Evaluation of Genes Encoding the Enzymes of the Kennedy Pathway in Soybeans with Altered Fatty Acid Profiles

by

Amy J. M. McNaughton

A Thesis
presented to
The University of Guelph

In partial fulfillment of requirements
for the degree of
Doctor of Philosophy
in
Plant Agriculture

Guelph, Ontario, Canada

ABSTRACT

EVALUATION OF GENES ENCODING THE ENZYMES OF THE KENNEDY
PATHWAY IN SOYBEANS WITH ALTERED FATTY ACID PROFILES

Amy J. M. McNaughton   Advisor:
University of Guelph, 2012   Dr. Istvan Rajcan

Soybean (*Glycine max* (L.) Merr) is the largest oil and protein crop in the world and it is grown for both oil and protein. To address the needs of both the edible oil market and industrial applications of soybean oil, fatty acid modification has been a focus of soybean breeding programs. Natural variation, mutagenesis and genetic engineering have been used to alter the fatty acid profile. Several genes, mostly desaturases, have been associated with altered fatty acid profiles but enzymes in the Kennedy Pathway have yet to be studied as another source of genetic variation for altering the fatty acid profiles. The Kennedy Pathway is also known as the oil producing pathway and consists of four enzymes: glycerol-3-phosphate acyltransferase (G3PAT); lysophosphatidic acid acyltransferase (LPAAT); phosphatidic acid phosphatase (PAP); and diacylglycerol acyltransferase 1 (DGAT1). The starting material for this pathway is glycerol-3-phosphate, which is produced from glycerol by glycerol kinase (GK), and the product of this pathway is triacylglycerol (TAG). The overall objective of this study was to elucidate the role that the Kennedy Pathway plays in determining the fatty acid profile in two ways: (1) sequencing the transcribed region of the

genomic genes encoding the enzymes of GK, G3PAT, LPAAT, and DGAT1 in soybean genotypes with altered fatty acid profiles; and (2) studying their expression over seed development, across three growing temperatures. The genetic material for the study consisted of four soybean genotypes with altered fatty acid profile: RG2, RG7, RG10, and SV64-53. Results from sequencing showed that the mutations identified in G3PAT, LPAAT, and DGAT1 in the four soybean genotypes did not explain the differences in the fatty acid profiles. The expression of G3PAT, LPAAT, and DGAT1 over seed development showed that G3PAT had the lowest levels, followed by LPAAT, then DGAT1, across the growing temperatures. The differences in expression among genotypes corresponded to differences in fatty acid accumulation, suggesting that expression rather than genetic mutations in the transcribed region of the genes influenced the fatty acid profile of the genotypes in this study. In conclusion, the enzymes of the Kennedy Pathway appear to contribute to the altered fatty acid profiles observed in the soybean mutant genotypes.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr. Istvan Rajcan, for accepting me into his lab as a Masters student and supporting me in my transfer to the PhD program. I would also like to thank him for his guidance throughout my studies. Secondly, I would like to thank my advisory committee – Dr. Gary Ablett, Dr. Barry Shelp, Dr. Peter Pauls, and Dr. Ian Tetlow – for their guidance and suggestions. Thirdly, I want to thank my examination committee Dr. Barry Micallef and Dr. Ian Tetlow, my external examiner Dr. David F. Hildebrand and Dr. Clarence Swanton for chairing my examination.

I would also like to thank Chris Grainger for all of his assistance in the molecular lab and for putting up with all of my questions. I would like to thank Dr. Jeff Boersma and Ron Fletcher for their assistance with the gas chromatograph and fatty acid analysis. Thank you to Julia Zilka and Lin Liao for their guidance in the growth room and growth cabinets and assistance with my plants. I want to thank the staff at the Genomics Facility, especially Jeff Gross and Jing Zhang, for all of their work and assistance with sequencing and real-time PCR.

I want to thank my labmates for listening to my frustrations and offering suggestions. I want to especially thank Milad Eskandari for his assistance with my statistical work and sequencing.

Thank you to my parents, Robbie and Ruth Ann, for supporting and believing in me and for encouraging me to achieve my dreams. Thank you to my "little" brother, Adam, for being you and fixing my car so many times. Thank you to my grandparents, Leonard and Donna, for supporting my dreams; I'm sorry

that you won't be there to see me achieve those dreams, may you rest in peace. Thanks to my cat, Molly, for always being on my side and being the cutest "human", may you rest in peace.

I want to thank my friends for being a shoulder to cry on, an ear to listen to all of the complaints, and providing a distraction when I needed it. Thank you to my family away from home, the Phillips', you have made me feel like I was never alone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	analysis of variance
Bps	base pairs
CRD	completely randomized design
cTP	chloroplast transit peptide
DAF	days after flowering
DGAT1	diacylglycerol acyltransferase 1
DNA	deoxyribonucleic acid
EMS	ethylmethane sulfonate
ER	endoplasmic reticulum
FA	fatty acid
G3PAT	glycerol-3-phosphate acyltransferase
GC	gas chromatography
GK	glycerol kinase
LPAAT	lysophosphatidic acid acyltransferase
lsmeans	least squares means
MS	mean squares
mTP	mitochondrial transit peptide
NMR	nuclear magnetic resonance
PAP	phosphatidic acid phosphatase
PCR	polymerase chain reaction
pdiff	least significant difference
qRT-PCR	quantitative real-time polymerase chain reaction

RNA	ribonucleic acid
SD	standard deviation
SP	signal peptide
SE	standard error
TAG	triacylglycerol

CHAPTER 1: GENERAL INTRODUCTION

1.1 Soybeans (*Glycine max* (L.) Merr.)

Soybeans were domesticated between the 17th and 11th centuries B.C. in

China and today soybeans are the world's leading oilseed crop, producing

approximately 9.5 billion bushels in 2010 (Hymowitz, 1970; American Soybean

Association, 2010). The United States is the leading producer of soybeans, with

35% of the world's production, followed by Brazil (27%) and Argentina (19%),

with Canada ranked 7th (2%) (American Soybean Association, 2010). Within

Canada, approximately 160 million bushels of soybeans were grown last year,

with Ontario growing approximately 70% of that or 112 million bushels (Statistics

Canada, 2010).

Soybeans contain approximately 40% protein and about 20% oil

(Hildebrand *et al.*, 2008). Protein meal is one use of the protein and soy protein

meal makes up 69% of the world consumption of protein meal (American

Soybean Association, 2010). The oil has many uses. One use is vegetable oil,

with soy oil accounting for 29% of the world's consumption of such oils (American

Soybean Association, 2010). Soy oil is also used as an industrial lubricant and

as a feedstock for producing automotive parts.

The soybean genome was completed in 2010 and is now available online

(Schmutz *et al.*, 2010). This information allows researchers to search for genes

important to their studies. Since it is known that the soybean genome underwent

a whole genome duplication, multiple copies of each gene may be present

making genetic studies more complicated (Schmutz *et al.*, 2010).

1.2 The Kennedy Pathway

The Kennedy pathway is also known as the oil producing pathway or the triacylglycerol biosynthetic pathway. The pathway was first described by Eugene P. Kennedy in 1961 in mouse liver and adipose tissue but is also known to be present in plants (Nelson & Cox, 2005). In plants, triacylglycerol serves as a storage material and is found in lipid bodies within the plant cell (Ohlrogge & Browse, 1995). An overview of the pathway is shown in Figure 1.1, and a depiction of a generic triacylglycerol molecule is provided in Figure 1.2.

The initial substrate of the pathway is glycerol-3-phosphate produced from glycerol, a reaction catalyzed by glycerol kinase (E.C. 2.7.1.30) (Coleman & Lee, 2004). Glycerol-3-phosphate can also be produced from dihydroxyacetone phosphate via glycerol-3-phosphate dehydrogenase (E.C. 1.1.1.8) (Shen *et al.*, 2006). The glycerol backbone undergoes a series of reactions that attach three fatty acids, one to each of the three *sn*-positions in a step-wise fashion. These reactions are catalyzed by three enzymes – namely, glycerol-3-phosphate acyltransferase (E.C. 2.3.1.15), lysophosphatidic acid acyltransferase (E.C. 2.3.1.51), and diacylglycerol acyltransferase (E.C. 2.3.1.20) (Martin & Wilson, 1983). Before the third fatty acid is attached, the *sn*-3 position must be dephosphorylated by phosphatidic acid phosphatase (E.C. 3.1.3.4) (Frentzen, 1998). Glycerol kinase and the three acyltransferase enzymes are the focus of this study and they will be reviewed and discussed in further detail in subsequent chapters.
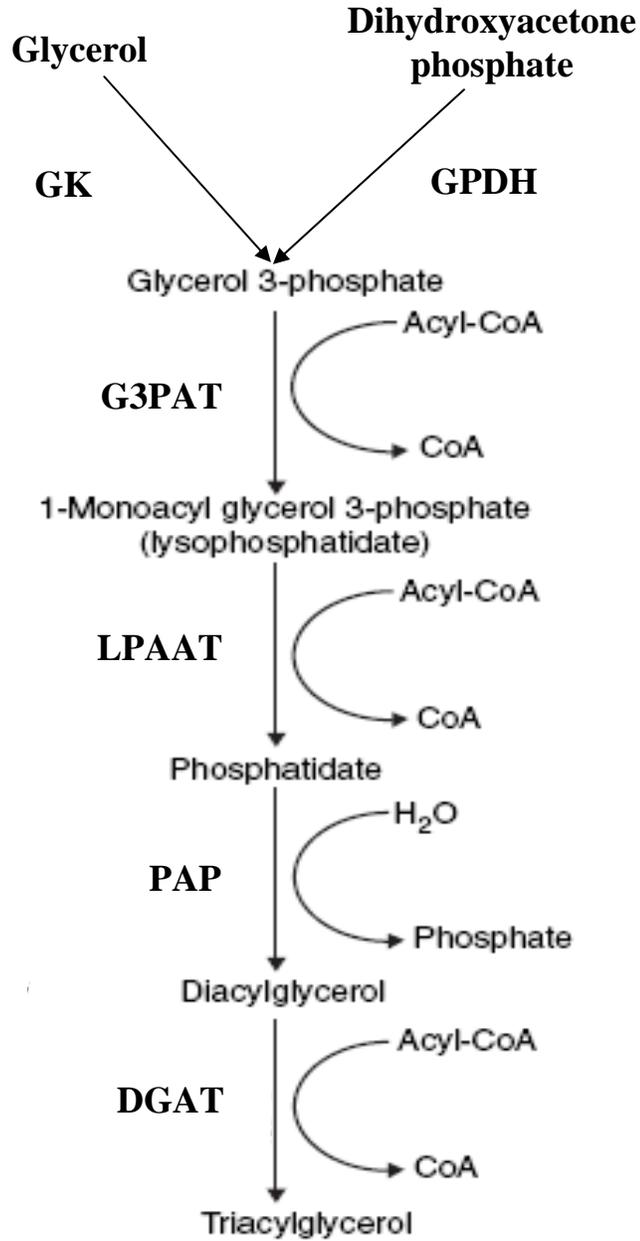
Figure 1.1: Kennedy Pathway schematic. Simplified schematic of the triacylglycerol biosynthesis pathway, also known as the Kennedy Pathway – including the reaction directly upstream, which is catalyzed by GK (adapted from Ramli *et al.*, 2005). GK – glycerol kinase; GPDH – glycerol-3-phosphate dehydrogenase; G3PAT – glycerol-3-phosphate acyltransferase; LPAAT – lysophosphatidic acid acyltransferase; PAP – phosphatidic acid phosphatase; DGAT1 – diacylglycerol acyltransferase

sn-1 position $CH_2-O-\overset{\overset{\textstyle O}{\|}}{C}$〰〰〰〰$R_1$

sn-2 position $CH-O-\overset{\overset{\textstyle O}{\|}}{C}$〰〰〰〰$R_2$

sn-3 position $CH_2-O-\overset{\overset{\textstyle O}{\|}}{C}$〰〰〰〰$R_3$
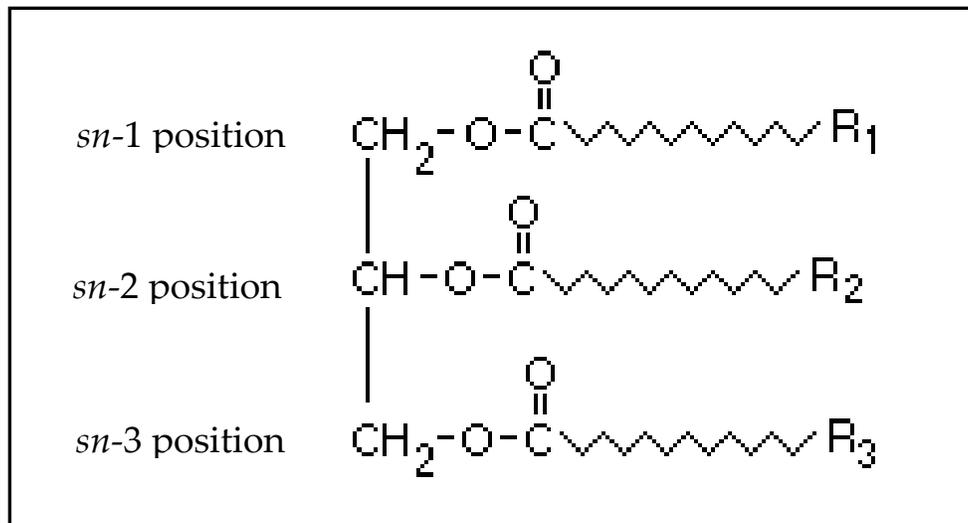
Figure 1.2: Schematic of triacylglycerol molecule. Generalized schematic of the triacylglycerol molecule showing the 3 positions on the glycerol backbone where the fatty acids are attached in a stepwise fashion. (Adapted from http://library.med.utah.edu/NetBiochem/FattyAcids/7_2.html)

1.3 Bioproducts

Soybean oil has several applications besides products for human consumption. One such application is industrial lubricants, especially in situations where its biodegradability is important to decrease environmental impacts – for example, ground drilling and marine engines (Kinney, 1998). Another application is a renewable feedstock for producing automotive fuel, specifically diesel (Durrett *et al.*, 2008). One application pertinent to this study is the use of soy oil as a base for polyol production (John *et al.*, 2002). Polyols are used to make materials like polyurethane foam, which are used to build car seats (Sain, 2006).

This study was part of the larger project known as The BioCar Initiative, within the oil stream. The overall goal of the Initiative was to make bioproducts competitive, both in functional value and in price, with the current synthetic products used in the automotive industry (Sain, 2006).

CHAPTER 2: LITERATURE REVIEW

2.1 History of Soybeans

Soybean originated in China, where it was domesticated between the 17th and 11th centuries B.C. (Hymowitz, 1970). It was not until an American seaman named Samuel Bowen returned to Savannah, Georgia from a trip to China, via London, in 1765 that soybeans were introduced to North America (Hymowitz & Harlan, 1983). In 1881 soybeans obtained from John A. Bruce & Co. in Hamilton, Ontario were first grown in Canada at the Ontario Agricultural College, but this did not continue (Shurtleff & Aoyagi, 2010). Finally, in 1893 soybeans were again grown at the Ontario Agricultural College using seeds obtained by Charles A. Zavitz from Prof. Charles C. Georgeson at the Kansas Agricultural Experiment Station (Shurtleff & Aoyagi, 2010). From this, Charles A. Zavitz started to develop soybeans that would be suitable for growing in Ontario; the first Canadian variety, OAC21, was registered in 1923 (Shurtleff & Aoyagi, 2010).

2.2 Soybean Oil

There are five main fatty acids in soybean seeds that occur in typical levels: palmitic acid (16:0) (11%); stearic acid (18:0) (4%); oleic (18:1Δ^9) (25%); linoleic acid (18:2$\Delta^{9,12}$) (52%); and α-linolenic acid (18:3$\Delta^{9,12,15}$) (8%) (Fehr, 2007). These same five fatty acids are also found in lipid membranes, with palmitic and linoleic acids predominating, however the amount of each fatty acid present is different from the amount in seeds (Rivera & Penner, 1978). In soybean root plasmamembranes, the amounts of each fatty acid present in

plants grown at 22°C are as follows: palmitic acid – 30%, stearic acid – 16%, oleic acid – 9%, linoleic acid – 25%, and linolenic acid – 21% (Rivera & Penner, 1978). Figure 2.1 gives a general overview of how these fatty acids are formed and stored in soybeans. Depending on the end use of the soy oil, the concentration of each fatty acid may need to be altered to fit specific needs. Studies have investigated genetic and environmental influences on the typical fatty acid profile.

2.2.1 Altered Fatty Acids

2.2.1.1 Palmitic Acid

Palmitic acid (16:0) is a saturated fatty acid that is detrimental to human health as it raises the so-called "bad cholesterol" (low density lipids, LDL); therefore, reducing the level of palmitic acid in soybean oil would help to reduce saturated fats in food products (Heart & Stroke Foundation, 2008). On the other hand, elevated palmitic acid helps to improve the oxidative stability in soybean oil (Shen *et al.*, 1997). Elevated palmitic acid genotypes have decreased yield, and other agronomic traits such as emergence are altered (Hayes *et al.*, 2002).

Several loci with differing alleles that contribute to altering this fatty acid have been identified. The first two loci studied were *fap1* and *fap2*, which are responsible for reduced and elevated palmitic acid levels, respectively (Erickson *et al.*, 1988). Fehr *et al.* (1991a) created a mutant soybean line that has elevated palmitic levels controlled by another allele at the *fap2* locus, named *fap2-b*.

The *fap3* (originally named *fapx*) locus was identified by Schnebly *et al*.

7
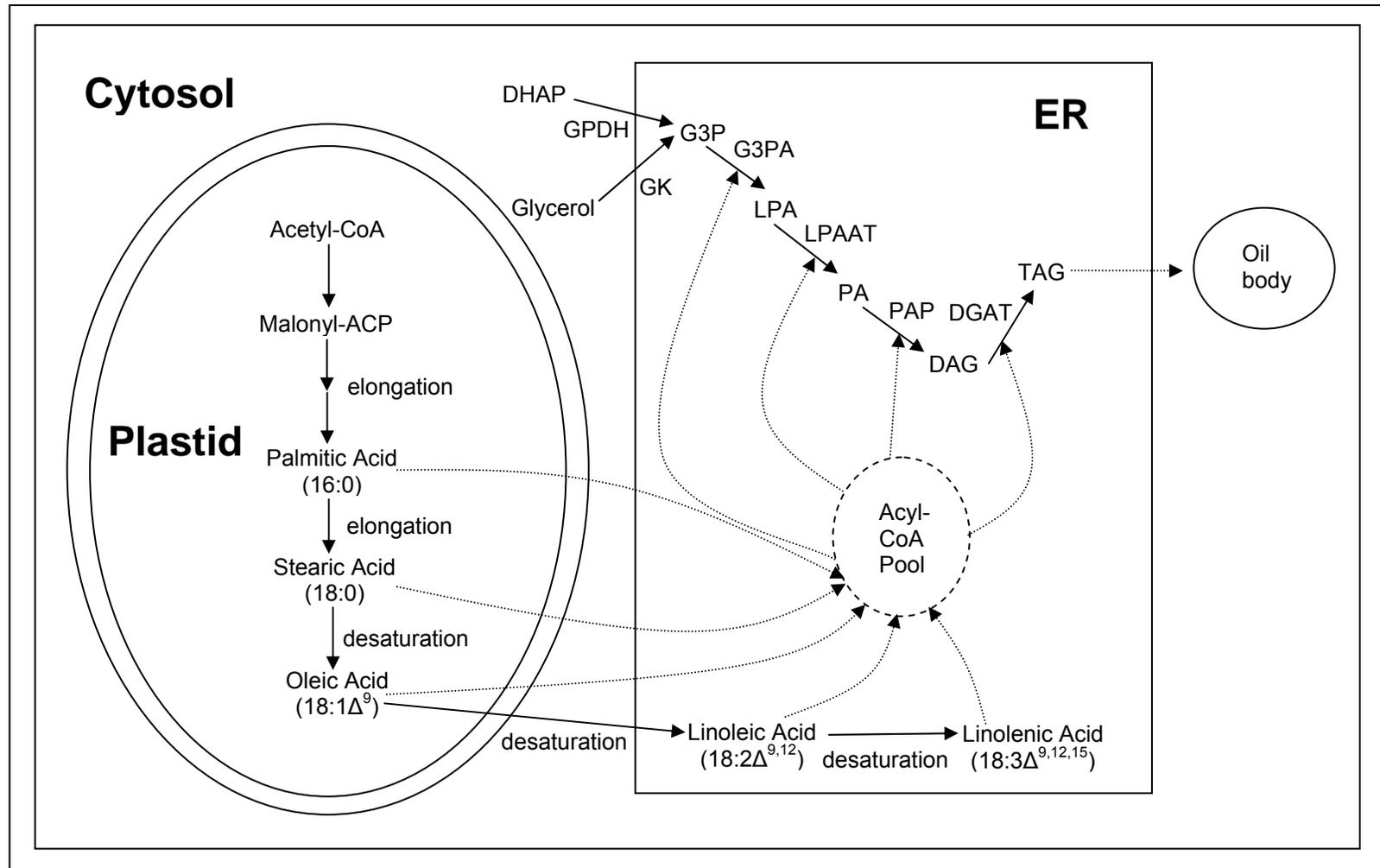
Figure 2.1: Fatty acid and triacylglycerol biosynthesis in soybean seeds. Simplified schematic of fatty acid and triacylglycerol biosynthesis in soybean seeds. GK – glycerol kinase; DHAP – dihydroxyacetone phosphate; GPDH – glycerol-3-phosphate dehydrogenase; G3P – glycerol-3-phosphate; G3PAT – glycerol-3-phosphate acyltransferase; LPA – lysophosphatidic acid; LPAAT – lysophosphatidic acid acyltransferase; PA – phosphatidic acid; PAP – phosphatidic acid phosphatase; DAG – diacylglycerol; DGAT – diacylglycerol acyltransferase; TAG – triacylglycerol; ACP – acyl carrier protein. (Adapted from Hildebrand *et al.*, 2008)

(1994) in a line that was developed by Fehr *et al.* (1991b) and was associated with a low palmitic acid phenotype. Two other alleles at the *fap3* locus have been identified in other mutant genotypes. The first was a natural mutation (meaning neither chemically induced or via transformation) reported by Burton *et al.* (1994) and was named *fap3-nc*. The second was a chemically-induced mutation named *fap3-ug*, originally named *fap** (Stojsin *et al.*, 1998a).

In addition to *fap2*, alleles at four other loci are associated with elevated levels of palmitic acid. One of these is named *fap4* (originally named *fap?*), which was a mutation created by chemical treatment (Schnebly *et al.*, 1994). The mutation at the *fap5* locus was also created by chemical treatment, and it was found that this locus is closely linked to the *fap2-b* locus (Stoltzfus *et al.*, 2000a). The fourth locus associated with elevated palmitic acid concentration was discovered in a chemically mutated line and was named *fap6* (Narvel *et al.*, 2000). The *fap7* locus was first noted by Stoltzfus *et al.* (2000b) in a chemically-mutated line, and was reported to be closely linked to the *fap6* locus.

2.2.1.2 Stearic Acid

Stearic acid (18:0) is a saturated fatty acid that is neutral to human health. The level of stearic acid is already fairly low in soybeans and advancements in lowering stearic acid were achieved in conjunction with lowering palmitic acid, resulting in lower levels of total saturated fatty acids (Fehr, 2007). Increasing stearic acid can decrease the need for hydrogenation for creating *trans*-fat free margarines and shortenings from soybean oil by the use of interesterification

(List *et al.*, 1997). Increasing stearic acid can be difficult because of a concurrent decrease in yield and unpredictable emergence rates (Fehr, 2007). It should be noted that the decrease in yield appears to occur only in genotypes with the *fas1-a* allele (Lundeen *et al.*, 1987).

Two different loci have been associated with elevated stearic acid levels. The first locus has several different alleles where all but one was created using chemical mutation. The first of these alleles was named *fas1* and was created using ethyl-methanesulfonate (EMS) (Graef *et al.*, 1985). The second, *fas1-a*, was created using sodium azide (Graef *et al.*, 1985). The third is *fas1-b*, and it was also created using EMS (Graef *et al.*, 1985). When these three alleles were tested to understand levels of dominance, it was shown that *fas1-a* and *fas1-b* were not dominant against each other but both were completely dominant over *fas1* (Graef *et al.*, 1985).

A natural mutation, named *fas1-nc*, was shown to be allelic to the chemically-induced *fas1-a* mutation (Pantalone *et al.*, 2002). Three additional alleles were also identified after the EMS treatment of three different genotypes and they were named *fas1*-(Ames 1), *fas1*-(Ames 2), and *fas1*-(Ames 3) (Bubeck *et al.*, 1989). In the same study, another line that was mutated using EMS also showed elevated stearic acid levels, but the allele in this case was not at the *fas1* locus – although the locus of this allele has not yet been named (Bubeck *et al.*, 1989).

2.2.1.3 Oleic Acid

Oleic acid ($18:1\Delta^6$) is unsaturated with one double bond and has a higher oxidative stability than linoleic acid ($18:2\Delta^{9,12}$) and linolenic acid ($18:3\Delta^{9,12,15}$); therefore, efforts have been made to increase this fatty acid (Fehr, 2007). Through conventional breeding, greater than 70% oleic acid has been achieved but through genetic engineering efforts, this maximum has exceeeded 80% and even over 90% (Alt *et al.*, 2005b; Kinney & Knowlton, 1998; Buhr *et al.*, 2002). Within conventional breeding, two mutant genotypes were created using X-ray radiation (Takagi & Rahman, 1996). The first mutant was found to have a deletion in a gene that controls the conversion of oleic acid to linoleic acid, and the allele was named *ol* (Takagi & Rahman, 1996). When the inheritance of this allele was studied, it was found that even in the homozygous state (*olol*), there was still a range of oleic acid concentration in the progeny, suggesting that there are other genes that modify this trait as well (Alt *et al.*, 2005a). The second mutant had a mutation at the same locus as the first but it had a different allele and this one was named *ol^a* (Rahman *et al.*, 1996). It was shown that the allele for low oleic acid (*Ol*) is partially dominant to *ol* but completely dominant to *ol^a*, whereas *ol^a* is completely dominant to *ol* (Rahman *et al.*, 1996). Low levels of oleic acid did not appear to be a focus of any breeding efforts.

2.2.1.4 Linoleic Acid

Linoleic acid ($18:2\Delta^{9,12}$) is negatively correlated with oleic acid, showing a strong inverse relationship (Fehr, 2007). Linoleic acid is already present at a

fairly high level in soybeans, and therefore there has been lower interest than in the other fatty acids in modifying linoleic acid level (Pantalone *et al.*, 2004).

2.2.1.5 α-Linolenic Acid

α-Linolenic acid ($18:3\Delta^{9,12,15}$), called linolenic acid from here on, has poor oxidative stability, leading to off-flavours and odors in many food applications (Dutton *et al.*, 1951; Primomo *et al.*, 2002). Efforts have been made to lower the levels of linolenic acid present so that hydrogenation of oil can be minimized – since this causes the production of *trans*-fats (Fehr, 2007). While linolenic acid is beneficial to humans on its own, it is the processing of this fatty acid that renders it detrimental. Four different loci have been identified as responsible for reduced levels of linolenic acid, with each locus having at least two different alleles.

The first of these loci is the Fan1 locus, which has three different alleles. The first allele is *fan1*, and this has five different variations based on the line in which they were developed. The first *fan1* allele was created through EMS mutagenesis in a line named C1640 and was named simply *fan1* (Wilcox & Cavins, 1985; Wilcox & Cavins, 1987). Using X-ray irradiation, two genotypes were created with mutations at the *fan1* allele, and they were named *fan1*(M-5) and *fan1*(IL-8) after the genotypes in which they are found (Rahman & Takagi, 1997; Rahman *et al.*, 1995). EMS chemical treatment was used to create the line A5, which also had a mutation at the Fan1 locus and was named *fan1*(A5) (Hammond & Fehr, 1983; Rennie & Tanner, 1991a). Another line created using X-ray irradiation was named J18 and it contained a mutation at the Fan1 locus;

the associated allele was named *fan1*(J18) (Anai *et al.*, 2005). The second allele at the Fan1 locus is called *fan1-b* and was found in the EMS-created line RG10 (Stojsin *et al.*, 1998b). A further study of RG10 showed that the mutations in the *fan1-b* allele are in two genes – *Fad3A* and *Fad3B* (Reinprecht *et al.*, 2009). The third allele at the Fan1 locus is currently known as *fan1-nc*, previously known as *fan1*(PI123440) (Rennie & Tanner, 1989a as cited in Fehr, 2007).

The second locus, named Fan2, has two different alleles. The first allele is known as *fan2* and was identified in an EMS-mutated line named A23 (Fehr *et al.*, 1992). The second allele was named *fan2-b* and was found in the line CX1512-44 (Bilyeu *et al.*, 2005).

The third locus in the Fan family, Fan3, has two different alleles. The first allele was discovered in EMS-treated line A26 and was named *fan3* (Fehr & Hammond, 2000). The second allele was named *fan3-b* and was noted in X-ray mutant line M24 (Anai *et al.*, 2005).

The fourth locus is named Fanx and it has two different alleles. The first allele is named *fanx* and was identified in KL-8, which was created using X-ray irradiation (Rahman & Takagi, 1997). The second allele was found in the X-ray mutant line mentioned above, M24 and was named *fanx[a]* (Rahman *et al.*, 1998). To date, up to three of these loci – namely *fan1*(A5), *fan2*, *fan3* – have been combined into one line, called A29, creating an ultra-low linolenic acid line containing approximately 1% linolenic acid (Ross *et al.*, 2000).

Despite the problems with higher linolenic levels for food production and oil stability, high linolenic acid can be useful in industrial applications (Fehr,

2007). Through increasing the expression of a key enzyme (*FAD3*), soybean

seeds have been produced with more than 50% linolenic acid for use in industrial

applications, but no commercial cultivars are currently available with elevated

levels of linolenic acid (Cahoon, 2003).

2.2.2 Effect of Temperature on Fatty Acids

The air temperature during the critical seed filling influences both the

overall oil concentration as well as the fatty acid composition. For example,

when the daytime temperature was increased from 70°F (21.1°C) to 85°F

(29.4°C) for one week during the seed filling period, the total oil concentration

was increased from 19.6% to 22% (Howell & Cartter, 1958). A larger experiment

using six temperature treatments (20, 26, 27, 29, 33, and 35°C) indicated that the

higher temperatures decreased the total concentration of oil present, and

decreased the polyunsaturated fatty acid concentration (Dornbos Jr. and Mullen,

1992). As for the individual fatty acids, linolenic acid and linoleic acid decreased

as the temperature increased using five temperature regimes (18/13, 24/19,

27/22, 30/25, and 33°C/28°C day/night), while oleic acid increased (Wolf *et al.*,

1982). Neither palmitic acid nor stearic acid were affected by increasing

temperature (Wolf *et al.*, 1982). When a high stearic line, A6, that contains the

fas1-a allele was tested under different growing temperatures (15/12, 28/22, and

40°C/30°C day/night), the levels of stearic acid were low at a low growing

temperature suggesting that the growing temperature influences the expression

of the *fas1-a* allele (Rennie & Tanner, 1989b). Hou *et al.* (2006) found that the

levels of linolenic acid was less stable over environments than those for the other

fatty acids.

Air temperature also influences the fatty acid concentration in leaves for

some of the fatty acids, such as linolenic and oleic acids, which is opposite to the

effect on seed oil. Linoleic and oleic acids decreased with increasing temperature

across three temperature regimes (15/12, 28/22, and 40°C/30°C day/night), while

linolenic acid increased and there were no changes for palmitic and stearic

(Rennie & Tanner, 1991b). There were also three other fatty acids present in the

leaves that are not present in seed oil – palmitoleic (16:1), γ-linolenic (18:3g),

and behenic (22:0), but the last one was only present at low temperatures

(15°C/12°C day/night) (Rennie & Tanner, 1991b).

2.3 The Kennedy Pathway

To date, fatty acid composition in soybeans has been altered by altering

the fatty acid pool through fatty acid biosynthesis. The effects that enzymes of

the Kennedy Pathway (see Fig. 1.1) can have on the final fatty acid profile of

soybeans has not been investigated. The Kennedy Pathway utilizes the fatty

acids present in the fatty acid pool to create triacylglycerol (see Fig. 1.2), which is

the oil storage molecule. This pathway generally occurs in the endoplasmic

reticulum, although there are other forms of these enzymes with different

subcellular localizations (Table 2.1). The enzymes in the Kennedy Pathway are

reviewed below.

New evidence suggests that the production of triacylglycerol in plants is

not as straightforward as first thought (Fig. 2.2) – via the Kennedy Pathway (Fig. 1.1). The extent to which plants utilize these alternative pathways differs among species (Chapman & Ohlrogge, 2012). Some of the enzymes involved in the alternate pathways are briefly reviewed below, as they are not included in the study presented herein.

2.3.1 Glycerol Kinase (GK)

Glycerol kinase (GK) creates the starting material – glycerol-3-phosphate – of the Kennedy Pathway (Coleman & Lee, 2004). An early study showed that GK has a cytoplasmic cellular localization and its activity is optimal at pH 9.0 (Sadava & Moore, 1987). It has been suggested that during triacylglycerol accumulation, the production of glycerol-3-phosphate may be the limiting step (Baud *et al.*, 2008). Glycerol-insensitive *Arabidopsis thaliana* mutants have one mutation (*gli1*), which mapped to the location of the glycerol kinase gene on Chromosome 1 and the cDNA was cloned (Eastmond, 2004). In this same study, it was shown that *GLI1* is expressed in all tissues, is upregulated during senescence and early post-germinative growth, and is cytosolic (Eastman, 2004). Huang (1975) showed that GK is also involved during post-germination, where it phosphorylates the glycerol released from the metabolism of triacylglycerol for further processing. Kang *et al.* (2003) studied non-host resistance in *Arabidopsis* and found that the gene *NHO1* is required for resistance. They found that this gene encodes a GK enzyme. When soybeans were subjected to drought stress, GK expression increased in the roots suggesting that it is involved in drought
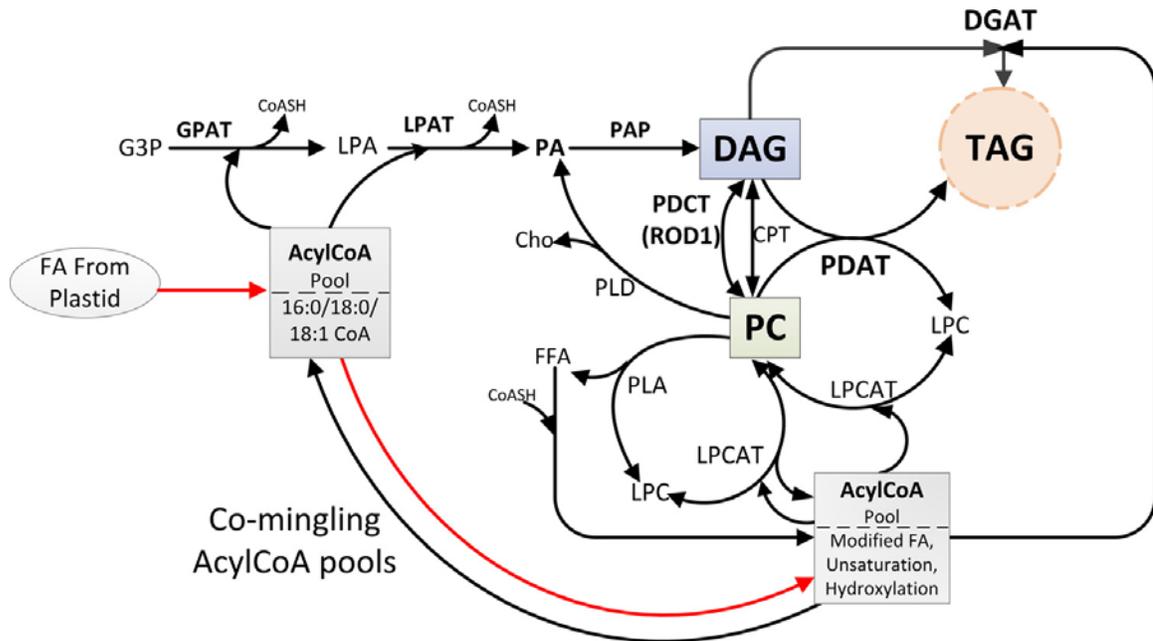
Figure 2.2: Alternative pathways to produce triacylglycerol. From Chapman & Ohlrogge (2012). G3P – glycerol-3-phosphate; GPAT – glycerol-3-phosphate acyltransferase; LPA – lysophosphatidic acid; LPAT – lysophosphatidic acid acyltransferase; PA – phosphatidic acid; PAP – phosphatidic acid phosphatase; DAG – diacylglycerol; DGAT – diacylglycerol acyltransferase; TAG – triacylglycerol; PC – phosphatidylcholine; LPCAT – lysophosphatidylcholine acyltransferase; LPC – lysophosphatidylcholine; PDCT – phosphatidylcholine:diacylglycerol phosphocholine transferase; PDAT – phospholipid:diacylglycerol acyltransferase; Cho – choline; CPT – choline-phosphotransferase; PLA – phospholipase A; PLD – phospholipase D; FFA – free fatty acid

Table 2.1: Subcellular localization of Kennedy Pathway enzymes. Enzyme isoforms pertinent to TAG biosynthesis in seeds are bolded.

Enzyme	Localization	Reference
Glycerol Kinase	**Cytosol**	Sadava & Moore, 1987
Glycerol-3-Phosphate Acyltransferase	Mitochondria	Coleman & Lee, 2004
	Cytosol	Coleman & Lee, 2004
	Chloroplast	Xu *et al.*, 2006
	Endoplasmic Reticulum	Slabas *et al.*, 2001
Lysophosphatic Acid Acyltransferase	Mitochondria	Coleman & Lee, 2004
	Endoplasmic Reticulum	Hares & Frentzen, 1987
	Plastid	Yu *et al.*, 2004
Phosphatidic Acid Phosphatase	Chloroplast	Joyard & Douce, 1979
	Cytosol	Perry *et al.*, 1999
	Endoplasmic Reticulum	Perry *et al.*, 1999
Diacylglycerol Acyltransferase	**Endoplasmic Reticulum**	Zhang *et al.*, 2005
	Chloroplast	Martin & Wilson, 1984

stress response (Alam *et al.*, 2010).

2.3.2 Glycerol-3-Phosphate Dehydrogenase (GPDH)

An alternate path for making glycerol-3-phosphate from dihydroxyacetone phosphate is catalyzed by glycerol-3-phosphate dehydrogenase (GPDH) (Shen *et al.*, 2006). It has been suggested that it is this reaction by GPDH that creates glycerol-3-phosphate for triacylglycerol, although the relative importance of GPDH and GK to provide glycerol-3-phosphate for triacylglycerol synthesis has yet to be determined (Frentzen, 1993; Vigeolas *et al.*, 2007). GPDH that creates glycerol-3-phosphate is localized to the cytosol and is NADH dependent (Finlayson & Dennis, 1980). The optimal pH for this reaction is 8.1 (Finlayson & Dennis, 1980). When a yeast GPDH was over-expressed in *Brassica napus*, there was a 40% increase in seed oil content, suggesting that the amount of glycerol-3-phosphate being produced limits, in part, the accumulation of oil in seeds (Vigeolas *et al.*, 2007). This enzyme has also been called dihydroxyacetone phosphate reductase in some literature, as well as NAD^+:glycerol-3-phosphate oxidoreductase (Kirsch *et al.*, 1992; Frentzen, 1993).

2.3.3 Glycerol-3-Phosphate Acyltransferase (G3PAT)

Glycerol-3-phosphate acyltransferase (G3PAT) catalyzes the first reaction in the Kennedy Pathway. Four G3PATs have been found in Arabidopsis, which differ in their subcellular location. One is bound to the mitochondrial membrane, another is cytosolic, the third is in the chloroplast stroma, and a fourth G3PAT is

bound to the endoplasmic reticulum (Slabas *et al.*, 2001; Coleman & Lee, 2004;

Xu *et al.*, 2006). It is thought that there are up to 20 genes in soybean that

encode a G3PAT (www.uky.edu/SoyLipidGenes).

Despite the many forms, there are four main motifs in the amino acid

sequences that are conserved across species (Fig. 2.3), and these four motifs

are located in the N-terminal region where the active site of the enzyme is

located (Coleman & Lee, 2004). G3PAT from avocados and safflower is most

active with saturated fatty acids as substrate, but G3PAT isolated from

chloroplasts of pea and spinach prefers oleic acid (Frentzen, *et al.*, 1983; Bafor

et al., 1990; Eccleston & Harwood, 1990). In safflower, G3PAT from the

particulate fraction was shown to have a higher level of specificity and selectivity

under different temperatures (Ichirara, 1984). The preference for oleic acid in

spinach creates a product that is more efficient for the subsequent reaction

compared to a product containing palmitic acid in the *sn*-1 position (Hares &

Frentzen, 1987). The plastidial G3PAT and the mitochondrial G3PAT have

different specificities and selectivities in terms of fatty acid preference; for

example, the plastidial form prefers oleic acid over palmitic acid whereas the

mitochondrial form prefers palmitic acid (Frentzen, 1993). Due to difficulties

isolating and purifying membrane-bound enzymes, no data regarding the fatty

acid preference for the ER bound G3PAT was available (Slabas *et al.*, 2001;

Chapman & Ohlrogge, 2012). Differing specificities among isoforms may

account for the discrepancies reported in different species (Frentzen, 1993). It

should be noted that specificty refers to the utilization and rate of incorporation of

	Motif I catalysis	Motif II G3P binding	Motif III G3P binding	Motif IV catalysis
GPAT				
E. coli	303 VPC**HR**SHM**D**YLL	348 GAFFI**RR**TF	382 Y**F**VE**GGR**SRTGRLLDPKTGTL	417 ITL**I**P**I**YI
S. cerevisiae GAT1[a]	77 CAP**H**AN**Q**FI**D**PA	131 GGIPVP**R**IQ	257 I**F**PE**GG**SHDRPSLLPIKAGVA	292 VAV**V**P**C**GL
S. cerevisiae GAT2[b]	87 AAP**H**AN**Q**FV**D**PV	134 MAIGVV**R**PQ	249 I**F**PE**GG**SHDRTNLLPLKAGVA	372 KLPL**P**LIV
Mouse	227 LPV**H**RSHI**D**YLL	272 GGFFI**RR**RL	312 I**F**LE**GTR**SRSGKTSCARAGLL	347 IL**V**IP**V**GI
A. thaliana	226 ISN**H**QSEA**D**PAV	254 KCVAG**D**RVI	319 IAPS**GGR**DRPNPSTGEWFPAP	359 GH**I**YP**M**SL

Figure 2.3: Functional motifs of G3PAT in the N-terminus. Motifs in the N-terminus that were determined to be involved in enzyme activity. Conserved amino acids are bolded. Adapted from Coleman & Lee, 2004.

21

a fatty acid when only one fatty acid is provided, whereas selectivity refers to the utilization and rate of incorporation of each fatty acid when multiple fatty acids are provided (Hares & Frentzen, 1987).

A family of G3PAT enzymes is present in arabidopsis, and one form (AtG3PAT1) is very important for pollen development (Zheng *et al.*, 2003). A deficiency in AtG3PAT1 is associated with several changes in fatty acid composition in seeds, but it does not affect the overall oil concentration (Zheng *et al.*, 2003).

A L261F substitution in squash G3PAT using site-directed mutagenesis causes the enzyme to become substrate selective, whereas it was non-selective previously (Slabas *et al.*, 2002). The authors show that Lys^{193}, His^{194}, Arg^{235}, and Arg^{237} are involved in binding the phosphate group; changes in Lys^{193}, Arg^{235}, and Arg^{237} inactivate the enzyme, whereas changes in His^{194} reduces activity (Slabas *et al.*, 2002). Frentzen *et al.* (1994) isolated and characterized the cDNA of the plastidial G3PAT from *Pisum sativum* and showed that when the first 35 amino acids on the N-terminus are eliminated, the protein loses functionality. The central region of G3PAT, specifically S243 and E305, confers substrate specificity and there was a loss of specificity when they were substituted for threonine and aspartic acid, respectively (Ferri & Toguri, 1997).

In Arabidopsis a mutant form of plastidial G3PAT was identified at the *ATS* locus (*ats1-1*), and the mutation was located at a typical splice-site of intron 5 and leading to three deviant transcripts (Xu *et al.*, 2006). This enzyme is regulated, at least in part, by dephosphorylation and phosphorylation (Lau &

Rodriguez, 1996; Coleman & Lee, 2004). Transforming Arabidopsis with a safflower G3PAT (plastidial form) or *Escherichia coli* G3PAT increases total oil concentration and seed weight (Jain *et al.*, 2000).   Not only are seeds affected by changes in G3PAT, wheat transformed with pea G3PAT has an altered membrane lipid composition including reduced levels of palmitic acid and oleic acid and elevated levels of linoleic acid (Edlin *et al.*, 2000).

2.3.4 Lysophosphatidic Acid Acyltransferase (LPAAT)

      Lysophosphatidic acid acyltransferase (LPAAT) catalyzes the second reaction in the Kennedy Pathway. This enzyme is located in mitochondria, and the endoplasmic reticulum and plastids (Coleman & Lee, 2004; Hares & Frentzen, 1987; Yu *et al.*, 2004). LPAAT is often highly substrate specific, with unsaturated fatty acids being preferred (Griffiths *et al.*, 1985). The level of selectivity of LPAAT has been shown to be species-specific; for example, safflower LPAAT is more selective in fatty acid utilization than turnip rape LPAAT (Bafor *et al.*, 1990). The length of the fatty acid is also important; for example, coconut LPAAT utilizes medium-length fatty acids more effectively than long-chain fatty acids (Knutzon *et al.*, 1995). These substrate specificities are quite stringent under *in vivo* conditions (Baud & Lepiniec, 2010). A full-length cDNA encoding LPAAT was isolated from meadowfoam seeds and a full length clone of coconut LPAAT was isolated using a cDNA library of coconut endosperm (Lassner *et al.*, 1995; Knutzon *et al.*, 1995). Sequence identity among plant LPAATs are about 57%, but that increases to 66% in the central region (Frentzen

& Wolter, 1998).

A LPAAT knockout mutant of *Arabidopsis thaliana* is embryo-lethal in the homozygous state (Kim & Huang, 2004). The gene identified in the mutant, named *LPAAT1*, represents the plastidial form of the enzyme (Kim & Huang, 2004). A similar study showed that the lethality occurs during the transition from the globular to heart stage in embryo development, which is the period of chloroplast formation (Yu *et al.*, 2004). This LPAAT, named *ATS2*, was demonstrated to play a role in the formation of plastid membranes (Yu *et al.*, 2004).

Site-directed mutagenesis of LPAAT isolated from *Brassica napus* resulted in loss of function when an alanine replaced a putative catalytic histidine, as well as when substitutions occurred in the central core (Maisonneuve *et al.*, 2000). The N-terminus of the protein was essential for optimal enzymatic activity – without the N-terminus enzyme activity was severely impaired (Maisonneuve *et al.*, 2000). There is a motif located in the C-terminus that, when truncated, renders the protein non-functional, indicating that it must be essential to protein function either for activity or structure (Maisonneuve *et al.*, 2000).

When soybeans were transformed with a yeast gene, *SLC1*, that encodes LPAAT activity, the total oil concentration in the seed increases and the concentration of protein decreases (Rao & Hildebrand, 2009). The same result was found when *Arabidopsis thaliana* and *Brassica napus* cv Hero were transformed with the mutant form of *SLC1* – namely *SLC1-1* (Zou *et al.*, 1997). It was found that more very-long-chain fatty acids are directed to the *sn*-2 position

in the transformed plants (Zou *et al.*, 1997). In a different study, a gene encoding a LPAAT from meadowfoam was inserted into a high-erucic acid line of *Brassica napus* and it was found that erucic acid is incorporated at the *sn-2* position whereas it is not in the control plants (Lassner *et al.*, 1995).

2.3.5 Phosphatidic Acid Phosphatase (PAP)

Phosphatidic acid phosphatase (PAP) catalyzes the third reaction in the Kennedy Pathway, where the phosphate group is removed from the *sn-3* position. PAP has been shown to be associated with the inner envelope of the chloroplast, the cytosol, and the endoplasmic reticulum (Joyard & Douce, 1979; Perry *et al.*, 1999).

The localization of PAP dictates the conditions under which it has proper activity. Chloroplast PAP from spinach prefers pH 9.0, it is inhibited by Mg^{2+}, and it possesses different biochemical properties than the extraplastidial form; for example, di- and tri-phosphates have no effect on plastidial PAP activity, while di-phosphates stimulate and tri-phosphates inhibit extraplastidial PAP (Block *et al.*, 1983; Malherbe *et al.*, 1995). Three *Arabidopsis* isoforms of this PAP have been identified (Nakamura *et al.*, 2007).

Cytosolic PAP from *Brassica napus* prefers pH 5.0, whereas ER PAP from safflower prefers pH of 6-7; both PAPs require Mg^{2+} for activity (Weselake, 2005; Ichihara *et al.*, 1989). It is thought that during times of high triacylglycerol (TAG) accumulation in seeds, PAP translocates from the cytosol to the ER and that this may be a rate limiting step in TAG biosynthesis in developing safflower seeds

(Ichihara *et al.*, 1990).

A study of *Ricinus communis* identified four putative PAP genes *in silico*, and it was found that all of these, along with numerous others from other species, have six transmembrane domains and three conserved phosphatase sequence motifs (Cagliari *et al.*, 2010).

2.3.6 Diacylglycerol Acyltransferase 1 (DGAT1)

Diacylglycerol acyltransferase 1 (DGAT1) catalyzes the fourth and final reaction in the Kennedy Pathway. DGAT1 was associated with the endoplasmic reticulum in seeds, and chloroplast envelopes in leaves (Martin & Wilson, 1984; Kaup *et al.*, 2002; Zhang *et al.*, 2005; McNaughton, 2007). The current understanding is that there are three DGAT1s in soybean: DGAT1a, DGAT1b, and DGAT1c (www.uky.edu/SoyLipidGenes).

DGAT1 was first cloned from mouse DNA and was originally thought to encode acyl CoA:cholesterol acyltransferase (ACAT), but was subsequently shown to be DGAT1 (Cases *et al.*, 1998). Using the mouse sequence, a homologue was sequenced from *Arabidopsis thaliana* (*AtDGAT1*) (Hobbs *et al.*, 1999). *AtDGAT1* was found to be located on the upper arm of Chromosome II, with an abnormally long 5' untranslated region, it contains 15 introns, and it encodes a 530 amino acid long protein (Hobbs *et al.*, 1999). This protein is predicted to have 9 to 10 transmembrane domains and is shown to have a N-terminal region that plays a role in interacting with the acyl-CoA that the fatty acid is attached to for transport (Weselake *et al.*, 2000; Lung & Weselake, 2006). A

mutant created by Katavic *et al.* (1995) was studied by two other groups who traced the mutation to a locus on the upper arm of Chromosome II, termed the *TRIACYGLYCEROL1* (*TAG1*) locus (Zou *et al.*, 1999; Routaboul *et al.*, 1999). *TAG1* was found to be expressed in the leaves, stems, embryos, and flowers of *Brassica napus* (Hobbs *et al.*, 1999). The expression of DGAT1 in *Arabidopsis thaliana* was tissue specific, as well as having a developmentally-dependent pattern of expression (Lu *et al.*, 2003). Soybean DGAT1 (*GmDGAT1*), which was cloned using the *AtDGAT1* sequence, was found to be 1880 bp long and codes for a 498 amino acid protein (Wang *et al.*, 2006). Comparison of sequences from several soybean accessions revealed that the nucleotide variation was large with most of that variation being found in the introns (Wang *et al.*, 2006). The sequence variation among plant species appears to be mainly in the N-terminal regions (Lung & Weselake, 2006). A developmental expression study including soybean DGAT1 showed that DGAT1 expression is highest at 45 days after flowering (DAF), with a substantial increase occurring between 25 and 35 DAF, and a substantial decrease in expression occurring between 55 and 65 DAF (Li *et al.*, 2010). This increase in DGAT1 expression coincided with oil accumulation (Li *et al.*, 2010). Li *et al.* (2010) found that DGAT1 was the major enzyme in seed oil accumulation across several species.

DGAT1 has been shown to have a preference for linoleic acid, palmitic acid, and oleic acid, with soybean DGAT1 showing the highest preference for linoleic acid (Cao & Huang, 1986; Lung & Weselake, 2006). It was shown by *in vitro* assay that the specificity of DGAT1 changes with temperature; for example,

the preference of canola DGAT1 for oleic acid at room temperature changes to a preference for erucic acid at 40°C (Cao & Huang, 1987). Byers *et al.* (1999) showed that several other factors can increase the level of DGAT1 activity in cell culture, including ATP, Co-enzyme A, and magnesium, along with other unidentified small organic compounds. When bovine serum albumin (BSA) was added to microsomal mixtures containing DGAT from maturing sunflower seeds, oleic acid was incorporated at a higher rate along with overall enzymatic activity (Triki *et al.*, 2000).

A maize QTL associated with high oil and oleic acid concentration is present in ancestral genotypes and possesses phenylalanine inserted at amino acid 469, but this amino acid is missing from modern corn genotypes (Zheng *et al.*, 2008). When the expression of this ancestral form was studied via transformation into modern maize lines, there was a marked increase in both oil and oleic acid concentration (Zheng *et al.*, 2008).

DGAT1 has been shown to be important for total oil concentration, as well as fatty acid profile. This was demonstrated by the silencing of *DGAT1* in tobacco, which resulted in a decline of up to 49% in total seed oil concentration and seed weight (Zhang *et al.*, 2005). When canola was transformed with *DGAT1* from Arabidopsis or from another *B. napus* line, there was an increase in total oil concentration, up to 7% under both greenhouse and field environments (Taylor *et al.*, 2009). Overexpression of *DGAT1* in the seeds of wild-type arabidopsis increases total oil concentration by 10% to 70% (Jako *et al.*, 2001). In castor bean development, changes in DGAT1 activity are positively correlated

with the accumulation of TAG (He *et al.*, 2004). Differences in DGAT activity

were correlated with differences in the total oil concentration in soybean

genotypes (Settlage *et al.*, 1998).

2.3.7 Diacylglycerol Acyltransferase 2 (DGAT2)

Diacylglycerol acyltransferase 2 (DGAT2) catalyzes the same reaction as

DGAT1 but DGAT2 is more important for incorporating unusual fatty acids into

triacylglycerol (Li *et al.*, 2010). DGAT2 has been shown to be localized to a

different subdomain of the ER than DGAT1 (Shockey *et al.*, 2006). It is predicted

that there are five DGAT2 genes in soybeans (www.uky.edu/SoyLipidGenes).

While DGAT1 is predicted to have 9 to 10 transmembrane domains, DGAT2 is

predicted to have only two transmembrane domains (Weselake *et al.*, 2000;

Shockey *et al.*, 2006). When soybean was transformed with DGAT2A from a soil

fungus, total oil content in seeds was increased 1.5%, whiile protein content and

yield were unchanged (Lardizabal *et al.*, 2008). In tung tree, DGAT2 expression

is induced during oil biosynthesis in seeds and has a preference for eleostearic

acid – the main fatty acid in tung oil (Shockey *et al.*, 2006). Across seed

development in castor bean, DGAT2 was more highly expressed than DGAT1

and it is thought that DGAT2 plays a major role in TAG synthesis in castor bean

than DGAT1 (Kroon *et al.*, 2006; Cagliari *et al.*, 2010).

2.3.8 Alternative Pathway Key Enzymes

Phospholipid:diacylglycerol acyltransferase (PDAT) (E.C. 2.3.1.158) uses

phosphatidylcholine and diacylglycerol to create triacylglycerol with no additional acyl-CoA fatty acids (Chapman & Ohlrogge, 2012). There are six predicted PDATs in soybean (www.uky.edu/SoyLipidGenes). This enzyme has been shown to be important for the incorporation of unusual fatty acids, like the epoxy and hydroxy fatty acids (Li *et al*., 2010). *Ricinus communis* (castor bean) PDAT shows a preference for ricinoleic acid and vernolic acid, both of which are unusual fatty acids (Dahlqvist *et al.*, 2000). Castor bean has a PDAT that is thought to be found only in this species, named PDAT1-2, it is specific for ricinoleic acid, and is localized in the ER (Kim *et al.*, 2011). There are several genes encoding PDAT in castor bean and all of these copies are expressed at different times during seed development (Cagliari *et al.*, 2010). In Arabidopsis, PDAT was shown to be most active with fatty acids that contained multiple double bonds, epoxy or hydroxy groups (Ståhl *et al.*, 2004). It was also shown that, in Arabidopsis, PDAT1 had a function that overlapped with DGAT1 and that they are essential for normal pollen and seed development (Zhang *et al.*, 2009).

Phosphatidylcholine:diacylglycerol phosphocholine transferase (PDCT) mediates the fatty acid composition of triacylglycerol by channeling unsaturated fatty acids into triacylglycerol via phosphatidylcholine (Chapman & Ohlrogge, 2012).

2.4 Objectives & Hypotheses

The overall objective of this study was to elucidate the role that the Kennedy Pathway plays in determining the fatty acid profile of soybeans. It was

hypothesized that gene sequences and expression patterns account for differences in seed fatty acid profiles among mutant genotypes – RG2, RG7, RG10, and SV64-53.

The first objective was to compare the genomic sequence of each of the four enzymes (GK, G3PAT, LPAAT, and DGAT1) to differences in fatty acid profiles of the mutant geotypes. It was hypothesized that mutations in the transcribed region of one of the genes encoding each of these four enzymes would lead to changes in the enzymes. It was hypothesized that the changes in the enzymes would differ among the mutant genotypes and would correspond to the differences in the fatty acid profiles.

The second objective was to compare the developmental profiles for G3PAT, LPAAT, and DGAT1 expression and fatty acids as a function of temperature. It was hypothesized that the expression of these genes and the fatty acid profile would differ among genotypes within a temperature, as well as among temperatures within a genotype. It was also hypothesized that these expression differences would correspond to differences seen in the fatty acid profile.

CHAPTER 3:

FATTY ACID COMPOSITION IN DEVELOPING AND MATURE SEEDS OF

FOUR SOYBEAN GENOTYPES WITH ALTERED FATTY ACID PROFILES

GROWN AT DIFFERENT TEMPERATURES

3.1 Abstract

Soybeans contain five main fatty acids: palmitic (16:0); stearic (18:0); oleic (18:1Δ^9); linoleic (18:2$\Delta^{9,12}$); and α-linolenic (18:3$\Delta^{9,12,15}$). Four fatty acid mutants were analyzed in this study – RG2, RG7, RG10, and SV64-53. Fatty acid profiles were analyzed from seeds at maturity grown at a normal (25$^{\text{o}}$C/20$^{\text{o}}$C) temperature, as well as over seed development across three temperatures – low (20$^{\text{o}}$C/15$^{\text{o}}$C), normal (25$^{\text{o}}$C/20$^{\text{o}}$C), and high (30$^{\text{o}}$C/25$^{\text{o}}$C). Differences were found among genotypes grown to maturity, with RG2 being low in palmitic acid, RG7 high in stearic acid, and RG10 low in linolenic acid. Several differences were found in fatty acid accumulation during seed development – both within genotypes across temperatures and within temperatures across genotypes. One of the most readily seen differences is in the accumulation of oleic acid, which began at 25 days after flowering (DAF) in RG2 grown at high temperature, but at 45 DAF when grown at normal temperature. Temperature affected the fatty acid profile during seed development, although not to the same extent across seed development or genotypes. This study confirms previous work on the influence of temperature and development on fatty acid profiles of seeds from the four fatty acid profile mutants utilized.

3.2 Introduction

Soybean oil is composed of five major fatty acids, with a typical profile as follows: palmitic acid (11%), stearic acid (4%), oleic acid (25%), linoleic acid (52%), and α-linolenic acid (8%) (Fehr, 2007).   This relative composition can be altered through various methods, including traditional breeding, biotechnology, and mutagenesis. Studies have investigated underlying genetic causes of altered composition, as well as the impact of environmental conditions. Manipulation of fatty acid composition has been the goal of breeding to address many applications. Each fatty acid has unique characteristics lending themselves to these applications; some are positive characteristics, whereas others are negative. The main characteristics of each fatty acid are described below.

Palmitic acid (16:0) is harmful to human health, and thus a reduction in its level is beneficial for human health (Heart & Stroke, 2008). Seven different loci have been identified that contribute to the alteration of this fatty acid have been identified (Erickson *et al.*, 1988; Schnebly *et al.*, 1994; Narvel *et al.*, 2000; Stoltzfus *et al.*, 2000a; Stoltzfus *et al.*, 2000b). Some of these loci are associated with elevated levels of palmitic acid, whereas others are associated with decreased levels. Several of these loci have multiple alleles.

Stearic acid (18:0) is a saturated fatty acid that is neutral to human health, but it is usually lowered in combination with palmitic acid to decrease the overall levels of saturated fatty acids (Fehr, 2007). However, by increasing stearic acid the need for hydrogenating soy oil to create margarines and shortenings is

decreased, which leads to decreased amounts of *trans*-fat in these products (List *et al.*, 1997). Only two loci have been identified that are associated with increased stearic acid concentrations, one of which has numerous alleles (Graef *et al.*, 1985; Bubeck *et al.*, 1989; Pantalone *et al.*, 2002).

Oleic acid ($18:1\Delta^9$) is an unsaturated fatty acid with high levels of oxidative stability (Fehr, 2007). Increasing levels of this fatty acid has been the focus of research, and it has been increased to over 90% (Buhr *et al.*, 2002). One locus, with three known alleles, has been identified for control of the desaturase enzyme that converts oleic acid to linoleic acid (Rahman *et al.*, 1996).

Linoleic acid ($18:2\Delta^{9,12}$) has not been well studied since it naturally occurs at high levels and therefore it lacks the need to be modified (Pantalone *et al.*, 2004).

A decrease in linolenic acid ($18:3\Delta^{9,12,15}$) is the focus of many breeding efforts. It is linolenic acid that is responsible for the off-flavours and odours in food products (Dutton *et al.*, 1951; Primomo *et al.*, 2002). Four loci, all with at least two alleles, have been identified that are associated with reduced levels of linolenic acid (Wilcox & Cavins, 1985; Wilcox & Cavins, 1987; Fehr *et al.*, 1992; Rahman & Takagi, 1997; Fehr & Hammond, 2000).

Not only does genetics influence the levels of fatty acids present in soybean oil, but air temperature during the seed filling period also has an influential role. For example, when temperature increases, the level of polyunsaturated fatty acids decreases, with the level of linolenic acid being the most affected (Dornobos Jr. & Mullen, 1992; Hou *et al.*, 2006). It has also been

35

shown that linoleic and linolenic acids decrease with increasing temperatures but oleic acid increased (Wolf *et al.*, 1982). Since temperature has an influence on fatty acids, this study used controlled environments (growth room and growth chamber) for growing the genotypes used.

3.2.1 Objectives & Hypotheses

The objective of this Chapter was to determine the fatty acid composition of mature seeds from four soybean genotypes with altered fatty acid profiles under "normal" temperature to compare to differences in the genomic transcribed region of the genes of the enzymes of the Kennedy Pathway. The second objective of the Chapter was to determine the fatty acid composition of developing seeds from four soybean genotypes with altered fatty acid compositions grown under three different growing temperatures to compare to the gene expression of three Kennedy Pathway genes.

It was hypothesized that the four soybean mutant lines have altered fatty acid profiles, which is further modified when the lines are grown in different temperature regimes.

3.3 Materials and Methods

3.3.1 Fatty Acid Analysis of Mature Seed

3.3.1.1 Genetic material

Four genotypes selected based on their fatty acid profiles were used in this study – RG2, RG7, RG10, and SV64-53 (Table 3.1). All four genotypes were created at the University of Guelph. RG2 came from a cross between CLP-1 (Stojsin *et al.*, 1998b) and ELLP-2 (Stojsin *et al.*, 1998a), RG7 was created through the chemical mutagenesis (ethyl methane sulfonate (EMS)) of Elgin 87 (Stojsin *et al.*, 1998a), RG10 was the result of the EMS mutagenesis (Stojsin *et al.*, 1998b) of C1640 (Wilcox & Cavins, 1987) , and SV64-53 was derived from a cross between ELLP-3 (Stojsin *et al.*, 1998a) and C1726 (Erickson *et al.*, 1988). These lines were previously described by Hou (2004) in her doctoral thesis and were selected based upon the fatty acid results from that study. These lines were identified in that study as having unique fatty acid profiles that would be useful to the bioproducts industry. Hou (2004) grew these lines under field conditions.

3.3.1.2 Growing conditions and sample collection

Two plants per pot were grown in 6 L, No. 2 nursery container pots (Kord Products, Toronto, ON), in LA4 Sunshine Professional Growing Mix (SunGro Horticulture Canada Ltd., Seba Beach, AB). The seeds were coated with HiCoat peat Bradyrhizobia (Becker Underwood for head-to-head trials, Saskatoon, SK) just prior to hand planting to encourage nodulation. The pots were arranged in a

Table 3.1: Origins and alleles of the four soybean genotypes used.

Genotype	Origins	Alleles	Reference
RG2	CLP-1 X ELLP-2	*fap1*, *fap3*	Stojšin *et al.*, 1998a; Stojšin *et al.*, 1998b
RG7	Elgin 87 mutation	Unknown, high stearic	Stojšin *et al.*, 1998a
RG10	C1640 mutation	*fan-b*	Wilcox & Cavins, 1987; Stojšin *et al.*, 1998b
SV64-53	ELLP-3 X C1726	*fap1*, *fap3*	Erickson *et al.*, 1988; Stojšin *et al.*, 1998a

completely randomized design (CRD), with three reps, each rep containing one pot per line. They were grown in a growth room at 25°C day/20°C night, 16 hr day/8 hr night, 80% humidity with drip irrigation (water only). These growing conditions are used on a regular basis in the soybean breeding program at the University of Guelph. The growth room was lit with both fluorescent lamps (Sylvania F96T12/CW/VHO 215 watt bulbs) and incandescent bulbs (Sylvania 40 watt, clear A19 bulb) with an intensity of 270 μmol/m^2/s (measured at 30 cm from lamps). Biological pest controls were used as needed (BioBest Biological Systems, Westerlo, Belgium). To ensure uniformity across pots, no fertilizer was used. Ten leaf punches, approximately 0.7cm in diameter, per plant were taken using a handheld hole punch adapted for leaf punches and then frozen overnight at -20°C. The leaf punches were freeze-dried (Savant Modulyo D, Savant Instruments Inc., Holbrook, NY) for 24 hrs, and then stored at -20°C. The plants were grown to maturity and the pods were collected from each pot and threshed by hand within each pot.

3.3.1.3 Total oil and fatty acid analysis

A 5 g subsample of seeds from each soybean line was dried for 24 hrs in an oven (Cenco Instruments Corporation, U.S.A.) at 110°C, and then analyzed for total oil concentration using nuclear magnetic resonance (NMR) according to the manufacturer's instructions (The Minispec mq10 NMR Analyzer, Bruker v.2.2; Bruker Corporation, Milton, ON). The instrument calibration was previously developed internally by the soybean breeding lab (Samii-Saket *et al.*, 2011).

For the determination of fatty acid composition, the dried sample of seed used for total oil analysis was first ground to a fine powder in a household Lancaster coffee grinder. The powdered samples were stored in a cool, dry, dark environment overnight until fatty acid extraction was performed. Fatty acid extraction was performed in a similar manner to that used by Primomo *et al.* (2002) with a few changes.   Approximately 100 mg of powdered sample was transferred to a 10 mL screw-top tube, and 0.6 mL 0.25 M KOH (Caledon, Georgetown, ON) in 1:1 methanol (HPLC grade; Fisher Scientific, Waltham, MA):diethyl ether (anhydrous; Fisher Scientific, Waltham, MA) was added. The samples were vortexed (Maxi Mix 1 Type 16700 Mixer, Thermolyne, Dubuque, IA) for 15 s and then placed in a 60°C waterbath (Precision Waterbath, GCA Corp., Chicago, IL) for 1 min. The samples were removed from the waterbath and allowed to cool on the counter for approximately 10 min. After cooling completely, 2 mL saturated sodium chloride (certified A.C.S., crystal; Fisher Scientific, Waltham, MA) and 2.5 mL iso-octane (HPLC grade; Fisher Scientific, Waltham, MA) were added to each sample. The samples were shaken vigorously by hand for 10 s. To separate the iso-octane layer and the water layers, the samples were transferred to a tabletop centrifuge (IEC HN-SII Centrifuge, Damon/IEC Division, Needham Heights, MA), and centrifuged for 1 min at approximately 1000 rpm. A 1 mL aliquot was removed from the upper layer and deposited into a 1.8 mL vial (Autosampler Vial, 1.8 mL LOV, Clear, 11 mm, Crimp, VWR, West Chester, PA) for gas chromatography (GC) analysis. Samples were loaded into the auto sampler and analyzed using a Hewlett

Packard HP 6890 Series GC System and Hewlett Packard HP 6890Series

Injector (Wilmington, DE), containing a DB-23 column (15 m length, 0.25 mm

internal diameter, 0.25 μm film; Agilent Technologies, Inc., Santa Clara, CA) with

one 2 μL injection per vial. The data were obtained and analyzed using Agilent

ChemStation Rev.B.04.02 (Agilent Technologies, Inc., Santa Clara, CA).

3.3.1.4 Statistical analysis

Percentage of fatty acid was transformed to g/kg using the following

equation: % FA X 10 = FA (g/kg). Using SAS v.9.1, an analysis of variance

(ANOVA) using PROC GLM was used for each fatty acid. Least squares means

(LSmeans) and standard errors (SE; SAS coding = stderr) were also calculated.

All possible pair-wise comparisons were made using least significant difference

(pdiff), and using the Scheffé method of adjustments (adjust=scheffe). A Type I

error of α = 0.05 was used.

3.3.2 Fatty Acid Analysis of Developing Seeds

3.3.2.1 Genetic material

The four genotypes used were described and grown as above. There

were two plants per pot, with nine pots per line. Three pots per line were

transferred to one of three growth chambers 8 to 10 d after emergence. The

plants were grown simultaneously in three separate chambers at one of three

temperatures: 30°C day/25°C night (Conviron E8 CMP3244; Controlled

Environments Ltd., Winnipeg, MB), 25°C day/20°C night (Conviron E8M Compact

"6"; Controlled Environments Ltd., Winnipeg, MB), 20°C day/15°C night (Conviron

E8M; Controlled Environments Ltd., Winnipeg, MB). These temperatures were chosen based on previous literature describing the influence of temperature on fatty acid profiles in mature seeds. All chambers had a 16 hr day/8 hr night, 80% humidity, and plants were hand watered (water only) as needed. All chambers were lit with fluorescent lamps (Sylvania F48T12/CW/VHO, 115 watt bulbs) and incandescent bulbs (Sylvania 40 watt, Clear A19) with an intensity of 200 $\mu mol/m^2/s$ (measured at 30 cm from lamps). The pots were arranged in reps, with three reps per chamber. The chambers were randomized individually, with one pot per line per rep. Biological pest controls were used as needed (BioBest Biological Systems, Westerlo, Belgium). To ensure uniformity among pots, no fertilizer was used. Individual flowers were tagged as they emerged over a three month period. Developing pods were then collected in the afternoon at 15, 25, 35, 45, 55, and 65 days after flowering (DAF), immediately snap-frozen in liquid nitrogen, and stored as whole pods at -80^oC. These time points were previously used by Li *et al.* (2010) in a similar study of gene expression and fatty acid accumulation. Plants were not grown to maturity. Li *et al.* (2010) showed that the gene expression dropped off at 65 DAF, and it was felt that this was an appropriate point to end the study.

3.3.2.2 Fatty acid analysis

Up to 100 mg of fine-ground sample remaining after RNA isolation (Chapter 5) was transferred to a 1.5 mL snap-cap microfuge tube (FisherBrand, Fisher Scientific, Waltham, MA). Samples were freeze-dried (Savant Modulyo D,

Savant Instruments Inc., Holbrook, NY) for 24 h, and then stored in a cool, dry,

dark location overnight until fatty acid analysis. Fatty acid extraction and analysis

were performed as described above (3.3.1.3).

3.3.2.3 Statistical analysis

Statistical analyses were performed as described above (3.3.1.4),

including pair-wise comparisons. They also included a calculation of Least

Significant Difference (LSD) using the arithmetic means in SAS v9.1

(MEANS/LSD lines). The analyses were performed by time point (DAF) to

evaluate genotype, temperature, and their interaction. Analyses were also

performed by genotype to evaluate time point (DAF), temperature, and their

interaction. T-tests were performed to compare fatty acid profiles at 65 DAF to

those at maturity (3.3.1), both grown under normal temperature, using Students t-

test free.

3.4 Results

3.4.1 Fatty Acid Composition of Mature Seed from Four Genotypes Grown at Normal Temperature

There was no difference in total oil concentration of mature seeds collected from the four genotypes grown at $25^{o}C/20^{o}C$. However, significant genotype effects were found for four of the five main fatty acids – palmitic, stearic, linoleic, and linolenic (Table 3.2). For palmitic acid, it was found that RG2 differed significantly from all the other genotypes, whereas RG7 significantly differed in stearic acid and RG10 was significantly differed in linolenic acid. RG7 and RG10 were significantly different for linoleic acid. SV64-53 had a fatty acid profile that was close to a normal profile, although it was originally selected for low palmitic acid based on the results of a study done by Hou (2004) using the same genotypes studied here. This deviation may be due to seed source and the line continuing to segregate for palmitic acid content.

3.4.2 Impact of Genotype on the Fatty Acid Composition of Developing Seeds from Plants Grown at Three Temperatures

In this study four genotypes, RG2, RG7, RG10, and SV64-53 were grown at low ($20^{o}C/15^{o}C$), normal ($25^{o}C/20^{o}C$), and high ($30^{o}C/25^{o}C$) temperatures. Genotypes reacted differently when they are compared to each other within a temperature and within a genotype across temperatures (Figs 3.2, 3.3, 3.4, and 3.6). Fatty acid profiles also changed over seed development. There were significant differences across time points for all fatty acids within all genotypes,

Table 3.2: Total oil and fatty acid composition of mature seed from four genotypes grown at 25°C/20°C. Data represent the means (± standard deviations (SD)) of three replicate experimental units (2 plants/pot). Significant differences among genotypes for each fatty acid are denoted: * at α = 0.05 level and ** at α = 0.01 level. **†** denotes genotypes with values for fatty acids differing significantly from other genotypes.

Genotype	Total Oil (g/kg) (mean ± SD, range)	Fatty Acid Profile (g/kg) (mean ± SD, range)				
		Palmitic **	Stearic **	Oleic	Linoleic *	Linolenic **
RG2	233 ± 9.0	**42 ± 1.9†**	22 ± 0.8	255 ± 19.3	594 ± 16.2	88 ± 0.9
	219 to 240	**41 to 44**	21 to 22	235 to 273	578 to 610	87 to 89
RG7	237 ± 4.7	102 ± 5.4	**57 ± 9.6†**	247 ± 34.8	**516 ± 38.0†**	79 ± 7.4
	232 to 243	98 to 106	**50 to 63**	222 to 271	**489 to 542**	74 to 85
RG10	228 ± 6.8	107 ± 1.3	27 ± 1.0	231 ± 17.9	**604 ± 36.4†**	**31 ± 17.7†**
	222 to 238	106 to 109	26 to 28	215 to 250	**564 to 629**	**20 to 52**
SV64-53	225 ± 14.2	104 ± 2.3	25 ± 1.5	258 ± 11.2	551 ± 6.7	63 ± 1.8
	206 to 238	101 to 105	24 to 27	250 to 271	544 to 558	61 to 64

45

Figure 3.1: Comparison of the fatty acid composition in mature seed from four altered fatty acid soybean mutants with typical values from the literature. Average fatty acid composition, with standard errors, is shown for each genotype. The means represent three biological replicates. "Normal" shows the typical fatty acid profile found in *Glycine max* (L.) Merr. according to Fehr (2007). Statistical comparisons to the "normal" profile were not possible.

with the exception of two: stearic and oleic acids in RG7 (Table 3.3). General trends are discussed first followed by specific statistical comparisons for each genotype.

3.4.2.1 RG2

Generally, the initial levels of fatty acids were similar at normal and high temperatures, but they were at different initial levels at low temperature (Fig. 3.2). The initial level of linolenic acid at the low temperature was higher than at other temperatures, and levels declined much earlier at the high temperature. It should be noted that the level of oleic acid increased at 25 DAF at high temperature and it continued at 35 DAF; this trend was not seen at the other temperatures. Concommitantly, there was a slight decrease in linoleic acid, demonstrating an inverse relationship between the two fatty acids. There also appeared to be a negative relationship between oleic and linolenic acids under normal temperature from 45 DAF to 65 DAF. Palmitic and stearic acids were at similar levels across all temperatures.

For RG2, the effect of temperature was only significant for oleic and linolenic acids and there were no significant interactions between temperature and time point (Table 3.3). Pair-wise comparisons for temperature (Table 3.4) showed that oleic acid was significantly higher at high temperature than low and normal temperatures, whereas linolenic acid was significantly lower at high temperature than at low and normal temperatures (Table 3.4). Pair-wise comparisons were performed for time point (Table 3.5) and it was found that at

Table 3.3: Analysis of Variance of the fatty acid composition of developing seed from four genotypes grown at three temperatures. Analysis of variance (ANOVA) Mean Squares (MS) are based on three different temperatures (Low, Normal, High) and six different time points (15, 25, 35, 45, 55, & 65 days after flowering). df – degrees of freedom; * - significant at $\alpha = 0.05$ level; ** - significant at $\alpha = 0.01$ level.

Genotype	Source	df	Palmitic	Stearic	Oleic	Linoleic	Linolenic
RG2	Temp (T)	2	170.3	67.1	34238.1**	3614.5	14026.9**
	Time (D)	5	4380.5**	3651.1**	11502.7*	64264.2**	30439.3**
	T X D	10	96.6	131.9	5424.0	832.2	1647.1
	Error	30	174.4	208.0	4582.1	1736.5	1980.2
RG7	Temp (T)	2	47.0	11297.8*	3178.7	474.3	8274.0**
	Time (D)	5	2840.4**	1211.4	1831.1	32516.1**	19740.3**
	T X D	9	93.7	1927.9	2210.9	3937.1	266.9
	Error	23	24.6	3049.4	2609.7	2154.2	1042.2
RG10	Temp (T)	2	46.1	202.0*	21233.6**	14903.3**	4253.9**
	Time (D)	5	3265.3**	1874.0**	14991.4**	14923.1**	14891.3**
	T X D	10	118.6**	113.6*	2519.5	3591.6	989.1
	Error	20	26.3	46.2	1586.6	2274.5	745.3
SV64-53	Temp (T)	5	46.0	145.2	10508.9*	1274.8	17106.7**
	Time (D)	5	4382.3**	6526.2**	11622.8**	43372.4**	29168.1**
	T X D	10	271.0	314.1**	2446.2	4549.4*	1426.8
	Error	26	172.7	99.0	2857.1	2079.5	1125.6

Figure 3.2: Fatty acid composition of developing seed in RG2 soybean grown at three temperatures. A – grown at "low" temperature; B – grown at "normal" temperature; C – grown at "high" temperature.

Table 3.4: Pair-wise comparison LSmeans for temperature of significant differences in fatty acid composition of developing seeds of four soybean genotypes grown at three temperatures. LSmeans could not be estimated for RG7 at low temperature due to no samples available for evaluation at 15 DAF. Comparisons are within a column for each genotype. Least Significant Difference (LSD) was based on arithmetic means. LSmeans with the same letter designation are not significantly different.

Genotype	Temperature	Palmitic (g/kg)	Stearic (g/kg)	Oleic (g/kg)	Linoleic (g/kg)	Linolenic (g/kg)
RG2	Low	61.5^a	40.8^a	214.6^a	526.5^a	156.3^a
	Normal	61.9^a	37.4^a	203.3^a	536.3^a	161.1^a
	High	55.9^a	41.6^a	291.5^b	505.1^a	105.9^b
	LSD	41.41	20.69	30.97	45.02	23.29
RG7	Low	non-est	non-est	non-est	non-est	non-est
	Normal	128.0^a	53.1^a	185.0^a	491.4^a	142.5^a
	High	129.0^a	104.1^b	160.0^a	488.7^a	118.2^a
	LSD	12.59	44.77	56.21	37.28	25.93
RG10	Low	136.1^a	42.3^{ab}	157.4^a	569.8^{ab}	94.5^a
	Normal	134.6^a	41.1^a	152.9^a	598.7^a	72.7^{ab}
	High	138.9^a	49.6^b	236.3^b	520.4^b	54.8^b
	LSD	4.31	7.31	33.48	40.01	22.95
SV64-53	Low	134.1^a	46.7^a	165.1^a	471.5^a	181.0^a
	Normal	132.2^a	48.7^a	216.9^b	479.9^a	122.3^b
	High	130.2^a	53.5^a	210.4^{ab}	492.1^a	114.0^b
	LSD	9.50	7.63	40.95	34.93	25.70

Table 3.5: Pair-wise comparison LSmeans for time point of significant differences in fatty acid composition of developing seeds of four soybean genotypes grown at three temperatures. LSmeans could not be estimated for RG7 at 15 DAF due to no samples available for evaluation at 15 DAF grown under low temperature. Comparisons are done within a column for each genotype separately. Least Significant Difference (LSD) was based on arithmetic means. LSmeans with the same letter designation are not significantly different.

Genotype	DAF	Palmitic (g/kg)	Stearic (g/kg)	Oleic (g/kg)	Linoleic (g/kg)	Linolenic (g/kg)
RG2	15	102.5^a	79.7^a	180.0^a	387.4^a	249.6^a
	25	66.7^b	39.0^b	248.8^a	477.3^b	168.2^b
	35	62.3^{bc}	46.4^b	269.8^a	473.9^b	147.6^b
	45	42.6^{bc}	23.9^b	276.9^a	564.5^c	92.0^b
	55	41.6^c	25.4^b	242.6^a	602.0^c	88.4^b
	65	43.0^{bc}	25.0^b	200.8^a	630.7^c	100.7^b
	LSD	58.56	29.26	43.80	63.66	33.08
RG7	15	non-est	non-est	non-est	non-est	non-est
	25	152.7^a	64.0^a	179.4^a	422.4^a	181.6^a
	35	116.9^b	68.4^a	173.9^a	523.9^b	117.0^b
	45	112.3^b	83.3^a	177.4^a	514.0^b	113.0^b
	55	110.0^b	54.1^a	171.7^a	560.5^b	103.7^b
	65	110.2^b	50.6^a	166.8^a	574.4^b	97.9^b
	LSD	18.35	65.25	81.92	54.35	37.80
RG10	15	168.3^a	81.9^a	108.5^a	471.0^a	170.3^a
	25	170.8^a	56.9^b	122.4^a	526.9^a	123.0^a
	35	131.1^b	35.9^c	243.0^b	557.5^a	32.4^b
	45	116.3^c	31.9^c	222.5^b	582.7^b	46.6^b
	55	116.1^c	31.4^c	189.7^a	619.3^b	43.5^b
	65	116.4^c	27.9^c	207.1^a	620.3^b	28.3^b
	LSD	6.25	10.59	48.57	58.15	33.29
SV64-53	15	163.6^a	114.5^a	127.5^a	331.7^a	259.3^a
	25	165.7^a	63.8^b	156.8^a	418.8^a	194.7^a
	35	128.2^b	33.2^c	216.4^a	515.0^b	107.1^b
	45	116.9^b	31.7^c	233.6^a	523.2^b	94.6^b
	55	110.5^b	28.5^c	220.0^a	543.7^b	97.4^b
	65	108.1^b	26.0^c	230.3^a	554.5^b	81.4^b
	LSD	13.59	10.91	58.59	49.98	36.77

15 DAF levels of palmitic, stearic, and linolenic acids were significantly higher than at all other time points, but the level of linoleic acid was significantly lower. For linoleic acid the levels at 15 DAF, 25 DAF and 35 DAF were significantly lower than at all other time points, but 25 DAF and 35 DAF linoleic acid levels were not significantly different from each other. It was also noted that palmitic acid levels were significantly higher at 25 DAF than at 55 DAF.

3.4.2.2 RG7

With RG7 (Fig. 3.3), there were no samples available for 15 DAF at the low temperature. Nevertheless, there appeared to be a similar trend across all temperatures for both linoleic (increasing) and linolenic (decreasing) acids. Oleic acid increased in a similar manner at normal and high temperatures but decreased at later dates at high temperature, in contrast to other temperatures. There was a marked decrease in linoleic acid at 45 DAF without a corresponding increase in any other fatty acid similar to that described above for RG2.

For RG7, the effect of temperature was significant for stearic and linolenic acids, and there were no significant interactions between temperature and time point (Table 3.3). Following pair-wise comparisons, it was found that high temperature resulted in a significantly higher level of stearic acid than that found at normal temperature (Table 3.4). Pair-wise comparisons performed for palmitic, linoleic, and linolenic acids across time points yielded the same result: 25 DAF was significantly different from all the other time points, with palmitic and linolenic acid being higher at 25 DAF and lower for linoleic acid (Table 3.5).
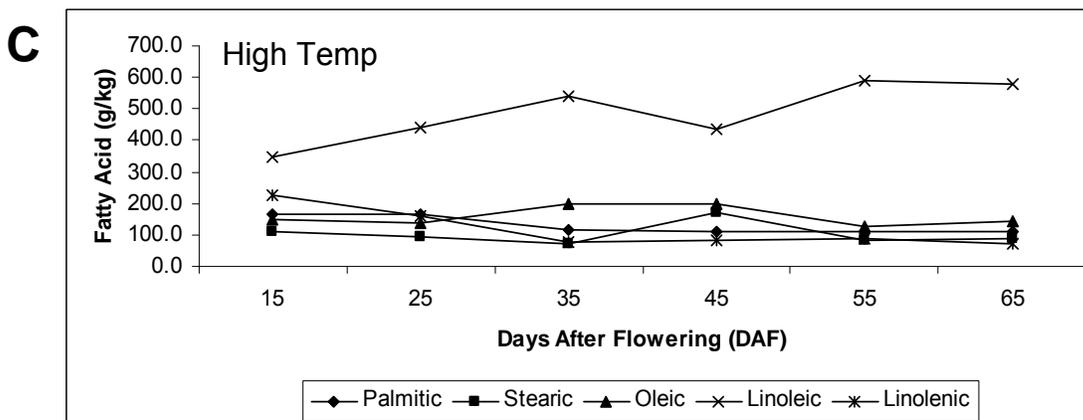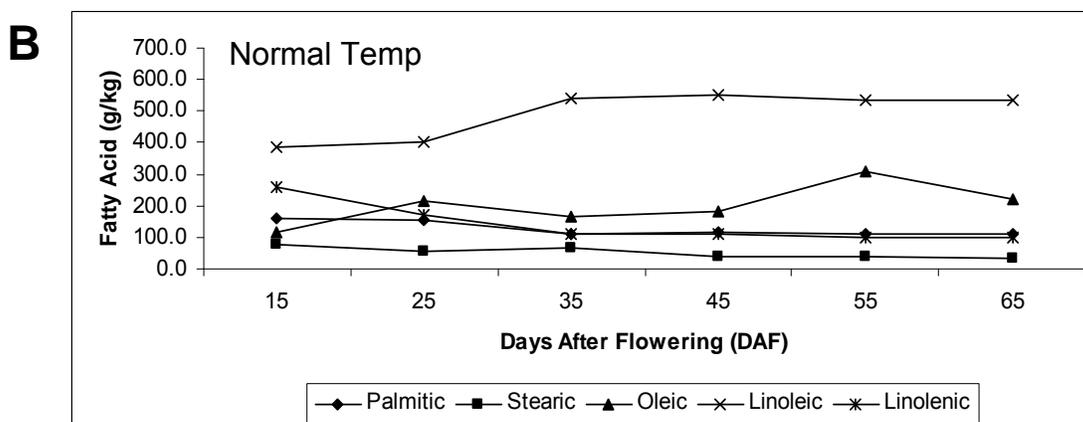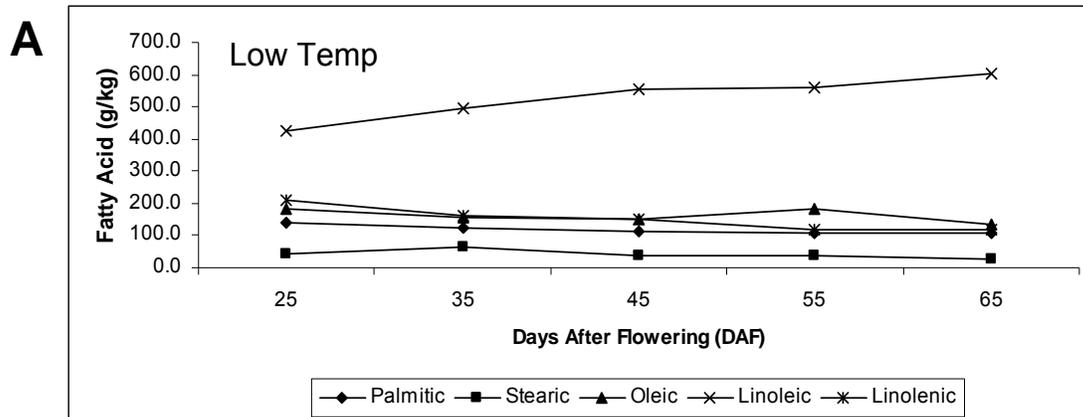
Figure 3.3: Fatty acid composition of developing seed in RG7 soybean grown at three temperatures. A – grown at "low" temperature; B – grown at "normal" temperature; C – grown at "high" temperature.

53

3.4.2.3 RG10

Linoleic acid in RG10 (Fig. 3.4) exhibited a slight upward trend that was the same for all temperatures, although there was a distinct reduction at 35 DAF under high temperature. This decrease in linoleic acid at 35 DAF under high temperature coincided with an increase in oleic acid, again suggesting an inverse relationship such as that seen in RG2. The level of linolenic acid declined earlier in development at the high temperature compared to normal temperature, but it declined steadily over development, whereas at normal temperature the levels declined distinctly between 25 and 35 DAF. At low temperature, linolenic acid levels increased again, albeit slightly, at 45 DAF and continued at the higher level at 55 DAF, which was not the case at the other temperatures.

The effect of time point was significant across all fatty acids for RG10 and pair-wise comparisons showed that 15 DAF and 25 DAF were significantly higher than all other time points for both palmitic and stearic acids (Table 3.5). Oleic acid levels at 15 and 25 DAF were only significantly lower than 35 and 45 DAF. Linoleic acid levels at 15 DAF was significantly lower than 45, 55, and 65 DAF. Linolenic acid levels at 15 DAF and 25 DAF were significantly higher than at other time points.

The effect of temperature was also significant in RG10 for all fatty acids except palmitic (Table 3.3). Pair-wise comparisons showed that at high temperature levels of stearic and oleic acids were significantly lower than at normal temperature, and levels of linoleic acid at high temperature were significantly higher than at normal temperature (Table 3.4). High temperature
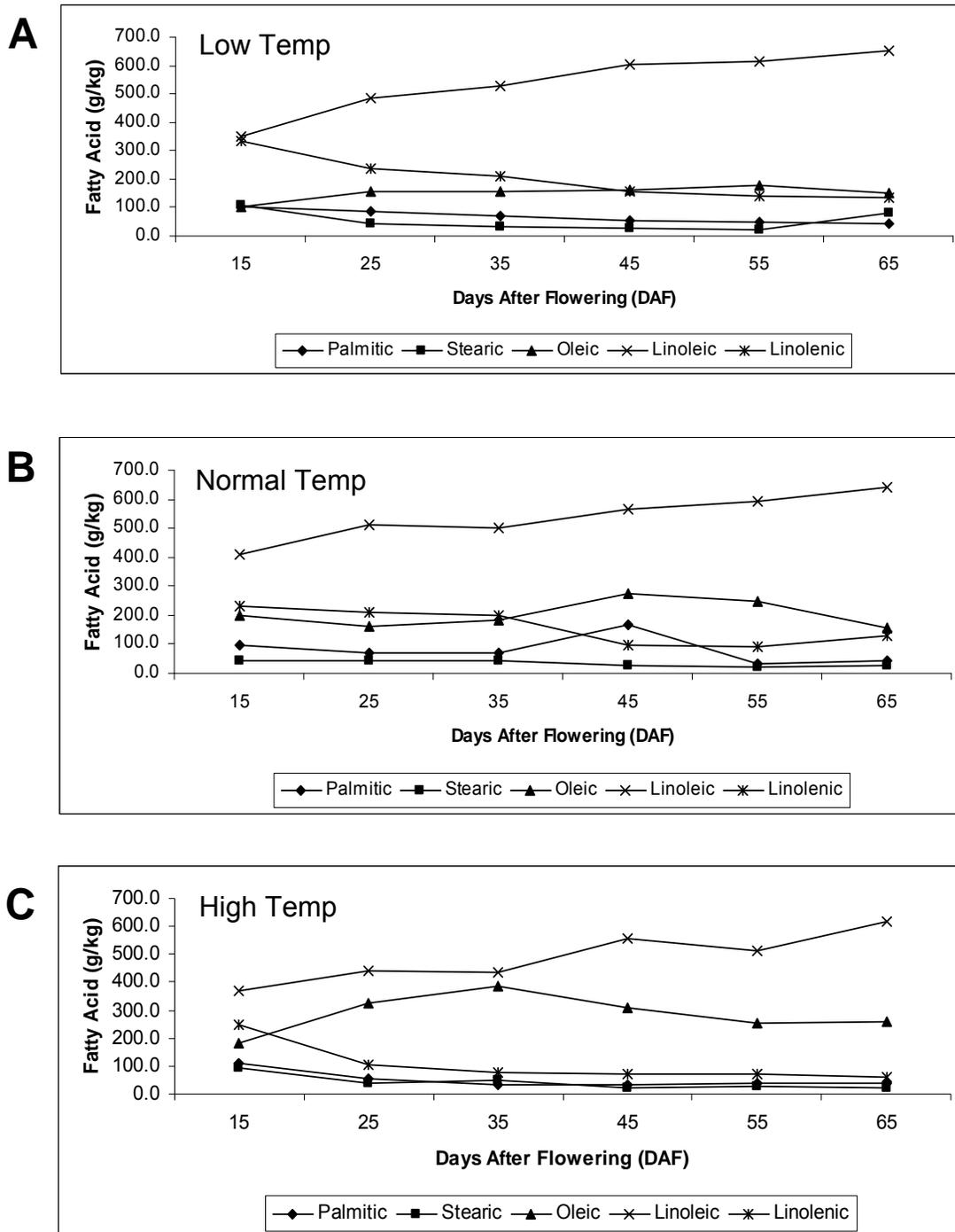
54
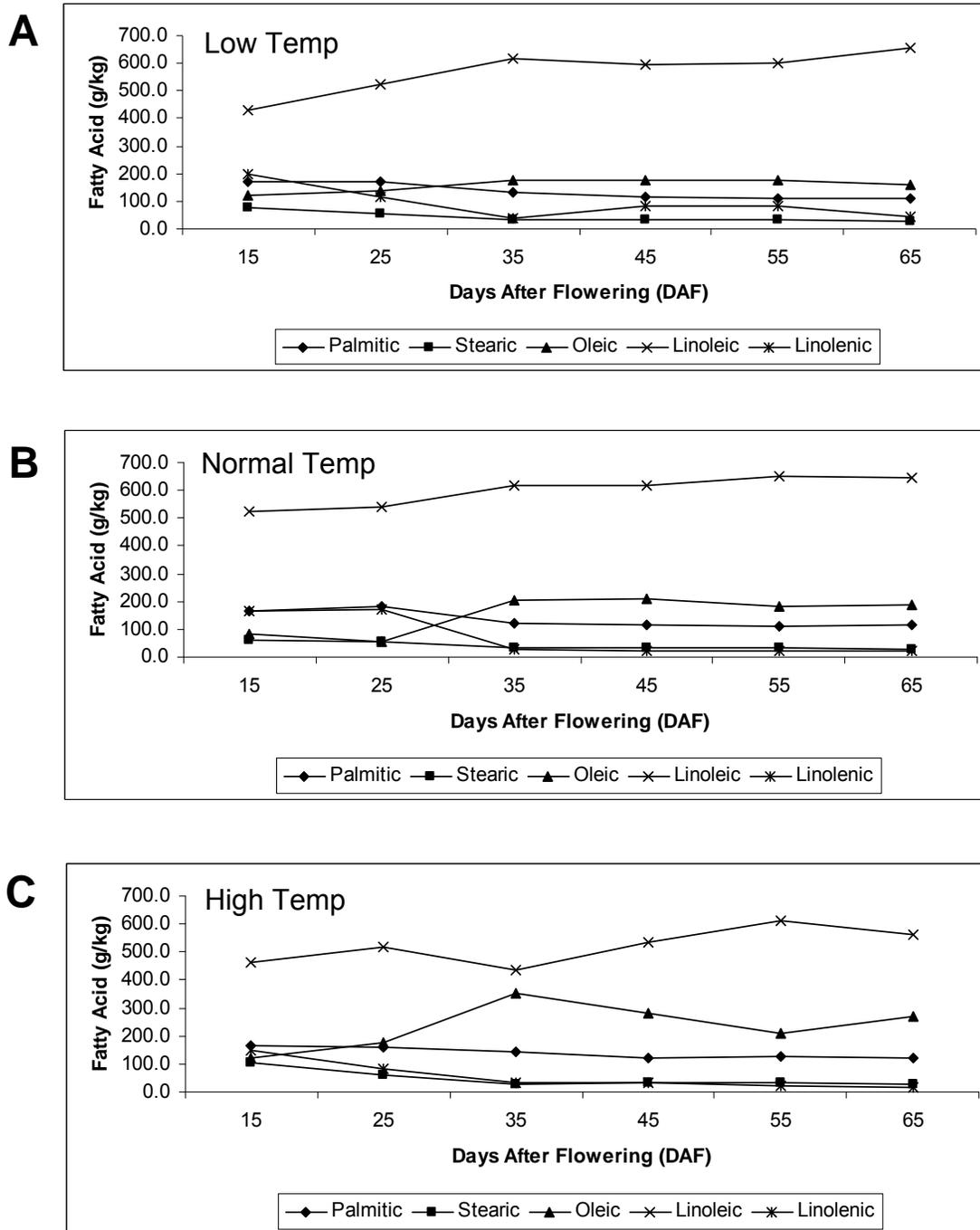
Figure 3.4: Fatty acid composition of developing seed in RG10 soybean grown at three temperatures. A – grown at "low" temperature; B – grown at "normal" temperature; C – grown at "high" temperature.

produced significantly higher levels of oleic acid and lower levels of linolenic acid than at low temperature.

There were significant temperature by time point interactions for both palmitic and stearic acids (Table 3.3). The pair-wise comparisons for palmitic acid showed a similar trend across all temperatures in that it decreased throughout development, although the decline occurred sooner at low temperature than the other temperatures and the levels at high temperature remained higher than at the other temperatures as they declined (Fig. 3.5A). The pair-wise comparisons for stearic acid also showed a decrease over development with the most drastic being at high temperature (Fig. 3.5B).

3.4.2.4 SV64-53

For SV64-53 (Fig. 3.6), the inverse relationship between linoleic and oleic was seen at normal temperature from 45 DAF through 65 DAF. At low temperature, the increase in linoleic was steady when compared to high temperature where there was a sudden increase in the linoleic level between 25 and 35 DAF followed by a plateau. At high temperature, the levels of stearic acid declined at 35 DAF, whereas at both low and normal temperatures it declined at 25 DAF. At low temperature, the levels of linolenic acid decreased steadily but declined dramatically under the other temperatures; 25 DAF at normal temperature; and 35 DAF at high temperature. It should also be noted that the initial level of linolenic acid was lower at high temperature than low and normal temperatures, which had similar initial levels.
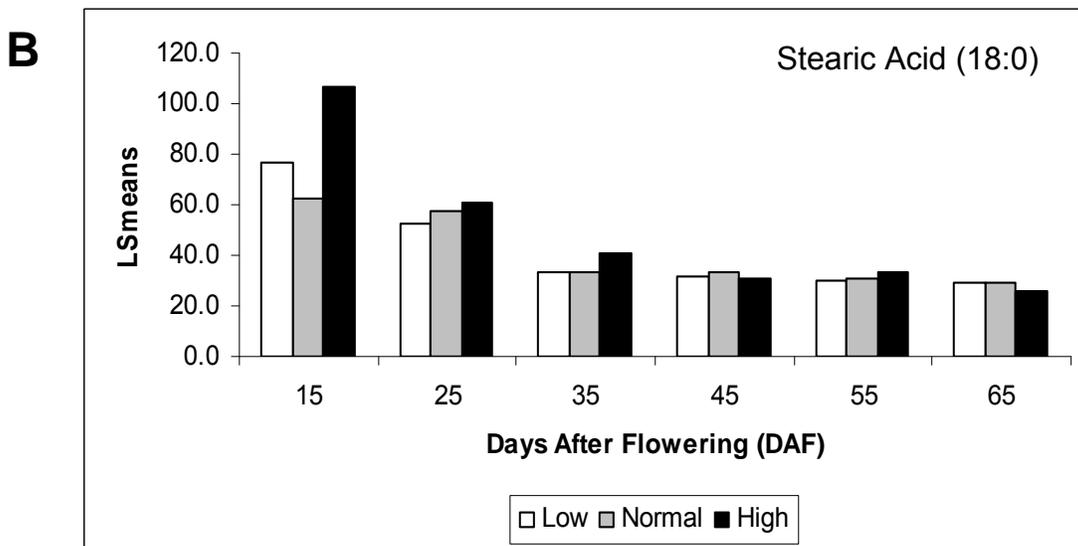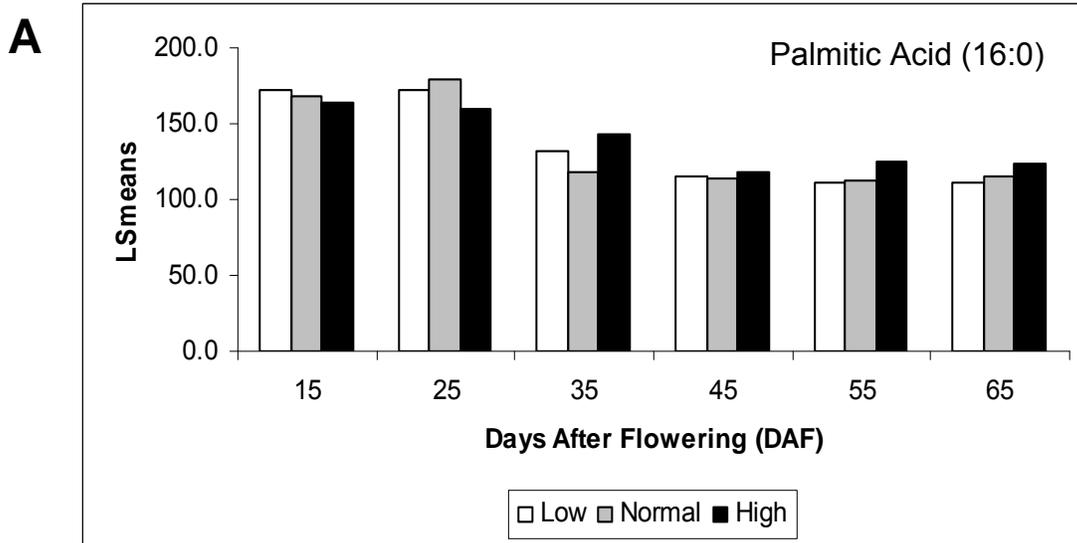
Figure 3.5: LSmeans of significant interactions between time point and growing temperature developing seeds of RG10 soybean. Interaction between time point and growing temperature in RG10 was significant for palmitic and stearic acids. A – LSmeans for interaction between time point and growing temperature for palmitic acid. B – LSmeans for interaction between time point and growing temperature for stearic acid.
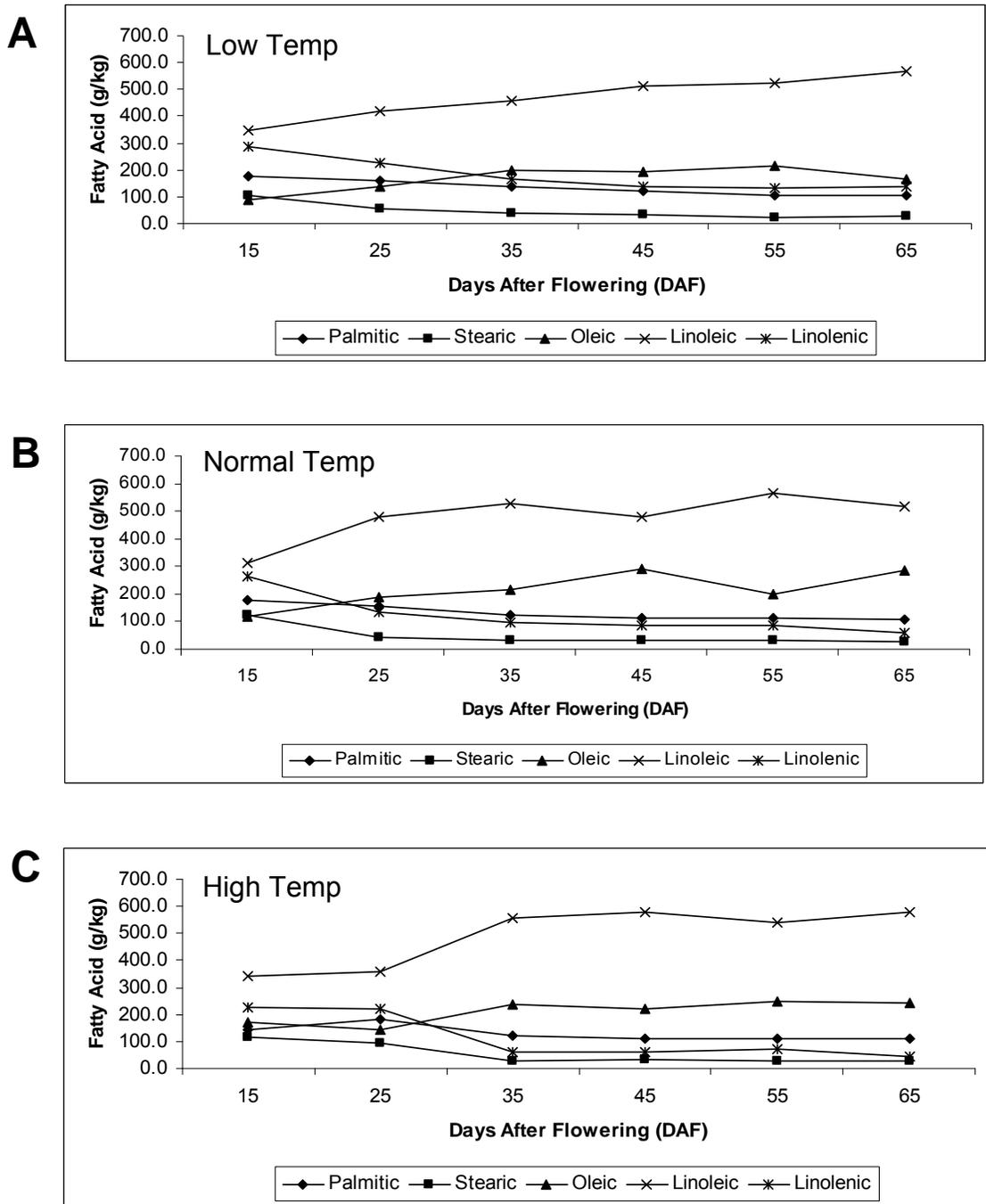
Figure 3.6: Fatty acid composition of developing seed in SV64-53 soybean grown at three temperatures. A – grown at "low" temperature; B – grown at "normal" temperature; C – grown at "high" temperature.

SV64-53 showed significant differences in time point across all fatty acids (Table 3.3). Stearic acid pair-wise comparisons for time points showed that both 15 and 25 DAF had significantly higher levels than at all other time points (Table 3.5). For palmitic and linolenic acids, time point comparisons showed that both 15 and 25 DAF had significantly higher levels than at all other time points except 15 DAF compared to 25 DAF, which had similar levels. Both time points had significantly lower levels of linoleic acid than at other time points (Table 3.5).

The temperature effect was significant for oleic and linolenic acids (Table 3.3), with low temperature resulting in significantly lower levels of oleic acid than at normal temperature and low temperature resulting in significantly higher levels of linolenic acid than both normal and high temperatures (Table 3.4). The interaction between temperature and time point was significant for stearic and linolenic acids. When pair-wise comparisons were evaluated, stearic acid decreased over development for all temperatures with the major decline occurring at 25 DAF for both low and normal temperatures, and at 35 DAF for high temperature (Fig. 3.7A). The pair-wise comparisons for linoleic acid showed that its accumulation increased over development in all growing temperatures (Fig. 3.7B).

3.4.3 Impact of Development on the Fatty acid Composition of Seeds from Four Genotypes Grown at Three Temperatures

Statistical analyses were performed within time points to better understand at what points in development temperature is critical to fatty acid accumulation

A Stearic Acid (18:0)

B Linoleic Acid (18:1)
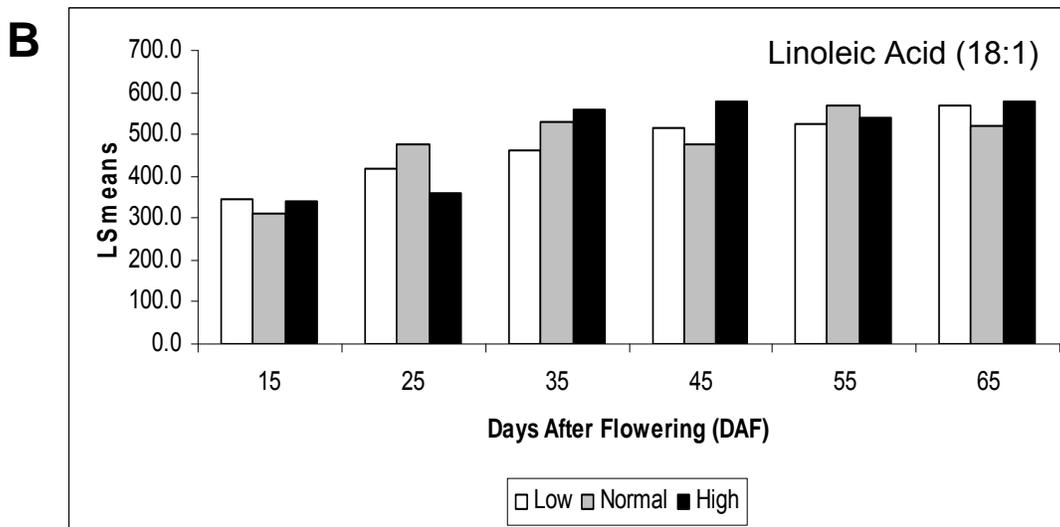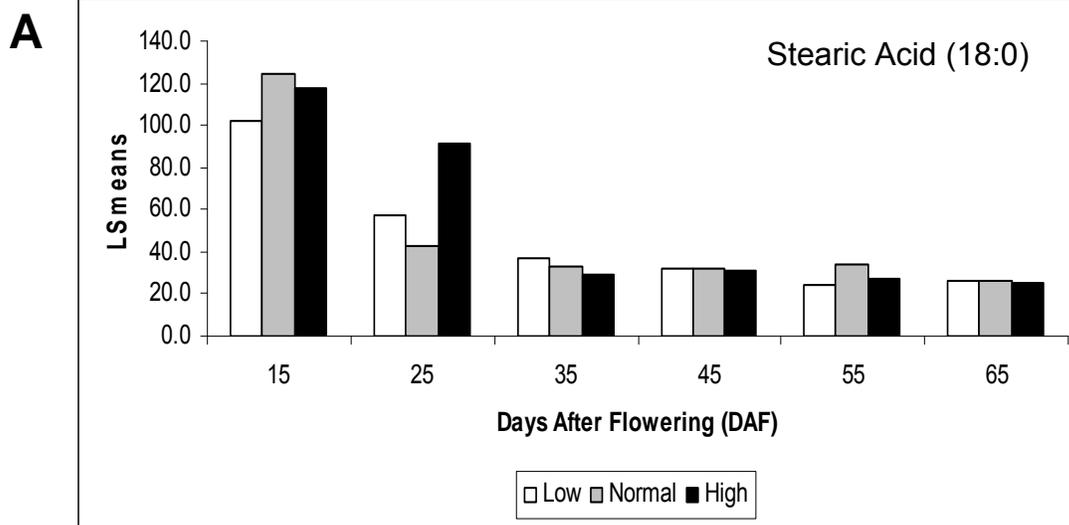
Figure 3.7: LSmeans of significant interactions between time point and growing temperature for fatty acid composition of SV64-53. Interaction between time point and growing temperature in SV64-53 was significant for stearic and linoleic acids. A – LSmeans for interaction between time point and growing temperature for stearic acid. B – LSmeans for interaction between time point and growing temperature for linoleic acid.

Table 3.6: Analysis of Variance of fatty acid profiles of developing seed from four soybean genotypes grown under different temperatures. Data respresent the Mean Squares (MS) of the fatty acid profiles at six developmental time points of four genotypes (RG2, RG7, RG10, SV64-53) grown under low, normal, and high temperatures. DAF – days after flowering; df – degrees of freedom; * - significant at α = 0.05 level; ** - significant at α = 0.01 level.

DAF	Source	df	Palmitic	Stearic	Oleic	Linoleic	Linolenic
15	Geno (G)	3	6283.6**	1358.0*	5720.3	14931.8**	7228.7*
	Temp (T)	2	78.9	1134.8	2271.3	1762.8	3170.1
	G X T	5	215.1	298.4	1307.4	1170.5	421.8
	Error	12	485.3	383.6	1805.8	2057.6	1924.0
25	Geno (G)	3	13041.3**	772.1	14810.7*	13604.6**	5093.2
	Temp (T)	2	61.4	1138.2*	4244.2	2661.6	3717.6
	G X T	6	370.7	454.0	7556.3	2035.4	4021.6
	Error	13	322.0	270.3	4064.5	1978.1	2804.3
35	Geno (G)	3	7960.1**	1977.3*	12886.6	9131.5*	18417.2**
	Temp (T)	2	258.4	25.8	30166.8**	7650.4	12356.6**
	G X T	6	282.2	54.9	5903.9	9654.3**	2989.3
	Error	20	188.1	617.5	4515.0	2881.0	1629.4
45	Geno (G)	3	11523.9**	6598.1	15028.3*	8887.4*	6444.5**
	Temp (T)	2	31.5	3588.2	11826.8	3095.6	10249.9**
	G X T	6	39.9	4427.1	2484.0	6447.0	513.7
	Error	22	74.3	2308.7	3793.4	2800.4	553.4
55	Geno (G)	3	10971.1**	1368.1**	8279.7**	10353.1**	5814.1**
	Temp (T)	2	41.7	566.2*	568.3	876.2	5982.3**
	G X T	6	52.1**	435.3**	2312.9	1527.6	469.7
	Error	22	13.7	129.4	1649.2	1100.0	740.3
65	Geno (G)	3	6727.4**	618.7	2652.4	6148.2**	4810.7**
	Temp (T)	2	21.3	313.7	7268.0*	1903.4	4120.4*
	G X T	6	23.2	588.1	4286.8	1579.7	868.9
	Error	10	19.5	577.1	1895.5	736.1	695.5

and the same can be said for genotype (Table 3.6). As mentioned above, there were no samples of RG7 at 15 DAF for fatty acid analysis.

3.4.3.1   15 Days After Flowering

At 15 DAF, the effect of genotype was significant for all fatty acids except oleic acid and pair-wise comparisons were performed (Table 3.6). These comparisons showed that RG2 had significantly lower levels of palmitic acid than RG10 and SV64-53 and significantly lower levels of stearic acid than SV64-53 (Table 3.7). RG10 had significantly higher levels of linoleic acid than RG2 and SV64-53 and had significantly less linolenic acid than the other genotypes (Table 3.7).

3.4.3.2   25 Days After Flowering

At 25 DAF, the effect of genotype was significant for palmitic, oleic, and linoleic acids and pair-wise comparisons were performed (Table 3.6). RG2 had significantly less palmitic acid than all other genotypes but significantly higher levels of oleic acid than RG10. RG10 had significantly higher levels of linoleic acid than RG7 and SV64-53 (Table 3.7). The effect of temperature was significant for stearic acid and pair-wise comparisons revealed that high temperature had significantly higher levels of stearic acid than low temperature (Table 3.8).

3.4.3.3   35 Days After Flowering

At 35 DAF, the effect of genotype was significant for all fatty acids but oleic acid (Table 3.6). RG2 had significantly less palmitic acid than all other genotypes but only had significantly lower levels of linoleic acid than RG10 (Table 3.7). RG10 had significantly lower levels of linolenic than the other genotypes (Table 3.7). The effect of temperature was significant for oleic and linolenic acids at 35 DAF (Table 3.6). At high temperature, oleic acid levels were significantly higher than both normal and low temperatures and linolenic acid was significantly lower at high temperature than the other temperatures according to the pair-wise comparisons (Table 3.8). For linoleic acid, the interaction between temperature and genotype was significant and pair-wise comparisons showed that linoleic acid levels were lower at high temperature in RG2 and RG10 but were higher in RG7 and SV64-53 (Fig. 3.8). It was also shown that levels of linoleic acid were lower at low temperature than at normal temperature for three of the four genotypes with RG10 being the exception where levels were the same at both low and normal temperatures (Fig. 3.8).

3.4.3.4   45 Days After Flowering

Evaluating 45 DAF showed that the effect of genotype was significant for all fatty acids except stearic acid (Table 3.6). Pair-wise comparisons for palmitic acid demonstrated that RG2 had significantly lower levels than the other three genotypes but had significantly higher level of oleic acid than RG7 (Table 3.7). RG10 had significantly less linolenic acid than all other genotypes. At 45 DAF,

Table 3.7: Pair-wise comparison LSmeans for genotype of significant differences in fatty acid composition of developing seeds of four genotypes grown at three temperatures. LSmeans could not be estimated for RG7 at 15 DAF due to no samples available for evaluation at 15 DAF grown under low temperature. Comparisons are done within a column for each time point separately. Least Significant Difference (LSD) was based on arithmetic means. LSmeans with the same letter designation are not significantly different.

DAF	Genotype	Palmitic	Stearic	Oleic	Linoleic	Linolenic
15	RG2	102.5^a	79.7^a	180.0^a	387.4^a	249.6^a
	RG7	non-est	non-est	non-est	non-est	non-est
	RG10	168.3^b	81.9^a	108.5^a	471.0^b	170.3^b
	SV64-53	163.6^b	114.5^b	127.5^a	331.7^a	259.3^a
	LSD	29.23	25.99	56.38	60.19	58.20
25	RG2	66.7^a	39.0^a	248.8^a	477.3^a	168.2^a
	RG7	152.7^b	64.0^a	179.4^{ab}	422.4^a	181.6^a
	RG10	170.8^b	56.9^a	122.4^b	526.9^b	123.0^a
	SV64-53	165.7^b	63.8^a	156.8^{ab}	418.8^a	194.7^a
	LSD	22.40	20.53	79.59	55.52	66.11
35	RG2	62.3^a	46.4^a	269.8	473.9^a	147.6^a
	RG7	116.9^b	68.4^a	173.9	523.9^{ab}	117.0^a
	RG10	131.1^b	35.9^a	243.0	557.5^b	32.4^b
	SV64-53	128.2^b	33.2^a	216.4	515.0^{ab}	107.1^a
	LSD	14.30	25.92	70.08	55.98	42.10
45	RG2	42.6^a	23.9^a	276.9^a	564.5^a	92.0^a
	RG7	112.3^b	83.3^a	177.4^b	514.0^a	113.0^a
	RG10	116.3^b	31.9^a	222.5^{ab}	582.7^a	46.6^b
	SV64-53	116.9^b	31.7^a	233.6^{ab}	523.2^a	94.6^a
	LSD	8.69	48.42	62.07	53.33	23.71
55	RG2	41.6^a	25.4^a	242.6^a	602.0^a	88.4^a
	RG7	110.0^b	54.1^b	171.7^b	560.5^{ab}	103.7^a
	RG10	116.1^c	31.4^a	189.7^{ab}	619.3^c	43.5^b
	SV64-53	110.5^b	28.5^a	220.0^{ab}	543.7^b	97.4^a
	LSD	3.73	11.46	40.92	33.12	27.42
65	RG2	43.0^a	25.0^a	200.8	630.7^a	100.7^a
	RG7	110.2^b	50.6^a	166.8	574.4^{bc}	97.9^a
	RG10	116.4^b	27.9^a	207.1	620.3^{ab}	28.3^b
	SV64-53	108.1^b	26.0^a	230.3	554.5^c	81.4^{ab}
	LSD	6.00	32.62	59.12	36.84	35.81

Table 3.8: Pair-wise comparison LSmeans for temperature of significant differences in fatty acid composition of developing seeds of four soybean genotypes grown at three temperatures. LSmeans could not be estimated for RG7 at 15 DAF due to no samples available for evaluation at 15 DAF grown under low temperature. Comparisons are done within a column for each time point separately. Least Significant Difference (LSD) was based on arithmetic means. LSmeans with the same letter designation are not significantly different.

| DAF	Temperature	Palmitic	Stearic	Oleic	Linoleic	Linolenic
15	Low	non-est	non-est	non-est	non-est	non-est
	Normal	151.2[a]	81.6[a]	128.9[a]	406.7[a]	231.6[a]
	High	145.7[a]	106.7[a]	157.0[a]	379.6[a]	211.0[a]
	LSD	24.69	21.95	47.62	50.83	49.16
25	Low	135.9[a]	48.2[a]	170.5[a]	460.5[a]	184.9[a]
	Normal	139.7[a]	49.1[ab]	155.1[a]	482.5[a]	173.5[a]
	High	141.2[a]	70.5[b]	205.0[a]	441.1[a]	142.3[a]
	LSD	19.20	17.59	68.21	47.58	56.65
35	Low	115.2[a]	44.4[a]	192.7[a]	514.7[a]	133.0[a]
	Normal	106.9[a]	45.9[a]	192.0[a]	546.6[a]	108.5[a]
	High	106.8[a]	47.7[a]	292.6[b]	491.5[a]	61.5[b]
	LSD	12.61	22.86	61.81	49.37	37.13
45	Low	98.5[a]	30.9[a]	191.1[a]	559.4[a]	120.1[a]
	Normal	97.4[a]	33.0[a]	238.4[a]	552.4[a]	78.8[b]
	High	95.1[a]	64.2[a]	253.3[a]	526.6[a]	60.8[b]
	LSD	7.54	42.02	53.86	46.27	20.57
55	Low	92.3[a]	29.4[a]	197.9[a]	571.3[a]	109.0[a]
	Normal	95.4[a]	31.8[ab]	209.9[a]	586.1[a]	76.8[b]
	High	95.9[a]	43.3[b]	210.2[a]	586.6[a]	64.0[b]
	LSD	3.23	9.95	35.51	29.00	23.79
65	Low	92.6[a]	27.1[a]	161.7[a]	615.8[a]	102.8[a]
	Normal	94.3[a]	29.1[a]	212.9[a]	585.0[a]	78.7[ab]
	High	96.4[a]	40.8[a]	229.2[a]	584.2[a]	49.7[b]
	LSD	5.15	28.01	50.76	31.63	30.75
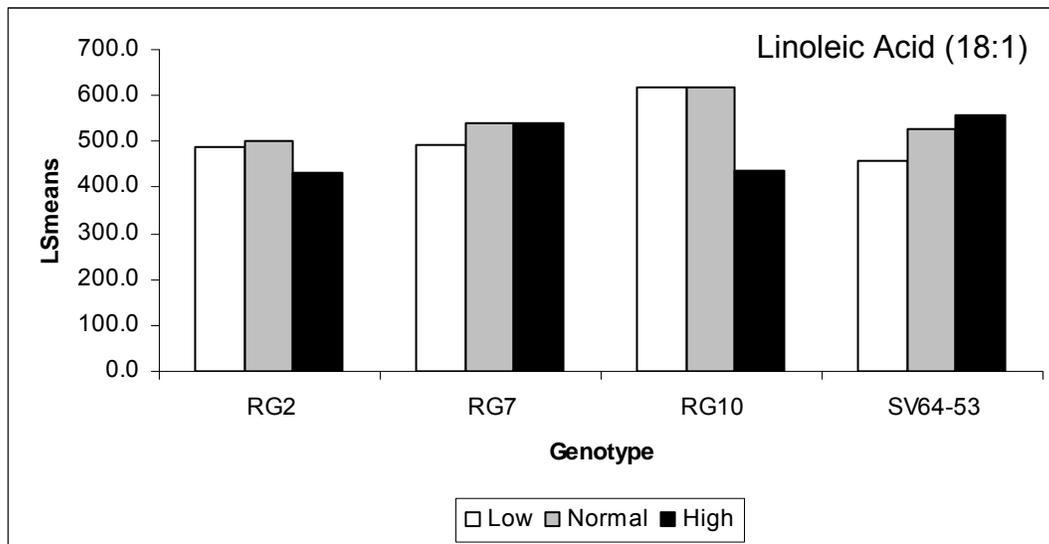
Figure 3.8: LSmeans of significant interactions between time point and growing temperature for fatty acid composition at 35 DAF. Interaction between genotype and growing temperature at 35 DAF was significant for linoleic acid.

66

the effect of temperature was significant for linolenic acid with high temperature having significantly less than both low and normal temperatures (Table 3.8).

3.4.3.5   55 Days After Flowering

At 55 DAF, the effect of genotype was significant for all five fatty acids (Table 3.6). Pair-wise comparisons were performed (Table 3.7). RG2 had significantly lower levels of palmitic acid than all other genotypes, while RG10 had significantly higher levels of palmitic acid than all the other genotypes. RG7 produced significantly more stearic acid than all other genotypes but RG7 only produced significantly more oleic acid than RG2. Looking at pair-wise comparisons for linoleic acid, RG10 had significantly higher levels than RG7 and SV64-53, as well as RG2 had significantly higher levels of linoleic acid than SV64-53. RG10 had significantly less linolenic acid than all the other genotypes.

Temperature had a significant effect on fatty acid levels at 55 DAF for only stearic and linolenic acids (Table 3.6) and pair-wise comparisons were performed (Table 3.8). At low temperature stearic acid accumulated in significantly lower levels than at high temperature but linolenic acid was significantly higher at low temperature than at either normal and high temperatures. There were significant interactions between genotype and temperature for palmitic and stearic acids and pair-wise comparisons were reviewed. In RG2 palmitic acid levels were much lower than the other genotypes. In RG10, palmitic acid levels were higher at high temperature than at other temperatures, a trend that was not seen in the other genotypes (Fig. 3.9A). In RG7 at high temperature, stearic acid

67

A   Palmitic Acid (16:0)
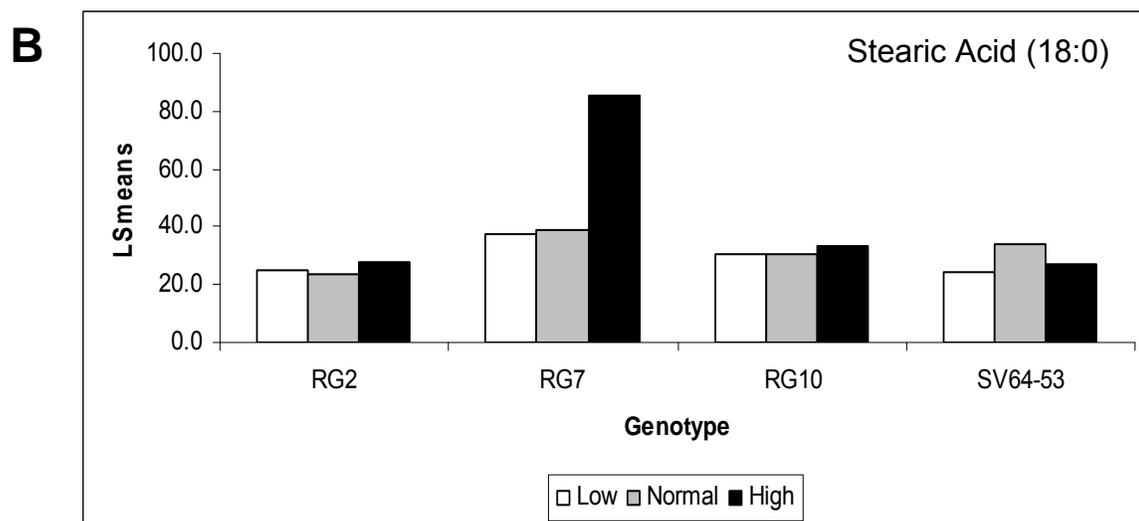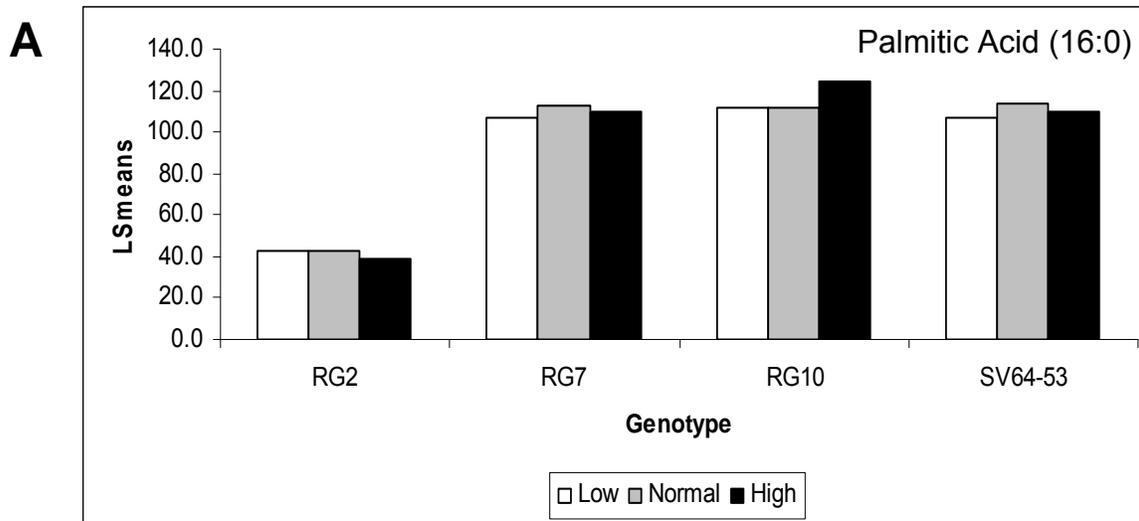
B   Stearic Acid (18:0)

Figure 3.9: LSmeans of significant interactions between time point and growing temperature for fatty acid composition at 55 DAF. Interaction between genotype and growing temperature at 55 DAF was significant for palmitic and stearic acids. A – LSmeans for interaction between genotype and growing temperature for palmitic acid. B – LSmeans for interaction between genotype and growing temperature for stearic acid.

accumulated to a much higher level than in any other genotype or temperature, while levels remained relatively low in the other genotypes (Fig. 3.9B).

3.4.3.6 65 Days After Flowering

At 65 DAF, the effect of genotype was significant for palmitic, linoleic, and linolenic acids (Table 3.6). Pair-wise comparisons showed that RG2 had significantly lower levels of palmitic acid than all other genotypes but had significantly higher levels of linoleic acid than RG7 and SV64-53 (Table 3.7). RG10 had significantly higher levels of linoleic acid than SV64-53 and significantly lower levels of linolenic acid than RG2 and RG7 (Table 3.7). Temperature had an effect at 65 DAF on oleic and linolenic acids, with high temperature producing significantly lower levels of linolenic acid than low temperature (Table 3.8). When fatty acid profiles at 65 DAF at normal temperature was compared to those of mature seeds from plants grown under normal temperature, it was apparent that there were no significant differences with few exceptions (Table 3.9). For this reason, including time points up to maturity was not included.

Table 3.9: P-values of t-test comparing fatty acid profiles at 65 DAF and maturity of plants grown under normal temperature. * - denotes significant differences at α = 0.05

Genotype	Palmitic	Stearic	Oleic	Linoleic	Linolenic
RG2	0.2153	**0.0021***	**0.0064***	0.0629	**0.0438***
RG7	0.3826	0.2873	0.6362	0.7215	0.2687
RG10	**0.0335***	0.2439	0.1884	0.4208	0.7068
SV64-53	0.1974	0.4168	0.1567	0.1244	0.8751

3.5 Discussion

The genotypes studied here – RG2, RG7, RG10, and SV64-53 –
demonstrated that the fatty acid profile can be altered using chemical
mutagenesis and breeding. Hou (2004) studied these lines under field conditions
and obtained similar fatty acid results, with the exception of palmitic acid in
SV64-53. SV64-53 in this study showed a fatty acid profile close to a typical
profile suggesting that there may still be some segregation in this line for fatty
acids and the seed received for this study was close to a typical fatty acid profile.

The final fatty acid composition determined under normal temperatures in
a growth room (Table 3.1) was not statistically different (Table 3.9) from the fatty
acid composition measured at 65 DAF under normal conditions in a growth
chamber (Figs. 3.2, 3.3, 3.4, and 3.6 panel B). RG2 seemed to be the exception
with three of the five fatty acids significantly different.

In three of the four genotypes – RG2, RG10, and SV64-53 – the levels of
oleic acid increased with development, whereas the levels of linoleic acid
decreased (see Figs. 3.2C, 3.4C, 3.6B). The level of a fatty acid measured at
any given time reflects a balance in the rates of synthesis, conversion to other
fatty acids, and incorporation. Since the levels of fatty acids were not the same
across genotypes, suggested that these dynamics are influenced by genotype.
The same can be said for temperature.

Li *et al.* (2010) also investigated the fatty acid accumulation in soybean
seed over development, but they used only one cultivar (cv Jack) and a normal
growing temperature. Comparing their results to the results in this study for

normal growing temperature (Figs. 3.2, 3.3, 3.4, and 3.6 panel B), the genotypes in this study showed similar trends in fatty acid accumulation over development, but the level of each fatty acid and the levels of accumulation differed. For example, in Li *et al.* (2010) the initial concentration of linoleic acid was much lower than what was seen in this study with a steeper increase over development, whereas in this study the initial concentration of linoleic acid was higher and increased at a slower pace. This was likely due to the cultivar used in the literature (cv Jack), since the growing conditions used were identical to those used in this study.

Most of the previous studies looking at the influence of growing temperature on the fatty acid profile of soybean seeds measured the composition at maturity only. One of these studies showed that the levels of palmitic and stearic acids do not change, whereas oleic acid increases and linoleic and linolenic acids decrease with increasing temperatures (18°C to 33°C) (Wolf *et al.*, 1982). In the present study, within a genotype across temperatures some trends emerge. Palmitic acid does not change with changing temperature, which agrees with the literature (Wolf *et al.*, 1982). Stearic acid appears to increase in only RG7 with increasing temperature, which is contrary to the study by Wolf, who used the cultivar Fiskeby V (Wolf *et al.*, 1982). Oleic acid showed a trend that was in contrast to literature in RG7 and SV64-53, in that it increased from low temperature to normal temperature but decreased again at high temperature, while in RG2 and RG10 the trend agreed with previous studies (Wolf *et al.*, 1982). There was no change in the level of linoleic acid in RG2 and RG7 due to

72

temperature, but a decrease was evident in RG10. Linolenic acid levels

decreased with increasing growing temperature in all genotypes, following

literature (Wolf *et al.*, 1982).

Both genotype and temperature influenced fatty acid accumulation, but not

to an equal extent. As the seed developed, the relative importance of these two

factors changed (Table 3.5). For palmitic acid, genotype was the major

contributor to the differences seen across the entire developmental period,

whereas temperature did not influence its accumulation. Temperature was a

significant contributing factor in linolenic acid accumulation from 35 DAF to 65

DAF.

The major factor influencing stearic acid in RG7 was temperature,

whereas seed development was the major factor in the other genotypes (Table

3.2). It is not clear if genotype or temperature was the major factor contributing

to the divergence from the literature in oleic acid accumulation observed in both

RG7 and SV64-53. Regarding linoleic acid, neither RG2 nor RG7 had

temperature as a contributing factor, which was in contrast to that observed in

SV64-53. Across all genotypes, the decrease in linolenic acid can be attributed

to both development and temperature.

This study confirmed that temperature does influence fatty acid

accumulation, although not necessarily in the same manner as previously

reported in the literature (Wolf *et al.*, 1982). It also showed that the four

genotypes studied here had fatty acid profiles that differed from the typical fatty

acid profile (Fehr, 2007), and fatty acid levels accumulated during seed

development differentially under different growing conditions. These findings will help future breeding in the face of climate change by understanding better how these genotypes react to different growing temperatures. More informed decisions can then be made regarding which genotype would be best to grow for bioproduct feedstocks.

CHAPTER 4:

EVALUATION OF THE GENOMIC SEQUENCES OF KENNEDY PATHWAY

ENZYMES IN SEED OF SOYBEAN GENOTYPES WITH ALTERED FATTY

ACID COMPOSITIONS

4.1 Abstract

The Kennedy Pathway creates triacylglycerol, the main oil storage molecule in plants. In this study the genomic sequence of glycerol kinase (GK) (E.C. 2.7.1.30), glycerol-3-phosphate acyltransferase (G3PAT) (E.C. 2.3.1.15), lysophosphatidic acid acyltransferase (LPAAT) (E.C. 2.3.1.51), and diacylglycerol acyltransferase 1 (DGAT1) (E.C. 2.3.1.20) were determined in four fatty acid profile mutant soybean genotypes – RG2, RG7, RG10, and SV64-53. Numerous mutations were identified within the introns of the genes, and only four mutations were found within the coding regions. Of these four mutations, two were silent mutations and the other two changed the amino acid sequence, but the amino acids changed were not pertinent to enzyme function. *In silico* analyses revealed predicted glycosylation and phosphorylation sites, as well as the subcellular localizations and the number of transmembrane domains, if any are present. From the results presented here, it can be concluded that the differences in fatty acid profile in the genotypes studied are not due to mutations in the transcribed genomic sequences of GK, G3PAT, LPAAT, and DGAT1 isoforms selected in this study.

4.2 Introduction

The Kennedy Pathway creates the main oil storage molecule in plants –
triacylglycerol (Ohlrogge & Browse, 1995). This pathway consists of four
enzymes: glycerol-3-phosphate acyltransferase (G3PAT); lysophosphatidic acid
acyltransferase (LPAAT); phosphatidic acid phosphatase (PAP); and
diacylglycerol acyltransferase (DGAT). Glycerol kinase (GK), while not
considered part of the pathway, creates the starting material used , glycerol-3-
phosphate – although new data suggests that this may not be the case
(Frentzen, 1993; Coleman & Lee, 2004; Vigeolas *et al.*, 2007).

GK has been cloned and sequenced from soybean by Huang *et al.* (2003)
(NCBI accession AY492004), and the sequence was submitted directly to
EMBL/GenBank/DDBJ databases. Sequences of GK from other species have
also been directly submitted to the NCBI/GenBank databases. These include
Medicago truncatula (XM_003604305), *Pandanus amaryllifolius* (AY496707),
Oryza sativa (AY377833), *Ricinus communis* (XM_002517635), and *Zea mays*
(EU970286). Two *Arabidopsis thaliana* GKs have been identified – NHO1
(NM_106694) and GLI1 (AY234854). NHO1 was shown to be required for non-
host resistance in arabidopsis (Kang *et al.*, 2003). GLI1 was identified in a
glycerol-insensitive arabidopsis mutant and was shown to be ubiquitous across
all tissues, with an upregulation during senscence and early post-germinative
growth (Eastmond, 2004).

G3PAT catalyzes the first reaction in the Kennedy Pathway by adding an
acyl group to the *sn*-1 position to create lysophosphatidic acid from glycerol-3-

77

phosphate (Frentzen, 1998). G3PAT has been studied in several plant species and it has been found to have various isoforms that have differing subcellular localizations. These isoforms all possess four main motifs that are involved in catalysis and glycerol-3-phosphate (G3P) binding, which reside in the N-terminus where the active site is located (Coleman & Lee, 2004). The importance of the N-terminus to the activity of the enzyme was shown by eliminating the first 35 amino acids in a plastidial G3PAT from *Pisum sativum* (Frentzen, 1994). Other amino acids have also been shown to be important to the activity of the enzyme, and the amino acids S243 and E305 confers substrate specificity, at least in squash (Slabas *et al.*, 2002). Not only do mutations in the coding region of a gene change the protein produced, but it was shown that a mutation located in a typical splice-site of intron 5 in an *Arabidopsis thaliana* plastidial G3PAT led to three alternate transcripts, giving rise to abnormal proteins (Xu *et al.*, 2006). G3PAT is also partially regulated by phosphorylation and dephosphorylation (Lau & Rodriguez, 1996; Coleman & Lee, 2004).

LPAAT catalyzes the second reaction in the Kennedy Pathway. LPAAT has three isoforms based on their subcellular locations (Hares & Frentzen, 1987; Coleman & Lee, 2004; Yu *et al.*, 2004). The central core of the enzyme is the location of the active site, and substituting an alanine to replace a putative catalytic histidine renders the protein inactive (Maisonneuve *et al.*, 2000). Both the N-terminal region and the C-terminal region play a role in enzyme activity as well. When the N-terminus is disrupted, the activity of the enzyme is impaired but when the C-terminus is removed the protein becomes non-functional

(Maisonneuve *et al.*, 2000).

Diacylglycerol acyltransferase 1 (DGAT1) catalyzes the final reaction in the Kennedy pathway by adding an acyl group to the *sn*-3 position of the glycerol backbone, and it is the only enzyme unique to the pathway (Martin & Wilson, 1983; Frentzen, 1998; Hobbs *et al.*, 1999). This protein is an integral membrane protein with 9 to 10 transmembrane domains (Lung & Weselake, 2006). It has been shown that the N-terminal region interacts with the acyl-CoA end of the fatty acid, although the variation in DGAT1 among plant species is mainly in this region (Weselake *et al.*, 2000; Lung & Weselake, 2006). There is sequence variation within a species as well, soybeans for example, but the differences are localized to the introns (Wang *et al.*, 2006). In ancestral corn genotypes, a phenylalanine present at amino acid 469 increases the concentration of oleic acid present compared to modern corn genotypes, which does not contain the phenylalanine (Zheng *et al.*, 2008). Other enzymes, namely DGAT2 and PDAT, can add the third fatty acid to diacylglycerol to form triacylglycerol, but it has been shown that DGAT1 is the predominant enzyme in soybeans (Li *et al.*, 2010; Chapman & Ohlrogge, 2012).

4.2.1 Objectives & Hypothesis

The objective of this Chapter was to identify mutations in the transcribed genomic sequences of GK, G3PAT, LPAAT, and DGAT1 and correlate them to the differences in fatty acid profiles in the genetic material used (Chapter 3). It was hypothesized that mutations will be identified that correspond to the

differences in fatty acid profiles.

4.3 Materials and Methods

4.3.1 Genetic material

      RG2, RG7, RG10, and SV64-53 were grown as described in Chapter 3, Section 3.3.1.1. Ten leaf punches, approximately 0.7cm, per plant were taken using a handheld hole punch adapted for leaf punches and frozen overnight at -20°C. The leaf punches were freeze-dried (Savant Modulyo D, Savant Instruments Inc., Holbrook, NY) for 24 hrs, and then stored at -20°C.

4.3.2 Isolation of genomic DNA

      Approximately 10 leaf punch discs were transferred to a 2.0 mL conical screw cap tube (FisherBrand, Fisher Scientific, Waltham, MA), with one grinding bead. The discs were ground to a fine powder in a tissue homogenizer (FastPrep FP120, Bio/Can Scientific Inc., Mississauga, ON) for 45 s at speed 4.0. Genomic DNA was extracted from the ground tissue using the GenElute^{TM} Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, Saint Louis, MO) following the manufacturers instructions. Quality and quantity was assessed using gel electrophoresis on a 1% agarose gel, and spectrophotometry (NanoDrop® Spectrophotometer ND-1000 v.3.5.2; NanoDrop Technologies, Inc., Wilmington, DE).

4.3.3 Genes selected for sequencing

      The cDNA sequence of glycerol kinase (GK) (NCBI accession AY492004) from soybean was obtained from NCBI. This sequence was BLASTed against

the whole soybean genome of Williams82 (Schmutz J, *et al.*, 2010) using

Phytozome (Phytozome v.7.0) – called the reference sequence – and the

genomic transcribed region of the gene on Chromosome 7 was selected.. The

cDNA sequence of glycerol-3-phosphate acyltransferase (G3PAT) from

Arabidopsis thaliana (NCBI accession NM_102953) was obtained from NCBI.

This sequence was BLASTed to the soybean genome (Phytozome v.7.0) and the

genomic sequence on Chromosome 9 was chosen for isolation and

characterization. The cDNA sequence of lysophosphatidic acid acyltransferase

(LPAAT) from *Ricinus communis* (NCBI accession XM_002526684) was

obtained from NCBI. This sequence was BLASTed to the soybean genome

(Phytozome v.7.0) and the genomic sequence on Chromosome 19 was chosen

for isolation and characterization. The sequence of and diacylglycerol

acyltransferase 1 (DGAT1) from *Arabidopsis thaliana* (NCBI accession

NM_127503) was obtained from NCBI. This sequence was BLASTed to the

soybean genome (Phytozome v.7.0) and the genomic sequence on

Chromosome 13 was chosen for isolation and characterization; this copy of

DGAT1 is known as DGAT1a. For each BLASTed sequence the results with the

lowest e-value (closest match) was selected for isolation and characterization.

These choices were based on the best information available at the time.

4.3.4 Amplification and sequencing of genes

   Using the reference sequence, primers were designed using Primer3

(Primer3 v.0.4.0; Rozen & Skaletsky, 2000), and they were ordered from

Laboratory Services at the University of Guelph. Gradient PCR (Stratagene Robocycler® Gradient 96, Agilent Technologies, Inc., Santa Clara, CA) was performed to determine the optimal annealing temperature and best set of primers for each gene section using a randomly selected line's isolated genomic DNA. Phusion High Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA) was used for PCR reactions that were assembled according to the manufacturer's guidelines. Information regarding the primers used for sequencing all four genes can be found in Appendix A. The PCR cycles were as follows: $98^{o}C$ for 2 min to activate the polymerase; 35 cycles of 1 min at $98^{o}C$, 1 min at $47^{o}C – 58^{o}C$ gradient, 1 min 30 s at $72^{o}C$; and a final extension of 5 min at $72^{o}C$. These reactions were run on a 1.5% agarose gel, imaged using a FluorChem ^{TM} Fc8800 imager (ver. 3.1, Alpha Innotech Corporation, Santa Clara, CA), and the three best were selected for sequencing following purification. Purification was performed using one of two methods: GenElute^{TM} PCR Clean-Up Kit (Sigma-Aldrich, Saint Louis, MO) or GenElute^{TM} Gel Extraction Kit (Sigma-Aldrich, Saint Louis, MO), both according to the manufacturer's instructions. The former was used if one discrete band was seen on the gel and the latter if more than one band was seen on the gel. Following purification, samples were checked for purity on an agarose gel and the concentration was determined using a NanoDrop® spectrophotometer (ND-1000 v.3.5.2; NanoDrop Technologies, Inc., Wilmington, DE). BigDye® Terminator v.3.1 (Applied Biosystems, Carlsbad, CA) was obtained from Advanced Analysis Centre (AAC) Genomics Facility (University of Guelph) and cycle sequencing reactions were

performed on the purified samples following the facility's guidelines. The

reactions were analyzed by the AAC Genomics Facility using an Applied

Biosystems 3730 DNA Analyzer (Applied Biosystems, Carlsbad, CA). The

sequence obtained was aligned with the reference sequence using BLASTN

2.2.25+ from NCBI (Zhang *et al.*, 2000) to locate any mutations. Reverse

reaction results were first put through Reverse Complement (Stothard, 2000)

before aligning to the reference sequence. Once the best set of primers at the

optimal annealing temperature was determined, the genomic DNA from the other

three genotypes was also sequenced using the above method. The sections of

the gene sequenced were contiged together and the final full sequence was

blasted against the whole soybean genome (Schmutz J, *et al.*, 2010) using

Phytozome (Phytozome v.7.0) as a final check.

4.3.5 *In silico* analysis

Since mutations in introns may lead to alternate transcripts, which could

give rise to alternate proteins being formed, the presence of alternate transcripts

was investigated using *in silico* analysis (Schuler, 2008). The gene sequence

was tested for alternate transcripts using GeneSeqer on the PlantGDB website

(Kleffe *et al.*, 1996; Brendel & Kleffe, 1998; Brendel *et al.*, 1998; Usuka *et al.*,

2000; Usuka & Brendel, 2000; Zhu & Brendel, 2001). *Medicago* was chosen as

the splice-site model, with strict alignment to the *Glycine max* cDNA library. The

coding sequence as identified by Phytozome (Phytozome v.7.0) was translated

to its corresponding protein sequence using ExPASY Bioinformatics Resource

Portal – Translate Tool (Gasteiger *et al.*, 2003). The subcellular targeting of the

deduced amino acid sequence was predicted using TargetP v1.1 (Emanuelsson

et al., 2000). Since proteins can be regulated via glycosylation and

phosphorylation, the predicted glycosylation and phosphorylation sites were

obtained using the YinOYang v1.2 program and NetPhos v.2.0 (Blom *et al.*,

1999; Gupta, 2001; Gupta & Brunak, 2002). The location of any transmembrane

domains, as well as the N-terminus orientation, in the deduced amino acid

sequence were analyzed using HMMTop v 2.0 (Tusnády & Simon, 1998;

Tusnády & Simon, 2001).

4.4 Results

4.4.1 Glycerol Kinase (GK)

After sequencing GK on Chromosome 7, no differences were found in the transcribed region of the gene among the genotypes tested, as well as against the Phytozome (Phytozome v.7.0) sequence, which is based on Williams 82. Using the coding region identified by Phytozome (Phytozome v.7.0), the nucleotide sequence was translated to a peptide (Fig. 4.1). Using this translated peptide, *in silico* analyses were performed. The first program used was YinOYang (v.1.2), which predicts glycosylation sites and any of those sites that are also phosphorylated. Seven glycosylation sites were predicted and one site was also predicted to have a phosphorylation site (Fig. 4.2). Further phosphorylation sites were predicted using NetPhos (v.2.0), with 28 phosphorylation sites predicted (Fig. 4.3); this number also includes the one identified by YinOYang (v.1.2). Using HMM Top (v.2.0), it was predicted that GK does not have any transmembrane helices and according to TargetP (v.1.1), it was predicted that GK is localized to an area of the cell that does not include the chloroplast, the mitochondria or the secretory pathway (Fig. 4.4).

4.4.2 Glycerol-3-Phosphate Acyltransferase (G3PAT)

Comparison to the reference sequence from Phytozome (Phytozome v.7.0) on Chromosome 9 suggested the presence of numerous mutations in the sequences from the four genotypes: three in the coding region of the gene, and the remainder in introns (non-coding region). Two of the mutations in the coding

86

M S K E D V F V G A I D Q G T S S S R F I I Y D A K T G V V G C H H V E F T
Q F Y P Q A G W V E H D P M E I L E S V K V C V A K A V D K A T A D G F N
V D K G L K A I G L T N Q R E T T L L W S K S T G L P L H N A I V W M D A R
T S S I C R R L E K E L S G G R N H F V E S C G L P I S T Y F S A L K L L W L
M E N V D A V K E A I K K K D A L F G T I D T W L I W N L T G G V N G G L H
V T D V S N A S R T M L M N L K T L D W D A S T L K T L N I P A E I L P N I V
S N A E I I G E V G S G W P I A G V P I A G C L G D Q H A A M L G Q S C R K
G E A K S T Y G T G A F I L L N T G E G I I Q S K H G L L S T I A F K L G P K
A P T N Y A L E G S V A I A G A A V Q W L R D G L G I I S S A A E I E E M A L
Q V E S T G G V Y F V P A F N G L F A P W W R E D A R G V C I G I T R F T S
K G H I A R A V L E S M C F Q V K D V L D S M H K D S G E S E S Q K T E F L
L R V D G G A T V N N L L M Q I Q A D L V G C P V I R P A D I E T T A L G A A
Y A A G L A T G I W K E D F I F N T E E K L K N A R V F R P V M T E E V R K
K K V D S W C K A V S K T F D L A D L A L **Stop**

Figure 4.1: The predicted amino acid sequence of GK. The protein
sequence was translated using ExPASY Bioinformatics Resource Portal –
Translate Tool (Gasteiger *et al.*, 2003). The protein sequence was found in 5'- 3'
orientation in frame 1. This was the sequence used for subsequent *in silico*
analyses.

A

B
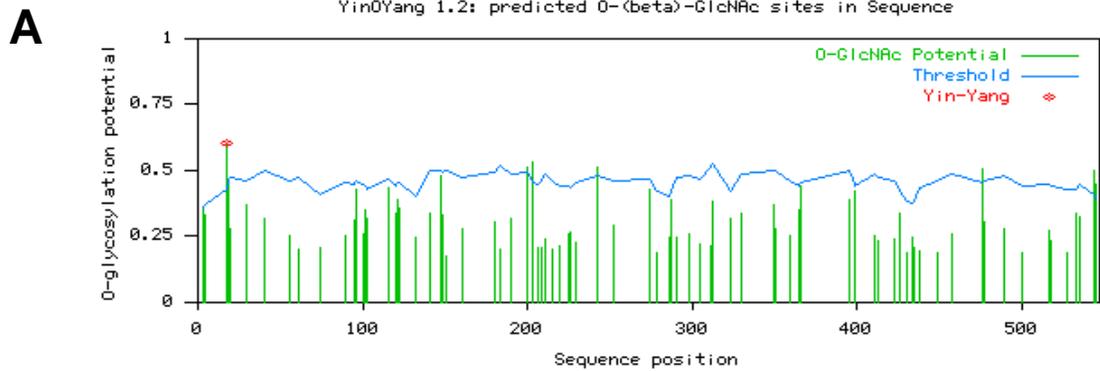
Residue No.	Amino Acid	Glycosylation	Phosphorylation
18	S	√	√
200	T	√	
203	S	√	
242	S	√	
476	T	√	
544	S	√	
545	T	√	

Figure 4.2: YinOYang predictions for glycosylation and phosphorylation sites of the GK protein. Predicted locations, including the residue number and amino acid at that position, of glycosylation and phosphorylation within the GK protein. Sites predicted to have glycosylation and phosphorylation are shown as a Yin-Yang site in A. A – YinOYang 1.2 output graph; B – table denoting the specific residues predicted to have glycosylation and phosphorylation sites. (Gupta, 2001; Gupta & Brunak, 2002)

A
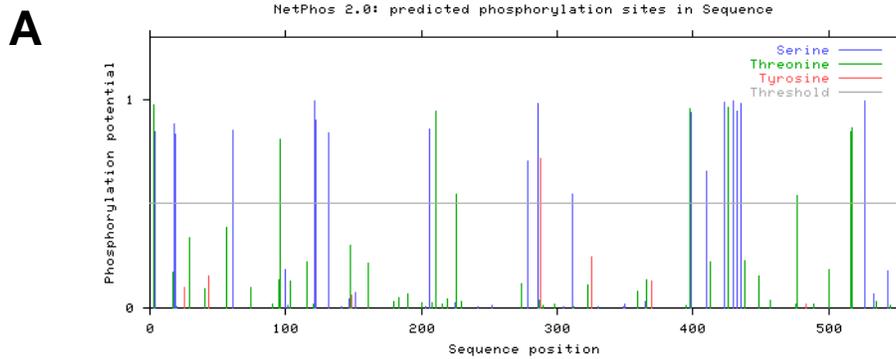
Residue No.	Amino Acid	Score
3	T	0.978
4	S	0.848
18	S	0.885
19	S	0.836
61	S	0.855
96	T	0.811
121	S	0.993
122	S	0.903
132	S	0.838
206	S	0.857
211	T	0.945
226	T	0.545
278	S	0.703
286	S	0.983
288	Y	0.717
311	S	0.547
398	T	0.958
399	S	0.937
410	S	0.658
423	S	0.989
426	T	0.961
430	S	0.992
433	S	0.947
435	S	0.984
477	T	0.538
516	T	0.847
517	T	0.867
527	S	0.992

B (table label at left)

Figure 4.3: Predicted phosphorylation sites in GK protein. Predicted locations, including the residue number and amino acid at that position, of phosphorylation sites in the GK protein. The higher the score number, the more confident the prediction as a phosphorylation site. A – NetPhos 2.0 output graph; B – table denoting the specific residues predicted to be phosphorylated. (Blom *et al.*, 1999)

	Length	cTP	mTP	SP	Other	Location
A	547	0.074	0.106	0.101	0.923	other

B N-terminus: OUT
Number of transmembrane helices: 0

Figure 4.4: Subcellular localization and transmembrane predictions for GK.
A – TargetP (v.1.1) predictions for the subcellular localization of GK. cTP –
chloroplast transit peptide; mTP – mitochondrial target peptide; SP – secretory
pathway signal peptide; Other – another subcellular localization; Location –
predicted localization within the cell. (Emanuelsson *et al.*, 2000) B – HMM Top
(v.2.0) prediction of transmembrane helices and N-terminus orientation.
(Tusnády & Simon, 1998; Tusnády & Simon, 2001)

region were found in RG7, one in RG2, and one in SV64-53 (Figs. 4.5A-C). The

translations of RG2, RG7, and SV64-53 were aligned with the reference

translation revealing a mutation located in exon 9 in RG2 and exon 3 in RG7.

Both of these mutations resulted in a change in the amino acid sequence; in

RG2, glutamic acid was changed to lysine and in RG7, glutamine was changed

to arginine thereby changing the amino acid sequence (Fig. 4.5D). The mutation

in SV64-53 had no effect on the amino acid sequence. *In silico* analysis for

alternate transcripts revealed that mutations in introns would not change the

transcript splicing.

The subsequent *in silico* analyses was performed using the reference,

RG2, and RG7 sequences (Fig. 4.6). The reference sequence was used since

RG10 had no mutations in the coding regions and the mutation in exon 10 in

SV64-53 was found to be silent.   After performing all *in silico* analyses in

parallel, it was found that the E $\rightarrow$ K substitution in RG2 and the Q $\rightarrow$ R

substitution in RG7 did not change any of the predictions for the G3PAT protein.

As such, only the results from the reference sequence were presented.

YinOYang analysis yielded 31 predicted glycosylation sites,

with 12 of those sites also predicted to be phosphorylated (Fig. 4.7). NetPhos

predictions showed that there were a total of 33 predicted phosphorylation

sites(Fig. 4.8). The predicted subcellular location of the G3PAT studied here was

the chloroplast and HMMTop predicted G3PAT to be a soluble protein with no

transmembrane helices (Fig. 4.9) (Tusnády & Simon, 1998; Tusnády & Simon,

2001).

```
A   Williams82  GAAGTGTACAATAACTATAAAAATGCAGTTATCCA AGTGGAGATCCCAAGGCAAAGGAG
    RG7         GAAGTGTACAATAACTATAAAAATGCAGTTATCCG AGTGGAGATCCCAAGGCAAAGGAG
                ********************************** ***********************

B   Williams82  CACCGGAG AATGGGCGCCGGTAAGTTCACTCTCTTGTTAAAGTAAAATAACAAAAGTGTC
    RG2         CACCGGAA AATGGGCGCCGGTAAGTTCACTCTCTTGTTAAAGTAAAATAACAAAAGTGTC
                ******* ****************************************************

C   Williams82  AGACTTGTTGAA CATTCTGGTCCACCAGGTCATGTATATCCTTTAGCGATATTGTGCCAT
    RG7         AGACTTGTTGAG CATTCTGGTCCACCAGGTCATGTATATCCTTTAGCGATATTGTGCCAT
    SV64-53     AGACTTGTTGAG CATTCTGGTCCACCAGGTCATGTATATCCTTTAGCGATATTGTGCCAT
                *********** ************************************************

D   Williams82  MSRTGSSAYYWVADATPPLRNSKTTMFMLSTPPTTTFFVTPTPRVPFLSSSKPSSVSLLR
    RG2         MSRTGSSAYYWVADATPPLRNSKTTMFMLSTPPTTTFFVTPTPRVPFLSSSKPSSVSLLR
    RG7         MSRTGSSAYYWVADATPPLRNSKTTMFMLSTPPTTTFFVTPTPRVPFLSSSKPSSVSLLR
    SV64-53     MSRTGSSAYYWVADATPPLRNSKTTMFMLSTPPTTTFFVTPTPRVPFLSSSKPSSVSLLR
                ************************************************************

    Williams82  SSTASYSPCCSSSSSITPKVKSNDNNNCYLVSAKHSPANTSASVSSRTFLNAKNEQELLA
    RG2         SSTASYSPCCSSSSSITPKVKSNDNNNCYLVSAKHSPANTSASVSSRTFLNAKNEQELLA
    RG7         SSTASYSPCCSSSSSITPKVKSNDNNNCYLVSAKHSPANTSASVSSRTFLNAKNEQELLA
    SV64-53     SSTASYSPCCSSSSSITPKVKSNDNNNCYLVSAKHSPANTSASVSSRTFLNAKNEQELLA
                ************************************************************

    Williams82  GIRKEVEAGSLPANVAAGMEEVYNNYKNAVIQS GDPKAKEIVLSNMIALLDRVFLDVTDP
    RG2         GIRKEVEAGSLPANVAAGMEEVYNNYKNAVIQS GDPKAKEIVLSNMIALLDRVFLDVTDP
    RG7         GIRKEVEAGSLPANVAAGMEEVYNNYKNAVIRS GDPKAKEIVLSNMIALLDRVFLDVTDP
    SV64-53     GIRKEVEAGSLPANVAAGMEEVYNNYKNAVIQS GDPKAKEIVLSNMIALLDRVFLDVTDP
                ****************************** :***************************

    Williams82  FVFQPHHKAKREPFDYYVFGQNYIRPLVDLKNSYVGNMPLFIEMEEKLKQGHNIILMSNH
    RG2         FVFQPHHKAKREPFDYYVFGQNYIRPLVDLKNSYVGNMPLFIEMEEKLKQGHNIILMSNH
    RG7         FVFQPHHKAKREPFDYYVFGQNYIRPLVDLKNSYVGNMPLFIEMEEKLKQGHNIILMSNH
    SV64-53     FVFQPHHKAKREPFDYYVFGQNYIRPLVDLKNSYVGNMPLFIEMEEKLKQGHNIILMSNH
                ************************************************************

    Williams82  QTEADPAIIALLLETRIPYIAESMTYVAGDRVITDPLSKPFSIGRNLICVYSKKHMLDDP
    RG2         QTEADPAIIALLLETRIPYIAESMTYVAGDRVITDPLSKPFSIGRNLICVYSKKHMLDDP
    RG7         QTEADPAIIALLLETRIPYIAESMTYVAGDRVITDPLSKPFSIGRNLICVYSKKHMLDDP
    SV64-53     QTEADPAIIALLLETRIPYIAESMTYVAGDRVITDPLSKPFSIGRNLICVYSKKHMLDDP
                ************************************************************

    Williams82  ALIEMKRNANIRALKEMAMLLRSGSQIVWIAPSGGRDRPDPHTCE APAPFDTSSVDNMR
    RG2         ALIEMKRNANIRALKEMAMLLRSGSQIVWIAPSGGRDRPDPHTCK APAPFDTSSVDNMR
    RG7         ALIEMKRNANIRALKEMAMLLRSGSQIVWIAPSGGRDRPDPHTCE APAPFDTSSVDNMR
    SV64-53     ALIEMKRNANIRALKEMAMLLRSGSQIVWIAPSGGRDRPDPHTCE APAPFDTSSVDNMR
                ********************************************: *************

    Williams82  RLVEHSGPPGHVYPLAILCHDIMPPPLKVEKEIGEKRIISFHGAGISVAPALSFSETTAT
    RG2         RLVEHSGPPGHVYPLAILCHDIMPPPLKVEKEIGEKRIISFHGAGISVAPALSFSETTAT
    RG7         RLVEHSGPPGHVYPLAILCHDIMPPPLKVEKEIGEKRIISFHGAGISVAPALSFSETTAT
    SV64-53     RLVEHSGPPGHVYPLAILCHDIMPPPLKVEKEIGEKRIISFHGAGISVAPALSFSETTAT
                ************************************************************

    Williams82  SENPEKAKELFSKALYDSVTEQYNVLKSAIHGKKGFEASTPVVSLSQPWK
    RG2         SENPEKAKELFSKALYDSVTEQYNVLKSAIHGKKGFEASTPVVSLSQPWK
    RG7         SENPEKAKELFSKALYDSVTEQYNVLKSAIHGKKGFEASTPVVSLSQPWK
    SV64-53     SENPEKAKELFSKALYDSVTEQYNVLKSAIHGKKGFEASTPVVSLSQPWK
                **************************************************
```

Figure 4.5: Mutations in exons 3, 9 and 10 of G3PAT on Chr. 9 and the amino acid sequence. Differences in sequence are indicated by boxes. A – point mutation in exon 3 in line RG7. B – point mutation in exon 9 in line RG2. C – point mutation in exon 10 in lines RG7 and SV64-53. D – change in amino acid sequence in RG2 and RG7 but not in SV64-53. Williams82 – reference sequence from Phytozome. * - matching residues.

A **M** S R T G S S A Y Y W V A D A T P P L R N S K T T M F M L S T P P T T T F
F V T P T P R V P F L S S S K P S S V S L L R S S T A S Y S P C C S S S S S
I T P K V K S N D N N N C Y L V S A K H S P A N T S A S V S S R T F L N A K
N E Q E L L A G I R K E V E A G S L P A N V A A G M E E V Y N N Y K N A V I
Q S G D P K A K E I V L S N M I A L L D R V F L D V T D P F V F Q P H H K A
K R E P F D Y Y V F G Q N Y I R P L V D L K N S Y V G N M P L F I E M E E K
L K Q G H N I I L M S N H Q T E A D P A I I A L L L E T R I P Y I A E S M T Y V
A G D R V I T D P L S K P F S I G R N L I C V Y S K K H M L D D P A L I E M K
R N A N I R A L K E M A M L L R S G S Q I V W I A P S G G R D R P D P H T G
E W A P A P F D T S S V D N M R R L V E H S G P P G H V Y P L A I L C H D I
M P P P L K V E K E I G E K R I I S F H G A G I S V A P A L S F S E T T A T S
E N P E K A K E L F S K A L Y D S V T E Q Y N V L K S A I H G K K G F E A S
T P V V S L S Q P W K **Stop**

B **M** S R T G S S A Y Y W V A D A T P P L R N S K T T M F M L S T P P T T T F F
V T P T P R V P F L S S S K P S S V S L L R S S T A S Y S P C C S S S S S I T
P K V K S N D N N N C Y L V S A K H S P A N T S A S V S S R T F L N A K N E
Q E L L A G I R K E V E A G S L P A N V A A G M E E V Y N N Y K N A V I Q S
G D P K A K E I V L S N M I A L L D R V F L D V T D P F V F Q P H H K A K R E
P F D Y Y V F G Q N Y I R P L V D L K N S Y V G N M P L F I E M E E K L K Q
G H N I I L M S N H Q T E A D P A I I A L L L E T R I P Y I A E S M T Y V A G D
R V I T D P L S K P F S I G R N L I C V Y S K K H M L D D P A L I E M K R N A
N I R A L K E M A M L L R S G S Q I V W I A P S G G R D R P D P H T G K W A
P A P F D T S S V D N M R R L V E H S G P P G H V Y P L A I L C H D I M P P
P L K V E K E I G E K R I I S F H G A G I S V A P A L S F S E T T A T S E N P E
K A K E L F S K A L Y D S V T E Q Y N V L K S A I H G K K G F E A S T P V V
S L S Q P W K **Stop**

C **M** S R T G S S A Y Y W V A D A T P P L R N S K T T M F M L S T P P T T T F
F V T P T P R V P F L S S S K P S S V S L L R S S T A S Y S P C C S S S S S I
T P K V K S N D N N N C Y L V S A K H S P A N T S A S V S S R T F L N A K N
E Q E L L A G I R K E V E A G S L P A N V A A G M E E V Y N N Y K N A V I R
S G D P K A K E I V L S N M I A L L D R V F L D V T D P F V F Q P H H K A K
R E P F D Y Y V F G Q N Y I R P L V D L K N S Y V G N M P L F I E M E E K L
K Q G H N I I L M S N H Q T E A D P A I I A L L L E T R I P Y I A E S M T Y V A
G D R V I T D P L S K P F S I G R N L I C V Y S K K H M L D D P A L I E M K R
N A N I R A L K E M A M L L R S G S Q I V W I A P S G G R D R P D P H T G E
W A P A P F D T S S V D N M R R L V E H S G P P G H V Y P L A I L C H D I M
P P P L K V E K E I G E K R I I S F H G A G I S V A P A L S F S E T T A T S E
N P E K A K E L F S K A L Y D S V T E Q Y N V L K S A I H G K K G F E A S T
P V V S L S Q P W K **Stop**

Figure 4.6: The predicted amino acid sequence of G3PAT. The protein
sequence was translated using ExPASY Bioinformatics Resource Portal –
Translate Tool (Gasteiger *et al.*, 2003). The protein sequence was found in 5'- 3'
orientation in frame 1. A – protein sequence translated from reference nucleotide
sequence. B – protein sequence translated from RG2 nucleotide sequence with
the amino acid difference highlighted in grey. C – protein sequence translated
from RG7 nucleotide sequence with the amino acid difference highlighted in grey.
Both sequences were used in subsequent *in silico* analyses.

93

A

B

Residue No.	Amino Acid	Glycosylation	Phosphorylation
2	S	√	
4	T	√	
34	T	√	
35	T	√	
36	S	√	√
49	S	√	
50	S	√	
51	S	√	
55	S	√	√
57	S	√	√
61	S	√	
62	S	√	√
63	T	√	
65	S	√	√
67	S	√	√
71	S	√	
72	S	√	
73	S	√	
92	S	√	
96	S	√	√
100	T	√	√
101	S	√	
153	S	√	
274	T	√	
278	S	√	√
343	T	√	
353	T	√	√
354	S	√	√
460	T	√	√
464	S	√	
466	S	√	
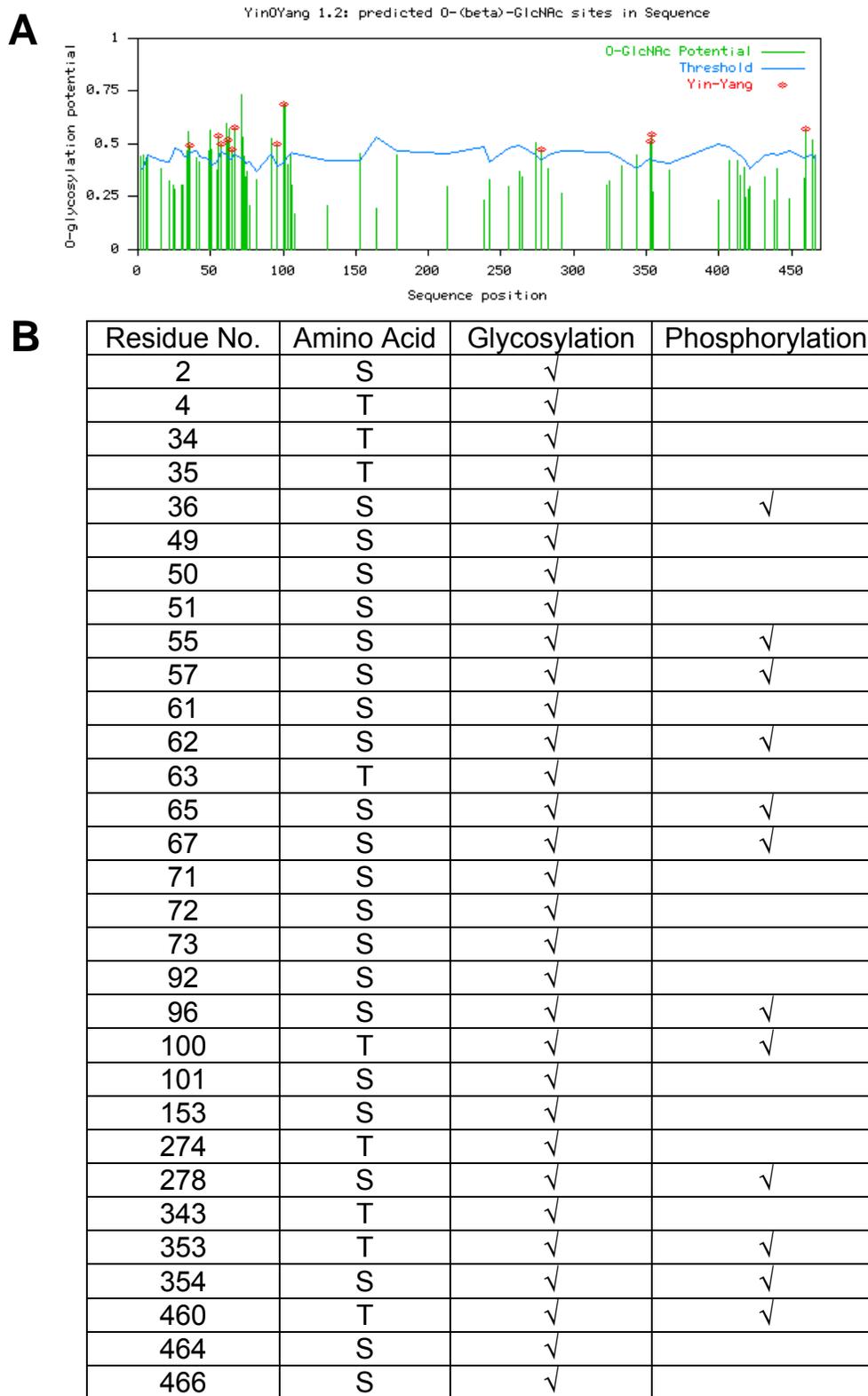
Figure 4.7: YinOYang predictions for glycosylation and phosphorylation sites if the G3PAT protein. Predicted locations, including residue number and amino acid at that position, glycosylation and phosphorylation sites. A – YinOYang 1.2 output graph; B – table denoting the specific residues to have glycosylation and phosphorylation sites. (Gupta, 2001; Gupta & Brunak, 2002)

A

NetPhos 2.0: predicted phosphorylation sites in Sequence

B

Residue No.	Amino Acid	Score
22	S	0.755
36	T	0.906
42	T	0.962
54	S	0.896
55	S	0.983
57	S	0.826
62	S	0.911
65	S	0.868
67	S	0.971
77	T	0.904
89	Y	0.552
96	S	0.985
100	T	0.740
103	S	0.720
105	S	0.921
143	Y	0.941
196	Y	0.685
197	Y	0.831
214	Y	0.518
242	T	0.554
266	Y	0.586
278	S	0.831
282	S	0.948
353	T	0.814
354	S	0.955
366	S	0.642
400	S	0.680
413	S	0.961
421	S	0.795
436	Y	0.949
438	S	0.988
448	S	0.605
460	T	0.916
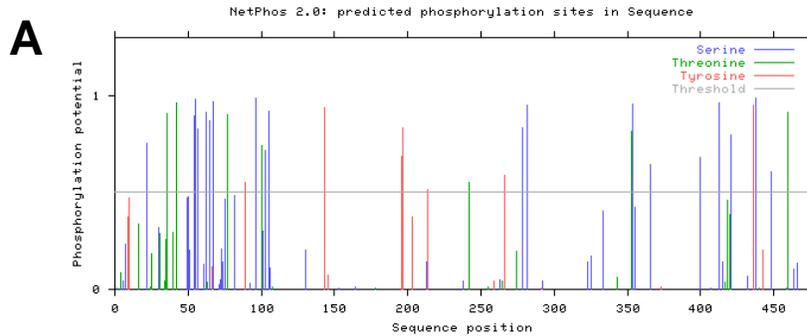
Figure 4.8: Predicted phosphorylation sites in G3PAT protein. Predicted locations, including the residue number and amino acid at that position, of phosphorylation sites in the G3PAT protein. The higher the score, the more confident the prediction as a phosphorylation site. A – NetPhos 2.0 output graph; B – table denoting the specific residues predicted to be phosphorylated. (Blom *et al.*, 1999)

A	Length	cTP	mTP	SP	Other	Location
	470	0.665	0.146	0.000	0.100	Chloroplast

B

```
Length:  470
N-terminus: OUT
Number of transmembrane helices: 0

seq  MSRTGSSAYY WVADATPPLR NSKTTMFMLS TPPTTTFFVT PTPRVPFLSS     50
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  SKPSSVSLLR SSTASYSPCC SSSSSITPKV KSNDNNNCYL VSAKHSPANT    100
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  SASVSSRTFL NAKNEQELLA GIRKEVEAGS LPANVAAGME EVYNNYKNAV    150
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  IQSGDPKAKE IVLSNMIALL DRVFLDVTDP FVFQPHHKAK REPFDYYVFG    200
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  QNYIRPLVDL KNSYVGNMPL FIEMEEKLKQ GHNIILMSNH QTEADPAIIA    250
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  LLLETRIPYI AESMTYVAGD RVITDPLSKP FSIGRNLICV YSKKHMLDDP    300
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  ALIEMKRNAN IRALKEMAML LRSGSQIVWI APSGGRDRPD PHTGEWAPAP    350
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  FDTSSVDNMR RLVEHSGPPG HVYPLAILCH DIMPPPLKVE KEIGEKRIIS    400
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  FHGAGISVAP ALSFSETTAT SENPEKAKEL FSKALYDSVT EQYNVLKSAI    450
pred 0000000000 0000000000 0000000000 0000000000 0000000000

seq  HGKKGFEAST PVVSLSQPWK    470
pred 0000000000 0000000000
```

Figure 4.9: Subcellular localization and transmembrane predictions for G3PAT. A – TargetP (v.1.1) predictions for the subcellular localization of G3PAT. cTP – chloroplast transit peptide; mTP – mitochondrial target peptide; SP – secretory pathway signal peptide; Other – another subcellular localization; Location predicted localization within the cell. (Emanuelsson *et al.*, 2000) B – HMMTop (v.2.0) prediction of transmembrane helices and N-terminus orientation; N-terminus – orientation of N-terminus, either in or out (cytosol); seq – protein sequence, pred – predicted location of amino acids, o/O – outside (cytosolic side). (Tusnády & Simon, 1998; Tusnády & Simon, 2001)

4.4.3 Lysophosphatidic Acid Acyltransferase (LPAAT)

Several mutations were identified in LPAAT located on Chromosome 19, and all were located in introns but none of these mutations gave rise to alternate transcripts (Fig 4.10). Since there were no changes in the splicing and no mutations in the exons, the amino acid sequence was translated from the reference sequence (Fig. 4.11). Using this peptide sequence, YinOYang predictions were performed but these resulted in a warning message that they could not be performed due to the possible presence of a signal peptide. The presence of a signal peptide was confirmed by the TargetP prediction that this protein is part of the secretory pathway (Fig. 4.13A). NetPhos revealed 20 predicted phosphorylation sites (Fig. 4.12). It was predicted to have five transmembrane helices (Fig. 4.13B). It was also predicted that the N-terminus was located on the cytosolic side of the membrane.

4.4.4 Diacylglycerol Acyltransferase 1 (DGAT1)

After sequencing DGAT1 on Chromosome 13 in all of the genotypes, several mutations were identified and all but two of these mutations were located in introns (non-coding region). (Fig. 4.14). Two mutations were located within the coding region in the line SV64-53, but these were found to be silent mutations (Fig. 4.15). The mutations in the introns did not change the splicing of the transcripts.

Since there were no changes in the splicing and the only mutations found in exons were silent mutations, the amino acid sequence was translated from the

A

Nucleotide No.	Mutation
853	A → G
2213	T → A
2670	Deletion
2792	Deletion
5429	T → A
7287	T → A
7289	C → G
8750	Deletion
9717	Insertion

B

Nucleotide No.	Mutation
853	A → G
2604	Deletion
2795	Deletion
2859	Insertion
5429	T → A
9018	Insertion

C

Nucleotide No.	Mutation
6079	Insertion
9030	Deletion

D

Nucleotide No.	Mutation
2795	Deletion
4596	T → G

Figure 4.10: Mutations found in LPAAT located on Chromosome 19. The mutations that were found in LPAAT on Chromosome 19 were all located within introns compared to the reference sequence – Williams82. The nucleotide number where the mutation was located, with the insertions occurring after the number denoted in the table. The substitution mutations are written with the nucleotide in the reference sequence before the arrow and the substituted nucleotide after the arrow. A – mutations in RG2; B – mutations in RG7; C – mutations in RG10; D – mutations in SV64-53.

M A I A A A A V V V P L G L L F F A S G L L V N L I Q A I C Y V V V R P V S K

N L Y R R I N R V V A E L L W L E L V W L I D W W A G V K V Q I F T D H E T

F H L M G K E H A L V I S N H R S D I D W L V G W V S A Q R S G C L G S T L

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

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

G L S V P R N V L I P R T K G F V S A V S H M R S F V P A I Y D I T V A I P K

S S P A P T M L R L F K G Q P S V V H V H I K R H L M K E L P E T D E A V A

Q W C R D I F V A K D V L L D K H I A E D S F S D Q D L Q D T G R P I K S L

L V V I S W V C L V V A G S V K F L Q R S S L L S S W K G V A F S A F G L A

V V T A L M Q I L I Q F S Q S E R S N P A K I A R A K S K N K G G Q L E A R

N D K Q Q **Stop**

Figure 4.11: The predicted amino acid sequence of LPAAT. The protein sequence was translated using ExPASY Bioinformatics Resource Portal – Translate Tool (Gasteiger *et al.*, 2003). The protein sequence was found in 5'- 3' orientation in frame 1. This was the sequence used for subsequent *in silico* analyses.

A

NetPhos 2.0: predicted phosphorylation sites in Sequence

B
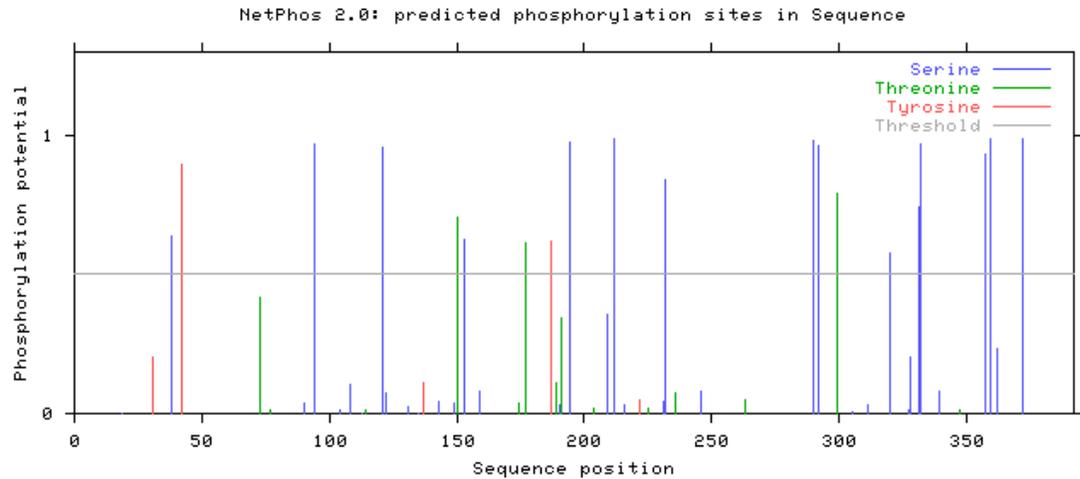
Residue No.	Amino Acid	Score
38	S	0.639
42	Y	0.893
94	S	0.968
121	S	0.954
150	T	0.704
153	S	0.626
177	T	0.616
187	Y	0.620
194	S	0.974
212	S	0.986
232	S	0.843
290	S	0.982
292	S	0.964
299	T	0.792
320	S	0.579
331	S	0.744
332	S	0.966
357	S	0.933
359	S	0.990
372	S	0.985

Figure 4.12: Predicted phosphorylation sites in LPAAT protein. Predicted locations, including the residue number and amino acid at that position, of phosphorylation sites in the LPAAT protein. The higher the score number, the more confident the prediction as a phosphorylation site. A – NetPhos 2.0 output graph; B – table denoting the specific residues predicted to be phosphorylated. (Blom *et al.*, 1999)

A

	Length	cTP	mTP	SP	Other	Location
	387	0.003	0.027	0.989	0.022	Secretory

B

```
Length:  387
N-terminus: OUT
Number of transmembrane helices: 5
Transmembrane helices: 4-23 48-67 99-118 305-324 335-354


seq  MAIAAAAVVV PLGLLFFASG LLVNLIQAIC YVVVRPVSKN LYRRINRVVA     50
pred OooHHHHHHH HHHHHHHHHH HHHiiiiiii iiiiiiiiii iiiiiiiHHH

seq  ELLWLELVWL IDWWAGVKVQ IFTDHETFHL MGKEHALVIS NHRSDIDWLV    100
pred HHHHHHHHHH HHHHHHHooo ooooooooo o ooOOOOOOOO OOOOOOooHH

seq  GWVSAQRSGC LGSTLAVMKK SSKFLPVIGW SMWFSEYLFL ERSWAKDEST    150
pred HHHHHHHHHH HHHHHHHHii iiiiiiiiii iiiIIIIIII IIIIIIIIII

seq  LKSGIQQLSD FPLPFWLALF VEGTRFTQAK LLAAQEYATS TGLSVPRNVL    200
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq  IPRTKGFVSA VSHMRSFVPA IYDITVAIPK SSPAPTMLRL FKGQPSVVHV    250
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq  HIKRHLMKEL PETDEAVAQW CRDIFVAKDV LLDKHIAEDS FSDQDLQDTG    300
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq  RPIKSLLVVI SWVCLVVAGS VKFLQRSSLL SSWKGVAFSA FGLAVVTALM    350
pred iiiiHHHHHH HHHHHHHHHH HHHHoooooo ooooHHHHHH HHHHHHHHHH

seq  QILIQFSQSE RSNPAKIARA KSKNKGGQLE ARNDKQQ    387
pred HHHHiiiiii iiiiiiiiiI IIIIIIIIII IIIIIII
```

Figure 4.13: Subcellular localization and transmembrane predictions for LPAAT. A – TargetP (v.1.1) predictions for the subcellular localization of LPAAT. cTP – chloroplast transit peptide; mTP – mitochondrial target peptide; SP – secretory pathway signal peptide; Other – another subcellular localization; Location – predicted localization within the cell. (Emanuelsson *et al.*, 2000) B – HMM Top (v.2.0) prediction of transmembrane helices and N-terminus orientation; N-terminus – orientation of N-terminus, either in or out (cytosol); seq – protein sequence, pred – prediction location of amino acids, i/I – inside, H – transmembrane helix, o/O – outside (cytosol side). (Tusnády & Simon, 1998; Tusnády & Simon, 2001)

A

Nucleotide No.	Mutation
2890	G → A
3113	T → C
3156	G → A
3201	C → T
3216	Deletion
5112	G → A

B

Nucleotide No.	Mutation
2716	G → A
2890	G → A

C

Nucleotide No.	Mutation
828	Deletion
2886	Deletion
2890	G → A
4648	Deletion
6260	A → G

D

Nucleotide No.	Mutation
1441	C → A
1534	T → A
1535	T → A
1536	T → A
1566	G → A
2422*	T → C
2440*	A → G
2619	G → C
2702	G → A
2716	G → A
2796	T → C
2890	G → A
2952	T → C
2961	T → A
2972	G → A
2988	A → T
4413	T → C
4479	G → T
5438	G → A
5438	Insertion
6345	C → T

Figure 4.14: Mutations found in DGAT1 located on Chromosome 13. The mutations that were found in DGAT1 on Chromosome 13 were all located within introns compared to the reference sequence – Williams82. The nucleotide number where the mutation was located, with the insertions occurring after the number denoted in the table. The substitution mutations are written with the nucleotide in the reference sequence before the arrow and the substituted nucleotide after the arrow. A – mutations in RG2; B – mutations in RG7; C – mutations in RG10; D – mutations in SV64-53; * - mutation in the coding sequence.

```
A  SV64-53    TAGATAAACCATCTGCTTTCCCTCCTTATTCAAACATGTTTCATGTGTTATAGGGAGAAG
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Williams82  TAGATAAACCATCTGCTTTCCCTCCTTATTCAAACATGTTTCATGTGTTATAGGGAGAAG

   SV64-53    CTCTGCCCGATACTCTGAACATGGACTATCCTTACAACGTAAGCTTCAAGAGCTTGGCAT
               |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||| |||
   Williams82  CTCTGCCCGATACTCTGAACATGGACTATCCTTACAATGTAAGCTTCAAGAGCTTAGCAT

   SV64-53    ATTTCCTGGTTGCCCCTACATTATGTTACCAGGTAGCAGTACTTTCAAGTGATTTAGTTA
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Williams82  ATTTCCTGGTTGCCCCTACATTATGTTACCAGGTAGCAGTACTTTCAAGTGATTTAGTTA


B  Williams82  MAISDEPETVATALNHSSLRRRPTAAGLFNSPETTTDSSGDDLAKDSGSDDSISSDAANS
   SV64-53     MAISDEPETVATALNHSSLRRRPTAAGLFNSPETTTDSSGDDLAKDSGSDDSISSDAANS
               ************************************************************

   Williams82  QPQQKQDTDFSVLKFAYRPSVPAHRKVKESPLSSDTIFRQSHAGLFNLCIVVLVAVNSRL
   SV64-53     QPQQKQDTDFSVLKFAYRPSVPAHRKVKESPLSSDTIFRQSHAGLFNLCIVVLVAVNSRL
               ************************************************************

   Williams82  IIENLMKYGWLIKSGFWFSSKSLRDWPLFMCCLSLVVFPFAAFIVEKLAQQKCIPEPVVV
   SV64-53     IIENLMKYGWLIKSGFWFSSKSLRDWPLFMCCLSLVVFPFAAFIVEKLAQQKCIPEPVVV
               ************************************************************

   Williams82  VLHIIITSASLFYPVLVILRCDSAFLSGVTLMLFACVVWLKLVSYAHTNYDMRALTKSVE
   SV64-53     VLHIIITSASLFYPVLVILRCDSAFLSGVTLMLFACVVWLKLVSYAHTNYDMRALTKSVE
               ************************************************************

   Williams82  KGEALPDTLNMDYPYNVSFKSLAYFLVAPTLCYQPSYPRTPYIRKGWLFRQLVKLIIFTG
   SV64-53     KGEALPDTLNMDYPYNVSFKSLAYFLVAPTLCYQPSYPRTPYIRKGWLFRQLVKLIIFTG
               ************************************************************

   Williams82  VMGFIIEQYINPIVQNSQHPLKGNLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAE
   SV64-53     VMGFIIEQYINPIVQNSQHPLKGNLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAE
               ************************************************************

   Williams82  LLRFGDREFYQDWWNAKTVEDYWRMWNMPVHKWMIRHLYFPCLRHGIPKAVALLIAFLVS
   SV64-53     LLRFGDREFYQDWWNAKTVEDYWRMWNMPVHKWMIRHLYFPCLRHGIPKAVALLIAFLVS
               ************************************************************

   Williams82  ALFHELCIAVPCHIFKLWAFGGIMFQVPLVFITNYLQNKFRNSMVGNMIFWFIFSILGQP
   SV64-53     ALFHELCIAVPCHIFKLWAFGGIMFQVPLVFITNYLQNKFRNSMVGNMIFWFIFSILGQP
               ************************************************************

   Williams82  MCVLLYYHDLMNRKGKLD
   SV64-53     MCVLLYYHDLMNRKGKLD
               ******************
```

Figure 4.15: Mutations in exon 7 of DGAT1 on Chr. 13 and the amino acid sequence DGAT1. A – Two point mutations – outlined in boxes – located in exon 7 of DGAT1 on Chromosome 13 in line SV64-53. Both point mutations were found to be silent mutations. B – Amino acid sequences of the reference and SV64-53 coding regions, showing no changes. Williams82 – reference sequence from Phytozome (Phytozome v.7.0). * - amino acids match.

reference sequence (Fig. 4.16). YinOYang analysis was performed since

DGAT1 is not predicted to contain a signal peptide, and it was found that 5 out of

9 glycosylation sites in it were predicted to be phosphorylated as well (Fig. 4.17).

When NetPhos predictions were performed, it was found that DGAT1 contains 25

predicted phosphorylation sites (Fig. 4.18). The subcellular location was

predicted to be "other" – meaning not the chloroplast, not the mitochondria, and

not the secretory pathway (Fig. 4.19A). DGAT1 was predicted to be an integral

membrane protein with 10 transmembrane domains and the N-terminus facing to

the inside (the non-cytosol side) (Fig. 4.19B).

M A I S D E P E T V A T A L N H S S L R R R P T A A G L F N S P E T T T D S S G D D L A K D S G S D D S I S S D A A N S Q P Q Q K Q D T D F S V L K F A Y R P S V P A H R K V K E S P L S S D T I F R Q S H A G L F N L C I V V L V A V N S R L I I E N L M K Y G W L I K S G F W F S S K S L R D W P L F M C C L S L V V F P F A A F I V E K L A Q Q K C I P E P V V V V L H I I I T S A S L F Y P V L V I L R C D S A F L S G V T L M L F A C V V W L K L V S Y A H T N Y D M R A L T K S V E K G E A L P D T L N M D Y P Y N V S F K S L A Y F L V A P T L C Y Q P S Y P R T P Y I R K G W L F R Q L V K L I I F T G V M G F I I E Q Y I N P I V Q N S Q H P L K G N L L Y A I E R V L K L S V P N L Y V W L C M F Y C F F H L W L N I L A E L L R F G D R E F Y Q D W W N A K T V E D Y W R M W N M P V H K W M I R H L Y F P C L R H G I P K A V A L L I A F L V S A L F H E L C I A V P C H I F K L W A F G G I M F Q V P L V F I T N Y L Q N K F R N S M V G N M I F W F I F S I L G Q P M C V L L <u>Y Y H D L</u> M N R K G K L D **Stop**

Figure 4.16: The predicted amino acid sequence of DGAT1a. The protein sequence was translated using ExPASY Bioinformatics Resource Portal – Translate Tool (Gasteiger *et al.*, 2003). The protein sequence was found in 5'- 3' orientation in frame 1. This was the sequence used for subsequent *in silico* analyses. This was the sequence used for subsequent *in silico* analyses. The underlined amino acids are a pentapeptide ER retrieval motif for ER localization (Shockey *et al.*, 2006).

A

YinOYang 1.2: predicted O-(beta)-GlcNAc sites in Sequence

B
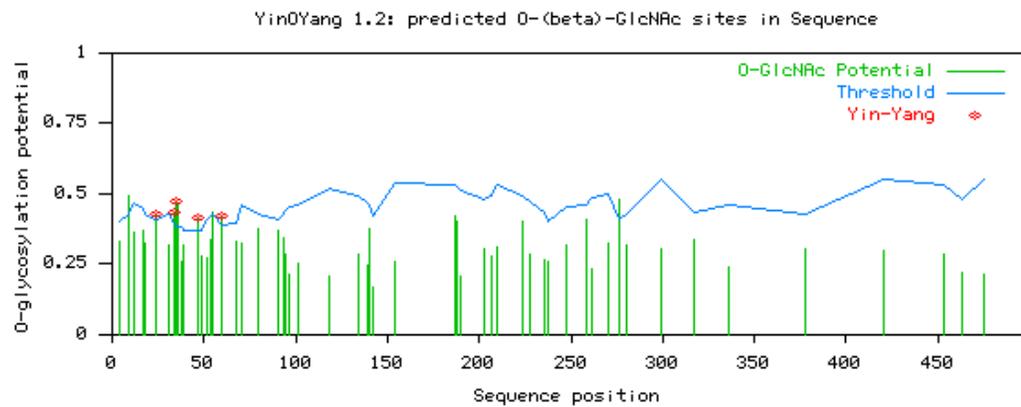
Residue No.	Amino Acid	Glycosylation	Phosphorylation
9	T	√	
24	T	√	√
34	T	√	√
35	T	√	√
36	T	√	
47	S	√	√
55	S	√	
60	S	√	√
276	S	√	

Figure 4.17: YinOYang predictions for glycosylation and phosphorylation sites of the DGAT1a protein. Predicted locations, including residue number and amino acid at that position, of glycosylation and phosphorylation are shown as a Yin-Yang in A. A – YinOYang 1.2 output graph; B – table denoting the specific residues predicted to have glycosylation and phosphorylation sites. (Gupta, 2001; Gupta & Brunak, 2002)

A

B
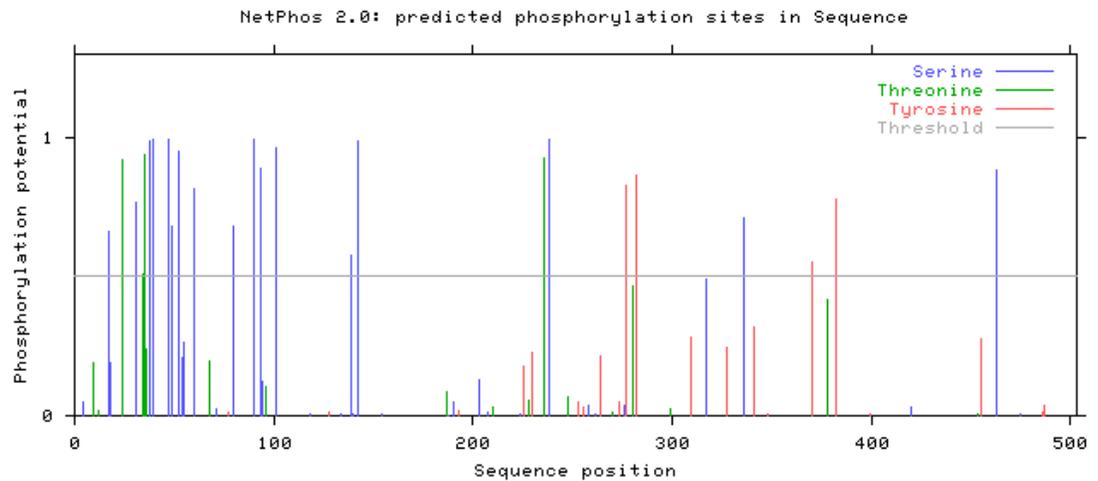
Residue No.	Amino Acid	Score
17	S	0.660
24	T	0.920
31	S	0.767
34	T	0.508
35	T	0.937
38	S	0.989
39	S	0.996
47	S	0.995
49	S	0.683
52	S	0.950
60	S	0.817
80	S	0.678
90	S	0.995
93	S	0.888
101	S	0.961
139	S	0.577
142	S	0.987
236	T	0.928
238	S	0.993
277	Y	0.825
282	Y	0.865
336	S	0.714
370	Y	0.554
382	Y	0.776
463	S	0.880

Figure 4.18: Predicted phosphorylation sites in DGAT1a protein. Predicted locations, including the residue number and amino acid at that position, of phosphorylation site in the DGAT1 protein. The higher the score, the more confident the prediction as a phosphorylation site. A – NetPhos 2.0 output graph; B – table denoting the specific residues predicted to be phosphorylated. (Blom *et al.*, 1999)

	Length	cTP	mTP	SP	Other	Location
A	498	0.205	0.109	0.062	0.677	Other

B
```
Length:  498
N-terminus:  IN
Number of transmembrane helices: 10
Transmembrane helices: 103-122 147-165 178-198 203-220 251-268
293-310 341-359 406-424 437-456 469-487

seq  MAISDEPETV ATALNHSSLR RRPTAAGLFN SPETTTDSSG DDLAKDSGSD    50
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII

seq  DSISSDAANS QPQQKQDTDF SVLKFAYRPS VPAHRKVKES PLSSDTIFRQ   100
pred IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIIIIII IIIIIiiiii

seq  SHAGLFNLCI VVLVAVNSRL IIENLMKYGW LIKSGFWFSS KSLRDWPLFM   150
pred iiHHHHHHHH HHHHHHHHHH HHooooooooo oooooooooo oooooooHHHH

seq  CCLSLVVFPF AAFIVEKLAQ QKCIPEPVVV VLHIIITSAS LFYPVLVILR   200
pred HHHHHHHHHH HHHHHiiiii iiiiiiiHHH HHHHHHHHHH HHHHHHHHoo

seq  CDSAFLSGVT LMLFACVVWL KLVSYAHTNY DMRALTKSVE KGEALPDTLN   250
pred ooHHHHHHHH HHHHHHHHHH iiiiiiiiii iiiiiiiiii iiiiiiiiii

seq  MDYPYNVSFK SLAYFLVAPT LCYQPSYPRT PYIRKGWLFR QLVKLIIFTG   300
pred HHHHHHHHHH HHHHHHHHoo oooooooooo oooooooooo ooHHHHHHHH

seq  VMGFIIEQYI NPIVQNSQHP LKGNLLYAIE RVLKLSVPNL YVWLCMFYCF   350
pred HHHHHHHHHH iiiiiiiiii iiiiiiiiii iiiiiiiiii HHHHHHHHHH

seq  FHLWLNILAE LLRFGDREFY QDWWNAKTVE DYWRMWNMPV HKWMIRHLYF   400
pred HHHHHHHHHo oooooooooo ooooOOOOOO OOOOOOOOOO oooooooooo

seq  PCLRHGIPKA VALLIAFLVS ALFHELCIAV PCHIFKLWAF GGIMFQVPLV   450
pred oooooHHHHH HHHHHHHHHH HHHHiiiiii iiiiiiHHHH HHHHHHHHHH

seq  FITNYLQNKF RNSMVGNMIF WFIFSILGQP MCVLLYYHDL MNRKGKLD   498
pred HHHHHHooooo ooooooooHH HHHHHHHHHH HHHHHHHiii iiiiiiii
```

Figure 4.19: Subcellular localization and transmembrane predictions for DGAT1a. A – TargetP (v.1.1) predictions for the subcellular localization of DGAT1. cTP – chloroplast transit peptide; mTP – mitochondrial target peptide; SP – secretory pathway signal peptide; Other – another subcellular localization; Location – predicted localization within the cell. (Emanuelsson *et al*., 2000) B – HMM Top (v.2.0) prediction of transmembrane helices and N-terminus orientation; N-terminus – orientation of N-terminus, either in or out (cytosol); seq – protein sequence, pred – prediction location of amino acids, i/I – inside, H – transmembrane helix, o/O – outside (cytosol side). (Tusnády & Simon, 1998; Tusnády & Simon, 2001)

108

4.5 Discussion

Across the four genes studied, only one, G3PAT, contained mutations in the coding region, which will be discussed later; all of the other mutations identified were located in introns. An extensive review by Schuler (2008) of genes in plants that have alternative splicing due to mutation in their non-coding regions (introns) demonstrated that the amino acid sequence, and thus the protein, can be affected by mutations in the introns. The sequences of G3PAT, LPAAT and DGAT1 were put into an alternative splicing prediction program and there were no alternate splice site(s) identified, meaning that the mutations found in the introns would not change the amino acid sequence and the protein. For each enzyme, the same *in silico* analyses were performed. One analysis of importance was the prediction of phosphorylation sites since it is known that G3PAT is partially regulated through phosphorylation and dephosphorylation (Coleman & Lee, 2004).

Glycerol kinase, located on Chromosome 7, was predicted to have a subcellular localization of "other" – meaning not chloroplast, mitochondria, or secretory pathway – which is in agreement with the literature, suggesting that GK is associated with the cytosol in arabidopsis and in developing groundnut seeds (Ghosh & Sastry, 1988; Eastmond, 2004). After reviewing the literature, there does not appear to be any documentation of post-translational modifications of GK, yet *in silico* analyses predicted glycosylation and phosphorylation sites. It may be possible that experiments have not yet been conducted to empirically deduce these predicted modifications and their role on the function of the protein.

109

G3PAT, located on Chromosome 9, contained a mutation in the coding region in RG2 (G → A) and RG7 (A → G) but they did not match the amino acids Slabas *et al.* (2002) deduced were important for the functioning of the protein. They used site directed mutagenesis to alter amino acids in the predicted binding domain and these altered proteins were subjected to specificity assays to confirm inactivity. These changes in the amino acid sequence did not change the outcome of any of the *in silico* analyses performed. The G3PAT sequenced in this study was predicted to be targeted to the chloroplast, which partially agrees with the literature, which suggests that G3PAT is found in all areas of the cell, including the chloroplast (Slabas *et al.*, 2001; Coleman & Lee, 2004; Xu *et al.*, 2006). It is possible that this is not the isoform that is active in the Kennedy Pathway, since it is not predicted to be localized to the ER. Also, the prediction by HMMTop that there are no transmembrane helices is also in agreement with the literature indicating that the chloroplast G3PAT is soluble (Coleman & Lee, 2004). Since no signal peptide was predicted for G3PAT, YinOYang predictions showed that G3PAT contains 31 glycosylation sites and 12 of those locations are also phosphorylated. Phosphorylation site prediction yielded 33 sites, which is important since it has been shown that the G3PAT enzyme is partly regulated through phosphorylation and dephosphorylation (Coleman & Lee, 2004).

LPAAT, located on Chromosome 19, contained mutations in the introns, but they did not change the splicing of the transcript. LPAAT was predicted to target to the secretory pathway, which includes the endoplasmic reticulum (ER). This prediction partially agrees with the literature, in that there is evidence that

LPAAT is found to associate with the ER, but it is also found in mitochondria and chloroplasts (Coleman & Lee, 2004; Hares & Frentzen, 1987; Yu *et al.*, 2004). It is possible that more than one LPAAT is present in soybeans, and that the isoforms are targeted to the mitochondria and chloroplasts, whereas the isoform sequenced in this study is the one targeted to the ER where the Kennedy Pathway has been shown to occur (Baud & Lepiniec, 2010). More than one isoform of LPAAT have been identified in various plant species, including Arabidopsis, and it is possible that soybeans also have multiple isoforms (Frentzen, 1993; Frentzen, 1998; Kim & Huang, 2004). Using HMMTop prediction software, it was predicted that LPAAT contains five transmembrane helices with the N-terminus oriented to the outer face of the membrane, and this concurs with the literature that LPAAT is an integral membrane protein (Frentzen, 1993). Since a signal peptide was predicted for targeting, YinOYang predictions for glycosylation could not be performed. Phosphorylation site predictions were able to be performed in the presence of a signal peptide and it was predicted that the LPAAT protein contains 20 phosphorylation sites. A review of the literature suggested that these predicted phosphorylation sites have not been elucidated using experimental evidence to deduce their influence on the protein.

DGAT1, located on Chromosome 13, contained two mutations in the coding sequence in SV64-53, but they were found to be silent mutations. The other mutations identified were located in the introns but these mutations did not result in alternate transcripts. The subcellular localization of DGAT1 was predicted to be "other" – meaning other than chloroplast, mitochondria, or

secretory pathway, which is contrary to the literature. DGAT1 has been localized

to the ER and the chloroplast (Martin & Wilson, 1984; Kaup *et al.*, 2002; Zhang *et*

al., 2005; Lung & Weselake, 2006; McNaughton, 2007). Shockey *et al.* (2006)

showed that DGAT1 in tung tree was localized to the ER by a pentapeptide ER

retrieval motif just upstream from the C-terminus, which is in agreement with the

peptide sequence of soybean DGAT1 in the current study because it appeared to

contain this same motif (Fig. 5.4). Since TargetP analyzes sequences for

"classical" or typical targeting sequences at the N-terminus of a peptide, it may

not be able to predict a signal peptide in DGAT1 since its targeting to the ER

does not appear to follow the "classical" or typical cellular targeting. Unlike

DGAT1 in tung tree in which the N-terminus is facing the cytosolic side (outside)

of the ER, the DGAT1 studied here was predicted by HMMTop to have an inside

facing N-terminus (Shockey *et al.*, 2006). HMMTop also predicted that the

DGAT1 studied here has ten transmembrane helices, which agrees with a

comparison among DGAT1s in nine different species showing nine to ten

transmembrane helices (Lung & Weselake, 2006). Since no signal peptide was

detected, YinOYang predictions found nine amino acids predicted to be

glycosylated with five of them also predicted to be phosphorylated.

Phosphorylation site predictions showed 25 predicted phosphorylation sites in

DGAT1. Phosphorylation sites have been predicted and identified in DGAT1 in

both castor bean and arabidopsis, in agreement with the predictions for soybean

DGAT1 (Hobbs *et al.*, 1999; He *et al.*, 2004).

In this study, no sequence changes in the transcribed region of the genes

were identified that lead to changes in the enzymes. Based on these results, it can be concluded that sequence mutations for the Kennedy Pathway enzymes analyzed in this study are not the basis for the differences seen in fatty acid profiles (Chapter 3). This may suggest that, in the genotypes studied here, other genes are responsible for the altered fatty acid profiles. These genes would encode the elongases and desaturases responsible for fatty acid biosynthesis. It is also possible that other isoforms of the Kennedy Pathway enzymes, or other copies of the genes encoding these enzymes, contribute to the fatty acid profile.

CHAPTER 5:

THE IMPACT OF GROWING TEMPERATURE ON THE EXPRESSION OF

KENNEDY PATHWAY ENZYMES IN DEVELOPING SEED OF SOYBEAN

GENOTYPES WITH ALTERED FATTY ACID COMPOSITION

5.1 Abstract

The main oil storage molecule, triacylglycerol, is created by the Kennedy Pathway. Here the expression of glycerol-3-phosphate acyltransferase (G3PAT), lysophosphatidic acid acyltransferase (LPAAT), and diacylglycerol acyltransferase 1 (DGAT1) in seed of four soybean genotypes with altered fatty acid compositions was investigated as a function of development and three growing temperatures. The expression of G3PAT steadily declined following 15 days after flowering (DAF), and therefore, it is likely to be more highly expressed earlier in development than was measured in the study. Therefore, it did not correspond to the fatty acid accumulation. LPAAT expression coincided with the accumulation of oleic acid ($18:1\Delta^9$) and linolenic acid ($18:3\Delta^{9,12,15}$), but these correspondences were temperature dependent. The expression of DGAT1 corresponded to the accumulation of linoleic acid ($18:2\Delta^{9,12}$), but the level of coincidence varied slightly among genotypes. This study suggested that the expression of the acyltransferase enzymes of the Kennedy Pathway influence the fatty acid composition in seeds from four fatty acid mutant soybean genotypes evaluated.

5.2 Introduction

The Kennedy Pathway produces triacylglycerol, an oil storage molecule in plants (Ohlrogge & Browse, 1995). The pathway consists of four enzymes: glycerol-3-phosphate acyltransferase (G3PAT); lysophosphatidic acid acyltransferase (LPAAT); phosphatidic acid phosphatase (PAP); and diacylglycerol acyltransferase (DGAT). New evidence suggests there are alternate pathways that produce triacylglycerol as well (Chapman & Ohlrogge, 2012).

G3PAT catalyzes the first reaction in the Pathway. G3PAT has differing preferences for which fatty acid it utilizes most efficiently. These differing preferences depends on the plant species and differing subcellular isoforms (Frentzen, *et al.*, 1983; Hares & Frentzen, 1987; Bafor *et al.*, 1990; Eccleston & Harwood, 1990; Frentzen, 1993). G3PAT fatty acid preferences can also be influenced by growing temperature, as was shown in a study on safflower G3PAT (Ichirara, 1984). When Arabidopsis was transformed with an additional G3PAT, the total seed oil concentration increased (Jain *et al.*, 2000).

LPAAT catalyzes the second reaction in the Pathway. LPAAT usually prefers unsaturated fatty acids but the level of that preference is species specific (Griffiths *et al.*, 1985; Bafor *et al.*, 1990). When an additional copy of LPAAT was transformed into soybeans, the total seed oil concentration increased (Rao & Hildebrand, 2009). In addition to increased oil concentration, the concentration of very-long chain fatty acids is be increased in arabidopsis transformed with an additional copy of LPAAT (Zou *et al.*, 1997).

DGAT1 catalyzes the final reaction in the Kennedy Pathway. DGAT1 specificity responds to temperature change in assays, as was shown in canola, where DGAT1 changed from preferring oleic acid to a preference for erucic acid (Cao & Huang, 1987). DGAT1 has a preference for linoleic, palmitic, and oleic acids overall (Cao & Huang, 1986; Lung & Weselake, 2006). DGAT1 is expressed in all aerial tissues of *Brassica napus* but was more tissue specific in expression in Arabidopsis (Hobbs *et al.*, 1999; Lu *et al.*, 2003). In Arabidopsis, DGAT1 also shows a developmentally-dependent expression pattern (Lu *et al.*, 2003). A study including soybeans showed that DGAT1 expression peaked at 45 days after flowering (DAF) (Li *et al.*, 2010). DGAT2 and PDAT are also capable of catalyzing this final reaction but it has been shown that DGAT1 is the predominant enzyme in soybeans (Li *et al.*, 2010; Chapman & Ohlrogge, 2012).

5.2.1 Objectives & Hypothesis

The objective of this Chapter was to determine the developmental expression of G3PAT, LPAAT, and DGAT1 genes in seed of four soybean fatty acid profile mutant genotypes grown under three temperature conditions and to compare these expression profiles with the corresponding fatty acid profiles (Chapter 3). It was hypothesized that expression of these genes will differ among the four genotypes, as well as differ across temperatures, and that differences in expression will correspond to differences in the fatty acid profiles.

117

5.3 Materials and Methods

5.3.1 Genetic material

Plants were grown and material was collected as described in Chapter 3, Section 3.3.1.1.

5.3.2 Isolation of total RNA

Pods were opened and seeds were weighed individually as a measure of developmental stage. The seeds were ground by hand to a fine powder in liquid nitrogen using a mortar and pestle with the mortar and pestles being thoroughly cleaned between uses using ELIMINase (Decon Labs, Inc., King of Prussia, PA) and 0.1% v/v diethylpyrocarbonate-treated (Sigma-Aldrich, Saint Louis, MO) ultra-pure water. Approximately 100 mg of ground tissue was used for extraction of RNA. The extraction was performed using PureLink^{TM} RNA Mini Kit (Invitrogen, Carlsbad, CA), with the tissue first being homogenized in lysis buffer plus 1% β-mercaptoethanol for 30 s at approximately 200 rpm using a Stirrer RZR50 (Conframo Ltd., Wiarton, ON) and disposable pestles for 1.5 mL microfuge tubes (Diamed, Mississauga, ON). The 2 μg extracted RNA was treated with DNase I Amplification Grade (Invitrogen, Carlsbad, CA) to ensure it was free of genomic DNA contamination. To estimate the level of degradation of RNA, samples were run on a 1.5% agarose gel. Other quality checks – such as 260:280 ratio and 260:230 ratio – were performed in triplicate on a NanoDrop® spectrophotometer (ND-1000 v.3.5.2; NanoDrop Technologies, Inc., Wilmington, DE), as well as a measuring the concentration of the sample in ng/μL . To check

for genomic DNA contamination, PCR was performed on DNase treated samples. PCR reactions were prepared following the guidelines provided with JumpStart^{TM} Taq DNA Polymerase (Sigma-Aldrich, Saint Louis, MO), using 1 μL DNase-treated RNA. A positive control reaction of genomic DNA was included. Primers used were: *GmTfl1*-frag1 forward (GCACGTCCATGGTACTAGTAGTA) and *GmTfl1*-frag1 reverse (CAACAAGGAAGAGGAGGAGC) (Tian *et al.*, May 11, 2010). A Stratagene Robocycler® (Agilent Technologies, Inc., Santa Clara, CA) was used for performing the PCR reaction. The PCR cycles were as follows: 95°C for 2 min to activate the polymerase; 35 cycles of 1 min at 95°C, 1 min at 51°C, 1 min 30 s at 72°C; a final extension of 5 min at 72°C.

5.3.3 Quantitative real-time PCR (qRT-PCR)

Quantitative real-time polymerase chain reaction (qRT-PCR) was performed by The AAC Genomics Facility (University of Guelph). A high capacity cDNA reverse transcription kit (Applied Biosystems, Carlsbad, CA) was used to create cDNA from the total RNA extracted. RNA was first diluted in a dilution series prior to reverse transcription reactions. Each reaction contained: 2 μL 10X reverse transcription buffer, 0.8 μL 25X dNTPs (100mM), 2 μL 10X random primers, 1 μL MultiScribe^{TM} Reverse Transcriptase (50 U/μL), 600 ng total RNA, nuclease-free water to bring reaction volume up to 20 μL. The reactions were performed using a PTC-100 Programmable Thermal Controller (MJ Research Inc., St. Bruno, QC) using the following program: 10 min at 25°C, 120 min at 37°C, 5 min at 85°C, then held at 4°C until removal from the machine. After the

reverse transcription was completed, 80 µL of water was added to each reaction and mixed well. Primers used for real-time PCR were designed using Primer Express 3.0 such that they spanned an exon-exon junction (Applied Biosystems, Carlsbad, CA), were ordered from Laboratory Services (University of Guelph) and information on the primers for G3PAT, LPAAT, and DGAT1 are presented in Appendix B. The specificity of these primers were checked using Primer-BLAST and they were determined to be specific for the target genes (Rozen &Skaletsky, 2000). 18S was selected as the housekeeping gene based on a previous study on DGAT1 expression during seed development which used 18S as the housekeeping gene (Li *et al*., 2010). Real-time PCR reactions were set up as follows: 10 µL 2X PerfeCTa SYBR Green FastMix ROX (Quanta BioSciences Inc., Gaithersburg, MD), 0.2 µL forward primer, 0.2 µL reverse primer, 5 µL template, and 4.2 µL water. The real-time PCR reactions were performed using StepOnePlus^{TM} Real-Time PCR System (Applied Biosystems, Carlsbad, CA), with the following program: 30 sec at 95°C to activate the reaction; 45 cycles of 1 sec at 95°C, 30 sec at 60°C; a melting curve followed the cycling using 15 sec at 95°C, 1 min at 60°C, 0.3^{o}C increasing increments held for 15 sec each until 95°C was reached. The melting curve phase was included after each run to confirm the specificity of the amplicons. One to three biological replicates were used depending on the number of samples available for that genotype at that growing temperature and time point. For each biological replicate, two technical replicates were performed.

<u>5.3.4 Analysis of qRT-PCR results</u>

Following real-time PCR, standard curves, melt curves (to check primer specificity) and C_T were determined and analyzed using Step One Plus v.2.2 (Applied Biosystems, Carlsbad, CA). Relative expression was determined using the $\Delta\Delta C_T$ method (Livak & Schmittgen, 2001). ΔC_T was determined by subtracting the C_T of 18S from the C_T of the target gene for each sample, these were then averaged. $\Delta\Delta C_T$ was determined by subtracting the ΔC_T for each sample from the average ΔC_T of 15 DAF sample for the genotype/temperature. The exception was RG7 at low temperature, which was compared to 25 DAF since no samples were available for 15 DAF. The ratio (relative expression) was calculated using the formula: $(1+\text{primer efficiency})^{-\Delta\Delta C_T}$.

5.4 Results

5.4.1 Glycerol-3-Phosphate Acyltransferase (G3PAT)

The expression profile of G3PAT across the four genotypes studied –
RG2, RG7, RG10, and SV64-53 – under three growing temperatures – low
($20^{o}C/15^{o}C$), normal ($25^{o}C/20^{o}C$), and high ($30^{o}C/25^{o}C$) – is shown in Fig. 5.1.
As a note, there were no samples available for 15 DAF for RG7 at low growing
temperature for analysis and the expression analysis for RG7 at low temperature
was relative to 25 DAF for that temperature (Fig 5.1B). All of the other samples
were relative to 15 DAF within their respective temperature and genotype. With
few exceptions, the expression of G3PAT across all temperatures and in all
genotypes was down-regulated compared to 15 DAF. It can be seen that the
level of down-regulation does differ across temperatures within a genotype. For
example, with SV64-53 the down-regulation is less at the high temperature
compared to low and normal temperatures (Fig. 5.1D). In RG2, G3PAT
expression is slightly upregulated at normal temperature at 65 DAF (Fig. 5.1A).
In RG7, expression was upregulated at 65 DAF under both low and high growing
temperatures. In RG10 at the normal temperature expression was upregulated
at 25 DAF and at 65 DAF at the high temperature (Fig. 5.1C).

5.4.2 Lysophosphatidic Acid Acyltransferase (LPAAT)

The results of the expression profiles of LPAAT in the four genotypes –
RG2, RG7, RG10, and SV64-53 – under three growing temperatures – low
($20^{o}C/15^{o}C$), normal ($25^{o}C/20^{o}C$), and high ($30^{o}C/25^{o}C$) are presented (Fig. 5.2).
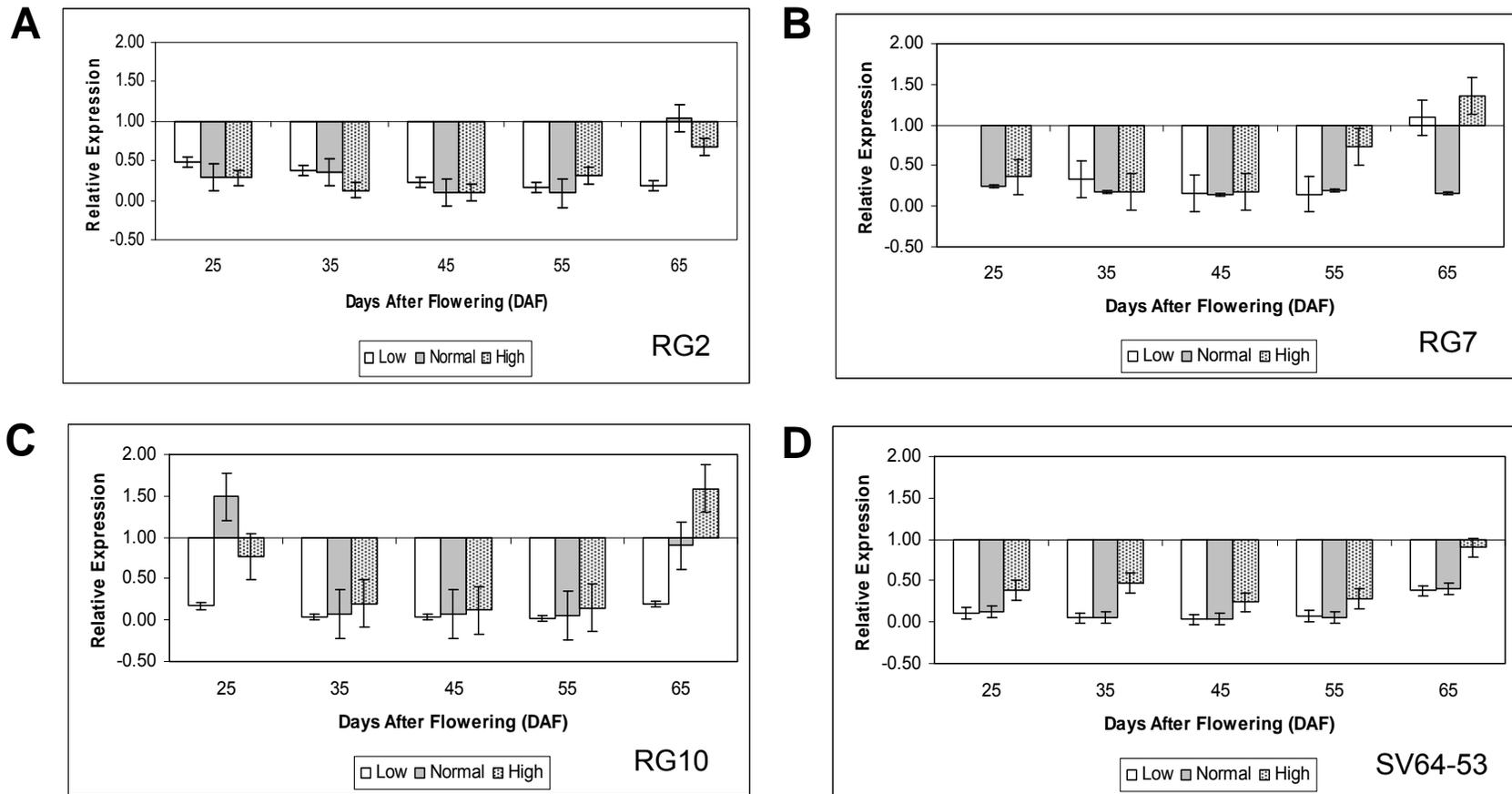
Figure 5.1: Relative expression profiles of G3PAT in developing seed from four soybean genotypes grown at three temperatures. The expression of G3PAT normalized to a housekeeping gene -18S. Expression levels shown are relative to the levels at 15 DAF for each temperature, except RG7 at low temperature which is relative to 25 DAF. Standard error is included. Relative expression was calculated using the $\Delta\Delta C_T$ method. Bars above one show an upregulation of expression; bars below one show a downregulation of expression. A – line RG2; B – line RG7; C – line RG10; D – line SV64-53.
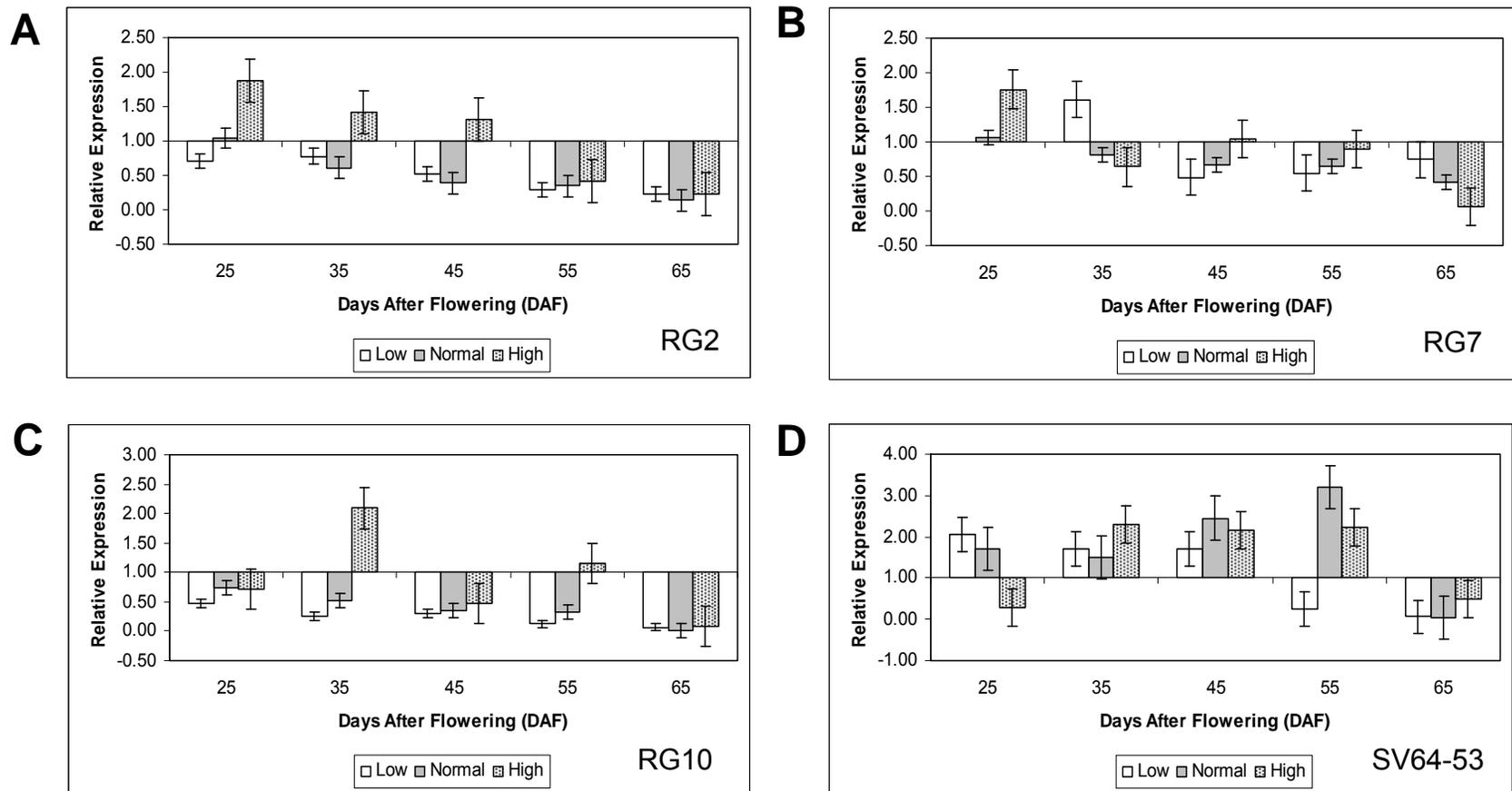
Figure 5.2: Relative expression profiles of LPAAT in developing seed from four soybean genotypes grown at three temperatures. The expression of LPAAT normalized to a housekeeping gene -18S. Expression levels shown are relative to the levels at 15 DAF for each temperature, except RG7 at low temperature which is relative to 25 DAF. Standard error is included. Relative expression was calculated using the $\Delta\Delta C_T$ method. Bars above one show an upregulation of expression; bars below one show a downregulation of expression. A – line RG2; B – line RG7; C – line RG10; D – line SV64-53.

At the low temperature, LPAAT expression was down-regulated relative to 15 DAF for both RG2 (Fig. 5.2A) and RG10 (Fig. 5.2C), whereas it was up-regulated at 25, 35, and 45 DAF and then became down-regulated at 55 and 65 DAF for SV64-53 (Fig. 5.2D). It should be noted that there were no data available for RG7 at 15 DAF under the low temperature, and therefore the expression shown in panel B (Fig. 5.2) for low temperature was relative to 25 DAF. All other samples were relative to 15 DAF within their respective genotype and growing temperature. It can be seen that a drop in expression occurred at 45 DAF. Similar expression trends were seen at the normal temperature as at the low temperature for both RG2 and RG10, while RG7 also showed a down-regulation at 35 DAF and beyond (Fig. 5.2A-C). SV64-53, again, showed a different expression pattern with LPAAT up-regulated throughout development until 65 DAF (Fig. 5.2D).

At the high temperature, expression patterns differed from other temperatures for all genotypes. RG2 showed a decreasing trend across development, with the highest level of expression at 25 DAF and the lowest at 65 DAF (Fig. 5.2A). RG7 oscillated between up- and down-regulation over development with the exception of 65DAF when it continued to be down-regulated, starting with up-regulation at 25 DAF (Fig. 5.2B). RG10 also showed oscillation between up- and down-regulation but started with down-regulation at 25 DAF and ended with down-regulation at 65 DAF (Fig. 5.2C). SV64-53 was down-regulated at 25 DAF but became up-regulated at 35 DAF, dropped slightly at 45 DAF where it leveled off for 55 DAF, then became down-regulated at 65

DAF (Fig. 5.2D).

5.4.3 Diacylglycerol Acyltransferase 1 (DGAT1)

The results of the expression of DGAT1 in the four genotypes studied –

RG2, RG7, RG10, and SV64-53 – under three growing temperatures – low

($20^{o}C/15^{o}C$), normal ($25^{o}C/20^{o}C$), and high ($30^{o}C/25^{o}C$) are presented (Fig. 5.3).

It should be noted that there were no samples available for analyses at 15 DAF

for RG7 under low growing temperature and the expression analysis for low

growing temperature was relative to 25 DAF for RG7 (Fig. 5.3B). All other

samples were relative to 15 DAF within their respective genotype and growing

temperature. At the low temperature, DGAT1 was upregulated throughout

development in all genotypes except RG10 (Fig. 5.3C), where it was down-

regulated at 25, 35, and 55 DAF but upregulated at 45 and 65 DAF. At normal

temperature, DGAT1 was upregulated over the entire development for both RG7

(Fig. 5.3B) and SV64-53 (Fig. 5.3D) but in RG2 (Fig. 5.3A) and RG10 (Fig. 5.3C),

expression became down-regulated at 65 DAF. At high temperature, DGAT1

expression was highly upregulated for the entire developmental period studied

for all genotypes but it was extremely upregulated in SV64-53 (Fig. 5.3D)

compared to the other genotypes – up to over 8 times higher expression.
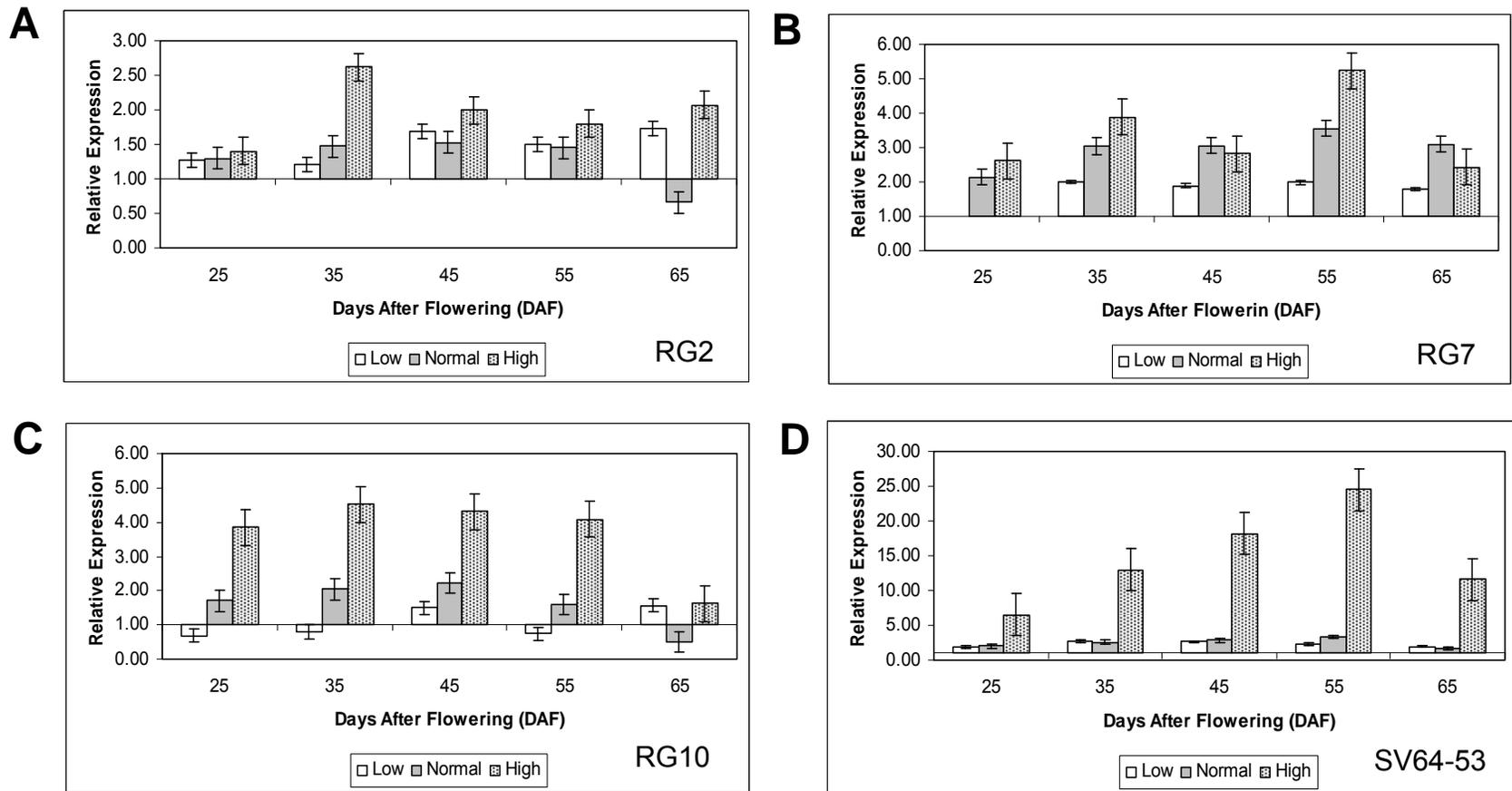
Figure 5.3: Relative expression profiles of DGAT1 in developing seed from four soybean genotypes grown at three temperatures. The expression of DGAT1 normalized to a housekeeping gene -18S. Expression levels shown are relative to the levels at 15 DAF for each temperature, except RG7 at low temperature which is relative to 25 DAF. Standard error is included. Relative expression was calculated using the $\Delta\Delta C_T$ method. Bars above one show an upregulation of expression; bars below one show a downregulation of expression. A – line RG2; B – line RG7; C – line RG10; D – line SV64-53.

5.5 Discussion

The expression patterns of G3PAT, LPAAT, and DGAT1 were studied here. Due to the unavailability of samples of RG7, at 15 DAF under low growing temperature, relative expression was measured in relation to 25 DAF. Since no genomic sequence mutations were identified (see Ch. 4), the differences in expression seen may be due to other factors, such as regulatory regions outside the gene.

No obvious trends were observed in the developmental expression patterns in G3PAT (Fig. 5.1) compared to the fatty acid profiles (Fig. 3.2, 3.3, 3.4, and 3.6). The reason may be that after 15 DAF, there is a sufficient amount of the enzyme produced and therefore expression can be down regulated. Since the product of G3PAT is utilized for the subsequent reactions, G3PAT would be needed the earliest in development to create this product – lysophosphatidic acid. In *B. napus* G3PAT expression was found to increase at 23 to 29 DAF from basal levels then it returned to basal levels shortly thereafter (O'Hara *et al.*, 2002). If this experiment were to be repeated, developmental time points prior to 15 DAF should be used to determine if G3PAT expression is higher at the earlier stages of seed development once the seed size is reasonable for experimentation. From the results of this study, it is difficult to determine the preferences for fatty acid that soybean G3PAT possesses and whether those preferences change under different growing temperatures. Ichirara (1984) showed that the preference of G3PAT for fatty acids was influenced by temperature in safflower. Further studies would be needed to obtain this

information for the soybean G3PAT and to determine if it was in line with

G3PATs isolated from chloroplasts of pea and spinach, both of which had a

preference for oleic acid (Frentzen *et al.*, 1983). This could be done through

specificity and selectivity studies, while including various temperatures at which

the assays are conducted. It may also be possible that the isoform of G3PAT

studied here is not the isoform involved in the Kennedy Pathway, and future

studies should take this into account and include other isoforms in those studies.

When the developmental expression patterns of LPAAT (Fig. 5.2) are

compared to the developmental fatty acid profiles (Fig. 3.2, 3.3, 3.4, and 3.6),

there are some patterns that emerge. Some of the expression profiles over the

development correspond to changes in the fatty acid profiles over the same

developmental period. At the low and normal temperatures, the decrease in

expression of LPAAT corresponded with the decrease of linolenic acid (18:3) for

RG2. At the high temperature for RG2 though, the expression pattern coincided

with the accumulation of oleic acid (18:1). However, in RG7 there did not appear

to be any correspondence between the expression pattern and the fatty acid

profile at both the low and high temperatures, but the expression pattern at

normal temperature did coincide with the accumulation of linolenic acid (18:3). In

RG10, there did not appear to be any correspondence between LPAAT

expression and fatty acid profile at the high temperature, but at the low

temperature the expression did correspond to a decrease then leveling off of

palmitic acid (16:0). It can also be seen that the decrease in expression at the

normal temperature agreed with changes in both palmitic (16:0) and linolenic

(18:3) acids. When SV64-53 was considered, at the low temperature the expression coincided with the accumulation of oleic acid (18:1). At the normal temperature, the increase and plateau in expression overlapped with the increase and plateau of linoleic acid (18:2). The expression pattern at high temperature corresponded to both oleic (18:1) and linoleic (18:2) acids.

There were differences found in the expression of DGAT1 within a temperature across genotypes over seed development (Fig. 5.3). There were also differences in expression within a genotype across temperatures, such as SV64-53 where DGAT1 expression increased dramatically at high temperature compared to low and normal temperatures (Fig. 5.3D). When the developmental expression patterns of DGAT1 (Fig. 5.3) were compared to the fatty acid profiles (Fig. 3.2, 3.3, 3.4, and 3.6), trends could be observed. The trends seen in the expression of DGAT1 corresponded to trends seen in the fatty acid profiles. The expression of DGAT1 at normal temperature in this study was similar to that found in the cultivar Jack in a study by Li *et al.* (2010), with some differences in RG7 and SV64-53 in that there was no drop in expression at 65 DAF. The accumulation of fatty acids in this study also followed the pattern reported by Li *et al.* (2010) except that the genotypes studied here had a higher level of linoleic acid (18:2) accumulated by 65 DAF than did cv Jack. In RG2 at the normal and high temperatures, the expression of DGAT1 corresponded similarly to oleic acid (18:1). At low temperature though, DGAT1 expression corresponded to linoleic acid (18:2). DGAT1 expression in RG7 corresponded with linoleic acid (18:2) accumulation at both normal and high temperature but corresponded to stearic

acid (18:0) accumulation at low temperature only. When RG10 was considered, there appeared to be no correspondence between fatty acid accumulation and DGAT1 expression at both low and high temperatures, but the expression corresponded to linoleic acid (18:2) accumulation at normal temperature. In SV64-53, DGAT1 expression coincided with the accumulation of linoleic acid (18:2) at all temperatures, although the accumulation of 18:2 at high temperature was not proportional to the substantially higher level of expression observed at that temperature. It should be noted that there was larger gap between 18:2 and other fatty acids in SV64-53 at the high temperature compared to the other temperatures.

Overall, no discernable substrate preference could be ascertained for G3PAT. It appears that LPAAT expression corresponds with the accumulation pattern of linolenic acid (18:3) and oleic acid (18:1), suggesting that the soybean LPAAT studied here had a preference for linolenic acid (18:3), but that preference appears to be altered at growing temperatures outside of normal. The apparent preference of the LPAAT in this study for linolenic (18:3) and oleic (18:1) acids agrees with the literature, which suggested that LPAAT is most active with unsaturated fatty acids (Griffiths *et al.*, 1985). It appears that DGAT1 expression corresponded to the accumulation pattern of linoleic acid (18:2) suggesting that the soybean DGAT1 studied here had a preference for linoleic acid (18:2) which, however, varied slightly across genotypes. The apparent preference of DGAT1 in this study for linoleic acid (18:2) agrees with the literature showing that DGAT1 in soybeans had a preference for linoleic acid

(18:2) (Cao & Huang, 1986). Further studies of the specificity and selectivity are needed to confirm the preferences found in this study, similar to those by Cao & Huang (1986). It is also possible that the differences seen among the genotypes could also be due to the enzyme activity or the output of the enzymes. Kinetic studies of these enzymes to understand the rate at which they are catalyzing the reactions should also be performed.

CHAPTER 6: GENERAL DISCUSSION

This thesis aimed to elucidate how the Kennedy Pathway contributes to the final fatty acid profile in four mutant soybean genotypes with altered fatty acid profiles. This goal was approached in two ways: (1) sequencing of genes encoding the enzymes of the Kennedy Pathway; and (2) studying the expression of some of these genes throughout seed development. The results from these two studies were compared to the fatty acid profiles measured. It should be noted that phosphatidic acid phosphatase (PAP) was not included in this study, as it is not well characterized or studied in the literature, in any organism and has only recently been cloned in mammals.

The transcribed genomic sequences of GK, G3PAT, LPAAT, and DGAT1 revealed only one change in the coding region that would result in a change in amino acid sequence. The change was located in G3PAT in RG2 and RG7, but *in silico* analyses suggested that enzyme function would not be altered. Numerous mutations were located in the intron regions but none led to alternate transcripts, according to *in silico* analysis. It appears that the differences in the fatty acid profiles of the genotypes in this study were not due to mutations in the transcribed region of the genes encoding the Kennedy Pathway enzymes.

GK was omitted from the gene expression study since differences were not found in the transcribed genomic sequence of the gene; it was also omitted due to resource constraints and it was decided to focus on the acyltransferases. Therefore, only G3PAT, LPAAT, and DGAT1 were included. Results from the expression analysis showed differences in the expression of each enzyme.

These differences were observed among genotypes and across growing temperatures. Since the RNA used for expression analysis came from the same samples it can be said that overall, G3PAT had the lowest levels of expression (Fig. 5.1), followed by LPAAT (Fig. 5.2), with DGAT1 having the highest expression levels (Fig. 5.3). This expression pattern was expected, as it follows the order in which the enzymes function in the pathway and a build up of each enzyme to sufficient levels would be needed in this order (Fig. 1.1). It is interesting to note that both G3PAT and LPAAT were down-regulated after 15 DAF in some of the genotypes, suggesting that they were expressed earlier in seed development creating a possible buildup of the enzymes to a sufficient level. It can be speculated that PAP would have an expression pattern that would fit between LPAAT and DGAT1, with a higher expression at 25 and 35 DAF then declining thereafter since it catalyzes the step between LPAAT and DGAT1 (Fig. 1.1). It should also be noted that SV64-53 had the highest levels of expression for both LPAAT and DGAT1 compared to the other genotypes, regardless of temperature. This result was observed especially under high temperature for DGAT1.

The line SV64-53 appears to be a unique line. During the course of collecting the material for the expression study, it was observed that all of the SV64-53 plants began flowering at approximately the same time, regardless of the growing temperature. The other genotypes began flowering under high temperature first, then normal temperature, and finally low temperature with a difference of about 2 weeks between flowering at high and low temperatures.

SV65-53 also had the most pods. It was notable that the expression levels were very high under high temperature for DGAT1 in SV64-53, which suggests that SV64-53 is a distinctive soybean line with altered fatty acids that warrants further study. One possible study would be to understand if the synchronized flowering time across growing temperature is a phenomenon of an artificial environment or not. This could be done by growing SV64-53 in different heat unit zones in the field and noting the number of days to flowering. The underlying molecular mechanism of this synchronized flowering should also be investigated so that it may be traced through the progeny of SV64-53 for the development of a cultivar that will flower at the same number of days after planting regardless of which heat zone it is grown in.

6.1 Limitations

Since it is known that the soybean genome underwent a whole genome duplication, at least two copies of a gene are usually present (Schmutz *et al.*, 2010). In this study, only one copy of each gene was sequenced due to time and resource constraints. Also, the expression of only one copy of each gene was studied. The primers used in the expression study were BLASTed against the whole genome and was found to be specific for the copy of the gene used in this study. In additional studies, including the other copies of these genes may be informative.

Some of the enzymes of the Kennedy Pathway are encoded by gene families and this study only looked at one member of each of those families. The

genes were selected based on the best information available at the time as to which member to select. It should also be noted that the soybean genome was not completed and published until 2010, making it difficult to obtain sequence information for this study. There was time-constrainsts that precluded sequencing other copies of the genes or other gene family members. In the future, studies should also be done to include the other family members to deduce what contributions, if any, they make to the final fatty acid profile. These studies could identify which members of the enzyme families are important to triacylglycerol biosynthesis in seeds.

In the expression study, it would have been ideal to repeat the study such that each of the three growth chambers used would have been low temperature, normal temperature, and high temperature. Due to time and resource constraints, this was not possible. The expression study should be repeated in the future using the same growth chambers and rotating the temperature regimes. The corresponding fatty acid profiles should also be measured during the repeated growth room study. Comparing the fatty acid profiles obtained in a growth room study to those obtained from field conditions would also be informative.

6.2 Contributions and Future Directions

To date, the Kennedy Pathway has not been well characterized in soybean and this study provides a foundation on which to build further studies. It is evident from the differences in expression that there is an underlying

mechanism causing the differences seen in this study. This mechanism may be a transcription factor or differences in regulatory regions outside the transcribed region of the genes themselves. Another possible mechanism may be the flux of carbon during seed development, as is seen in developing potato tubers, where an increase in sucrose leads to an upregulation of two enzymes used to process the sucrose into starch at the trasncriptional level (Geigenberger, 2003). This study found that expression differences coincided with differences in fatty acid accumulation over seed development, but further specificity/selectivity studies would need to be performed to confirm what was seen here. This could be done by cloning the cDNA of these genes and expressing them in bacteria or yeast to gain a purified protein that can then be subjected to specificity and/or selectivity studies such as the ones carried out by Frentzen *et al.* (1983).

It would also be informative to repeat the expression study done here but include time points earlier than 15 DAF to better capture the expression of G3PAT and LPAAT during the early development with the seed size being considered when selecting earlier time points. Other members of the gene families of G3PAT, LPAAT and DGAT should also be investigated to elucidate any additional roles they may play in the fatty acid profiles.

In addition to including earlier dates in a repeated expression study, it would also be informative to include time points later in development than 65 DAF, possibly up to maturity. This could assist in capturing the down-regulation of DGAT1 later in development, especially in SV64-53 grown under high temperature where the expression of DGAT1 was extremely high (Fig. 5.3D).

137

Little to no down-regulation of DGAT1 was observed in this study (Fig. 5.3) suggesting that in the genotypes studied here, DGAT1 expression remained high further into development than was expected.

It may be useful to include PAP in a future expression study. Despite not being well researched, it would be important to understand the expression pattern of PAP over seed development and how it fits between the expression of LPAAT and DGAT1. Understanding how PAP expression is affected by temperature would also be informative since the expression of the other enzymes – G3PAT, LPAAT, and DGAT1 – were affected by growing temperature.

This study also investigated the accumulation of fatty acids over seed development grown under three different temperature regimes. It was shown that both temperature and genotype influenced the fatty acid profile over seed development, although the factor that had the major influence changed over seed development. In the future, time points earlier than 15 DAF (while considering the seed size) and time points later than 65 DAF, up to maturity, should be included in a study repeating the experiments presented in this study. By including additional time points, a deeper understanding of what the major factors are influencing the accumulation of fatty acids over seed development. This would also provide insight into critical periods where temperature influences fatty acid accumulation during seed development.

6.3 Conclusions

This study provided a first look into the role the Kennedy Pathway plays in

fatty acid profiles in soybean. It appears that the enzymes in the Kennedy

Pathway studied here may play a role in determining the fatty acid profile in the

mutant soybean genotypes studied. However, the regulation of their expression

rather than mutation in their DNA sequence appears to result in the altered fatty

acid composition of the soybean oil. It is likely that the fatty acid profile observed

is a combination of the fatty acid biosynthesis and the triacylglycerol biosythesis.

REFERENCES

Alam, I., Sharmin, S.A., Kim, K.-H., Yang, J.K., Choi, M.S., Lee, B.-H. 2010.
Proteome analysis of soybean roots subjected to short-term drought
stress. *Plant Soil*, 333: 491-505.

Alt. J.L., Fehr, W.R., Welke, G.A., Sandhu, D. 2005a. Phenotypic and molecular
analysis of oleate content in the mutant soybean line M23. *Crop Sci*, 45
(5): 1997-2000.

Alt, J.L., Fehr, W.R., Welke, G.A., Shannon, J.G. 2005b. Transgressive
segregation for oleate content in three soybean populations. *Crop Sci*, 45
(5): 2005-2007.

American Soybean Association. 2011. Soy Stats Online.
http://www.soystats.com/2011/

Anai, T., Yamada, T., Kinoshita, T., Rahman, S.M., Takagi, Y. 2005. Identification
of corresponding genes for three low-α-linolenic acid mutants and
elucidation of their contribution to fatty acid biosynthesis in soybean seed.
Plant Sci, 168 (6): 1615-1623.

Bafor, M., Stobart, A.K., Stymne, S. 1990. Properties of the glycerol acylating
enzymes in microsomal preparations from developing seeds of safflower
(*Carthamus tinctorius*) and turnip rape (*Brassica campestris*) and their
ability to assemble cocoa-butter like fats. *J Am Oil Chem Soc*, 67 (4): 217-
225.

Baggott, J. 1998. Mobilization and transport of adipose fatty acid.
http://library.med.utah.edu/NetBiochem/FattyAcids/7_2.html, current as of

August 10, 2011.

Baud, S., Dubreucq, B., Miquel, M., Rochat, C., Lepiniec, L. 2008. Storage reserve accumulation in Arabidopsis: Metabolic and developmental control of seed filling. *The Arabidopsis Book*, 63 (1): 1-24.

Baud, S., Lepiniec, L. 2010. Physiological and developmental regulation of seed oil production. *Prog Lipid Res*, 49: 235-249.

Bilyeu, K., Palavalli, L., Sleper, D., Beuselinck, P. 2005. Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. *Crop Sci*, 45 (5): 1830-1836.

Block, M.A., Dorne, A.-J., Joyard, J., Douce, R. 1983. The phosphatidic acid phosphatase of the chloroplast envelope membrane. *FEBS J*, 164 (1): 111-115.

Blom, N., Gammeltoft, S., Brunak, S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*, 294 (5): 1351-1362.

Brendel, V., Kleffe, J. 1998. Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in *Arabidopsis thaliana* genomic DNA. *Nucl Acids Res*, 26 (20): 4748-4757.

Brendel, V., Kleffe, J., Carle-Urioste, J.C., Walbot, V. 1998. Prediction of splice sites in plant pre-mRNA from sequence properties. *J Mol Biol*, 276 (1): 85-104.

Brendel, V., Xing, L., Zhu, W. 2004. Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genetic locus.

Bioinformatics, 20 (7): 1157-1169.

Bubeck, D.M., Fehr, W.R., Hammond, E.G. 1989. Inheritance of palmitic and stearic acid mutants of soybean. *Crop Sci*, 29 (3): 652-656.

Buhr, T., Shirley, S., Ebrahim, F., Xing, A., Zhou, Y., Mathiesen, M., Schweiger, B., Kinney, A., Staswick, P., Clemente, T. 2002. Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *Plant J*, 30 (2): 155-163.

Burton, J.W., Wilson, R.F., Brim, C.A. 1994. Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. *Crop Sci*, 34 (1): 313.

Byers, S.D., Laroche, A., Smith, K.C., Weselake, R.J. 1999. Factors enhancing diacylglycerol acyltransferase activity in microsomes from cell-suspension cultures of oilseed rape. *Lipids*, 34 (11): 1143-1149.

Cagliari, A., Margis-Pinheiro, M., Loss, G., Mastroberti, A.A., Mariath, J.E.deA., Margis, R. 2010. Identification and expression analysis of castor bean (*Ricinus communis*) genes encoding enzymes from the triacylglycerol biosythesis pathway. *Plant Sci*, 179: 499-509.

Cahoon, E.B. 2003. Genetic enhancement of soybean oil for industrial uses: prospects and challenges. *AgBioForm*, 6 (1&2): 11-13.

Cao, Y-Z., Huang, A.H.C. 1986. Diacylglycerol acyltransferase in maturing oil seeds of maize and other species. *Plant Physiol*, 82: 813-820.

Cao, Y-Z., Huang, A.H.C. 1987. Acyl coenzyme A preference of diacylglycerol acyltransferase from the maturing seeds of *Cuphea*, maize, rapeseed, and

canola. *Plant Physiol*, 84: 762-765.

Cases, S. Smith, S.J., Zheng, Y.-W., Myers, H.M., Lear, S.R., Sande, E., Novak, S., Collins, C., Welch, C.B., Lusis, A.J., Erickson, S.K. 1998. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc Natl Acad Sci USA*, 95: 13018-13023.

Chapman, K.D., Ohlrogge, J.B. 2012. Compartmentation of triacylglycerol accumulation in plants. *J Biol Chem*, 287 (4): 2288-2294.

Coleman, R.A., Lee, D.P. 2004. Enzymes of triacylglycerol synthesis and their regulation. *Prog Lipid Res*, 43: 134-176.

Dahlqvist, A., Ståhl, U., Lenman, M., Banaś, A., Lee, M., Sandager, L., Ronne, H., Stymne, S. 2000. Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc Natl Acad Sci*, 97 (12): 6487-6492.

Dornbos Jr., D.L., Mullen, R.E. 1992. Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. *J Am Oil Chem Soc*, 69 (3): 228-231.

Durrett, T.P., Benning, C., Ohlrogge, J. 2008. Plant triacylglycerols as feedstocks for the production of biofuels. *The Plant Journal*, 54: 593-607.

Dutton, H.J., Lancaster, C.R., Evans, C.D., Cowan, J.C. 1951. The flavor problem of soybean oil. VIII. Linolenic acid. *J Am Oil Chem Soc*, 28 (3): 115-118.

Eastmond, P.J. 2004. Glycerol-insensitive *Arabidopsis* mutants: *gli1* seedlings lack glycerol kinase, accumulate glycerol and are more resistant to abiotic stress. *Plant J*, 37: 617-625.

Eccleston, V.S., Harwood, J.L. 1990. Acylation reactions in developing avocado. In: *Plant Lipid Biochemistry, Structure and Utilization* (Quinn, P.J., Harwood, J.L., eds.). Portland Press, London, pp. 178-180.

Edlin, D.A.N., Kille, P., Wilkinson, M.D., Jones, H.D., Harwood, J.L. 2000. Morphological and metabolic changes in transgenic wheat with altered glycerol-3-phosphate acyltransferase or acyl-acyl carrier protein (ACP) thioesterase activities. *Biocheml Soc T*, 28 (6): 682-683.

Emanualsson, O. Nielsen, H., Brunak, S. von Heijne, G. 2000. Predicting subcellar localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, 300: 1005-1016. http://www.cbs.dtu.dk/services/TargetP/

Erickson, E.A., Wilcox, J.R., Cavins, J.F. 1988. Inheritance of altered palmitic acid percentage in two soybean mutants. *J Hered*, 79 (6): 465-468.

Fehr, W.R. 2007. Breeding for modified fatty acid composition in soybean. *Crop Sci*, 47 (S3): S72-S87.

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

Fehr, W.R., Welke, G.A., Hammond, E.G., Duvick, D.N., Cianzio, S.R. 1991a. Inheritance of elevated palmitic acid content in soybean seed oil. *Crop Sci*, 31 (6): 1522-1524.

Fehr, W.R., Welke, G.A., Hammond, E.G., Duvick, D.N., Canzio, S.R. 1991b. Inheritance of reduced palmitic acid content in seed oil of soybean. *Crop Sci*, 31 (1): 88-89.

Fehr, W.R., Welke, G.A., Hammond, E.G., Duvick, D.N., Canzio, S.R. 1992. Inheritance of reduced linolenic acid content in soybean genotypes A16 and A17. *Crop Sci*, 32 (4): 903-906.

Ferri, S.R., Toguri, T. 1997. Substrate specificity modification of the stromal glycerol-3-phosphate acyltransferase. *Arch Biochem and Biophys*, 337 (2): 202-208.

Finlayson, S.A., Dennis, D.T. 1980. NAD$^+$-specific glycerol-3-phosphate dehydrogenase from developing castor bean endosperm. *Arch Biochem Biophys*, 199 (1): 179-185.

Frentzen, M. 1993. Acyltransferases and triacylglycerols. In: *Lipid Metabolism in Plants* (Moore, Jr., T.S., ed.). CRC Press, Ann Arbor, MI, pp. 195-230.

Frentzen, M. 1998. Acyltransferases from basic science to modified seed oils. *Lipids*, 100 (4-5): 161-166.

Frentzen, M., Heinz, E., McKeon, T.A., Stumpf, P.K. 1983. Specificities of glycerol-3-phosphate acyltransferase and monoacylglycerol-3-phosphate acyltransferase from pea and spinach chloroplasts. *Eur J Biochem*, 129: 629-636.

Frentzen, M., Peterek, G., Wolter, F.P. 1994 Properties and subcellular localization of a plastidal *sn*-glycerol-3-phosphate acyltransferase of *Pisum sativum* L. expressed in *Escherichia coli*. *Plant Sci*, 96: 45-53.

145

Frentzen, M., Wolter, F.P. 1998. Molecular biology of acyltransferases involved in glycerolipid synthesis. In: *Plant Lipid Biosynthesis: Fundamentals and Agricultural Applications* (Harwood, J.L. ed). Society of Experimental Biology, Seminar Series 67. Cambridge University Press. pp. 247-272.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31: 3784-3788. http://www.expasy.org/

Geigenberger, P. 2003. Regulation of sucrose to starch conversion in growing potato tubers. *J Exl Bot*, 54 (382): 457-465.

Ghosh, S., Sastry, P.S. 1988. Triacylglycerol synthesis in developing seeds of groundnut (*Arachis hypogaea*): Pathway and properties of enzymes of *sn*-glycerol-3-phosphate formation. *Arch Biochem Biophys* 262 (2): 508-516.

Graef, G.L., Fehr, W.R., Hammond, E.G. 1985. Inheritance of three stearic acid mutants of soybean. *Crop Sci*, 25 (6): 1076-1079.

Griffiths, G., Stobart, A.K., Stymne, S. 1985. The acylation of *sn*-glycerol-3-phospate and the metabolismof phosphatidate in microsomal preparations from developing cotyledons of safflower *Catharnus tinctorius* seed. *Biochem J* 230: 379-388.

Hammond, E.G., Fehr, W.R. 1983. Registration of A5 germplasm line of soybean (Reg. No. GP44). *Crop Sci*, 23 (1): 192.

Hares, W., Frentzen, M. 1987. Properties of the microsomal acyl-coA: *sn*-1-acyl-glycerol-3-phosphate acyltransferase from spinach (*Spinacia oleracea* L.) leaves. *J Plant Physiol*, 131: 49-59.

Hayes, M.F., Fehr, W.R., Welke, G.A. 2002. Associatioin of elevated palmitate with agronomic and seed traits of soybean. *Crop Sci*, 42 (4): 1117-1120.

He, X., Chen, G.Q., Lin, J.-T., McKeon, T.A. 2004. Regulation of diacylglycerol acyltransferase in developing seeds of castor. *Lipids*, 39 (9): 865-871.

He, X., Turner, C., Chen, G.Q., Lin, J-T., McKeon, T.A. 2004. Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. *Lipids*, 39 (4): 311-318.

Heart and Stroke Foundation. 2008. Dietary fats, oils, and cholesterol. http://www.heartandstroke.on.ca/site/c.pvl3leNWJwE/b.3581947/k.470F/Dietary_fats_oils_and_cholesterol.htm.

Hildebrand, D.F., Li, R., Hatanaka, T. 2008. Genomics of soybean oil traits. In: *Genetics and Genomics of Soybean* (Stacey, G., ed.). Springer Science+Business Media, LLC, NY, pp. 185-209.

Hippmann, H., Heinz, E. 1976. Glycerol kinase in leaves. *Z. Pflanzenphysiol* 79: 408-418.

Hobbs, D.H., Lu, C., Hills, M.J. 1999. Cloning of a cDNA encoding diacylglycerol acyltransferase from *Arabidopsis thaliana* and its functional expression. *FEBS Lett*, 452: 145-149.

Howell, R.W., Cartter, J.L. 1958. Physiological factors affecting composition of soybean: II. Response of oil and other constituents of soybeans to temperature under controlled conditions. *Agron J*, 50: 664-667.

Hou, G. 2004. Inheritance and genetic mapping of fatty acids in soybean seed oil. *University of Guelph – Department of Plant Agriculture, Guelph,*

Ontario, Canada.

Hou, G., Ablett, G.R., Pauls, K.P., Rajcan, I. 2006. Environmental effects on fatty acid levels in soybean seed oil. *J Am Oil Chem Soc*, 83 (9): 759-763.

Huang, A.H. 1975. Enzymes of glycerol metabolism in the storage tissues of fatty seedlings. *Plant Physiol*, 55: 555-558.

Huang T.C., Yen T.Y., Huang Y.W., Wu M.L. 2004. *Glycine max* glycerol kinase sequence. EMBL Database.

Hymowitz, T. 1970. On the domestication of the soybean. *Econ Bot*, 24 (4): 408-421.

Hymowitz, T., Harlan, J.R. 1983. Introduction of soybeans to North America by Samuel Bowen in 1765. *Econ Bot*, 37 (4): 371-379.

Ichihara, K. 1984. *Sn*-glycerol-3-phosphate acyltransferase in a particulate fraction from maturing sunflower seeds. *Arch Biochem Biophys*, 232 (2): 685-689.

Ichihara, K. 1991. The action of phosphatidate phosphatase on the fatty-acid composition of safflower triacylglycerol and spinach glycerolipids. *Planta*, 183 (3): 353-358.

Ichihara, K., Murota, N., Fujii, S. 1990. Intracellular translocation of phosphatidate phosphatase in maturing safflower seeds: a possible mechanism of feedforward control of triacylglycerol synthesis of fatty acids. *Biochim Biophys Acta*, 1043 (3): 227-234.

Ichihara, K., Norikura, S., Fujii, S. 1989. Microsomal phosphatidate phosphatase in maturing safflower seeds. *Plant Physiol*, 90: 413-419.

Jain, R.K., Coffey, M., Lai, K., Kumar, A., MacKenzie, S.L. 2000. Enhancement of seed oil content by expression of glycerol-3-phosphate acyltransferase genes. *Biochem Soc T*, 28 (6): 958-961.

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

John, J., Bhattacharya, M., Turber, R.B. 2002. Characterization of polyurethane foams from soybean oil. *J Applied Polymer Sci*, 86: 3097-3107.

Joyard, J., Douce, R. 1979. Characterization of phosphatidate phosphohydrolase activity associated with chloroplast envelope membranes. *FEBS Lett*, 102 (1): 147-150.

Kachroo, A.P., Kachroo, P., Hildebrand, D.F. http://www.uky.edu/SoyLipidGenes accessed June 23, 2012.

Kang, L., Li, J., Zhao, T. Xiao, F., Tang, X., Thilmony, R., He, SY., Zhou, J-M. 2003. Interplay of the *Arabidopsis* nonhost resistance gene *NHO1* with bacterial virulence. *P Natl A Sci USA*, 100 (6): 3519-3524.

Katavic, V., Reed., D.W., Taylor, D.C., Giblin, E.M., Barton, D.L., Zou, J., MacKenzie, S.L., Covello, P.S., Kunst, L. 1995. Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in *Arabidopsis thaliana* affecting diacylglycerol acyltransferase activity. *Plant Physiol*, 108: 399-409.

Kaup, M.T. Froese, C.D., Thompson, J.E., 2002. A role for diacylglycerol

acyltransferase during leaf senescence. *Plant Physiol*, 129: 1616-1626.

Kennedy, E.P. December 1961. Biosynthesis of complex lipids. *Proc Fed Amer Soc Experi Biol*, 20 (1): 934-940.

Kim, H.U., Huang, A.H.C. 2004. Plastid lysophosphatidyl acyltransferase is essential for embryo development in Arabidopsis. *Plant Physiol*, 134: 1206-1216.

Kim, H.U., Lee, K.-R., Go, Y.S., Jung, J.H., Suh, M.-C., Kim, J.B. 2011. Endoplasmic reticulum-located PDAT1-2 from castor bean enhances hydroxy fatty acid accumulation in transgenic plants. *Plant Cell Physiol*, 52 (6): 983-993.

Kinney, A.J. 1998. Production of specialised oils for industry. In: *Plant Lipid Biosynthesis: Fundamentals and Agricultural Applications* (Harwood, J.L., ed.). Society for Experimental Biology Seminar Series 67, Cambridge University Press, NY, pp. 273-285.

Kinney, A.J., Knowlton, S. 1998. Designer oils: the high oleic acid soybean. In: *Genetic modification in the food industry: a strategy for food quality improvement* (Roller, S., Harlander, S., eds.). Blackie Academic & Professional, London, pp.193-213.

Kirsch, T., Gerber, D.W., Byerrum, R.U., Tolbert, N.E. 1992. Plant dihydroxyacetone, phosphate reductases: Purification, characterization, and localization. *Plant Physiol*, 100: 352-359.

Kleffe, J., Hermann, K., Vahrson, W., Wittig, B., Brendel, V. 1996. Logitlinear models for the prediction of splice sites in plant pre-mRNA sequences.

Nucl Acids Res, 24 (23): 4709-4718.

Knutzon, D.S., Lardizabal, K.D., Nelsen, J.S., Bleibaum, J.L., Davies, H.M., Metz, J.G. 1995. Cloning of a coconut endosperm cDNA encoding a 1-acyl-*sn*-glycerol-3-phosphate acyltransferase that accepts medium-chain-length substrates. *Plant Physiol*, 109: 999-1006.

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

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

Lassner, M.W., Levering, C.K., Davies, H.M., Knutzon, D.S. 1995. Lysophosphatidic acid acyltransferase from meadowfoam mediates insertion of erucic acid at the *sn*-2 position of triacylglycerol in transgenic rapeseed oil. *Plant Physiol*, 109:1389-1394.

Lau, T.E., Rodriguez, M.A. 1996. A protein tyrosine kinase associated with the ATP-depemdemt inactivation of adipose diacylglycerol acyltransferase. *Lipids*, 31: 277-283.

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

Livak, K.J., Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods*, 25: 402-408.

List, G.R., Mounts, T.L., Orthoefer, F., Neff, W.R. 1997. Effect of interesterification on the structure and physical properties of high-stearic acid soybean oil. *J Am Oil Chem Soc*, 74 (3): 327-329.

Lu, C.L., de Noyer, S.B., Hobbs, D.H., Kang, J., Wen, Y., Krachtus, D., Hills, M.J. 2003. Expression pattern of diacylglycerol acyltransferase-1, an enzyme involved in triacylglycerol biosynthesis, in *Arabidopsis thaliana*. *Plant Mol Biol*, 52: 31-41.

Lundeen, P.O., Fehr, W.R., Hammond, E.G., Cianzio, S.R. 1987. Association of alleles for high stearic acid with agronomic characters of soybean. *Crop Sci*, 27 (6): 1102-1105.

Lung, S.-C., Weselake, R.J. 2006. Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. *Lipids*, 41 (12): 1073-1088.

Lyons, J.M., Wheaton, T.A., Pratt, H.K. 1964. Relationship between the physical nature of mitochondrial membranes and chilling sensitivity in plants. *Plant Physiol*, 39 (2): 262-268.

Maisonneuve, S., Bessoule, J.-J., Lessire, R., Delseny, M. Roscoe, T.J. 2000. Mutagenesis of a plastidial lysophosphatidic acid acyltransferase. *Biochem Soc T*, 28 (6): 961-964.

Malherbe, A., Block, M.A., Douce, R., Joyard, J. 1995. Solubilization and biochemical properties of phosphatidate phosphatase from spinach

chloroplast envelope membranes. *Plant Physiol Biochem*, 33 (2): 149-161.

Martin, B.A., Wilson, R.F. 1983. Properties of diacylglycerol acyltransferase from spinach leaves. *Lipids*, 18 (1): 1-6.

Martin, B.A., Wilson, R.F. 1983. Subcellular localization of triacylglycerol synthesis in spinach leaves. *Lipids*, 19 (2): 117-121.

McNaughton, A.J.M. 2007. The subcellular targeting of diacylglycerol acyltransferase 1 (DGAT1): a putative dual-localized endoplasmic reticulum-chloroplast protein. *Wilfrid Laurier University – Biology Department, Waterloo, Ontario, Canada*.

Nakamura, Y., Tsuchiya, M., Ohta, H. 2007. Plastidic phosphatidic acid phosphatases identified in a distinct subfamily of lipid phosphate phosphatases with prokaryotic origin. *J Biol Chem*, 282 (39): 29013-29021.

Narvel, J.M., Fehr, W.R., Ininda, J., Welke, G.A., Hammond, E.G., Duvick, D.N., Cianzio, S.R. 2000. Inheritance of elevated palmitate in soybean seed oil. *Crop Sci*, 40 (3):635-639.

Nelson, D.L., Cox, M.M. 2005. *Lehninger Principles of Biochemistry*, 4th ed. W.H. Freeman and Company. p. 804.

O'Hara, P., Slabas, A.R., Fawcett, T. 2002. Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. *Plant Physiol*, 129 (1): 310-320.

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

Pantalone, V.R., Walker, D.R., Dewey, R.E., Rajcan, I. 2004. DNA marker-

assisted selection for improvement of soybean concentration and quality. In: *Legume Crop Genetics* (Wilson, R.F., Stalker, H.T., Brummer, E.C., eds.). AOCS Press, Champaign, Illinois, pp. 283-311.

Pantalone, V.R., Wilson, R.F., Novitsky, W.P., Burton, J.W. 2002. Genetic regulation of elevated stearic acid concentration in soybean seed oil. *J Am Oil Chem Soc*, 79 (6): 549-553.

Perry, H.J., Bligny, R., Gout, E., Harwood, J.L. 1999. Changes in Kennedy pathway intermediates associated with increased triacylglycerol synthesis in oil-seed rape. *Phytochemistry*, 52: 799-804.

Primomo, V.S., Falk, D.E., Ablett, G.R., Tanner, J.W., Rajcan, I. 2002. Inheritance and interaction of low palmitic and low linolenic soybean. *Crop Sci*, 42 (1): 31-36.

Phytozome v.7.0. www.phytozome.net. University of California.

Rahman, S.M., Kinoshita, T., Anai, T., Arima, S., Takagi, Y., 1998. Genetic relationships of soybean mutants for different linolenic acid contents. *Crop Sci*, 38 (3): 702-706.

Rahman, S.M., Takagi, Y. 1997. Inheritance of reduced linolenic acid content in soybean seed oil. *Theor Appl Genet*, 94 (3-4): 299-302.

Rahman, S.M., Takagi, Y., Kinoshita, T. 1996. Genetic control of high oleic acid content in the seed oil of two soybean mutants. *Crop Sci*, 36 (5): 1125-1128.

Rahman, S.M., Takagi, Y., Kumamaru, T. 1995. Low linolenate sources at the *Fan* locus in soybean lines M-5 and IL-8. *Breed Sci*, 46 (2): 155-158.

Ramli, U. S., Salas, J. J., Quant, P. A., Harwood, J. L. 2005. Metabolic control analysis reveals an important role for diacylglycerol acyltransferase in olive but not in oil palm lipid accumulation. *FEBS J,* 272: 5764-5770.

Rao, S.S., Hildebrand, D. 2009. Changes in oil content of transgenic soybeans expressing the yeast *SLC1* gene. *Lipids*, 44: 945-951.

Reinprecht, Y., Luk-Labey, S.-Y., Larsen, J., Poysa, V.W., Yu, K., Rajcan, I., Ablett, G.R., Pauls, K.P. 2009. Molecular basis of the linolenic acid trait in soybean EMS mutant line RG10. *Plant Breeding*, 128: 253-285.

Rennie, B.D, Tanner, J.W. 1989a. Genetic analysis of low linolenic acid levels in the soybean line PI123440. *Soybean Genet Newsl*, 16: 25-26.

Rennie, B.D., Tanner, J.W. 1989b. Fatty acid composition of oil from soybean seeds grown at extreme temperatures. *J Am Oil Chem Soc*, 66 (11): 1622-1624.

Rennie, B.D., Tanner, J.W. 1991a. New allele at the *Fan* locus in the soybean line A5. *Crop Sci*, 31 (2): 297-301.

Rennie, B.D., Tanner, J.W. 1991b. Fatty acid composition of oil from soybean leaves grown at extreme temperatures. *J Am Oil Chem Soc*, 68 (2): 106-108.

Rivera, C.M., Penner, D. 1978. Rapid changes in soybean root membrane lipids with altered temperature. *Phytochemistry*, 17 (8): 1269-1272.

Ross, A.J., Fehr, W.R., Welke, G.A., Hammond, E.G., Cianzio, S.R. 2000. Agronomic and seed traits of 1%-linolenate soybean genotypes. *Crop Sci*, 40 (2): 383-386.

Routaboul, J-M., Benning, C., Bechtold, N., Caboche, M., Lepiniec, L. 1999. The *TAG1* locus of *Arabidopsis* encodes for a diacylglycerol acyltransferase. *Plant Physiol Biochem*, 37 (11): 831-840.

Rozen, S., Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz S., Misener S., eds.). Humana Press, Totowa, NJ, pp. 365-386.

Sadava, D., Moore, K. 1987. Glycerol metabolism in higher plants: glycerol kinase. *Biochem and Biophys Res Co*, 143 (3): 977-983.

Sain, M. 2006. Ontario bio-car initiative ORF-RE-063 research proposal to the ministry of research and innovation, research branch.

Samii-Saket, G., Boersma, J.G., Ablett, G.R., Falk, D.E., Fletcher, R., Rajcan, I. 2011. Single soybean seed NMR calibration for oil measurement using commercial cooking oils. *J Am Oil Chem Soc*, 88 (11): 1795-1798.

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

Schnebly, S.R., Fehr, W.R., Welke, G.A., Hammond, E.G., Duvick, D.N. 1994. Inheritance of reduced and elevated palmitate in mutant lines of soybean. *Crop Sci*, 34 (4): 829-833.

Schuler, M.A. 2008. Splice site requirements and switches in plants. In: *Nuclear pre-mRNA Processing in Plants: Current Topics in Microbiology and Immunology 326* (Reddy, A.S.N., Golovkin, M., eds.). Springer, Berlin, pp.39-59.

Settlage, S.B., Kwanyuen, P., Wilson, R.F. 1998. Relation between diacylglycerol acyltransferase activity and oil concentration in soybean. *J Am Oil Chem Soc*, 75 (7): 775-781.

Shen, N., Fehr, W.R., Johnson, L., White, P. 1997. Oxidative stabilities of soybean oils with elevated palmitate and reduced linolenate contents. *J Am Oil Chem Soc*, 74 (3): 299-302.

Shen, W., Wei, Y., Dauk, M., Tan, Y., Taylor, D.C., Selvaraj, G., Zou, J. 2006. Involvement of a glycerol-3-phosphate dehydrogenase in modulating the NADH/NAD$^+$ ratio provides evidence of a mitochondrial glycerol-3-phosphate shuttle in *Arabidopsis*. *Plant Cell*, 18: 422-441.

Shockey, J.M., Gidda, S.K., Chapital, D.C., Kuan, J-C., Dhanoa, P.K., Bland, J.M., Rothstein, S.J., Mullen, R.T., Dyer, J.M. 2006. Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell*, 18 (9): 2294-2313.

Shurtleff, W., Aoyagi, A. 2010. *History of Soybeans and Soyfoods in Canada*

(1831 – 2010). Soyinfo Center, Lafayette, CA. www.soyinfocenter.com

Slabas, A.R., Hanley, Z., Schierer, T., Rice, D., Turnbull, A., Rafferty, J., Brown, A. 2001. Acyltransferases and their role in the biosynthesis of lipids – opportunities for new oils. *J Plant Physiol*, 158: 505-513.

Slabas, A.R., Kroon, J.T.M., Scheirer, T.P., Gilroy, J.S., Hayman, M., Rice, D.W., Turnbull, A.P., Rafferty, J.B., Fawcett, T., Simon, W.J. 2002. Squash glycerol-3-phosphate (1)-acyltransferase alteration of substrate selectivity and identification of arginine and lysine residues important in catalytic activity. *J Biol Chem*, 277 (46): 43918-43923.

Ståhl, U., Carlsson, A.S., Lenman, M., Dahlqvist, A., Huang, B., Banaś, W., Banaś, A., Stymne, S. 2004. Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from Arabidopsis. *Plant Physiol*, 135 (3): 1324-1335.

Statistics Canada. 2010. Field Crop Reporting Series. Catalogue no. 22-002-X, Vol. 89, No. 8.

Stojsin, D., Luzzi, B.M., Ablett, G.R., Tanner, J.W. 1998a. Use of gene substitution values to quantify partial dominance in low palmitic acid soybean. *Crop Sci*, 38 (6): 1437-1441.

Stojsin, D., Luzzi, B.M., Ablett, G.R., Tanner, J.W. 1998b. Inheritance of low linolenic acid level in the soybean line RG10. *Crop Sci*, 38 (6): 1441-1444.

Stoltzfus, D.L., Fehr, W.R., Welke, G.A., Hammond, E.G., Cianzio, S.R. 2000a. A *fap5* allele for elevated palmitate in soybean. *Crop Sci*, 40 (3): 647-650.

Stoltzfus, D.L., Fehr, W.R., Welke, G.A., Hammond, E.G., Cianzio, S.R. 2000b. A

fap7 allele for elevated palmitate in soybean. *Crop Sci*, 40 (6): 1538-1542.

Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28: 1102-1104. http://www.bioinformatics.org/sms/, current as of June 12, 2011.

Takagi, Y., Rahman, S.M. 1996. Inheritance of high oleic acid content in the seed oil of soybean mutant M23. *Theor Appl Genet*, 92 (2): 179-182.

Taylor, D.C., Zhang, Y., Kumar, A., Francis, T., Giblin, E.M., Barton, D.L., Ferrie, J.R., Laroche, A., Shah, S., Zhu, W., Snyder, C.L., Hall, L., Rakow, G., Harwood, J.L., Weselake, R.J. 2009. Molecular modification of triacylglycerol accumulation by over-expression of *DGAT1* to produce canola with increased seed oil content under field conditions. *Botany*, 87: 533-543.

Tian, Z., Wang, X, Lee, R., Li, Y., Specht, J.E., Nelson, R.L., McClean, P.E., Qiu, L., Ma, J. May 11, 2011. Artificial selection for determinate growth habit in soybean. *P Natl A Sci USA*, 107 (19): 8563-8568.

Triki, S., Hamida, J.B., Mazliak, P. 2000. Diacylglycerol acyltransferase in maturing sunflower seeds. *Biochem Soc T*, 28 (6): 689-692.

Tusnády, G.E., Simon, I. 1998. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol*, 283: 489-506.

Tusnády, G.E., Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 17: 849-850.

http://www.enzim.hu/hmmtop/index.html

Usuka, J., Brendel, V. 2000. Gene structure prediction by spliced alignment of genomic DNA with protein sequences: Increase accuracy by differential splice site scoring. *J Mol Biol*, 297 (5): 1075-1085.

Usuka, J., Zhu, w., Brendel, V. 2000. Optimal spliced alignment of homologous cDNA to a genomic DNA template. *Bioinformatics*, 16 (3): 203-211.

Vigeolas, H., Waldeck, P., Zank, T., Geigenberger, P. 2007. Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotech J*, 5: 431-441.

Wang, H.-W., Zhnag, J.-S., Gai, J.-Y., Chen, S.-Y. 2006. Cloning and comparative analysis of the gene encoding diacylglycerol acyltransferase from wild type and cultivated soybean. *Theor Appl Genet*, 112: 1086-1097.

Weselake, R.J. 2005. Storage lipids. In: *Plant Lipids: Biology, Utilisation and Manipulation* (Murphy, D.J. ed.). Blackwell Publishing Ltd., Oxford, pp.162-225.

Weselake, R.J., Nykiforuk, C.L., Laroche, A., Patterson, N.A., Wiehler, W.B., Szarka, S.J., Moloney, M.M., Tari, L.W., Derekh, U. 2000. Expression and properties of diacylglycerol acyltransferase from cell-suspension cultures of oilseed rape. *Biocheml Soc T*, 28 (6): 684-686.

Wilcox, J.R., Cavins, J.F. 1985. Inheritance of low linolenic acid content of the seed oil of a mutant in *Glycine max*. *Theor Appl Genet*, 71 (1): 74-78.

Wilcox, J.R., Cavins, J.F. 1987. Gene symbol assigned for linolenic acid mutant

in soybean. *J Hered*, 78 (6): 410.

Wolf, R.B., Cavins, J.F., Kleiman, R., Black, L.T. 1982. Effect of temperature on
soybean seed constituents: oil, protein, moisture, fatty acids, amino acids,
and sugars. *J Am Oil Chem Soc*, 59 (5): 230-232.

Xu, C., Yu, B., Cornish, A.J., Froehlich, J.E., Benning, C. 2006.
Phosphatidylglycerol biosynthesis in chloroplasts of Arabidopsis mutants
deficient in acyl-ACP glycerol-3-phosphate acyltransferase. *Plant J*, 47:
296-309.

Yu, B., Wakao, S., Fan, J., Benning, C. 2004. Loss of plastidic lysophosphatidic
acid acyltransferase causes embryo-lethality in *Arabidopsis*. *Plant Cell
Physiol*, 45 (5): 503-510.

Zhang, F.-Y., Yang, M.-F., Xu, Y.-N. 2005. Silencing of *DGAT1* in tobacco
causes a reduction in seed oil content. *Plant Sci*, 169: 689-694.

Zhang, M., Fan, J., Taylor, D.C., Ohlrogge, J.B. 2009. *DGAT1* and *PDAT1* have
overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are
essential for normal pollen and seed development. *Plant Cell*, 21: 3885-
3901.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000. A greedy algorithm
for aligning DNA sequences, *J Comput Biol*, 7 (1-2): 203-14.
BLASTN2.2.25+

Zheng, P., Allen, W.B., Roesler, K., Williams, M.E., Zhang, S., Li, J., Glassman,
K., Ranch, J., Nubel, D., Solawetz, W., Bhattramakki, D., Llaca, V.,
Deschamps, S., Zhong, G.Y., Tarczynski, M.C., Shen, B. 2008. A

161

phenylalanine in DGAT is a key determinant of oil content and
composition in maize. *Nat Genet*, 40 (3): 367-372.

Zheng, Z., Xia, Q., Duak, M., Shen, W., Selvaraj, G., Zou, J. 2003. Arabidopsis
AtGPAT1, a member of the membrane-bound glycerol-3-phosphate
acyltransferase gene family, is essential for tapetum differentiation and
male fertility. *Plant Cell*, 15: 1872-1887.

Zou, J., Katavic, V., Giblin, E.M., Barton, D.L., MacKenzie, S.L., Keller, W.A., Hu,
X., Taylor, D.C. 1997. Modification of seed oil content and acyl
composition in the Brassicaceae by expression of a yeast *sn*-2
acyltransferase gene. *Plant Cell*, 9: 909-923.

Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G., Taylor, D.C. 1999. The
Arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol
acyltransferase gene. *Plant J*, 19 (6): 645-653.

APPENDIX A:

PRIMERS FOR GENE SEQUENCING

Table A1: Sequencing primers for GK. Primers used in sequencing GK; Bps, the approximate base pairs covered in that section of the gene based on GK on Chromosome 07 sequence obtained through Phytozome (Phytozome v.7.0).

Gene Section	Bps	Primer Name	Annealing Temp	Sequence (5' → 3')
1 forward	1 – 2715	Gkf2	58°C	CGCGATCGACCAAGGAAC
		Gkr2		AGCAAGATCAGCTAAGTCGAAA
1 reverse	1 – 1100	Gkf2	53°C	CGCGATCGACCAAGGAAC
		Gkr6		TGTGTTTGGATTCATGTTGGA
2 forward	800 – 2715	Gkf8	48°C	GGGATCACTTTTCCCGTATG
		Gkr2		AGCAAGATCAGCTAAGTCGAAA
2 reverse	800 – 1900	Gkf8	52°C	GGGATCACTTTTCCCGTATG
		Gkr9		CACTGCAGCTCCAGCAATAG
3 forward	1700 – 2715	Gkf10	57°C	AAAGGGGAGGCTAAGAGCAC
		Gkr2		AGCAAGATCAGCTAAGTCGAAA
3 reverse	1 – 2715	Gkf2	58°C	CGCGATCGACCAAGGAAC
		Gkr2		AGCAAGATCAGCTAAGTCGAAA

Table A2: Sequencing primers for G3PAT. Primers used in sequencing glycerol-3-phosphate acyltransferase (G3PAT); Bps, the approximate base pairs covered in that section of the gene based on G3PAT on Chromosome 09 sequence obtained through Phytozome (Phytozome v.7.0).

Gene Section	Bps	Primer Name	Annealing Temp	Sequence (5' → 3')
1	5 – 378	09GPIIF1	55°C	GGCTTTGATCTTTGAACAAT
		09GPIIR1		AGAAGTGAAGTGAACCTTGC
2	314 – 637	09GPVIF1	53°C	ATCTTCTACTGCGTCCTACT
		09GPVIR1		AAATCTCTACAAACATCACATAG
2aa	550 – 750	GP2aaF2	49°C	TTTAGATGCTTACTCAAGGAGACC
		GP2aaR2		TTATGCCACCAATTTAACGGTAT
3	596 – 851	09GPVIIF1-2	53°C	GCTTCTCTCGTATCTTTATC
		09GPVIIR1-2		TTTTCACTCTATCCTTTCTTA
4	702 – 1116	09GPVIF2	56°C	CTGGTTGTTGGAAATAAATAC
		09GPVIR2		TCTTCGGCAAATAAATAGTC
5	1080 – 1311	09GPIIIF2	50°C	GACTGAGGTTTGACATTTAGAA
		09GPIIIR2		AAGGCTAGTAAAATACAGGTAACA
6	1292 – 1767	09GPIVF2	51°C	CATTTTATCTCTTCTCTGTTTT
		09GPIVR2		TAACTTGATTGCTTGTGAG
7	1738 – 2238	09GPIIF3	54°C	TCTCTGCCTGCTAATGTTG
		09GPIIR3		ATACAATCTCCTTTGCCTTG
7a	1600 – 2300	GP5F2	56°C	ATTTCTTTACTTCATGTGTTGATTCG
		GP5R2		CCTGAAATATGATTGCCCACA
8	2093 – 2610	09GPIIIF3	50°C	CTGCCTTTTTCATGCTTTAC
		09GPIIIR3		TCGACTAAAGGACGGATATAAT
9	2295 – 3117	09GPIIIF4	52°C	AGGCAAAGGAGATTGTATTG
		09GPIIIR4		ACCAAGCAAAAAGTCAAAGA
9a	2500 – 3200	GP7aF2	58°C	CCACACCACAAAGCAAAGAG
		GP7aR2		GCCCCCTGGAATCTAAATGT
10	3118 – 3225	09GPVIF4-1	56°C	GAGTCACCACCTTCCAATAG
		09GPVIR4-1		CCAAGACACTACAAACCATAGA
11	3139 – 3580	09GPVF2-1	56°C	GAGTCACCACCTTCCAATAG
		09GPVR2-1		GGTTAACCACGACCAACTT
12	3467 – 3770	09GPVIIIF1	55°C	CCTGTATTGGTGTCTACTGTGA
		09GPVIIIR1		GCATAACTCCAATGAGGAAC
13	3758 – 4089	09GPVIIF2-2	55°C	TGGTTTGTAGTGTCTTGGAA
		09GPVIIR2-2		GAGAAGGATGCCCAGTAAC
14	4014 – 4916	09GPIVF5	55°C	ACACAATGGGGTCTCTATCA
		09GPIVR5		ATTGCATAATTTGGGTCAAC
15	4729 – 5125	09GPVF3-2	53°C	ACAAAGTAGGGTGAGGGTAGT
		09GPVR3-2		GCAGCAGGATAAGAAACATC
16	4940 – 5847	09GPIIIF6	50°C	AGTGCAGATTGTGGAACATT
		09GPIIIR6		TGGTTAGTTCGCCTAAAGAA

Table A2 continued

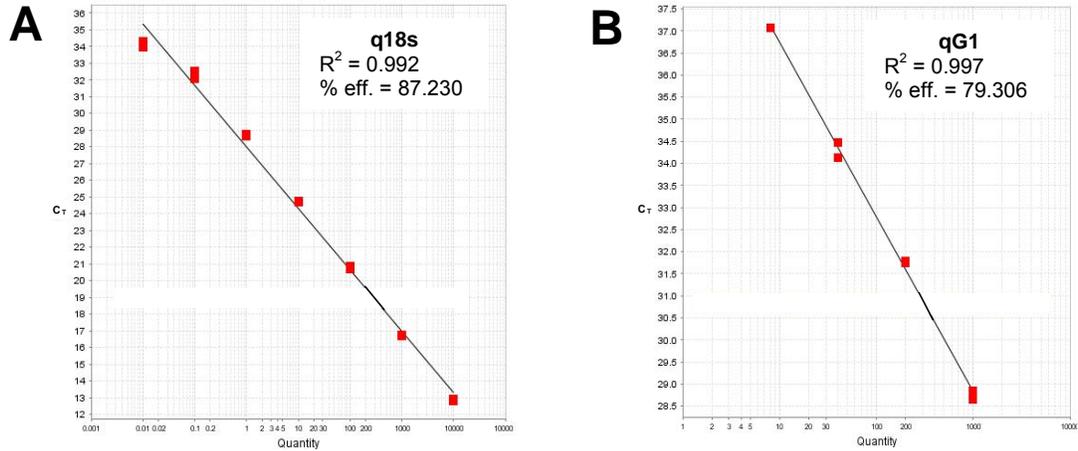| 17 | 5821 – 6636 | 09GPIIIF7 | 48°C | ATAGAGATGAAAAGAAATGC |
		09GPIIIR7		ACCATAAAGAATGAGAGTTG
17a	6100 – 6400	GP12aF2	58°C	ACAGAATTGGGCATGTTTACTTT
		GP12aR2		CCAAGCTGTCCCCTAGTCAA
18	6581 – 7459	09GPIVF6	52°C	GTCGGGGAAGATTGATTTA
		09GPIVR6		AACCAGCACAAGACCATAAC
19	7446 – 8062	09GPIIIF9	49°C	TGGGTTCTCTGTATTATCTGC
		09GPIIIR9		AAACAAGGCAAACTCTCG
19a	7400 – 7700	GP14aF1	49°C	ATTCCATGCTTATTGGGTTCTC
		GP14aR1		CTTGTGACTCCTTTTATCCCACA
20	7931 – 8446	09GPIVF7	49°C	TATTCAGGCACCCTTTGATA
		09GPIVR7		CTCTTTCTCTCTCATTGTTGTAAA
21	8260 – 9046	09GPIIIF10	53°C	TCCTGAAAAGGTAAGTTGAGTT
		09GPIIIR10		TTGTAATGAAGAAATGCCAAT
22	8899 – 9261	09GPIIF8	50°C	TTCTTTTACTTCTTCAACTTCA
		09GPIIR8		ACAAGCCATCCATCTAATAA
23	9154 – 9987	09GPIIIF11	49°C	TCTGATGCAGGTACAATCCT
		09GPIIIR11		ACAGATTCCAGATGAATTGAA
24	9830 – 10316	09GPIVF8-1	50°C	CCCCACTGACTTTATTTTCA
		09GPIVR8-1		CTTGCATTAGCATTTGGAC
24a	9700 – 10316	GP19aF2	56°C	AACGGTAAGAAGTAGATGGAAGC
		GP19aR2		TTGCATTAGCATTTGGACCAG

Table A3: Sequencing primers for LPAAT. Primers used in sequencing lysophosphatidic acid acyltransferase (LPAAT); Bps, the approximate base pairs covered in that section of the gene based on LPAAT on Chromosome 19 sequence obtained through Phytozome (Phytozome v.7.0).

Gene Section	Bps	Primer Name	Annealing Temp	Sequence (5' → 3')
1	1 – 1800	19Lf68	51°C	TGTTAACGACAATTCGATTCCA
		19Lr68		ATCCATACCAATCATTCCAAGC
1.5	1700 – 2000	19Lf76	57°C	AATCAAGCTACTCTAGCAGGTTTTTAT
		19Lr76		CCTCCCAATCCTAACCAATC
2	1800 – 3600	19Lf71	51°C	GCTTGGAATGATTGGTATGGAT
		19Lr71		CAATCAGCATTTGAAGCAAAAA
2.5	2700 – 3400	19Lf78	49°C	TATTTTTGTTTATTTAAATGGGATGAC
		19Lr78		TCATTTTGAAACTATAAGTCAGAAGCA
3	3600 – 5400	19Lf37	55°C	GGTTGTCTTGGCAGCACTCTA
		19Lr37		ATGTTAAATTTCCGCCCCTTGTA
3a	4500 – 5200	19Lf48	55°C	GGGATACATGCTTGTGTTTGTC
		19Lr48		TGTCTTTGACTACATCTCTTGGTCA
3.5	5100 – 5500	19Lf50	52°C	TTGATTTGAGAAATGTTTTGAGGA
		19Lr50		TATGTTAAATTTCCGCCCCTTG
4	5400 – 7200	19Lf40	49°C	GGTTTTATTGGCTCAAGGTTAGATT
		19Lr40		ACAGTCCTCAAATCCATCTGTAAAA
4a	5800 – 6100	19Lf53	50°C	CATCCGAGAAAATTCAGGAAAG
		19Lr53		GTATCCACCGGTCAGTAATTGG
4.5	6900 – 7300	19Lf56	54°C	CCCCTAACTTATCTCAACTCCTCA
		19Lr56		CACTAGTGAGACTCGCTGGAAA
5	7200 – 9000	19Lf41	53°C	GGATTTGAGGACTGTAGAAGTTTCA
		19Lr41		CCCCAATCTACTTTTCTTACCAAAT
5a	7600 – 8500	19Lf61	56°C	TCAGTGTCAGCATTTCGTTGTA
		19Lr61		GGAAAGTCAATCAAATCACAAATG
5.5	8600 – 9100	19Lf64	58°C	CGTGCCTTTATTTGTGCTTAAT
		19Lr64		AAATTCTTCTGATAGAACTGGATCG
6	9000 – 10544	19Lf45	48°C	AGAGAGTTCCCACTGTATCATTCTG
		19Lr45		ATTTTTCTAGGCCATCCTAGATTGT
6a	9400 – 9700	19Lf65	51°C	GGGATTGGATGGCTCTATTATG
		19Lr65		AACTGTGCAAGTAGAACATTGTG

Table A4: Sequencing primers for DGAT1. Primers used in sequencing diacylglycerol acyltransferase 1 (DGAT1); Bps, the approximate base pairs covered in that section of the gene based on DGAT1 on Chromosome 13 sequence obtained through Phytozome (Phytozome v.7.0).

Gene Section	Bps	Primer Name	Annealing Temp	Sequence (5' → 3')
1	15 – 577	13DGIIIF1-1	48°C	GCTCGCTCGGTCTTCTTTT
		13DGIIIR1-1		AAAACCTGATCCACCAATTCA
2	61 – 870	13DGIF1	50°C	CGATGAGCCTGAAACTGTA
		13DGIR1		GCTAAACCAAAAGCCAGAT
3	869 – 1821	13DGIIF1	57°C	TTTATGACCTTGATTTTGAC
		13DGIIR1		AACTGATTTGGTAAGTGCT
4	1796 – 2400	13DGIF3	54°C	AATGCTATTTGCTTGTGTTGTA
		13DGIR3		AAGGATAGTCCATGTTCAGAGTA
5	2311 – 2714	13DGIIIF2-2	48°C	TTACTTCCATCCTGTGTTCC
		13DGIIIR2-2		TCATACTCTGTTTCTGGGTTC
6	2443 – 3692	13DGIF4	57°C	CTCTGCCCGATACTCTGA
		13DGIR4		AAATCCCATAACTCCTGTAAAT
6a	2750 – 3450	D6aF1	57°C	GACAACAAAGAAAAGGGGAAAA
		D6aR1		ATTTTATGCCGGGGAGTCAT
7	3320 – 4154	13DGIIF3	55°C	AGTCTATGTGTTCCTCAATCTG
		13DGIIR3		TACAAAAGTCAAGGCATCAA
8	3828 – 4640	13DGIF5	58°C	TTACAGGAGTTATGGGATTT
		13DGIR5		GAAGAAGCTCTGCCAATA
9	4522 – 5049	13DGIIF4	50°C	GACTATTGGTGTCTTGAGTGTA
		13DGIIR4		TTCTCTTTCTTCCGTTGTTA
10	4795 – 5710	13DGIF6	57°C	CTACCAGGATTGGTGGAAT
		13DGIR6		GAAGGCAATTAAAAGAGCAA
11	5702 – 6398	13DGIVF1-2	54°C	TTGCACTAATTGTATGGTTTGATA
		13DGIVR1-2		TTTCCTCTTAATCCCCGTAAT
11a	6000 – 6400	D11aF2	58°C	GGTTGTTCTGCTTGAGTGTTTG
		D11aR1		AATTCTCCTCCAAGGGTTTCC
12	6332 – 7233	13DGIIF6	50°C	CTGTTCCTTGCCACATATTC
		13DGIIR6		GTCAAGTTTGCCTTTCCTATT
13	7055 – 7833	13DGIIF7-2	48°C	TACTTTCTTCTCAAAACACG
		13DGIIR7-2		TTGCCTTTCCTATTCATTA

APPENDIX B:

PRIMERS FOR GENE EXPRESSION

A

q18s
$R^2 = 0.992$
% eff. = 87.230

C_T

Quantity

B

qG1
$R^2 = 0.997$
% eff. = 79.306

C_T

Quantity

C

Gene Name	Chrom No.	Primer Name	Forward/ Reverse	Sequence
G3PAT	09	qG1f	Forward	GATGTCAAATCACCAAACTGAAGCT
		qG1r	Reverse	CCGGCTACATAGGTCATGCTTT
18S	**	q18Sf	Forward	CGGCTACCACATCCAAGGAA
		q18Sr	Reverse	GCTGGAATTACCGCGGCT

Figure B1: Standard curve and primers used to measure expression of G3PAT. Standard curves for each set of primers used for quantitative real-time PCR (qRT-PCR). Correlation coefficients(R^2) values and the percentage of efficiency of each primer pair are indicated. A – primer pair used for 18s, the housekeeping gene for qRT-PCR; B – primer pair used for glycerol-3-phosphate acyltransferase (G3PAT); C – Primers used for quantitative real-time polymerase chain reaction. Chrom. No. is the Chromosome number the copy of the gene studied is located on, based on Phytozome (Phytozome v.7.0). G3PAT = glycerol-3-phosphate acyltransferase; 18S = the 18S subunit of ribosomal RNA. ** = sequence and information for primers from Li *et al.*, 2010.

170

ATGAGCAGGACCGGTTCTTCAGCTTACTACTGGGTGGCAGACGCCACCCCT
CCGCTCCGTAATAGTAAGACTACTATGTTTATGCTCTCCACGCCGCCAACAA
CCACATTCTTCGTTACGCCAACTCCTAGGGTTCCTTTTCTCTCTTCATCAAAA
CCTTCTTCCGTTTCTCTGTTGCGATCTTCTACTGCGTCCTACTCGCCGTGTT
GCTCCTCCTCCTCCATCACTCCCAAGGTTAAATCCAACGATAACAACAA
CTGCTACCTCGTCTCCGCTAAACATTCTCCCGCTAACACGTCCGCTTCGGTT
TCTTCACGCACCTTCCTCAACGCGAAGAACGAGCAA**GA**GCTTCTAGCTGGA
ATCAGGAAAGAAGTAGAAGCTGGATCTCTGCCTGCTAATGTTGCTGCAGGA
ATGGAAGAAGTGTACAATAACTATAAAAATGC**AG**TTATCCAAAGTGGAGATC
CCAAGGCAAAGGAGATTGTATTGTCGAATATGATTGCTTTATTAGATCGCGT
ATTTTTGGATGTGAC**AG**ATCCTTTTGTCTTTCAACCACACCACAAAGCAAAG
AGAGAGCCTTTTGACTACTACGTGTTTGGTCAGAATTATATCCGTCCTTTAGT
CGATCTCA**AA**AATTCTTATGTTGGCAACATGCCCCTTTTCATTGAAATGGAA
GAGAAACTTAAGCA**GG**GGCACAACATCATCTTGATGTCAAATCACCAAACTG
AAGCTGATCCAGCCATCATTGCTTTGCTGCTTGAAACACGAATCCCATATAT
TGCTGAAAGCAT**GA**CCTATGTAGCCGGAGATAGAGTTATAACTGATCCTCTG
TCCAAACCATTCAGTATTGGCA**GG**AATCTCATTTGTGTTTACTCTAAAAAGCA
CATGCTTGATGATCCAGCTCTTATAGAGATGAAAAGAAATGCAAATATACGA
GCTCTGAAGGAAATGGCTATGCTTTTAA**GG**AGTGGTTCACAAATAGTCTGGA
TTGCCCCAAGTGGTGGTAGGGATCGCCCAGATCCCCACACCGGAGAATGG
GCGCC**GG**CACCCTTTGATACTTCTTCGGTAGATAATATGAGAAGACTTGTTG
AACATTCTGGTCCACCAGGTCATGTATATCCTTTAGCGATATTGTGCCATGA
TATAATGCCCCCTCCACTGAA**GG**TTGAGAAAGAAATTGGGGAGAAAAGAATT
ATATCCTTTCATGGGGCTGGCATATCAGTGGCTCCAGCATTAAGCTTTTCTG
AAACCACTGCTACTAGTGAAAATCCTGAAAA**GG**CTAAGGAGTTATTCTCAAA
AGCCCTGTACGATTCTGTGACCGAGCAATATAATGTGCTGAAATCTGCGATA
CATGGCAAAAAAGGATTTGAAGCATCAACTCCTGTAGTTTCTTTGTCACAGC
CATGGAAGTAGTTGAAATTGGCATTTCTTCATTACAATTTGCTCTGATGCA**GA**
GGCTCCAACTGCTTTGTTCCCTTTCAGAAGCAAGTTACAAGACTTCAGTCAA
ACAATTTCAACTGATTTACTTCTGAGGGACTGCCTATTACTACACCGGTCAC
CGAATGATTTGAGCTTGATGGAAGTTTGCAGTCAAATACATATTTTTCATTTC
ATTTGTCCTTTTGCTCTTGGTTGCCGTTATCAGCATTCAATTCATCTGGAATC
TGTTTCAGTTCAGAAGGTTCAAATTCAGCTGCGTACTGTACAGGTCTCTCTT
AGTTCGGTGTCAGATTCAGTTCATTGACTGATAAAATACTAAATTTTTTCCCT
GCAATTTTGTGATCAGGCTTAGCTAGCTGAATAGATAAAATCTAATTACTTCC
GTTTGTATTTTAAGTTAACTTTGTTCCATTATAGATGAATAGATGTTAATATTA
CATGTTCAGAATCAAGTTGTTCAGACGGGGTCAGTGAATAAACTGGTCCAAA
TGCTAATGCAAG

Figure B2: Coding sequence of G3PAT on Chr. 9 with expression primer locations. The coding region of G3PAT showing the location of the primers used for gene expression. Exon-exon junctions are underlined and bolded. Primer locations are shaded in grey.
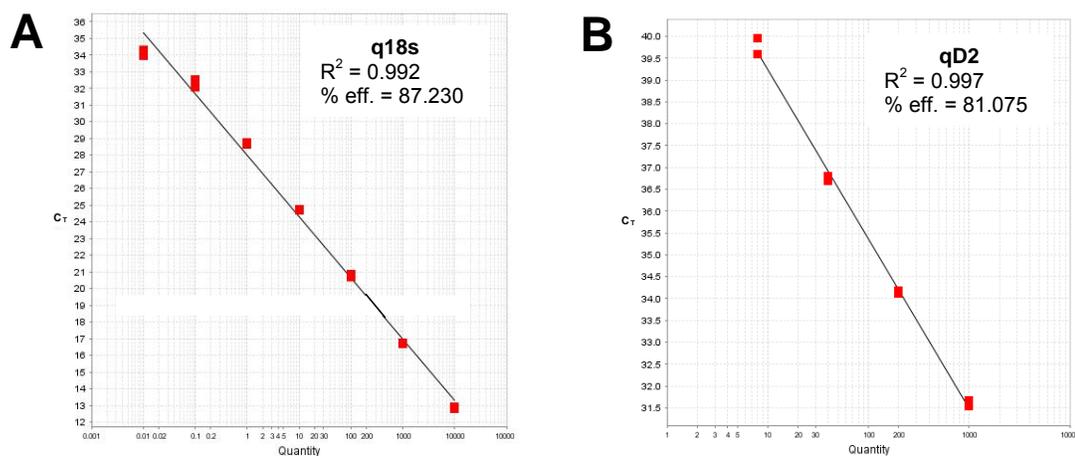
A q18s
$R^2 = 0.992$
% eff. = 87.230

B q19L2
$R^2 = 0.998$
% eff. = 89.241

C
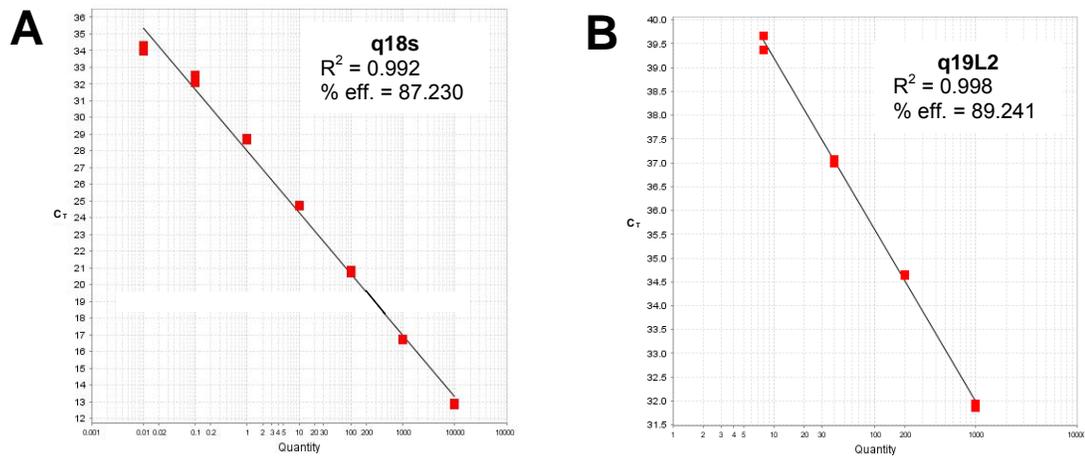
Gene Name	Chrom No.	Primer Name	Forward/ Reverse	Sequence
LPAAT	19	q19L2f	Forward	GGCTGGAACTTGTATGGCTTATTG
		q19L2r	Reverse	AGTGCATGCTCTTTACCCATTAAAT
18S	**	q18Sf	Forward	CGGCTACCACATCCAAGGAA
		q18Sr	Reverse	GCTGGAATTACCGCGGCT

Figure B3: Standard curve and primers used to measure expression of LPAAT. Standard curves for each set of primers used for quantitative real-time PCR (qRT-PCR). Correlation coefficients(R^2) values and the percentage of efficiency of each primer pair are indicated. A – primer pair used for 18s, the housekeeping gene for qRT-PCR; B – primer pair used for lysophosphatidic acid acyltransferase (LPAAT); C – Primers used for quantitative real-time polymerase chain reaction. Chrom. No. is the Chromosome number the copy of the gene studied is located on, based on Phytozome (Phytozome v.7.0). LPAAT = lysophosphatidic acid acyltransferase; 18S = the 18S subunit of ribosomal RNA. ** = sequence and information for primers from Li *et al.*, 2010.

172

ATGGCTATTGCAGCAGCGGCTGTGGTGGTGCCATTGGGCCTGCTCTTCTTC
GCCTCCGGCCTCCTTGTTAATCTCATTCA**GG**CAATATGCTATGTCGTGGTAA
GGCCGGTGTCCAAGAATTTGTACAGACGGATCAACCGGGTAGTGGCGGAA
CTCTTGTGGCTGGAACTTGTATGGCTTATTGATTGGTGGGCAGGAGTTAA**G**
GTCCAAATATTCACAGATCATGAAACCTTTCATTTAATG**GG**TAAAGAGCATG
CACTTGTGATAAGCAATCACAGAAGTGATATTGATTGGCTTGTTGGATGGGT
TTCAGCTCA**GC**GTTCAGGTTGTCTTGGCAGCACTCTAGCTGTGATGAAGAAA
TCTTCAAAGTTTCTGCC**GG**TCATTGGCTGGTCAATGTGGTTTTCTGAGTATC
TTTTTCTGGAGAGAAGTTGGGCCAAGGATGAAAGCACATTAAA**GT**CAGGCA
TCCAGCAACTGAGTGATTTCCCTCTTCCCTTTTGGTTGGCTCTCTTTGTAGA
AGGAACACGCTTTACACAGGCCAAACTATTAGCTGCTCAGGAATATGCAACT
TCCACTGGATTGTCTGTTCCCAGAAATGTTTTGATTCCAAGAACTAA**GG**GTT
TTGTTTCTGCAGTAAGTCATATGCGCTCATTTGTTCCTGCCATTTATGATATA
ACAGTAGCTATCCCTAAGAGTTCCCTGCTCCTACAATGCTAAGACTCTTCA
AGGGACAACCTTCAGT**GG**TGCATGTTCATATCAAGAGGCATTTAATGAAGGA
ATTGCCAGAAACAGATGAGGCTGTTGCTCAATGGTGTCGAGATATATTTGTG
GCCAA**GG**ATGTTTTGTTAGACAAACATATAGCTGAGGACTCTTTTAGTGATC
AAGATTTGCAGGATACTGGTCGACCAATAAAGTCTCTTCT**GG**TAGTTATATC
TTGGGTTTGTCTGGTTGTTGCGGGGTCTGTAAAGTTCCTACAACGGTCTTCG
CTACTCTCTTCCTGGAAGGGTGTTGCATTTTCAGCTTTTGGTTTGGCAGTTG
TTACTGCACTTATGCAAATTCTGATTCAATTCTCACAGTCAGAGCGTTCAAAC
CCTGCCAAGATCGCCCGGGCAAAGTCAAAAAACAAGGGGGGACAACTAGA
GGCTAGGAATGACAAACAACAGTAG

Figure B4: Coding sequence of LPAAT on Chr. 19 with expression primer locations. The coding region of LPAAT showing the location of the primers used for gene expression. Exon-exon junctions are underlined and bolded. Primer locations are shaded in grey.

C	Gene Name	Chrom No.	Primer Name	Forward/ Reverse	Sequence
	DGAT1	13	qD2f	Forward	TGAAGATTATTGGAGGATGTGGAA
			qD2r	Reverse	CCGTGCCTTAAACATGGAAAA
	18S	**	q18Sf	Forward	CGGCTACCACATCCAAGGAA
			q18Sr	Reverse	GCTGGAATTACCGCGGCT

Figure B5: Standard curve and primers used to measure expression of DGAT1. Standard curves for each set of primers used for quantitative real-time PCR (qRT-PCR). Correlation coefficients(R^2) values and the percentage of efficiency of each primer pair are indicated. A – primer pair used for 18s, the housekeeping gene for qRT-PCR; B – primer pair used for diacylglycerol acyltransferase 1 (DGAT1); C – Primers used for quantitative real-time polymerase chain reaction. Chrom. No. is the Chromosome number the copy of the gene studied is located on, based on Phytozome (Phytozome v.7.0). DGAT1 = diacylglycerol acyltransferase 1; 18S = the 18S subunit of ribosomal RNA. ** = sequence and information for primers from Li *et al.*, 2010.

ATGGCGATTTCCGATGAGCCTGAAACTGTAGCCACTGCTCTCAACCACTCTT
CCCTGCGCCGCCGTCCCACCGCCGCTGGCCTCTTCAATTCGCCCGAGACG
ACCACCGACAGTTCCGGTGATGACTTGGCCAAGGATTCCGGTTCCGACGAC
TCCATCAGCAGCGACGCCGCCAATTCGCAACCGCAACAAAAACAAGACACT
GATTTCTCCGTCCTCAAATTCGCCTACCGTCCTTCCGTCCCCGCTCATCGCA
AAGTGAAGGAAAGTCCGCTCAGCTCCGACACCATTTTCCGTCA**GA**GTCACG
CGGGCCTCTTCAACCTCTGTATAGTAGTCCTTGTTGCTGTGAATAGCCGACT
CATCATTGAGAATTTAATGAA**GT**ATGGTTGGTTGATCAAATCTGGCTTTTGGT
TTAGCTCAAAGTCATTGAGAGACTGGCCCCTCTTCATGTGTT**GT**CTTTCTCTT
GTGGTATTTCCTTTTGCTGCATTTATAGTGGAGAAGTTGGCACAGCAGAAGT
GTATACCCGAACC**AG**TTGTTGTTGTACTTCATATAATCATTACCTCAGCTTCA
CTTTTCTATCCAGTTTTAGTAATTCTCA**GG**TGTGATTCTGCTTTTCTATCAGG
TGTTACGTTAATGCTATTTGCTTGTGTTGTATGGTTAAAATTGGTGTCTTATG
CACATACAAACTATGATATGAGAGCACTTACCAAATCAGTTGAAAA**GG**GAGA
AGCTCTGCCCGATACTCTGAACATGGACTATCCTTACAATGTAAGCTTCAAG
AGCTTAGCATATTTCCTGGTTGCCCCTACATTATGTTACCA**GC**CAAGCTATC
CTCGCACACCTTATATTCGAAAGGGTTGGCTGTTTCGCCAACTTGTCAAGCT
GATAATATTTACAGGAGTTATGGGATTTATAATAGAACA**AT**ACATTAATCCCA
TTGTACAAAATTCACAGCATCCTCTCAAGGGAAACCTTCTTTACGCCATCGA
GAGAGTTCTGAAGCTTTCTGTTCCAAATTTATATGTGTGGCTCTGCATGTTCT
ATTGCTTTTTCCACCTTT**GG**TTAAATATATTGGCAGAGCTTCTTCGATTTGGT
GATCGTGAATTCTACCAGGATTGGTGGAATGCCAAAACTGT**TGAAGATTA**TT
**GGAGGATGTGGAATATGCCTGTTCACAAATGGATGATCCGCCACCTATATTT
TCCATGTTTAAGGCACGG**TATACCAAA**GG**CCGTTGCTCTTTTAATTGCCTTC
CTGGTTTCTGCTTTATTCCATGA**GC**TGTGCATCGCTGTTCCTTGCCACATATT
CAAGTTGTGGGCTTTCGGTGGAATTATGTTTCA**GG**TTCCTTTGGTCTTCATC
ACTAATTATCTGCAAAATAAATTCAGAAACTCGAT**GG**TTGGAAATATGATTTT
TTGGTTCATATTCAGTATTCTTGGTCAACCTATGTGCGTACTGCTATATTACC
ATGACTTAATGAATAGGAAAGGCAAACTTGACTGAAGGTGCACGTGGATAA
GCTTTTCTGTTTTTGGAGTGTATAATTGATGTCGATATGTTGATCAATATTGT
TTTCCACGAGTACTTTCATCTACCATGGCAGTGGCTGCTCTGAAGGATTTCC
ACCTGATATACCAGGTCGCGAGGCTAATTCATCTTGATCTATGTACTTATCA
ACTCTCCTCTGGCAATTGTATCGATATATGCAATTTTGAGAGCCATACACTG
GCATTGATAACTGCCAAGGAACAGTGTAGCTGTTTTCTGTTAAATGTTAATTA
GTAGAGAGCTAGATGTAAATAATTTATGCTC

Figure B6: Coding sequence of DGAT1 on Chr. 13 with expression primer locations. The coding region of DGAT1 showing the location of the primers used for gene expression. Exon-exon junctions are underlined and bolded. Primer locations are shaded in grey.