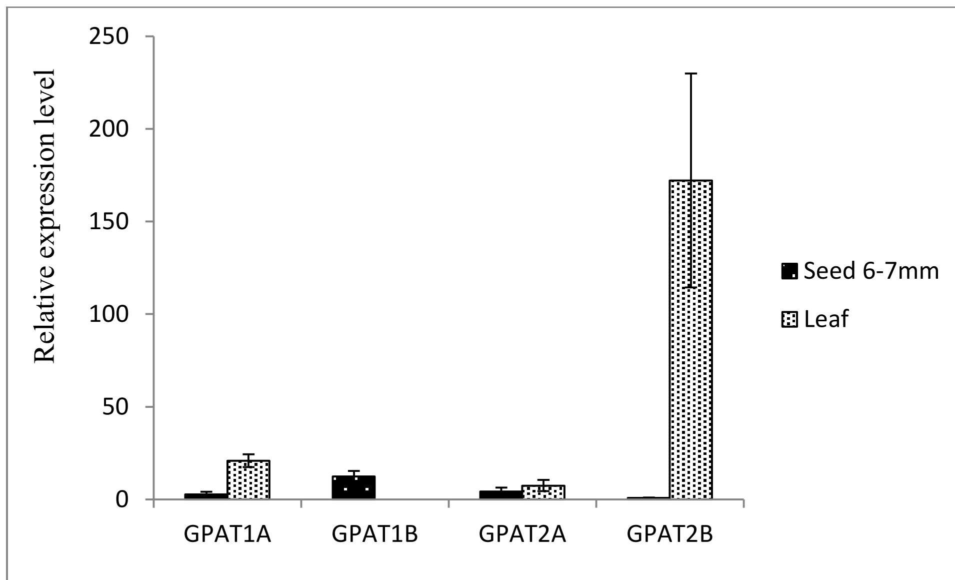


Figure 5.3: Relative expression of putative lysophosphatidic acid acyltransferase (*LPAAT*) genes in leaf and seeds at different sizes of cultivar Williams 82. The five putative *LPAAT* genes are along the x-axis. Bar heights represent the average values from three replicates expressed relative to the housekeeping gene (cons7) control. Error bars represent plus and minus one standard deviation from the mean.

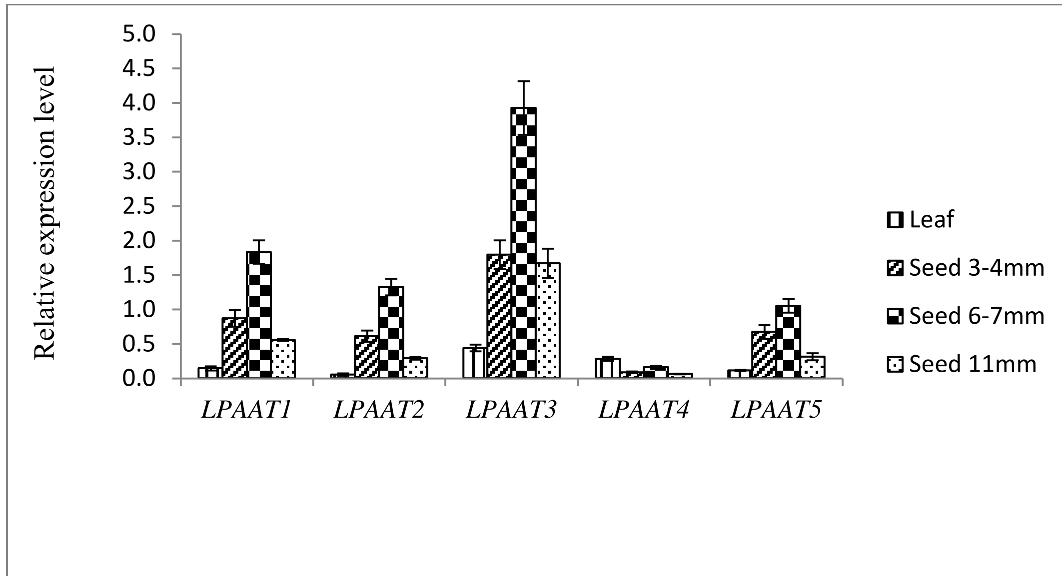
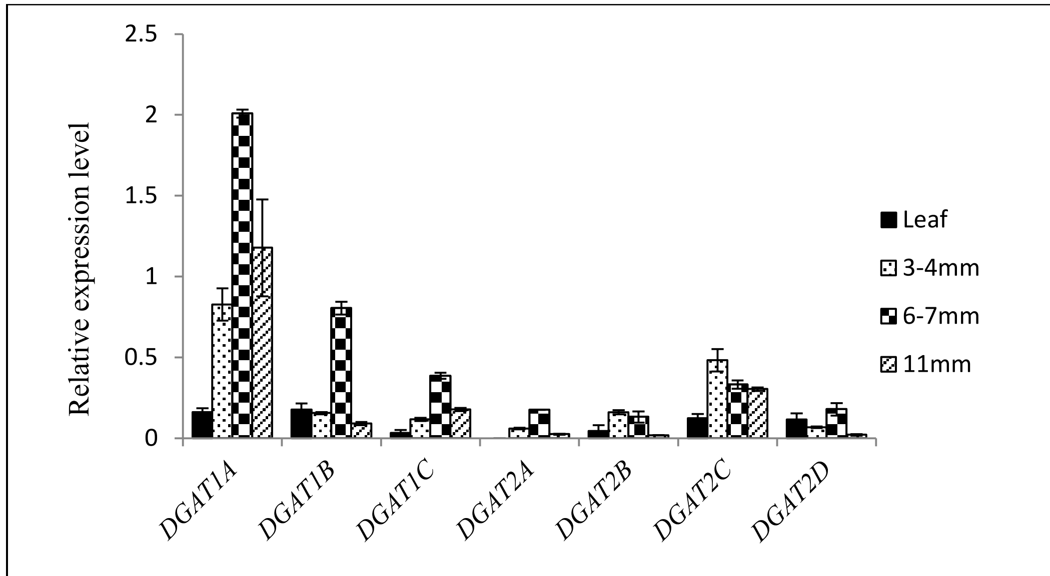


Figure 5.4: Relative expression of putative diacylglycerol acyltransferase (*DGAT*) genes in leaf and seeds at different sizes of cultivar Williams 82. The seven putative *DGAT* genes are along the x-axis. Bar heights represent the average values from three replicates expressed relative to the housekeeping gene (cons7) control. Error bars represent plus and minus one standard deviation from the mean.



TABLES

Table 5.1: Information of homologous gene of *GPAT*, *LPAAT* and *DGAT* genes used in the RT-PCR analysis.

Gene	Gene ID	Gene size (bp)	Protein size (aa)	Exon/Intron
GPAT1A	Glyma14g03210	2,149	540	2/1
GPAT1B	Glyma02g45600	2,151	539	2/1
GPAT2A	Glyma02g41660	2,057	461	2/1
GPAT2B	Glyma14g07290	2,770	512	2/1
LPAAT1	Glyma02g31320	7,820	384	11/10
LPAAT2	Glyma03g29600	7,044	377	11/10
LPAAT3	Glyma10g12560	13,387	383	11/10
LPAAT4	Glyma15g03880	3,877	381	11/10
LPAAT5	Glyma19g32420	10,544	387	11/10
DGAT1A	Glyma13g16560	8,171	498	16/15
DGAT1B	Glyma17g06120	8,114	504	16/15
DGAT1C	Glyma09g07520	5,477	394	9/8
DGAT2A	Glyma01g36010	5,066	329	9/8
DGAT2B	Glyma11g09410	2,493	318	9/8
DGAT2C	Glyma16g21960	12,682	323	9/8
DGAT2D	Glyma09g32790	6,286	307	9/8

Table 5.2 Expression levels of all soybean homologous genes of *GPAT*, *LPAAT* and *DGAT* genes in leaf and different stages of seed development obtained from microarray data in soybase database.

		leaf	1cm pod	seed 10DAF	seed 14DAF	seed 21DAF	seed 25DAF	seed 28DAF	seed 35DAF	seed 42DAF
GPAT										
Glyma14g03210	*GPAT*1A	0	0	0	1	0	0	0	0	0
Glyma02g45600	*GPAT*1B	0	0	0	0	0	0	0	0	0
Glyma02g41660	*GPAT*2A	0	0	0	0	0	0	0	0	0
Glyma14g07290	*GPAT*2B	0	0	0	0	0	0	0	0	0
LPAAT										
Glyma02g31320	*LPAAT*1	5	3	3	3	3	6	3	5	3
Glyma03g29600	*LPAAT*2	3	7	4	5	7	4	2	3	2
Glyma10g12560	*LPAAT*3	8	5	2	4	5	4	1	4	2
Glyma15g03880	*LPAAT*4	5	0	0	1	1	1	0	0	0
Glyma19g32420	*LPAAT*5	3	5	2	4	5	4	1	2	1
DGAT1										
Glyma13g16560	DGAT1A	2	2	2	1	1	5	3	11	13
Glyma17g06120	DGAT1B	1	3	0	1	0	1	1	2	2
Glyma09g07520	DGAT1C	1	0	0	1	1	6	3	8	8
DGAT2										
Glyma01g36010	DGAT2A	2	3	3	2	2	3	2	4	2
Glyma11g09410	DGAT2B	0	2	0	1	0	0	0	1	0
Glyma16g21960	DGAT2C	4	5	2	4	4	5	2	5	4
Glyma09g32790	DGAT2D	12	6	2	2	2	2	1	2	1

LITERATURE CITED

Bilyeu K., Palavalli L., Sleper D., Beuselinck P. (2005) Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science 45:1830-1836.

Bilyeu K.D., Gillman J.D., LeRoy A.R. (2011) Novel *FAD3* mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Science 51:259-264

Bouvier-Navé P., Benveniste P., Oelkers P., Sturley S.L., Schaller H. (2000) Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase. European Journal of Biochemistry 267:85-96.

Bruner A., Jung S., Abbott A., Powell G. (2001) The naturally occurring high oleate oil character in some peanut varieties results from reduced oleoyl-PC desaturase activity from mutation of Aspartate 150 to Asparagine. Crop Science 41:522 - 526.

Cases S., Smith S.J., Zheng Y.-W., Myers H.M., Lear S.R., Sande E., Novak S., Collins C., Welch C.B., Lusis A.J., Erickson S.K., Farese R.V. (1998) Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of the National Academy of Sciences of the United States of America 95:13018-13023.

Chi X., Yang Q., Lu Y., Wang J., Zhang Q., Pan L., Chen M., He Y., Yu S. (2011) Genome-Wide Analysis of Fatty Acid Desaturases in Soybean (*Glycine max*). Plant Molecular Biology Reporter:1-15.

Courchesne N.M.D., Parisien A., Wang B., Lan C.Q. (2009) Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. Journal of Biotechnology 141:31-41.

Dierking E., Bilyeu K. (2008) Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome 1:135 - 145.

Frentzen M. (1998) Acyltransferases from basic science to modified seed oils. Lipid / Fett 100:161-166.

Gidda S.K., Shockey J.M., Rothstein S.J., Dyer J.M., Mullen R.T. (2009) *Arabidopsis* thaliana *GPAT*8 and *GPAT*9 are localized to the ER and possess distinct ER retrieval signals: Functional divergence of the dilysine ER retrieval motif in plant cells. Plant Physiology and Biochemistry 47:867-879.

Griffin T.J., Gygi S.P., Ideker T., Rist B., Eng J., Hood L., Aebersold R. (2002) Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in Saccharomyces cerevisiae. Molecular & Cellular Proteomics 1:323-333.

He X., Turner C., Chen G.Q., Lin J.-T., McKeon T.A. (2004) Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean Lipid 39:311-319.

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

Hildebrand D., Li R., Yu K., Hatanaka T. (2009) Accumulation of expoxy fatty acids in plant oils, in: C. T. Hou and J. F. Shaw (Eds.), Biocatalysis and Agriculture Biotechnology, CRC Press, Boca Raton, FL. pp. 33-41.

Hildebrand D.F., Li R., Hatanaka T. (2008) Genomics of soybean oil traits, in: G. Stacey (Ed.), Genetics and genomics of soybean, Springer New York, New York. pp. 185-209.

Ideker T., Thorsson V., Ranish J.A., Christmas R., Buhler J., Eng J.K., Bumgarner R., Goodlett D.R., Aebersold R., Hood L. (2001) Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. Science 292:929-934.

Jain R.K., Coffey M., Lai K., Kumar A., MacKenzie S.L. (2000) Enhancement of seed oil content by expression of glycerol-3-phosphate acyltransferase genes. Biochem. Soc. Trans. 28:958-961.

Jako C., Kumar A., Wei Y., Zou J., Barton D.L., Giblin E.M., Covello P.S., Taylor D.C. (2001) Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol. 126:861-874.

Kim M., Go Y., Lee S., Kim Y., Shin J., Min M., Hwang I., Suh M. (2010) Seed-expressed casein kinase I acts as a positive regulator of the Se*FAD2* promoter via phosphorylation of the SebHLH transcription factor. Plant Molecular Biology 73:425-437.

Kim M., Kim J.-K., Shin J., Suh M. (2007) The SebHLH transcription factor mediates trans-activation of the Se*FAD2* gene promoter through binding to E- and G-box elements. Plant Molecular Biology 64:453-466.

Kroon J.T.M., Wei W., Simon W.J., Slabas A.R. (2006) Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. Phytochemistry 67:2541-2549.

Lardizabal K., Effertz R., Levering C., Mai J., Pedroso M.C., Jury T., Aasen E., Gruys K., Bennett K. (2008) Expression of *Umbelopsis ramanniana* DGAT2A in seed increases oil in soybean. Plant Physiol. 148:89-96.

Lardizabal K.D., Mai J.T., Wagner N.W., Wyrick A., Voelker T., Hawkins D.J. (2001) DGAT2 Is a new diacylglycerol acyltransferase gene family. Purification, cloning and expression in insect cells of two polypeptides from Mortierella ramanniana with diacylglycerol acyltransferase activity. J. Biol. Chem. 276:38862-38869.

Libault M., Thibivilliers S., Bilgin D.D., Radwan O., Benitez M., Clough S.J., Stacey G. (2008) Identification of Four Soybean Reference Genes for Gene Expression Normalization. Plant Gen. 1:44-54.

Maisonneuve S., Bessoule J.-J., Lessire R., Delseny M., Roscoe T.J. Expression of Rapeseed Microsomal Lysophosphatidic Acid Acyltransferase Isozymes Enhances Seed Oil Content in *Arabidopsis*. Plant Physiol. 152:670-684.

Milcamps A., Tumaney A.W., Paddock T., Pan D.A., Ohlrogge J., Pollard M. (2005) Isolation of a gene encoding a 1,2-Diacylglycerol-sn-acetyl-CoA Acetyltransferase fromd eveloping seeds of Euonymus alatus. J. Biol. Chem. 280:5370-5377.

Mounts T.L., Warner K., Usta G.R., Neffa W.E., Wilson R.F. (1994) Low linolenic acid soybean oils-Alternatives to frying oils. Journal of the American Oil Chemists' Society 71:495-499.

Mozaffarian D., Katan M.B., Ascherio A., Stampfer M.J., Willett W.C. (2006) Trans fatty acids and cardiovascular disease. New England Journal of Medicine 354:1601-1613.

Nykiforuk C.L., Furukawa-Stoffer T.L., Huff P.W., Sarna M., Laroche A., Moloney M.M., Weselake R.J. (2002) Characterization of cDNAs encoding diacylglycerol acyltransferase from cultures of Brassica napus and sucrose-mediated induction of enzyme biosynthesis. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1580:95-109.

Oakes J., Brackenridge D., Colletti R., Daley M., Hawkins D.J., Xiong H., Mai J., Screen S.E., Val D., Lardizabal K., Gruys K., Deikman J. (2011) Expression of Fungal diacylglycerol acyltransferase2 Genes to Increase Kernel Oil in Maize. Plant Physiology 155:1146-1157.

Pfaffl M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucl. Acids Res. 29:e45.

Raneses A., Glaser L., Price J., Duffield J. (1999) Potential biodiesel markets and their economic effects on the agricultural sector of the United States. Industrial Crops and Products 9:151 - 162.

Rao S., Hildebrand D. (2009) Changes in Oil Content of Transgenic Soybeans Expressing the Yeast SLC1 Gene. Lipids 44:945-951.

Reinprecht Y., Luk-Labey S.Y., Larsen J., Poysa V.W., Yu K., Rajcan I., Ablett G.R., Pauls K.P. (2009) Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. Plant Breeding 128:253-258.

Salt F.G. (1957) Vegetable Oils' Use Keeps Growing. Chemical & Engineering News 35:92-93.

Scherder C.W., Fehr W.R. (2008) Agronomic and seed characteristics of soybean lines with increased oleate content. Crop Sci. 48:1755-1758.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.-C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Shockey J.M., Gidda S.K., Chapital D.C., Kuan J.-C., Dhanoa P.K., Bland J.M., Rothstein S.J., Mullen R.T., Dyer J.M. (2006) Tung Tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 18:2294-2313.

Stahl U., Carlsson A.S., Lenman M., Dahlqvist A., Huang B., Banas W., Banas A., Stymne S. (2004) Cloning and Functional Characterization of a Phospholipid:Diacylglycerol Acyltransferase from *Arabidopsis*. Plant Physiol. 135:1324-1335.

Wang H.-W., Zhang J.-S., Gai J.-Y., Chen S.-Y. (2006) Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean. Theoretical Applied Genetics 112.

Weselake R.J. (2005) Storage lipids, in: D. J. Murphy (Ed.), Plant lipids: biology, utilization and manipulation, Blackwell Publishing, Oxford. pp. 162-225.

Weselake R.J., Shah S., Tang M., Quant P.A., Snyder C.L., Furukawa-Stoffer T.L., Zhu W., Taylor D.C., Zou J., Kumar A., Hall L., Laroche A., Rakow G., Raney P., Moloney M.M., Harwood J.L. (2008) Metabolic control analysis is helpful for

informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. J. Exp. Bot. 206.

Xu J., Francis T., Mietkiewska E., Giblin E.M., Barton D.L., Zhang Y., Zhang M., Taylor D.C. (2008) Cloning and characterization of an acyl-CoA-dependent diacylglycerol acyltransferase 1 (*DGAT1*) gene from Tropaeolum majus, and a study of the functional motifs of the *DGAT* protein using site-directed mutagenesis to modify enzyme activity and oil content. Plant Biotechnology Journal 6:799-818.

Xu K., Yang Y., Li X. (2010) Ectopic expression of Crambe abyssinica lysophosphatidic acid acyltransferase in transgenic rapeseed increases its oil content. African Journal of Biotechnology 9:3904-3910.

Yadav N.S., Wierzbicki A., Aegerter M., Caster C.S., Perez-Grau L., Kinney A.J., Hitz W.D., Booth Jr. J.R., Schweiger B., Stecca K.L., Allen S.M., Blackwell M., Reiter R.S., Carlson T.J., Russell S.H., Feldmann K.A., Pierce J., Browse J. (1993) Cloning of higher plant [omega]-3 fatty acid desaturases. Plant Physiol. 103:467-476.

Yen C.-L.E., Stone S.J., Koliwad S., Harris C., Farese R.V., Jr. (2008) Thematic Review Series: Glycerolipids. *DGAT* enzymes and triacylglycerol biosynthesis. J. Lipid Res. 49:2283-2301

Zheng P., Allen W.B., Roesler K., Williams M.E., Zhang S., Li J., Glassman K., Ranch J., Nubel D., Solawetz W., Bhattramakki D., Llaca V., Deschamps S., Zhong G.-Y., Tarczynski M.C., Shen B. (2008) A phenylalanine in *DGAT* is a key determinant of oil content and composition in maize. Nat Genet 40:367-372.

Zou J., Katavic V., Giblin E.M., Barton D.L., MacKenzie S.L., Keller W.A., Hu X., Taylor D.C. (1997) Modification of Seed Oil Content and Acyl Composition in the Brassicaceae by Expression of a Yeast sn-2 Acyltransferase Gene. Plant Cell 9:909-923.

VITA

Anh Tung Pham was born January 30, 1983 in Hue and grew up in Hanoi, the capital of Vietnam. She graduated from Yen Hoa High School in 2001 and study biotechnology at Hanoi University of Science. Anh earned her bachelors of science in biotechnology in 2005. She was awarded the Vietnam Education Fellowship to study for a M.S. degree in plant pathology at the Universit of Illinois, Urbana-Champaign. After receiving the M.S. degree, she moved to Columbia, Missouri to complete her PhD at the University of Missouri-Columbia in the Division of Plant Science under the advisement of Dr. Kristin Bilyeu. She is married to Giang Truong Nguyen and has one son, Long Thang Nguyen.