GENETICS OF SOYBEAN SEED LIPOXYGENASES
AND LINOLENIC ACID CONTENT IN SEEDS OF THE SOYBEAN WILD ANCESTOR

A Dissertation presented to the Faculty of the Graduate School at the University of Missouri-Columbia

In Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

by

JULIÁN MARIO LENIS

Dr. Kristin D. Bilyeu and Dr. J. Grover Shannon, Dissertation Supervisors

JULY 2011
The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

GENETICS OF SOYBEAN SEED LIPOXYGENASES
AND LINOLENIC ACID CONTENT IN SEEDS OF THE SOYBEAN WILD ANCESTOR

Presented by Julián Mario Lenis,
a candidate for the degree of Doctor of Philosophy,
and hereby certify that, in their opinion, it is worthy of acceptance.

__________________________________________
Dr. Kristin D. Bilyeu

__________________________________________
Dr. J. Grover Shannon

__________________________________________
Dr. Michael D. McMullen

__________________________________________
Dr. Emmanuel Liscum
Dedicated to

Mi Florcita Masha

My children Uliana and Sasha (and those that might come in the future)

My parents Julián and Gely

My siblings and their families

Grandma Gringa

My friends from the U.S. (my third family)

My friends from Argentina
ACKNOWLEDGEMENTS

I would like to thank my dissertation supervisors, Dr. Kristin Bilyeu and Dr. Grover Shannon, for their constant support and useful advice. Kristin, thank you for helping me better understand how exciting molecular biology can be and for teaching me how to generate basic science knowledge and use it to develop applied tools. With you, the pretended distinction between basic and applied science disappears…or rather transforms into a bridge that connects both. Grover, thanks for believing in me even before we met. I also want to thank you for helping me realize that failure does not represent the “end of the world” and that success in research is many times the result of the lessons learned after failing. I thank you the most for reminding me that faith in the Divine Providence never disappoints.

I would also like to extend my appreciation to the other members of my dissertation committee: Dr. Mike McMullen and Dr. Mannie Liscum. Thanks for devoting part of your time and research talent to my education and training as a scientist.

I also want to offer a big thank you to Dr. Mark Ellersieck, who has always been an invaluable help to carry out the analysis and interpretation of those results that require the use of Statistics.

Also, I want to express my gratitude to the Bilyeu lab members (The Green Bandana Team): Paul Little, thanks for your help in getting everything ready for the field as well as organizing the space for harvested plants and threshing. Also, thanks for sharing your creativity and good suggestions in the planning of our research activities
(and the cowboy movies!). Jason Gillman, I am and will always be extremely grateful for your guidance and role as an “older brother” to me. You were Kristin’s right hand in the endeavor of adapting my field breeder/agronomist mind setting to that of a molecular biologist/geneticist. Thanks for all the times you let me interrupt your own research to ask you the most basic questions and always give thoughtful and kind answers. Katy Hagely, I cannot find words that are meaningful enough to tell you how much I value all that you have done for me. Thanks for enduring the many hours under the unforgiving Missouri summer sun tying up the wild plants from my populations. Most importantly, I thank you for your spontaneous and fresh approach to life and, of course, your friendship. Anh Pham, I want to thank you for the wonderful example you gave us every day as a bright, hardworking, young mother and wife. It has been an inspiring experience to work with you. Ashley Tetlow, thanks for your assistance in the various lab activities, from DNA extraction to reagent preparation, and for modeling a young professional who aims to make good science and bring improvements to the world community.

In addition, I extend special thanks to Kerry Clark and Christi Cole for their assistance during planting and throughout the crop season. Christi, I am very grateful for the numerous seed samples you analyzed for me.

Finally, I thank Estación Experimental Agroindustrial “Obispo Colombes” (EEAOC) and particularly my fellow researchers Mario Devani and Fernando Ledesma, for their permanent effort and support to make all this happen.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ........................................................................................................... ii

**LIST OF ILLUSTRATIONS** ...................................................................................................... xi

**LIST OF TABLES** ...................................................................................................................... xiii

Chapter

1. Literature Review .................................................................................................................. 1
   - Origin and dissemination of soybean .................................................................................... 2
   - Economic importance of soybean ....................................................................................... 3
   - Soybean products and applications .................................................................................... 3
   - Soyfoods ............................................................................................................................... 5
   - Soybean seed composition ................................................................................................. 11
   - Lipoxygenases .................................................................................................................... 14
   - Occurrence of lipoxygenases in plants ................................................................................ 15
   - Role of lipoxygenases ........................................................................................................ 16
   - Importance and effect of lipoxygenases on food quality .................................................... 17
   - Lipoxygenases in soybean .................................................................................................. 18
   - Genetics of soybean seed lipoxygenases ......................................................................... 19
   - Development of a triple lipoxygenase mutant genotype in soybean ............................... 21
   - Molecular basis of the lipoxygenase 2-null phenotype in PI 86023 ............................ 22
   - Screening for seed lipoxygenases ...................................................................................... 22
   - Soybean oil and essential fatty acids .............................................................................. 24
   - Essential fatty acids and health ....................................................................................... 25
Sources of ω-3 fatty acids ................................................................. 26
Soybean oil, 18:3 content and the nutritional quality vs. functionality
dilemma .......................................................................................... 27
Potential uses of soybean oil enriched in ω-3 fatty acids ................ 28
Fatty acid and Triacylglycerol synthesis ........................................... 29
Fatty Acid Desaturase (FAD) genes in plants .................................. 30
Microsomal FAD genes in the complex genome of soybean ............ 32
Contribution of FAD3 genes to seed 18:3 content ........................ 33
Genetic modification of fatty acid profile in soybean oil ................ 34
Acyltransferase genes in plants ....................................................... 37
Cholinephosphotransferase genes in plants ..................................... 38
Role of acyltransferase and cholinephosphotransferase genes in oil
quality .............................................................................................. 38
Soybean homologs of AtDGAT1, AtPDAT1 and AtPDCT genes .... 42
Associations among some seed quality traits in soybean ................ 43
Associations among some seed quality traits in G. max x G. soja
populations ..................................................................................... 45
Soybean seed fatty acid profile and environment ............................ 47
Soybean seed fatty acid profile and maturity groups ...................... 49
Winter nurseries ............................................................................. 50
Soybean germplasm collection ........................................................ 52
Breeding populations ...................................................................... 53
Linolenic acid content in the soybean wild ancestor ....................... 54
Figures ............................................................................................. 56
Tables ............................................................................................. 59
2. Soybean Seed Lipoxygenase Genes: Molecular Characterization and Development of Molecular Marker Assays

Introduction

Materials and Methods

DNA isolation and PCR for sequencing Lox1 and Lox3
Sequence analysis
Plant material and development of segregating populations
Sampling DNA and seed harvest
Lox1 genotyping assay
Lox2 and Lox3 genotyping assays
Phenotyping assays

Results

Genomic arrangement of soybean seed lipoxygenase genes
Genetic mutations in Lox1 and Lox3 genes
Development of molecular marker assays
Genetic recombination within Lox1 gene locus in Jinpumkong 2
Association analysis for seed lipoxygenase phenotype and Lox1 and Lox3 genotypes

Discussion

Conclusions
3. Genetics of High Linolenic Acid in Seed Oil of *Glycine soja*: Role of ω-6 and ω-3 Desaturase Genes ...................................................................................................................... 121

Introduction ...................................................................................................................................... 122

Materials and Methods ...................................................................................................................... 131
  Plant material and development of segregating populations ......................................................... 131
  Sampling DNA and determinations on seeds ................................................................................. 132
  Candidate genes .............................................................................................................................. 133
  DNA isolation and PCR for sequencing candidate genes ............................................................. 134
  Sequence analysis .......................................................................................................................... 135
  Genotyping assays ........................................................................................................................ 135
  Expression analysis of candidate genes ....................................................................................... 139

Results ................................................................................................................................................. 141
  Polymorphisms in candidate genes of *Glycine soja* line PI 366121........................................ 141
  Association between high 18:3 content in seeds and genotype at the candidate genes ............ 141
  Gene expression studies .............................................................................................................. 143
  Segregation study ......................................................................................................................... 144

Discussion ......................................................................................................................................... 146
Conclusions ........................................................................................................... 149

Figures .................................................................................................................. 152

Tables .................................................................................................................... 156

Bibliography .......................................................................................................... 165

4. Genetics of High Linolenic Acid in Seed Oil of *Glycine soja*: Role of
   Acyltransferase and Cholinephosphotransferase Genes................................. 170

Introduction ............................................................................................................ 171

Materials and Methods .......................................................................................... 182
   Plant material and development of segregating populations ....................... 182
   Sampling DNA and determinations on seeds ................................................. 182
   Candidate genes ............................................................................................... 183
   DNA isolation and PCR for sequencing candidate genes ............................ 184
   Sequence analysis ............................................................................................ 184
   Genotyping assays ............................................................................................ 185
   Expression analysis of candidate genes ........................................................ 186

Results .................................................................................................................... 188
   Polymorphisms in candidate genes of *Glycine soja* line PI 366121 ............... 188
   Association between high 18:3 content in seeds and genotype at the
   candidate genes ................................................................................................. 189
   Gene expression studies ................................................................................. 189

Discussion ............................................................................................................... 191
Figures .........................................................................................................................248

Tables ............................................................................................................................254

Bibliography ..................................................................................................................262

6. Conclusions and Future Directions .........................................................................266

Bibliography ..................................................................................................................273

VITA .................................................................................................................................275
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Comparison of the fatty acid composition of various plant-based oils typically used for food</td>
<td>57</td>
</tr>
<tr>
<td>1.2</td>
<td>Fatty acid (FA) and triacylglycerol (TAG) synthesis in oilseed plants developing seeds</td>
<td>58</td>
</tr>
<tr>
<td>2.1</td>
<td>Genomic arrangement of soybean seed lipoxygenase genes</td>
<td>109</td>
</tr>
<tr>
<td>2.2</td>
<td>\textit{Lox1} gene structure and polymorphisms observed in \textit{lox1} mutant alleles</td>
<td>110</td>
</tr>
<tr>
<td>2.3</td>
<td>\textit{Lox3} gene structure and polymorphisms observed in \textit{lox3} mutant allele</td>
<td>111</td>
</tr>
<tr>
<td>2.4</td>
<td>Genotyping results using molecular marker assays for soybean seed lipoxygenase genes</td>
<td>112</td>
</tr>
<tr>
<td>2.5</td>
<td>Comparison of Jinpumkong 2, PI 86023 and PI 408251 alleles at the \textit{Lox1} locus</td>
<td>113</td>
</tr>
<tr>
<td>3.1</td>
<td>Fatty acid (FA) and triacylglycerol (TAG) synthesis in oilseed plants developing seeds</td>
<td>153</td>
</tr>
<tr>
<td>3.2</td>
<td>Phenotype and genotype of F\textsubscript{2} plants segregating for \textit{FAD3A}, \textit{FAD3B} and \textit{FAD3C}</td>
<td>154</td>
</tr>
<tr>
<td>3.3</td>
<td>Relative expression of \textit{FAD3A} (3A), \textit{FAD3B} (3B) and \textit{FAD3C} (3C) genes in seeds of Williams 82 (Gm) and PI 366121 (Gs) at 25, 30, 35 and 40 days after flowering (DAF)</td>
<td>155</td>
</tr>
<tr>
<td>4.1</td>
<td>Fatty acid (FA) and triacylglycerol (TAG) synthesis in oilseed plants developing seeds</td>
<td>200</td>
</tr>
<tr>
<td>4.2</td>
<td>Relative expression of \textit{DGAT1A}, \textit{DGAT1B}, \textit{PDAT7} and \textit{PDAT13} genes in seeds of Williams 82 (Gm) and PI 366121 (Gs) at 25 and 40 days after flowering (DAF)</td>
<td>201</td>
</tr>
</tbody>
</table>
5.1. Frequency distribution of RILs for seed dry weight in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) in an F₅ population resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja)........249

5.2. Frequency distribution of RILs for relative seed linolenic acid content (18:3) in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) in an F₅ population resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja) .................................................................250

5.3. Frequency distribution for relative seed linolenic acid content (18:3) in a population of F₂ individuals resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja).................................................................251

5.4. Frequency distribution of RILs for relative seed oleic acid content (18:1) in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) in an F₅ population resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja).................................................................252

5.5. Frequency distribution of RILs for seed oil content in Upala, Costa Rica (CR) in an F₅ population resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja) .................................................................253
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of <em>Glycine max</em> and <em>Glycine soja</em> accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean</td>
<td>60</td>
</tr>
<tr>
<td>2.1. Primers used for amplification and sequencing of <em>Lox1</em> and <em>Lox3</em> genes</td>
<td>115</td>
</tr>
<tr>
<td>2.2. Listing of mutant alleles at the three soybean seed <em>Lox</em> genes</td>
<td>116</td>
</tr>
<tr>
<td>2.3. Colorimetric determination of presence (+) or absence (-) of Lox1 and Lox3 activity in progeny seeds derived from genotyped parental plants</td>
<td>117</td>
</tr>
<tr>
<td>3.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of <em>Glycine max</em> and <em>Glycine soja</em> accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean</td>
<td>157</td>
</tr>
<tr>
<td>3.2. Primers used for amplification and sequencing of <em>FAD2-1</em> and <em>FAD3</em> genes</td>
<td>158</td>
</tr>
<tr>
<td>3.3. Primers and SimpleProbes (and their final concentration in the PCR mix) used to track Williams 82 and PI 366121 alleles of the <em>FAD3A</em> gene (7% lin population) and B1-52 and PI 366121 alleles of <em>FAD2-1B</em>, <em>FAD3A</em>, <em>FAD3B</em> and <em>FAD3C</em> genes (1% lin population)</td>
<td>161</td>
</tr>
<tr>
<td>3.4. Primers used in qRT-PCR experiments to determine the expression of <em>FAD3A</em>, <em>FAD3B</em> and <em>FAD3C</em> genes of <em>Glycine max</em> variety Williams 82 and <em>Glycine soja</em> accession PI 366121 relative to the expression of the reference gene <em>cons7</em></td>
<td>162</td>
</tr>
<tr>
<td>3.5. Polymorphisms in coding and non coding sequence of <em>FAD2-1</em> and <em>FAD3</em> genes between <em>Glycine soja</em> line PI 366121 and reference <em>Glycine max</em> cultivar Williams 82</td>
<td>163</td>
</tr>
<tr>
<td>3.6. Mean differences in seed 18:3 level in F2 progeny of B1-52 x PI 366121 (1% lin) cross. Effect of substitution of PI 366121 and B1-52 alleles of <em>FAD3A</em>, <em>FAD3B</em> and <em>FAD3C</em> genes on 18:3 content</td>
<td>164</td>
</tr>
</tbody>
</table>
4.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of *Glycine max* and *Glycine soja* accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean .................................................................203

4.2. Primers used for amplification and sequencing of *DGAT1A*, *DGAT1B*, *PDAT7*, *PDAT13*, *PDCT7* and *PDCT8* genes ..................................................................................204

4.3. Primers used for amplification and sequencing of the transferase genes that were polymorphic between *Glycine max* cultivar Williams 82 and *Glycine soja* accession PI 366121 .................................................................................209

4.4. Primers used in qRT-PCR experiments to determine the expression of *DGAT1A*, *DGAT1B*, *PDAT7* and *PDAT13* genes of *Glycine max* variety Williams 82 and *Glycine soja* accession PI 366121 relative to the expression of the reference gene *cons7* ................................................................................210

4.5. Polymorphisms in coding and non coding sequence of *DGAT1A*, *DGAT1B*, *PDAT7*, *PDAT13*, *PDCT7* and *PDCT8* genes between *Glycine soja* line PI 366121 and reference *Glycine max* cultivar Williams 82 ..........................................................................211

5.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of *Glycine max* and *Glycine soja* accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean .................................................................255

5.2. Fatty acid composition of the ten individuals that accumulated the highest 18:3 contents in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) ..................................................................................256

5.3. Fatty acid composition of the ten individuals that accumulated the lowest 18:3 contents in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) ..................................................................................257

5.4. Regression coefficients (b) and $R^2$ values (%) for seed quality traits of RILs in an $F_5$ population resulting from the cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*) grown in Upala, Costa Rica ........................................258
5.5. Regression coefficients (b) and $R^2$ values (%) for seed quality traits of RILs in an $F_5$ population resulting from the cross Williams 82 ($Glycine max$) x PI 366121 ($Glycine soja$) grown in Columbia, Missouri ..........................259

5.6. Regression coefficients (b) and $R^2$ values (%) for seed quality traits of sibling $F_4$ RILs’ progeny grown in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) ........................................................................................................260

5.7. Parent-offspring ($F_{2,2:3}$) narrow-sense heritability ($h^2$) and $R^2$ values (%) for seed oil 18:3 and 18:1 contents in the interspecific cross Williams 82 ($Glycine max$) x PI 366121 ($Glycine soja$) population.................................261
Chapter 1

Literature Review
Origin and dissemination of soybean

Evidence suggests that soybean \textit{[Glycine max (L.) Merr.]} was domesticated in the Northeastern region of China between the years 1700 and 1100 B.C. (Hymowitz 1970; Hymowitz and Newell 1980). From about the first century A.D. to the 17\textsuperscript{th} century, soybeans were introduced into several countries such as Japan, Indonesia, Vietnam, Thailand, Nepal, India, Philippines, Myanmar and Malaysia. Landraces eventually developed in these regions, establishing a secondary gene center (Hymowitz 2008). Soybean probably reached Europe in the 18\textsuperscript{th} century, and it was described by Linnaeus in 1737. Samuel Bowen is credited with the introduction of soybean into North America, where soybeans were grown for the first time in 1765 (Hymowitz and Harlan 1983). At the end of the 19\textsuperscript{th} century, soybeans were grown at almost every agricultural station in the U.S. and tested for use in pastures as hay and silage (Hymowitz 2004). Most of the early U.S. soybeans were used as a forage crop rather than harvested for seed. The disruption of trade routes during World War II made it very difficult to satisfy the demand of edible fats and oils in the U.S since \(\sim 40\%\) was imported. As the U.S. looked for alternatives to these imports, there was a rapid increase of the area cultivated with soybean, which was now being grown for the oil of its seeds (Gibson and Benson 2005). By 1950, nearly 100\% of the U.S. soybean crop was grown for seed (Wilson 2008). In the early '50s soybean meal became available as a low-cost, high protein feed ingredient, triggering an explosion in U.S. livestock and poultry production (SoyStats 2011).
Economic importance of soybean

Currently, soybean is the leading oilseed crop produced and consumed in the world, representing 58% and 43% of oilseed production and consumption, respectively (SoyStats 2011; Wilcox 2004). In addition, soybean provides 69% of the protein meal consumed worldwide (SoyStats 2011).

About 50 countries in the world produce soybean (Wilcox 2004). However, over 87% of the 258.4 million metric Tons (MMT) harvested in 2010 were produced in only four countries. The U.S. was the largest producer with 90.6 MMT (35% of the global production), followed by Brazil with 70.0 MMT (27%), Argentina with 49.5 MMT (19%) and China with 15.2 MMT (6%) (USDA-FAS 2011). In the U.S., soybean is the second largest crop in cash sales after corn and the number one value crop export (SoyStats 2011).

Soybean products and applications

The demand for soybean remains strong and continues to grow since it is used as an ingredient in the formulation of a whole host of food, feed and industrial products. It is the functional utility of the soybean seed components in a wide array of products that
keeps this demand high and increasing (Wilson 2008). Typically, commodity soybean seeds contain 40% protein, 20% oil, 35% carbohydrates and 5% ash (Liu 1999).

About 85% of the world’s soybeans are processed into soybean meal and oil. Approximately 98% of the soybean meal that is crushed is further processed into animal feed with the balance used to make soy flour and proteins for food use. Of the oil fraction, 95% is consumed as edible oil, and the rest is used for industrial products. Also, approximately 10% of soybeans are used directly as human food, animal feed and seeds for the next planting season (Goldsmith 2008; Soyatech 2011).

Soybean meal provides a high-protein feed ingredient that triggered an explosion in U.S. livestock production. Soybean meal is used as a main component in diets of poultry, swine, beef, dairy and pets. Also, due to its good amino acid profile, the aquaculture industry is making use of soybean protein concentrate as an alternative to fishmeal (Naylor et al. 2009; SoyStats 2011; Wilson 2008).

Soybean oil is a byproduct of protein feed production. Human health concerns about cholesterol and heart diseases made food manufacturers shift away from animal-derived fat or oil first and saturated fats from tropical oils later. These events created great opportunities for soybean oil to become the preferred oil for food manufacturers (Goldsmith 2008). In 2010, ~68% of edible fats and oils consumed in the U.S. derived from soybean. Products from soybean oil include salad or cooking oil, baking and frying
fats, margarine and shortenings (SoyStats 2011; Wilson 2008). Also, about 14% of the soybean oil consumed in the U.S. is utilized for industrial applications and products such as soap, cosmetics, plastics, inks, solvents, resins, clothing and biodiesel (SoyStats 2011; Wilson 2008). Soybean is also a source of high-value secondary co-products such as lecithin, vitamins, nutraceuticals and antioxidants (Wilson 2008).

**Soyfoods**

Soybeans have long been consumed by humans in Asia in such traditional foods as tofu, soymilk, tempeh and natto. Some of these foods have also become popular in other parts of the world (Liu 2008). In addition, in recent years, breakthroughs in food science and processing have made it possible to use soybean ingredients in new ways, creating foods that are familiar to consumers but that incorporate parts of the soybean for functional or nutritional purposes. This has greatly expanded the food processing industry’s use of soybeans and soy-based ingredients.

Approximately 6% of soybeans are used directly as human food, mostly in Asia (Goldsmith 2008; Soyatech 2011). In addition, ~2% of the soybean meal is further processed into soy flours and proteins for food use (Goldsmith 2008; Soyatech 2011). The highly regarded nutritional value of soyfoods derives from the quantity and quality of the protein and the oil found in soybean seeds. Soybean proteins have excellent amino
acid profiles and contain all eight amino acids essential for human health (Murphy 2008; Stein et al. 2008). In addition to the nutritional qualities, soybean proteins have cholesterol-lowering ability and anticancer activities (FDA 1999; Kennedy 1995; Kennedy 1998; Murphy 2008). Soybean seeds also contain a high amount of oil. The oil found naturally in soybean seeds, which ends up in whole soybean-based foods, contains ~60-65% polyunsaturated fatty acids and ~20-25% monounsaturated fatty acid. The consumption of soybean oil has also been associated to multiple health benefits including reduction of cholesterol, blood pressure and cardiovascular risk factors (Hill et al. 2008).

In the U.S., the soyfood market has shown a dramatic increase over the last fifteen years, with sales totaling ~US$ 1 billion in 1996 and growing up to US$ 4.5 billion in 2009 (Soyfoods 2011). Much of this increase took place after the Food and Drug Administration (FDA) (1999) officially recognized the cholesterol-lowering effects of soy protein. Another factor driving this growth is the permanent introduction of new soyfood products in the market. It is estimated that over 2700 new products with soybean as an ingredient have been introduced in the market between 2000 and 2007 (Soyfoods 2011).

Soyfoods can be classified as traditional and second generation types. Traditional soyfoods, in turn, can be classified as fermented and non-fermented (Liu 2008; Soyatech 2011; USB 2010). Traditional non-fermented soyfoods include:
Tofu: a soft, cheese-like food made by curdling fresh, hot soymilk with a coagulant. Tofu is a bland product that easily absorbs the flavors of other ingredients with which it is cooked. Tofu is rich in both high-quality protein and B vitamins. Firm tofu is higher in protein, fat, and calcium than other forms of tofu. Soft tofu is good for recipes that call for blended tofu. Silken tofu is a creamy product and can be used as a replacement for sour cream in many dip recipes.

Soy milk: soybeans soaked, ground fine, and strained produce a fluid called soybean milk. Plain, unfortified soymilk is an excellent source of high-quality protein, B vitamins, iron, unsaturated fatty acids and niacin. It contains no cholesterol or lactose. Compared to cow or human milk, soy milk has lower amounts of fat, carbohydrates, calcium, riboflavin, thiamine, methionine and lysine (Kosikowski 1971).

Soy sprouts: made from soybeans germinated in the dark, soy sprouts are an excellent source of nutrition since they are packed with protein, vitamin C, and β-carotene (Bates and Matthews 1975; Xu et al. 2005). Also, compared to dry soybeans, soy sprouts have reduced levels of flatulence-causing oligosaccharides and phytic acid (Liu 2008).

Yuba: made by lifting and drying the thin layer formed on the surface of cooling hot soymilk. It has 55% protein and 28% lipids content (Wu and Bates 1972) and the protein in yuba has ~100% digestibility (Ikeda et al. 1995).
Okara: is a pulp fiber by-product of soymilk. It has high fiber content and considerable amounts of protein (28%) and oil (10%) (van der Riet et al. 1989). Okara tastes similar to coconut and can be baked or added as fiber to granola and cookies.

Soy nuts: are whole soybeans that have been soaked in water and then baked until browned. Soy nuts can be covered with a variety of flavor and coatings, including chocolate. Higher in protein and isoflavones and lower in lipids, soy nuts are similar in texture and flavor to peanuts (Liu 2008).

Toasted soy flour: made from roasted soybeans ground into a fine powder. Soy flour is 50% protein and gluten-free. There are three kinds of soy flour available: natural or full-fat, which contains the natural oils found in the soybean; defatted, which has the oils removed during processing; and lecithinated, which has lecithin added to it.

Green vegetable soybeans: these large soybeans are harvested when the beans are still green and sweet tasting and can be served as a snack or a main vegetable dish after boiling in slightly salted water for 15-20 minutes. They are high in protein (11-16%) and fiber and contain ~8-11% oil. Green soybeans are sold frozen in the pod (edamame) or shelled (mukimame).
Traditional fermented soyfoods include:

*Soy sauce:* is a dark-brown liquid made from soybeans that have undergone a fermenting process. Soy sauces have a salty taste, but are lower in sodium than traditional table salt. Specific types of soy sauce are shoyu, tamari and teriyaki.

*Tempeh:* is a tender white cake of cooked soybeans. The solid cake is formed because of fermentation which binds the soy and other grains together. The texture of tempeh is firm, tender and chewy. High quality tempeh has a mild “mushroom-like” flavor and aroma. Tempeh contains about 19% protein, is higher in fiber than tofu and is a significant source of vitamins and minerals.

*Miso:* is a rich, salty condiment made from soybeans and a grain such as rice, plus salt and a mold culture and then aged in cedar vats for one to three years.

*Natto:* is made of fermented, cooked whole soybeans. Because the fermentation process breaks down the beans’ complex proteins, natto is more easily digested than whole soybeans. Natto also contains important amounts of vitamin $K_2$ (Yanagisawa and Sumi 2005), a cofactor of osteocalcin, a bone protein (Yamaguchi et al. 2001). It has a sticky, viscous coating with a cheesy texture. It is traditionally served as a topping for rice, in miso soups and with vegetables.
Other traditional fermented soyfoods include *soy yogurt* (fermented soy milk), *sufu* (fermented tofu) and *soy nuggets* (fermented whole soybeans).

Second generation soyfoods include:

*Meat alternatives*: this category contains hundreds of products made from tofu, tempeh, textured soy flour, textured soy concentrate, isolated soy protein and wheat gluten. Products may take the shape of burgers, hot dogs, sausages, luncheon meats, ground meat and meatballs.

*Cheese alternatives*: block, sliced, spreadable and grated cheese alternatives may be made from soymilk, tofu or other protein ingredients. They can be flavored like American, Mozzarella, Cheddar, Monterey Jack, Parmesan and others.

*Frozen desserts*: produced in much the same manner as their dairy counterparts. They may be prepared from a base of soymilk, soy yogurt, tofu or isolated soy protein.

*Baked goods*: soy protein is used in the manufacturing of breads, cookies, crackers and other baked goods.

*Cereals*: Soy protein is used extensively as an ingredient in hot cereal mixes and breakfast bars to boost protein value and quantity.

*Pasta*: pasta products can be fortified with soy protein to increase nutritional value.
Soybean seed composition

Nutritional factors

Soybean seeds typically contain 40% protein, 20% oil, 35% carbohydrates and 5% ash on a dry weight basis (Liu 1999). There is a wide range of genetic variation in seed protein and oil concentrations among accessions of the USDA Soybean Germplasm Collection, with protein and oil levels ranging from 34.1 to 56.8% and from 8.3 to 27.9%, respectively (Wilson 2004).

The largest mass of the seed protein (65-80%) is comprised of the storage proteins glycinin and β-conglycinin. Storage proteins have no biological activities but they serve as nitrogen stores for the germinating seeds. The other soybean seed proteins have biological activity and include the lipoxygenases, the trypsin inhibitor family and lectins (Murphy 2008). Soybean proteins have excellent amino acid profiles for humans but they are deficient in sulfur amino acids for livestock and rodents (Murphy 2008).

Lipids, one of the most important components of soybean seeds, are primarily found in the cotyledons (Gerde and White 2008). Most lipids in soybean seeds act as energy reserve while the rest are structural components of membranes. The predominant fatty acid (FA) composition of commodity soybean oil is 11% palmitic (16:0), 4% stearic (18:0), 23% oleic (18:1), 54% linoleic (18:2) and 8% linolenic (18:3) (Fehr 2007; Hill et al. 2008; Wilson 2004). For reference, the first number of these FA abbreviations
indicates the number of carbon atoms, and the second number represents the number of
double bonds in the molecule. 16:0 and 18:0 are saturated FAs, 18:1 is a
monounsaturated FA, and 18:2 and 18:3 are polyunsaturated FAs (PUFA). Also, 18:2 and
18:3 are ω-6 and ω-3 types of FAs, respectively, which are essential to humans and all
animals. The term essential refers to any compound that is required for life but cannot be
synthesized by the organism and therefore must be incorporated with diet. Compared to
other popular vegetable oils, soybean oil contains substantially less 18:1 than olive or
canola oil (Fig. 1.1). The proportion of saturated FAs is similar in soybean and olive
while canola contains half the levels observed in soybean and olive oil. Soybean oil is
high in 18:2, and it has 18:3 contents similar to walnut and wheat germ. Canola and
hempseed oils have slightly higher 18:3 contents while flaxseed oil is considerable higher
in 18:3 than soybean oil. Sunflower, corn, olive, peanut and palm oils provide very small
quantities of 18:3 (Fig. 1.1). Since ~70% of the edible fats and oils consumed in the U.S.
derive from soybean, soybean oil is a major contributor of dietary 18:3 (Hill et al. 2008).

Carbohydrates make up a significant part of soybean seeds, typically 30-35% of the
seed weight (Middelbos and Fahey Jr. 2008). Based on their physicochemical properties,
these carbohydrates are divided into structural and non structural. The first group
includes dietary fiber components (indigestible by endogenous enzymes) such as
cellulose, hemicelluloses, pectins and glycoproteins (Bach Knudsen et al. 1987;
Selvendran et al. 1987). Non structural carbohydrates comprise starch, sucrose and other
low molecular weight sugars and oligosaccharides (raffinose and stachyose) (Karr-Lilienthal et al. 2005; Wilson 2004). Sucrose, which is nutritionally useful in monogastric animals, can be as high as 9% whereas stachyose and raffinose, not digestible in monogastric animals, can reach 6.4 and 1.4% of the dehulled soybean meal, respectively (Middelbos and Fahey Jr. 2008).

Soybean seeds contain various highly valued minor constituents. Among those associated to lipids, the most important are tocopherols (antioxidants) (Wang 2008; White and Xing 1997), phytosterols (associated to reductions of low density lipoprotein cholesterol and antioxidants) (Tian and White 1994; Wang 2008), phospholipids (provide choline, precursor of the neurotransmitter acetylcholine; reduce cardiovascular disease and liver cancer) (da Costa et al. 2005; Innis et al. 2007; Orthoefer and List 2006), sphingolipids (mediators of cell growth and differentiation and thus cancer) (Merrill and Schmelz 2001) and carotenoids (pro-vitamin A) (Wang 2008). Minor components of value associated to proteins include lectins (inhibit tumor growth; however, they also destroy intestinal cell organization with negative impact on absorption) (Gonzalez De Mejia and Prisecaru 2005; Sharon 2007) and lunasin (cancer preventive) (Galvez et al. 2001; Jeong et al. 2007). Finally, soybean seeds also contain isoflavones, which are receiving increasing attention due to the multiple health benefits associated with its consumption (Hirose et al. 2005; Messina and Hughes 2003; Rochfort and Panozzo 2007).
Antinutritional factors

Some of the soybean seed constituents have adverse effects on nutrient digestibility. Among those, trypsin inhibitors are one of the most important. These compounds block the protease activity of trypsin and other enzymes, decreasing protein digestibility and bioavailability (Liener 1994). Phytate, another antinutritional factor, has high affinity for di- and trivalent cations including calcium, magnesium, zinc, iron and phosphorus. Since monogastric animals lack phytase, the enzyme responsible for breaking down this storage molecule, there is a decrease in nutrient bioavailability and feed conversion when phytate is in the meal (Selle et al. 2000; Traylor et al. 2001). Raffinose and stachyose cannot be digested by monogastric animals due to the lack of α-galactosidase in the gut. However, the intestinal micro flora is capable of utilizing those oligosaccharides, generating gases and acids that lead to bloating, diarrhea and general discomfort (Coon et al. 1990; Lowell and Kuo 1989; Zuo et al. 1996).

Lipoxygenases

Lipoxygenases are enzymes belonging to a group of non-heme iron containing proteins widely distributed in plants and mammals; only recently, they have been detected in coral, moss, fungi and a number of bacteria as well (Andreou et al. 2009; Brash 1999; Oliw 2002). The substrates for lipoxygenases are absent in most bacteria
and yeast, and so are lipoxygenases (Brash 1999; Watanabe et al. 1997). In insects, there is no definitive account of the presence of these enzymes (Brash 1999).

Lipoxygenases catalyze the incorporation of molecular oxygen into PUFAs that contain one or more (1Z,4Z)-pentadiene moieties to yield the corresponding hydroperoxy FAs (Brash 1999; Liavonchanka and Feussner 2006). In plants, oxygenation can occur on carbon number 13 (catalyzed by the 13-Lox family) or on carbon number 9 (catalyzed by the 9-Lox family) of 18:2 and 18:3, the most important substrates for lipoxygenase catalysis (Andreou and Feussner 2009; Feussner and Wasternack 2002; Liavonchanka and Feussner 2006; Song et al. 1990). In mammals, the substrate for lipoxygenases is the ω-6 FA arachidonic acid (20:4) (Brash 1999).

**Occurrence of lipoxygenases in plants**

In plants, lipoxygenases can be found in all organs, representing up to 2% of soybean seed protein (Loiseau et al. 2001). Siedow (1991) reported that most plant lipoxygenases are soluble enzymes predominantly located in the cytoplasm. However, lipoxygenases can be also found in other cell compartments and in association with membranes (Loiseau et al. 2001). In fact, lipoxygenases are classified according to their subcellular localization as type 1 and type 2 (Shibata et al. 1994). Type 1 lipoxygenases include extraplastidial enzymes with no transit peptide and high sequence similarity to
one another (>75%). Most plant lipoxygenase genes isolated so far belong to the type 1 class (Loiseau et al. 2001). Type 2 lipoxygenases comprise plastidial enzymes that carry a chloroplast transit peptide and show moderate overall sequence similarity (>35%) (Shibata et al. 1994).

**Role of lipoxygenases**

Lipoxygenase-derived fatty acid hydroperoxides can be further metabolized into volatile aldehydes and jasmonates in plants (Mosblech et al. 2009), in diols and lactones in fungi (Tsitsigiannis and Keller 2007) and in lipoxins and leukotrienes in mammals (Samuelsson et al. 1987; Sigal et al. 1994). The metabolites that derive from the oxidation of PUFAs via a lipoxygenase-catalyzed step or from alternative oxidation reactions and subsequent reactions are collectively named oxylipins. Oxylipins are involved in inflammation, asthma and heart disease in mammals (Samuelsson et al. 1987; Sigal et al. 1994) while in fungi they play a role in regulation of mycotoxin production and of the sexual and asexual life cycle (Tsitsigiannis and Keller 2007).

In plants, products of the lipoxygenase pathway have been shown to have a role in a variety of processes such as vegetative growth (Kolomiets et al. 2001; Porta et al. 1999; Vellosillo et al. 2007); wounding, herbivore and pathogen attack responses (Bell et al. 1995; Creelman and Mullet 1997; Rancé et al. 1998; Rustérucci et al. 1999) and
mobilization of storage lipids during germination (Feussner et al. 2001; Porta and Rocha-Sosa 2002). In addition, lipoxygenase-derived oxylipins also play a role in fertility and flower development (Caldelari et al. 2011); signaling, as inducers of defense gene expression (Glazebrook 2001; Wasternack 2007) and as antimicrobial compounds (Prost et al. 2005).

**Importance and effect of lipoxygenases on food quality**

The lipoxygenase-mediated formation of FA hydroperoxides is the starting point for a series of reactions that result in the production of short chain aldehydes, ketone alcohols, acids, cyclopentanoids and epoxides, among others. Many of these compounds, especially the aldehyde hexanal, are responsible for the development of off flavors and aroma in soybean-containing products (Gardner 1991; Hamilton et al. 1997; Sekiya et al. 1986; Sessa 1979). The undesirable flavors, characterized as beany, grassy, astringent and bitter, reduce the acceptance of soybean products by many consumers who prefer a bland flavor (Gerde and White 2008; Robinson et al. 1995; Sessa 1979; Torres-Penaranda et al. 1998; Wilson 1996). Tofu and soy milk made with lipoxygenase-free soybeans were reported to have reduced beany flavor and aroma (Torres-Penaranda et al. 1998).

Strategies followed to reduce the off flavors include preventing the beany flavor formation, stripping off the responsible volatiles once they have formed, masking the off
flavor with flavorings and genetic elimination of seed lipoxygenases (Liu 2008). Since the use of soybean seeds lacking the three lipoxygenase enzymes is the only alternative that allows reducing time, labor and expenses during the processing of soybean for soyfoods, lipoxygenase-free soybean varieties are in demand.

Lipoxygenases in soybean

Lipoxygenases constitute a gene family with six predicted Lox genes in the model species Arabidopsis (Bannenberg et al. 2009). Shin et al. (2008) identified 19 Lox genes in soybean, many more than those found in Arabidopsis. There is evidence suggesting that soybean underwent two whole-genome duplication events, which resulted in a highly duplicated genome with approximately 75% of the genes existing as paralogues (Schmutz et al. 2010). Not all the Lox genes found in the soybean genome are expressed in seeds. Mature soybean seeds contain primarily three lipoxygenases, Lox1, Lox2 and Lox3 (Axelrod et al. 1981). These enzymes are localized in the cytoplasm of cotyledon cells and in their protein-storage vacuoles (Song et al. 1990; Wang et al. 1999). Soybean Lox1 promotes oxidation on carbon 13 only while Lox2 and Lox3 catalyze the production of both the 13- and the 9-hydroperoxides (Axelrod et al. 1981; Christopher et al. 1972; Fukushige et al. 2005). The physiological role of soybean seed lipoxygenases are not understood yet. Song et al. (1990) proposed they oxygenate FAs to facilitate their
transport to the glyoxisomes. However, other studies suggested that lipoxygenases are not involved in lipid mobilization during germination and that they instead were recruited as storage proteins (Siedow 1991; Wang et al. 1999). It cannot be ruled out that seed lipoxygenases play a role in the defense of the seed during development by catalyzing the formation of various oxylipins (Wang et al. 1999). Regardless of the actual role lipoxygenases may play in the physiology of soybean seeds, lipoxygenase-free seeds have been able to develop into normal plants without defect (Hajika et al. 1992; Wang et al. 1999).

**Genetics of soybean seed lipoxygenases**

Diverse soybean germplasm was screened to identify genotypes with missing Lox1, Lox2 and Lox3 activity (Davies and Nielsen 1986; Hildebrand and Hymowitz 1981, 1982; Kitamura et al. 1983; Kitamura et al. 1985). The plant introductions PI 133226 from Indonesia and PI 408251 from Korea were phenotyped as missing Lox1 activity (Hildebrand and Hymowitz 1981). Using both immunological and electrophoretic methods, Kitamura et al. (1983) identified the soybean cultivars ‘Wasenatsu’ and ‘Ichigowase’, from the Japanese germplasm collection, as lacking Lox3 activity. The soybean genotypes Wasenatsu and Ichigowase are currently part of the USDA soybean germplasm collection, and they can be found under the accession numbers PI 417458 and
PI 205085, respectively (http://www.ars-grin.gov/npgs/index.html). Finally, the accession PI 86023, which was introduced from Japan, was identified as a Lox2 null and is the only known source of the \textit{lox2} mutant allele (Davies and Nielsen 1986; Kitamura et al. 1985).

Genetic studies demonstrated that the absence of each enzyme is under the control of three null alleles, \textit{lox1}, \textit{lox2} and \textit{lox3}, which are inherited as simple recessive alleles (Davies and Nielsen 1986; Hildebrand and Hymowitz 1982; Kitamura et al. 1983; Kitamura et al. 1985).

The \textit{Lox1} and \textit{Lox2} loci were found to be in tight genetic linkage on chromosome 13 (LG F), with \textit{lox1} and \textit{lox2} mutant alleles being in the repulsion phase (Davies and Nielsen 1986; Hildebrand and Hymowitz 1982; Kitamura et al. 1985). The \textit{Lox3} locus, on chromosome 15 (LG E), segregates independently of \textit{Lox1} and \textit{Lox2} (Davies and Nielsen 1986; Hajika et al. 1992; Kitamura et al. 1985). Two types of double mutants were created by crossing single mutant genotypes: one type lacked Lox1 and Lox3 activity while the other lacked Lox2 and Lox3 activity (Davies and Nielsen 1986; Kitamura et al. 1985).
Development of a triple lipoxygenase mutant genotype in soybean

Due to the tight linkage in repulsion phase, a \textit{lox1, lox2} double mutant or a \textit{lox1, lox2, lox3} triple mutant could not be developed through traditional breeding. The original repulsion-phase linkage in independent sources of mutant alleles at the \textit{Lox1} and \textit{Lox2} loci was eventually broken, resulting in a coupling-phase linkage that ultimately led to the development of a triple null lipoxygenase genotype (Hajika et al. 1991; Kitamura 1991). The development of the triple \textit{Lox} mutant included gamma-ray mutagenesis of F$_2$ seeds from crosses between \textit{lox1, lox3} and \textit{lox2, lox3} double mutants, (Hajika et al. 1991). After screening more than 1800 M$_3$ individuals, only one seed lacking all the seed lipoxygenases was identified (Hajika et al. 1991). However, the mechanism involved in the induction of the genotype lacking the three seed lipoxygenases was unknown (Hajika et al. 1991). Point or gene mutation, chromosomal deletion or even cross-over between the \textit{Lox1} and \textit{Lox2} loci were all possible causes leading to the creation of a triple mutant genotype (Hajika et al. 1991).
Molecular basis of the lipoxygenase 2-null phenotype in PI 86023

The molecular basis of the null Lox2 phenotype was reported to be a T2849A missense mutation (Wang et al. 1994). This change resulted in the substitution of glutamine for histidine at position 532 in a highly conserved histidine-rich motif of the Lox2 protein. The histidine residue at position 532 of Lox2 was shown to be one of the iron-binding ligands essential for Lox activity (Steczko et al. 1992; Wang et al. 1994). Although the mutation did not prevent gene transcription (transcripts were detected in mid-maturation seeds), it severely affected the structure and function of the protein, resulting eventually in the degradation of the mutated enzyme before the seeds reached maturity (Wang et al. 1994). The genetic basis of both Lox1- and Lox3-free genotypes has not been previously reported.

Screening for seed lipoxygenases

The development of seed-lipoxygenase-free soybean cultivars requires accurate and repeatable evaluation assays to select for the appropriate genotypes. The presence or absence of lipoxygenases in breeding programs has been determined by laborious and time consuming seed phenotyping methods. In addition, these procedures are not able to discriminate between individuals that are homozygous wild type and heterozygous in the same generation. Consequently, tests to study the genetic makeup of the parents must be
performed on the progeny, adding both time and expense to the selection process. The methods that have been utilized include detection of the Lox proteins in single dimension SDS-polyacrylamide gels using electrophoresis and crude seed extracts with either total protein staining or by Westerns with specific lipoxygenase antibodies and colorimetric enzymatic assays, based on the pH of the test solution for which the corresponding lipoxygenase has optimum activity (Kitamura et al. 1983; Narvel et al. 2000; Suda et al. 1995). A single nucleotide polymorphism (SNP) marker that co-segregated perfectly with the Lox2 locus was also previously developed, but it was not specific for the lox2 causative mutation (Kim et al. 2004).

After the publication of the paper “Soybean seed lipoxygenase genes: molecular characterization and development of molecular marker assays” (Lenis et al. 2010), a study analyzing the molecular basis of the seed Lox null line OX948 was published (Reinprecht et al. 2011). The lox1 mutant allele present in soybean accession OX948 was reported to have the same causative mutation as that previously described for the lox1 single mutant genotype PI 408251 as well as the triple mutant Jinpumkong 2. Also, the reported T2849A missense mutation that rendered Lox2 inactive in PI 86023 and present in all Lox2 null accessions was found in the triple mutant OX948. The mutation in the lox3 allele of OX948 was different from that described for the single lox3 mutants PI
205085 and PI 417458 and the triple mutant Jimpumkong 2. In the soybean line OX 948, the \textit{lox3} mutant allele presented, among other mutations, two single base substitutions (T-584A and C-635T) in an AAATAC paired box in the promoter region of the gene (Reinprecht et al. 2011). Even though Reinprecht et al. (2011) suggest that the mutations in the promoter region are responsible for the dysfunctional \textit{Lox3} gene in OX948, they could not demonstrate it unambiguously since some lines carrying such mutations presented Lox3 activity. The authors attribute the contradictory results to two possible causes: cross reaction among the Lox enzymes and/or the presence of an additional copy of the \textit{Lox3} gene. It seems also possible that the lack of Lox3 activity in OX948 is due to any of the other(s) polymorphisms found.

\textbf{Soybean oil and essential fatty acids}

Oil from commodity soybean genotypes has an average content of ~60-65\% PUFAs, with 18:2 and 18:3 representing 54 and 8\% of total oil, respectively (Fehr 2007; Hill et al. 2008; Wilson 2004). 18:2 and 18:3 are \(\omega\)-6 and \(\omega\)-3 types of FAs, respectively, which are essential to humans and all animals and must be incorporated with diet for normal growth and development (Brenna 2011; Gerde and White 2008; Griffiths and Morse 2006; UMMC 2011). In recent years, \(\omega\)-3 FAs have received much attention in the U.S. by nutritionists who are concerned about a decreasing ratio of dietary \(\omega\)-3 to \(\omega\)-6
Essential fatty acids and health

Fat constitutes between 15 and 22% of the body weight of an average healthy person (Griffiths and Morse 2006). It has typically being considered that the purpose of this fat was mainly as energy storage, but it is known now that fat plays important roles in various metabolic processes of importance to body function. FAs are structural components of the phospholipids of cell membranes, affecting membrane fluidity and modulating the behavior of membrane-bound receptors, enzymes and ion channels (Brown et al. 2005; Li et al. 2005). FAs are also involved in gene regulation and are responsible for the impermeability of the skin to water as well as for the regulation of permeability in the gut and other tissues (de Jager et al. 2004; Norlen and Al-Amoudi 2004; Sampath and Ntambi 2004). Also, FAs are precursors for compounds that regulate blood pressure, immunity and inflammation, among other functions (FitzGerald 2004; Francois and Coffman 2004; Plat and Mensink 2005).
Multiple studies show that diets consisting of oils rich in 18:2 and very low in 18:3 as the exclusive source of fat through pregnancy and lactation result in visual, cognitive and behavioral deficits in the offspring of primates, mice, rats and other species (Brenna 2011). Such diets consistently induced functional deficits in vision, reflex responses, learning, behavior and motor development, while feeding ω-3 FAs enhanced visual and cognitive functions in the offspring (Brenna 2011).

ω-3 FAs are important for normal growth and development. They also reduce inflammation and may help lower risk of diseases such as heart disease, cancer, type II diabetes and arthritis (UMMC 2011). ω-3 FAs are highly concentrated in the brain and appear to be important for cognitive (brain memory and performance) and behavioral function (UMMC 2011).

Sources of ω-3 fatty acids

Plant-based foods, such as oils, nuts and seeds are rich in 18:3. The most important sources of dietary 18:3 are flax (*Linum usitatissimum*) seeds and flax oil, walnuts (*Juglans* sp.) and walnut oil, canola (*Brassica napus*) oil, soybeans and soybean oil, olive oil and pumpkin seeds. 18:3 is also found in some green leafy vegetables, such as Brussels sprouts, kale, spinach, and salad greens (Gebauer et al. 2006; Sacks 2011).
Cold water fatty fish such as salmon, herring, mackerel, tuna and anchovies are the primary sources of eicosapentaenoic acid (EPA, 20:5) and docosahexaenoic acid (DHA, 22:6), other types of ω-3 FAs. Although fish are a dietary source of ω-3 FAs, they do not synthesize them; instead, EPA and DHA originate in algae and plankton and become increasingly concentrated in organisms as they move up the food chain (Falk-Petersen et al. 1998; Gebauer et al. 2006; Sacks 2011).

Soybean oil, 18:3 content and the nutritional quality vs. functionality dilemma

Since soybean oil is the predominant vegetable oil in the U.S. diet, it is a major contributor of dietary 18:3 and 18:2 (Hill et al. 2008; Soyconnection 2011; SoyStats 2011). Nonetheless, in soybean oil used for commercial frying and the manufacture of shortenings and margarines, much of the 18:2 and 18:3 is purposely reduced either through the process of hydrogenation or by genetic means (Warner 2008). Partially hydrogenated soybean oil contains 8.6% 18:2 and 0.2% 18:3, while the fully hydrogenated soybean oil contains no unsaturated FAs (Hill et al. 2008). There are also low and ultra low 18:3 soybean varieties which contain ~3 and 1% 18:3, respectively (Bilyeu et al. 2005; Bilyeu et al. 2006; Fehr 2007). Even though PUFAs are nutritionally desirable, the polyunsaturation contributes to oxidative instability and degradation during frying, which results in unpleasant flavor and odor (Fehr 2007; Gerde and White 2008;
O'Brien 2008). Therefore, the restaurant and food industries demand frying oils with improved oxidative stability, which unfortunately comes at the expense of the nutritionally desirable 18:3.

**Potential uses of soybean oil enriched in ω-3 fatty acids**

All the health benefits associated to a diet rich in ω-3 FAs make soybean genotypes with elevated 18:3 in the seed oil desirable, especially in food grade soybeans (Lee et al. 2007).

Also, there has been a growing interest from the aquaculture industry in using plant-based proteins and oils as components of fish diets as a way to reduce both feeding costs and, especially, the pressure on marine resources. Soybean protein concentrate and oil, due to their excellent nutritional characteristics, are considered very suitable for aquafeed. Increasing global demand for fish high in long-chain ω-3 oils creates perfect market opportunities for soybeans with enhanced ω-3 content (Naylor et al. 2009).

In addition, oils low in saturates and high in PUFAs would have applications as drying oils for paints, coatings, inks, resins and other industrial products (Fehr 2007; Wilson 2004).
**Fatty acid and Triacylglycerol synthesis**

Fatty acid biosynthesis is part of the triacylglycerol (TAG) synthetic pathway. A general scheme of FA and TAG synthesis in developing seeds is presented in Fig. 1.2. In plastids, FAs are synthesized from acetyl-CoA, which is carboxylated to form malonyl-CoA through the action of acetyl-CoA carboxylase (ACCase). A second enzyme system, fatty acid synthase (FAS), transfers the malonyl moiety to acyl carrier protein (ACP) and catalyzes the extension of the growing acyl chain with malonyl-ACP to form 16:0-ACP. Finally, elongation and desaturation of 16:0-ACP to form 18:0-ACP and 18:1-ACP, respectively, results in *de novo* synthesized FAs. By the action of ACP-thioesterases, the FA molecules are released from ACP in the plastid stroma and cross the membrane by an unknown mechanism. On the plastid outer membrane, an acyl-CoA synthetase (ACS) assembles acyl-CoA esters that are then available for acyltransferase reactions in the endoplasmic reticulum (ER). In the ER, synthesized FAs are incorporated onto a glycerol backbone to form TAG through the Kennedy pathway. Two successive acyl-CoA-dependent acylation reactions transfer FAs to glycerol-3-phosphate (G3P) to produce lysophosphatidic (LPA) and phosphatidic (PA) acids, respectively. PA dephosphorylation results in diacylglycerol (DAG), which is at the branch point of the pathway between TAG and phosphatidylcholine (PC) formation. The acyl-CoA-dependent acylation of DAG, catalyzed by diacylglycerol acyltransferase (DGAT), leads to TAG, while the transfer of a phosphocholine into DAG, catalyzed by CDP-choline:1,2-diacylglycerol
cholinephosphotransferase (CPT), results in PC formation. A recently discovered enzyme, phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT) catalyzes the interconversion between DAG and PC by phosphocholine headgroup exchange (Lu et al. 2009). The FA desaturation (FAD) pathways, catalyzed by FAD2 and FAD3 desaturases, utilize PC as a substrate increasing the number of unsaturations of its acyl chains. The PUFA-enriched PC, by the reverse reaction of CPT and PDCT, can be transformed back into DAG. Also, the acyl chain at position 2 of PC may undergo exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). PC can also participate in an acyl-CoA-independent reaction to produce TAG. In this case, phospholipid:diacylglycerol acyltransferase (PDAT) catalyzes the formation of TAG using DAG and PC as substrates. Finally, a second acyl-CoA-independent reaction leading to TAG synthesis has been described, where diacylglycerol transacylase (DGTA) catalyzes the transfer of an acyl moiety between two DAG molecules to form TAG and monoacylglycerol (MAG) (Lung and Weselake 2006; Ohlrogge and Browse 1995; Ohlrogge and Jaworski 1997).

**Fatty Acid Desaturase (FAD) genes in plants**

Fatty acids are the main components of plant membrane and seed storage lipids (Töpfer et al. 1995). In plants, FAs are synthesized *de novo* in the stroma of plastids
through a complex series of condensation reactions to produce either C16 of C18 FAs (Browse and Somerville 1991; Lung and Weselake 2006). In most plant tissues, over 75% of the FAs are unsaturated (Hildebrand et al. 2008). In non photosynthetic tissues, PUFAs can reach up to 90% of the oil (Miquel and Browse 1994). Fatty acid desaturases are the enzymes that introduce double bonds, i.e., unsaturations, into the hydrocarbon chains of fatty acids. The activity of these fatty acid desaturases is critical for the function of biological membranes by maintaining their proper fluidity. These enzymes are encoded by nuclear genes and differ in their substrate specificity and subcellular localization. FAD2 and FAD3 are located in the ER while FAB2, FAD4, FAD5, FAD6, FAD7 and FAD8 are targeted to the plastids (Wallis and Browse 2002). FAB2, also referred to as Stearoyl ACP Desaturase (SACP), is a soluble desaturase and catalyzes the synthesis of 18:1 from 18:0; the other desaturases are membrane bound. FAD2 and FAD6 are ω-6 desaturases. FAD3, FAD7 and FAD8 are ω-3 desaturases. Finally, FAD4 and FAD5 produce 16:1 from 16:0 (Gibson et al. 1994; Hildebrand et al. 2008; Murphy and Piffanelli 1998). While the microsomal FAD2 (ω-6 desaturase) and FAD3 (ω-3 desaturase) enzymes can only add double bonds on 18:1 and 18:2, respectively (Browse et al. 1993; Miquel and Browse 1992), the plastidial FAD6 (ω-6 desaturase) and FAD7 and FAD8 (ω-3 desaturases) can use as substrates both 16:1 or 18:1 and 16:2 or 18:2, respectively (Browse et al. 1989; Browse et al. 1986; McConn et al. 1994). Both microsomal and plastidial ω-3 and ω-6 desaturases can act on FAs attached at either
position 1 or position 2 of the glycerol backbone. Also, FAD2 and FAD3 can only use as substrate FAs esterified to PC, whereas the plastidial FAD6, FAD7 and FAD8 have no apparent specificity for the nature of the lipid headgroup (Browse et al. 1989; Browse et al. 1993; Browse et al. 1986; McConn et al. 1994; Miquel and Browse 1992; Ohlrogge and Browse 1995).

**Microsomal FAD genes in the complex genome of soybean**

Most PUFAs, especially in non photosynthetic tissues, are produced by desaturases located in the ER (Miquel and Browse 1994). The soybean genome is complex and has a predicted size of 1,115 Mb distributed across 20 chromosomes (Arumuganathan and Earle 1991; Schmutz et al. 2010). There is evidence suggesting that soybean underwent two whole-genome duplication events followed by gene diversification and loss and chromosome rearrangements. The result is a highly duplicated genome with approximately 75% of the genes existing as paralogues (Schmutz et al. 2010). Therefore, in soybean, it is very likely to find multiple copies of a gene that is present as a single copy in other species like *Arabidopsis*.

In soybean, there are at least five FAD2 and three FAD3 genes. *FAD2-1A* (Glyma10g42470) and *FAD2-1B* (Glyma20g24530) are primarily expressed in developing seeds, and play an important role in determining the PUFA content in soybean
oil (Heppard et al. 1996; Pham et al. 2010; Schlueter et al. 2007; Tang et al. 2005). The FAD2-2 genes, consisting of FAD2-2A (Glyma19g32930), FAD2-2B (Glyma19g32940) and FAD2-2C (Glyma03g30070), are constitutively expressed in both vegetative tissue as well as developing seeds (Li et al. 2007; Schlueter et al. 2007). The exception is FAD2-2A, which produced amplicon when using genomic DNA but failed to amplify any of the RNA samples collected from diverse tissues. This would suggest that FAD2-2A is not expressed (Schlueter et al. 2007). Three microsomal ω-3 desaturase genes, namely FAD3A (Glyma14g37350), FAD3B (Glyma02g39230) and FAD3C (Glyma18g06950), have been identified in soybean (Bilyeu et al. 2003). FAD3A was characterized as the most highly expressed of the three homologs in developing seeds. FAD3A, FAD3B and FAD3C genes were also expressed in leaf tissue and flowers (Bilyeu et al. 2003). cDNA from a putative fourth microsomal ω-3 desaturase gene, herein referred to as FAD3D (Glyma11g27190), was obtained and characterized (Anai et al. 2005). Spatial and temporal expression of this gene seems to indicate it is a minor contributor in the desaturation of FAs in developing seeds (Anai et al. 2005).

**Contribution of FAD3 genes to seed 18:3 content**

The relative contribution of FAD3A, FAD3B and FAD3C genes to the total seed 18:3 content was quantified in soybean (Bilyeu et al. 2005; Bilyeu et al. 2006). 18:3
concentration was measured in seeds of individuals segregating for mutant and wild type alleles of the three \textit{FAD3} genes (Bilyeu et al. 2005; Bilyeu et al. 2006). An additive association was found, so that as the mutant alleles replaced their wild type counterparts, a stepwise decrease in seed 18:3 content could be observed. Also, it was noticed that the \textit{FAD3} genes make unequal contributions to the total 18:3 content, with mutations in \textit{FAD3A} associated with the greatest reductions (~50% of the total 18:3 level). \textit{FAD3B} and \textit{FAD3C} make nearly equivalent contributions and smaller than that of \textit{FAD3A} (~20% of the total each). The combination of mutant \textit{FAD3A}, \textit{FAD3B} and \textit{FAD3C} genes results in soybean seeds with ~1% 18:3 content, a phenotype that is a stably inherited by the offspring (Bilyeu et al. 2005; Bilyeu et al. 2006; Ross et al. 2000).

\textbf{Genetic modification of fatty acid profile in soybean oil}

Genetic modification of the typical FA makeup of soybean oil has been conducted for a number of years to improve oil nutritional and functional quality (Fehr 2007; Gerde and White 2008). Research priorities were based on targeted FA profiles identified by consumers and end-users as having the highest probability to expand the use of soybean oil in edible and industrial applications (Wilson 1998). Three different oil phenotypes were chosen based on the relative volume of soybean oil used for specific applications (Lee et al. 2007; Wilson 1998; Wilson 2004). In the U.S., soybean oil is currently used
mainly as salad or cooking oil (58%), baking or frying fats (23%) and for various industrial applications (14%). First, soybean oil containing low saturates (<7%), mid to high 18:1 (60% or higher) and low 18:3 (~2%) would be a competitive product for the salad and cooking oil, baking and frying fats markets (Wilson 2004). Oil with such profile has good oxidative stability, meaning that there is no need for hydrogenation with the consequent formation of the unhealthy trans-fats, and in addition, it meets the FDA guidelines for a “low in saturated fat” label (Lee et al. 2007; Wilson 2004). Diets low in saturated fats and high in 18:1 are associated to reduced cholesterol and heart disease (Chang and Huang 1998; Wardlaw and Snook 1990). In addition, oils with even higher 18:1 levels would be of use in the manufacture of lubricants, hydraulic oil, soy diesel as well as many pharmaceuticals and cosmetic products (Dunn et al. 1996; USDA 1996). Second, soybean oil high in saturated FAs (~40%) would minimize the need for hydrogenation to form ingredients for products such as margarine and could be used to make trans-fat-free margarine and shortening based products (List et al. 1995; List et al. 2000). Third, soybean oil with higher levels of PUFAs (~85%) would represent a good alternative to petroleum ingredients for industrial applications such as the manufacture of oil-based paints, coatings, inks, among others (Wilson 2004).

Genetic changes in soybean oil composition were achieved by conventional breeding and genetic engineering. Numerous genes that influence the contents of each of the FAs typically found in soybean have been identified, and in the majority of cases, the
changes in FA profile were associated with loss-of-function mutations within genes that encode biosynthetic enzymes (Gillman and Bilyeu 2011). For example, mid to high 18:1 (50-80% 18:1) as well as low and ultra low 18:3 (<3% and ~1% 18:3, respectively) genotypes were obtained by breeding with mutant alleles of FAD2 and FAD3 genes (Alt et al. 2005a; Alt et al. 2005b; Bilyeu et al. 2005; Bilyeu et al. 2006; Bilyeu et al. 2003; Ross et al. 2000). These non-functional alleles cause the accumulation of substrates and reduction of corresponding products in their respective desaturation points. To illustrate, soybean genotype M23, which contains 46% 18:1, was developed by treatment of cultivar Bay with X-ray irradiation (Rahman et al. 1994). Deletion of the gene at the FAD2-1A locus was associated with the mid-oleic phenotype of M23 (Sandhu et al. 2007). Recently, it was observed that combination of mutant alleles at both the FAD2-1A and FAD2-1B loci further increases 18:1 content up to 80% of oil (Hoshino et al. 2010; Pham et al. 2010). Soybean oil with >80% 18:1 was also developed through genetic engineering (Kinney and Knowlton 1998; Parrot and Clemente 2004) by silencing the soybean FAD2-1 genes via expression of transgenic FAD2-1 constructs (Buhr et al. 2002; Kinney and Knowlton 1998; Parrot and Clemente 2004). High 18:1 genotypes typically present a reduction in the other four main FAs commonly found in soybean oil but most significantly in the fraction of PUFAs (Kinney and Knowlton 1998; Pham et al. 2010). Soybean oils with such profile have an exceptional oxidative stability and are very suitable for food and industrial applications (Warner and Gupta 2005). In addition to the
high 18:1 genotypes, soybean lines carrying mutant alleles of \textit{FAD2-1A}, \textit{FAD2-1B}, \textit{FAD3A} and \textit{FAD3C} genes were developed (K. Bilyeu, personal communication, 2011). The oil phenotype that results from stacking these four mutations is very high 18:1 (~80%) and very low 18:3 (<3%) contents (K. Bilyeu, personal communication, 2011).

\textbf{Acyltransferase genes in plants}

The FA esterification reactions that lead to the formation of TAG are catalyzed by five types of acyltransferases. GPAT, LPAAT and DGAT are the acyltransferases involved in the Kennedy pathway (Fig. 1.2). In a second pathway, the Lands’ cycle, acyl groups are incorporated into PC through the activity of LPCAT, and then subsequently channeled from PC to TAG via PDAT (Fig. 1.2). It has been suggested that the reverse reaction of LPCAT may play a key role in exchanging “remodeled” FAs from phospholipids such as PC back into the cytosolic acyl-CoA pool, making these FAs available for any reaction catalyzed by acyl-CoA dependent transferases (Cahoon et al. 2007; Dyer et al. 2008; Napier and Graham 2010). However, plant orthologues of the recently identified mammalian \textit{LPCAT} genes have yet to be discovered (Napier and Graham 2010). Two of the acyltransferases discussed above, DGAT and PDAT, have been shown to be important in both the determination of total oil content as well as fatty acid composition in seeds (Katavic et al. 1995; Zhang et al. 2009; Zheng et al. 2008).
Cholinephosphotransferase genes in plants

There are at least two types of cholinephosphotransferases in plants, and both of them catalyze reactions involving DAG and PC: CPT and PDCT (Fig. 1.2). CPT plays a role in the production of de novo PC, and it transfers a phosphocholine group from cytidine diphosphate (CDP)-choline to DAG, generating as a result cytidine monophosphate and the phospholipid PC (Lu et al. 2009; Lung and Weselake 2006). The reverse reaction of CPT moves a more unsaturated DAG into the mainstream of TAG synthesis (Weselake et al. 2009). PDCT catalyzes the transfer of the phosphocholine headgroup from PC to DAG, and hence allowing the incorporation of 18:1 into PC for further desaturation as well as the transfer of 18:2 and 18:3 into DAG molecules that then enter the TAG synthesis pathway (Lu et al. 2009). In developing seeds, most of the conversion of DAG to PC would be carried out by PDCT (Lu et al. 2009). Of the two cholinephosphotransferases, only PDCT has been shown to play a major role in the FA composition of Arabidopsis seed oil (Lu et al. 2009).

Role of acyltransferase and cholinephosphotransferase genes in oil quality

Studies on the role of DGAT1 and PDAT1 in TAG biosynthesis have shown these genes contributing the most to TAG accumulation in Arabidopsis thaliana developing seeds (Zhang et al. 2009). While pdat1 single mutants presented unaltered oil content
and FA profile, mutants for the DGAT1 gene showed a 20 to 40% decrease in TAG content and a significant increase in 18:3 content at the expense of 18:1 and 20:1 (eicosenoic acid) (Katavic et al. 1995; Mhaske et al. 2005; Zhang et al. 2009). PDAT1 RNAi in a dgat1 background as well as the reciprocal combination resulted in a severe phenotype with an absolute oil reduction of ~84% compared with the dgat1 control. Also, oil in these individuals showed a decrease in 18:3 but retained the low 18:1 and 20:1 proportions compared with dgat1 control (Zhang et al. 2009).

Even though both DGAT1 and PDAT1 can catalyze the final DAG acylation, the acyl composition of the resulting TAG will depend on the relative contribution of these two enzymes to the process, provided that identical DAG molecules are utilized as substrate. In Arabidopsis, DGAT1 shows a preference for either saturated or long-chain FAs (>C18) (Katavic et al. 1995). On the contrary, Stahl et al. (2004) demonstrated that Arabidopsis PDAT1 can utilize different phospholipids as acyl donors, exhibiting the highest activity for acyl groups containing several double bonds, epoxy or hydroxyl groups. The prevalence of saturated and long-chain FAs at the third position of the TAG molecule (~70%) suggests that DGAT1 contributes more than PDAT1 in Arabidopsis seeds (Katavic et al. 1995; Zhang et al. 2009).

A second family of DGAT genes (DGAT2) was first identified in Mortierella ramanniana, an oleaginous fungus, and its members have no sequence similarity with
DGAT1 (Lardizabal et al. 2001). DGAT1 and DGAT2 are localized in different regions of the ER, and they appear to have non redundant functions in TAG biosynthesis (Shockey et al. 2006). In vitro assays showed minor differences in substrate selectivity between DGAT1 and DGAT2 enzymes of Vernicia fordii (tung tree). However, when selectivity was evaluated in vivo, DGAT2 was able to catalyze the incorporation of higher amounts of eleostearate, an unusual PUFA that makes up to 80% of tung oil, into TAG (Shockey et al. 2006). Other studies support the idea that DGAT2 is involved in the accumulation of unusual FAs such as epoxy and hydroxy FAs (Burgal et al. 2008; Kroon et al. 2006; Li et al. 2010).

Comparison between transcript levels of DGAT1, DGAT2 and PDAT1 with oil and FA accumulation in developing seeds of several plant species suggests that DGAT1 is the major enzyme for oil accumulation, especially in soybean and Arabidopsis, while DGAT2 and PDAT1 enzymes play important roles in the accumulation of unusual FAs (Li et al. 2010). Also, according to Li et al. (2010), DGAT2 and PDAT1 low expression levels at stages of high oil synthesis would suggest no association between oil accumulation and the function of these two genes in soybean and Arabidopsis seeds. However, Zhang et al. (2009) previously demonstrated that on a dgat1 mutant background, PDAT1 could complement up to 80% the function of DGAT1 in Arabidopsis developing seeds. Similarly, the relative contribution of DGAT1 and DGAT2 to seed oil accumulation and composition is at the moment uncertain as only DGAT1 knockouts
have been characterized in plants (Katavic et al. 1995; Zhang et al. 2009; Zou et al. 1999).

PDCT has been shown to play a major role in the final FA composition of seed TAG as well. A comprehensive analysis of an *Arabidopsis* PDCT mutant indicated that the mutation significantly reduces the transfer of 18:1 into PC for desaturation as well as the reverse transfer of 18:2 and 18:3 into DAG. As a result of this deficiency, there is a 40% decrease in the content of PUFA and a corresponding increase in 18:1 in TAG. Genetic studies on the PDCT mutant indicate that the mutation behaves as a single, recessive Mendelian allele (Lu et al. 2009).

The relative contribution of GPAT, LPAAT and CPT to the overall TAG composition in seeds is not very well known. GPAT seems to have a broad specificity, using saturated, monounsaturated and polyunsaturated FAs (Christie et al. 1991; Lísa and Holcapek 2008). Therefore, acyl composition at position 1 in TAGs would depend on the acyl-CoA availability rather than GPAT specificity (Voelker and Kinney 2001). After studying the substrate selectivity of plant and microbial LPAATs, Brown et al. (2002) concluded that there is little selectivity between 18:1- and 18:2-CoA donor substrates, both incorporated into PA at significant rates, but 16:0 and 18:0 are selected against. However, no clear correlation was observed between the amount of 18:2 or 18:1 at position 2 of TAGs and the major LPAAT activity towards these substrates, leading
Brown et al. (Brown et al. 2002) to conclude that the availability of substrates or subsequent remodeling reactions are more important factors in determining the final oil composition.

**Soybean homologs of AtDGAT1, AtPDAT1 and AtPDCT genes**

Several studies report DGAT, PDAT and PDCT playing significant roles in seed oil quality (Katavic et al. 1995; Li et al. 2010; Lu et al. 2009; Zhang et al. 2009; Zheng et al. 2008), and thus the involvement of those genes in the high 18:3 content in *G. soja* seeds was investigated.

The Arabidopsis genome contains one *DGAT1* (At2g19450) one *PDAT1* (At5g13640) and one *PDCT* (At3g15820) (Hobbs et al. 1999; Lu et al. 2009; Stahl et al. 2004; Zou et al. 1999). In soybean, there are at least two *DGAT1* genes, namely *DGAT1A* (Glyma13g16560) and *DGAT1B* (Glyma17g06120), and their full length cDNAs were cloned from developing seeds (Hildebrand et al. 2008; Wang et al. 2006). These genes showed maximum transcript levels at stages of high TAG synthesis (Hildebrand et al. 2008; Wang et al. 2006). In addition, alignment of the DGAT1A deduced amino acid sequence from seven *G. max* and five *G. soja* accessions showed similarities ranging from 97.8 to 100%; in other words, there is a high conservation of *DGAT1* genes in these two species. Also, the transcript abundance of the *DGAT1A* gene
between *G. max* and *G. latifolia* (one of the 23 perennial wild relatives) accessions had no significant difference. Results from the gene structural and expression studies led Wang et al. (2006) to conclude that *DGAT1A* is not associated to differences in oil content between soybean accessions.

It is not known how many functional *PDAT* and *PDCT* genes are present in the soybean genome. Li et al. (2010) reported that a soybean homolog of *AtPDAT1* (At5g13640) was expressed at very low levels throughout seed development, ruling out this gene as an important player in oil accumulation in soybean seeds. However, there are at least six soybean homologs of At5g13640, whose similarities to *AtPDAT1* ranges from 60 to 86%, and we were not able to identify which of those genes was studied by Li et al. (2010).

**Associations among some seed quality traits in soybean**

The three major factors that determine soybean seed quality are protein and oil contents, protein and oil composition and seed physical appearance. Protein and oil contents as well as their composition are important because of their economical and nutritional value. Also, the relative composition of soybean oil and protein determines their physicochemical properties. Features that are important for seed appearance, such as seed coat and hilum color and seed size play a role in consumer acceptance as well as
food uses. For example, many factors affect the quality of natto, a Japanese fermented soyfood. The preferred soybeans to produce natto are round, small seeded, with high contents of soluble sugars, yellowish seed coat and clear hilum. On the contrary, for soymilk and tofu preparation, large-seeded soybeans with clear hilum and seed coat and high protein contents are preferred (Liu 2008). During selection of soybeans for a particular use, it is important to know the relationships between the different quality attributes.

In a study comprising 10 soybean genotypes selected based on differences in seed size, Liu et al (1995) reported no significant correlations between seed mass and oil content. Seed size, however, was associated to the contents of unsaturated fatty acids: there was a positive correlation with 18:1 and negative correlations with both 18:2 and 18:3. Among the unsaturated FAs, 18:1 showed a significant and strong negative association with 18:3 content in seeds, which agrees with the direction of the association previously reported by Ross et al. (2000). The study by Ross et al. (2000) was carried out with three BC1F2:4 populations developed with an ultra low 18:3 genotype (~1% 18:3) as the common donor parent, and three recurrent parents with low 18:3 content (~2.5%). Each population consisted of 27 1% lines and 27 2% lines (Ross et al. 2000).

Reinprecht et al. (2006) also analyzed the relationship among seed traits in a population of 169 RILs segregating for 18:3. Seed mass and oil content showed no
significant association in three of the six environments of evaluation whereas in the other three environments, the association between these two traits was negative. In addition, while seed mass showed a positive correlation with 18:1, the association with 18:3 was not significant. Ross et al. (2000) also reported no correlation between seed mass and 18:3 contents in soybean seeds. However, the lines with the highest 18:3 levels consistently recorded the heaviest seeds across three different populations. In contrast to the observations by Liu et al. (1995) and Ross et al. (2000), Reinprecht (2006) found no association between seed 18:1 and 18:3.

The associations of seed oil-18:3 contents and seed oil-18:1 contents ranged from not significant to significant with very low r values depending on the environment considered (Reinprecht et al. 2006). In the study carried out by Ross et al. (2000), seed oil and 18:3 contents were directly associated, so that lines with the higher 18:3 contents had significantly higher oil content than lines with the lower levels of the PUFA.

**Associations among some seed quality traits in G. max x G. soja populations**

In many crop species, the utilization of their respective wild relatives has been used to broaden their genetic base (Stalker 1980). Previous studies suggest that G. soja, the wild relative of cultivated soybean, carries all the genes to accumulate the highest 18:3 levels observed in the two species while the cultivated soybean has no genes that
contribute to high 18:3 levels (Pantalone et al. 1997b; Rebetzke et al. 1997; Shibata et al. 2008). The opposite could be said about seed oil and 18:1 contents, i.e., *G. soja* did not make any contribution towards the accumulation of oil and/or 18:1 beyond the levels typically observed in cultivated soybean (Rebetzke et al. 1997; Shibata et al. 2008). Pantalone et al. (1997b) analyzed the relationships between seed traits in three interspecific *G. max* x *G. soja* populations and reported that while seed mass presented a strong negative association with seed 18:3 content ($R^2 = 89\%$), the association between seed mass and oil content was positive ($R^2 = 50\%$). Also, the seed oil-18:3 content and seed oil-18:1 content associations were analyzed in three interspecific populations (Shibata et al. 2008). Seed oil and 18:3 contents presented a negative correlation in both F$_2$ and F$_0$ generations ($r = -0.45$) whereas seed oil and 18:1 levels showed weak or no correlation at all (Shibata et al. 2008).

To summarize, the associations observed between different pairs of seed traits are in general not consistent across studies (Liu et al. 1995; Pantalone et al. 1997b; Reinprecht et al. 2006; Ross et al. 2000; Shibata et al. 2008). This could indicate that such relationships are highly dependent on the genotypes under consideration. For example, seed mass and 18:3 content were inversely associated in two of the studies considered (Liu et al. 1995; Pantalone et al. 1997b) while in the other two studies the association was not significant (Reinprecht et al. 2006; Ross et al. 2000). Also, seed oil and 18:3 contents showed a direct association in Ross et al. (2000), an inverse association
in Shibata et al. (2008) and no association in Reinprecht et al. (2006). Yet, even when analyzing the same set of genotypes, seed traits that were correlated in one environment might show no association in another one, which probably reflects the importance of the interaction between these traits and the growing environment (Reinprecht et al. 2006).

Soybean seed fatty acid profile and environment

It has been very well documented that the environment in which the plant grows influences the expression of soybean seed composition traits (Kumar et al. 2006; Wilson 2004). Temperature during seed development seems to be the most influential environmental factor affecting seed composition. Higher temperatures have been associated to higher oil and lower protein concentrations in soybean seeds (Piper and Boote 1999; Wilcox and Cavins 1992; Wolf et al. 1982). As with protein and oil, temperature also has an effect on fatty acid composition. While 18:2 and 18:3 accumulation is reduced under elevated temperatures, 18:1 levels are significantly increased. The saturated FAs 16:0 and 18:0 were mostly unaffected by changes in temperature (Byfield and Upchurch 2007b; Howell and Collins 1957; Rennie and Tanner 1989; Wilson 2004; Wolf et al. 1982). The increase of 18:1 at the expense of 18:2 and 18:3 in soybeans seeds that develop at elevated temperatures has been linked to the inhibitory effect of high temperatures on the desaturases’ gene expression as well as
enzymatic activity (Byfield and Upchurch 2007a, b; Cheesbrough 1989; Heppard et al. 1996; Tang et al. 2005). Other factors such as rainfall during seed development, latitude, soil fertility or irrigation showed little or no effect on soybean seed FA composition (Howell and Collins 1957; Kumar et al. 2006; Lee et al. 2008).

It is normal and expected to observe some variation in the specific FA composition of the oil of soybean genotypes from year to year and location to location, largely due to temperature differences during seed filling and maturation (Weiss et al. 1952; Wilcox and Cavins 1992). However, if the FA profile of soybean oil is not stable and changes substantially based on the environment, the oil might end up presenting a FA composition different from what was predicted or desired. This lack of stability can result in huge economic losses for soybean farmers and oil companies (Gerde and White 2008). Thus, finding genes affecting FA composition that are stable to environmental factors is very important. For example, N97-3363-4 and N98-4445A, soybean lines characterized as having mid 18:1 contents, showed very low stability in 18:1 levels when grown in different environments while M23, another genotype with mid 18:1 content, was significantly more stable across the same environments (Oliva et al. 2006). Also, soybean genotypes with reduced 18:3 presented very consistent levels of this PUFA across environments (Oliva et al. 2006; Primomo et al. 2002). It is important to note that the genotypes that showed the highest stability were those carrying one mutated gene or a
combination of mutated genes encoding for desaturase enzymes and thus reduced or eliminated enzyme activity.

**Soybean seed fatty acid profile and maturity groups**

Soybean cultivars are classified by maturity group (MG). There are thirteen MGs, with MG 000 being the earliest and MG X being the latest, and they are used to identify the region of adaptation of soybean genotypes. MG zones represent defined areas where a cultivar is best adapted. Cultivars of two to three MGs are often grown successfully within a MG zone. Plant development, from germination through the onset of flowering and maturity, is controlled by photoperiod and temperature (Major et al. 1975). How cultivars respond to these factors determines which MG they fall into. Soybean is a short-day plant species because floral induction happens only when days are shorter than some critical length. Thus, soybeans adapted to northern U.S. have longer minimum day length requirements for the onset of flowering than cultivars adapted to southern regions of the country. Therefore, northern U.S. cultivars flower and mature too early when grown in the shorter photoperiods and warmer temperatures of the south. On the contrary, in southern U.S. cultivars, flower induction and maturation are delayed by the long days and cooler temperatures of northern U.S. (Heatherly and Elmore 2004). As a result of MG effect, the FA profile of a given cultivar planted on the same day can vary
considerably with location, with the highest contents of PUFAs in northern MG zones and the lowest levels observed in the southern MG zones. In the northern locations, the cultivar will flower later, and seed development will take place with cooler ambient temperatures. Contrarily, in the southern locations, the soybean genotype will flower earlier and seeds will grow and mature with higher temperatures. Several QTL studies show the importance of MG in the determination of FA composition, which is evidenced by the fact that QTLs for maturity usually overlap with those for different FAs (Hyten et al. 2004; Reinprecht et al. 2006; Spencer et al. 2004).

**Winter nurseries**

It has become a common practice among public and private breeding programs to use winter nurseries to advance breeding populations for multiple generations in a single year. Winter nurseries allow advancing populations as well as increasing seed quantities during the non-growth season of summer crops such as soybean and corn. Common winter nurseries for U.S. breeding programs are located in tropical territories such as Hawaii and Puerto Rico and countries like Costa Rica. Winter nurseries in temperate locations such as central Argentina and Chile are also widely used, especially by seed companies that prefer to grow their populations under conditions that resemble more the area where their products will be marketed.
Tropical winter nurseries are very good to quickly advance generations since temperatures are adequate for vegetative growth all year around. Also, due to their proximity to the Equator and the short day lengths, soybeans have a shortened cycle, which allows replanting every ~90 days. In addition, also as a result of the short photoperiod, soybeans from all MGs are induced to flower and reach maturity at approximately the same time, and thus genotypes adapted to different latitudes are synchronized to start the reproductive stage.

Usually, before advancing a population to produce one more generation, selection is carried out to keep only the individuals that gather the genotype or phenotype one is breeding for. Because of the unique climate in tropical winter nurseries, especially the constant high temperatures, phenotype-based selection should not be carried out for traits whose expression is influenced by the environment since it can lead to choosing the wrong individuals. This is particularly true when selecting for any of the unsaturated FAs given that temperature influences so strongly desaturases’ transcription and activity (Byfield and Upchurch 2007a, b; Cheesbrough 1989; Heppard et al. 1996; Tang et al. 2005).
**Soybean germplasm collection**

Soybean germplasm collections are listed and/or maintained by several countries (FAO 2010). The headquarters of the U.S. soybean germplasm collection are located at the University of Illinois. The U.S. soybean germplasm collection currently has over 21,000 accessions including soybeans, wild annual soybeans and wild perennial *Glycine* species. As of 2006, there were 718 public cultivars, 16,791 PI accessions, 196 T-lines, 641 isolines, 1,116 wild annual soybeans (*G. soja*) and 919 accessions of wild perennial *Glycine* species (Hymowitz 2008). Information on the accessions in the collection can be found on the GRIN system (http://www.ars-grin.gov/npgs/) maintained by the USDA-ARS. All of the accessions within the U.S. soybean germplasm collection are freely available to scientists worldwide. These soybean lines are an important source for new genes for disease resistance, seed composition modification, tolerance to various abiotic stresses and, in recent years, the source of new genes for high yield. Less than 1% of the lines in the soybean germplasm collection have made any contribution to the commercial soybean varieties grown in the U.S., so it remains a vast reservoir of unknown genetic diversity to be discovered and utilized to improve soybean varieties.
Breeding populations

Several different populations may be utilized for genetic studies within a given plant species, with each type possessing advantages and disadvantages (Burr and Burr 1991; Collard et al. 2005; Schneider 2005). Among the most commonly types used in soybean research are:

$F_2$ populations: derived from selfing or intermating $F_1$ individuals. Their main advantages are that they are easy to construct and require a short time to produce. However, since the linkage established is based on only one meiotic cycle, $F_2$ populations are not good for fine mapping. Also, quantitative traits cannot be precisely mapped using $F_2$ population as each individual is genetically different and cannot be evaluated in replicated trials over locations and years. Thus, the effect of the $G \times E$ interaction on the expression of quantitative traits cannot be precisely estimated. In addition, $F_2$ populations are not long-term populations.

$F_3$ populations: derived from selfing $F_2$ individuals one more generation. $F_3$ populations allow replicate measurements, and can be used for mapping QTLs. However, $F_3$ populations are good for mapping large distances only; since $F_3$ families are still very heterozygous, the precision of the estimates can be low. $F_{2,3}$ families can be used to reconstitute the genotype of $F_2$ plants. Like $F_2$ populations, $F_3$ are not ‘eternal’.
Recombinant Inbred Lines (RILs): inbreeding from individual F₂ plants allows the construction of RILs, each containing a unique combination of chromosomal segments from the original parents. RILs, once they achieved homozygosity, can be multiplied and reproduced indefinitely without further segregation. RILs can be replicated over locations and years and therefore are of immense value in mapping QTL. Also, RILs are very useful in identifying tightly linked markers since they are obtained after several meiotic cycles. The time needed to develop a population of RILs is the major disadvantage.

Near Isogenic Lines (NILs): generated either by repeated selfing or backcrossing the F₁ or BCₙ plants to the recurrent parent. NIL populations are very useful for fine mapping and QTL validation. Once the entire genome of the NILs is exactly like the recurrent parent except in the region around the marker locus, any phenotypic differences among NILs are then most probably due to the QTL linked to the marker locus. Once homozygosity has been achieved, NILs also constitute immortal populations for mapping. The disadvantage of NIL populations is the length of time needed for producing the NILs.

Linolenic acid content in the soybean wild ancestor

Accessions of *Glycine soja* (Siebold & Zucc.), the wild ancestor of cultivated soybean, have been reported to contain as much as 23% 18:3, representing potentially
useful genetic resources for identifying either new alleles or genes that govern 18:3 accumulation (Pantalone et al. 1997a; Wilson 2004). However, very little is known about the genetic regulation of high 18:3 in G. soja. Results of a survey on the average oil content and FA profile of G. max and G. soja accessions in the USDA’s National Plant Germplasm System (NPGS) (http://www.ars-grin.gov/npgs/) are shown in Table 1.1. On average, G. soja produces twice as much 18:3 but almost half the amount of 18:1 and oil as cultivated soybean. Also, G. soja presents higher variability for 18:3 than G. max, whereas G. max surpasses G. soja in variability for oil, 18:1 and 18:2. Analysis of relative 18:1 and 18:2 desaturation patterns, which would reflect ω-6 and ω-3 desaturases’ activity, respectively, in four G. max x G. soja populations suggested that G. soja carries superior alternative FAD2 and FAD3 alleles directing the expression of the high 18:3 trait (Pantalone et al. 1997a). Finally, increasing the expression of the FAD3 genes by genetic engineering allowed the generation of soybean seeds that accumulated 18:3 in excess of 50% of the total oil (Cahoon 2003).
Figures
Figure 1.1. Comparison of the fatty acid composition of various plant-based oils typically used for food. ‘PH Soybean oil’ stands for partially hydrogenated soybean oil.
Figure 1.2. Fatty acid (FA) and triacylglycerol (TAG) synthesis in oilseed plants developing seeds. FA synthesis occurs in plastids. Synthesized FAs enter the cytosolic acyl-CoA pool and are used to form TAG from glycerol-3 phosphate (G3P) in the ER. The FA desaturation (FAD) pathways utilize phosphatidylcholine (PC) as a substrate. FAD2 enzymes catalyze the unsaturation of 18:1-PC to form 18:2-PC; FAD3 enzymes synthesize 18:3-PC from 18:2-PC. PUFA-enriched PC, can be transformed back into diacylglycerol (DAG). Also, the acyl chain at position 2 of PC may undergo exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; ACS, acyl-CoA synthetase; CPT, CDP-choline: 1,2-diacylglycerol cholinephosphotransferase; DGAT, diacylglycerol acyltransferase; DGTA, diacylglycerol transacylase; FAE, FA elongase; FAS, FA synthase; GPAT, phosphatidate phosphatase; LDAT, phospholipid:diacylglycerol acyltransferase; PDCT, phosphatidylcholine:diacylglycerol cholinephosphotransferase; PLA2, phospholipase A2; TS, acyl-ACP thioesterase (modified from Lung and Weselake 2006)
Tables
Table 1.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of *Glycine max* and *Glycine soja* accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean

<table>
<thead>
<tr>
<th></th>
<th>Oil (%)</th>
<th>16:0 (%)</th>
<th>18:0 (%)</th>
<th>18:1 (%)</th>
<th>18:2 (%)</th>
<th>18:3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>G. max</em>&lt;sup&gt;a&lt;/sup&gt;</td>
<td>20 ± 2.1</td>
<td>11 ± 1.0</td>
<td>4 ± 0.6</td>
<td>23 ± 3.8</td>
<td>54 ± 3.7</td>
<td>8 ± 1.5</td>
</tr>
<tr>
<td><em>G. soja</em>&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11 ± 1.4</td>
<td>12 ± 0.9</td>
<td>4 ± 0.5</td>
<td>14 ± 2.4</td>
<td>54 ± 1.7</td>
<td>16 ± 2.2</td>
</tr>
</tbody>
</table>

<sup>a</sup>All observations for *G. max* accessions in the NPGS were used in the calculations (n=16,588)

<sup>b</sup>All observations for *G. soja* accessions in the NPGS were used in the calculations (n=1,242)
Bibliography


Chang NW, Huang PC (1998) Effects of the ratio of polyunsaturated and monounsaturated fatty acid on rat plasma and liver lipid concentration. Lipids 33:481-487


Christopher JP, Pistorius EK, Regnier FE, Axelrod B (1972) Factors influencing the positional specificity of soybean lipoxygenase. Biochim Biophys Acta 289:82-87


FAO (2010) The second report on the state of the world's plant genetic resources for food and agriculture. Rome


Hoshino T, Takagi Y, Anai T (2010) Novel GmFAD2-1b mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with GmFAD2-1a mutant alleles. Breeding Sci 60:419-425


Kitamura K (1991) Spontaneous and induced mutations of seed proteins in soybean (Glycine max L. Merrill). Gamma Field Symposia No 30, Japan, pp 61-69


Lee JD, Oliva ML, Sleper DA, Shannon JG (2008) Irrigation has little effect on unsaturated fatty acid content in soya bean seed oil within genotypes differing in fatty acid profile. J Agron Crop Sci 194:320-324


Li R, Yu K, Hildebrand D (2010) DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. Lipids 45:145-157


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 Biol 10:195


Shockey JM, Gidda SK, Chapital DC, Kuan J-C, Dhanoa PK, Bland JM, Rothstein SJ, Mullen RT, Dyer JM (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


USDA (1996) Industrial uses of agricultural materials situation and outlook report. IUS 6:22-23


Zhang M, Fan J, Taylor DC, Ohlrogge JB (2009) DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21:3885-3901


Chapter 2

Soybean Seed Lipoxygenase Genes: Molecular Characterization and Development of Molecular Marker Assays
Introduction

Lipoxygenases are enzymes belonging to a group of non-heme iron containing proteins widely distributed in plants, fungi and animals (Brash 1999). In plants, lipoxygenases can be found in all organs, representing up to 2% of soybean seed protein (Loiseau et al. 2001). Lipoxygenases catalyze the oxidation of polyunsaturated fatty acids such as linoleic (18:2) and α-linolenic (18:3) to produce unsaturated fatty acid hydroperoxides (Brash 1999; Liavonchanka and Feussner 2006; Song et al. 1990). Products of the lipoxygenase pathway have shown to have a role in a variety of plant processes such as vegetative growth; wounding, herbivore and pathogen attack responses and also in mobilization of storage lipids during germination (Porta and Rocha-Sosa 2002).

In germinating soybean seeds, lipoxygenases were found not to be involved in lipid mobilization (Wang et al. 1999). The lack of substantial oxygenation of polyunsaturated fatty acids during the germination process supports the idea that soybean seed lipoxygenases might have been recruited to function as storage proteins despite their intact but obsolete enzymatic capacity (Siedow 1991; Wang et al. 1999). Regardless of the actual role lipoxygenases may play in the physiology of soybean seeds, lipoxygenase-
free seeds have been demonstrated to develop into normal plants without defect (Hajika et al. 1992; Wang et al. 1999).

Seed lipoxygenases are of significant importance to the food industry, and lipoxygenase-free soybean varieties are in demand. The oxidation products resulting from seed lipoxygenase activity have been associated with the development of undesirable grassy and beany flavors in products containing protein from soybean and other legumes’ seeds (Gerde and White 2008; Robinson et al. 1995; Sessa 1979; Wilson 1996).

Mature soybean seeds contain primarily three lipoxygenases, Lox1, Lox2 and Lox3 (Axelrod et al. 1981). Genetic studies demonstrated that the absence of each enzyme is under the control of three null alleles, lox1, lox2 and lox3, which are inherited as simple recessive alleles (Davies and Nielsen 1986; Hildebrand and Hymowitz 1982; Kitamura et al. 1983; Kitamura et al. 1985). In addition to a single characterized source of the lox2 mutant allele, two single mutant soybean lines for both Lox1 and Lox3 genes have been reported and are available in the USDA’s National Plant Germplasm System (Davies and Nielsen 1986; Hildebrand and Hymowitz 1982; Kitamura et al. 1983; Kitamura et al. 1985; KRDA 2009). The Lox1 and Lox2 loci were found to be in tight genetic linkage on chromosome 13 (LG F), with lox1 and lox2 mutant alleles being in the repulsion phase since they were identified in independent germplasm (Davies and Nielsen 1986;
Hildebrand and Hymowitz 1982; Kitamura et al. 1985). The Lox3 locus, on chromosome 15 (LG E), segregates independently of Lox1 and Lox2 (Davies and Nielsen 1986; Hajika et al. 1992; Kitamura et al. 1985). The original repulsion-phase linkage in independent sources of mutant alleles at the Lox1 and Lox2 loci was broken, resulting in a coupling-phase linkage that eventually led to the development of a triple null lipoxygenase genotype (Hajika et al. 1991; Kitamura 1991). The mechanism involved in the induction of the genotype lacking the three seed lipoxygenases, however, was unknown (Hajika et al. 1991).

The soybean lipoxygenase gene family is thought to have expanded by ancient polyploidy followed by a recent soybean-specific duplication. Consequently, the level of similarity and sequence conservation between regions containing Lox genes is high (Shin et al. 2008). Nineteen soybean Lox genes were analyzed and found to be distributed across only four chromosomes, with each one containing between three to seven Lox genes (Shin et al. 2008).

The molecular basis of the null lox2 genotype was reported to be a T2849A missense mutation, which resulted in the substitution of glutamine for histidine in a highly conserved histidine-rich motif. Although the mutation did not prevent gene transcription, the structure and function of the protein was severely affected, resulting
eventually in its degradation (Wang et al. 1994). The genetic basis of both Lox1- and Lox3-free genotypes has not been previously reported.

The development of seed-lipoxygenase-free soybean cultivars requires accurate evaluation assays to select for the appropriate genotypes. The presence or absence of lipoxygenases in breeding programs has been determined by laborious and time consuming seed phenotyping methods. In addition, these procedures are not able to discriminate between individuals that are homozygous wild type and heterozygous in the same generation. Consequently, tests to study the genetic makeup of the parents must be performed on the progeny, adding both time and expense to the selection process.

Methods to phenotype for seed lipoxygenase status include SDS-polyacrylamide gel electrophoresis, immunological assays, and the relatively faster and simpler colorimetric assays (Kitamura et al. 1983; Narvel et al. 2000; Suda et al. 1995). A single nucleotide polymorphism (SNP) marker that co-segregated perfectly with the Lox2 locus was previously developed, but it was not specific for the lox2 causative mutation (Kim et al. 2004).
The primary objective of this work was to determine the molecular basis of the null mutations in soybean $Lox1$ and $Lox3$ genes. In addition, we wanted to design molecular marker assays perfectly associated with the causative mutations for the three soybean seed $Lox$ genes as well as to investigate the genetic process involved in the generation of a genotype lacking the three soybean seed lipoxygenases, ‘Jinpumkong 2’ (Kim et al. 1997).

In this study, two independent mutations were identified as the genetic causes of the lack of $Lox1$ activity in seeds of $lox1$ mutants, while a common mutation was observed to be responsible for the $Lox3$-free phenotype in all $lox3$ mutants. Perfect molecular marker assays were designed to distinguish mutant from wild type alleles for $Lox1$, $Lox2$ and $Lox3$ genes and showed a complete association between the inheritance of homozygous $lox$ mutant alleles and the lack of lipoxygenase activity. Finally, it was observed that Jinpumkong 2 has the same mutations found in the single mutant sources and that recombination within $Lox1$ locus led to the creation of this triple mutant cultivar.
Materials and Methods

DNA isolation and PCR for sequencing *Lox1* and *Lox3*

Genomic DNA was isolated from ~30 mg dried seed tissue using the DNeasy Plant Mini Kit (Qiagen Sciences Inc., Germantown, MD) and used at 5 to 50 ng per PCR amplification. PCR was carried out using Ex *Taq* according to manufacturer’s recommendations (TaKaRa, Otsu, Shiga, Japan) in a PTC-200 thermocycler (MJ Research/Bio-Rad, Hercules, CA) using the following conditions: 95°C for 5 min followed by 35 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1.5 min. PCR products were analyzed by gel electrophoresis (1% agarose gel) to verify for size and ensure specific amplification. PCR products were isolated with the QIAprep Spin Miniprep kit (Qiagen Sciences Inc.) and sequenced at the University of Missouri DNA Core facility. Primers used for amplification and sequencing are listed in Table 2.1.

Sequence analysis

Sequences were imported into ContigExpress program of Vector NTI Advance 10 (Invitrogen, Carlsbad, CA), trimmed, assembled and manually evaluated for
disagreements between ‘Williams 82’ (Bernard and Cremeens 1988) reference sequence and imported sequence contigs. Putative SNPs and deletions were verified by at least two independent PCR amplifications. To evaluate the effect of changes in coding sequence at the protein level, the program ExPASy translate tool was used (http://ca.expasy.org/tools/dna.html). Sequence alignments were generated using the AlignX software (Invitrogen). \textit{Lox2} and \textit{Lox3} gene models were obtained from the whole soybean genome assembly (http://www.phytozome.net/soybean) and verified for consistency with accessions [GenBank:D13949.1] and [GenBank:EU028322.1], respectively. The \textit{Lox1} gene model was determined through the alignment of genomic and coding sequence [GenBank:EU028320.1] using the AlignX software (Invitrogen) since we found discrepancies in both sequence and predicted gene model between NCBI entries and the whole soybean genome assembly.

**Plant material and development of segregating populations**

Study of the genetic basis of the mutations at the soybean seed \textit{Lox} genes and development of molecular marker assays to identify wild type and mutant alleles were carried out with soybean lines PI 408251, PI 547877 and PI 133226 (single mutants for \textit{Lox1} gene); PI 86023 (single mutant for the \textit{Lox2} gene); PI 205085 and PI 417458 (single mutants for \textit{Lox3} gene) and Jinpumkong 2 as well as breeding lines IA2040LF, 8AR-
56061 and 935F203 (null for the three soybean seed lipoxygenase genes \textit{Lox1}, \textit{Lox2} and \textit{Lox3}). Seeds of single mutant accessions were obtained from the USDA’s National Plant Germplasm System (http://www.ars-grin.gov/npgs/index.html), while seeds of Jinpumkong 2 were donated by Dr. Young-Hyun Hwang (Division of Plant Biosciences, Kyungpook National University, Republic of Korea). Jinpumkong 2 is a Korean cultivar that was released for soyfood uses, such as soymilk, due to its lack of beany taste (KRDA 2009). Seeds of breeding lines IA2040LF, 8AR-56061 and 935F203 were provided by Dr. John Schillinger (Schillinger Genetics).

In the summer of 2006, Jinpumkong 2 was crossed with ‘M23’ (Takagi and Rahman 1996), a line with normal lipoxygenase activity. F\textsubscript{1} seeds were sent to Costa Rica, and F\textsubscript{2} seed was produced during the winter of 2006-2007. F\textsubscript{2} recombinant inbred lines (RILs) were advanced to F\textsubscript{5} by single seed descent method in Costa Rica from the summer of 2007 to early 2008. In 2008, F\textsubscript{5} seed from single F\textsubscript{4} plants was planted at the Lee farm, University of Missouri-Delta Center, MO.

An additional F\textsubscript{3:4} NIL population derived from the cross M23 x Jinpumkong 2 was developed. Seven F\textsubscript{2} plants were randomly selected, and F\textsubscript{3} seed from each F\textsubscript{2} plant was used for further analysis.
Sampling DNA and seed harvest

FTA PlantSaver cards (Whatman Inc., Florham Park, NJ) were used to take DNA samples from 129 plants of the F₅ RIL population and from six plants from each of the seven F₃ NILs, according to manufacturer's recommendations. Progeny of each single plant from which a DNA sample was collected were harvested in the fall of 2008 for lipoxygenase activity assays.

Lox1 genotyping assay

PCR was carried out in a PTC-200 thermocycler (MJ Research/Bio-Rad) using the following conditions: 95°C for 5 min followed by 35 cycles of 95°C for 20 s, 60°C for 20 s, and 72°C for 20 s, followed by a melting curve from 70 to 85°C, with readings taken every 0.1°C. Reactions were carried out in 20 μl containing: 5-50 ng DNA template, primers 5’-ACCGACATCTTAGCGTGCTT-3’ and 5’-AAAAAGGTTGTCTCTATTATGCCAT-3’ (0.375 μM final concentration), buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 μg ml⁻¹ BSA, 200 μM dNTPs], 1.25 μM EvaGreen (Biotium Inc., Hayward, CA), 5% DMSO and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Samples homozygote for the lox1 mutant allele produced a melting peak at 76°C compared to a peak at 81.5°C for wild type plants. Heterozygotes produced both the 76°C and 81.5°C melting peaks. To
verify PCR product size differences, products were separated on 2% agarose gel and approximate size was evaluated by comparison to a Flashgel 100 bp DNA marker (Lonza, Rockland, ME). Expected product size was 130 and 56 bp for wild type and mutant alleles, respectively.

**Lox2 and Lox3 genotyping assays**

SimpleProbe assays are based on the disassociation kinetics of SimpleProbe oligonucleotides (Roche Applied Sciences, Indianapolis, IN) transitioning from a fluorescent bound to target state to a nonfluorescent unbound state. SimpleProbes were designed using the Lightcycler Probe Design Software, version 1 (Roche Applied Sciences) and purchased from Roche Applied Sciences. Asymmetric PCR was carried out to generate more single stranded DNA for probe binding. *Lox2* genotyping assays were performed with an asymmetric mixture of primers [forward (5’-CTGGCCAAAGCTTATGTGGT-3’) and reverse (5’-GCCTTGATGTTCATGGTGCT-3’) primers at 0.1 and 1 µM final concentration, respectively]; the *Lox2* SimpleProbe was 5’-SPC-GTTAAATACTCAAGCGGTGATTGAGCCATTCA-phosphate-3’. *Lox3* genotyping reactions were carried out similarly: forward (5’-CATAGTAGTGGTGGTGGTGC-3’) and reverse (5’-TGTTGAGCCAACTAAGTCGAGA-3’) primers at 1 and 0.1 µM final concentration,
respectively; the \textit{Lox3} SimpleProbe was 5’-SPC- TAATTCCCCCAACGCTGGTTACGCTA-phosphate-3’. Reactions were carried out in a total volume of 20 µl containing template, primers, SimpleProbe (0.4 and 0.2 µM final concentration for \textit{Lox2} and \textit{Lox3}, respectively), buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 µg ml\(^{-1}\) BSA, 200 µM dNTPs], 5% DMSO and 0.2X Titanium \textit{Taq} polymerase (BD Biosciences). Genotyping reactions were performed using a Lightcycler 480 II real time PCR instrument (Roche Applied Sciences), using the following PCR parameters: 95°C for 5 min followed by 40 cycles of 95°C for 20 s, 60°C for 20 s, 72°C for 20 s and then a melting curve from 55 to 72°C in the \textit{Lox2} assay and from 58 to 73°C in the \textit{Lox3} assay. Fluorescence was read after each cycle and every 0.1°C during the melting curve analysis (MCA). A mismatch between the probe and the target sequence results in altered disassociation kinetics of the probe. Consequently, each genotype produces a characteristic melting profile, as measured by the melting temperature (\(T_m\)) of the first negative derivative of the fluorescence value change with increasing temperature. In the SimpleProbe assays, \textit{Lox2} alleles produced a \(T_m\) of 64°C and \textit{lox2-a} alleles produced a \(T_m\) of 68°C; \textit{Lox3} alleles produced a \(T_m\) of 66°C and \textit{lox3-a} produced a \(T_m\) of 62°C.
Phenotyping assays

Colorimetric assays were used to detect the activity of Lox1 and Lox3 in two segregating populations. Phenotypic assays were performed with F_{4:6} and F_{3:4} seeds of the RIL and NIL populations, respectively. Dyes, substrate and test solutions were prepared as outlined by Narvel et al. (2000), with slight modifications: 1) Reactions were performed in 96 deep well 2 ml plates (Fisherbrand, Denmark), 2) The volume of soaking water as well as of test solutions were halved and 3) Lox3 assays were scored after 15 min. Four individual progeny seeds from each plant were tested. Preliminary experiments showed that for the Lox3 assay, scoring samples as long as 30 min after addition of test solution was very difficult as differences in coloration of samples were not always clear. However discrimination of samples was best between 5 and 15 min after adding test solution. Also, for the Lox3 test, in addition to triple-null seeds, single mutants for Lox3 were included as controls. The inclusion of the additional control was needed to correct lox3 single mutant samples for non-specific bleaching of the test solution.
Results

Genomic arrangement of soybean seed lipoxygenase genes

With the availability of the whole soybean genome assembly (http://www.phytozome.net/soybean), we were able to determine the disposition of soybean seed lipoxygenase genes in the genome (Fig. 2.1). *Lox1* and *Lox2* (Glyma13g42320 and Glyma13g42310 in the phytozome soybean browser, respectively) are located on chromosome 13/LG F along with two other genes with lipoxygenase domains (Glyma13g42330 and Glyma13g42340) in a cluster that spans 41.54 kb. The distance between *Lox1* and *Lox2* is only 2,998 bp (Fig. 2.1a). *Lox3* (Glyma15g03030) is located on chromosome 15/LG E in another lipoxygenase-rich gene cluster that covers 23.74 kb and includes two other genes that also have lipoxygenase domains (Glyma15g03040 and Glyma15g03050) (Fig. 2.1b). The gene models for *Lox1*, *Lox2* and *Lox3* represent 3,830, 3,939 and 4,095 bp of sequence, respectively, from start to stop codon, and each gene contains nine exons.
Genetic mutations in *Lox1* and *Lox3* genes

We used gene-specific PCR to amplify and sequence *lox1* from PI 408251 and PI 133226 (*lox1*, *Lox2*, *Lox3*) and *lox3* from PI 205085 and PI 417458 (*Lox1*, *Lox2*, *lox3*). Genomic sequence from cultivar Williams 82 (*Lox1*, *Lox2*, *Lox3*) was used as reference sequence (http://www.phytozome.net/soybean).

For the *lox1* allele in PI 408251, we identified a 74 bp deletion that starts in exon 8 at genomic position 2752, relative to start codon (GenBank accession GQ227538). The deletion creates an immediate premature stop codon and a truncated protein of 524 residues, compared to the reference Lox1 protein containing 839 amino acid residues (Fig. 2.2). For the *lox1* allele in PI 133226, a nonsense mutation, C2880A relative to start codon, results in an S568STOP change (Fig. 2.2; GenBank accession GQ227539). The *lox1* 74 bp deletion found in PI 408251 was absent in PI 133226, and PI 408251 did not contain the C2880A mutation found in the PI 133226 *lox1* allele. No other changes were identified in the *lox1* alleles compared to the reference Williams 82 *Lox1* sequence.

Characterization of the *lox3* allele in PI 205085 and PI 417458 revealed an identical mutation in both sources: a single base deletion of a guanine in a run of five guanine nucleotides, within exon 1, from position 97 to 101 relative to start codon, compared to the wild type *Lox3* reference sequence (GenBank accessions GQ227541 and GQ227542). This guanine deletion results in a frame shift at position 101 that prematurely truncates
the protein after only 41 amino acids (Fig. 2.3). No other changes were identified in the *lox3* allele compared to the reference Williams 82 *Lox3* sequence.

We will refer to the *lox1* mutant alleles from PI 408251 and PI 133226 as *lox1-a* and *lox1-b*, respectively. Similarly, we will refer to the *lox3* mutant allele found in PI 205085 and PI 417458 as *lox3-a*. Finally, the *lox2* mutant allele present in PI 86023, previously reported by Wang et al. (1994), we will term *lox2-a*.

**Development of molecular marker assays**

Molecular marker assays were designed to distinguish the mutant *lox1-a*, *lox2-a*, and *lox3-a* alleles from the wild type alleles for *Lox1*, *Lox2* and *Lox3* genes (Fig. 2.4). Due to the high degree of sequence similarity among the genes, we performed gene-specific PCR and then discriminated between wild type and mutant alleles of each gene using the characteristic melting profile produced from SimpleProbe disassociation for *Lox2* and *Lox3* assays (see Materials and Methods). For the *Lox1* assay, gene-specific primers were designed to flank the region containing the 74 bp deletion in *lox1-a* mutants and discriminate genotypes using Melting Curve Analysis (MCA) of the amplification products in the presence of the fluorescent dye EvaGreen or based on resolution of product size differences. MCA using DNA from plants homozygous for the *lox1-a* allele gave a distinctive melting peak of 76°C compared to a peak at 81.5°C for DNA from
Williams 82. Heterozygotes produced both the 76°C and 81.5°C melting peaks (Fig. 2.4a). For the lox2-a and Lox2 alleles, MCA using the SimpleProbe assay yielded melting peaks at 68°C and 64°C for individuals that were homozygous, respectively. Heterozygous lox2-a/Lox2 plants showed both melting peaks at 68°C and 64°C (Fig. 2.4b). Finally, a SimpleProbe assay designed to distinguish Lox3 from lox3-a showed, in a MCA, melting peaks at 66°C and 62°C for homozygous wild type and mutant plants, respectively, while both melting peaks were observed in heterozygous individuals (Fig. 2.4c). The molecular marker assays reliably distinguished homozygous mutant, homozygous wild type, and heterozygous genotypes for the Lox1, Lox2 and Lox3 genes.

Genetic recombination within Lox1 gene locus in Jinpumkong 2

The hurdle to the development of a lox1, lox2, lox3 triple mutant genotype was the tight linkage between the Lox1 and Lox2 loci. A single lipoxygenase-free soybean line was recovered after a complex experiment that included gamma-ray mutagenesis of mixed F2 seeds of crosses between lox1, lox3 and lox2, lox3 double mutants (Hajika et al. 1991). However, the mechanism involved in the induction of the genotype lacking the three seed lipoxygenases was unknown (Hajika et al. 1991). Subsequently, one additional triple mutant lipoxygenase genotype was developed from mutagenesis of a
**loxl, lox2, lox3** double mutant, but that line appears to be unrelated to the cultivar Jinpumkong 2 and the lipoxygenase-free breeding lines available for investigation (Kitamura 1991).

Sequence characterization and genotyping of the Jinpumkong 2 seed lipoxygenase genes enabled the reconstruction of the events leading to the combination of **loxl, lox2, and lox3** mutant alleles. The Jinpumkong 2 **lox2** allele was genotyped to contain the H532Q missense mutation from PI 86023. The Jinpumkong 2 **lox3** allele was identical to the **lox3-a** allele found in PI 205085 and PI 417458 (GenBank accession GQ227544).

Characterization of **loxl** from Jinpumkong 2 revealed a unique **loxl** mutant allele. Although Jinpumkong 2 contained the **loxl** 74 bp deletion identical to that in PI 408251, an additional seven SNPs and a 3 bp deletion were also present in the 5’ end of the Jinpumkong 2 **loxl** gene (Fig. 2.5; GenBank accession GQ227543).

The **Lox2** gene resides ~3 kb upstream of the **Loxl** gene (Fig. 2.1). The source of the Jinpumkong 2 **lox2** mutation was PI 86023 (**Loxl, lox2, Lox3**), which produces active Lox1 enzyme. We verified that the PI 86023 **Loxl** allele does not contain either the PI 408251 74 bp deletion or the PI 133226 C2880A mutation. However, sequence characterization of the PI 86023 **Loxl** gene revealed a number of missense, silent and intron SNPs compared to the reference Williams 82 **Loxl** sequence and the PI 408251 and PI 133226 **lox1** mutant alleles (Fig. 2.5; GenBank accession GQ227540). A contiguous subset of these **Loxl** polymorphisms from PI 86023 (seven SNPs and a 3 bp
deletion) was present in the Jinpumkong 2 loxl allele upstream from the 74 bp deletion (Fig. 2.5).

The existence of a novel mutant allele of loxl in Jinpumkong 2 containing elements of both the PI 86023 Lox1 gene and the PI 408251 loxl gene is consistent with a recombination event which occurred within the Lox1 locus that broke the repulsion linkage to couple the two mutant loxl and lox2 alleles from independent sources in a single line (Fig. 2.5). The last shared Lox1 SNP present in both Jinpumkong 2 and PI 86023 is at position 472 relative to start codon, and sixteen SNPs and two single base deletions beginning at position 1185 relative to start codon are present only in the PI 86023 Lox1 allele (Fig. 2.5). The novel mutant allele at the Lox1 locus from Jinpumkong 2 will be referred to as loxl-c.

Finally, genotyping of three different Lox-null soybean breeding lines (i.e., IA2040LF, 8AR-56061 and 935F203) that are currently used to develop marketable lipoxygenase-free soybean varieties revealed they contain identical combination of lox mutations to that found in Jinpumkong 2.

Together with the previously identified mutation in Lox2 present in the Lox2-free accession PI 86023 (Wang et al. 1994), there are now a total of five seed lox mutant alleles, and three of them have perfect molecular markers (i.e., specific to the causative mutation), one for each of the three seed Lox loci (Table 2.2).
Association analysis for seed lipoxygenase phenotype and *Lox1* and *Lox3* genotypes

To determine the consistency of the molecular marker assays developed to distinguish mutant from wild type alleles for *Lox1, Lox2* and *Lox3* genes, we carried out an association analysis between the genotype of individuals of two segregating populations and the seed lipoxygenase phenotype of their progeny. The segregating populations were derived from a cross between M23 (*Lox1, Lox2, Lox3*) and Jinpumkong 2 (*lox1, lox2, lox3*). Population 1 was an F4:5 RIL population containing 129 individuals, and population 2 was a set of seven F2:3 NILs with 42 individuals; two individuals of the NIL population were not analyzed. We genotyped samples from these populations for the *lox1, lox2*, and *lox3* alleles and phenotyped for seed Lox1 and Lox3 activity in the progeny (see Materials and Methods). Results indicated a perfect association between the inheritance of homozygous *lox1* or *lox3* alleles and the lack of either Lox1 or Lox3 activity, respectively (Table 2.3). It was also observed that the *Lox3* locus segregated independently of the *Lox1* and *Lox2* loci. On the contrary, *Lox1* and *Lox2* loci were inherited as if they were on the same locus, i.e., no independent segregation occurred, but a tight coupling-phase linkage existed (data not shown).
Recent phylogenetic analysis has revealed that the lipoxygenase gene family expanded by two rounds of whole genome duplication, resulting in gene members with a high degree of sequence and structure similarity (Shin et al. 2008). In our study, we characterized three different seed-expressed lipoxygenase genes (Lox1, Lox2 and Lox3), which possess a high degree of identity at the amino acid level (81.1% between Lox1 and Lox2, 70.7% between Lox1 and Lox3 and 74.5% between Lox2 and Lox3). Based on the soybean genome assembly (http://www.phytozome.net/soybean), we determined that the distance between Lox1 and Lox2 loci is ~3 kb. The minute separation between the two loci explains the tight linkage reported in previous genetic studies (Davies and Nielsen 1986; Hajika et al. 1992; Kitamura et al. 1985). Hajika et al. (1991) were able to obtain a soybean line lacking all the seed lipoxygenases by gamma-ray irradiation of F2 plants from a cross between two double mutants: Lox1- and Lox3-free x Lox2- and Lox3-free. However, the mechanism involved in the induction of the triple mutant was unknown to the authors. Point or gene mutation, chromosomal deletion and crossing-over are among the possible events that could have allowed them to obtain the triple mutant line.
Jinpumkong 2 is a null for the three seed lipoxygenases (Kim et al. 1997). By means of gene-specific PCR and sequence comparison, we determined that no new mutation or chromosomal deletion containing any of the three seed Lox genes took part in the triple null quality of Jinpumkong 2, leaving genetic recombination between Lox1 and Lox2 loci as a viable alternative. Comparison of Lox1 sequence between Jinpumkong 2 and PI 86023 (donor of lox2-a allele) indicates that these two accessions share unique sequence polymorphisms in the 5’ end of the Lox1 gene that are not present either in Williams 82 (reference sequence) or in lox1 single mutants (PI 408251 and PI 133226). Moreover, after the last common polymorphism between Jinpumkong 2 and PI 86023 at position 472, relative to start codon, the sequence between the triple mutant and PI 408251 becomes identical, while the PI 86023 Lox1 sequence contains additional unique polymorphisms. This strongly implicates that genetic recombination within the Lox1 locus, creating a novel lox1 mutant allele, is the mechanism that allowed the combination of lox1, lox2 and lox3 alleles in Jinpumkong 2 despite the donor repulsion-phase linkage between the alleles at the Lox1 and Lox2 loci (Lox1, lox2 and lox1, Lox2).

Mutant soybean lines lacking Lox1 (Hildebrand and Hymowitz 1981, 1982), Lox2 (Davies and Nielsen 1986; Kitamura et al. 1985) and Lox3 (Kitamura et al. 1983) have been identified, and the inheritance of the activity of each of the enzymes was studied. Also, the molecular basis for Lox2 deficiency was previously investigated (Wang et al. 1994) and reported to be a missense T2849A mutation, which results in a H532Q
substitution in a highly conserved histidine-rich motif. Such a substitution does not prevent the expression of the gene, but affects protein structure, function and stability (Wang et al. 1994). This is the first report describing the molecular basis for Lox1- and Lox3-null phenotypes.

Seed lipoxygenases affect food quality since they are involved in the production of undesirable grassy and beany aroma and flavor in soybean-containing foods (Gerde and White 2008). The genetic elimination of seed lipoxygenases represents a solution to this problem. To develop Lox-free soybean cultivars, accurate evaluation assays to select for appropriate genotypes are required. The methods that have been used to characterize soybean lines for presence/absence of seed lipoxygenases include detection of the Lox proteins in single dimension SDS-polyacrylamide gels using electrophoresis and crude seed extracts with either total protein staining or by Westerns with specific lipoxygenase antibodies, and colorimetric enzymatic assays, based on the pH of the test solution for which the corresponding lipoxygenase has optimum activity (Kitamura et al. 1983; Narvel et al. 2000; Suda et al. 1995). While these methods test phenotypes, none of them analyze the genetic makeup of evaluated soybean lines. In each case of the phenotypic assays, samples which have a heterozygous genotype are not distinguishable from homozygous wild-type individuals. We report herein the development of perfect molecular marker assays for each of the three soybean seed Lox genes. Association
analysis showed a complete agreement between genotype and phenotype of evaluated individuals, validating the robustness of our assays.

In our hands, colorimetric assays were consistent except when evaluating Lox2 activity. We were unable to assign the correct phenotype to several known wild type samples for Lox2, which failed to react in the test solution. One factor affecting the accuracy of the colorimetric assay for Lox2 was seed age: the older the seed, the more inconsistent the test was. However, even with recently harvested seeds, the assay remained inconsistent. Seed age had no impact on the call accuracy of our molecular marker assays. An additional advantage of our molecular markers is that homozygous mutant, heterozygous, and homozygous wild type genotypes could be clearly differentiated in the same generation. On the contrary, for the phenotypic assays, progeny tests must be performed to definitely discern the genetic makeup of individuals, representing a considerable increase in both the number of samples to analyze and time.

As a result of this work, three lox mutant alleles, one for each of the soybean seed Lox genes, can be easily and quickly detected by means of perfect molecular markers. Furthermore, the potential of these assays becomes more evident by the fact that these lox mutant alleles were found to be present in lipoxygenase-free soybean breeding lines currently in use. The utilization of molecular markers herein reported will allow soybean breeders to directly select for mutant alleles in early generations of segregating
populations allowing more efficient introgression of the lipoxygenase-free trait in their soybean varieties.
Conclusions

In this work, the genetic basis of mutations in \textit{Lox1} and \textit{Lox3} soybean genes were investigated. Two independent mutations, a 74 bp deletion and a C2880A nonsense mutation, were responsible for the premature truncation of the Lox1 protein in mutants. In contrast, all \textit{lox3} mutants showed a single base deletion introducing a frame shift at position 101 which resulted in a premature stop codon. Co-dominant molecular marker assays perfectly associated with the causative mutations were developed for \textit{Lox1}, \textit{Lox2} and \textit{Lox3} genes, providing a quick, inexpensive and accurate means to assess the genetic makeup of individuals. Finally, it was elucidated that genetic recombination was the mechanism that broke the tight repulsion-phase linkage between \textit{Lox1} and \textit{Lox2} loci, allowing the combination of three independent \textit{lox} mutant alleles in the lipoxygenase-free variety Jinpumkong 2.
Figures
Fig. 2.1. Genomic arrangement of soybean seed lipoxygenase genes. Cluster of Lox genes in a chromosome 13 and b chromosome 15. Genes are represented by white boxes. Arrows in white boxes indicate orientation of genes. Lox genes covered in this work are denoted with their name, i.e., Lox1, Lox2 and Lox3; Lox genes not covered in this study are denoted as Lox. Code below white boxes representing genes corresponds to gene name in the whole soybean genome assembly (http://www.phytozome.net/soybean). Gene clusters are drawn to scale.
Fig. 2.2. *Lox1* gene structure and polymorphisms observed in *lox1* mutant alleles. Williams 82 (wild type), PI 408251 (*lox1* single mutant) and PI 133226 (*lox1* single mutant) alleles at the *Lox1* locus. PI 408251 (*lox1-a*) presents a 74 bp deletion starting at genomic position 2752 relative to start codon. The deletion creates an immediate premature stop codon and a truncated protein of 524 residues compared to the reference *Lox1* protein of 839 residues. PI 133226 (*lox1-b*) has a C2880A nonsense mutation relative to start codon that results in a S568STOP change and premature truncation of *Lox1* protein to 567 amino acids. Boxes represent exons; lines between boxes represent introns; boxes with dashed lines represent exons with untranslated regions in predicted *Lox1* proteins from *lox1* mutant alleles.
Fig. 2.3. *Lox3* gene structure and polymorphisms observed in *lox3* mutant allele. Williams 82 (wild type) and PI 205085 and PI 417458 (*lox3* single mutants) alleles at the *Lox3* locus. PI 205085 and PI 417458 (*lox3*-a) present a single guanine deletion in a run of five guanines starting at genomic position 97 relative to start codon. As a result, a frame shift is introduced at position 101 (bold nucleotide) that prematurely truncates the protein at position 42. Amino acid residues resulting from the frame shift are in bold. Numbers in the left margin of genomic and protein sequences indicate nucleotide and amino acid residue number, respectively. Boxes represent exons; lines between boxes represent introns, boxes with dashed lines represent exons with untranslated regions in predicted Lox3 protein from *lox3*-a mutant allele.
Fig. 2.4. Genotyping results using molecular marker assays for soybean seed lipoxygenase genes. a Typical results for the \textit{Lox1} genotyping assay. Homozygous \textit{Lox1} and PI 408251-derived \textit{lox1} individuals give peaks at 81.5 and 76°C, respectively. Heterozygotes show both peaks. b Typical results for the \textit{Lox2} genotyping assay. Homozygous \textit{Lox2} and \textit{lox2} individuals give peaks at 64 and 68°C, respectively. Heterozygotes show both peaks. c Typical results for the \textit{Lox3} genotyping assay. Homozygous \textit{Lox3} and \textit{lox3} individuals give peaks at 66 and 62°C, respectively. Heterozygotes show both peaks.
Fig. 2.5. Comparison of Jinpumkong 2, PI 86023 and PI 408251 alleles at the Lox1 locus. Jinpumkong 2 contains lox1-c, a unique lox1 mutant allele with elements of both the PI 86023 Lox1 gene and the PI 408251 lox1-a gene. Jinpumkong 2 has the lox1 74 bp deletion identical to that in PI 408251 as well as seven SNPs and a 3 bp deletion present in the 5’ end of the PI 86023 Lox1 gene. Asterisks represent SNPs found in PI 86023. The symbol Δn represents a deletion of ‘n’ nucleotides when ‘n’ is higher than one. Boxes represent exons; lines between boxes represent introns. Black boxes represent genomic regions from PI 86023; white boxes represent genomic regions from PI 408251; light grey boxes and lines represent recombination region within Lox1 locus; boxes with dashed lines represent exons with untranslated regions in predicted Lox1 proteins from Jinpumkong 2 and PI 408251
Table 2.1. Primers used for amplification and sequencing of \textit{Lox1} and \textit{Lox3} genes. Primers are listed in 5’→3’ configuration.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence</th>
<th>Reverse</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lox1</strong> gene</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lox1f1</td>
<td>CATGGGGGCCTTAATGACACT</td>
<td>Lox1r1</td>
<td>GGATGGTTCTTGGGTCTTGA</td>
</tr>
<tr>
<td>Lox1f2</td>
<td>GTGCTCATCGTACCAGCaa</td>
<td>Lox1r2</td>
<td>GCGCACGCTTTTGTAAAGTT</td>
</tr>
<tr>
<td>Lox1f3</td>
<td>GCCGTAGTGTCTCCTCCAG</td>
<td>Lox1r3</td>
<td>CTGTTGGACCTCGACAGTT</td>
</tr>
<tr>
<td>Lox1f4</td>
<td>CCAGCACCACCTCTGAGTA</td>
<td>Lox1r4</td>
<td>CACGAATTACGCAGGATT</td>
</tr>
<tr>
<td>Lox1f5</td>
<td>TTGGGCAACCCAGATAAGAG</td>
<td>Lox1r5</td>
<td>GCTTCGATGTCTACCTCGT</td>
</tr>
<tr>
<td>Lox1f6</td>
<td>AAATCCCTGCGTAATTCGTG</td>
<td>Lox1r6</td>
<td>ATTCGAGCCACCTACCTCCAG</td>
</tr>
<tr>
<td>Lox1f7</td>
<td>CTGCCAAGACTTATGCGACA</td>
<td>Lox1r7</td>
<td>GGCTAATGAGAGTCGGCAAC</td>
</tr>
<tr>
<td>Lox1f8</td>
<td>ATGGAGCCATTCGCTCATAGC</td>
<td>Lox1r8</td>
<td>TCGGATCATTTGCTCTCCTC</td>
</tr>
<tr>
<td>Lox1f9</td>
<td>ACCGACATCTTACGCTGTT</td>
<td>Lox1r9</td>
<td>TTTGCAAAACCACAAACAAAC</td>
</tr>
<tr>
<td><strong>Lox3</strong> gene</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lox3f1</td>
<td>CAGTTTGGTTGACATTTGG</td>
<td>Lox3r1</td>
<td>TGGCATTGAAATCCACGAA</td>
</tr>
<tr>
<td>Lox3f2</td>
<td>AAGATGCTTTGGGGTCTTCT</td>
<td>Lox3r2</td>
<td>TGGAGCTGTTGTCACCTTG</td>
</tr>
<tr>
<td>Lox3f3</td>
<td>CAAGTGACACCAGAGCTCCA</td>
<td>Lox3r3</td>
<td>GAATTGGCTTTTCCAACCAAA</td>
</tr>
<tr>
<td>Lox3f4</td>
<td>AGAACGCACAGAGTGGGAAA</td>
<td>Lox3r4</td>
<td>CCTTTGGGAGGTTGCATT</td>
</tr>
<tr>
<td>Lox3f5</td>
<td>AAGCATTTCTGAGTTGTCCTT</td>
<td>Lox3r5</td>
<td>CGTGCACAGAGGTTGCTTC</td>
</tr>
<tr>
<td>Lox3f6</td>
<td>TGGGGAAGGTATTCTCTGGA</td>
<td>Lox3r6</td>
<td>CGTGCACAGAGGTTGCTTC</td>
</tr>
<tr>
<td>Lox3f7</td>
<td>AGGTGGGTACAGGAGAGAAC</td>
<td>Lox3r7</td>
<td>TGCAAATTAACGATTTGTC</td>
</tr>
<tr>
<td>Lox3f8</td>
<td>AAGGGGCTACTGATGAGGT</td>
<td>Lox3r8</td>
<td>TTCATGGACCACATGTTGA</td>
</tr>
<tr>
<td>Lox3f9</td>
<td>CTTTGCAACCCAGGTCAAGT</td>
<td>Lox3r9</td>
<td>CTCCCCCTACGAGGATAAGG</td>
</tr>
<tr>
<td>Lox3f10</td>
<td>TTTTCTTTGGCAAACACAGTTC</td>
<td>Lox3r10</td>
<td>CCCCCTACGAGGATAAGGAA</td>
</tr>
<tr>
<td>Lox3f11</td>
<td>TTTGTCAATACCGAGTGTCA</td>
<td>Lox3r11</td>
<td>TTCCCACTCTTTGGCCTTC</td>
</tr>
<tr>
<td>Lox3f12</td>
<td>CCGAGTGTCAATTAGTGGTTT</td>
<td>Lox3r12</td>
<td>ACACCTATTCCGAGTACCT</td>
</tr>
</tbody>
</table>
**Table 2.2.** Listing of mutant alleles at the three soybean seed *Lox* genes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Lox1(^a)</th>
<th>Lox2</th>
<th>Lox3</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 133226</td>
<td>lox1-b(^c2880A)</td>
<td>Lox2</td>
<td>Lox3</td>
<td>this work</td>
</tr>
<tr>
<td>PI 408251</td>
<td>lox1-a(^c2880A)</td>
<td>Lox2</td>
<td>Lox3</td>
<td>this work</td>
</tr>
<tr>
<td>PI 547877</td>
<td>lox1-a(^c2880A)</td>
<td>Lox2</td>
<td>Lox3</td>
<td>this work</td>
</tr>
<tr>
<td>PI 86023</td>
<td>Lox1</td>
<td>lox2-a(^T2849A)</td>
<td>Lox3</td>
<td>Wang et al. (1994)</td>
</tr>
<tr>
<td>PI 417458</td>
<td>Lox1</td>
<td>Lox2</td>
<td>Lox3-a(^G101\Delta)</td>
<td>this work</td>
</tr>
<tr>
<td>PI 205085</td>
<td>Lox1</td>
<td>Lox2</td>
<td>Lox3-a(^G101\Delta)</td>
<td>this work</td>
</tr>
<tr>
<td>Jinpumkong 2</td>
<td>lox1-c(^c74bp)</td>
<td>lox2-a(^T2849A)</td>
<td>Lox3-a(^G101\Delta)</td>
<td>this work</td>
</tr>
</tbody>
</table>

\(^a\)Superscript designates type of mutation;\(^b\)C2880A = C to A change at position 2880,\(^d\)Δ74bp = 74 bp deletion starting at position 2752,\(^e\)T2849A = T to A change at position 2849,\(^f\)G101Δ = deletion of a guanine resulting in a frame shift at position 101. All mutations are relative to start codon, where A = 1. Uppercase designates wild-type alleles. Mutant *lox* alleles molecularly characterized in this work are indicated by black background and white fonts. Novel PI 86023 *Lox1* allele, characterized in this work, is indicated by light grey background and black fonts.
Table 2.3. Colorimetric determination of presence (+) or absence (-) of Lox1 and Lox3 activity in progeny seeds derived from genotyped parental plants

<table>
<thead>
<tr>
<th>Parent genotype</th>
<th>Observed n° of progeny seeds with each possible Lox phenotype a</th>
</tr>
</thead>
<tbody>
<tr>
<td>N° of plants</td>
<td>Genotype</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Lox1 Lox1 Lox3 Lox3</td>
</tr>
<tr>
<td>8</td>
<td>Lox1 Lox1 Lox3 lox3</td>
</tr>
<tr>
<td>32</td>
<td>Lox1 Lox1 lox3 lox3</td>
</tr>
<tr>
<td>2</td>
<td>Lox1 lox1 Lox3 Lox3</td>
</tr>
<tr>
<td>3</td>
<td>Lox1 lox1 Lox3 lox3</td>
</tr>
<tr>
<td>10</td>
<td>Lox1 lox1 lox3 lox3</td>
</tr>
<tr>
<td>28</td>
<td>lox1 lox1 Lox3 Lox3</td>
</tr>
<tr>
<td>14</td>
<td>lox1 lox1 Lox3 lox3</td>
</tr>
<tr>
<td>45</td>
<td>lox1 lox1 lox3 lox3</td>
</tr>
</tbody>
</table>

aFor phenotype determination, four seeds of each parental plant were evaluated. Homozygous wild type genotypes are indicated by white background and black fonts. Homozygous mutant genotypes are indicated by black background and white fonts. Heterozygous genotypes are indicated by light grey background and black fonts.

bχ² for 3 Lox3(+) : 1 Lox3(-) = 2.67; not significant at p<0.05

cχ² for 3 Lox1(+) : 1 Lox1(-) = 0.17; not significant at p<0.05

dχ² for 9 Lox1(+) Lox3(+) : 3 Lox1(-) Lox3(-) : 3 Lox1(-) Lox3(+) : 1 Lox1(-) Lox3(-) = 6.14; not significant at p<0.05

eχ² for 3 Lox1(+) : 1 Lox1(-) = 0.00; not significant at p<0.05

fχ² for 3 Lox3(+) : 1 Lox3(-) = 0.38; not significant at p<0.05


Kitamura K (1991) Spontaneous and induced mutations of seed proteins in soybean (Glycine max L. Merrill). Gamma Field Symposia No 30, Japan, pp 61-69


Chapter 3

Genetics of High Linolenic Acid in Seed Oil of *Glycine soja*:

Role of ω-6 and ω-3 Desaturase Genes
Introduction

Importance of soybean and soybean oil

Soybean [Glycine max (L.) Merr.] is the leading oilseed crop produced and consumed in the world, representing 58% and 43% of world oilseed production and consumption, respectively (SoyStats 2011; Wilcox 2004). Soybean is the United States' second largest crop in cash sales and the number one value crop export (SoyStats 2011). Economic importance of soybean oil is given by its utilization in the manufacture of both edible and industrial products. Approximately 70% of the edible fats and oils consumed in the US as salad and cooking oil, baking and frying fats and margarine and spreads derive from soybean. Also, ~14% of the soybean oil utilized in the US is for industrial purposes such as the production of biodiesel, inks, paints, varnishes, resins and plastics (Cahoon 2003; SoyStats 2011).

Lipids represent one of the most important components in soybean seeds, comprising about 20% of their weight (Gerde and White 2008). The predominant fatty acid (FA) composition of commodity soybean oil is about 11% palmitic acid (16:0), 4% stearic acid (18:0), 23% oleic acid (18:1), 54% linoleic acid (18:2) and 8% linolenic acid (18:3) (Fehr 2007; Hill et al. 2008; Wilson 2004). For reference, the first number of these FA abbreviations indicates the number of carbon atoms, and the second number represents the number of double bonds in the molecule. 16:0 and 18:0 are saturated FAs,
18:1 is a monounsaturated FA, and 18:2 and 18:3 are polyunsaturated FAs (PUFA). Also, 18:2 and 18:3 are ω-6 and ω-3 types of FAs, respectively, which are essential to humans and all animals. The term essential refers to any compound that is required for life but cannot be synthesized by the organism and therefore must be incorporated with diet.

**Potential uses of soybean oil enriched in ω-3 fatty acids**

The balance between dietary ω-3 and ω-6 FAs strongly affects their function in the body. Currently, nutritionists are concerned over the reduction of the nutritionally desirable 18:3 levels in western diets and more particularly the ratio of ω-3:ω-6 FAs, which should ideally be 1:2.3 (Gerde and White 2008; Kris-Etherton et al. 2000). According to reports of the University of Maryland Medical Center (UMMC), the typical American diet tends to contain 14-25 times more ω-6 than ω-3 FAs (http://www.umm.edu/altmed/articles/omega-3-000316.htm).

ω-3 FAs play a crucial role in brain function as well as normal growth and development. They also reduce inflammation and may help lower risk of diseases such as heart disease, cancer, type II diabetes and arthritis (UMMC 2011). ω-3 FAs are highly concentrated in the brain and appear to be important for cognitive (brain memory and performance) and behavioral function (UMMC 2011). Because of these health benefits, elevated 18:3 soybean genotypes are desirable, especially in food grade soybeans (Lee et al. 2007).
Also, there has been a growing interest from the aquaculture industry in using plant-based proteins and oils as components of fish diets as a way to reduce both feeding costs and, especially, the pressure on marine resources. Soybean protein concentrate and oil, due to their excellent nutritional characteristics, are considered very suitable for aquafeed. Increasing global demand for fish high in long-chain ω-3 oils creates perfect market opportunities for soybeans with enhanced ω-3 content (Naylor et al. 2009). In addition, oils low in saturates and high in PUFAs would have applications as drying oils for paints, coatings and other industrial products (Wilson 2004).

**Fatty acid and Triacylglycerol synthesis**

Fatty acid biosynthesis is part of the triacylglycerol (TAG) synthetic pathway. A general scheme of FA and TAG synthesis in developing seeds is presented in Fig. 3.1. In plastids, FAs are synthesized from acetyl-CoA, which is carboxylated to form malonyl-CoA through the action of acetyl-CoA carboxylase (ACCase). A second enzyme system, fatty acid synthase (FAS), transfers the malonyl moiety to acyl carrier protein (ACP) and catalyzes the extension of the growing acyl chain with malonyl-ACP to form 16:0-ACP. Finally, elongation and desaturation of 16:0-ACP to form 18:0-ACP and 18:1-ACP, respectively, results in de novo synthesized FAs. By the action of ACP-thioesterases, the FA molecules are released from ACP in the plastid stroma and cross the membrane by an unknown mechanism. On the plastid outer membrane, an acyl-CoA synthetase (ACS) assembles acyl-CoA esters that are then available for acyltransferase reactions in the
endoplasmic reticulum (ER). In the ER, synthesized FAs are incorporated onto a glycerol backbone to form TAG through the Kennedy pathway. Two successive acyl-CoA-dependent acylation reactions transfer FAs to glycerol-3-phosphate (G3P) to produce lysophosphatidic (LPA) and phosphatidic (PA) acids, respectively. PA dephosphorylation results in diacylglycerol (DAG), which is at the branch point of the pathway between TAG and phosphatidylcholine (PC) formation. The acyl-CoA-dependent acylation of DAG, catalyzed by diacylglycerol acyltransferase (DGAT), leads to TAG, while the transfer of a phosphocholine into DAG, catalyzed by CDP-choline:1,2-diacylglycerol cholinephosphotransferase (CPT), results in PC formation. A recently discovered enzyme, phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT) catalyzes the interconversion between DAG and PC by phosphocholine headgroup exchange (Lu et al. 2009). The FA desaturation (FAD) pathways, catalyzed by FAD2 and FAD3 desaturases, utilize PC as a substrate increasing the number of unsaturations of its acyl chains. The PUFA-enriched PC, by the reverse reaction of CPT and PDCT, can be transformed back into DAG. Also, the acyl chain at position 2 of PC may undergo exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). PC can also participate in an acyl-CoA-independent reaction to produce TAG. In this case, phospholipid:diacylglycerol acyltransferase (PDAT) catalyzes the formation of TAG using DAG and PC as substrates. Finally, a second acyl-CoA-independent reaction leading to TAG synthesis has been described, where diacylglycerol transacylase (DGTA) catalyzes the transfer of an acyl moiety between two DAG molecules to form TAG and
monoacylglycerol (MAG) (Lung and Weselake 2006; Ohlrogge and Browse 1995; Ohlrogge and Jaworski 1997).

**Fatty Acid Desaturase (FAD) genes in plants**

Fatty acids are the main components of plant membrane and seed storage lipids (Töpfer et al. 1995). In most plant tissues, over 75% of the FAs are unsaturated (Hildebrand et al. 2008). Fatty acid desaturases are the enzymes that introduce double bonds, i.e., unsaturations, into the hydrocarbon chains of fatty acids. These enzymes are encoded by nuclear genes and differ in their substrate specificity and subcellular localization. FAD2 and FAD3 are located in the ER while FAB2, FAD4, FAD5, FAD6, FAD7 and FAD8 are targeted to the plastids (Wallis and Browse 2002). FAB2, also referred to as SACPD, is a soluble desaturase and catalyzes the synthesis of 18:1 from 18:0; the other desaturases are membrane bound. FAD2 and FAD6 are ω-6 desaturases that catalyze the unsaturation of 18:1 to form 18:2. FAD3, FAD7 and FAD8 are ω-3 desaturases that synthesize 18:3 from 18:2. Finally, FAD4 and FAD5 produce 16:1 from 16:0 (Gibson et al. 1994; Hildebrand et al. 2008; Murphy and Piffanelli 1998).

**Microsomal FAD genes in soybean**

Most PUFAs, especially in nonphotosynthetic tissues, are produced by desaturases located in the ER (Miquel and Browse 1994). The soybean genome is complex and has a
predicted size of 1,115 Mb distributed across 20 chromosomes (Arumuganathan and Earle 1991; Schmutz et al. 2010). There is evidence suggesting that soybean underwent two whole-genome duplication events followed by gene diversification and loss and chromosome rearrangements. The result is a highly duplicated genome with approximately 75% of the genes existing as paralogues (Schmutz et al. 2010). Therefore, in soybean, it is very likely to find multiple copies of a gene that is present as a single copy in other species like Arabidopsis.

In soybean, there are at least five FAD2 and three FAD3 genes. FAD2-1A (Glyma10g42470) and FAD2-1B (Glyma20g24530) are primarily expressed in developing seeds, and play an important role in determining the PUFA content in soybean oil (Heppard et al. 1996; Pham et al. 2010; Schlueter et al. 2007; Tang et al. 2005). The FAD2-2 genes, consisting of FAD2-2A (Glyma19g32930), FAD2-2B (Glyma19g32940) and FAD2-2C (Glyma03g30070), are constitutively expressed in both vegetative tissue as well as developing seeds (Li et al. 2007; Schlueter et al. 2007). The exception is FAD2-2A, which produced amplicon when using genomic DNA but failed to amplify any of the RNA samples collected from diverse tissues. This would suggest that FAD2-2A is not expressed (Schlueter et al. 2007). Three microsomal ω-3 desaturase genes, namely FAD3A (Glyma14g37350), FAD3B (Glyma02g39230) and FAD3C (Glyma18g06950), have been identified in soybean (Bilyeu et al. 2003). FAD3A was characterized as the most highly expressed of the three homologs in developing seeds. FAD3A, FAD3B and FAD3C genes were also expressed in leaf tissue and flowers (Bilyeu et al. 2003). cDNA from a putative fourth microsomal ω-3 desaturase gene, herein
referred to as *FAD3D*, was obtained and characterized (Anai et al. 2005). Spatial and temporal expression of this gene seems to indicate it is a minor contributor in the desaturation of FAs in developing seeds (Anai et al. 2005).

**Modification of fatty acid profile in soybean oil**

Genetic modification of the typical FA makeup of soybean oil has been conducted for a number of years to improve oil nutritional and functional quality (Fehr 2007; Gerde and White 2008). Mid- to high-oleic (50-80% 18:1) as well as low- and ultralow-linolenic (<3% and ~1% 18:3, respectively) genotypes were obtained by breeding with mutant alleles of *FAD2* and *FAD3* genes (Alt et al. 2005a; Alt et al. 2005b; Bilyeu et al. 2005; Bilyeu et al. 2006; Bilyeu et al. 2003; Ross et al. 2000). These non-functional alleles cause the accumulation of substrates and reduction of corresponding products in their respective desaturation points. To illustrate, soybean genotype M23, which contains 46% 18:1, was developed by treatment of cultivar Bay with X-ray irradiation (Rahman et al. 1994). Deletion of the gene at the *FAD2-1A* locus was associated with the mid-oleic phenotype of M23 (Sandhu et al. 2007). Recently, it was observed that combination of mutant alleles at both the *FAD2-1A* and *FAD2-1B* loci further increases 18:1 content up to 80% of oil (Hoshino et al. 2010; Pham et al. 2010). Also, mutations were discovered in *FAD3A*, *FAD3B* and *FAD3C* genes, and soybean germplasm containing 1% 18:3 was developed through the combination of mutant alleles at those three loci (Bilyeu et al. 2006).
Linolenic acid content in the soybean wild ancestor

Accessions of Glycine soja (Siebold & Zucc.), the wild ancestor of cultivated soybean, have been reported to contain as much as 23% 18:3, representing potentially useful genetic resources for identifying either new alleles or genes that govern 18:3 accumulation (Pantalone et al. 1997; Wilson 2004). However, very little is known about the genetic regulation of high 18:3 in G. soja. Results of a survey on the average oil content and FA profile of G. max and G. soja accessions in the USDA’s National Plant Germplasm System (NPGS) (http://www.ars-grin.gov/npgs/) are shown in Table 3.1. On average, G. soja produces twice as much 18:3 but almost half the amount of 18:1 and oil as cultivated soybean. Also, G. soja presents higher variability for 18:3 than G. max, whereas G. max surpasses G. soja in variability for oil, 18:1 and 18:2. Analysis of relative 18:1 and 18:2 desaturation patterns, which would reflect ω-6 and ω-3 desaturases’ activity, respectively, in four G. max x G. soja populations suggested that G. soja carries superior alternative FAD2 and FAD3 alleles directing the expression of the high 18:3 trait (Pantalone et al. 1997). Finally, increasing the expression of the FAD3 genes by genetic engineering, allowed the generation of soybean seeds that accumulated 18:3 in excess of 50% of the total oil (Cahoon 2003).

The objective of this work was to gain a better understanding of the genetics of 18:3 accumulation in G. soja in general, and the role of the microsomal ω-6 (FAD2-1A and FAD2-1B) and ω-3 (FAD3A, FAD3B, FAD3C and FAD3D) desaturase genes in the high 18:3 content of PI 366121 in particular. We hypothesized that superior FAD3 and/or FAD2 alleles were driving the elevated 18:3 content in PI 366121. Structural and
functional characterization of four ω-3 and two ω-6 desaturase genes suggests that PI 366121 carries similar desaturase alleles to those found in soybean.
Materials and Methods

Plant material and development of segregating populations

Study of the genetic basis of high 18:3 in seeds of soybean wild ancestor was carried out with *G. soja* accession PI 366121 and *G. max* genotypes Williams 82 and B1-52. PI 366121 is listed as a high 18:3 content entry (up to 23% 18:3) in the USDA’s National Plant Germplasm System ([http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1274060](http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1274060)). Williams 82 (Bernard and Cremeens 1988) is the reference soybean genotype in genomic studies and the chosen cultivar for the construction of the soybean whole-genome shotgun sequence assembly (Schmutz et al. 2010). Williams 82 has functional *FAD3A*, *FAD3B* and *FAD3C* genes (Bilyeu et al. 2005; Bilyeu et al. 2006), and it presents a typical 18:3 content (~7-8%) for a soybean line. B1-52 is an ultralow 18:3 content (~1%) genotype (Bilyeu et al. 2006), and perfect molecular markers that allow tracking its mutant alleles for the *FAD3A*, *FAD3B* and *FAD3C* genes were designed in our laboratory and are available to use.

Also, *Glycine soja* accession PI 366121 was crossed with *G. max* genotypes Williams 82 and B1-52 to generate two segregating populations, which will be referred to as the 7% lin and 1% lin populations, respectively. F$_2$ seeds were produced in growth chamber and planted to generate F$_3$ seeds in Columbia, Missouri in the summer of 2009. For the 7% lin population, F$_3$ seeds were sent to Upala, Costa Rica and advanced to F$_5$ by
single seed descent method to generate recombinant inbred lines (RILs). Also, a sibling set of the Costa Rica-produced F₄ population was planted in Columbia, Missouri in the summer of 2010. Seeds of the F₅ RILs obtained in both Upala, Costa Rica and Columbia, Missouri were used in further analyses.

**Sampling DNA and determinations on seeds**

FTA PlantSaver cards (Whatman Inc., Florham Park, NJ) were used to take DNA samples from ~110 plants from each the 7% lin and the 1% lin populations in F₂ generational advance, according to manufacturer's recommendations. Before planting, dry weight and oil content on a dry weight basis were determined on individual seeds. Also, single seeds were evaluated for oil composition by cutting a seed chip (~1/3 of the seed) from each seed. The remainder of the seed was used for planting.

Progeny of each single plant from which a DNA sample was collected were harvested in the fall of 2009, and relative fatty acid content was determined on three individual F₃ seeds per plant.

Also, from each of the 7% lin RILs produced in both Upala, Costa Rica (CR) and Columbia, Missouri (CoMo), a composite sample of five F₅ seeds was used for seed dry weight and relative fatty acid content determination.
In addition, from the 1% lin population, 214 individual F$_3$ seeds derived from a single F$_2$ plant were analyzed for relative fatty acid content to determine the nature of the genetics governing 18:3 content in seeds (qualitative vs. quantitative).

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 as explained in Bilyeu et al. (2005). The individual fatty acid contents are reported as the relative percents of palmitic, stearic, oleic, linoleic, and linolenic acids in the extracted oil.

Seed oil and moisture content was determined through nuclear magnetic resonance (NMR) using a Bruker Minispec Mq10 NMR machine (Bruker Optics Inc., Billerica, MA).

**Candidate genes**

We considered two types of candidate genes based on where in the TAG pathway their products carry out the catalytic action: desaturases (ω-3 and ω-6) and transferases (acyltransferases and cholinephosphotransferase). In this chapter, only desaturases will be considered.
Taking into account previous successful efforts to modify soybean oil composition using mutants of either ω-3 (FAD3 genes) or ω-6 desaturases (FAD2 genes) as well as the results of the paper by Pantalone et al. (1997), we considered as very strong candidates the following genes: FAD3A (Glyma14g37350), FAD3B (Glyma02g39230), FAD3C (Glyma18g06950), FAD3D (Glyma11g27190), FAD2-1A (Glyma10g42470) and FAD2-1B (Glyma20g24530).

**DNA isolation and PCR for sequencing candidate genes**

Genomic DNA was isolated from ~30 mg dried seed tissue using the DNeasy Plant Mini Kit (Qiagen Sciences Inc., Germantown, MD) and used at 5 to 50 ng per PCR amplification. PCR was carried out using Ex Taq according to manufacturer’s recommendations (TaKaRa, Otsu, Shiga, Japan) in a PTC-200 thermocycler (MJ Research/Bio-Rad, Hercules, CA) using the following conditions: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for as long as needed in each case (for the extension step of PCR, we added 30 s/500 bp of genomic sequence). PCR products were analyzed by gel electrophoresis (1% agarose gel) to verify for size and ensure specific amplification. PCR products were isolated with the QIAprep Spin Miniprep kit (Qiagen Sciences Inc.) and sequenced at the University of Missouri DNA Core facility. Primers used for amplification and sequencing are listed in Table 3.2.
Sequence analysis

Sequences were imported into ContigExpress program of Vector NTI Advance 10 (Invitrogen, Carlsbad, CA), trimmed, assembled and manually evaluated for disagreements between Williams 82 (Bernard and Cremeens 1988) reference sequence and imported sequence contigs. Putative SNPs and deletions were verified by at least two independent PCR amplifications. To evaluate the effect of changes in coding sequence at the protein level, the program ExPASy translate tool was used (http://ca.expasy.org/tools/dna.html). Sequence alignments were generated using the AlignX software (Invitrogen). Gene models were obtained from the whole soybean genome assembly (http://www.phytozome.net/soybean). Exon-intron boundaries were verified for accuracy using TBLASTN program with the protein sequence of the corresponding homologous genes in Arabidopsis as queries and translated genomic sequence in six frames.

Genotyping assays

Molecular marker assays were designed to track alleles from PI 366121, Williams 82 and B1-52 for FAD2-1 and FAD3 genes in the 7 and 1% lin F2 populations.

SimpleProbe assays

SimpleProbe assays are based on the disassociation kinetics of SimpleProbe oligonucleotides (Roche Applied Sciences, Indianapolis, IN) transitioning from a
fluorescent bound to target state to a nonfluorescent unbound state. SimpleProbes were designed using the Lightcycler Probe Design Software, version 1 (Roche Applied Sciences) and purchased from Roche Applied Sciences. Asymmetric PCR was carried out to generate more single stranded DNA for probe binding. SimpleProbe assays were designed to distinguish B1-52 and PI 366121 alleles of \textit{FAD2-1B}, \textit{FAD3A}, \textit{FAD3B} and \textit{FAD3C} genes. Also, a SimpleProbe assay was developed to identify Williams 82 and PI 366121 alleles of \textit{FAD3A} gene. Sequence of primers and SimpleProbes as well as their final concentration in the PCR mix are listed in Table 3.3.

Reactions were carried out in a total volume of 20 µl containing template, primers, SimpleProbe, buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl\textsubscript{2}, 3.75 μg ml\textsuperscript{-1} BSA, 200 μM dNTPs], 5% DMSO and 0.2X Titanium Taq polymerase (BD Biosciences). Genotyping reactions were performed using a Lightcycler 480 II real time PCR instrument (Roche Applied Sciences), using the following PCR parameters: 95°C for 5 min followed by 40 cycles of 95°C for 20 s, 60°C for 20 s, 72°C for 20 s and then a melting curve from 50 to 75°C. Fluorescence was read after each cycle and every 0.1°C during the melting curve analysis (MCA). A mismatch between the probe and the target sequence results in altered disassociation kinetics of the probe. Consequently, each genotype produces a characteristic melting profile, as measured by the melting temperature (T\textsubscript{m}) of the first negative derivative of the fluorescence value change with increasing temperature. In the \textit{FAD3A} (7% lin) assay, Williams 82 allele produced a T\textsubscript{m} of 60°C and PI 366121 allele produced a T\textsubscript{m} of 65°C; in the \textit{FAD2-1B} assay, B1-52 allele produced a T\textsubscript{m} of 65°C and PI 366121 allele produced a T\textsubscript{m} of 56°C; in the \textit{FAD3A}
(1% lin) assay, B1-52 allele produced a Tₘ of 54°C and PI 366121 allele produced a Tₘ of 64°C; in the FAD3B assay, B1-52 allele produced a Tₘ of 53°C and PI 366121 allele produced a Tₘ of 59°C; finally, in the FAD3C assay, B1-52 allele produced a Tₘ of 64°C and PI 366121 allele produced a Tₘ of 68°C.

**CAPS assays**

CAPS (Cleaved Amplified Polymorphic Sequences) markers are based on the locus-specific amplification of genomic DNA followed by digestion with restriction endonucleases that generates allele-specific digestion patterns (Konieczny and Ausubel 1993). A 703 bp fragment of FAD2-1B from Williams 82 and PI 366121 was amplified with forward and reverse primers 5’-GGCAGACCCCTATGATGGTTTTGCT-3’ and 5’-TTACATTATAGCCATGGATCGCTAC-3’, respectively. PCR reactions were carried out in a total volume of 15 µl containing 5-50 ng DNA template, primers (0.5 µM), buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 µg ml⁻¹ BSA, 200 µM dNTPs], 5% DMSO and 0.2X Titanium Taq polymerase (BD Biosciences). The following PCR parameters were followed: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1 min. After amplification, digestion of PCR products with restriction enzyme EcoRI (New England BioLabs, Inc., Ipswich, MA) was carried out for 5 hs following manufacturer’s instructions. Following digestion, samples were run on a 1.5% agarose gel. Individuals that were homozygous for the Williams 82 allele showed two bands of size 395 bp and 308 bp, while homozygotes for the PI 366121 allele presented a single band of 703 bp. Heterozygotes showed three bands (703, 395 and 308 bp).
**Fragment analysis**

This type of genotyping assay uses a fluorescence-based detection system to determine differences in size of DNA fragments. Fluorescent dyes are incorporated into the DNA fragment using fluorescence-labeled primers. Fragment analysis assays are useful to distinguish alleles differing in size due to small insertions/deletions. Alleles from Williams 82 and PI 366121 for *FAD3B* and *FAD3C* genes were distinguished with this kind of assay. PCR reactions were carried out under the following conditions: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 60°C (*FAD3B*) or 62°C (*FAD3C*) for 30 s and 72°C for 40 s and 72°C for 5 min after the last cycle. Reactions were performed in a total volume of 15 µl containing 5-50 ng DNA template, 0.5 µM of FAM-labeled primers (5’-FAM-CTCAAAGATTTTGTTGTGTCCT-3’ and 5’-CAGTCACTTTAAAGATCTCAATG-3’ for *FAD3B* and 5’-GTAGTGACTGCATTGGTAGCTG-3’ and 5’-FAM-CGAACAATCATGCATAACCAAGTG-3’ for *FAD3C*, forward and reverse, respectively), buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 µg ml⁻¹ BSA, 200 µM dNTPs], 5% DMSO and 0.2X Titanium *Taq* polymerase (BD Biosciences). Fluorescently-labeled PCR-amplified fragments were sent to the University of Missouri DNA Core for detection and interpretation. Williams 82 alleles were 98 bp and 226 bp while PI 366121 alleles were 107 bp and 221 bp for *FAD3B* and *FAD3C*, respectively.


**Sequence analysis**

Williams 82 and PI 366121 alleles for \( FAD2-1A \) and \( FAD3D \) genes were identified in individuals of the 7\% lin population by sequence analysis. Polymorphic fragments were amplified by means of PCR and sent to the University of Missouri DNA Core for sequencing. The primers used to amplify and then sequence the polymorphic fragments were forward 5’-GAAATGAAACCATAAATAAACC-3’ and reverse 5’-TTCTAAATCCCTTGATTCAGCCGT-3’ for \( FAD2-1A \) and forward 5’-CCAATGTGGGTTAAGTAATGCC-3’ and reverse 5’-CAGAGAATTAATACCGCTTCAACC-3’ for \( FAD3D \). PCR reactions were carried out under the following conditions: 95°C for 5 min, 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1 min and 72°C for 5 min after the last cycle. Reactions were performed in a total volume of 15 µl containing 5-50 ng DNA template, primers (0.5 µM), 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 and 0.2X Titanium Taq polymerase (BD Biosciences).

**Expression analysis of candidate genes**

Standard curves to determine the efficiency of each primer pair were constructed using serial dilutions of genomic DNA as template (Pfaffl 2001). Total RNA was isolated from powdered Williams 82 or PI 366121 developing seeds by means of PureLink™ Plant RNA reagent (Invitrogen) following the manufacturer’s instructions. Total RNA was treated with DNase I and the supplied 10X DNase I buffer (Ambion Inc.,
Austin, TX) by incubating at 37°C for 30 min and then heating at 75°C for 15 min to inactivate the DNase. Effectiveness of DNase treatment on total RNA was evaluated by comparing amplification of samples using RT-PCR master mix with or without reverse transcriptase enzyme following the protocol described below.

The reaction mix consisted of QuantiTect SYBR Green RT-PCR master mix (Qiagen, Valencia, CA), supplemented with 0.2X Titanium *Taq* polymerase (BD Biosciences, Palo Alto, CA), DNase-treated RNA template and 0.5 μM of each primer in 20 μL reactions. The experimental conditions were reverse transcription at 50°C for 30 min followed by 95°C for 15 min, then 35 to 40 cycles of 95°C for 20 s, 60°C for 20 s, 72°C for 20 s and an ending hold at 4°C. After cycling was completed and before the ending hold, a melting curve program was run to confirm the specificity of the amplification.

Cycle threshold measurements were mathematically converted into genome equivalents by application of the standard curves. Expression levels of candidate genes were expressed as a ratio to the expression level of reference gene *cons7*, a putative insulin-degrading enzyme (Libault et al. 2008). Primers used for qRT-PCR experiments are listed in Table 3.4.
Results

Polymorphisms in candidate genes of *Glycine soja* line PI 366121

Gene-specific PCR was used to amplify and sequence *FAD2-1A* and *FAD2-1B*; *FAD3A*, *FAD3B*, *FAD3C* and *FAD3D* genes from *G. soja* line PI 366121. Genomic sequence from cultivar Williams 82 was used as reference sequence (http://www.phytozome.net/soybean).

Out of the six candidate genes whose genomic sequence was analyzed, three presented no polymorphisms at the coding sequence level (*FAD2-1A*, *FAD3C* and *FAD3D*), while the remaining three (*FAD2-1B*, *FAD3A* and *FAD3B*) showed silent and/or missense polymorphisms in the coding sequence (Table 3.5). The M126V missense polymorphism observed in *FAD2-1B* is not exclusive of PI 366121, and it has been observed in several *G. max* accessions (Pham et al. 2010).

Association between high 18:3 content in seeds and genotype at the candidate genes

The identification of polymorphisms in the candidate genes between the parents of the two populations presented the opportunity to directly test the association of each candidate gene with the fatty acid profile of the seed oil. Molecular marker assays were designed to track the alleles from PI 366121, Williams 82 and B1-52 for the desaturase
genes in the F₂ populations. Also, for FAD2-1A and FAD3D, sequencing of genomic DNA of selected F₂ individuals from the 7% lin population was carried out to determine if PI 366121 alleles were associated with high 18:3 content in seeds. Association analysis between genotype and phenotype of F₂ individuals yielded different results depending on the population and gene considered (Fig. 3.2).

In the 1% lin population, the presence of the PI 366121 alleles for the FAD3A, FAD3B and FAD3C genes was associated with an increase in seed 18:3 content. Individuals homozygous for the PI 366121 FAD3A, FAD3B and FAD3C alleles accumulated, in average, the highest 18:3 content, although it was lower than that of PI 366121, while individuals with the B1-52 alleles showed the lowest 18:3 content, which was identical to that of B1-52 (Fig. 3.2a). Interestingly, in the 7% lin population, no association was observed between the presence of PI 366121 FAD3A, FAD3B and FAD3C alleles and increments in 18:3 in seeds (Fig. 3.2b). Moreover, no association was found between the presence of the PI 366121 FAD2-1B allele and seed 18:3 content in any of the two populations (data not shown).

Also, ten F₂ individuals of the 7% lin population, five from the group with highest and lowest 18:3 content, respectively, were sequenced for two polymorphisms, one in an intron of FAD3D and the other in a region ~7kb upstream of FAD2-1A. Results indicated no association between the presence of a particular allele at any of these two genes and 18:3 content (data not shown).
Oil profile of three separate F3 seeds collected from individual F2 plants was determined, and association between F2 genotype at all desaturase genes considered and F3 phenotype was analyzed in both the 1% and the 7% lin populations. The pattern of genotype-phenotype association or lack of it that had been previously observed in F2 seeds was confirmed in F3 (data not shown).

Allele substitution calculations in the 1% lin population indicated that PI 366121 FAD3A, FAD3B and FAD3C genes contribute unequally to 18:3 content in seeds (Table 3.6). An increase in 18:3 levels occurred as PI 366121 alleles replaced their B1-52 counterparts. Also, it was observed that PI 366121 FAD3A gene associated with greater increases in 18:3 levels than FAD3B and FAD3C, and that FAD3B and FAD3C have similar effects on the proportion of 18:3 in seeds.

Gene expression studies

To investigate if the higher 18:3 content in PI 366121 seeds was due to over expression of any of the desaturase genes, quantitative RT-PCR experiments were carried out. Developing seeds from cultivar Williams 82 and accession PI 366121 were collected at 25, 30, 35 and 40 days after flowering (DAF) and used for total mRNA extraction and quantification. A standard curve was constructed for each pair of gene-specific primers to determine the efficiency of each primer pair. The data was obtained in technical triplicates and biological duplicates. The expression levels of the investigated genes were
expressed as a ratio to the expression level of the reference gene cons7 (Libault et al. 2008).

The results indicate that higher 18:3 content in the G. soja accession PI 366121 is not due to an over expression of FAD3A, FAD3B or FAD3C genes as compared to Williams 82. On the contrary, the expression of FAD3 genes was very similar (30 and 35 DAF) or even higher (FAD3A, 40DAF and FAD3A, FAD3B and FAD3C, 25 DAF) in the G. max cultivar than in PI 366121 (Fig. 3.3). Also, there were no differences in transcript level of either FAD2-1A or FAD2-1B between PI 366121 and Williams 82 (data not shown).

**Segregation study**

In the 1% lin population, no genotypic class recovered the PI 366121 18:3 content level (Fig. 3.2a). Also, in some of the F2:3 individuals that had the FAD3 genes fixed, 18:3 seed content showed standard deviations bigger than any of the parents, which could suggest the involvement of additional genes in the determination of this trait. For example, 13 F2:3 individuals derived from the same F2 plant (which had functional FAD3 genes fixed), were analyzed for relative fatty acid levels and showed 18:3 contents with standard deviation values ~3 and 8 times higher than in PI 366121 and B1-52, respectively. To further investigate if just a few or many genes were affecting 18:3 content and segregating in progeny of this F2 plant, an additional 200 F3 seeds were used for fatty acid determination. A continuous distribution of phenotypes rather than discrete
classes was observed, which would indicate the involvement of multiple additional genes in the determination of 18:3 content in seeds (data not shown).

Also, to determine the effect of different environments on seed 18:3 content, two sets of sibling RILs in F₄ generation from the 7% lin population were planted in summer 2010. One set was planted in CR and the other in CoMo. At maturity, plants were individually threshed and F₅ seeds were obtained. Relative fatty acid composition was determined for each line on a sample composed of five seeds. Analysis of phenotypic frequency distribution shows that 18:3 levels were normally distributed in the two locations (Chapter 5 of this work). Also, 18:3 content was higher in CoMo than in CR, with population averages of 10.7 and 8.1%, respectively. Additionally, the range of phenotypic values was different in the two locations, with a smaller difference between the lowest and highest value in CR (6.7%) than in CoMo (9.6%). Parental genotypes also showed different 18:3 content in the two locations; Williams 82 and PI 366121 reached 6.2 and 10.7% 18:3, respectively in CR and 7.6 and 13.7%, respectively in CoMo (Chapter 5 of this work).
Discussion

Seed oil content and composition are the final result of a complex and not completely understood network of biochemical reactions taking place during TAG biosynthesis in the ER. In addition to the inherent complexity of the process, it seems likely that the channeling of acyl substrates during TAG synthesis is configured differently from one plant species to the other, which makes difficult the extrapolation of results between plant species (Ruiz-López et al. 2009; Taylor et al. 2009). However, there are examples of successful modification of oil composition in crop species such as soybean by applying the knowledge obtained from research in model species (Bilyeu et al. 2003; Pham et al. 2010).

Accessions of *Glycine soja* have been reported to contain as much as 23% 18:3 in the seed oil, representing potentially useful genetic resources for identifying either new alleles or genes that govern 18:3 accumulation (Pantalone et al. 1997; Wilson 2004). Analysis of relative 18:1 and 18:2 desaturation patterns in four *G. max* x *G. soja* populations suggested that *G. soja* carries superior *FAD2* and *FAD3* alleles directing the expression of the high 18:3 trait (Pantalone et al. 1997). Our results, however, indicate that *G. max* and its wild ancestor carry equivalent alleles for both the ω-3 (Fig. 3.2) and ω-6 desaturases (not shown), meaning that their contributions to 18:3 content in seeds are very similar. Also, a major finding of this research is that having all *FAD3* genes functional is an essential requirement to achieve the highest 18:3 levels in seeds. In the
1% lin population, the genotypic class with the maximum 18:3 was that with functional and homozygous *FAD3A, FAD3B* and *FAD3C* genes (Fig. 3.2a). Moreover, the importance of the individual contributions of *G. soja* *FAD3A, FAD3B* and *FAD3C* to 18:3 accumulation in seeds shows *FAD3A* contributing more to 18:3 content than *FAD3B* and *FAD3C*, and *FAD3B* and *FAD3C* having similar effects on the proportion of 18:3 in the oil. Identical observations were previously reported for soybean (Bilyeu et al. 2005; Bilyeu et al. 2006).

The gene expression studies indicate that the high 18:3 content in seeds of PI 366121 is not the result of more transcriptionally active ω-3 (Fig. 3.3) or ω-6 (not shown) desaturase genes. What is more, the expression level of these genes in *G. max* cultivar Williams 82 was similar to or higher than in *G. soja* accession PI 366121. Even though functional desaturase genes are important and essential to achieve high 18:3 content in seeds, it is evident they are not exclusively responsible for the 18:3 accumulation levels observed in PI 366121.

Segregation studies in F3 generation (1% lin population) indicate that linolenic acid content in *G. soja* is a trait determined by multiple genes since it presents a continuous variation of phenotypic values rather than distinct classes. Additional support to the multigenic nature of 18:3 in seeds of *G. soja* comes from the analysis of the phenotypic frequency of RILs in F5 generation (7% lin population), which shows a normal and continuous distribution. In addition, 18:3 content in seed oil of both RILs and parental lines showed to be influenced by the growing environment, another key feature of quantitative traits (Bernardo 2010). Higher 18:3 levels were observed in CoMo, where
temperatures are cooler during the seed development stages, than in CR. Another piece of evidence of the effect of the environment on 18:3 content in seeds is the wider range of phenotypes in the RIL population grown in CoMo than in that grown in CR. This is explained by differences in latitude and consequently photoperiod in one and the other location: while distinct maturity groups are expressed in CoMo, they are minimized in CR.
Conclusions

Soybean is valued on the marketplace as a source of protein and oil. Soybean oil is broadly incorporated into food, feed and industrial products. Increased demand for vegetable oils and growing awareness of health issues around dietary fats make necessary a better understanding of the genetics underlying oil content and composition. Enhancement of ω-3 FA content of the oil for both food and feed applications is highly desired. Accessions of G. soja, the wild ancestor of soybean, contain, in average, twice as much ω-3 FAs (18:3) as cultivated soybean, representing a potential source of superior genes/alleles. However, little is known about the genetics of the elevated 18:3 trait in wild soybean.

Significant progress has been made in understanding the role of FAD2-1 and FAD3 genes in the high 18:3 phenotype of PI 366121. The data suggests that the G. soja accession PI 366121 and G. max cultivar Williams 82 carry equivalent functional alleles for FAD2-1A, FAD2-1B, FAD3A, FAD3B, FAD3C and FAD3D genes. Moreover, elevated 18:3 in PI 366121 seeds is not the result of higher expression of either FAD3 or FAD2-1 genes. Since the desaturase genes do not explain the differences in 18:3 content between G. soja and G. max, it is necessary to change the focus towards genes that encode enzymes catalyzing other steps in the TAG synthesis pathway and that have the potential to modify the composition of oil. Enzymes participating directly or indirectly in the incorporation of 18:3 in TAG, where ~90% of 18:3 is, would be good candidates.
In recent years, efforts have been made to better understand FA and TAG synthesis regulation and accumulation. Research on the model plant *Arabidopsis* shows that oil content and composition in seeds vary dramatically depending on whether DGAT1 or PDAT1 contributes more to the final acylation step to form TAG (Katavic et al. 1995; Routaboul et al. 1999; Zhang et al. 2009; Zou et al. 1999). Changes in oil composition are explained, at least in part, by the fact that while DGAT1 shows higher activity with saturated and long-chain FAs, PDAT1 shows a preference for acyl groups containing several double bonds and unusual FAs (Katavic et al. 1995; Stahl et al. 2004; Zhang et al. 2009). In addition, Lu et al. (2009) showed evidence of a new enzyme, PDCT, which interconverts DAG and PC during TAG synthesis in developing seeds of *Arabidopsis*. PDCT seems to catalyze an important step for PUFA enrichment of TAG since the null allele was associated with a reduction of ~40% in the seed content of PUFA. Also, it has been shown that the majority of the fatty acids synthesized in plastids enter PC by exchanging with PUFA in a process known as acyl editing (Bates et al. 2009; Bates et al. 2007). The enzyme catalyzing the incorporation of *de novo* FAs into PC, LPCAT, by the reverse reaction releases PUFAs that in turn will be part of the cytoplasmic acyl-CoA pool (Stymne and Stobart 1984). In soybean, it has been shown that LPCAT has a preference for 18:1-CoA and 18:2-CoA as acyl donors, and those acyl chains are incorporated into PC for further desaturation (Stymne and Glad 1981). Because of the key role that LPCAT has in regenerating PC from LPC, and since PC is the only substrate for the desaturases, the involvement of LPCAT in determining oil composition could be significant.
Detailed understanding of TAG metabolism and its regulation is still lacking. However, research on oil traits has been dynamic, particularly in the model plant Arabidopsis, and that will certainly have a positive impact in the improvement of oilseeds.
Figures
Figure 3.1. Fatty acid (FA) and triacylglycerol (TAG) synthesis in oilseed plants developing seeds. FA synthesis occurs in plastids. Synthesized FAs enter the cytosolic acyl-CoA pool and are used to form TAG from glycerol-3 phosphate (G3P) in the ER. The FA desaturation (FAD) pathways utilize phosphatidylcholine (PC) as a substrate. FAD2 enzymes catalyze the unsaturation of 18:1-PC to form 18:2-PC; FAD3 enzymes synthesize 18:3-PC from 18:2-PC. PUFA-enriched PC, can be transformed back into diacylglycerol (DAG). Also, the acyl chain at position 2 of PC may undergo exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; ACS, acyl-CoA synthetase; CPT, CDP-choline:1,2-diacylglycerol cholinephosphotransferase; DGAT, diacylglycerol acyltransferase; DGTA, diacylglycerol transacylase; FAE, FA elongase; FAS, FA synthase; GPAT, glycerol-3-phosphate acyltransferase; LPA, lysophosphatidic acid; LPAAT, lysophosphatidic acid acyltransferase; LPC, lysophosphatidylcholine; PAP, phosphatidate phosphatase; PDAT, phospholipid:diacylglycerol acyltransferase; PDCT, phosphatidylcholine:diacylglycerol cholinephosphotransferase; PLA2, phospholipase A2; TS, acyl-ACP thioesterase (modified from Lung and Weselake 2006)
Figure 3.2. Phenotype and genotype of F₂ plants segregating for FAD3A, FAD3B and FAD3C. Histograms representing average 18:3 content (% of oil content) of genotypic classes in F₂ generation. a B1-52 x PI 366121 (1% lin) population. b Williams 82 x PI 366121 (7% lin) population. Average phenotypes of parental lines are represented in black or white bars. Average phenotypes of F₂ individuals with indicated genotypes are represented in gray bars. Error bars represent one standard deviation from the mean. Upper case letters represent Glycine soja (PI 366121) alleles; lower case letters represent Glycine max (either B1-52 or Williams 82) alleles. Spaces with missing gray bar and with the letters “na” represent genotypic classes not recovered in F₂. Lack of error bar represents genotypic classes for which only one individual was recovered in F₂.
Figure 3.3. Relative expression of *FAD3A* (3A), *FAD3B* (3B) and *FAD3C* (3C) genes in seeds of Williams 82 (Gm) and PI 366121 (Gs) at 25, 30, 35 and 40 days after flowering (DAF). Steady state mRNA levels for each gene were quantitated by quantitative PCR following reverse transcription of total RNA. The histograms represent the expression of *FAD3A* (white bars), *FAD3B* (gray bars) and *FAD3C* (black bars) as the ratio to the expression level of the reference gene *cons7*.
Tables
Table 3.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of *Glycine max* and *Glycine soja* accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean. Information on wild soybean accession PI 366121 and *G. max* genotypes Williams 82 and B1-52 is also included.

<table>
<thead>
<tr>
<th></th>
<th>Oil (%)</th>
<th>16:0 (%)</th>
<th>18:0 (%)</th>
<th>18:1 (%)</th>
<th>18:2 (%)</th>
<th>18:3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>G. max</em>^a</td>
<td>20 ± 2.1</td>
<td>11 ± 1.0</td>
<td>4 ± 0.6</td>
<td>23 ± 3.8</td>
<td>54 ± 3.7</td>
<td>8 ± 1.5</td>
</tr>
<tr>
<td><em>G. soja</em>^b</td>
<td>11 ± 1.4</td>
<td>12 ± 0.9</td>
<td>4 ± 0.5</td>
<td>14 ± 2.4</td>
<td>54 ± 1.7</td>
<td>16 ± 2.2</td>
</tr>
<tr>
<td>PI 366121</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>11</td>
<td>54</td>
<td>20</td>
</tr>
<tr>
<td>(G. soja)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams 82</td>
<td>20</td>
<td>11</td>
<td>3</td>
<td>23</td>
<td>56</td>
<td>7</td>
</tr>
<tr>
<td>(G. max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1-52</td>
<td>25</td>
<td>11</td>
<td>4</td>
<td>23</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>(G. max)^c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a All observations for *G. max* accessions in the NPGS were used in the calculations (n=16,588)

^b All observations for *G. soja* accessions in the NPGS were used in the calculations (n=1,242)

^c Data generated in the lab. B1-52 carries mutant non-functional alleles of *FAD3A*, *FAD3B* and *FAD3C* genes (Bilyeu et al. 2006)
Table 3.2. Primers used for amplification and sequencing of FAD2-1 and FAD3 genes. Primers are listed in 5’→3’ configuration

<table>
<thead>
<tr>
<th>Primers used for both amplification and sequencing</th>
<th>forward sequence</th>
<th>reverse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FAD3A gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3AF1 CAAAATTCATCCTCGCTGGTTTT</td>
<td>3AR1 CCCAGGATTCATCCTTCTCTCA</td>
<td></td>
</tr>
<tr>
<td>3AF2 CTATTGGCCCATTCAAGGAA</td>
<td>3AR2 TTCTCTCAGTTGAGGAAC</td>
<td></td>
</tr>
<tr>
<td>3AF3 GGTAGTTTGATCCCGGTCTTT</td>
<td>3AR3 GCGATACCAAGGCAGTTTCT</td>
<td></td>
</tr>
<tr>
<td>3AF4 TTTCCACCCAGTGAGGACAA</td>
<td>3AR4 ATATGGGTGGAAAGCGAA</td>
<td></td>
</tr>
<tr>
<td>3AF5 GTATCGCGCAAGGTAACCA</td>
<td>3AR5 CGAAGTGGGTGTGAAAACATGG</td>
<td></td>
</tr>
<tr>
<td>3AF6 TACCTCGGTAAAGCGGTATTT</td>
<td>3AR6 TCAACGGTGGAACACTCTGTTC</td>
<td></td>
</tr>
<tr>
<td><strong>FAD3B gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3BF1 ACCGTAATTGGGGTTGTTCCA</td>
<td>3BR1 TCGAATGGGGTTCTAGCACC</td>
<td></td>
</tr>
<tr>
<td>3BF2 GCAGCAATTTCACCTGCAAA</td>
<td>3BR2 AGGCCATTTCAGTTGGTGAAC</td>
<td></td>
</tr>
<tr>
<td>3BF3 TGCTTCTGAGGCTTTTCTT</td>
<td>3BR3 AATGGGTTGGAACAGCAGAA</td>
<td></td>
</tr>
<tr>
<td>3BF4 CCATGTCACCACCAAGAAAC</td>
<td>3BR4 CAACAAAAACAGCGAAGC</td>
<td></td>
</tr>
<tr>
<td>3BF5 CAAGCCCCGGTTTTTCTTA</td>
<td>3BR5 GAAATGGTAATGGCGCAGAT</td>
<td></td>
</tr>
<tr>
<td>3BF6 CCTCGTTGAAGCGGTACATT</td>
<td>3BR6 TCCAGCGGTGTTGAATCTG</td>
<td></td>
</tr>
<tr>
<td>ALB9 ATGAAACATGTACCTTACAC</td>
<td>ALB10 GATTTAAGATGGAGAGAGCATTAG</td>
<td></td>
</tr>
<tr>
<td>ALB11 CTTTCTAGGTGTACATGAC</td>
<td>ALB12 CCAATAAATATAACCCCAAAGTTTG</td>
<td></td>
</tr>
<tr>
<td><strong>FAD3C gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3CF1 AGGCCATTGGAGGTGACATGTATG</td>
<td>3CR1 AGGGCTTCTGCTCCACTGAA</td>
<td></td>
</tr>
<tr>
<td>3CF2 GAAAAGGGGTTATCTAGAGGAG</td>
<td>3CR2 TGTGGAGCTGGGAATCAGTC</td>
<td></td>
</tr>
<tr>
<td>3CF3 GTTCCAATAGAAGAAGGGG</td>
<td>3CR3 GCATGCTTTGAAAGTAAGTGC</td>
<td></td>
</tr>
<tr>
<td><strong>FAD3D gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3DF1 GCATGTAACGTGATTAATACCCA</td>
<td>3DR1 GTGGTGAAAGGAAGAGGAAG</td>
<td></td>
</tr>
<tr>
<td>3DF2 CTTCTTCTTCTCCCCCTTCCC</td>
<td>3DR2 GCCATATTAAACCACATTGG</td>
<td></td>
</tr>
<tr>
<td>3DF3 CCAATGTGGGTATTAAGTATGCC</td>
<td>3DR3 CAGAGAATTAATACCAGCTTCACC</td>
<td></td>
</tr>
<tr>
<td>3DF4 TCCACAATAGACTAAAGCACGC</td>
<td>3DR4 CTACGCCATCGTACGCC</td>
<td></td>
</tr>
<tr>
<td><strong>FAD2-1A gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-1AF ACTGCAATCGAATAATACGACC</td>
<td>2-1AR TGATATTGTCCCGTGACG</td>
<td></td>
</tr>
</tbody>
</table>
**Table 3.2.** Primers used for amplification and sequencing of FAD2-1 and FAD3 genes. Primers are listed in 5′→3′ configuration (Continued)

### Primers used for both amplification and sequencing

<table>
<thead>
<tr>
<th>FAD2-1B gene</th>
<th>forward sequence</th>
<th>reverse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1BF</td>
<td>CCCGCTGTCCCTTTAACT</td>
<td>2-1BR</td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>TTACATTATAGGCCATGGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CGCTAC</td>
</tr>
</tbody>
</table>

### Primers used for sequencing only

<table>
<thead>
<tr>
<th>FAD3A gene</th>
<th>forward sequence</th>
<th>reverse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3AF7</td>
<td>GTCATTATCTTAGTTAACC</td>
<td>3AR7</td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>GCACTCACCACAAATAATTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GATAC</td>
</tr>
<tr>
<td>3AF8</td>
<td>CCGGAAGGAAGGTTCTC</td>
<td>3AR8</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>GGAATTCCATAGAGCTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AGCA</td>
</tr>
<tr>
<td>3AF9</td>
<td>GTAATTGTGGAGACGCTG</td>
<td>3AR9</td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>TTGTTACCTTGCCCGGATA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>3AF10</td>
<td>CTTGAACATTTTTTTCTGC</td>
<td>S3AR3-1</td>
</tr>
<tr>
<td></td>
<td>ATTACCTTAGTTAACCGAGA</td>
<td>GGGATTCCAAACACGCAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TA</td>
</tr>
<tr>
<td>3AF11</td>
<td>TTTGTCAATTCTTAGTTA</td>
<td>S3AR3-2</td>
</tr>
<tr>
<td></td>
<td>ACCGAGAAGGATTAC</td>
<td>TGTGTTGATGCTTTCAAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAAAAA</td>
</tr>
<tr>
<td>S3AF3-1</td>
<td>GGGATTCCAAACACGCAA</td>
<td>S3AR3-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCATGGTGATGCAAGTAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTG</td>
</tr>
<tr>
<td>S3AF3-2</td>
<td>GCCACAGAACTCACCATC</td>
<td>S3AR1</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>GGGATAACAAAGCAGTGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTG</td>
</tr>
<tr>
<td>S3AF3-3</td>
<td>TGGTTGGGTACCAGTTTTC</td>
<td>S3AR2</td>
</tr>
<tr>
<td></td>
<td>TTC</td>
<td>CAGATTTGTTCCACATCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAA</td>
</tr>
<tr>
<td>S3AF2</td>
<td>GCCATGGAAGCTTTTCAGAC</td>
<td>S3AR3</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>GGGAACAGATGCTGTAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GG</td>
</tr>
<tr>
<td>S3AF3</td>
<td>TCCTGGGTTCCAGTATGTGA</td>
<td>S3AR4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGGGCTCTTGAGAACTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAA</td>
</tr>
<tr>
<td>S3AF3-2B</td>
<td>TTCGAGGAAGAATTAGCCA</td>
<td>S3AR3-2B</td>
</tr>
<tr>
<td></td>
<td>CA</td>
<td>TTTCAGTAAGATGTTGAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AGGATG</td>
</tr>
<tr>
<td>S3AF3-4B</td>
<td>GTATCGCGGGAAGGTACA</td>
<td>S3AR3-3B</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>ACGAAGTGGGTTGAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CATGG</td>
</tr>
<tr>
<td>S3AF3-5B</td>
<td>CTTGCATTCCCCACCATATT</td>
<td>S3AR3-4B</td>
</tr>
<tr>
<td></td>
<td>TCT</td>
<td>ATTTTGCTGTGTTGCTG</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>TC</td>
</tr>
<tr>
<td>S3AF3-6B</td>
<td>CAAAATGGGACGAGAATG</td>
<td>S3AR3-5B</td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>TGGGTTGAAACTTTCCCAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>S3AF3-7B</td>
<td>TATATACAGTTTTTACCAT</td>
<td>S3AR3-5B</td>
</tr>
<tr>
<td></td>
<td>GCTTTGGG</td>
<td>TGGGTTGAACCTTTCCCAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>S3AF3-8B</td>
<td>ACTGCACCGGAGACTGAG</td>
<td>S3AR3-5B</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>TGGGTTGAAACTTTCCCAA</td>
</tr>
</tbody>
</table>

159
Table 3.2. Primers used for amplification and sequencing of FAD2-1 and FAD3 genes. Primers are listed in 5’→3’ configuration (Continued)

<table>
<thead>
<tr>
<th>Primers used for sequencing only</th>
<th>forward sequence</th>
<th>reverse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FAD3B gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3BF7</td>
<td>GCTTATATGGTAAAGTGATTTTC</td>
<td>S3BR1 TGCATCTTAACCGGTGGAG</td>
</tr>
<tr>
<td>3BF8</td>
<td>ATTGAACAGTGGCCATGGAA</td>
<td>S3BR2 GCCAGTATGAAACTAACCATCCA</td>
</tr>
<tr>
<td>S3BF2</td>
<td>GAAATAGGGGGAGGAGCAG</td>
<td>S3BR3 AAGAACAGCCTCAGGAAGCA</td>
</tr>
<tr>
<td>S3BF3</td>
<td>TTCAAGGCACAATGTCTTG</td>
<td>S3BR4 TGCACATGATCAAAACAATTACA</td>
</tr>
<tr>
<td>S3BF4</td>
<td>TGCTAGGAAACCATTCA GG</td>
<td>S3BR5 CGCAATAAAAGAAAGAGAACAAA</td>
</tr>
<tr>
<td>S3BF5</td>
<td>TGACTTTGGAAGACAACAA CACA</td>
<td></td>
</tr>
<tr>
<td>S3BF6</td>
<td>GGTGAACACTTTAAATGC TAGATAGTA</td>
<td></td>
</tr>
<tr>
<td><strong>FAD2-1B gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S21BF1</td>
<td>CCCGCTGTCCCTTTTAAACT</td>
<td>S21BR1 TGCACGATTCAATTTGGAGA</td>
</tr>
<tr>
<td>S21BF2</td>
<td>CATTTTGCTGTTACTTTTGG</td>
<td>S21BR2 AATGAAGTGAGGAAGGAAGCG</td>
</tr>
<tr>
<td>S21BF3</td>
<td>CACCATGCCTTCAGCAAGTA</td>
<td>S21BR3 CAAAAGCCTCAGCATTTGA</td>
</tr>
<tr>
<td>S21BF4</td>
<td>GCACACACACTATGCCCTTGC</td>
<td>S21BR4 CACTATGGCCATTTGCTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S21BR5 AAGTCATTACGCAGGAACAAA TC</td>
</tr>
</tbody>
</table>
Table 3.3. Primers and SimpleProbes (and their final concentration in the PCR mix) used to track Williams 82 and PI 366121 alleles of the *FAD3A* gene (7% lin population) and B1-52 and PI 366121 alleles of *FAD2-1B, FAD3A, FAD3B* and *FAD3C* genes (1% lin population). The underlined nucleotide in the SimpleProbe represents the polymorphic SNP.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
<th>SimpleProbe</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>FAD3A</em> (7% lin)</td>
<td>5’-GCCATGGAAGCTTTTCAGAC-3’ (0.5 µM)</td>
<td>5’-AAAGCCGGATGAATACCAC-3’ (0.1 µM)</td>
<td>5’-Fluorescein-SPC- GTGCCCACCAGGCTATTCAAG-Phosphate-3’ (0.2 µM)</td>
</tr>
<tr>
<td><em>FAD2-1B</em> (1% lin)</td>
<td>5’-TCCCCAAAACCAAATCCAAA-3’ (0.1 µM)</td>
<td>5’-TTGCAACCGGTAGCAAG-3’ (0.5 µM)</td>
<td>5’-Fluorescein-SPC- TGCTCCCATATTTCAACTCGAGG-Phosphate-3’ (0.2 µM)</td>
</tr>
<tr>
<td><em>FAD3A</em> (1% lin)</td>
<td>5’-TTGCATCATCACCATGGTACATC-3’ (0.5 µM)</td>
<td>5’-AGCTATTACGTAGATACCTAC-3’ (0.2 µM)</td>
<td>5’-Fluorescein-SPC- GTATCCTGCGCCTACACCA-Phosphate-3’ (0.2 µM)</td>
</tr>
<tr>
<td><em>FAD3B</em> (1% lin)</td>
<td>5’-AGCCACAGAAGCTCACCATCAA-3’ (0.5 µM)</td>
<td>5’-TGCGACGATCTAATG-3’ (0.2 µM)</td>
<td>5’-Fluorescein-SPC- GTCCTTTGTTGAACAGAATG-Phosphate-3’ (0.2 µM)</td>
</tr>
<tr>
<td><em>FAD3C</em> (1% lin)</td>
<td>5’-GTCTTTTGTGAACAGCATT-3’ (0.5 µM)</td>
<td>5’-CTCTGCAAAATTCATAGTTGT-3’ (0.2 µM)</td>
<td>5’-Fluorescein-SPC- AGGACCACGACATCATGGTACAGAAT-Phosphate-3’ (0.2 µM)</td>
</tr>
</tbody>
</table>
Table 3.4. Primers used in qRT-PCR experiments to determine the expression of \textit{FAD3A}, \textit{FAD3B} and \textit{FAD3C} genes of \textit{Glycine max} variety Williams 82 and \textit{Glycine soja} accession PI 366121 relative to the expression of the reference gene \textit{cons7}.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{FAD3A}</td>
<td>5'‐AGCGACACAAAGCAGCAAAAT-3'</td>
<td>5'‐GTCTCGGTGCGAGTGAAGGT-3'</td>
</tr>
<tr>
<td>\textit{FAD3B}</td>
<td>5'‐CTGCAGCAATTCACTTGCAC-3'</td>
<td>5'‐CCAGAACATTGTGCCTTGAA-3'</td>
</tr>
<tr>
<td>\textit{FAD3C}</td>
<td>5'‐CTCAGAAATCTGGGCCATTG-3'</td>
<td>5'‐TCGCTAACGAAGTGCCTTGAA-3'</td>
</tr>
<tr>
<td>\textit{cons7}</td>
<td>5'‐ATGAATGACGTTCCCATGTA-3'</td>
<td>5'‐GGCATTAAGGCAGCTCCTCTC-3'</td>
</tr>
</tbody>
</table>
Table 3.5. Polymorphisms in coding and non coding sequence of FAD2-1 and FAD3 genes between *Glycine soja* line PI 366121 and reference *Glycine max* cultivar Williams 82. Changes at the protein level are in parenthesis after the corresponding polymorphism.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coding</th>
<th>Non coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAD2-1A</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>FAD2-1B</td>
<td>A782G (M126V), A946G, C1063T, T1075C, T1088C, A1324G</td>
<td>C1677T</td>
</tr>
<tr>
<td>FAD3A</td>
<td>A494G</td>
<td>T313C, -1035T, -1064T, -1190T, T1550A, TATT deleted at nt 1562, A1571T, G3502C, -3924T</td>
</tr>
<tr>
<td>FAD3B</td>
<td>T1392C</td>
<td>G610T, C738T, TCTTTACAC inserted at nt 1931, G3166C, T3168C, C3190T, C3262T</td>
</tr>
<tr>
<td>FAD3C</td>
<td>none</td>
<td>ATGTT deleted at nt 338</td>
</tr>
<tr>
<td>FAD3D</td>
<td>none</td>
<td>A2013T</td>
</tr>
</tbody>
</table>

*Positions are relative to start codon, where A = 1
Table 3.6. Mean differences in seed 18:3 level in F₂ progeny of B1-52 x PI 366121 (1% lin) cross. Effect of substitution of PI 366121 (uppercase) and B1-52 (lowercase) alleles of FAD3A, FAD3B and FAD3C genes on 18:3 content. Differences were calculated by subtraction of 18:3 level in one genotypic class from the 18:3 level in the other genotypic class. Negative (-) and positive (+) signs represent decrease and increase in 18:3, respectively.

<table>
<thead>
<tr>
<th>Primary locus</th>
<th>Secondary loci</th>
<th>Difference in 18:3 content (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA to aa</td>
<td>AA to Aa</td>
<td>Aa to aa</td>
</tr>
<tr>
<td>BBCC</td>
<td>ND</td>
<td>-2.4</td>
<td>ND</td>
</tr>
<tr>
<td>BBc</td>
<td>-4.3</td>
<td>-2.2</td>
<td>-2.1</td>
</tr>
<tr>
<td>BBcc</td>
<td>-3.9</td>
<td>-0.5</td>
<td>-3.4</td>
</tr>
<tr>
<td>BBcC</td>
<td>-3.9</td>
<td>-1.2</td>
<td>-2.7</td>
</tr>
<tr>
<td>BbCc</td>
<td>-4.2</td>
<td>-1.4</td>
<td>-2.8</td>
</tr>
<tr>
<td>BbcC</td>
<td>-2.9</td>
<td>-1.3</td>
<td>-1.6</td>
</tr>
<tr>
<td>bbCC</td>
<td>-4.0</td>
<td>-0.1</td>
<td>-3.9</td>
</tr>
<tr>
<td>bbCc</td>
<td>-4.4</td>
<td>-1.1</td>
<td>-3.3</td>
</tr>
<tr>
<td>bbcc</td>
<td>-3.4</td>
<td>-1.2</td>
<td>-2.2</td>
</tr>
<tr>
<td>BB to bb</td>
<td>BB to Bb</td>
<td>Bb to bb</td>
<td></td>
</tr>
<tr>
<td>AACC</td>
<td>-2.2</td>
<td>-1.5</td>
<td>-0.7</td>
</tr>
<tr>
<td>AACc</td>
<td>-1.1</td>
<td>-0.7</td>
<td>-0.4</td>
</tr>
<tr>
<td>AAcc</td>
<td>-2.3</td>
<td>-1.2</td>
<td>-1.1</td>
</tr>
<tr>
<td>AaCC</td>
<td>+0.1</td>
<td>-0.3</td>
<td>+0.4</td>
</tr>
<tr>
<td>AaCc</td>
<td>0</td>
<td>+0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>Aacc</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>aaCC</td>
<td>ND</td>
<td>ND</td>
<td>-0.8</td>
</tr>
<tr>
<td>aaCc</td>
<td>-1.2</td>
<td>-0.6</td>
<td>-0.6</td>
</tr>
<tr>
<td>aacc</td>
<td>-1.8</td>
<td>-0.2</td>
<td>-1.6</td>
</tr>
<tr>
<td>CC to cc</td>
<td>CC to Cc</td>
<td>Cc to cc</td>
<td></td>
</tr>
<tr>
<td>AABB</td>
<td>-1.8</td>
<td>-1.2</td>
<td>-0.6</td>
</tr>
<tr>
<td>AAbb</td>
<td>-1.5</td>
<td>-0.4</td>
<td>-1.1</td>
</tr>
<tr>
<td>AAbb</td>
<td>-1.9</td>
<td>-0.1</td>
<td>-1.8</td>
</tr>
<tr>
<td>AaBB</td>
<td>+0.1</td>
<td>-1</td>
<td>+1.1</td>
</tr>
<tr>
<td>AaBb</td>
<td>-1.6</td>
<td>-0.6</td>
<td>-1</td>
</tr>
<tr>
<td>Aabb</td>
<td>-3</td>
<td>-1.1</td>
<td>-1.9</td>
</tr>
<tr>
<td>aaBB</td>
<td>ND</td>
<td>ND</td>
<td>-0.2</td>
</tr>
<tr>
<td>aaBb</td>
<td>-0.5</td>
<td>-0.7</td>
<td>+0.2</td>
</tr>
<tr>
<td>aabb</td>
<td>-1.3</td>
<td>-0.5</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

ND=not determined


Hoshino T, Takagi Y, Anai T (2010) Novel GmFAD2-1b mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with GmFAD2-1a mutant alleles. Breeding Sci 60:419-425


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 Biol 10:195


Zhang M, Fan J, Taylor DC, Ohlrogge JB (2009) DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21:3885-3901

Chapter 4

Genetics of High Linolenic Acid in Seed Oil of *Glycine soja*: Role of Acyltransferase and Cholinephosphotransferase Genes
Introduction

Importance of soybean and soybean oil

Soybean [*Glycine max* (L.) Merr.] is the leading oilseed crop produced and consumed in the world, representing 58% and 43% of world oilseed production and consumption, respectively (SoyStats 2011; Wilcox 2004). Soybean is the United States' second largest crop in cash sales and the number one value crop export (SoyStats 2011). Economic importance of soybean oil is given by its utilization in the manufacture of both edible and industrial products. Approximately 70% of the edible fats and oils consumed in the US as salad and cooking oil, baking and frying fats and margarine and spreads derive from soybean. Also, ~14% of the soybean oil utilized in the US is for industrial purposes such as the production of biodiesel, inks, paints, varnishes, resins and plastics (Cahoon 2003; SoyStats 2011).

Lipids represent one of the most important components in soybean seeds, comprising about 20% of their weight (Gerde and White 2008). The predominant fatty acid (FA) composition of commodity soybean oil is about 11% palmitic acid (16:0), 4% stearic acid (18:0), 23% oleic acid (18:1), 54% linoleic acid (18:2) and 8% linolenic acid (18:3) (Fehr 2007; Hill et al. 2008; Wilson 2004). For reference, the first number of these FA abbreviations indicates the number of carbon atoms, and the second number represents the number of double bonds in the molecule. 16:0 and 18:0 are saturated FAs,
18:1 is a monounsaturated FA, and 18:2 and 18:3 are polyunsaturated FAs (PUFA). Also, 18:2 and 18:3 are ω-6 and ω-3 types of FAs, respectively, which are essential to humans and all animals. The term essential refers to any compound that is required for life but cannot be synthesized by the organism and therefore must be incorporated with diet.

**Potential uses of soybean oil enriched in ω-3 fatty acids**

The balance between dietary ω-3 and ω-6 FAs strongly affects their function in the body. Currently, nutritionists are concerned over the reduction of the nutritionally desirable 18:3 levels in western diets and more particularly the ratio of ω-3:ω-6 FAs, which should ideally be 1:2.3 (Gerde and White 2008; Kris-Etherton et al. 2000). According to reports of the University of Maryland Medical Center (UMMC), the typical American diet tends to contain 14-25 times more ω-6 than ω-3 FAs (http://www.umm.edu/altmed/articles/omega-3-000316.htm).

ω-3 FAs play a crucial role in brain function as well as normal growth and development. They also reduce inflammation and may help lower risk of diseases such as heart disease, cancer, type II diabetes and arthritis (UMMC 2011). ω-3 FAs are highly concentrated in the brain and appear to be important for cognitive (brain memory and performance) and behavioral function (UMMC 2011). Because of these health benefits, elevated 18:3 soybean genotypes are desirable, especially in food grade soybeans (Lee et al. 2007).
Also, there has been a growing interest from the aquaculture industry in using plant-based proteins and oils as components of fish diets as a way to reduce both feeding costs and, especially, the pressure on marine resources. Soybean protein concentrate and oil, due to their excellent nutritional characteristics, are considered very suitable for aquafeed. Increasing global demand for fish high in long-chain ω-3 oils creates perfect market opportunities for soybeans with enhanced ω-3 content (Naylor et al. 2009). In addition, oils low in saturates and high in PUFAs would have applications as drying oils for paints, coatings and other industrial products (Wilson 2004).

**Fatty acid and Triacylglycerol synthesis**

Fatty acid biosynthesis is part of the triacylglycerol (TAG) synthetic pathway. A general scheme of FA and TAG synthesis in developing seeds is presented in Fig. 4.1. In plastids, FAs are synthesized from acetyl-CoA, which is carboxylated to form malonyl-CoA through the action of acetyl-CoA carboxylase (ACCase). A second enzyme system, fatty acid synthase (FAS), transfers the malonyl moiety to acyl carrier protein (ACP) and catalyzes the extension of the growing acyl chain with malonyl-ACP to form 16:0-ACP. Finally, elongation and desaturation of 16:0-ACP to form 18:0-ACP and 18:1-ACP, respectively, results in *de novo* synthesized FAs. By the action of ACP-thioesterases, the FA molecules are released from ACP in the plastid stroma and cross the membrane by an unknown mechanism. On the plastid outer membrane, an acyl-CoA synthetase (ACS) assembles acyl-CoA esters that are then available for acyltransferase reactions in the
In the endoplasmic reticulum (ER), synthesized FAs are incorporated onto a glycerol backbone to form TAG through the Kennedy pathway. Two successive acyl-CoA-dependent acylation reactions transfer FAs to glycerol-3-phosphate (G3P) to produce lysophosphatidic (LPA) and phosphatidic (PA) acids, respectively. PA dephosphorylation results in diacylglycerol (DAG), which is at the branch point of the pathway between TAG and phosphatidylcholine (PC) formation. The acyl-CoA-dependent acylation of DAG, catalyzed by diacylglycerol acyltransferase (DGAT), leads to TAG, while the transfer of a phosphocholine into DAG, catalyzed by CDP-choline:1,2-diacylglycerol cholinephosphotransferase (CPT), results in PC formation. A recently discovered enzyme, phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT) catalyzes the interconversion between DAG and PC by phosphocholine headgroup exchange (Lu et al. 2009). The FA desaturation (FAD) pathways, catalyzed by FAD2 and FAD3 desaturases, utilize PC as a substrate increasing the number of unsaturations of its acyl chains. The PUFA-enriched PC, by the reverse reaction of CPT and PDCT, can be transformed back into DAG. Also, the acyl chain at position 2 of PC may undergo exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). PC can also participate in an acyl-CoA-independent reaction to produce TAG. In this case, phospholipid:diacylglycerol acyltransferase (PDAT) catalyzes the formation of TAG using DAG and PC as substrates. Finally, a second acyl-CoA-independent reaction leading to TAG synthesis has been described, where diacylglycerol transacylase (DGTA) catalyzes the transfer of an acyl moiety between two DAG molecules to form TAG and
monoacylglycerol (MAG) (Lung and Weselake 2006; Ohlrogge and Browse 1995; Ohlrogge and Jaworski 1997).

**Acyltransferase genes in plants**

Fatty acids are the main components of plant membrane and seed storage lipids (Töpfer et al. 1995). Plant oils consist mainly of TAGs, the major storage form, which are composed of three fatty acids bound to a glycerol backbone (Lu et al. 2011). In most plant tissues, over 75% of the FAs are unsaturated (Hildebrand et al. 2008). The FA esterification reactions that lead to the formation of TAG are catalyzed by five types of acyltransferases. GPAT, LPAAT and DGAT are the acyltransferases involved in the Kennedy pathway (Fig. 4.1). In a second pathway, the Lands’ cycle, acyl groups are incorporated into PC through the activity of LPCAT, and then subsequently channeled from PC to TAG via PDAT (Fig. 4.1). It has been suggested that the reverse reaction of LPCAT may play a key role in exchanging “remodeled” FAs from phospholipids such as PC back into the cytosolic acyl-CoA pool, making these FAs available for any reaction catalyzed by acyl-CoA dependent transferases (Cahoon et al. 2007; Dyer et al. 2008; Napier and Graham 2010). However, plant orthologues of the recently identified mammalian LPCAT genes have yet to be discovered (Napier and Graham 2010). Two of the acyltransferases discussed above, DGAT and PDAT, have shown to be important in both the determination of total oil content as well as fatty acid composition in seeds (Katavic et al. 1995; Zhang et al. 2009; Zheng et al. 2008).
Cholinephosphotransferase genes in plants

There are at least two types of cholinephosphotransferases in plants, and both of them catalyze reactions involving DAG and PC: CPT and PDCT (Fig. 4.1). CPT plays a role in the production of de novo PC, and it transfers a phosphocholine group from cytidine diphosphate (CDP)-choline to DAG, generating as a result cytidine monophosphate and the phospholipid PC (Lu et al. 2009; Lung and Weselake 2006). The reverse reaction of CPT moves a more unsaturated DAG into the mainstream of TAG synthesis (Weselake et al. 2009). PDCT catalyzes the transfer of the phosphocholine group from PC to DAG, allowing the incorporation of 18:1 into PC for further desaturation as well as the transfer of 18:2 and 18:3 into DAG molecules that then enter the TAG synthesis pathway (Lu et al. 2009). In developing seeds, most of the conversion of DAG to PC would be carried out by PDCT (Lu et al. 2009). Of the two cholinephosphotransferases, only PDCT has been shown to play a major role in the FA composition of Arabidopsis seed oil (Lu et al. 2009).

Role of acyltransferase and cholinephosphotransferase genes in oil quality

Studies on the role of DGAT1 and PDAT1 in TAG biosynthesis have shown these genes contributing the most to TAG accumulation in Arabidopsis thaliana developing seeds (Zhang et al. 2009). While pdat1 single mutants presented unaltered oil content and FA profile, mutants for the DGAT1 gene showed a 20 to 40% decrease in TAG content and a significant increase in 18:3 content at the expense of 18:1 and 20:1
(eicosenoic acid) (Katavic et al. 1995; Zhang et al. 2009). *PDAT1* RNAi on *dgatl* background as well as the reciprocal combination resulted in a severe phenotype with an absolute oil reduction of ~84% compared with the *dgatl* control. Also, oil in these individuals showed a decrease in 18:3 but retained the low 18:1 and 20:1 proportions compared with *dgatl* control (Zhang et al. 2009).

Even though both DGAT1 and PDAT1 can catalyze the final DAG acylation, the acyl composition of the resulting TAG will depend on the relative contribution of these two enzymes to the process, provided that identical DAG molecules are utilized as substrate. In *Arabidopsis*, DGAT1 shows a preference for either saturated or long-chain FAs (>C18) (Katavic et al. 1995). On the contrary, Stahl et al. (2004) demonstrated that *Arabidopsis* PDAT1 can utilize different phospholipids as acyl donors, exhibiting the highest activity for acyl groups containing several double bonds, epoxy or hydroxyl groups. The prevalence of saturated and long-chain FAs at the third position of the TAG molecule (~70%) suggests that DGAT1 contributes more than PDAT1 in *Arabidopsis* seeds (Katavic et al. 1995; Zhang et al. 2009).

Comparison between transcript levels of *DGAT1*, *DGAT2* and *PDAT1* with oil and FA accumulation in developing seeds of several plant species suggests that DGAT1 is the major enzyme for oil accumulation, especially in soybean and *Arabidopsis*, while DGAT2 and PDAT1 enzymes play important roles in the accumulation of unusual FAs (Li et al. 2010). Also, according to Li et al. (2010), *DGAT2* and *PDAT1* low expression levels at stages of high oil synthesis would suggest no association between oil accumulation and the function of these two genes in soybean and *Arabidopsis* seeds.
However, Zhang et al. (2009) previously demonstrated that on a \textit{dgat1} mutant background, \textit{PDAT1} could complement up to 80\% the function of \textit{DGAT1} in \textit{Arabidopsis} developing seeds.

PDCT has shown to play a major role in the final FA composition of seed TAG as well. A comprehensive analysis of an \textit{Arabidopsis PDCT} mutant indicated that the mutation significantly reduces the transfer of 18:1 into PC for desaturation as well as the reverse transfer of 18:2 and 18:3 into DAG. As a result of this deficiency, there is a 40\% decrease in the content of PUFA and a corresponding increase in 18:1 in TAG. Genetic studies on the \textit{PDCT} mutant indicate that the mutation behaves as a single, recessive Mendelian allele (Lu et al. 2009).

The relative contribution of GPAT, LPAAT and CPT to the overall TAG composition in seeds is not very well known. GPAT seems to have a broad specificity, using saturated, monounsaturated and polyunsaturated FAs (Christie et al. 1991; Lísa and Holcapek 2008). Therefore, acyl composition at position 1 in TAGs would depend on the acyl-CoA availability rather than GPAT specificity (Voelker and Kinney 2001). After studying the substrate selectivity of plant and microbial LPAATs, Brown et al. (2002) concluded that there is little selectivity between 18:1- and 18:2-CoA donor substrates, both incorporated into PA at significant rates, but 16:0 and 18:0 are selected against. However, no clear correlation was observed between the amount of 18:2 or 18:1 at position 2 of TAGs and the major LPAAT activity towards these substrates, leading Brown et al. (2002) to conclude that the availability of substrates or subsequent remodeling reactions are more important factors in determining the final oil composition.
Acyltransferase and cholinephosphotransferase genes in soybean

The soybean genome is complex and has a predicted size of 1,115 Mb distributed across 20 chromosomes (Arumuganathan and Earle 1991; Schmutz et al. 2010). There is evidence suggesting that soybean underwent two whole-genome duplication events followed by gene diversification and loss and chromosome rearrangements. The result is a highly duplicated genome with approximately 75% of the genes existing as paralogues (Schmutz et al. 2010). Therefore, in soybean, it is very likely to find multiple copies of a gene that is present as a single copy in other species like *Arabidopsis*.

Based on the available scientific evidence at the moment this project started, it was decided to focus the research efforts on those transferases that appeared to be important in the determination of FA composition of seed oil. Several studies report *DGAT*, *PDAT* and *PDCT* playing significant roles in seed oil quality (Katavic et al. 1995; Li et al. 2010; Lu et al. 2009; Zhang et al. 2009; Zheng et al. 2008), and thus the involvement of those genes in the high 18:3 content in *G. soja* seeds was investigated.

In soybean, there are at least two *DGAT1* genes, namely *DGAT1A* and *DGAT1B*, and their full length cDNAs were cloned from developing seeds (Hildebrand et al. 2008; Wang et al. 2006). These genes showed maximum transcript levels at stages of high TAG synthesis (Hildebrand et al. 2008; Wang et al. 2006). In addition, alignment of the DGAT1A deduced amino acid sequence from seven *G. max* and five *G. soja* accessions showed similarities ranging from 97.8 to 100%, in other words, there is a high conservation of *DGAT1* genes. Also, the transcript abundance of the *DGAT1A* gene
between *G. max* and *G. latifolia* (one of the 23 perennial wild relatives) accessions had no significant difference. Results from the gene structural and expression studies led Wang et al. (2006) to conclude that *DGAT1A* is not associated to differences in oil content between soybean accessions.

It is not known how many functional *PDAT* and *PDCT* genes are present in the soybean genome. Li et al. (2010) reported that a soybean homolog of *AtPDAT1* (At5g13640) was expressed at very low levels throughout seed development, ruling out this gene as an important player in oil accumulation in soybean seeds. However, there are at least six soybean homologs of At5g13640, whose similarities to *AtPDAT1* ranges from 60 to 86%, and we were not able to identify which of those genes was studied by Li et al. (2010).

**Linolenic acid content in the soybean wild ancestor**

Accessions of *Glycine soja* (Siebold & Zucc.), the wild ancestor of cultivated soybean, have been reported to contain as much as 23% 18:3, representing potentially useful genetic resources for identifying either new alleles or genes that govern 18:3 accumulation (Pantalone et al. 1997; Wilson 2004). However, very little is known about the genetic regulation of high 18:3 in *G. soja*. Results of a survey on the average oil content and FA profile of *G. max* and *G. soja* accessions in the USDA’s National Plant Germplasm System (NPGS) (http://www.ars-grin.gov/npgs/) are shown in Table 4.1. On average, *G. soja* produces twice as much 18:3 but almost half the amount of 18:1 and oil
as cultivated soybean. Also, *G. soja* presents higher variability for 18:3 than *G. max*, whereas *G. max* surpasses *G. soja* in variability for oil, 18:1 and 18:2. Analysis of relative 18:1 and 18:2 desaturation patterns, which would reflect ω-6 and ω-3 desaturases’ activity, respectively, in four *G. max x G. soja* populations suggested that *G. soja* carries superior alternative *FAD2* and *FAD3* alleles directing the expression of the high 18:3 trait (Pantalone et al. 1997). Also, increasing the expression of the *FAD3* genes by genetic engineering, allowed the generation of soybean seeds that accumulated 18:3 in excess of 50% of the total oil (Cahoon 2003). However, structural and expression studies of the *G. max* and *G. soja* alleles of *FAD2* and *FAD3* desaturase genes showed that their contributions to 18:3 content in seed oil are similar (Lenis, Chapter 3 of this work).

The objective of this work was to gain a better understanding of the genetics of 18:3 accumulation in seeds of *G. soja* in general, and the role of some of the microsomal acyltransferase (*DGAT1* and *PDAT1*) and cholinephosphotransferase (*PDCT*) genes in the high 18:3 content of PI 366121 in particular. We hypothesized that elevated 18:3 in *G. soja* is the result of differences in gene(s) that code for enzymes that either directly or indirectly favor the incorporation of 18:3 into TAG.
Materials and Methods

Plant material and development of segregating populations

Study of the genetic basis of high 18:3 in seeds of soybean wild ancestor was carried out with G. soja accession PI 366121 and G. max genotype Williams 82. PI 366121 is listed as a high 18:3 content entry (up to 23% 18:3) in the USDA’s National Plant Germplasm System (http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1274060). Williams 82 (Bernard and Cremeens 1988) is the reference soybean genotype in genomic studies and the chosen cultivar for the construction of the soybean whole-genome shotgun sequence assembly (Schmutz et al. 2010). Williams 82 presents a typical 18:3 content (~7-8%) for a soybean line.

Glycine soja accession PI 366121 was crossed with G. max genotype Williams 82 to generate a segregating population, which will be referred to as the 7% lin population. F2 seeds were produced in growth chamber and planted to generate F3 seeds in Columbia, Missouri in the summer of 2009.

Sampling DNA and determinations on seeds

FTA PlantSaver cards (Whatman Inc., Florham Park, NJ) were used to take DNA samples from ~110 plants from the 7% lin population in F2 generational advance,
according to manufacturer’s recommendations. Before planting, single seeds were evaluated for oil composition by cutting a seed chip (~1/3 of the seed) from each seed. The remainder of the seed was used for planting.

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 as explained in Bilyeu et al. (2005). The individual fatty acid contents are reported as the relative percents of palmitic, stearic, oleic, linoleic, and linolenic acids in the extracted oil.

Candidate genes

We considered two types of candidate genes based on where in the TAG pathway their products carry out the catalytic action: desaturases (ω-3 and -6) and transferases (acyltransferases and cholinephosphotransferase). In this chapter, only transferases will be considered.

Literature on *Arabidopsis* has shown that various enzymes participating in TAG formation may influence fatty acid composition through substrate preference (Snyder et al. 2009). Considering the available information at the moment this research was initiated, the following soybean genes were considered strong candidate genes: *DGAT1A* (Glyma13g16560), *DGAT1B* (Glyma17g06120), soybean homologs of *Arabidopsis PDAT1* (Glyma07g04080 and Glyma13g16790) and soybean homologs of *Arabidopsis PDCT* (Glyma07g03350 and Glyma08g22750).
DNA isolation and PCR for sequencing candidate genes

Genomic DNA was isolated from ~30 mg dried seed tissue using the DNeasy Plant Mini Kit (Qiagen Sciences Inc., Germantown, MD) and used at 5 to 50 ng per PCR amplification. PCR was carried out using Ex Taq according to manufacturer’s recommendations (TaKaRa, Otsu, Shiga, Japan) in a PTC-200 thermocycler (MJ Research/Bio-Rad, Hercules, CA) using the following conditions: 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for as long as needed in each case (for the extension step of PCR, we added 30 s/500 bp of genomic sequence). PCR products were analyzed by gel electrophoresis (1% agarose gel) to verify for size and ensure specific amplification. PCR products were isolated with the QIAprep Spin Miniprep kit (Qiagen Sciences Inc.) and sequenced at the University of Missouri DNA Core facility. Primers used for amplification and sequencing are listed in Table 4.2.

Sequence analysis

Sequences were imported into ContigExpress program of Vector NTI Advance 10 (Invitrogen, Carlsbad, CA), trimmed, assembled and manually evaluated for disagreements between Williams 82 (Bernard and Cremeens 1988) reference sequence and imported sequence contigs. Putative SNPs and deletions were verified by at least two independent PCR amplifications. To evaluate the effect of changes in coding sequence at the protein level, the program ExPASy translate tool was used (http://ca.expasy.org/tools/dna.html). Sequence alignments were generated using the
AlignX software (Invitrogen). Gene models were obtained from the whole soybean genome assembly (http://www.phytozome.net/soybean). Exon-intron boundaries were verified for accuracy using TBLASTN program with the protein sequence of the corresponding homologous genes in Arabidopsis as queries and translated genomic sequence in six frames.

**Genotyping assays**

Williams 82 and PI 366121 alleles for the candidate transferase genes were identified in selected individuals of the 7% lin population by sequence analysis. Polymorphic fragments from the five individuals with the highest and lowest 18:3 content were amplified by means of PCR and sent to the University of Missouri DNA Core for sequencing. The primers used to amplify and then sequence the polymorphic fragments are listed in Table 4.3. PCR reactions were carried out under the following conditions: 95°C for 5 min, 40 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 1 min and 72°C for 5 min after the last cycle. Reactions were performed in a total volume of 15 µl containing 5-50 ng DNA template, primers (0.5 µM), buffer [40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 µg ml⁻¹ BSA, 200 µM dNTPs], 5% DMSO and 0.2X Titanium *Taq* polymerase (BD Biosciences).
Expression analysis of candidate genes

Standard curves to determine the efficiency of each primer pair were constructed using serial dilutions of genomic DNA as template (Pfaffl 2001). Total RNA was isolated from powdered Williams 82 or PI 366121 developing seeds by means of PureLink™ Plant RNA reagent (Invitrogen) following the manufacturer’s instructions. Total RNA was treated with DNase I and the supplied 10X DNase I buffer (Ambion Inc., Austin, TX) by incubating at 37°C for 30 min and then heating at 75°C for 15 min to inactivate the DNase. Effectiveness of DNase treatment on total RNA was evaluated by comparing amplification of samples using RT-PCR master mix with or without reverse transcriptase enzyme following the protocol described below.

The reaction mix consisted of QuantiTect SYBR Green RT-PCR master mix (Qiagen, Valencia, CA), supplemented with 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA), DNase-treated RNA template and 0.5 µM of each primer in 20 µL reactions. The experimental conditions were reverse transcription at 50°C for 30 min followed by 95°C for 15 min, then 35 cycles of 95°C for 20 s, 60°C for 20 s, 72°C for 20 s and an ending hold at 4°C. After cycling was completed and before the ending hold, a melting curve program was run to confirm the specificity of the amplification.

Cycle threshold measurements were mathematically converted into genome equivalents by application of the standard curves. Expression levels of candidate genes were expressed as a ratio to the expression level of reference gene cons7, a putative
insulin-degrading enzyme (Libault et al. 2008). Primers used for qRT-PCR experiments are listed in Table 4.4.
Results

Polymorphisms in candidate genes of *Glycine soja* line PI 366121

We used gene-specific PCR to amplify and sequence *DGAT1A* (Glyma13g16560) and *DGAT1B* (Glyma17g06120); two *PDAT* (Glyma07g04080 and Glyma13g16790) and two *PDCT* (Glyma07g03350 and Glyma08g22750) genes from *G. soja* line PI 366121. Genomic sequence from cultivar Williams 82 was used as reference sequence (http://www.phytozome.net/soybean).

Out of the 6 candidate genes for which genomic sequence was analyzed, two presented no polymorphisms at the coding sequence level (*DGAT1A* and *PDCT8*), while the remaining four showed silent and/or missense polymorphisms in the coding sequence (Table 4.5). The two missense polymorphisms observed in each *DGAT1B* and *PDCT7* occur in non conserved regions of their respective protein products, making it very unlikely that these changes affect either enzymatic activity or stability. *PDAT7* is the only gene that presented a missense polymorphism that has never been reported before (A649V); the alanine residue at position 649 is completely conserved in all plant proteins present in the non redundant protein database in NCBI when using PI 366121 *PDAT7* protein product as a query.
Association between high 18:3 content in seeds and genotype at the candidate genes

Sequence analysis of genomic DNA from selected F₂ individuals from the 7% lin population was carried out to determine if PI 366121 transferases were associated with high 18:3 content in seeds. The presence of the PI 366121 alleles for the DGAT1, PDAT and PDCT genes were not associated with an increased accumulation of 18:3 in seeds. Both groups, the one with the highest and the one with the lowest 18:3 contents, presented individuals with the G. soja line alleles (data not shown).

Gene expression studies

To investigate if the higher 18:3 content in PI 366121 seeds was due to overexpression of any of the candidate genes, quantitative RT-PCR experiments were carried out. Developing seeds from cultivar Williams 82 and accession PI 366121 were collected at 25 and 40 days after flowering (DAF) and used for total mRNA extraction and quantification. A standard curve was constructed for each pair of gene-specific primers to determine the efficiency of each primer pair. The data was obtained in technical triplicates and biological duplicates. The expression levels of the investigated genes were expressed as a ratio to the expression level of the reference gene cons7 (Libault et al. 2008).

Our results indicate that DGAT1 genes are expressed at higher levels in Williams 82 than in PI 366121, particularly DGAT1A. The differences in expression are more
important at earlier stages of seed development (25 DAF), when transcripts of \textit{DGAT1A} and \textit{DGAT1B} are 29- and 20-fold more abundant in \textit{G.max} than in \textit{G. soja}, respectively. At later stages of seed development (40 DAF), differences in expression level are smaller (4- and 9-fold higher for \textit{DGAT1A} and \textit{DGAT1B}, respectively) (Fig. 4.2a).

\textit{PDAT13} and \textit{PDAT7} transcripts were low in abundance. RNA levels for \textit{PDAT13} were 8-fold and 3-fold higher in \textit{G. max} than in \textit{G. soja} at early (25 DAF) and late (40DAF) seed developmental stages, respectively. No differences in transcript level were observed between soybean and its wild ancestor for \textit{PDAT7} (Fig. 4.2b).

In \textit{G. max}, \textit{DGAT1A} transcript accumulation was 109- and 81-fold higher than that of \textit{PDAT13} at 25 and 40 DAF, respectively while in \textit{G. soja}, \textit{DGAT1A} expression was 31 and 53 times higher than that of \textit{PDAT13} for the same developmental stages (Fig. 4.2). When comparing gene expression between \textit{DGAT1A} and \textit{PDAT7}, the differences are even more significant, with 1437- and 476-fold higher in \textit{G. max} and 83- and 87-fold higher in \textit{G. soja} at early and late seed development, respectively (Fig. 4.2).
Discussion

Seed oil content and composition are the final result of a complex and not completely understood network of biochemical reactions taking place during TAG biosynthesis in the ER. In addition to the inherent complexity of the process, it seems likely that the channeling of acyl substrates during TAG synthesis is configured differently from one plant species to the other, which makes difficult the extrapolation of results between plant species (Ruiz-López et al. 2009; Taylor et al. 2009). However, there are examples of successful modification of oil composition in crop species such as soybean by applying the knowledge obtained from research in model species (Bilyeu et al. 2003; Pham et al. 2010).

Accessions of *Glycine soja* (Siebold & Zucc.) have been reported to contain as much as 23% 18:3 in the seed oil, representing potentially useful genetic resources for identifying either new alleles or genes that govern 18:3 accumulation (Pantalone et al. 1997; Wilson 2004). Analysis of relative 18:1 and 18:2 desaturation patterns in four *G. max* x *G. soja* populations suggested that *G. soja* carries superior FAD2 and FAD3 alleles directing the expression of the high 18:3 trait (Pantalone et al. 1997). However, research carried out as part of this work indicates that *G. max* and its wild ancestor carry equivalent alleles for both the ω-3 and ω-6 desaturases, meaning that their contributions to 18:3 content in seeds are very similar (Lenis, Chapter 3 of this work). Also, gene expression studies indicate that the high 18:3 content in seeds of PI 366121 is not the
result of more transcriptionally active ω-3 and ω-6 desaturase genes (Lenis, Chapter 3 of this work).

The fact that differences in 18:3 content in seeds of *G. max* and *G. soja* are not due to differences either in structure or expression of the microsomal desaturases prompted us to look at biochemical steps committed in the incorporation of unsaturated FAs into TAG. Several reports have shown that the proportion of 18:3 in TAG can be affected by genes other than the desaturases (Katavic et al. 1995; Routaboul et al. 1999; Zhang et al. 2009; Zou et al. 1999). Katavic et al. (1995) and Zhang et al. (2009) have shown that genotypes of *Arabidopsis* carrying mutant *DGAT1* have decreased seed oil content and increased proportion of 18:3 at the expense of 18:1. Also, Zhang et al. (2009) demonstrated that *PDAT* plays an important role in TAG biosynthesis and that DGAT and PDAT are the primary enzymes for oil accumulation in *Arabidopsis* seeds. Based on the reports cited above, it was hypothesized that the differences in 18:3 and oil content in seeds of *G. soja* and *G. max* might be associated to differences in either structure or expression of *DGAT* and/or *PDAT* genes between those two species. In addition, the recently discovered enzyme PDCT played a major role in the incorporation of 18:1 into PC (substrate of desaturases) as well as the transfer of 18:2 and 18:3 into DAG in *Arabidopsis* (Lu et al. 2009). As much as 40% of the 18:1 that is further desaturated to 18:2 and 18:3 enters PC trough the reaction catalyzed by PDCT (Lu et al. 2009).

The *Arabidopsis* genome contains one *DGAT1*, one *DGAT2* and six *PDATs*, but only *DGAT1* and *PDAT1* have been shown to play a major role in TAG synthesis in *Arabidopsis* developing seeds (Li et al. 2010; Zhang et al. 2009). Also, only one *PDCT*
gene has been identified in *Arabidopsis* so far (Lu et al. 2009). Due to the two rounds of whole-genome duplication, soybean is very likely to carry multiple copies of a gene that is present as a single copy in *Arabidopsis*.

Based on amino acid sequence identity with the *Arabidopsis* homologs and expression data, *DGAT1A* (Glyma13g16560), *DGAT1B* (Glyma17g06120), soybean homologs of *Arabidopsis PDAT1* (Glyma07g04080 and Glyma13g16790, herein referred to as *PDAT7* and *PDAT13*, respectively) and soybean homologs of *Arabidopsis PDCT* (Glyma07g03350 and Glyma08g22750, herein referred to as *PDCT7* and *PDCT8*, respectively) were selected as candidates potentially underlying differences in 18:3 content between *G. max* and *G. soja*. A high degree of sequence conservation between *G. max* and *G. soja* transferase genes was found. Similar observations were reported by Wang et al. (2006), who compared the coding sequence of the *DGAT1A* gene from seven *G. max* and five *G. soja* accessions. Only one out of six chosen candidate genes showed a missense polymorphism in an extremely conserved region (A649V in *PDAT7*). However, this polymorphism did not associate with higher 18:3 contents in seeds of segregating individuals. Likewise, PI 366121 alleles of *DGAT1B*, *PDAT13* and *PDCT7* showed no association with increased 18:3 accumulation. In addition, a preliminary QTL analysis with F$_5$ RILs from a Williams 82 x PI 366121 cross seems to indicate that no region of chromosome 13, where *DGAT1A* is located, is associated to differences in 18:3 between *G. soja* and *G. max* (data not shown). The gene *PDCT8* remains untested for its association with the high 18:3 phenotype due to a lack of polymorphism between the parents.
Analysis of the expression of the transferase genes shows $DGAT1A$ and $DGAT1B$ expressed at higher levels in $G.\ max$ than in $G.\ soja$ at both early and late seed developmental stages. In $Arabidopsis$, mutation of $DGAT1$ results in a 20 to 40% decrease in oil content (Katavic et al. 1995; Routaboul et al. 1999). Also, in both $Arabidopsis$ and maize, a decrease of DGAT1 activity led to altered fatty acid composition, with increases in 18:3 at the expense of 18:1 and other fatty acids (Katavic et al. 1995; Routaboul et al. 1999; Zheng et al. 2008). By comparing the phenotypes observed in $Arabidopsis$ and maize and their corresponding $DGAT1$ status with the gene expression data from this study, it is suggested that the higher oil content and the lower 18:3 levels observed in seeds of $G.\ max$ as compared to $G.\ soja$ is explained, at least in part, by the higher expression of the $G.\ max\ DGAT1$ genes. Contrary to these findings, Wang et al. (2006) reported no significant difference in transcript abundance of $DGAT1A$ gene between $G.\ max$ and $G.\ latifolia$, a perennial wild relative. Interestingly, expression of $PDAT13$ was higher in $G.\ max$ than in $G.\ soja$ at all developmental stages while no differences were observed for $PDAT7$ between soybean and its wild ancestor. Based on the higher 18:3 content in seeds of $G.\ soja$, it was expected to see the $PDAT$ genes expressed at higher levels in the soybean wild ancestor: PDAT esterifies position 3 of DAG with acyl chains taken from position 2 of PC. Since PC is the substrate for the desaturases, there is a great chance that the acyl chains transferred by PDAT to DAG in the process of TAG synthesis are rich in unsaturations. Even though the selected $PDAT$ genes are not transcribed at higher levels in $G.\ soja$ than in $G.\ max$, the differences in gene expression between $PDAT$ and $DGAT$ are much smaller in $G.\ soja$ than in the cultivated soybean. As a result, the proportion of $PDAT$ transcripts relative to those of
DGAT is several times higher in *G. soja* than in *G. max* (Fig. 4.2). Therefore, the relative higher *PDAT* transcript abundance could explain the superior 18:3 content in seeds of the soybean wild ancestor *G. soja*.
Conclusions

Soybean is valued in the marketplace as a source of protein and oil. Soybean oil is broadly incorporated into food, feed and industrial products. Increased demand for vegetable oils and growing awareness of health issues around dietary fats make necessary a better understanding of the genetics underlying oil content and composition. Enhancement of \( \omega-3 \) FA content of the oil for both food and feed applications is highly desired. Accessions of \( G. \ soja \), the wild ancestor of soybean, contain, in average, twice as much \( \omega-3 \) FAs (18:3) as cultivated soybean, representing a potential source of superior genes/alleles. However, little is known about the genetics of the elevated 18:3 trait in wild soybean.

After learning that the \( FAD2-1 \) and \( FAD3 \) desaturase genes do not explain the differences in 18:3 content between \( G. \ soja \) and \( G. \ max \) (Lenis, Chapter 3 of this work), it was decided to change the focus towards genes that encode enzymes catalyzing other steps in the TAG synthesis pathway and that have the potential to modify the composition of oil. Enzymes participating directly or indirectly in the incorporation of 18:3 in TAG, where ~90% of 18:3 is, would be good candidates. The \( PDCT \) genes chosen for this research do not explain why \( G. \ soja \) contains almost twice as much 18:3 as \( G. \ max \). On the other hand, the transcriptionally more active \( G. \ max \ DGAT1A \) and \( DGAT1B \) genes is consistent with the higher oil content and lower 18:3 levels observed in \( G. \ max \) when compared with \( G. \ soja \). The higher expression of the soybean \( DGAT1A \) and \( DGAT1B \)
genes could be the result of differences in regulating factor(s) acting in “trans”, which remains to be tested. Additionally, the higher PDAT:DGAT transcript ratios observed in *G. soja* relative to *G. max* agree with the elevated 18:3 and low oil contents typically measured in seeds of the wild ancestor.

The relative contribution of the acyl editing mechanisms as well as other enzymes that take part in TAG synthesis to overall TAG composition still needs to be completely understood. As presented by Lu et al. (2011), the FA fluxes through PC are of great importance in the determination of TAG composition since PC is the source of all the PUFAs. The majority of the fatty acids synthesized in plastids enter PC by exchanging with PUFA in the acyl editing process (Bates et al. 2009; Bates et al. 2007). Also, LPCAT, by the reverse reaction, releases PUFAs that in turn will be part of the cytoplasmic acyl-CoA pool (Stymne and Stobart 1984). Because of the key role that LPCAT has in regenerating PC from LPC, and since PC is the only substrate for the desaturases, the involvement of LPCAT in determining oil composition could be significant. Another enzyme that could play an important role in TAG composition is PLA₂, which catalyzes the hydrolysis of acyl groups in position 2 of PC producing free PUFAs for activation by ACS (Weselake et al. 2009). Finally, the contribution of acyltransferases on FA composition through substrate preferences has been recently reviewed (Snyder et al. 2009). However, the actual input of each of these enzymes to seed oil quality is still uncertain as only *DGAT1* and *PDAT1* knockouts have been characterized in plants.
Detailed understanding of TAG metabolism and its regulation is still lacking. However, research on oil traits has been dynamic, particularly in the model plant *Arabidopsis*, and that will certainly have a positive impact in the improvement of oilseeds.
Figures
Figure 4.1. Fatty acid (FA) and triacylglycerol (TAG) synthesis in oilseed plants developing seeds. FA synthesis occurs in plastids. Synthesized FAs enter the cytosolic acyl-CoA pool and are used to form TAG from glycerol-3 phosphate (G3P) in the ER. The FA desaturation (FAD) pathways utilize phosphatidylcholine (PC) as a substrate. FAD2 enzymes catalyze the unsaturation of 18:1-PC to form 18:2-PC; FAD3 enzymes synthesize 18:3-PC from 18:2-PC. PUFA-enriched PC, can be transformed back into diacylglycerol (DAG). Also, the acyl chain at position 2 of PC may undergo exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; ACS, acyl-CoA synthetase; CPT, CDP-choline: 1,2-diacylglycerol cholinephosphotransferase; DGAT, diacylglycerol acyltransferase; DGTA, diacylglycerol transacylase; FAE, FA elongase; FAS, FA synthase; GPAT, glycerol-3-phosphate acyltransferase; LPA, lysophosphatidic acid; LPAAT, lysophosphatidic acid acyltransferase; LPC, lysophosphatidylcholine; PAP, phosphatidate phosphatase; PDAT, phospholipid:diacylglycerol acyltransferase; PDCT, phosphatidylcholine:diacylglycerol cholinephosphotransferase; PLA₂, phospholipase A₂; TS, acyl-ACP thioesterase (modified from Lung and Weselake 2006)
Figure 4.2. Relative expression of a DGAT1A and DGAT1B and b PDAT7 and PDAT13 genes in seeds of Williams 82 (Gm) and PI 366121 (Gs) at 25 and 40 days after flowering (DAF). Steady state mRNA levels for each gene were quantitated by quantitative PCR following reverse transcription of total RNA. The histograms represent the expression of DGAT1A and PDAT13 (white bars) and DGAT1B and PDAT7 (gray bars) as the ratio to the expression level of the reference gene cons7.
Tables
Table 4.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of *Glycine max* and *Glycine soja* accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean. Information on wild soybean accession PI 366121 and *G. max* genotype Williams 82 is also included.

<table>
<thead>
<tr>
<th></th>
<th>Oil (%)</th>
<th>16:0 (%)</th>
<th>18:0 (%)</th>
<th>18:1 (%)</th>
<th>18:2 (%)</th>
<th>18:3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G. max</strong></td>
<td>20 ± 2.1</td>
<td>11 ± 1.0</td>
<td>4 ± 0.6</td>
<td>23 ± 3.8</td>
<td>54 ± 3.7</td>
<td>8 ± 1.5</td>
</tr>
<tr>
<td><strong>G. soja</strong></td>
<td>11 ± 1.4</td>
<td>12 ± 0.9</td>
<td>4 ± 0.5</td>
<td>14 ± 2.4</td>
<td>54 ± 1.7</td>
<td>16 ± 2.2</td>
</tr>
<tr>
<td>PI 366121</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>11</td>
<td>54</td>
<td>20</td>
</tr>
<tr>
<td>(G. soja)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams 82</td>
<td>20</td>
<td>11</td>
<td>3</td>
<td>23</td>
<td>56</td>
<td>7</td>
</tr>
<tr>
<td>(G. max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* All observations for *G. max* accessions in the NPGS were used in the calculations (n=16,588)

*b* All observations for *G. soja* accessions in the NPGS were used in the calculations (n=1,242)
Table 4.2. Primers used for amplification and sequencing of DGAT1A, DGAT1B, PDAT7, PDAT13, PDCT7 and PDCT8 genes. Primers are listed in 5’→3’ configuration.

<table>
<thead>
<tr>
<th>Primers used for both amplification and sequencing</th>
<th>forward</th>
<th>sequence</th>
<th>reverse</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGAT1A gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F1</td>
<td>TCTTTTCCAATGGCGATTTCA</td>
<td>DGAT1A_R1-2</td>
<td>TCAAAACAAAGGTGCCTCA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_G2</td>
<td>GTGGTTGTGCACTTGGTTC</td>
<td>DGAT1A_R2-2</td>
<td>AATGAGGCAACACCACAA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F3</td>
<td>AGCTCTGGCCGATACTCTAT</td>
<td>DGAT1A_R3</td>
<td>CTGAAGATGCCAAGGCA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F4</td>
<td>ACGGTGTAGGAGCCACAAAG</td>
<td>DGAT1A_R4</td>
<td>GTTTTCGATTCACCCAT</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F5</td>
<td>TGGCAGAGCTCTTCTTCTGT</td>
<td>DGAT1A_R5</td>
<td>CAACGGCCTGTACATGA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F6</td>
<td>CATCATTGAGAGGCCAG</td>
<td>DGAT1A_R6</td>
<td>ATTTCCACCTGGCATGAA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F7</td>
<td>GGTGGCTGTTCGCCAACAT</td>
<td>DGAT1A_R7</td>
<td>GATGGCAAGACCCAGGA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F8</td>
<td>CTGTGGTTTCGCTGCTTCAT</td>
<td>DGAT1A_R8</td>
<td>CCAAGCGCCACCTTAAGT</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F9</td>
<td>CAACCGCAACAAAAACAC</td>
<td>DGAT1A_R9</td>
<td>TCCCTCAGCATTCATCTTC</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F11</td>
<td>GAAAACTGAGTTAGTAAACGTGC</td>
<td>DGAT1A_R10</td>
<td>ATTAATGCAAGACCGGAA</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F13</td>
<td>GAAGACTGAGTTAGTAAACGTGC</td>
<td>DGAT1A_R13</td>
<td>TAATGCAGAACGGGAAG</td>
<td></td>
</tr>
<tr>
<td>DGAT1A_F13</td>
<td>GAAGACTGAGTTAGTAAACGTGC</td>
<td>DGAT1A_R14</td>
<td>TAATGCAGAACGGGAAG</td>
<td></td>
</tr>
<tr>
<td><strong>DGAT1B gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F2</td>
<td>AGTGAAGGAAAGTCCGCTACT</td>
<td>DGAT1B_R2</td>
<td>CTCCCCCTCATTTTCTCTC</td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F3</td>
<td>TTTAAAGCCTGTTGGTGAAGTT</td>
<td>DGAT1B_R3</td>
<td>ACAAGGTCAATGCGGCA</td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F4</td>
<td>GGGGAAGCCTGGCTGCTGA</td>
<td>DGAT1B_R4</td>
<td>CTCCGTCGTCGTAAGAG</td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F5</td>
<td>AGTGAAGGAAAGTCCGCTACT</td>
<td>DGAT1B_R5</td>
<td>TCCATTGCAGAACACAGG</td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F6</td>
<td>AAGCATTGCTTCTGTTGCT</td>
<td>DGAT1B_R6</td>
<td>CGAAGGCACCAACTTGA</td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F7</td>
<td>TTTCTTTGCTTGGCTGC</td>
<td>DGAT1B_R7</td>
<td>AAAATGGACCGACCTG</td>
<td></td>
</tr>
<tr>
<td>DGAT1B_F8</td>
<td>TCCATTTCATTCTCCTAC</td>
<td>DGAT1B_R8-2</td>
<td>CAAGCGCACAACCCACA</td>
<td></td>
</tr>
</tbody>
</table>

204
Table 4.2. Primers used for amplification and sequencing of DGAT1A, DGAT1B, PDAT7, PDAT13, PDCT7 and PDCT8 genes. Primers are listed in 5’→3’ configuration (Continued)

<table>
<thead>
<tr>
<th>Primers used for both amplification and sequencing</th>
<th>forward sequence</th>
<th>reverse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDAT7 gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDAT7_&lt;sub&gt;F1&lt;/sub&gt;</td>
<td>CGGATCAGTTTCCGTCA</td>
<td>PDAT7_&lt;sub&gt;R1&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT7_&lt;sub&gt;F2&lt;/sub&gt;</td>
<td>TGGTGCTTTGGGTGA</td>
<td>PDAT7_&lt;sub&gt;R2&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT7_&lt;sub&gt;F3&lt;/sub&gt;</td>
<td>TAGCCAGAGAAAGCGA</td>
<td>PDAT7_&lt;sub&gt;R3&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT7_&lt;sub&gt;F4&lt;/sub&gt;</td>
<td>AGTTCCGAGCAAGGTT</td>
<td>PDAT7_&lt;sub&gt;R4&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT13 gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDAT13_&lt;sub&gt;F3&lt;/sub&gt;</td>
<td>GGGATGAAATTCCCTC</td>
<td>PDAT13_&lt;sub&gt;R3&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT13_&lt;sub&gt;F4&lt;/sub&gt;</td>
<td>GTTCTTCATTGGCCTC</td>
<td>PDAT13_&lt;sub&gt;R4&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT13_&lt;sub&gt;F5&lt;/sub&gt;-&lt;sub&gt;4&lt;/sub&gt;</td>
<td>CAGAGCTCTGGTCTCAT</td>
<td>PDAT13_&lt;sub&gt;R5&lt;/sub&gt;-&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDAT13_&lt;sub&gt;F5&lt;/sub&gt;-&lt;sub&gt;6&lt;/sub&gt;</td>
<td>TCTAAGCTCAAACGCACT</td>
<td>PDAT13_&lt;sub&gt;R5&lt;/sub&gt;-&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDCT7 gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDCT7_&lt;sub&gt;F1&lt;/sub&gt;</td>
<td>CGTCACCACTTTCCTTT</td>
<td>PDCT7_&lt;sub&gt;R1&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDCT7 gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDCT8 gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDCT8_&lt;sub&gt;F1&lt;/sub&gt;</td>
<td>TCATGGTAAAGTGGCTAC</td>
<td>PDCT8_&lt;sub&gt;R1&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDCT8 &lt;sub&gt;F4&lt;/sub&gt;</td>
<td>ATCCACTTCCTCTCCC</td>
<td>PDCT8_&lt;sub&gt;R4&lt;/sub&gt;</td>
</tr>
<tr>
<td>PDCT8 gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDGAT1A gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F1&lt;/sub&gt;</td>
<td>CACAACCATGTGGAGTTG</td>
<td>SDGAT1A_&lt;sub&gt;R1&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F2&lt;/sub&gt;</td>
<td>GCCATTGACCTCGTCCTC</td>
<td>SDGAT1A_&lt;sub&gt;R2&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F3&lt;/sub&gt;</td>
<td>GAAAGGTGGTGGCTTTC</td>
<td>SDGAT1A_&lt;sub&gt;R3&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F4&lt;/sub&gt;</td>
<td>TCCCTACATGTACTCAAGG</td>
<td>SDGAT1A_&lt;sub&gt;R4&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F5&lt;/sub&gt;</td>
<td>ATTCAGTTTGGGCTTTC</td>
<td>SDGAT1A_&lt;sub&gt;R5&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F6&lt;/sub&gt;</td>
<td>CAATTGGGCAATTCCGAT</td>
<td>SDGAT1A_&lt;sub&gt;R6&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F7&lt;/sub&gt;</td>
<td>TCCGCTCAAATTCGGCTT</td>
<td>SDGAT1A_&lt;sub&gt;R7&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F8&lt;/sub&gt;</td>
<td>TGTTGCTGTGAATAGCAG</td>
<td>SDGAT1A_&lt;sub&gt;R8&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F9&lt;/sub&gt;</td>
<td>TTGGGCACTTGGTGGAAG</td>
<td>SDGAT1A_&lt;sub&gt;R9&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F10&lt;/sub&gt;</td>
<td>GCCATGGGCAATTCCGAT</td>
<td>SDGAT1A_&lt;sub&gt;R10&lt;/sub&gt;</td>
</tr>
<tr>
<td>SDGAT1A_&lt;sub&gt;F11&lt;/sub&gt;</td>
<td>GAGAGCACTTACCAAATC</td>
<td>SDGAT1A_&lt;sub&gt;R11&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

205
**Table 4.2.** Primers used for amplification and sequencing of *DGAT1A, DGAT1B, PDAT7, PDAT13, PDCT7* and *PDCT8* genes. Primers are listed in 5’→3’ configuration (Continued)

<table>
<thead>
<tr>
<th>Primers used for sequencing only</th>
<th>forward</th>
<th>sequence</th>
<th>reverse</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGAT1A gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDGAT1A_ F13</td>
<td>ATCCTTCTGCCTTGTGCTGAA T</td>
<td>SDGAT1A_ R13</td>
<td>TTCTGTGTATGTACCTTT T</td>
<td>TCAGGA</td>
</tr>
<tr>
<td>SDGAT1A_ F14</td>
<td>TTTCTTCTGTCTTTGCCAT C</td>
<td>SDGAT1A_ R14</td>
<td>TAGCTAATCAACTACTCTA T</td>
<td>ACCATCTG</td>
</tr>
<tr>
<td>SDGAT1A_ F15</td>
<td>ATCCTTTCTGAGGGGGAA AC</td>
<td>SDGAT1A_ R15</td>
<td>AAAGACAAAGTAGCATGGG T</td>
<td>TCCT</td>
</tr>
<tr>
<td>SDGAT1A_ F16</td>
<td>CTTTTGAGTGCATAAACAT GAC</td>
<td>SDGAT1A_ R16</td>
<td>GTTTTGGCATTCACCAAT C</td>
<td></td>
</tr>
<tr>
<td>SDGAT1A_ F17</td>
<td>CTCGAGGGATATGAGTGA AGA</td>
<td>SDGAT1A_ R17</td>
<td>TTGTGAACAGGCTGAAT GT</td>
<td></td>
</tr>
<tr>
<td>SDGAT1A_ F18</td>
<td>TGGGACTTGTGAGTTGAGG TAG</td>
<td>SDGAT1A_ R18</td>
<td>AATACACTTGGAAAGTACTG CTACCTG</td>
<td></td>
</tr>
<tr>
<td>SDGAT1A_ F19</td>
<td>CCTGCTTGACTTTAGCAGTA GCTTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDGAT1A_ F20</td>
<td>TCCGCTTTATTGGCTTCCTT C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGAT1B gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F1</td>
<td>TTGGGTGGCATATGCTCTTCT GAGG</td>
<td>SDGAT1B_ R1</td>
<td>GCAGGATGTTTCTGGTTCA AAA</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F2</td>
<td>ATTTTGGCCATTTGACCTT GG</td>
<td>SDGAT1B_ R2</td>
<td>ACTGCACATGGTTGTCCTG</td>
<td>TA</td>
</tr>
<tr>
<td>SDGAT1B_ F3</td>
<td>GTCTTTGCCTTTTCTCTGCA T</td>
<td>SDGAT1B_ R3</td>
<td>TGGAAGGAAAGACAGAGA TGA</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F4</td>
<td>GGTGGATTCACACATTCT ACGGA</td>
<td>SDGAT1B_ R4</td>
<td>TGGGAACCTTAATTCTAGC AGTG</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F5</td>
<td>GACTATATGCTCTGACTT CTGG</td>
<td>SDGAT1B_ R5</td>
<td>GGGGCTATGGAATACCC ATCC</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F6</td>
<td>TTTCTGTGTGTTGGCTGCTT C</td>
<td>SDGAT1B_ R6</td>
<td>TCCCTTGAGGAGTAGTGCTG TG</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F7</td>
<td>TGTTTGATGCACTTGAGG GAGG</td>
<td>SDGAT1B_ R7</td>
<td>TGCAACTTGGATTTGGA AA</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F8</td>
<td>CAGGAACATCTTTGCAAT CA</td>
<td>SDGAT1B_ R8</td>
<td>AGCAGGCACCTGTACAAT GA</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F9</td>
<td>CCTGCTCTGCTCTTGGCAT GT</td>
<td>SDGAT1B_ R9</td>
<td>CAAAACAAAGCAAGAAGA AG</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F10</td>
<td>GATTTGTTCTTGCAGCAGAC TCGAC</td>
<td>SDGAT1B_ R10</td>
<td>TATGTTGAAAAGAAAC AG</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F11</td>
<td>TGAATCGGTTCAGAGTC AC</td>
<td>SDGAT1B_ R11</td>
<td>ATGCAACAGCGCTCAGCAG AT</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F12</td>
<td></td>
<td>SDGAT1B_ R12</td>
<td>GAAATGCGCATAGGAGAA GA</td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_ F13</td>
<td></td>
<td>SDGAT1B_ R13</td>
<td>CAAAGAATACCTAGCAGTA GCATACAC</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.2. Primers used for amplification and sequencing of DGAT1A, DGAT1B, PDAT7, PDAT13, PDCT7 and PDCT8 genes. Primers are listed in 5’→3’ configuration (Continued)

<table>
<thead>
<tr>
<th>Primers used for sequencing only</th>
<th>forward sequence</th>
<th>reverse sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGAT1B gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDGAT1B_R14</td>
<td>TCAAGAAGCAACCGTCAA</td>
<td>AA</td>
</tr>
<tr>
<td>SDGAT1B_R15</td>
<td>TGGCAACCTGAGATGGTA</td>
<td>AA</td>
</tr>
<tr>
<td>SDGAT1B_R16</td>
<td>TTCCTCCAGTGCAATCAA</td>
<td>A</td>
</tr>
<tr>
<td>SDGAT1B_R17</td>
<td>GGAGTTCTTGTCAGTTCT</td>
<td>TG</td>
</tr>
<tr>
<td><strong>PDAT7 gene</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDAT7_F1</td>
<td>TGTTGCTTTGGTGACCA</td>
<td>GA</td>
</tr>
<tr>
<td>SPDAT7_R1</td>
<td>TCTAGGCCCTTTGGGAACG</td>
<td>TA</td>
</tr>
<tr>
<td>SPDAT7_F2</td>
<td>AGGCCCTGTCTGGACTT</td>
<td>GA</td>
</tr>
<tr>
<td>SPDAT7_R2</td>
<td>AGCTCGGTGAGCTCAGAA</td>
<td>TG</td>
</tr>
<tr>
<td>SPDAT7_F3</td>
<td>CCGGGATATTGCTGGTC</td>
<td>GA</td>
</tr>
<tr>
<td>SPDAT7_R3</td>
<td>AGCCACATCTCTGCAAA</td>
<td>AT</td>
</tr>
<tr>
<td>SPDAT7_F4</td>
<td>TTCAAATTGCATCTGCT</td>
<td>AT</td>
</tr>
<tr>
<td>SPDAT7_R4</td>
<td>CAAAATTGCTTACTTTTG</td>
<td>TA</td>
</tr>
<tr>
<td>SPDAT7_F5</td>
<td>CGGATGCTTTCGGTCA</td>
<td>GT</td>
</tr>
<tr>
<td>SPDAT7_F6</td>
<td>GAACGGAAACTCTCGAGG</td>
<td>AA</td>
</tr>
<tr>
<td>SPDAT7_F7</td>
<td>GGGTGCTACCTTCCGAGA</td>
<td>A</td>
</tr>
<tr>
<td>SPDAT7_F8</td>
<td>TCATCTGAATAGTGTGG</td>
<td>TG</td>
</tr>
<tr>
<td>SPDAT7_F9</td>
<td>TAGGAGAAGACGACAG</td>
<td>CA</td>
</tr>
<tr>
<td>SPDAT7_F10</td>
<td>GTGCTATCAAGGTCGCA</td>
<td>GT</td>
</tr>
<tr>
<td>SPDAT7_F11</td>
<td>TGGAGGAAAATCCCTTTG</td>
<td>TA</td>
</tr>
<tr>
<td>SPDAT7_F12</td>
<td>AAGATTCAACCCATCCC</td>
<td>G</td>
</tr>
</tbody>
</table>
Table 4.2. Primers used for amplification and sequencing of DGAT1A, DGAT1B, PDAT7, PDAT13, PDCT7 and PDCT8 genes. Primers are listed in 5’→3’ configuration (Continued)

<table>
<thead>
<tr>
<th>Primers used for sequencing only</th>
<th>forward</th>
<th>reverse</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PDAT13 gene</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F1</td>
<td>GGGAATGAATTTCCCATCT</td>
<td>CCAGGCAGATGGACTTCAATA</td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F2</td>
<td>TTTTATTTACCTCTGACA</td>
<td>CAGAAGGCTTTTGATGAA CGA</td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F3</td>
<td>CAGATGGTGTTGCAAAACAT</td>
<td>TACCCCAAGCAAAAACCTTG A</td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F4</td>
<td>TCAAGTTTTGCTTGGTA</td>
<td>CGA</td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F5</td>
<td>CCACTCCTTTTTTAATGTTGG</td>
<td>TGAAACTCATAATGCCATT CG</td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F6</td>
<td>GTTCCTTCATTTGCCCTCAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F7</td>
<td>GGGTAGGAACCTTGCAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F8</td>
<td>TGGATTTGGAACTAATAGATTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDAT13_ F9</td>
<td>GAGCCTCTGGTGAAGACTTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PDCT7 gene</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT7_ F1</td>
<td>CTTTCGATCTGGGCTTCTAT</td>
<td>GGCCCCGTTGGCTATCTTTA</td>
<td></td>
</tr>
<tr>
<td>SPDCT7_ F2</td>
<td>GGAACCTGGCTTGGACCTTG</td>
<td>GCGGTAATTAAGAATTGAGACTCC</td>
<td></td>
</tr>
<tr>
<td>SPDC7_ F3</td>
<td>TGGAGATTAACATCTAGGAGCAAA</td>
<td>GGGTAGGGTACTATTTCC AAA</td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F4</td>
<td>CCGAATCTCAACAGCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F5</td>
<td>GCGGTAATTAAGAATTGAGACTCC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers used for sequencing only</th>
<th>forward</th>
<th>reverse</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PDCT8 gene</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F1</td>
<td>CCGAATCTCAACAGCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F2</td>
<td>GGAAATAGTCTACCTACCAGGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F3</td>
<td>TGGAAGTTAATCCTAGGAGCAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F4</td>
<td>GGGTAGGGTACTCTGCTCCAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPDCT8_ F5</td>
<td>CAGAGCCACATTTCCAGC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3. Primers used for amplification and sequencing of the transferase genes that were polymorphic between *Glycine max* cultivar Williams 82 and *Glycine soja* accession PI 366121

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGAT1B</td>
<td>5’-AGAGACTGGCCCCTTTTCAT-3’</td>
<td>5’-TCATCAAAGCAGGCTTAGA-3’</td>
</tr>
<tr>
<td>PDAT7</td>
<td>5’-CCCTTTGCGGAGTGTTACAT-3’</td>
<td>5’-AAGTCATGGGGGCTTATCC-3’</td>
</tr>
<tr>
<td>PDAT13</td>
<td>5’-CAGTGCGGAAGAGCAAGAA-3’</td>
<td>5’-TTGAGGGGAATGAAGGAAAC-3’</td>
</tr>
<tr>
<td>PDCT7</td>
<td>5’-GTCACACCACCTTTCCCTCTTCC-3’</td>
<td>5’-GTCGTAAGCAGACGTCGAG-3’</td>
</tr>
</tbody>
</table>
Table 4.4. Primers used in qRT-PCR experiments to determine the expression of *DGAT1A*, *DGAT1B*, *PDAT7* and *PDAT13* genes of *Glycine max* variety Williams 82 and *Glycine soja* accession PI 366121 relative to the expression of the reference gene *cons7*

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGAT1A</td>
<td>5’-CAACGCAACAAACAAAGA-3’</td>
<td>5’-TGAGCGGACTTTCTTTCACT-3’</td>
</tr>
<tr>
<td>DGAT1B</td>
<td>5’-CAGCCTCCGACGACAATG-3’</td>
<td>5’-GGAACGGAGGGAGGTAAG-3’</td>
</tr>
<tr>
<td>PDAT7</td>
<td>5’-AGCTGTCTGCAAGTTGAGT-3’</td>
<td>5’-AAGTTTGCTGGAGAGAAT-3’</td>
</tr>
<tr>
<td>PDAT13</td>
<td>5’-GCGAAACGATGAGGAAAG-3’</td>
<td>5’-CCACGTACTGAGGAAACGAC-3’</td>
</tr>
</tbody>
</table>
Table 4.5. Polymorphisms in coding and non coding sequence of DGAT1A, DGAT1B, PDAT7, PDAT13, PDCT7 and PDCT8 genes between Glycine soja line PI 366121 and reference Glycine max cultivar Williams 82. Changes at the protein level are in parenthesis after the corresponding polymorphism.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coding</th>
<th>Non coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGAT1A</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>PDAT7</td>
<td>T2917C, C4207T (A649V)</td>
<td>A620T, -1704T, T2298A</td>
</tr>
<tr>
<td>PDAT13</td>
<td>T2755C</td>
<td>none</td>
</tr>
<tr>
<td>PDCT7</td>
<td>A32G (H11R), G49A (A17T), A189G, G246A, T288C</td>
<td>T518C</td>
</tr>
<tr>
<td>PDCT8</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

*Positions are relative to start codon, where A = 1*
Bibliography


Li R, Yu K, Hildebrand D (2010) DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. Lipids 45:145-157


Zhang M, Fan J, Taylor DC, Ohlrogge JB (2009) DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21:3885-3901


Chapter 5

High Linolenic Acid in Seed Oil of *Glycine soja*:
Phenotypic Study of Important Seed Quality Traits and Their Relationship with Linolenic Acid
Introduction

Importance of soybean and soybean oil

Soybean \([Glycine \text{ max} \text{ (L.) Merr.}]\) is the leading oilseed crop produced and consumed in the world, representing 58\% and 43\% of world oilseed production and consumption, respectively (SoyStats 2011; Wilcox 2004). Soybean is the United States' second largest crop in cash sales and the number one value crop export (SoyStats 2011). Economic importance of soybean oil is given by its utilization in the manufacture of both edible and industrial products. Approximately 70\% of the edible fats and oils consumed in the US as salad and cooking oil, baking and frying fats and margarine and spreads derive from soybean. Also, \sim 14\% of the soybean oil utilized in the US is for industrial purposes such as the production of biodiesel, inks, paints, varnishes, resins and plastics (Cahoon 2003; SoyStats 2011).

Lipids represent one of the most important components in soybean seeds, comprising about 20\% of their weight (Gerde and White 2008). The predominant fatty acid (FA) composition of commodity soybean oil is about 11\% palmitic acid (16:0), 4\% stearic acid (18:0), 23\% oleic acid (18:1), 54\% linoleic acid (18:2) and 8\% linolenic acid (18:3) (Fehr 2007; Hill et al. 2008; Wilson 2004). For reference, the first number of these FA abbreviations indicates the number of carbon atoms, and the second number represents the number of double bonds in the molecule. 16:0 and 18:0 are saturated FAs,
18:1 is a monounsaturated FA, and 18:2 and 18:3 are polyunsaturated FAs (PUFA). Also, 18:2 and 18:3 are ω-6 and ω-3 types of FAs, respectively, which are essential to humans and all animals. The term essential refers to any compound that is required for life but cannot be synthesized by the organism and therefore must be incorporated with diet.

**Potential uses of soybean oil enriched in ω-3 fatty acids**

The balance between dietary ω-3 and ω-6 FAs strongly affects their function in the body. Currently, nutritionists are concerned over the reduction of the nutritionally desirable ω-3 levels in western diets and more particularly the ratio of ω-3:ω-6 FAs, which should ideally be 1:2.3 (Gerde and White 2008; Kris-Etherton et al. 2000). According to reports of the University of Maryland Medical Center (UMMC), the typical American diet tends to contain 14-25 times more ω-6 than ω-3 FAs ([http://www.umm.edu/altmed/articles/omega-3-000316.htm](http://www.umm.edu/altmed/articles/omega-3-000316.htm)).

ω-3 FAs play a crucial role in brain function as well as normal growth and development. They also reduce inflammation and may help lower risk of diseases such as heart disease, cancer, type II diabetes and arthritis (UMMC 2011). ω-3 FAs are highly concentrated in the brain and appear to be important for cognitive (brain memory and performance) and behavioral function (UMMC 2011). Because of these health benefits, elevated 18:3 soybean genotypes are desirable, especially in food grade soybeans (Kumar et al. 2006; Lee et al. 2007; Shibata et al. 2008).
Also, there has been a growing interest from the aquaculture industry in using plant-based proteins and oils as components of fish diets as a way to reduce both feeding costs and, especially, the pressure on marine resources. Soybean protein concentrate and oil, due to their excellent nutritional characteristics, are considered very suitable for aquafeed. Increasing global demand for fish high in long-chain ω-3 oils creates perfect market opportunities for soybeans with enhanced ω-3 content (Naylor et al. 2009). In addition, oils low in saturates and high in PUFAs would have applications as drying oils for paints, coatings and other industrial products (Wilson 2004).

**Soybean seed composition and environment**

It has been very well documented that the environment in which the plant grows influences the expression of soybean seed composition traits (Kumar et al. 2006; Wilson 2004). Temperature during seed development seems to be the most influential environmental factor affecting seed composition. Higher temperatures have been associated to higher oil and lower protein concentrations in soybean seeds (Piper and Boote 1999; Wilcox and Cavins 1992; Wolf et al. 1982). As with protein and oil, temperature also has an effect on fatty acid composition. While 18:2 and 18:3 accumulation is reduced under elevated temperatures, 18:1 levels are significantly increased. The saturated FAs 16:0 and 18:0 were mostly unaffected by changes in temperature (Byfield and Upchurch 2007b; Howell and Collins 1957; Rennie and Tanner 1989; Wilson 2004; Wolf et al. 1982). The increase of 18:1 at the expense of 18:2 and
18:3 in soybeans seeds that develop at elevated temperatures has been linked to the inhibitory effect of high temperatures on the desaturases’ gene expression as well as enzymatic activity (Byfield and Upchurch 2007a, b; Cheesbrough 1989; Heppard et al. 1996; Tang et al. 2005). Other factors such as rainfall during seed development, latitude, soil fertility or irrigation showed little or no effect on soybean seed FA composition (Howell and Collins 1957; Kumar et al. 2006; Lee et al. 2008).

**Associations among some seed quality traits in soybean**

The three major factors that determine soybean seed quality are protein and oil contents, protein and oil composition and seed physical appearance. Protein and oil contents as well as their composition are important because of their economical and nutritional value. Also, the relative composition of soybean oil and protein determines their physicochemical properties. Features that are important for seed appearance, such as seed coat and hilum color and seed size play a role in consumer acceptance as well as food uses. For example, many factors affect the quality of natto, a Japanese fermented soyfood. The preferred soybeans to produce natto are round, small seeded, with high contents of soluble sugars, yellowish seed coat and clear hilum. On the contrary, for soymilk and tofu preparation, large-seeded soybeans with clear hilum and seed coat and high protein contents are preferred (Liu 2008). During selection of soybeans for a particular use, it is important to know the relationships between the different quality attributes.
In a study comprising 10 soybean genotypes selected based on differences in seed size, Liu et al (1995) reported no significant correlations between seed mass and oil content. Seed size, however, was associated to the contents of unsaturated fatty acids: there was a positive correlation with 18:1 and negative correlations with both 18:2 and 18:3. Among the unsaturated FAs, 18:1 showed a significant and strong negative association with 18:3 content in seeds, which agrees with the direction of the association previously reported by Ross et al. (2000). The study by Ross et al. (2000) was carried out with three BC$_1$F$_{2:4}$ populations developed with an ultra low 18:3 genotype (~1% 18:3) as the common donor parent, and three recurrent parents with low 18:3 content (~2.5%). Each population consisted of 27 1% lines and 27 2% lines (Ross et al. 2000).

Reinprecht et al. (2006) also analyzed the relationship among seed traits in a population of 169 RILs segregating for 18:3. Seed mass and oil content showed no significant association in three of the six environments of evaluation whereas in the other three environments, the association between these two traits was negative. In addition, while seed mass showed a positive correlation with 18:1, the association with 18:3 was not significant. Ross et al. (2000) also reported no correlation between seed mass and 18:3 contents in soybean seeds. However, the lines with the highest 18:3 levels consistently recorded the heaviest seeds across three different populations. In contrast to the observations by Liu et al. (1995) and Ross et al. (2000), Reinprecht (2006) found no association between seed 18:1 and 18:3.

The associations of seed oil-18:3 contents and seed oil-18:1 contents ranged from not significant to significant with very low $r$ values depending on the environment.
considered (Reinprecht et al. 2006). In the study carried out by Ross et al. (2000), seed oil and 18:3 contents were directly associated, so that lines with the higher 18:3 contents had significantly higher oil content than lines with the lower levels of the PUFA.

**Associations among some seed quality traits in G.max x G. soja populations**

In many crop species, the utilization of their respective wild relatives has been used to broaden their genetic base (Stalker 1980). Previous studies suggest that *G. soja*, the wild relative of cultivated soybean, carries all the genes to accumulate the highest 18:3 levels observed in the two species while the cultivated soybean has no genes that contribute to high 18:3 levels (Pantalone et al. 1997b; Rebetzke et al. 1997; Shibata et al. 2008). The opposite could be said about seed oil and 18:1 contents, i.e., *G. soja* did not make any contribution towards the accumulation of oil and/or 18:1 beyond the levels typically observed in cultivated soybean (Rebetzke et al. 1997; Shibata et al. 2008). Pantalone et al. (1997b) analyzed the relationships between seed traits in three interspecific *G. max x G. soja* populations and reported that while seed mass presented a strong negative association with seed 18:3 content ($R^2= 89\%$), the association between seed mass and oil content was positive ($R^2= 50\%$). Also, the seed oil-18:3 content and seed oil-18:1 content associations were analyzed in three interspecific populations (Shibata et al. 2008). Seed oil and 18:3 contents presented a negative correlation in both $F_2$ and $F_9$ generations ($r= -0.45$) whereas seed oil and 18:1 levels showed weak or no correlation at all (Shibata et al. 2008).
To summarize, the associations observed between different pairs of seed traits are in general not consistent across studies (Liu et al. 1995, Ross et al. 2000, Reinprecht et al. 2006, Pantalone 1997b, Shibata et al. 2008). This could indicate that such relationships are highly dependent on the genotypes under consideration. For example, seed mass and 18:3 content were inversely associated in two of the studies considered (Pantalone et al. 1997b, Liu et al. 1995) while in the other two studies the association was not significant (Ross et al. 2000, Reinprecht et al. 2006). Also, seed oil and 18:3 contents showed a direct association in Ross et al. (2000), an inverse association in Shibata et al. (2008) and no association in Reinprecht et al. (2006). Yet, even when analyzing the same set of genotypes, seed traits that were correlated in one environment might show no association in another one, which probably reflects the importance of the interaction between these traits and the growing environment (Reinprecht et al. 2006).

Inheritance of high 18:3 content in seeds

Rahman et al. (1994) reported that the high 18:3 content in the soybean mutant line B739 was due to a single recessive allele. This conclusion was based on the observation of a bimodal pattern of segregation in an F2 population resulting from the cross between the high 18:3 mutant line B739 and the typical 18:3 content variety Bay. Even though no gene was identified as responsible for the high 18:3 level in seeds of B739, Rahman et al. (1994) suggested that the mutation might influence the activity of desaturases. QTL studies carried out with soybean as well as interspecific populations show several
genomic regions associated to variations in seed 18:3 concentration (Diers and Shoemaker 1992; Hyten et al. 2004; Panthee et al. 2006; Reinprecht et al. 2006). What is more, the distribution of the progeny for seed 18:3 content is typically normal and continuous, characteristic of a multigenic trait (Panthee et al. 2006; Reinprecht et al. 2006; Shibata et al. 2008). The reported narrow-sense heritability estimates for seed 18:3 in interspecific populations are low, ranging from 0.10 to 0.47 depending on the population considered (Rebetzke et al. 1997). The low heritability values are in complete agreement with the low response to selection for 18:1 and 18:3 reported by Burton et al. (1983).

**Linolenic acid content in the soybean wild ancestor**

Accessions of *Glycine soja* (Siebold & Zucc.), the wild ancestor of cultivated soybean, have been reported to contain as much as 23% 18:3, representing potentially useful genetic resources for identifying either new alleles or genes that govern 18:3 accumulation (Pantalone et al. 1997a; Wilson 2004). However, very little is known about the genetic regulation of high 18:3 in *G. soja*. Results of a survey on the average oil content and FA profile of *G. max* and *G. soja* accessions in the USDA’s National Plant Germplasm System (NPGS) (http://www.ars-grin.gov/npgs/) are shown in Table 5.1. In average, *G. soja* produces twice as much 18:3 but almost half the amount of 18:1 and oil as cultivated soybean. Also, *G. soja* presents higher variability for 18:3 than *G. max*, whereas *G. max* surpasses *G. soja* in variability for oil, 18:1 and 18:2. Analysis of
relative 18:1 and 18:2 desaturation patterns, which would reflect \( \omega-6 \) and \( \omega-3 \) desaturases’ activity, respectively, in four \( G. \ max \times G. \ soja \) populations suggested that \( G. \ soja \) carries superior alternative \( FAD2 \) and \( FAD3 \) alleles directing the expression of the high 18:3 trait (Pantalone et al. 1997a). Finally, increasing the expression of the \( FAD3 \) genes by genetic engineering, allowed the generation of soybean seeds that accumulated 18:3 in excess of 50% of the total oil (Cahoon 2003). However, structural and expression studies of the \( FAD2 \) and \( FAD3 \) desaturase genes showed that \( G. \ max \) and \( G. \ soja \) alleles make similar contributions to 18:3 content in seed oil (Lenis, Chapter 3 of this work).

The objective of this work was to better understand the genetics underlying the elevated 18:3 accumulation in seeds of \( G. \ soja \) in general and in the accession PI 366121 in particular. Phenotypic data from segregating populations grown in different environments at different generations of inbreeding were utilized to achieve the goal. Descriptive statistics parameters were used to both characterize seed 18:3 content and compare it with other seed traits with known quantitative inheritance. Also, the existing relationships between seed 18:3 and other seed traits were analyzed to determine if breeding for high 18:3 would affect other traits of economic importance.
Materials and Methods

Plant material and development of segregating populations

Study of the genetic basis of high 18:3 in seeds of soybean wild ancestor was carried out with *G. soja* accession PI 366121 and *G. max* genotypes Williams 82 and B1-52. PI 366121 is listed as a high 18:3 content entry (up to 23% 18:3) in the USDA’s National Plant Germplasm System ([http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1274060](http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1274060)). Williams 82 (Bernard and Cremeens 1988) is the reference soybean genotype in genomic studies and the chosen cultivar for the construction of the soybean whole-genome shotgun sequence assembly (Schmutz et al. 2010). Williams 82 presents a typical 18:3 content (~7-8%) for a soybean line. B1-52 is an ultralow 18:3 content (~1%) genotype (Bilyeu et al. 2006). Such low 18:3 levels in seed of B1-52 is the result of the combination of non functional mutant alleles for the *FAD3A*, *FAD3B* and *FAD3C* genes.

*Glycine soja* accession PI 366121 was crossed with *G. max* genotypes Williams 82 and B1-52 to generate two segregating populations, which will be referred to as the 7% lin and 1% lin populations, respectively. F2 seeds were produced in growth chamber and planted to generate F3 seeds in Columbia, Missouri in the summer of 2009. For the 7% lin population, F3 seeds were sent to Upala, Costa Rica and advanced to F5 by single seed descent method to generate recombinant inbred lines (RILs). Also, a sibling set of the
Costa Rica-produced F₄ population was planted in Columbia, Missouri in the summer of 2010. Seeds of the F₃ RILs obtained in both Upala, Costa Rica and Columbia, Missouri were used in further analyses.

**Determinations on seeds**

Before planting, seed dry weight and oil content were determined on ~110 individual seeds from each the 7% lin and the 1% lin populations in F₂ generational advance. Also, single seeds were evaluated for oil composition by cutting a seed chip (~1/3 of the seed) from each seed. The remainder of the seed was used for planting.

Progeny of each single plant were collected in the fall of 2009, and relative fatty acid content was determined on three individual F₃ seeds per plant.

Also, from each of the 7% lin RILs as well as the parental lines produced in both Upala, Costa Rica (CR) and Columbia, Missouri (CoMo), a composite sample of five F₅ seeds was used for seed dry weight and relative fatty acid content determination. Total seed oil content was also determined in F₅ seeds of the 7% lin RILs produced in CR.

In addition, from the 1% lin population, seed chips from 214 individual F₃ seeds derived from a single F₂ plant were analyzed for relative fatty acid content to gain further insight on the nature of the genetics governing 18:3 content in seeds (qualitative vs. quantitative).
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 as explained in Bilyeu et al. (2005). The individual fatty acid contents are reported as the relative percents of palmitic, stearic, oleic, linoleic, and linolenic acids in the extracted oil.

Seed oil and moisture content was determined through nuclear magnetic resonance (NMR) using a Bruker Minispec Mq10 NMR machine (Bruker Optics Inc., Billerica, MA).

**Statistical analyses**

All statistical analyses were carried out using the SAS 9.2 software (SAS Institute Inc., Cary, NC). Basic statistics parameters for each data set were obtained using the MEANS procedure. Goodness-of-Fit tests for normal, lognormal, weibull or gamma distributions were carried out using the UNIVARIATE procedure. Regression analyses were performed with the REG procedure. Narrow-sense heritability was estimated on an individual plant basis from the regression of the F2:3 progeny values on original F2 parental values using the REG procedure.
Results

Descriptive statistics

A population of 279 F4 RILs was generated by crossing the *G. max* cultivar Williams 82 with the *G. soja* accession PI 366121. F4 siblings were planted in Columbia, Missouri (CoMo) and Upala, Costa Rica (CR), where F5 seeds were produced. Wide segregation for many morphological features was observed. Segregation was also observed for some seed quality traits.

Seed dry weight of parental lines and RILs was determined on a composite sample of seeds. The RILs segregated following a continuous distribution that is skewed to the right (Fig. 5.1). The population’s mean dry weight for 100 seeds was 7.0 and 5.5 g in CR and CoMo, respectively while the mid-parent values were 10.85 and 7.65 g in the same locations (Fig. 5.1). The dry weight of 100 seeds in the parental lines Williams 82 and PI 366121 were 17.2 and 4.5 g in CR and 13.0 and 2.3 g in CoMo, respectively (Fig. 5.1). No individual recovered the Williams 82 seed dry weight in either of the test locations while 5 and 1% of the RILs presented seed dry weights similar to PI 366121 in CR and CoMo, respectively (Fig. 5.1). The population mean was 3.85 and 2.15 g off to the left of the mid-parent values in CR and CoMo, respectively (Fig. 5.1).

Relative fatty acid content was determined for each F5 RIL and the parental lines. The individuals segregated for seed 18:3 content following a normal distribution both in
CR and CoMo (Fig. 5.2). The population’s mean for 18:3 was 8.1% in CR and 10.7% in CoMo, and the mid-parent values were 8.4 and 10.6% in CR and CoMo, respectively (Fig. 5.2). The 18:3 contents of both parents were recovered by the progeny and just one individual with transgressive segregation was on the low 18:3 end in both locations (Fig. 5.2). Approximately 13% (37 individuals) and 4% (11 individuals) of the RILs presented 18:3 contents similar to PI 366121 in CR and CoMo, respectively (Fig. 5.2).

When the population was analyzed in the earlier F2 generation, seed 18:3 content had also shown a normal and continuous distribution, with a mean of 9.3%. The 18:3 content in Williams 82 and PI 366121 was 7.6 and 11.7%, respectively, and the mid-parent value was 9.7% (Fig. 5.3). The phenotype of the two parental lines was recovered by individuals of the progeny, and 36% of the F2s showed 18:3 levels like those of PI 366121 (Fig. 5.3).

When considering the ten RILs that accumulated the highest levels of 18:3 in CR and CoMo, there were four individuals in common between the two locations (Table 5.2). The maturity dates of these RILs ranged from September 21 to October 29 (Table 5.2). In addition, when considering the ten individuals with the lowest levels of 18:3 in CR and CoMo, there were 3 RILs in common between the two locations (Table 5.3). These RILs matured from August 30 to October 12 (Table 5.3).

Seed 18:1 content presented a normal distribution in CoMo, but in CR, it did not fit any of the distribution types tested (Fig. 5.4). The offspring recovered the 18:1 levels observed in the parental lines, and transgressive segregation was evident, especially in
CR. The mean seed 18:1 content in CR was 18.3% while in CoMo this FA was 16.9% of the oil (Fig. 5.4). The mid-parent values were 20.1 and 18.0% in CR and CoMo, respectively. Something to highlight regarding seed 18:1 content is the difference observed in the range of values; in CoMo, 18:1 ranged from 13.15 to 21.40% while in CR 18:1 transitioned from 11.84 to 48.63% (Fig. 5.4).

Total seed oil content was determined on the RILs grown in CR only. The progeny showed a normal distribution with a mean oil content of 15.2% while the mid-parent value was 17.3% (Fig. 5.5). Oil content of both parental lines was recovered by progeny individuals. However, there was only 1 RIL showing the Williams 82 oil levels (0.3% of the population) while 65 RILs (~23% of the population) presented oil contents comparable to that of PI 366121. No transgressive segregation was observed (Fig. 5.5).

**Relationships between seed quality traits**

The dependence/independence relationship between pairs of seed quality traits was analyzed to gain a better understanding of the direction and magnitude of the changes that may happen in a trait as the result of variations in another one. Regression analysis was used to achieve this purpose.

Seed dry weight (DW), was inversely associated to seed 18:3 content. For every unit increase in DW, there was a reduction in 18:3 accumulation that ranged between 0.36 to 0.44% in CR and CoMo, respectively (Tables 5.4 and 5.5). On the contrary, seed
DW and 18:1 content were directly associated (Tables 5.4 and 5.5). The gains in 18:1 content ranged from 0.5 to 1.0% for every gram increase in DW of 100 seeds. The higher increments were observed in CR (Tables 5.4 and 5.5). Also, seed DW and oil content presented a direct association, with an average increase of 0.22% in oil content per every gram increase in the mass of 100 seeds (Table 5.4).

There was an inverse association between seed 18:1 and 18:3, with reductions that ranged from 0.15 to 0.30% in 18:3 per % unit increase in 18:1. The magnitude of reduction in 18:3 caused by increases in 18:1 was bigger in CoMo than in CR (Tables 5.4 and 5.5).

Seed oil content showed an inverse relationship with 18:1 and 18:3 seed contents. The regression analysis indicates that 18:1 decreases 0.53% per % unit increase in oil content while 18:3 is reduced 0.09% only (Table 5.4).

**Prediction of performance**

The F5 progeny resulting from growing sibling F4 RILs in CR and CoMo was used to determine how well the performance in one of the test locations can predict the outcome in the other test location. The traits evaluated were seed 18:3 and 18:1 contents and seed DW (Table 5.6).

The regression analysis indicates that for every unit increase in 18:3 observed in CR, there is almost a 0.5% increase in 18:3 content in CoMo. However, only 22% of the
variation observed in 18:3 content in CoMo is explained by the changes in 18:3 observed in CR (Table 5.6).

Every percent unit 18:1 increased in CR was corresponded by a 0.14% increase in CoMo, with only 16% of the variation observed in CoMo being explained by the changes in CR (Table 5.6).

Finally, every unit change in seed DW in CR was associated to a 0.34 change in the same direction in CoMo. Also, variations in seed DW in CR do only explain 17% of the total variation observed in CoMo (Table 5.6).

**Heritability estimates**

Narrow-sense heritability (h²) was estimated for 18:3 and 18:1 on an individual plant basis by means of regression of the F₂:₃ progeny values on the F₂ parental values. h² was not significant for 18:3 at the probability level of 0.05, but it was at p<0.1. However, 18:3 in this population presented a low h² estimate and even lower R² value (Table 5.7). Oleic acid, on the other hand, presented a significant h² estimate, but both the h² and R² values were small (Table 5.7).
Discussion

The RILs resulting from the *G. max* x *G. soja* cross showed a wide segregation for seed DW, with a range of seed mass values similar to those observed in other interspecific populations (Pantalone et al. 1997b). Both in CR as well as in CoMo, ~60% of the individuals presented DW below the population mean, giving a skewed shape to the frequency distribution curve (Fig. 5.1). In addition, a shift of the population means from the mid-parent phenotype towards the PI 366121 values was observed, so that the population means were 36 and 28% lower than the mid-parent values in CR and CoMo, respectively (Fig. 5.1). In other words, a higher number of individuals were more similar to PI 366121 than to Williams 82. While a number of RILs presented the PI 366121 seed DW, no individual showed Williams 82 seed mass (Fig. 5.1). The fact that no individual was able to recover the Williams 82 DW indicates that PI 366121 did not make any contribution towards increased seed mass. The population’s mean seed DW was higher in CR than in CoMo (Fig. 5.1); however, the values are not significantly different. Parental lines Williams 82 and PI 366121 also presented heavier seeds in CR than in CoMo. Obtaining bigger soybean seeds in CR is usual; due to the proximity to the Equator, the photoperiod is short (~12 h year round), so that soybean plants are induced to flower with relatively few nodes, the structures of the plant that will later bear the pods. The result is soybean plants that will develop relatively few but big seeds.

The continuous and normal distribution presented by seed 18:3 content indicates that it segregates as a multigenic trait (Fig. 5.2) (Bernardo 2010). The same type of
distribution had been already observed in F$_2$ (Fig. 5.3) and F$_3$ (data not shown) generations of this population. Unlike seed DW, the population’s mean 18:3 content was lined up with the mid-parent estimate in both locations, and the RILs were spread evenly to both sides of the population mean (Fig. 5.2). The population’s mean seed 18:3 accumulation was higher in CoMo than in CR, resembling the pattern observed in the parental lines (Fig. 5.2). It is well known the negative association of temperature with the levels of polyunsaturated fatty acids in soybean seeds (Heppard et al. 1996; Rennie and Tanner 1989; Wilson 2004), which explains the higher 18:3 contents in the cooler location (CoMo). No individual in the population showed significantly higher 18:3 than PI 366121, indicating that all genes for elevated 18:3 were contributed by the *G. soja* accession and none by Williams 82 (Fig. 5.2). Previously, Pantalone et al. (1997a) and Shibata et al (2008) reported similar findings (normal distribution without transgressive segregation) when evaluating the relative 18:2 desaturation of interspecific *G. max* x *G. soja* populations.

The mean 18:1 content of the RILs was nearly similar in both CR (18.3%) and CoMo (16.9%) (Fig. 5.4) and agrees with the mean 18:1 levels observed in the F$_2$ generation (18.1%). However, the range between the minimum and maximum values as well as the distribution of phenotypes was very different in the two locations. While minimum and maximum values were separated by 8.3 percent units in CoMo (from 13.15 to 21.40%), the spread was of 36.8 percent units in CR (from 11.84 to 48.63%) (Fig. 5.4). The wider range in CR is mainly due to the high 18:1 contents of the individuals in the top 10%, which presented 18:1 levels superior to 25% and a mean of 30.2%. Nine of the
30 RILs with the highest 18:1 in CR are among the individuals with the top 18:1 contents in CoMo as well (data not shown). Higher temperatures during seed development promote 18:1 accumulation in soybean seeds (Rennie and Tanner 1989; Wilson 2004), a relationship that was observed in Williams 82 and PI 366121. Accumulation of 18:1 in Williams 82 and PI 366121 increased from 20.7 to 23.0% and from 15.3 to 17.1% in CoMo and CR, respectively (Fig. 5.4). The unproportional increase in 18:1 contents that was observed in the seeds of RILs produced in CR might indicate the existence of genotype x environment interactions. One possible explanation for these interactions could be that ω-6 and/or ω-3 desaturases have a genotype-dependent sensitivity to temperature. In other words, the regulation of these desaturase genes (transcriptional and/or post-transcriptional) might be influenced by growth temperature at a greater extent in some genotypes than in others. For example, the F₂ individuals that generated the RILs of this study were obtained in growth chamber. In the growth chamber, the environmental conditions were constant throughout the plant cycle and thus seed filling occurred with higher temperatures than those typically observed in a location like CoMo. On the contrary, such conditions are similar, at some extent, to those observed in CR. In these F₂ seeds, mean 18:1 content was similar to the means observed in the RILs (18.1%), but the standard deviation was 1.6 times higher than the one observed in CoMo and 1.8 times smaller than standard deviation observed in CR (data not shown). It seems likely that the unusually high levels of 18:1 accumulated by some individuals in CR (and some F₂ individuals as well) are the result of depressed unsaturation. As a reference, the ultra-low linolenic genotype B1-52, which carries nonfunctional alleles of the FAD3A, FAD3B and FAD3C genes, accumulated 37.8% 18:1 in CR vs. 28.6% in CoMo, and the increase
in 18:1 was solely at the expense of 18:2 (data not shown). Increasing evidence suggests that post-translational mechanisms are involved in the control of desaturase activity (Heppard et al. 1996; Horiguchi et al. 2000; Matsuda et al. 2005; Tang et al. 2005). It has been reported that ω-6 and ω-3 desaturases are highly unstable at high temperatures (Cheesbrough 1989; Tang et al. 2005). Also, Byfield and Upchurch (Byfield and Upchurch 2007a, b) reported that the reduced 18:3 and elevated 18:1 contents observed in soybeans seeds grown at elevated temperature were associated with a decreased accumulation of FAD2-1A, FAD2-1B, FAD3A, FAD3B and FAD3C transcripts.

Seed oil content of the RILs showed a continuous and normal distribution (Fig. 5.5), which is in accordance with the reported multigenic nature of the trait (Shibata et al. 2008; Wilson 2004). Similar results were observed in F2 as well (data not shown). Even though parental phenotypes were recovered, no transgressive segregation was observed, which would indicate that all the genes necessary for high oil were contributed by Williams 82 (Fig. 5.5). The parental lines presented typical oil contents, with 22.7% in Williams 82 and 11.9% in PI 366121 (Fig. 5.5). The population mean was 15.2%, which is lower than the reported 19.5% mean oil content for the USDA Soybean Germplasm Collection (Wilson 2004). Also, as it was previously noted for seed DW, the RILs’ mean oil content was shifted to the left towards the PI 366121 parent by 2.1% from the mid-parent value (Fig. 5.5). In contrast, population mean and mid-parent value for oil content were identical in F2 seeds (data not shown).

This G. max x G. soja cross generated a population segregating for many traits, some of which were measured during this study. It was evident that most of the RILs
resembled much more the PI 366121 parent than Williams 82 for almost all the traits evaluated. To illustrate, the proportion of the population that falls either to the left or right of the mid-parent value will be used. Seed DW was the most striking case since 98 and 94% of the RILs were to the left of the mid-parent point (more similar to PI 366121) in CR and CoMo, respectively (Fig. 5.1). The unbalance of the population was also evident for seed oil content (Fig. 5.5) and days to maturity (data not shown), for which 86 and 68% of the RILs were on the same side as PI 366121, respectively. For seed 18:1 content, 80% and 75% of the individuals showed 18:1 levels inferior to the mid-parent value in CR and CoMo, respectively and thus more similar to PI 366121 (Fig. 5.4).

Finally, seed 18:3 content was different from the other traits since the RILs showed a more even distribution to both sides of the mid-parent value. In CoMo, 54% of the RILs were on the PI 366121 side (Fig. 5.2) and χ² test indicated no significant difference with a 1:1 distribution (data not shown). In CR, however, only 40% of the individuals presented 18:3 contents above the mid-parent estimate (Fig. 5.2), and the distribution of the RILs was statistically different from a 1:1 ratio. This was the only case in which the minority of the RILs was on the PI 366121 side, and the reason for it could be explained by the depressed unsaturation usually observed in CR that would prevent the expression of a PI 366121-like phenotype in the RILs. The reason for the majority of the RILs to resemble the G. soja parent is not known, but it could be that PI 366121 is contributing more “genetic factors” to the offspring than the G. max parent. These “genetic factors” could be extra copies of genes and/or more or stronger regulators of gene expression that were lost during domestication. Furthermore, it is not known if this unequal distribution of phenotypes between the parental values is a common phenomenon in G. max x G. soja.
interspecific populations or it is specific for this population. Even though Shibata et al. (2008) did not make any reference to this type of observations in their report, it would appear that the seed oil content of the RILs was biased towards the *G. soja* parent whereas seed 18:3 and 18:1 contents were evenly distributed between the parental phenotypes.

Soybean gets its value as a crop from the two main components of the seed: protein and oil. Soybean breeders, in order to add value to the crop, are permanently trying to increase not only seed yield but also the valuable seed components. Breeding for a simultaneous increase of seed yield, oil and protein has been hurdled by the negative correlations existing between seed oil and protein and between seed yield and protein. On the other hand, there is a positive association between seed yield and oil (Hurburgh et al. 1990; Wilcox and Guodong 1997). It was among the objectives of this work a further understanding of the relationships between 18:3 and 18:1 contents with seed oil and seed size in *G. max* x *G. soja* interspecific crosses. Such associations might have practical ramifications to soybean breeding.

The negative association between seed DW and 18:3 content indicates that breeding for big seeds, such as those wanted for tofu, and high 18:3 levels would present difficulties. On the other hand, breeding for small seeds, a requirement for natto, and elevated content of the ω-3 FA would be easier to achieve. However, despite the significant and negative association between seed size and 18:3 content, less than a quarter of the total variation for 18:3 (23% in CR and 21% in CoMo) was explained by changes in seed DW (Tables 5.4 and 5.5). Similar findings were observed by Liu et al.
(1995) and Pantalone et al. (1997b), but in the populations they studied, increases in seed mass explained between 36 to 89% of the reduction in 18:3 content. Reinprecht et al (2006) and Ross et al. (2000) found no significant correlation between seed size and seed 18:3 in *G. max* lines. However, the mean seed mass of the soybean lines with 2% 18:3 content was significantly and consistently higher than the seed mass of lines with 1% 18:3 across the three evaluated populations (Ross et al. 2000).

Contrary to what was observed for 18:3, seed mass and 18:1 presented a significant and direct association (Tables 5.4 and 5.5), a fact that had been previously observed by Liu et al. (1995) when analyzing seeds of ten soybean genotypes. However, while only ~10% of the increase in 18:1 is explained by bigger seeds in the *G. max* x *G. soja* RILs, a much higher proportion of the variability (between 65 and 81%) was reported in the study carried out with soybean cultivars (Liu et al. 1995). In another study, where the associations between several seed traits were determined in *G. max* RILs, seed mass and 18:1 presented association levels similar to what was found in this study (between 5 and 15%, depending on location) (Reinprecht et al. 2006). It would seem that seed dry weight and 18:1 content are more tightly linked in soybean than in its wild ancestor.

As expected, seed 18:3 and 18:1 contents were inversely related (Tables 5.4 and 5.5). It is very well known that environmental conditions that promote the buildup of 18:1 depress the accumulation of PUFA, including 18:3. This is especially true when considering the effect of temperature during seed development on fatty acid composition of the oil (Byfield and Upchurch 2007a, b; Wilson 2004). Even though the negative association was significant in the two locations of evaluation, the negative relationship
was more patent in CR ($R^2=45\%$) than in CoMo ($R^2=21\%$). These data would suggest that the depressing effects elevated temperatures have on desaturases’ transcription and/or activity and consequently PUFA accumulation are more important than the promoting effects lower temperatures have on FA desaturation. The negative association between 18:1 and 18:3 contents in seeds has also been observed in cultivated soybeans (Liu et al. 1995), in which increases in 18:1 explain reductions in 18:3 much better than in the interspecific population ($R^2$ values ranged between 66 and 82%). These results suggest that breeding for simultaneous high levels of 18:1 and 18:3 in soybean might be easier if *G. soja* is one of the parental lines. The relative FA profile of an interspecific population of NILs developed by crossing a *G. max* cultivar with the PI 366121 supports the relatively higher independence between 18:1 and 18:3 when using the *G. soja* accession as a parent. Approximately 13% of the individuals presented 18:1 levels ≥20% whereas the 18:3 contents ranged between 12.1 and 15.5%, which is unusually high in commodity soybeans (data not shown).

Even though seed oil showed a negative association with both 18:1 and 18:3 contents, the proportion of the changes in these two FAs that are actually explained by variations in oil are so minute that such relationship should not have practical implications when breeding for high oil and 18:1 or high oil and 18:3 (Table 5.4). Shibata et al. (2008) analyzed such relationships in $F_2$ as well as in RILs in $F_9$ from interspecific crosses and found a negative correlation between seed oil and 18:3 contents ($r=-0.45$) and no correlation between seed oil and 18:1 contents. In an analysis of the relationships among different traits in *G. max* $F_5$ RILs, seed oil showed no significant
association with either 18:3 or 18:1 in some locations or a positive, though weak, association in other locations (Reinprecht et al. 2006). The absolute values of the correlation coefficients reported by Reinprecht et al. (2006) are identical to the ones obtained in this research. Ross et al. (2000) evaluated the effect of having 1% or 2% 18:3 content in seeds on other seed traits in three G. max backcross populations. The populations with 2% 18:3 consistently presented higher seed oil concentration than the 1% populations, with correlations ranging from 0.35 to 0.60.

The quantitative nature of seed 18:3 and 18:1 contents and seed DW is not only reflected in the continuous distribution of phenotypes but also in the great proportion of genetically unexplained phenotypic variability that is observed when the RILs are grown in two distinct locations (Table 5.6). For the three seed traits under consideration, there were significant associations between the values obtained in one and the other location. The significance of these associations results from the common genetic background of the RILs grown in the two locations. However, the low R² coefficients, which ranged from 16 to 22%, indicate that the environment plays a big role in the final value of these traits. As described by Bernardo (2010), quantitative traits are often influenced to a greater extent by the environment than by the underlying genes themselves.

The low narrow-sense heritability estimates as well as low R² values from those regression analyses also support the notion of quantitative inheritance for both 18:3 and 18:1 (Table 5.7). Burton et al. (1983) reported that 18:3 and 18:1 show poor response to selection, which is in complete agreement with the also low heritability estimates for these FAs observed by Rebetzke et al. (1997).
Despite the low $h^2$ estimates and the significant influence of the environment on the final expression of 18:3 content, obtaining individuals with consistent relative phenotypes across locations seems possible. For example, when considering seed 18:3 levels in CR and CoMo, no individual achieved the same 18:3 content in the two locations. However, analysis of the top ten 18:3 accumulators in the two environments shows four RILs in common (Table 5.2). Comparable results were observed when studying the bottom ten 18:3 accumulators (Table 5.3). Also, the same proportion of common phenotypes was observed when seed 18:1 and DW were considered (data not shown). The size of the population is important in the identification of genotypes with consistent relative phenotypes. If only the top five 18:3 accumulators from CR and CoMo had been chosen for the comparison (1.8% of the population), no common individual would have been found. Instead, increasing the sample size to the top ten individuals allowed the identification of RILs with consistent relative phenotype (Table 5.2).

Finally, maturity date was determined for all the RILs grown in CoMo. A wide range of days to maturity was observed with the earliest genotypes being ready for harvest on August 30 and the latest on October 29. These dates represent 118 and 178 days from planting, respectively. All the top ten 18:3 accumulators matured after the mid-point maturity range (September 29, 148 days), with the exception of two individuals which matured 1 and 8 days earlier (Table 5.2). Also, all the bottom ten 18:3 accumulators matured before the mid-point maturity range, with the exception of one RIL that matured 13 days later (Table 5.3). This reflects the existing link between maturity group and FA composition that has been previously noticed (Reinprecht et al. 2006).
later maturing genotypes seeds develop with cooler temperatures favoring unsaturation and thus accumulation of 18:3.
Conclusions

The data presented in this work very clearly show the quantitative nature of seed 18:3 content in *G. soja* and its cultivated relative *G. max*. In spite of being a multigenic trait and as such highly influenced by the growing environment, the use of *G. soja* to increase seed 18:3 content in cultivated soybean seems possible. In CoMo, ~4% of the RILs presented 18:3 levels comparable to that of PI 366121, and a third of those top scorers were among the ten highest 18:3 accumulators in CR. In CR as well as among the F$_2$s obtained in growth chamber there was a higher proportion of individuals recovering the PI 366121 18:3 contents: 13 and 36%, respectively. A combination of lower than the average 18:3 contents and unusually higher standard deviations in PI 366121 resulted in more individuals with 18:3 levels comparable to those of the *G. soja* parent than in CoMo. This fact highlights the importance of the environment where phenotype-based selection is to be made. Selection of individuals based on phenotype for high 18:3 content should be carried out in the target environments. Otherwise, carrying out selection in environments like CR will lead to choosing a number of individuals that later, when evaluated in locations where the potential of the high 18:3 check is achieved, will fail to accumulate the expected levels of 18:3.

Given the low h$^2$ estimates for 18:3 and the diversity of the cross when using *G. soja* as one of the parents, which means many polymorphic genomic regions segregating,
selecting for high 18:3 would be more effective in late generations, when genes are for the most part fixed.

Breeding programs usually have multiple goals. When breeding soybean for high 18:3 content using *G. soja* as the donor, it will be important to consider not only the relationships among the different seed traits in the resulting population but also the effects that using wild soybean as a parent may have on these traits individually. To illustrate, seed 18:3 has shown to be associated to reductions in seed DW, seed 18:1 and oil contents in the Williams 82 x PI 366121 population. However, the observed $R^2$ values between 18:3 and any of those traits were never superior to 25% with the exception of 18:1 in CR, which suggests it would be possible to simultaneously breed, for example, for high 18:3 contents and bigger seeds. Nevertheless, getting big seeds relative not to the population but to appropriate controls, and high 18:3 levels, will be only possible if the *G. max* parent used in the cross can counter the strong bias *G. soja* confers to the resulting population. In other words, the *G. max* parental line should contribute to shift the population mean for seed dry weight towards the bigger seeds end.
Figures
Figure 5.1. Frequency distribution of RILs for seed dry weight in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) in an F$_5$ population resulting from the cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*). The black arrow indicates the PI 366121 phenotype, the white arrow indicates the Williams 82 phenotype and the patterned arrow indicates the mid-parent (MP) value. Line under a parental arrow indicates ± standard deviation. A bell-shaped line was drawn to indicate the frequency of phenotypes if the distribution were normal.
Figure 5.2. Frequency distribution of RILs for relative seed linolenic acid content (18:3) in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) in an F₅ population resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja). The black arrow indicates the PI 366121 phenotype, the white arrow indicates the Williams 82 phenotype and the patterned arrow indicates the mid-parent (MP) value. Line under a parental arrow indicates ± standard deviation. A bell-shaped line was drawn to indicate the frequency of phenotypes if the distribution were normal.
Figure 5.3. Frequency distribution for relative seed linolenic acid content (18:3) in a population of F$_2$ individuals resulting from the cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*). The black arrow indicates the PI 366121 phenotype, the white arrow indicates the Williams 82 phenotype and the patterned arrow indicates the mid-parent (MP) value. Line under a parental arrow indicates ± standard deviation. A bell-shaped line was drawn to indicate the frequency of phenotypes if the distribution were normal.
Figure 5.4. Frequency distribution of RILs for relative seed oleic acid content (18:1) in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo) in an F$_3$ population resulting from the cross Williams 82 (Glycine max) x PI 366121 (Glycine soja). The black arrow indicates the PI 366121 phenotype, the white arrow indicates the Williams 82 phenotype and the patterned arrow indicates the mid-parent (MP) value. Line under a parental arrow indicates ± standard deviation. A bell-shaped line was drawn to indicate the frequency of phenotypes if the distribution were normal.
Figure 5.5. Frequency distribution of RILs for seed oil content in Upala, Costa Rica (CR) in an F$_5$ population resulting from the cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*). The black arrow indicates the PI 366121 phenotype, the white arrow indicates the Williams 82 phenotype and the patterned arrow indicates the mid-parent (MP) value. Line under a parental arrow indicates ± standard deviation. A bell-shaped line was drawn to indicate the frequency of phenotypes if the distribution were normal.
Tables
Table 5.1. Average oil (% of seed dry mass) and individual fatty acid contents (% of total oil) of *Glycine max* and *Glycine soja* accessions in the USDA’s National Plant Germplasm System (NPGS) ± one standard deviation from the mean. Information on wild soybean accession PI 366121 and *G. max* genotypes Williams 82 and B1-52 is also included.

<table>
<thead>
<tr>
<th></th>
<th>Oil (%)</th>
<th>16:0 (%)</th>
<th>18:0 (%)</th>
<th>18:1 (%)</th>
<th>18:2 (%)</th>
<th>18:3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G. max</strong></td>
<td>20 ± 2.1</td>
<td>11 ± 1.0</td>
<td>4 ± 0.6</td>
<td>23 ± 3.8</td>
<td>54 ± 3.7</td>
<td>8 ± 1.5</td>
</tr>
<tr>
<td><strong>G. soja</strong></td>
<td>11 ± 1.4</td>
<td>12 ± 0.9</td>
<td>4 ± 0.5</td>
<td>14 ± 2.4</td>
<td>54 ± 1.7</td>
<td>16 ± 2.2</td>
</tr>
<tr>
<td>PI 366121</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>11</td>
<td>54</td>
<td>20</td>
</tr>
<tr>
<td>(G. soja)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Williams 82</strong></td>
<td>20</td>
<td>11</td>
<td>3</td>
<td>23</td>
<td>56</td>
<td>7</td>
</tr>
<tr>
<td>(G. max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B1-52</strong></td>
<td>25</td>
<td>11</td>
<td>4</td>
<td>23</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>(G. max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* All observations for *G. max* accessions in the NPGS were used in the calculations (n=16,588)

*b* All observations for *G. soja* accessions in the NPGS were used in the calculations (n=1,242)

*c* Data generated in the lab. B1-52 carries mutant non-functional alleles of *FAD3A, FAD3B* and *FAD3C* genes (Bilyeu et al. 2006)
Table 5.2. Fatty acid composition of the ten individuals that accumulated the highest 18:3 contents in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo). Maturity date (MD) for the CoMo location has been included.

<table>
<thead>
<tr>
<th>CR</th>
<th>Fatty acids (%)</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant ID</td>
<td>16:0</td>
<td>18:0</td>
</tr>
<tr>
<td>409</td>
<td>12.96</td>
<td>4.62</td>
</tr>
<tr>
<td>352</td>
<td>12.02</td>
<td>3.65</td>
</tr>
<tr>
<td>274&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12.92</td>
<td>3.69</td>
</tr>
<tr>
<td>314&lt;sup&gt;b&lt;/sup&gt;</td>
<td>12.20</td>
<td>3.86</td>
</tr>
<tr>
<td>427</td>
<td>11.99</td>
<td>4.36</td>
</tr>
<tr>
<td>176</td>
<td>12.57</td>
<td>3.74</td>
</tr>
<tr>
<td>175&lt;sup&gt;c&lt;/sup&gt;</td>
<td>11.56</td>
<td>3.78</td>
</tr>
<tr>
<td>569</td>
<td>12.31</td>
<td>3.64</td>
</tr>
<tr>
<td>416</td>
<td>12.57</td>
<td>3.92</td>
</tr>
<tr>
<td>539&lt;sup&gt;d&lt;/sup&gt;</td>
<td>11.91</td>
<td>3.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CoMo</th>
<th>Fatty acids (%)</th>
<th>MD&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant ID</td>
<td>16:0</td>
<td>18:0</td>
</tr>
<tr>
<td>424</td>
<td>10.97</td>
<td>3.54</td>
</tr>
<tr>
<td>539&lt;sup&gt;d&lt;/sup&gt;</td>
<td>11.81</td>
<td>3.38</td>
</tr>
<tr>
<td>466</td>
<td>10.42</td>
<td>3.13</td>
</tr>
<tr>
<td>305</td>
<td>11.22</td>
<td>3.32</td>
</tr>
<tr>
<td>438</td>
<td>11.42</td>
<td>3.38</td>
</tr>
<tr>
<td>175&lt;sup&gt;c&lt;/sup&gt;</td>
<td>10.59</td>
<td>2.65</td>
</tr>
<tr>
<td>314&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.69</td>
<td>3.48</td>
</tr>
<tr>
<td>274&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11.01</td>
<td>3.08</td>
</tr>
<tr>
<td>481</td>
<td>10.68</td>
<td>3.75</td>
</tr>
<tr>
<td>206</td>
<td>11.14</td>
<td>3.60</td>
</tr>
</tbody>
</table>

<sup>a,b,c,d</sup> Individuals that are among the top ten 18:3 accumulators in both CR and CoMo.
<sup>e</sup>Maturity dates for the RIL population ranged from 30 August to 29 October.
Table 5.3. Fatty acid composition of the ten individuals that accumulated the lowest 18:3 contents in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo). Maturity date (MD) for the CoMo location has been included.

<table>
<thead>
<tr>
<th>CR Plant ID</th>
<th>Fatty acids (%)</th>
<th>CoMo Plant ID</th>
<th>Fatty acids (%)</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16:0</td>
<td>18:0</td>
<td>18:1</td>
<td>18:2</td>
</tr>
<tr>
<td>166a</td>
<td>13.23</td>
<td>4.35</td>
<td>19.77</td>
<td>58.36</td>
</tr>
<tr>
<td>213</td>
<td>10.36</td>
<td>2.75</td>
<td>48.63</td>
<td>33.53</td>
</tr>
<tr>
<td>560b</td>
<td>11.51</td>
<td>3.97</td>
<td>13.34</td>
<td>66.44</td>
</tr>
<tr>
<td>48</td>
<td>12.00</td>
<td>4.22</td>
<td>36.08</td>
<td>42.48</td>
</tr>
<tr>
<td>445c</td>
<td>11.13</td>
<td>3.24</td>
<td>37.62</td>
<td>42.66</td>
</tr>
<tr>
<td>194</td>
<td>12.18</td>
<td>3.24</td>
<td>22.67</td>
<td>56.26</td>
</tr>
<tr>
<td>434</td>
<td>11.81</td>
<td>3.26</td>
<td>29.34</td>
<td>49.94</td>
</tr>
<tr>
<td>474</td>
<td>11.16</td>
<td>3.71</td>
<td>36.56</td>
<td>42.75</td>
</tr>
<tr>
<td>503</td>
<td>12.07</td>
<td>3.42</td>
<td>31.40</td>
<td>47.22</td>
</tr>
</tbody>
</table>

a, b, c Individuals that are among the bottom ten 18:3 accumulators in CR and CoMo.

d Maturity dates for the RIL population ranged from 30 August to 29 October.
Table 5.4. Regression coefficients (b) and $R^2$ values (%) for seed quality traits of RILs in an F$_5$ population resulting from the cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*) grown in Upala, Costa Rica. Traits in column X were used as the independent variable and traits in row Y as the dependent variable.

<table>
<thead>
<tr>
<th></th>
<th>Oil</th>
<th>18:1</th>
<th>18:3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>$R^2$</td>
<td>$b$</td>
</tr>
<tr>
<td>18:1</td>
<td>0.22$^b$</td>
<td>3</td>
<td>-0.53$^a$</td>
</tr>
<tr>
<td>Oil</td>
<td></td>
<td></td>
<td>-0.15$^c$</td>
</tr>
<tr>
<td>DW</td>
<td></td>
<td></td>
<td>1.05$^c$</td>
</tr>
</tbody>
</table>

$^a$Regression significant at p<0.05
$^b$Regression significant at p<0.01
$^c$Regression significant at p<0.001
**Table 5.5.** Regression coefficients (b) and $R^2$ values (%) for seed quality traits of RILs in an F$_5$ population resulting from the cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*) grown in Columbia, Missouri. Traits in column X were used as the independent variable and traits in row Y as the dependent variable.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>18:1</th>
<th></th>
<th>18:3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>18:1</td>
<td>b</td>
<td>$R^2$</td>
<td>-0.30$^c$</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>DW</td>
<td>0.48$^c$</td>
<td>11</td>
<td>-0.44$^c$</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Regression significant at $p<0.05$
$^b$Regression significant at $p<0.01$
$^c$Regression significant at $p<0.001$
Table 5.6. Regression coefficients (b) and $R^2$ values (%) for seed quality traits of sibling F_{4} RILs' progeny grown in Upala, Costa Rica (CR) and Columbia, Missouri (CoMo). Traits in column X were used as the independent variable and traits in row Y as the dependent variable.

<table>
<thead>
<tr>
<th></th>
<th>18:3 CoMo</th>
<th>18:1 CoMo</th>
<th>DW CoMo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>$R^2$</td>
<td>b</td>
</tr>
<tr>
<td>18:3 CR</td>
<td>0.48(^a)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>18:1 CR</td>
<td>0.14</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>DW CR</td>
<td></td>
<td></td>
<td>0.34</td>
</tr>
</tbody>
</table>

\(^a\)All regressions were significant at p<0.001
Table 5.7. Parent-offspring (F$_{2,2,3}$) narrow-sense heritability (h$^2$) and R$^2$ values (%) for seed oil 18:3 and 18:1 contents in the interspecific cross Williams 82 (*Glycine max*) x PI 366121 (*Glycine soja*) population

<table>
<thead>
<tr>
<th>Trait</th>
<th>$h^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:3</td>
<td>0.13$^a$</td>
<td>1</td>
</tr>
<tr>
<td>18:1</td>
<td>0.15$^a$</td>
<td>9</td>
</tr>
</tbody>
</table>

$^a$Regression not significant at p<0.05

$^a$Regression significant at p<0.001
Bibliography


Lee JD, Oliva ML, Sleper DA, Shannon JG (2008) Irrigation has little effect on unsaturated fatty acid content in soya bean seed oil within genotypes differing in fatty acid profile. J Agron Crop Sci 194:320-324


264


Chapter 6

Conclusions and Future Directions
Soybean seed protein and oil are broadly incorporated into feed, food and industrial products, which makes soybean a highly valuable crop. Increased demand for vegetable oils and growing awareness of health issues around dietary fats necessitate a better understanding of the genetics underlying oil content and composition in plants. In addition, soybean-containing food products have the opportunity to be improved by reducing or removing the unpleasant flavors and aroma that result from the oxidation of PUFAs by the action of seed lipoxygenases.

The genetic basis of mutations in *Lox1* (Glyma13g42320) and *Lox3* (Glyma15g03030) soybean genes were investigated. Two independent mutations, a 74 bp deletion followed by a stop codon and a C2880A nonsense mutation, were responsible for the premature truncation of the Lox1 protein in single mutant lines PI 408251 and PI 133226, respectively. In contrast, *lox3* mutants PI 205085 and PI 417458 showed a single base deletion introducing a frame shift at position 101 which resulted in a premature stop codon (Lenis et al. 2010).

Co-dominant molecular marker assays perfectly associated with the causative mutations were developed for *Lox1, Lox2* (Glyma13g42310) and *Lox3* genes, providing a quick, inexpensive and accurate means to assess the genetic makeup of individuals. The utilization of these molecular marker assays will allow soybean breeders to directly select for mutant alleles in early generations of segregating populations enabling a more
efficient development of seed-lipoxygenase-free soybean varieties with improved flavor and aroma (Lenis et al. 2010).

In addition, it was elucidated that genetic recombination was the mechanism that broke the tight repulsion-phase linkage between $Lox1$ and $Lox2$ loci, allowing the combination of three independent $lox$ mutant alleles in the lipoxygenase-free variety Jinpumkong 2 (Lenis et al. 2010).

Enhancement of $\omega$-3 FA content of the oil for both food and feed applications is highly desired (Brenna 2011; Lee et al. 2007; Naylor et al. 2009). Accessions of $G. soja$, the wild ancestor of soybean, contain, in average, twice as much $\omega$-3 FAs (18:3) as cultivated soybean, representing a potential source of superior genes/alleles. However, little is known about the genetics of the elevated 18:3 trait in wild soybean.

Significant progress has been made in understanding the role of $FAD2-1$ and $FAD3$ genes in the high 18:3 phenotype of PI 366121. The data from this research work suggest that the $G. soja$ accession PI 366121 and $G. max$ cultivar Williams 82 carry equivalent functional alleles for $FAD2-IA$ (Glyma10g42470), $FAD2-1B$ (Glyma20g24530), $FAD3A$ (Glyma14g37350), $FAD3B$ (Glyma02g39230), $FAD3C$ (Glyma18g06950) and $FAD3D$ (Glyma11g27190) genes. Moreover, elevated 18:3 in PI 366121 seeds is not the result of higher expression of either $FAD3$ or $FAD2-1$ genes. Higher 18:3 content in seeds of $G.$
soja, therefore, is not the result of superior ω-6 and/or ω-3 desaturases as had been previously suggested in the literature.

Since the desaturase genes did not explain the differences in 18:3 content between G. soja and G. max, the focus was changed towards genes that encode enzymes catalyzing other steps in the TAG synthesis pathway and that have the potential to modify the composition of oil. Enzymes participating directly or indirectly in the incorporation of 18:3 in TAG, where ~90% of 18:3 is, were considered good candidates.

Research on the model plant Arabidopsis shows that oil content and composition in seeds vary dramatically depending on whether DGAT1 or PDAT1 contributes more to the final acylation step to form TAG (Katavic et al. 1995; Routaboul et al. 1999; Zhang et al. 2009; Zou et al. 1999). Changes in oil composition are explained, at least in part, by differences in substrate specificity between the two enzymes. While DGAT1 shows higher activity with saturated and long-chain FAs, PDAT1 shows a preference for acyl groups containing several double bonds and unusual FAs (Katavic et al. 1995; Stahl et al. 2004; Zhang et al. 2009). In addition, Lu et al. (2009) showed evidence of a new enzyme, PDCT, which interconverts DAG and PC during TAG synthesis in developing seeds of Arabidopsis. PDCT seems to catalyze an important step for PUFA enrichment of TAG since the null allele was associated with a reduction of ~40% in the seed content of PUFA. The present study shows that the G. soja alleles of DGAT1B (Glyma17g06120), PDAT7 (Glyma7g04080), PDAT13 (Glyma13g16790) and PDCT7 (Glyma7g03350)
genes are not associated with higher 18:3 content in seeds. In addition, a preliminary QTL analysis seems to indicate that no region of chromosome 13, where DGAT1A (Glyma13g16560) is located, is associated to differences in 18:3 between G. soja and G. max.

Transcriptionally more active G. max DGAT1A and DGAT1B genes, as it was observed during this research work, is consistent with the higher oil content and lower 18:3 levels in G. max when compared with G. soja. The higher expression of the soybean DGAT1A and DGAT1B could be the result of differences in regulating factor(s) acting in “trans”, which remains to be tested. Additionally, the higher PDAT transcript abundance relative to DGAT transcripts observed in G. soja as compared to G. max might explain, at least in part, the elevated 18:3 and low oil contents typically measured in seeds of the wild ancestor.

The relative contribution of the acyl editing mechanisms as well as other enzymes that take part in TAG synthesis to overall TAG composition still needs to be completely understood. As presented by Lu et al. (2011), the FA fluxes through PC are of great importance in the determination of TAG composition since PC is the source of all the PUFAs. The majority of the fatty acids synthesized in plastids enter PC by exchanging with PUFA in the acyl editing process (Bates et al. 2009; Bates et al. 2007). Also, LPCAT, by the reverse reaction, releases PUFAs that in turn will be part of the cytoplasmic acyl-CoA pool (Stymne and Stobart 1984). Because of the key role that
LPCAT has in regenerating PC from LPC, and since PC is the only substrate for the desaturases, the involvement of LPCAT in determining oil composition could be significant. Another enzyme that could play an important role in TAG composition is PLA2, which catalyzes the hydrolysis of acyl groups in position 2 of PC producing free PUFAs for activation by ACS (Weselake et al. 2009). Finally, the contribution of acyltransferases on FA composition through substrate preferences has been recently reviewed (Snyder et al. 2009). However, the actual input of each of these enzymes to seed oil quality is still uncertain as only DGAT1 and PDAT1 knockouts have been characterized in plants.

Detailed understanding of TAG metabolism and its regulation is still lacking. However, research on oil traits has been dynamic, particularly in the model plant Arabidopsis, and that will certainly have a positive impact in the improvement of oilseeds.

The content of 18:3 in seeds of G. soja has shown to be a quantitative trait, highly influenced by the environment and with very low narrow-sense heritability. These features indicate that when using G. soja as donor of the high 18:3 trait, very small response to selection should be expected per breeding cycle.

Also, in the G. max x G. soja RIL population generated for this study, a higher proportion of the individuals resembled more the G. soja than the G. max parent for
almost all the traits analyzed including seed dry weight, seed oil and oleic content, days to maturity and general appearance of the plants. The effects of using *G. soja* as donor of a valuable phenotypic feature on other traits of economic importance should be considered in the breeding strategy.

Finally, no individual in the RIL population showed significantly higher seed 18:3 content than the *G. soja* parental line PI 366121. This fact indicates that all genes and/or alleles for elevated seed 18:3 were contributed by the *G. soja* accession and none by the *G. max* cultivar Williams 82.
Bibliography


Zhang M, Fan J, Taylor DC, Ohlrogge JB (2009) DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21:3885-3901

VITA

Julián Mario Lenis was born January 11, 1977 in San Miguel de Tucumán, Tucumán, Argentina to Julián Mario Lenis and Victoria Ángela Gutiérrez. Julián attended “Escuela General Belgrano”, a primary school, and “Escuela de Agricultura y Sacarotecnia”, a 6-year technical high school, graduating in 1995 as an Agronomist. Julián started his college education in 1996 at “Universidad Nacional de Tucumán” (UNT), earning the degree of Engineer in Agronomy in 2003. In 1998, Julián started working on different aspects of soybean research in the Grain Crop Department at Estación Experimental Agroindustrial “Obispo Colombres”, Tucumán, Argentina, under the supervision of the investigator Mario Devani. Having always had the desire to pursue graduate studies, Julián moved to the U.S. in 2005 and started working on a M.Sc. degree under the supervision of Drs. Grover Shannon and David Sleper. During his M.Sc., Julián carried out experiments to study the tolerance to salinity of different accessions of the genus *Glycine* as well as identify physiological traits that might explain the ability of some genotypes to grow under saline soil conditions. In 2008, right after completion of his M.Sc., Julián joined Dr. Kristin Bilyeu lab to pursue a Ph.D. degree under her supervision and co advised by Dr. Grover Shannon.

Julián holds the highest GPA ever obtained in the “Facultad de Agronomía y Zootecnia”, UNT. He is the main author of two research articles, has co edited a book and co authored book chapters as well as research and extension publications on soybean.