GENETIC AND BIOCHEMICAL ANALYSIS OF ESSENTIAL ENZYMES IN 
TRIACYLGLYCEROL SYNTHESIS IN ARABIDOPSIS

By

KIMBERLY LYNN COTTON

A dissertation submitted in partial fulfillment of 
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY 
Molecular Plant Sciences

DECEMBER 2015

© Copyright by KIMBERLY LYNN COTTON, 2015
All Rights Reserved
To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of

KIMBERLY LYNN COTTON find it satisfactory and recommend that it be accepted.

____________________________________________
John A. Browse, Ph.D., Chair

____________________________________________
Michael M. Neff, Ph.D.

____________________________________________
Thomas W. Okita, Ph.D.

____________________________________________
John J. Wyrick, Ph.D.
ACKNOWLEDGMENTS

To my PI John Browse, thank you for your support, guidance and patience over the years. To my committee, John Browse, Michael Neff, Thomas Okita, and John Wyrick, thank you for your insightful discussions and suggestions. To my collaborators Jay Shockey, Anushobha Regmi, Neil Adhikari, John Browse, and Phil Bates, thanks for the great work and great times. To Browse Lab members past and present, thanks for the help, guidance and wonderful memories. To the undergraduate students who have helped me over the years, thank you for all your help in making this project possible; and with a special thank you to Arlene, Tesa, Breana, and “no-longer-undergrads” David and Barbara Cotton for their particular dedication to this project. And finally, a huge thank you to my family and friends who have loved and supported me throughout this venture. Go Cougs!
GENETIC AND BIOCHEMICAL ANALYSIS OF ESSENTIAL ENZYMES IN
TRIACYLGLYCEROL SYNTHESIS IN ARABIDOPSIS

Abstract

by Kimberly Lynn Cotton, Ph.D.
Washington State University
December 2015

Chair: John A. Browse

Plant oils are used in food, fuel, and feedstocks for many consumer products, and so understanding the process by which they are made and modified will help us to make plant oils more healthy, useful, and sustainable. While some of the genes encoding the ER-localized enzymatic steps to triacylglycerol (TAG) have been well understood and documented, several are still in need of study. The glycerol-3-phosphate acyl transferase (GPAT) enzymatic activity is the first step in the pathway to TAG, and it acylates glycerol 3-phosphate to produce lysophosphatidic acid. *GPAT9* (AT5G60620) is conserved across land plants and is homozygous lethal, indicating an essential function. Transcript level in knockdown mutants correlates with GPAT activity and with oil levels, and the protein interacts with other enzymes in the TAG biosynthesis pathway. These data suggest that *GPAT9* encodes the main GPAT involved in membrane lipid and TAG synthesis. The phosphatidic acid phosphatase (PAP) step in TAG synthesis is responsible for the hydrolysis of inorganic phosphate from phosphatidic acid and creation of diacylglycerol (DAG). There are 13 putative PAPs in Arabidopsis which are homologous to known PAPs. Most of these are involved in other processes, including the plastidial lipid synthesis pathway and signaling pathways. The Arabidopsis gene *LPPβ* (At4g22550) is expressed in seed tissue, its protein product is localized to the ER, and it encodes PAP activity, indicating that it is a likely candidate for the PAP involved in oil synthesis. At the conclusion of this work, questions remain about the role of
LPPβ in oil synthesis and which genes encode the major enzymes involved in the steps generating phosphatidylcholine and converting it back to DAG; but the main Kennedy Pathway enzymes generating TAG have been identified and characterized.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER ONE. INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>12</td>
</tr>
<tr>
<td>CHAPTER TWO. INITIAL ASSESSMENT OF POTENTIAL PAP GENES INVOLVED IN OIL SYNTHESIS IN ARABIDOPSIS</td>
<td>17</td>
</tr>
<tr>
<td>2.1 INTRODUCTION</td>
<td>18</td>
</tr>
<tr>
<td>2.2 MATERIALS AND METHODS</td>
<td>25</td>
</tr>
<tr>
<td>2.3 RESULTS AND DISCUSSION</td>
<td>30</td>
</tr>
<tr>
<td>2.4 CONCLUSION</td>
<td>41</td>
</tr>
<tr>
<td>2.5 REFERENCES</td>
<td>54</td>
</tr>
<tr>
<td>CHAPTER THREE. GENETIC, MOLECULAR AND BIOCHEMICAL CHARACTERIZATION OF PUTATIVE PAP GENES</td>
<td>61</td>
</tr>
<tr>
<td>3.1 INTRODUCTION</td>
<td>62</td>
</tr>
<tr>
<td>3.2 MATERIALS AND METHODS</td>
<td>65</td>
</tr>
<tr>
<td>3.3 RESULTS</td>
<td>73</td>
</tr>
<tr>
<td>3.4 DISCUSSION</td>
<td>79</td>
</tr>
<tr>
<td>3.5 CONCLUSION</td>
<td>83</td>
</tr>
</tbody>
</table>
CHAPTER FOUR. IDENTIFICATION OF ARABIDOPSIS GPAT9 (At5g60620) AS AN ESSENTIAL GENE INVOLVED IN TRIACYLGLYCEROL BIOSYNTHESIS .................................................. 99

4.1 INTRODUCTION .................................................................................. 100
4.2 MATERIALS AND METHODS .............................................................. 104
4.3 RESULTS ............................................................................................. 112
4.4 DISCUSSION ......................................................................................... 124
4.5 CONCLUSIONS ................................................................................... 131
4.6 REFERENCES ...................................................................................... 148
4.7 SUPPLEMENTAL FIGURES, METHODS, AND REFERENCES ..................... 158
4.8 SUPPLEMENTAL REFERENCES ............................................................ 175

CHAPTER FIVE. CONCLUSIONS AND PERSPECTIVES .................................... 176

REFERENCES .......................................................................................... 185
LIST OF TABLES

CHAPTER 2

1. Literature review and protein information of putative PAPs...........................................45
2. Homologs of Arabidopsis putative PAP proteins in Castor Bean ......................................46
3. Homologs of Arabidopsis putative PAP proteins in Soybean ............................................47
4. Homologs of Arabidopsis putative PAPs in seed tissue of oilseeds ........................................48
5. Primers used in this work........................................................................................................53

CHAPTER 3

1. Primers used in this work........................................................................................................93

CHAPTER 4

1. Non-Mendelian segregation of GPAT9/gpat9-2 under sulfadiazine selection ...........136
2. Transmittance of sulfadiazine resistance from reciprocal crosses … .................................137
3. Ovule viability within GPAT9-2/gpat9-2 half siliques..............................................................138
4. Pollen germination and tube growth of GPAT9-2/gapt9-2 within the qrt1 ...............................140
S1. Primers utilized for PCR, RT-PCT, & qPCR ........................................................................173
S2. Gene identifiers for GPAT9, LPEAT1, and LPEAT2 sequences........................................174
LIST OF FIGURES

CHAPTER 1

1. TAG biosynthesis in the eukaryotic pathway in Arabidopsis.................................11

CHAPTER 2

1. Phylogenetic tree of LPP homologs........................................................................42
2. Conserved domains in the PAP2 family of PAPs....................................................43
3. Bioinformatics Information on LPPβ .................................................................49
4. Tissue expression of genes of interest ...................................................................50
5. Subcellular localization of LPPβ, SPP1, and LPAP1 .............................................51
6. Non-Diurnal Expression Pattern of LPPβ .............................................................52

CHAPTER 3

1. RT-PCR of SALK Insertional Mutants.................................................................84
2. Oil Levels in LPPβ RNAi Lines ..........................................................................85
3. LPPβ Expression in RNAi Plants .........................................................................87
4. LPPβ Expression in Multiple Mutant Lines .........................................................88
5. Seed Oil and Fatty Acid Content of Multiple Mutants .........................................89
6. PAP Complementation Assays ...........................................................................90
7. PAP Activity Assays .............................................................................................91
8. LPPβ is Not Regulated by WRII at the Transcript Level ......................................92

CHAPTER 4

1. Sequence Comparison of Selected GPAT9 Enzymes from Plants......................132
2. Phylogenetic Comparison of GPAT9.....................................................................134
CHAPTER 4

3. Aborted Ovules in GPAT9/gpat9-2 Siliques..........................135
5. Genetic Complementation of gpat9-2 Mutants with Transgenic Plant GPAT9 ....141
6. Segregating Red and Brown T2 GPAT9-amiRNA Seed Oil Content and Size.142
7. Oil Quantity and Composition of Homozygous T3 GPAT9 amiRNA Lines ...........143
8. Developing Silique Microsome GPAT Activity........................................144
9. Silique Microsome LPAAT Assays.............................................................145
10. Testing of Protein:Protein Interactions Between AtGPAT9 and Other Lipid ....146

S1. Sequence Alignment and Identity/Similarity Shading of Plant LPEAT1 Proteins ...158
S2. Sequence Alignment and Identity/Similarity Shading of Various Plant .............161
S3. Characterization of GPAT9 T-DNA Insertions ........................................164
S4. Analysis of Viability in Wild-Type and GPAT9/gpat9-2 Pollen......................165
S5. Lipid Characterization of Wild-Type and GPAT9-2/gpat9-2 Plants...............166
S6. Seed Fatty Acid Composition of GPAT9 amiRNA Knockdown Lines..............167
S7. Expression of GPAT9 in Developing Seeds of Wild-Type and GPAT9 ............168
S8. Developing Seed Microsome GPAT Assays of Wild-Type and amiRNA ..........169
S9. Testing of Protein:Protein Interactions Between AtLPAAT2, AtDGAT1, and ....170
S10. Sequence Identity Comparisons Between AtGPAT1 and AtGPAT9 to the ....171
LIST OF FIGURES (Continued)

CHAPTER 5

1. TAG Biosynthesis in the eukaryotic pathway of Arabidopsis ............................184
CHAPTER ONE

INTRODUCTION

Preface

Herein represents my contributions to the understanding of how oil is produced in plants. In Chapters 2 and 3, I will discuss my work analyzing putative phosphatidic acid phosphatases (PAPs) in Arabidopsis. To date, only two PAP genes, AtPAH1 and AtPAH2, have been shown to affect oil synthesis: in plants containing a double null mutant at these loci, the fatty acid content of the seeds is reduced only 15% relative to wild type and each of the single mutants, therefore other PAPs must be present to generate the remaining oil [1]. In Chapter 2, I will identify genes as potential PAPs and then analyze them based on published literature, bioinformatics data, and subcellular localization to determine their fitness for study as PAPs involved in oil synthesis. Chapter 3 will include work to characterize genes identified in Chapter 2, and whether or not they act redundantly with some of the other putative PAPs identified in Chapter 2. In Chapter 4, I describe my contributions to the identification and characterization of the main glycerol-3-phosphate acyltransferase (GPAT) in Arabidopsis, GPAT9. GPATs involved in cuticle formation and the plastidial lipid synthesis pathway have been identified, but this work represents the first identification and characterization of an essential GPAT involved in both oil and ER membrane synthesis.

1.1. Introduction to Lipid Metabolism

Lipids have diverse and vital roles in metabolism, signaling, and structure. Therefore, lipid biosynthesis has been an area of intense research for decades. Structural lipids form the barrier
between an organism and its environment, as well as partition organelles within the cell. As a compact source of energy and carbon for growth, the nonpolar lipid triacylglycerol (TAG) is vital to the developing plant embryo and pollen. Lipids also form signaling compounds and hormones for regulation of cellular responses.

Plant-based oils are used by humans both as a source of food calories and in commercial products. Oil from both plants and animals constitutes up to 40% of the caloric intake of a western pattern diet. Lipids are the main component of fats and oils, and as Americans increasingly suffer from obesity and related health problems, ingesting healthier fats by reducing saturated fats and increasing omega-3 fatty acids has become a priority among mainstream Americans. Additionally, as we reduce our dependence on non-renewable natural resources, it is increasingly important to engineer oilseed crops which generate fatty acids useful for industrial processes [2]. Commodities like cooking oil, plastics, paints, pharmaceuticals, lubricants, and dyes all rely on plant-derived oils. Many oil products could be sustainably and renewably sourced from plants, with the proper engineering and a keen eye to scale. Model systems such as Arabidopsis thaliana allow us to study how plants make oil, and then apply this knowledge to improve other plant species, in an effort to produce healthier foods or generate useful fuel alternatives.

In plants, TAG constitutes over 90% of the fatty acids in the seeds of Arabidopsis, and is an essential energy source for both germinating seeds and pollen [3]. A deficiency in TAG leads to sterile pollen and disruptions in embryo development, as seen in the dgat1-1 pdat1-2 double mutant [4]. While the pathways generating lipids are complex, many of the biochemical activities have been elucidated, and there is now a basic understanding of the reactions that generate galactolipids, phospholipids, and TAG in both plants and other organisms [3].
1.2. GLYCEROLIPID SYNTHESIS IN ARABIDOPSIS

1.2.1. LIPID BIOSYNTHESIS PATHWAYS AND TAG SYNTHESIS

Higher plants have two distinct pathways for the synthesis of glycerolipids [3]. The prokaryotic pathway is localized to the plastid and is responsible for the synthesis of the major plastidial membrane lipids phosphatidylglycerol (PG), monogalactosyl-diacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), and sulfoquinovosyldiacylglycerol (SQDG) [3]. The eukaryotic pathway is localized to the endoplasmic reticulum (ER) and generates extraplastidial membrane lipids such as phosphatidylcholine (PC), phosphatidylethanolamine (PE), and phosphatidylinositol (PI), as well as the storage lipid TAG [3].

Fatty acids are synthesized de novo in the plastid using acetyl-CoA as the initial substrate [3]. They are then exported to the ER for the synthesis of lipids in the eukaryotic pathway [3]. This pathway of lipid production was first described by Kennedy and colleagues more than 60 years ago, and describes the steps of lipids synthesis from glycerol-3-phosphate to TAG [5]. As seen in Figure 1, fatty acids are esterified to the sn-1 and sn-2 positions of glycerol-3-phosphate to generate first lyso-phosphatidic acid (LPA) using the enzyme glycerol-3-phosphate acyl transferase (GPAT), and then phosphatidic acid (PA), by the activity of a lysophosphatidic acid acyl transferase (LPAT) [3]. Phosphatidic acid phosphatase (PAP) then dephosphorylates PA to generate diacylglycerol (DAG) [3]. The Kennedy pathway describes DAG as being converted directly to TAG; however more recent studies suggest that de novo DAG is converted to phosphatidylcholine (PC) using PC:DAG choline phosphotransferase (PDCT) [6]. Triacylglycerol can be generated directly from PC and DAG using PC:DAG acyl transferase (PDAT) [7], or either phospholipase C (PLC) or PDCT can generate PC-derived DAG which can be converted to TAG through diacylglycerol acyl transferase (DGAT).
Upon synthesis, TAG will form oil bodies, consisting of a monolayer phospholipid membrane filled with hydrophobic TAG. These oil bodies can bud off from the ER and be transported elsewhere for storage and eventual use as an energy source [3]. While each enzymatic function in this pathway is known to occur, many of the genes encoding these functions have not been identified. Chapters 2 and 3 in this dissertation will focus on identifying the PAP involved in oil synthesis, while Chapter 4 will focus on identifying and characterizing the GPAT.

The eukaryotic pathway is also used to generate membrane phospholipids. PC is the main structural lipid in the ER, and the desaturation and acyl editing pathways branch off from PC. PA can be used to make phosphatidyl serine (PS) directly, or can produce CDP-DAG, an activated form of DAG, which is then used to generate phosphatidyl glycerol (PG) and phosphatidylinositol (PI). DAG can be converted to phosphatidylethanolamine (PE), which can also generate PS or PC. These phospholipids function as membrane components in the plant, where PC is the main ER structural lipid along with PE, and PS, PI, and PG make minor structural contributions. PI and its phosphorylated derivatives also play a major role in signaling. PA and DAG are present only in minute amounts in the ER membrane. Additionally, PC generated in the eukaryotic pathway can be converted to DAG and either used for TAG synthesis or transported back to the plastid to generate MGDG, DGDG, and SQDG

1.2.2. WRI1 CONTROL OF CARBON FLUX

Wrinkled1 (wri1) was first characterized in 1998 as a mutant whose seeds were impaired in the incorporation of carbohydrates into seed TAG, causing a wrinkled phenotype [8]. The wri1 mutant is compromised in glycolysis, rendering it deficient in converting sucrose into the precursors of TAG [9]. This is because WRI1 is a transcriptional activator which activates a subset of sugar responsive genes and controls the flow of carbon from sucrose import to oil accumulation
WRI1 is also controlled by the transcription factor LEC2, which plays a central role in embryo development, and WRI1 is required for LEC2 to properly control FA metabolism [11]. WRI1 regulates the expression of genes involved in late glycolysis, fatty acid synthesis, biotin synthesis, and lipoic acid synthesis [12], but it is not known to control genes involved in oil synthesis or packaging. It is a member of the plant-specific transcription family AP2/EREBP family with two AP2 DNA-binding domains [9], and functions by binding the AW box domain (CnTnGnγCG) in the promoter [13]. Two enhancer domains have also been identified in the promoters of two targets of WRI1 [12], but it is not yet known whether these enhancer domains are found in promoters of other WRI1 targets.

1.3. THE ROLE OF GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE IN LIPID SYNTHESIS

1.3.1. GPAT ACTIVITY IN LIPID SYNTHESIS

The sn-glycerol-3-phosphate acyltransferase activity is responsible for the first step in glycerolipid synthesis: the acylation of the glycolysis product glycerol-3-phosphate with an acyl-CoA or acyl-ACP to produce lysophosphatidic acid. This begins the pathway of lipid synthesis which leads to structural lipids that form the membranes of a cell, triacylglycerol that is used as an energy reserve, and signaling molecules which regulate cellular processes. Lipid synthesis occurs mainly in the ER and plastid, although there is also a minor pathway in the mitochondria.

1.3.2. KNOWN GPATS IN ARABIDOPSIS

The GPAT reaction, like most enzymatic activities in the Kennedy Pathway, was first characterized in animals over 60 years ago [5,14-16]. In mice, there are four homologous isoforms of GPAT, which are responsible for the synthesis of all glycerolipids and are the rate limiting step in TAG synthesis [17], but it has not been conclusively identified in humans. Arabidopsis contains
a plastid-localized GPAT (At1g32200; [18]), a plant-specific GPAT family consisting of 8 genes [19-21], and the ER-localized GPAT9 [21], the focus of our studies. The plastidial GPAT is encoded by the ATS1 locus, and is involved in PG synthesis in the plastid. The ats1-1 mutation eliminated the prokaryotic pathway of lipid synthesis, and when it was combined with a mutant in tgd1-1, chloroplast membrane synthesis was abolished and resulted in lethality [22].

The GPAT1 to -8 family forms a clade distinct from the other two GPATs, and members are involved in the synthesis of cutin and suberin, not phospholipids [19-21,23]. Cutin and suberin serve as a barrier to prevent dehydration, uncontrolled gas exchange, and physical damage from pathogens and herbivores. GPAT4 and GPAT8 are required for the accumulation of C16 and C18 cutin monomers in stems and leaves [24], while GPAT6 is required for incorporation of C16 monomers, and functions in stamen development and pollen fertility [25,26]. These three genes form the first subclade of the family. GPAT5 and GPAT7 also form a subclade, where GPAT5 is responsible for accumulation of C22 and C24 suberin monomers in root and seed coat [20], and GPAT7 appears to also produce C22 and C24 monomers for suberin synthesis in response to wounding [27]. GPAT1 to -3 form the final subclade, and GPAT1 displays sn-2 acyltransferase activity on saturated, mono unsaturated, and ω-oxidized acyl CoA’s which is essential for male fertility and tapetum differentiation. GPAT2 and GPAT3 have not yet been proven as acyltransferases in vitro [19,27].

While there has been significant study of Arabidopsis GPAT enzymes, the ER-localized isoform involved in structural, signaling, and storage lipid synthesis has not yet been identified. It is expected that a GPAT involved in these processes will be essential for viable plants.
1.4. The Role of Phosphatidic Acid Phosphatase in Lipid Synthesis

Phosphatidic acid phosphatases (PAP) are also essential for three main functions in the plant: membrane lipid synthesis, regulation of the signaling functions of PA and DAG, and synthesis of storage lipids. Several of these functions have been elucidated, including the PAPs involved in the prokaryotic pathway of lipid synthesis in the thylakoid membranes. However, the main PAP which is involved in the eukaryotic pathway of the ER has not yet been discovered.

1.4.1. PAP Forms (PAP1 and PAP2)

Phosphatidic acid phosphohydrolases (PAH), lipid phosphate phosphatases (LPP), and diacylglycerol pyrophosphate phosphatases (DPP) all share a common enzymatic activity: the hydrolysis of a phosphate group from PA, resulting in DAG [28]. They are generally broken into two groups: the “PAP1” enzymes, which are specific for phosphatidic acid as a substrate, and include the PAHs in yeast and mammals, and the “PAP2” enzymes, including LPPs and DPPs, which are often not specific for a particular lipid substrate and are often involved in signaling functions [28]. These two enzyme classes have a single enzymatic activity: the cleavage of a single phosphate group from PA (or other substrates in the case of LPP) [28]. DPPs are more specific for diacylglycerol pyrophosphate (DGPP) than for PA and catalyze two reactions: first they remove the β-phosphate group from diacylglycerol pyrophosphate, and subsequently hydrolyze the α-phosphate to convert the intermediate PA to DAG [29]. All PAPs are localized to the membrane or are cytosolic proteins with a peripheral association with the membrane, since the substrate PA is almost exclusively found in the membrane [28]. The differentiation between PAHs and LPPs has been well documented in yeast [28]; however studies in plants have been less conclusive, mainly due to the lack of knowledge of the PAP involved in oil synthesis.
1.4.2. PAPs in Yeast and Mammals

Until recently, research on phosphatidic acid phosphatase (PAP) enzymes involved in lipid synthesis had been hampered by a lack of molecular information on the genes encoding this enzymatic activity [30]. LPPs and DPPs have been studied in yeast and mammals in some detail, but these enzymes have little effect on TAG synthesis [28]. The yeast gene ScPAH1 allowed the molecular identification of the lipin family in mammals, which are PAPs involved in fat accumulation and regulation in mice, rats and humans. Its homologs in Arabidopsis, however, are involved in membrane remodeling and have only a minor effect on TAG. The final PAP in yeast is APP1, which has no close homologs in Arabidopsis. The lipin family of proteins are the only PAP enzymes in yeast and mammals which has been shown to affect TAG synthesis [1,31]. The main function of the proteins are the generation of the storage lipid TAG, although it also has a lesser effect on membrane growth, phosphatidic acid signaling, and regulation of overall lipid biosynthesis [1,32,33].

1.4.3. PAPs in Arabidopsis

Three PAPs in Arabidopsis (LPPγ, LPPε1, and LPPε2) have been identified and characterized as being involved in plastidial lipid synthesis, and are therefore not involved in oil synthesis [34]. Three other PAPs in Arabidopsis (LPP1-3) have been characterized or implicated as being involved in signaling pathways, and are unlikely to be involved in oil synthesis [35,36]. Only two redundant PAP genes (PAH1 and PAH2) have been discovered to be involved in TAG synthesis, and these together only account for ~15% of TAG in the seed [1]. Arabidopsis must have other PAP genes which account for the remaining PAP activity. Other than pah1pah2, no putative PAP mutants in Arabidopsis have been analyzed for a change in seed oil quantity. In this study, I will analyze all putative PAPs for possible involvement in the oil synthesis pathway.
1.4.4. PA AND DAG IN LIPID SIGNALING

In addition to their roles as Kennedy pathway intermediates and precursors for membrane lipids, PA, and DAG both have signaling functions, and therefore their accumulation is expected to be tightly regulated. Unlike structural lipids, signaling lipids are present in low quantities in the membrane until a stimulus causes their rapid accumulation; they are then turned over rapidly to attenuate their signaling functions.

Phosphatidic acid that is involved in signaling is generated not by the LPAAT activity mentioned above, but instead by two distinct pathways. In the first pathway, PLD removes the head group primarily from PC or PE to generate phosphatidic acid and the free head group [37]. In the second pathway, Phospholipase C removes both the head group and phosphate from phosphatidylinositol lipids to release DAG, which is then phosphorylated by DAG kinase to produce phosphatidic acid [37]. PA produced from different PLDs can be differentiated by the plant and have different roles in plant response, providing a means to specialize the response to different stresses. The PAP, diacylglycerol pyrophosphate phosphatase, or PA kinase activity may then attenuate the signaling functions of PA by converting it to DAG or DAG pyrophosphate [37]. Phosphatidic acid is produced as a signaling molecule in response to different conditions, including pathogen attack, ABA, osmotic stress, and temperature stress [37]. It acts through two main mechanisms: recruitment of proteins to the membrane and activation or deactivation of the protein’s activity [38-40].

The understanding of the role of DAG in signaling has lagged behind the understanding of PA as a signaling molecule, but there is compelling evidence that it should not be discarded from consideration. The interacting proteins for DAG have not yet been identified, and those plant proteins homologous to known DAG interactors in animals appear to interact with protein kinases
instead of DAG [41]. Because of this, many believe that PA is the main signaling lipid in plants; and indeed the two are interconvertible, making discrimination difficult at times. However, there is important evidence arguing for DAG acting as a signaling molecule, including that DAG accumulates in a manner consistent with an activity in signaling pathways, under various conditions including salinity stress [42], phosphate starvation [43-45], pollen tube tip growth [46], and brassinolide treatment [47]. DAG is known to bind to the C1 domain in protein kinase C (PKC) proteins; and while there are no PKCs in plants, there remains the possibility that it binds to a protein that is not a PKC homolog. Thus, DAG appears to be involved in signaling pathways, and there are many possible targets whose DAG binding have not yet been explored.

1.5. CONCLUSION

Two enzymatic activities in the pathway to TAG and other lipids lack knowledge of their molecular basis. The glycerol-3-phosphate acyl transferase (GPAT) and phosphatidic acid phosphatase (PAP) enzymatic reactions are both known to occur in lipid synthesis in Arabidopsis, and are essential in generating structural, signaling, and storage lipids. In Chapter 2, I will describe my work identifying potential PAP candidates in Arabidopsis. Chapter 3 contains information characterizing the PAP candidates. In Chapter 4, I will describe the characterization of GPAT9 as the main GPAT in membrane and oil synthesis.
Figure 1: TAG biosynthesis in the eukaryotic pathway of Arabidopsis. Pictured is a simplified diagram of relevant pathways to TAG. The names of the intermediates and a cartoon of their structure are shown in black, as are the arrows showing flux into TAG. Enzymes are shown in red, and green highlights where membrane lipids diverge off from the oil synthesis pathway. The entire pathway occurs in the membrane of the endoplasmic reticulum, but after synthesis, TAG will exit the ER in oil bodies. [6,48,49]
REFERENCES


CHAPTER TWO
INITIAL ASSESSMENT OF POTENTIAL PAP GENES INVOLVED IN OIL SYNTHESIS IN ARABIDOPSIS

ABSTRACT

Even though the enzymes involved in membrane lipid and oil synthesis have been studied for over 60 years, many of the genes associated with these enzymatic activities have yet to be determined. In plants, phosphatidic acid phosphatase (PAP) genes have been studied in plastidial lipid synthesis and in signaling, but the isoform involved in ER membrane lipid and oil synthesis has yet to be elucidated. Here, I use a bioinformatics approach to identify genes in the Arabidopsis genome which could be PAPs involved in membrane lipid and oil synthesis. Thirteen genes in the Arabidopsis genome were identified using homology to known yeast, mammalian, and plant PAPs. A PAP involved in these processes should show PAP activity, be expressed in the seed, and should not already be implicated in plastidial lipid synthesis or signaling. Most of the 13 candidate genes were either already shown to be involved in other pathways, or were not localized to the seed. Only one gene, \textit{LPP}\textsubscript{\beta} (At4g22550), showed characteristics warranting further study. Four other genes, \textit{PAH1}, \textit{PAH2}, \textit{SPP1}, and \textit{LPAPI} showed several characteristics indicating they could act redundantly with the PAP involved in oil synthesis, and therefore will be studied in that context. At the end of this study, we have shown that \textit{LPP}\textsubscript{\beta} is the only remaining PAP homolog in Arabidopsis which could be the main PAP involved in oil and membrane lipid synthesis.
2.1. INTRODUCTION

While the enzymatic activities involved in lipid synthesis have been known for over 60 years, the genes encoding some of those functions still need to be elucidated. The phosphatidic acid phosphatase (PAP) reaction catalyzes hydrolysis of the phosphate group from phosphatidic acid (PA) to generate diacylglycerol (DAG). While PAP activity has been characterized in the plastidial prokaryotic pathway of lipid synthesis and in stress and signaling pathways, the gene (or genes) encoding the PAP involved in the ER-localized eukaryotic pathway of lipid synthesis has not yet been identified. This enzyme is expected to catalyze an essential step in lipid and oil synthesis, and is therefore expected to be essential and well conserved. Essential genes are often encoded redundantly in genomes to safeguard against possible mutations, deletions, or inactivation (e.g. [1]); however, many highly conserved, essential housekeeping genes are actually maintained as single genes in a genome despite duplication [2]. Thus, either a single gene or several genes could encode the PAP enzymatic step in the eukaryotic pathway of lipid synthesis.

Two basic types of PAP activity are found in eukaryotic organisms: a “PAP1” type enzyme that is specific for PA, dependent on Mg$^{2+}$, and containing a $\text{DXXD(T/V)}$ motif; and a “PAP2” or lipid phosphate phosphatase (LPP) type activity that is not dependent on any divalent cation, is not specific for PA, and is directed by a three-motif domain ($\text{KXX}_{6}\text{RP}$, $\text{PSGH}$, and $\text{SRX}_{3}\text{HX}_{3}\text{D}$, reviewed by Carman and Han [3]). The PAP1 enzyme type is the only form that appears to be involved in phospholipid synthesis and oil accumulation; to date, no LPP type enzyme has been implicated in oil synthesis.

2.1.1. PAP HOMOLOGS IN YEAST

Yeast ($\text{Saccharomyces cerevisiae}$) has four genes encoding enzymes with PAP activity: $\text{LPP1}$ [4], $\text{DPP1}$ [5], $\text{PAH1}$ [6], and $\text{APP1}$ [7]. Both the Pah1p and App1p proteins are Mg$^{2+}$-
dependent and use a DXDX(T/V) motif for catalysis [8]. The two proteins are not homologous, however; Pah1p is a lipin homolog and the DXDX(T/V) motif is found within the haloacid dehalogenase-like (HAD-like) domain characteristic of the lipin family of proteins [6,9]. The DXDX(T/V) motif of the App1p protein is found within a region of weak sequence similarity to the HAD-like domain, but the two protein sequences are not similar enough to allow alignment [7]. In contrast, the Lpp1p and Dpp1p proteins use the PAP2 consensus sequences KX₆RP, PSGH, and SRX₃HX₃D found in lipid phosphatases, mammalian glucose-6-phosphatases, and bacterial nonspecific acid phosphatases [10,11]. Additionally, Pah1p is specific for PA [6], while App1p, Lpp1p and Dpp1p are not specific for PA and have been shown to also dephosphorylate DAG pyrophosphate, lysoPA, and, in the case of Lpp1p and Dpp1p, isoprenoid phosphates [4,5,7,12]. Lpp1p and Dpp1p both have six transmembrane domains, and are localized to the vacuole and golgi, respectively [4,5,13,14]. App1p is localized to the cytosol, microsome, and is most prominent in the mitochondria [7]. Pah1p exists mainly in the cytosol in its phosphorylated form, but for catalysis it is translocated to the nuclear/ER membrane through dephosphorylation [15].

The yeast Pah1p protein plays an important role in lipid metabolism and is the only PAP known to effect phospholipid and TAG synthesis. Compared to the Δpah1 mutant, the Δapp1 single mutant and Δlpp1 Δdpp1 double mutant did not show significant changes in the DAG, TAG, or phospholipid levels which are characteristic of the Δpah1 mutant (and likely contributes to its temperature-sensitive phenotype), indicating that PAH1 is the only PAP involved in mediating lipid synthesis in yeast [7]. In addition to the reduction in TAG and misregulation of phospholipid synthesis, the loss of Pah1p activity in the Δpah1 mutant leads to abnormal expansion of the nuclear/ER membrane, susceptibility to fatty acid induced lipotoxicity, defects in lipid droplet formation and disruption of vacuole homeostasis [6,9,16-19].
Unlike Pah1p, the Lpp1p, Dpp1p, and App1p proteins are not involved in de novo lipid synthesis that occurs in the ER [3-5]. While the roles of these proteins have not been fully explored, Lpp1p and Dpp1p appear to play roles in signaling by regulating the relative amounts of PA, DAG, diacylglycerol pyrophosphate, and lysoPA at the vacuole and golgi, respectively [4,5,13]. Additionally, both PA and DAG are known to facilitate membrane fission and fusion events in model systems [20-23]. App1p localizes to cortical actin patches and could regulate the local concentrations of PA and DAG [24], although its nonspecific nature indicates it could also regulate concentrations of other phospholipids as well. App1p also shows interactions with endocytic proteins through PXXP domains, and could regulate the enzymes involved in vesicular trafficking [7,25,26].

A quadruple knockout of the four PAPs has no PAP activity, indicating that there are no remaining undiscovered PAPs in yeast [7]. Pah1p is essential for oil synthesis and is the only PAP involved in TAG synthesis or the regulation of phospholipid synthesis. Thus, the PAPs in yeast form two groups: a “PAP1” lipin homolog that is specific for PA and is involved in phospholipid and TAG synthesis, and substrate-nonspecific PAPs that are involved in signaling and other processes.

2.1.2. PAP HOMOLOGS IN MAMMALS

Mammals have two main forms of PAP activity, which generally follow the division in yeast: nonspecific lipid phosphate phosphatases (LPPs) involved in signaling, and the lipin family, which is homologous to ScPAH1 and is involved in TAG synthesis as well as the regulation of lipid metabolism.

The LPP family of proteins are integral membrane proteins that are capable of dephosphorylating a wide variety of substrates, including phosphatidic acid (PA), diacylglycerol
pyrophosphate (DGPP), lysoposphatidic acid (LPA), sphingosine 1-phosphate (S1P), and ceramide 1-phosphate (C1P) [27]. LPPs localize predominantly to the ER and golgi, but also localize to the plasma membrane and to other membrane organelles. There are three LPP proteins in both mice and humans, and they demonstrate overlapping catalytic activities and substrate specificities [28]. The LPP1 protein in mouse appears to be the main LPA phosphatase; although a knockout did not cause any strong visible phenotypes, multiple tissues showed reduced LPA phosphatase activity. The LPP2 mutants are similarly unremarkable, and its role in vivo has not yet been established. The LPP3 protein appears to be involved in vascular development in mice, and null mutants are lethal due to an inability to properly form vasculature in utero [28]. While the LPP family is important for the organism, they clearly are not the main PAPs involved in oil or phospholipid synthesis.

The lipin proteins were first studied when Langer and colleagues described a mutation in mice that caused fatty liver dystrophy (fld) [29]; later, the gene that encoded the mutant enzyme was identified and named lipin [30]. Like the yeast homolog, both the human and mouse isoforms contain a DXDX(T/V) motif in a C-terminal HAD-like domain [31]. Lipin proteins are PAPs specific for PA, and regulate fat metabolism in mammalian cells; lipin deficiency prevents normal development of adipose tissue and results in lipodystrophy and insulin resistance in mice, and an excess of lipin1 leads to obesity and insulin sensitivity [30,32].

There are three lipin genes each in mouse and human; the lipin1 gene is the most studied and has several splice variants (reviewed in [33]). The isoforms show differential transcript expression in tissues and across development, and the isoforms exhibit different levels of PAP activity and PA affinity [31,34]. Lipin1 contributes almost all of the PAP activity in skeletal muscle [34]. Deficiency causes accumulation of PA and phospholipids, possibly leading to the
severe cardiac muscle damage and rhabdomyolysis in early childhood that is seen in humans with mutations in lipin1 [35,36]. Lipin1 is also a transcriptional coactivator of lipid metabolism genes, and physically interacts with several receptors and transcription factors in different pathways to activate or repress gene expression (e.g. PGC-1α [37], HNF-4α [38]; reviewed in [33]). Unlike the yeast Pah1p, whose activity is regulated through phosphorylation at key phosphorylation sites, mammalian lipins do not have those sites and are not regulated by phosphorylation [39,40].

2.1.3. Known PAP Homologs in Arabidopsis

At first glance, the PAP homologs in Arabidopsis follow the strict partitioning of yeast and mammals. There are two lipin homologs, and at least nine homologs of mammalian or cyanobacterial LPPs [41,42].

Of the nine LPP homologs that have been published so far, the biological functions of several have been well established in the literature. AtLPPδ, AtLPPε1, and AtLPPε2 are homologous to cyanobacterial LPP’s and are localized to the plastid, the site of the prokaryotic pathway of lipid synthesis [41]. While their specificity for PA was not analyzed, their PAP activity was assayed using the yeast Δdpp1Δlpp1Δpah1 mutant, and all three were able to complement the mutant, indicating that they exhibit PAP activity [41]. A homozygous knockout of LPPδ could not be isolated, indicating that it is an essential gene, while a double knockout of lppε1lppε2 showed no significant phenotype [41]. This indicates that the LPPδ gene in Arabidopsis encodes the main plastidic PAP involved in the prokaryotic pathway of lipid synthesis.

Additionally, AtLPP1 and AtLPP2 are eukaryotic LPP homologs that have been implicated in signaling pathways in Arabidopsis. Both LPP1 and LPP2 proteins showed diacylglycerol pyrophosphate (DGPP) phosphatase activity and PAP activity; LPP2 did not show a preference for either substrate, and LPP1 preferred DGPP [43]. AtLPP1 was differentially expressed in
response to stress conditions: expression was transiently induced after genotoxic stress and after elicitor treatments. Based on these results, *LPP1* is expected to be a PAP involved in attenuating the signaling functions of PA in response to stress conditions [43]. Seeds mutant for *AtLPP2* are more sensitive to ABA and the gibberellic acid biosynthesis inhibitor uniconazole during germination, though adult plants did not show this sensitivity [44]. Studies also suggest that LPP2 acts on the PA signaling pool which is downstream of ABA but upstream of ABI4, and parallel with ABI3 [44]. Thus, LPP2 is a negative regulator acting in the ABA signaling network regulating seed germination [44].

Based on the data from yeast and mammals, one would expect that one of the two lipin homologs would be involved in oil synthesis. However, this does not appear to be the case. Individual knockouts of the At*PAH1* and At*PAH2* genes do not cause a phenotype, and only when both are knocked out is there a 15% decrease in seed oil [45]. The leaves and roots of *pah1pah2* plants show an increase in total lipid and phospholipid content when measured based on fresh weight [45,46]. The mutant showed induced expression of phospholipid synthesis genes, leading to an increase in total ER membrane, similar to the yeast *PAH1* mutant [45,46]. During phosphate starvation, the mutants are significantly affected and galactolipid biosynthesis was impaired, indicating that *PAH1* and *PAH2* could act as mediators of membrane lipid remodeling during phosphate starvation [42]. These data indicate that, similar to the yeast *ScPAH1*, the Arabidopsis *AtPAH1* is involved in phospholipid synthesis and membrane management in the ER. However, the modest effect on oil synthesis starkly contrasts with the TAG synthesis phenotype in yeast and mammals, and indicates that the PAP involved in oil synthesis in Arabidopsis does not fit with the precedent set by yeast and mammals. There must be another non-lipin PAP in Arabidopsis that
accounts for the remaining 85% of the seed oil in \textit{pah1pah2}, which could be encoded either by an LPP homolog or by a yet-unidentified PAP protein sequence.

In this paper, I will screen the Arabidopsis genome for genes that are homologous to known PAPs, and analyze their known properties to determine if they could be PAPs involved in oil synthesis. I also confirm that expression of these genes is consistent with a PAP producing oil in the seed. Four genes have the potential to act redundantly with the main PAP involved in oil syntheses, and will be studied in that capacity. Only one of the genes homologous to known PAPs could still be the main PAP involved in oil synthesis, \textit{LPPβ}. However, there remains the possibility that the PAP involved in oil synthesis could be divergent from known sequences and therefore would not be identified in this project.
2.2. MATERIALS AND METHODS

2.2.1. PLANT MATERIALS AND GROWTH CONDITIONS

For tissue expression analysis, wild-type seeds of Arabidopsis (ecotype Columbia-0) were planted on soil and stratified for 2-3 d at 4°C in the dark. Plants were moved to a growth chamber and grown at 22°C under continuous light with a light intensity of 100-150 µmol m⁻² s⁻¹ light. For photoperiod expression analysis, wild type (WT) seeds were sterilized by mixing seeds in 70% ethanol for 10-15 minutes, then were washed with 100% ethanol and allowed to dry. Dry seeds were sprinkled on ½ MS + 1% sucrose plates and stratified for 4 d, then placed in a 22°C growth chamber with continuous light (100-150 µmol m⁻² s⁻¹ light). To explore subcellular localization of LPPβ, SPP1, and LPAP1, transient expression was performed in Nicotiana tabacum ‘Petit Havana’. Seeds were sterilized and grown on 3x MS medium with 0.6% agar for two weeks, transferred to rockwool and grown two weeks more, then planted in soil. Plants were grown at 25°C under continuous light with light intensity of 120-150 µmol m⁻² s⁻¹ light.

2.2.2. GENE IDENTIFICATION AND LITERATURE REVIEW

Mouse, human, and yeast PAP, LPP, DPP, APP, and lipin protein sequences were used to search the Arabidopsis proteome using the protein Basic Local Alignment Search Tool (BLASTp) [47]. Databases were also searched for genes annotated as PAPs, LPPs and DPPs, and the literature was searched for LPP, DPP, and PAP homologs which have been studied. Literature was screened to rule out PAPs which were involved in other pathways and to gain bioinformatics data detailed below.

Protein sequences of the 11 “PAP2” type genes were compared with other known PAPs; protein sequences were aligned with LPP protein sequences from human used by Nakamura et al.
[41]: human (“Hs”), mouse (“Mm”), Anabaena (Nostoc) sp. PCC7120 (“Ana”), Chlorobaculum tepidum (“Ct”), Synechocystis sp. PCC6803 (“Syn”), Saccharomyces cerevisiae (“Sc”), and Arabidopsis thaliana (“At”). In comparison with the previous study, we used the full protein sequence instead of only the domain sequences, two additional Arabidopsis sequences were used (R1L and LPAP1), and the sequence for the Escherichia coli (“Ec”) PgpB protein was included. A phylogenetic tree was constructed using ClustalW (www.genome.jp/tools/clustalw).

2.2.3. BIOINFORMATICS ANALYSIS

Known properties of all gene products were analyzed by searching relevant databases for experimental data first, and then examined for data generated by prediction programs. Localization was examined using the Subcellular Localization of Proteins in Arabidopsis Database (SUBA). This database includes data from published chimeric fusion studies and proteomic localization experiments, as well as predictive programs such as WoLF PSORT and TargetP [48-50].

Tissue expression analysis was used to identify PAP candidates whose transcripts are differentially expressed in tissues where TAG synthesis is high, using the Electronic Fluorescent Pictograph for visualization [51]. Expression of putative PAPs was analyzed in developing seed tissue in comparison with leaf and overall expression levels.

Coexpression with the known TAG synthesis genes in Arabidopsis could also indicate if the genes are expressed at the same time as TAG is being produced. Coexpression of PAP genes with those known to be involved in TAG synthesis, specifically DGAT1 and PDAT1, was analyzed using GeneCAT [52].

Expression patterns of a PAP involved in oil synthesis should be similar in Arabidopsis and other oilseed species because the PAP enzymatic function is expected to be required for
survival, and thus it is expected to be conserved among oilseed species. Therefore, transcript expression of the 13 PAP genes was analyzed on several platforms. Published pyrosequencing studies have examined expression of all 13 gene homologs in the oilseed species *Brassica napus*, *Ricinus communis*, *Euonymus alatus* and *Tropaeolum majus* [53]. Additionally, the 13 Arabidopsis protein sequences were searched against the transcriptomes of soybean and castor bean to determine if all of the genes have homologs and to better understand their function from their tissue expression. The soybean proteome at www.soybase.org and the castor proteome at www.phytozome.net were each searched using BLASTp with of the 13 putative PAP protein sequences. Top soybean and castor hits were then searched back against the Arabidopsis proteome at www.arabidopsis.org to determine best matches. Expression data was not available for any of the soybean PAP homologs.

To determine membrane integration and general protein structure, transmembrane regions in the protein were identified using TMHMM [54].

### 2.2.4. Silique Tissue Expression

To confirm seed transcript expression, 7-10 day old siliques and mature leaves were ground under liquid nitrogen, and RNA was isolated from tissue using the TRIzol Plus RNA Purification Kit (Life Technologies). DNA contamination was removed using the DNA-Free RNA Kit (Zymo Research), and complementary DNA generated using the SuperScript III First Strand Synthesis kit (Life Technologies). *LPPβ* full-length sequence was amplified using LBF1 and LBR1 primers and 35 cycles of PCR, with Actin8 as a control (see Table 5 for all primer sequences found in this chapter). Full length transcripts were amplified using LDF4 and LDR3 for *SPP1*, and R1L F2 and R1L R2 for *RIL*.
2.2.5. SUBCELLULAR LOCALIZATION

Coding sequence were amplified from silique cDNA ([55]; Quanta) and cloned into pENTR: LPPβ was amplified using the LBF1-G and LBR1 primers; SPP1 was amplified using LDF1-G and LDR3, and LPAP1 was amplified using UP1F1G and UP1R3. The sequenced gene was then cloned into pB7WGF2 using the LR reaction of Gateway cloning, to produce an N-terminal fusion under control of the strong, constitutive 35S cauliflower mosaic virus promoter. BiP, a HSP70 molecular chaperone known to localize to the ER, was a kind gift from Dr. Hwang (Pohang University, Korea [56]) and was sub-cloned into pB7WGR2 to create a fusion with RFP. pENTR-GUS was obtained from the Gateway cloning kit. BiP and GUS were then cloned into pB7WGR2 and pB7WGF2 as ER and cytoplasm controls, respectively [57]. Sequenced plasmids were transformed into Agrobacterium GV3101 using electroporation.

Tobacco leaves (Nicotiana tabacum cv. Petit Havana) were infiltrated with a mixture of the experimental- and ER-control-containing bacteria in the presence of acetosyringone (100 uM final concentration). Epidermal peels were visualized on a confocal microscope 48 hours after infiltration. An argon laser (488 nm) and HeNe laser (584 nm) were used to excite GFP and RFP, respectively. Filters isolated 490-520 nm for GFP, 595-630 nm for RFP, and 650-750 nm for chlorophyll autofluorescence.

2.2.6. PHOTOPERIOD-INDUCED EXPRESSION ANALYSIS

To explore whether or not LPPβ exhibits the same photoperiod expression as PnFL-1, transcript levels were examined after dark induction. Sterilized seeds were germinated on ½ MS plates and allowed to grow until full cotyledon expansion, about 4 days. Seedlings were placed in the dark, and tissue was harvested by freezing the cotyledons under liquid nitrogen; for samples harvested during dark induction, tissue was harvested in complete darkness. RNA was extracted
using the Trizol method (Invitrogen) and DNA contamination removed by digestion with DNaseI (Zymo Research). Complimentary DNA was prepared using Quanta’s qScript DNA SuperMix cDNA synthesis kit. Expression level was quantified using the Quanta SYBR FastMix kit on the MX3005P qPCR system (Stratagene). LPPβ specific primers LB qPCR-F3 and LB qPCR-R2 amplified a ~200 bp segment of LPPβ and the TUB2 qPCR-F and TUB2 qPCR-R primers amplified ~200bp of the Tubulin2 gene. LPPβ gene expression was normalized to AtTUB2 and the relative fold change between brown seeds and red seeds was calculated (=1/(2^exp-control)).
2.3. RESULTS AND DISCUSSION

2.3.1. GENE IDENTIFICATION

The LPP and DPP BLASTp sequence searches consistently resulted in eleven protein sequences (AtLPPβ, AtLPPγ, AtLPPε1, AtLPPε2, AtSPP1, AtLPP1, AtLPP2, AtLPP3, AtLPP4, LysoPhosphatidic Acid Phosphatase 1 (LPAP1), and Reduced Oleate Desaturation1-Like (R1L); see Table 1). When *Saccharomyces cerevisiae* PAH1 and *Homo sapiens* LPIN protein sequences were used as search queries against the Arabidopsis proteome, two proteins were identified (AtPAH1 and AtPAH2). The fourth and final PAP in *Saccharomyces cerevisiae* is Actin Patch Protein 1 (ScAPP1), which did not have any homologs in Arabidopsis. Through these combined approaches I identified a total of 13 putative Arabidopsis PAPs.

The 13 genes identified fall naturally into two groups. The PAP1 family is Mg\(^{2+}\) dependent and, in this case, is comprised of only the two genes homologous to lipins from mammals and yeast: AtPAH1 and AtPAH2. The second group contains the remaining 11 genes, which are part of the Mg\(^{2+}\)-independent PAP2 (LPP) family. While *PAH1* and *PAH2* share high homology with each other and with other lipins, the LPP family of 11 genes shares much lower homology. This can be seen in an unrooted tree in Figure 1 and alignment in Figure 2. Indeed, Nakamura et al. used domain sequences instead of the full protein sequences to generate an unrooted phylogenetic tree, and separated the Arabidopsis sequences into the “prokaryotic” and “eukaryotic” sequences [41]. They noted that the four “eukaryotic” LPP genes (LPP1-4) share more homology with mammalian and yeast LPP’s, while the “prokaryotic” LPP genes (LPPγ, LPPε1, LPPε2, and SPP1) form a second clade and share more homology with cyanobacterial LPP’s [41]. In their analysis, LPPβ was even more closely related to the cyanobacterial genes, and did not fall within either of the primarily-plant clades [41]. When the full protein sequences were used in my analysis, the
“prokaryotic” clade (LPPε1, LPPε2, LPPγ, SPP1, and LPAP1) was most divergent from known sequences, the “eukaryotic” clade (LPP1-4) was most similar to yeast and Chlorobaculum sequences, and the LPPβ and R1L sequences were most similar to bacterial and mammalian clades (Figure 1).

Most of the 13 genes identified also contained the expected conserved domains for their class of PAP. PAH1 and PAH2 shared a DXDXT catalytic motif as well as conserved N-terminal and C-terminal domains, as described in Eastmond et al. [45]. The PAP2 family has three conserved domains [10], which is present in most of the LPP family members identified. The *E. coli* protein PgpB is the only PAP2-like protein whose crystal structure has been solved. In PgpB, the three domains form a cluster on the periplasmic side of the membrane, just above the transmembrane domains, and at the top of a cleft where the substrate (such as PA) would sit, allowing access to and cleavage of the head group (phosphate, in the case of a true PAP; [58]). While the PgpB is not highly conserved with LPPβ or the other PAPs, this gives insight as to the mode of action of the PAP2 family. Of the Arabidopsis homologs identified, only 9 of the 11 genes identified by homology shared the three conserved domains (Figure 2). LPAP1 and R1L did not have any of the domains, while all of the remaining 9 genes had all three domains (See Figure 2). This was a definite indication that LPAP1 and R1L may not be prime candidates as PAPs; but since the homology to PgpB is so low, and so much is still to be learned about the mode of action in PAPs, I was hesitant to rule them out as possibly having PAP activity. Indeed, despite the lack of PAP2 domains, LPAP1 does show low PAP activity [59].

2.3.2. LITERATURE REVIEW

Three characteristics of the putative PAP genes excluded them from consideration in this study: a lack of PAP activity, localization to the plastid (or another location other than the
cytoplasm or ER), or lack of transcript expression in the seed. First, a protein that under rigorous scrutiny showed no PAP activity would be of little interest to this study, especially if it showed another enzymatic activity. However, a gene which showed even low PAP activity could still act redundantly with other PAPs and possibly compensate for a knockout of the main PAP. Second, the oil synthesis pathway in Arabidopsis is part of the eukaryotic pathway of lipid synthesis and is therefore localized to the ER membrane. It can involve proteins that are either membrane associated or cytoplasmic. A protein which localized to the plastid would likely be involved in the prokaryotic pathway of lipid synthesis, and would not be involved in oil synthesis. Localization to organelles other than the ER or cytoplasm would also likely not be associated with oil synthesis. Lastly, we are interested in oil synthesis in the context of oilseeds that are harvested for food, fuel, or feedstocks. Genes which are not expressed in the seed are likely not involved in seed oil synthesis, and therefore were ruled out. A summary of bioinformatics and literature review results can be found in Table 1.

Several genes were ruled out summarily. \textit{AtLPP}\textsubscript{γ} (At5g03080), \textit{AtLPP}\textsubscript{ε1} (At3g50920), and \textit{AtLPP}\textsubscript{ε2} (At5g66450) have all been shown to localize to the plastid, and are published as the primary plastidic PAPs forming thylakoid membrane lipids [41]. In their studies of \textit{Reduced Oleate Desaturation 1} (\textit{ROD1}), Lu et al [60] analyzed At3g15830. They found that this gene, which I am here calling \textit{ROD1-Like} (\textit{R1L}), was not expressed in the silique; and my RT-PCR studies also confirmed that the gene is not expressed in the seed. (See “Seed Tissue Expression,” section 2.3.5.) Similarly, \textit{AtLPP4} (At3g18220), is expressed only in the pollen, indicating that it may be involved in oil synthesis in the pollen, but not in seeds.

\textit{AtLPP1} (At2g01180) and \textit{AtLPP2} (At1g15080) are not expressed in the seed, and both have roles in signaling processes: \textit{LPP1} is involved in PA and DGPP signaling [43] and \textit{LPP2} is
involved in ABA signaling [44]. *AtLPP3* (At3g02600) is highly homologous to *LPP1* and *LPP2* [43]. These three genes form an LPP clade, which is further separated into two sub-groups: one composed of *LPP1*, and one composed of *LPP2* and *LPP3* [43]. The authors chose to study one gene from each sub-group, and therefore chose *LPP2* to represent both *LPP2* and *LPP3* [43]. I chose to follow the grouping set forth by Pierrugues et al., and assumed that *LPP3* is also involved in lipid signaling [43].

Five genes remained after the exclusion of putative PAPs clearly not involved in oil synthesis. *AtPAH1* (At3g09560) and *AtPAH2* (At5g42870) were shown to have PAP activity and are expressed in the seed, but have been shown to be involved in membrane homeostasis and lipid remodeling under phosphate stress. While I did not characterize PAH1 or PAH2 except as controls to my experiments, I did use the *pah1pah2* mutant background to determine if other putative PAPs were redundant with *PAH1* and *PAH2*.

Two genes, *AtSPP1* and an unnamed lysophosphatidic acid phosphatase I am here calling *Lysophosphatidic acid phosphatase* (*AtLPAP1*), showed interesting characteristics which warranted further study. *AtLPAP1* (At3g03520) had lysophosphatidic acid phosphatase activity when expressed in yeast, but it also showed low levels of PAP activity [59]. This activity, coupled with its expression in the seed and lack of a clear physiological role, warranted further study. The *sphingoid phosphate phosphatase 1* (*SPP1*) gene (At3g58490; formerly LPPδ) has slightly more information available, and is implicated in ABA-mediated stomatal closure [61]. Measurements of long-chain base-1-phosphate levels in yeast expressing *SPP1* and mutant Arabidopsis suggested that it encodes a long-chain base-1-phosphate phosphatase, but rigorous enzyme assays, including PAP activity, were not completed [61]. Due to the lack of molecular data on *SPP1*’s enzymatic activity, it also warranted further study. My goal with these two genes was to determine if they
were redundant with $PAH1$ and $PAH2$ and the remaining putative PAPs; and to confirm, if convenient, their lack of PAP activity.

The final and most promising candidate was $AtLPP\beta$ ($At4g22550$). Information about expression was limited because there are no probes for it on the AffyMetrix chip used to determine tissue expression for most genes in Arabidopsis, and it had never been studied in depth. The only information available was that it was not found in the chloroplast [41]. Unpublished reports also suggest that $LPP\beta$ encodes a protein that displays PAP activity [62]. A BLASTp analysis of putative homologs produced many putative or unknown proteins, but only one protein with any published information. $PnFL-1$ from the short day plant Japanese Morning Glory ($Ipomea nil$, formerly $Pharbitis nil$) has 58% homology to $LPP\beta$, and is expressed during the inductive dark period necessary for flowering, but expression does not occur if the plant is exposed to a night break (15 min of light during dark period) [63]. The two proteins are similar in size ($LPP\beta$ is 213 amino acids; $PnFL-1$ 209 aa); and a reciprocal BLASTp shows that $LPP\beta$ is the top homolog of $PnFL-1$ in the Arabidopsis genome (43% identity covering 93% of $LPP\beta$), followed by $LPP\gamma$ (32% identity covering 60% of $LPP\gamma$). These reports have led some to believe that $LPP\beta$ is a photoperiodically expressed homolog of $PnFL-1$ [62], but this has not been experimentally confirmed. My further analysis of whether $LPP\beta$ expression is diurnally controlled is described in Section 2.3.7, below.

2.3.3 Bioinformatics Analysis

All of the bioinformatics data analyzed can be found in Tables 1-4, but my analysis will focus on those six genes which have not yet been excluded from consideration: $LPP\beta$, $RIL$, $LPAP1$, $SPP1$, $PAH1$, and $PAH2$. Tissue expression was analyzed to determine if genes were expressed in developing seeds, the main site of oil synthesis. Online resources are largely based
on data produced from the Affymetrix ATH1 microarray chip which does not represent all the genes in Arabidopsis. Analysis of the database showed that \textit{LPAP1}, \textit{PAH1}, and \textit{PAH2} are all expressed in the developing seed, but \textit{SPP1} and \textit{LPPβ} are not on the chip (Column 8, Table 1). \textit{ROD1-Like} also showed seed expression, which conflicted with the expression data in Lu et al. [60]; therefore \textit{LPPβ}, \textit{SPP1}, and \textit{R1L} required confirmation of their expression (See section 2.3.4, “Seed Tissue Expression”).

Subcellular localization of the 13 identified putative PAPs was analyzed to determine if the identified genes localized to organelles other than the ER or cytoplasm, and therefore could be ruled out. Several PAPs have published localization data based on experimental evidence in conjunction with the data from the predictive programs found in the SUBA database. \textit{PAH1} and \textit{PAH2} are localized to the cytoplasm [42], and \textit{SPP1} is localized to the ER [61] (see Table 1, column 5), consistent with possible involvement in the eukaryotic pathway of lipid synthesis. None of these three proteins showed a consistent localization from the predictive programs. \textit{LPPβ} had 7 of 25 predictions for the ER, and only 1 of 25 predictions indicated plastidial localization. With some data indicating ER localization and without a confident prediction to the plastid, \textit{LPPβ} remains a potential PAP involved in oil synthesis.

Coexpression with known Arabidopsis TAG synthesis genes was examined to determine if the putative PAPs are involved in oil synthesis. Although two genes may have similar expression patterns simply by chance, coexpression often indicates that the genes are involved in similar processes. \textit{DGAT1} (At2g19450) and \textit{PDAT1} (At5g13640) both produce TAG, and together represent the last step generating TAG [64]. Coexpression was examined using GeneCat, which queries published expression data based on the Affymetrix ATH1 gene chip. As was seen in earlier analyses, \textit{LPPβ} and \textit{SPP1} are not on the ATH1 chip, and therefore data was not available
on these genes. Data from the remaining genes showed that they are not strongly coexpressed with DGAT1 and PDAT1; coexpression values were between 0.22 and -0.10, with 1.00 being perfectly correlated and -1.00 being perfectly anti-correlated (See Table 1, column 9).

Other oilseed species were analyzed to determine identity and, if possible, expression levels of PAP homologs in those species. The castor bean (*Ricinus communis*) is an oilseed species which produces ricinoleic acid, a hydroxylated fatty acid important in manufacturing many items including lubricants, paints, plastics, and pharmaceuticals. When the thirteen PAPs of interest were searched against the Castor genome, many had multiple homologs (see Table 2). LPPβ had only one primary homolog, meaning that when the homolog to LPPβ was searched back against the Arabidopsis genome, LPPβ was the top candidate. This homolog is annotated as a sphingosine-1-phosphate phosphohydrolase, because of its (and LPPβ’s) homology to SPP1; however this annotation, found in many species, is not based on experimental evidence.

Homologs of the identified putative PAPs were also analyzed in the soybean plant, *Glycine max*. Soybean is a crop plant grown for both its oil and protein in the seed; after extraction of the oil, the pressed meal is high in protein and is a common addition to human and animal food. As in Castor, LPPβ only had one homolog in Soybean (see Table 3). LPP4 and LPPε1 had no primary homolog, but the remaining 10 Arabidopsis genes all had 2 or 4 homologs.

Many gene expression analyses, even those done in species other than Arabidopsis, have been based on microarray data, and LPPβ homologs are rarely on the microarrays. One data set that did include LPPβ homologs is 454 pyrosequencing data done by Troncoso-Ponce and Kilaru and colleagues [53] (see Table 4). They sequenced transcripts from the oil producing seed tissues in the oilseeds *Brassica napus, Tropaeolum majus, Ricinus communis, and Euonymus alatus*. The oil-producing tissues of all four species showed low expression of LPPβ homologs, but expression
was present. In the Brassicales species *B. napus* and *T. majus*, oil is stored in the embryo; but in the Fabids *R. communis* and *E. alatus*, oil is stored in the endosperm. *R. communis* and *E. elatus* provide an interesting opportunity to compare the oil producing tissue (endosperm) with non-producing tissue, the embryo. Interestingly, virtually all of the PAPs studied had higher in the embryo than in the endosperm, except SPP1 in *E. alatus*. This could indicate that none of the PAPs studied are the main oil-producing PAP in these species, but a PAP that is involved in membrane lipid synthesis as well as oil synthesis may not be upregulated in oil producing tissues.

### 2.3.4. Bioinformatics of LPPβ

After exploring data on all of the putative PAPs, it was clear that LPPβ was the primary PAP of interest for this study. Therefore, the remaining bioinformatics analyses focused only on LPPβ.

Since LPPβ is not on the ATH1 microarray chip, we investigated other sources for tissue expression data in Arabidopsis. The AGRONOMICS tiling array is designed to allow expression profiling of more than 30,000 genes in Arabidopsis, with results that are comparable to the ATH1 chip [65]. LPPβ is represented on the AGRONOMICS array and the four anatomical tissues tested (juvenile and adult leaf, seedling, and flower) all show high expression levels (See Figure 3A). Unfortunately, neither the seed nor silique were tested, so expression of LB in the seed was confirmed using RT-PCR (See “Seed Tissue Expression,” section 2.3.5).

The protein structure of LPPβ was also investigated. Analysis using TMHMM indicated that LPPβ has four transmembrane domains, and is probably an integral membrane protein. It was then compared with the only known PAP2-type protein whose crystal structure has been solved, the phosphatidylglycerol-phosphate phosphatase B in *E. coli* (ecPgpB) [58]. PgpB and LPPβ share very little homology, but LPPβ does share the three signature PAP2 motifs: C1, “KX₆RP”; C2,
“PSGH”; and C3, “SRX\textsubscript{3}HX\textsubscript{3}D” (Figure 2) [10]. The four transmembrane domains indicated by TMHMM could be capable of forming a transmembrane core, as is seen in the PgpB structure (Figure 3B). These data indicate an integral membrane structure and suggest PAP or related activity.

### 2.3.5. Seed Tissue Expression

$LPP\beta$, $SPP1$, and $R1L$ required confirmation that they actually were expressed in the seed, which could be accomplished by looking at expression from the entire silique using RT-PCR. While the genomic reaction worked for $R1L$, indicating that the PCR reaction worked, the coding region of $R1L$ was not able to be amplified from either leaf or silique cDNA, indicating that it is not expressed in these tissues (see Figure 4A). This is consistent with the data from Lu et al., which showed expression only in flowers and not in seedlings, young leaves, or siliques [60]. Therefore, $R1L$ was conclusively dismissed from this study. $SPP1$ and $LPP\beta$ both showed bands in the leaf and silique cDNA reactions, indicating that they are expressed in these tissues and warrant further study (See Figures 4). The $LPP\beta$ gene does not have introns, so the genomic and cDNA bands are the same size, but the Actin8 reaction showed that the cDNA was not contaminated with genomic DNA. Therefore, after tissue expression analysis, $LPP\beta$ remains the top PAP candidate, $SPP1$ remains a gene of interest because of its possible redundancy with $LPP\beta$, and $R1L$ is no longer a gene of interest.

### 2.3.6. Subcellular Localization

To confirm that the LPP\textsubscript{β}, SPP1, and LPAP1 proteins localize to the ER or cytoplasm, they were expressed in tobacco epidermal cells as N-terminal GFP fusions. GUS was fused to GFP as a cytosolic control, and a RFP-BiP fusion served as an ER control. Chlorophyll autofluorescence allowed visualization of the plastid. Plastidial localization would indicate involvement in the
prokaryotic pathway of lipid synthesis, as in the case of LPPγ, LPPε1 and LPPε2 [41] and would be cause to set that gene aside from consideration. Similarly, localization to other organelles such as the peroxisome, vacuole, mitochondrion or nucleus would indicate that the gene product is not involved in oil synthesis. However, PAPs involved in oil synthesis could be localized to either the cytoplasm or the endoplasmic reticulum. As seen in Figure 5, the GFP-GUS cytoplasmic control appears to co-localize to the RFP-BiP endoplasmic control, so discriminating between cytoplasmic and ER localization is not possible with this method. The four transmembrane domains in LPPβ argues for an integral membrane association, so cytoplasmic localization is not expected. LPPβ, as well as LPAP1 and SPP1, appear to localize to the cytoplasm/ER (Figure 5): the green gene fusion and the red ER appear to colocalize. The blue chlorophyll autofluorescence does not colocalize with the green fusions, and in some cases even fills in spaces in the ER occupied by the plastids. Thus, LPPβ appears to localize to the ER and remains a strong candidate for the PAP involved in oil synthesis. LPAP1 and SPP1 also appear to localize to the ER/cytosol, and could still be acting redundantly with LPPβ in oil synthesis.

2.3.7. PHOTOPERIOD EXPRESSION OF LPPβ

The LPPβ gene has not been carefully studied and there are few homologs of LPPβ with any published data. The only published information on LPPβ homologs is on the Pharbitis nil (Ipomea nil) gene FL-1, which is expressed during the flower-inducing dark period [63]. Expression during the light hours was negligible, but after 10 hours of darkness, PnFL-1 expression was increased 100 fold. A night break (15 min light at the 8 hour time point) extinguished expression, and expression ended shortly after dawn. Thus, the authors surmised that PnFL-1 could be involved in inducing flowering in the short-day plant (see [63]).
Since the only homolog of LPPβ that has been studied was published as a photoperiod-expressed gene involved in the inductive dark period, examining the photoperiod expression of LPPβ was warranted. Since morning glory is a short day plant and Arabidopsis is a long day plant, both long day and short day light conditions were tested; a continuous light condition was also tested as a control. Tissue from fully expanded cotyledons was harvested before light treatment, at the end of an 8 hour dark period and two hours after dawn, at the end of a 16 hour dark period and 2 hours after dawn, at the 24 hour time point after both treatments, and after 24 hours without any dark treatment. While the PnFL-1 gene showed about a 100-fold increase in expression from 0h to 10h dark, LPPβ showed less than a 1-fold decrease in all conditions tested (Figure 6). Thus, expression of LPPβ does not appear to change enough to warrant involvement in any photoperiod-induced flowering pathway.

Photoperiod expression in a PAP involved in oil synthesis is not expected but also would not be unexpected. Many metabolic enzymes show diurnal expression patterns (e.g. see [66,67]), and one would expect that the substrates for fatty acid synthesis, and therefore lipid and oil synthesis, would also have similar regulation. Indeed, it has been known for many years that fatty acid synthesis levels and fatty acid composition undergo diurnal fluctuations [68], and more recently proteins involved in lipid synthesis (ACBP3) and oil synthesis (DGAT1 and DGAT3) have been shown to be diurnally regulated [69]. However, several oil synthesis genes do not show diurnal regulation (e.g.DGAT2 and PDAT1), indicating that not all steps in the pathway are similarly regulated. Thus, the consistent expression level of LPPβ throughout the day does exclude it from being a photoperiod regulator, but it does not affect its status as a possible PAP involved in oil synthesis.
2.4. CONCLUSION

The Arabidopsis genome contains thirteen genes which are homologous to known PAP genes. Most have traits inconsistent with a PAP involved in oil synthesis, and have been set aside from consideration. \textit{AtLPP\gamma}, \textit{AtLPP\epsilon1}, and \textit{AtLPP\epsilon2} are localized to the plastid and are involved in the prokaryotic pathway of lipid synthesis. \textit{AtLPP1}, \textit{AtLPP2}, and \textit{AtLPP3} are involved in signaling pathways. Neither \textit{AtLPP4} nor \textit{AtR1L} are expressed in the seed. Several other genes display characteristics that indicate they could be redundant with the PAP involved in oil synthesis. \textit{AtPAH1} and \textit{AtPAH2} are involved in oil synthesis, and together account for 15\% of seed oil. \textit{AtLPAP1} showed some PAP activity, and \textit{AtSPP1} was not assayed for PAP activity. The top candidate for a PAP involved in oil synthesis is \textit{AtLPP\beta}, a gene whose tissue expression and subcellular localization are consistent with a PAP involved in oil synthesis in seeds. Therefore, \textit{AtLPP\beta} will be analyzed in Chapter 3 in conjunction with \textit{AtPAH1}, \textit{AtPAH2}, \textit{AtLAP1}, and \textit{AtSPP1} to determine if they function redundantly.
Figure 1. Phylogenetic tree of LPP homologs. Rooted phylogenetic tree of human, mouse, yeast, bacterial, and plant LPPs drawn based on full protein sequences. Arabidopsis sequences are noted in bold, and LPPβ is highlighted in yellow. The horizontal branch lengths are proportional to sequence divergence. Hs, Homo sapiens; Mm, Mus musculus; Ana, Anabaena (Nostoc) sp. PCC7120; Ct, Chlorobaculum tepidum; Syn, Synechocystis sp. PCC6803; Ec, Escherichia coli; Sc, Saccharomyces cerevisiae; At, Arabidopsis thaliana.
E. coli PgpB   ---------------------------- ----------------------------------------
Yeast ScLPP   ---------------------------- ----MISVMADEKHKEYFKLYYFQYMIIGLCTILFLYSEI
AtLPPb; AT4G22550.1   ---------------------------- ----------------------------------------
AtLPP1; AT2G01180.1   -------------------------MTI GSFFSSLLFWRNSQDQEAQRGRMQEIDLSVHTIKSHGGRV
AtLPP2; AT1G15080   ARFHMDWILLHILIEVILNHFPPGEVDULTQDNPITFAPVAVILAVPFPIVACY
AtLPP3; AT3G02600   ARTHMDWILLViviclLIlIHHFPPGKDMTDLPSKNTPIWSVPPAVYAVPLIPEFVC
AtLPP4; AT3G18220.1   AREHLCDWILLLILLLIVIEIGLVNFFHPRPMFDTPMVAPWICILPICIFIVY
AtLPPe1; AT3G50920   LFVSSGKSMDOLVTKNAQDDRGPDQAFQAGISNSLSELQLVQQQDGIAGIAALRKLKVAIVA
AtLPPe2; AT5G66450.1   IWSVSGFKSMADLVKTNARRDGEDRFQALQQEAFNSSSELQNELVSDAGDGIEAIAANRLSKWIVAA
AtSPP1; AT3G58490.1   LVGIVTWICASSYLKFTHKFRSLLQWVRQVGVGLPLIILRQCNQVGLDFSSGLCVSVPSFYTA
AtLPPg; AT5G03080   ---------------------------M DLIPQQLKAVTLTHVRYRPGDQLGHFLAWISLVPVFISLG
E. coli PgpB   ---MRSIARRTAVGAALLLVMPVAVWIS GWRWQPGEQSWLLKAAFWVTETVTQPWGVITHLILFGWFL
Yeast ScLPP   SLVFRGQNIEFSLDDPSISKRYVPNELV GPLECLILSVGLSNMVVFWTCMFDKDLLKKNRVKRLRERP
AtLPPb; AT4G22550.1   SLSFLLSPPLRSFLVPFLLGLLFDLIFV GIVKLIFRRARPAYNHPSMSA-------------------
AtLPP1; AT2G01180.1   YLKRTCVYDLHHSILGLLFAVLITGVIT DSIKVATGRPRPNFYWRCFPD------------------G
AtLPP2; AT1G15080   YFIRNDVYDLHHAILLGLLFSVLITGVIT DAIKDAVGRPRPDFFWRCFPD-----------------GI
AtLPP3; AT3G02600   YFRRRDVYDLHHAVLGLLYSVLVTAVLT DAIKNAVGRPRPDFFWRCFPD------------------G
AtLPP4; AT3G18220.1   YYYYRRDVYDLHHAILGIGFSCLVTGVTTDSIAVGRFNYFFCRPFDGK
AtLPPe1; AT3G50920   LFGSIILLRHDGAALWAVIGSISNSALS VVLKRILNQERPTTTLRSDP--------------------
AtLPPe2; AT5G66450.1   LFGSVLLLRHDGAALWAVIGSVSNSVLS VALKRILNQERPVATLRSDP--------------------
AtSPP1; AT3G58490.1   FLPLLFWSGHGRLARQMTLLIAFCDYLG NCIKDVVSAPRPSCPPVRRITAT-----------------
AtLPPg; AT5G03080   GFVSHFILFRELQIGFFPGILVLQINFQINEFATVESGACT
Domain 1
E. coli PgpB    WPRRRTLTIAILLVWATGVMGSRLLLGM HWPRDLVVATLISWALVAVATWLAQRICGPLTPEEVN
Yeast ScLPP   FTTRNTRSCIWCPLLALVVMVSRVIDHR HHWYDVVSGAVLAFLVIYCCWKWTFTNLAKRDILPSPVSV
AtLPPb; AT4G22550.1   DGDVKVEVVVVVWIWATVTAISRILLGR HYVLDVAAGAFLGIVEALFALRFLRFDEMIFGR-------
AtLPP1; AT2G01180.1   GHVAKLCLVIFPLLAACLVGISRVDDYW HHWQDVFGGAIIGLTVAAFCYRQFYPNPYHEEGWGPYAYF
AtLPP2; AT1G15080   GHVAKLCIVILPLLVAALVGISRVDDYW HHWQDVFAGGLLGLAISTICYLQFFPPPYHTEGWGPYAYF
AtLPP3; AT3G02600   GHVAKLCLVFLPILISILIGFVDDHAYQVFAGAIGLVATPVCYQVFPPYDPENNGHYPAYF
AtLPP4; AT3G18220.1   PKHFDTPDKCCHVGGKIIKEGKSYCPHTSWSFAGLTFLAWY
AtLPPe1; AT3G50920   TNGVSLFLSGLILALGSYFIRLRVSQKL HTSSQVVVGAIVGSLFCILWYTMWNSLLREAFEASLLVQI
AtLPPe2; AT5G66450.1   TNVLSLFLSGFILALGSYFTWLRVSQKL HTTSSQVVVGAIVGSVYSTLWYVTWNSLVLEAFTSFVSQI
AtSPP1; AT3G58490.1   SVSIQYYGFALACLLVALIAFGRVYLGM HSVVDIVSGLAIGVLILGLWLTVNEKLDDFITSKQNVSSF
AtLPPg; AT5G03080   GLRSWPMNLLWMLAVTVMYBRVYLYHTVAQTAGALCQVGVSAWFQVFVNSVVLVVFPPFVEESVL
Domain 3
Figure 2. Conserved Domains in the PAP2 family of PAPs. Alignment of full protein sequences in known LPPs and newly identified LPPs. LPPβ is highlighted in yellow. The three conserved PAP2 domains are highlighted in blue. Underlined regions indicate transmembrane domains in LPPβ and in EcPgpB, the only PAP2 enzyme whose crystal structure has been solved.
<table>
<thead>
<tr>
<th>At Accession</th>
<th>Name</th>
<th>aa</th>
<th>PAP Activity?</th>
<th>Localization</th>
<th>KO decreases FA/lipid levels?</th>
<th>Involved in other pathway?</th>
<th>Tissue Expression</th>
<th>Coexpression*</th>
</tr>
</thead>
<tbody>
<tr>
<td>At4g22550</td>
<td>AtLPPβ (LB)</td>
<td>213</td>
<td>PAP Activity?</td>
<td>not chloroplast [41]</td>
<td></td>
<td></td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>At3g02600</td>
<td>AtLPP3</td>
<td>364</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td></td>
<td>--</td>
</tr>
<tr>
<td>At3g15830</td>
<td>Rod1Like (RIL)</td>
<td>296</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>not seed [51]</td>
<td>-0.103</td>
</tr>
<tr>
<td>At3g18220</td>
<td>AtLPP4</td>
<td>308</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>pollen [44]</td>
<td>0.216</td>
</tr>
<tr>
<td>At2g01180</td>
<td>AtLPP1 (PAP1)</td>
<td>327</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>leaves [51]</td>
<td>0.060</td>
</tr>
<tr>
<td>At1g15080</td>
<td>AtLPP2 (PAP2)</td>
<td>290</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>SAM (embryo) [51]</td>
<td>-0.091</td>
</tr>
<tr>
<td>At3g50920</td>
<td>AtLPPε1</td>
<td>279</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>not seed [51]</td>
<td>-0.069</td>
</tr>
<tr>
<td>At5g66450</td>
<td>AtLPPε2</td>
<td>286</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>At5g03080</td>
<td>AtLPPγ</td>
<td>226</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>ubiq, dry seed [51]</td>
<td>-0.017</td>
</tr>
<tr>
<td>At3g03520</td>
<td>AtLPAP1</td>
<td>523</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>Seed [51]</td>
<td>0.071</td>
</tr>
<tr>
<td>At3g58490</td>
<td>AtSPP1 (LPPδ)</td>
<td>416</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>dry seed [51]</td>
<td>0.005</td>
</tr>
<tr>
<td>At3g09560</td>
<td>AtPAH1</td>
<td>904</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>dry seed [51]</td>
<td>0.005</td>
</tr>
<tr>
<td>At5g42870</td>
<td>AtPAH2</td>
<td>930</td>
<td>PAP Activity?</td>
<td>chloroplast [41]</td>
<td></td>
<td>chloroplast PAP [41]</td>
<td>dry seed [51]</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 1: Literature review and protein information of putative PAPs. LPPβ, the main putative PAP of interest, is highlighted in yellow, while the four (putative) PAPs that could act redundantly with LPPβ are highlighted in grey. Pink background indicates negative attributes; tan background indicates neutral attributes, and green background indicates positive attributes. * Column 9: Coexpression with DGAT and PDAT: value given is an average of the two coexpression values, as calculated by GeneCat [52].
Table 2: Homologs of Arabidopsis putative PAP proteins in Castor Bean. \(LPP\), indicated in yellow, has only one homolog in Castor, while most other Arabidopsis putative PAPs have more than one. X: Primary homolog (reciprocal BLASTp results indicated Arabidopsis homolog as first candidate); *: Secondary homolog (reciprocal BLASTp results indicated Arabidopsis homolog as a candidate, but not the top choice).

<table>
<thead>
<tr>
<th>Phytozome Accession Number</th>
<th>Name Description</th>
<th>Castor Bean (Ricinus communis) PAP Homologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>28455.m000359</td>
<td>Sphingosine-1-phosphate phosphohydrolase, putative</td>
<td>AlLPP</td>
</tr>
<tr>
<td>29586.m000620</td>
<td>Phosphatidic acid phosphatase, putative</td>
<td>X</td>
</tr>
<tr>
<td>29660.m000759</td>
<td>Phosphatidic acid phosphatase, putative</td>
<td>X</td>
</tr>
<tr>
<td>29747.m001075</td>
<td>Phosphatidic acid phosphatase, putative</td>
<td>*</td>
</tr>
<tr>
<td>29660.m000760</td>
<td>ER Phosphatidate Phosphatase</td>
<td>*</td>
</tr>
<tr>
<td>29841.m002865</td>
<td>Phosphatidate-kholine diacylglycerol cholinephosphotransferase (PDCT)</td>
<td>*</td>
</tr>
<tr>
<td>29805.m001476</td>
<td>Dolichylphosphatase, putative</td>
<td>X</td>
</tr>
<tr>
<td>29634.m002101</td>
<td>Conserved hypothetical protein</td>
<td>*</td>
</tr>
<tr>
<td>29634.m002100</td>
<td>Palmitoyl-protein thioesterase/dolichylphosphatase 1</td>
<td>*</td>
</tr>
<tr>
<td>27574.m000218</td>
<td>Sphingosine-1-phosphate phosphohydrolase, putative</td>
<td>X</td>
</tr>
<tr>
<td>30148.m001447</td>
<td>Phospholipase C3 precursor, putative</td>
<td>X</td>
</tr>
<tr>
<td>30147.m014488</td>
<td>Phospholipase C4 precursor, putative</td>
<td>X</td>
</tr>
<tr>
<td>29681.m001334</td>
<td>Hydrolase, acting on ester bonds, putative</td>
<td>X</td>
</tr>
<tr>
<td>30142.m000628</td>
<td>Hydrolase, acting on ester bonds, putative</td>
<td>X</td>
</tr>
<tr>
<td>30170.m013896</td>
<td>PAH1 conserved hypothetical protein</td>
<td>*</td>
</tr>
<tr>
<td>30170.m013897</td>
<td>PAH2 conserved hypothetical protein</td>
<td>*</td>
</tr>
</tbody>
</table>
### Soybean (Glycine max) PAP Homologs

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>JGI Soybean Gene Call</th>
<th>Name</th>
<th>LPP</th>
<th>β</th>
<th>ε1</th>
<th>ε2</th>
<th>α</th>
<th>γ</th>
<th>ββ</th>
<th>βδ</th>
<th>γβ</th>
<th>γδ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyma.05G056700</td>
<td>Sphingosine-1-phosphate phosphohydrolase</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.07G029800</td>
<td>phosphatidic acid phosphatase-related / PAP2-related</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.08G213100</td>
<td>Sphingosine-1-phosphate phosphohydrolase</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.02G089100</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>X</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.18G293800</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>X</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.10G270100</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>X</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.20G121200</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>X</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.09G115800</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.10G270000</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.11G165300</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.18G058400</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.20G213500</td>
<td>LIPID PHOSPHATE PHOSPHATASE</td>
<td>*</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.10G050300</td>
<td>Palmitoyl-Protein Thioesterase/Dolichylphosphatase 1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.13G137800</td>
<td>Palmitoyl-Protein Thioesterase/Dolichylphosphatase 1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.01G171800</td>
<td>Palmitoyl-Protein Thioesterase/Dolichylphosphatase 1</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.11G071400</td>
<td>Palmitoyl-Protein Thioesterase/Dolichylphosphatase 1</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.03G070700</td>
<td>Sphingosine-1-phosphate phosphohydrolase</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.19G204500</td>
<td>Sphingosine-1-phosphate phosphohydrolase</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.03G092800</td>
<td>Phosphoesterase family</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.16G081200</td>
<td>Phosphoesterase family</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.18G064000</td>
<td>Phosphoesterase family</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.20G040600</td>
<td>Phosphoesterase family</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.03G174900</td>
<td>LIPIN</td>
<td>X</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.06G039200</td>
<td>LIPIN</td>
<td>X</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.10G046400</td>
<td>LIPIN</td>
<td>X</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.19G175600</td>
<td>LIPIN</td>
<td>X</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.04G038200</td>
<td>LIPIN</td>
<td>X</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.13G134500</td>
<td>LIPIN</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.17G242900</td>
<td>LIPIN</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glyma.20G085600</td>
<td>LIPIN</td>
<td>*</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Homologs of Arabidopsis putative PAP proteins in Soybean.** *LPP/β,* indicated in yellow, has only one homolog in soybean while all the other putative PAPs have at least two homologs. *X:* Primary homolog (reciprocal BLASTp results indicated Arabidopsis homolog as first candidate); *:* Secondary homolog (reciprocal BLASTp results indicated Arabidopsis homolog as a candidate, but not the top choice).
Table 4. Homologs of Arabidopsis putative PAPs in seed tissue of oilseeds. \textit{LPPβ} (highlighted in yellow) is expressed in all four species who were analyzed using 454 pyrosequencing. It is not on the microarray chip used to obtain the Arabidopsis data. Adapted from Troncoso-Ponce and Kilaru et al. [53].
Figure 3. Bioinformatics Information on LPPβ. (A) Tissue expression of LPPβ based on data from the AGRONOMICS whole genome tiling array [65] and visualized using Genevestigator (https://genevestigator.com/gv/index.jsp). LPPβ shows high expression in all tissues tested, but seed or silique tissue was not examined. (B) Transmembrane domain prediction from TMHMM [54].
Figure 4. Tissue expression of genes of interest. (A) The first set of tissue expression experiments: Actin8 control; RIL is not expressed in the leaf or silique; and SPP1 is expressed in both the leaf and silique. (B) Second set of tissue expression experiments: Actin8 control; and LPPβ is expressed in both the leaf and silique. LPPβ has no introns, but the Actin8 control shows that there is no genomic contamination in the leaf or silique samples. G, Genomic DNA (PCR positive control); L, Leaf cDNA; S, Silique cDNA; --, water (PCR negative control).
Figure 5. Subcellular Localization of LPPβ, SPP1, and LPAP1. Fusion proteins are shown in green, ER-localized BiP in red, and chloroplast autofluorescence in blue. The identity of the GFP-Fusion* is given in the row headings. GUS cytosolic localization is indistinguishable from BiP ER localization, but clearly different than chloroplast localization. LPPβ, SPP1, and LPAP1 all appear to localize to either the ER or cytosol.
Figure 6. Non-Diurnal Expression Pattern of LPPβ. Grey boxes indicate dark periods. Error bars represent Standard Error of three biological replicates. qPCR data was normalized to Tubulin2. (A) Short day expression level of LPPβ. (B) Long day expression of LPPβ. (C) Expression of LPPβ under continuous light.
<table>
<thead>
<tr>
<th>Name</th>
<th>Primer Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBF1</td>
<td>ATGTCTCCGTCGACGACG</td>
<td>This Work</td>
</tr>
<tr>
<td>LBF1-G</td>
<td>CACCATGTTCTCCGTCGACGACG</td>
<td>This Work</td>
</tr>
<tr>
<td>LBR1</td>
<td>ACCTCCAAAGATCATTTCA</td>
<td>This Work</td>
</tr>
<tr>
<td>LB qPCR-F3</td>
<td>TGCTGAAGCTTCCATGACTG</td>
<td>This Work</td>
</tr>
<tr>
<td>LB qPCR-R2</td>
<td>CTCAGAAAAATCGAAGACGGCAAAGACG</td>
<td>This Work</td>
</tr>
<tr>
<td>LDF1-G</td>
<td>CACCATGAAGAGATTTCTAGAAAAGATCAAATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LDR3</td>
<td>GGCACTGATTTATATATACACAAATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LDF4</td>
<td>AATCTACCAAACGTCTCGAAGTCTCT</td>
<td>This Work</td>
</tr>
<tr>
<td>LDR3</td>
<td>GGCACTGATTTATATATACACAAATG</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1F1G</td>
<td>CACCATGTTGGAGAAAACGAGCT</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1R3</td>
<td>TCAATTATACACAAATCAACAGA</td>
<td>This Work</td>
</tr>
<tr>
<td>R1L F2</td>
<td>AGCATCTTGGCATCTTCACC</td>
<td>This Work</td>
</tr>
<tr>
<td>R1L R2</td>
<td>TCATTTACAACTTCAAATGTTGTTTC</td>
<td>This Work</td>
</tr>
<tr>
<td>ACT8F</td>
<td>ATGGCCGATGCTGATGAC</td>
<td>Wayne 2012 [70]</td>
</tr>
<tr>
<td>ACT8R</td>
<td>TTGAAGCATTGCTGTGGACATG</td>
<td>Wayne 2012 [70]</td>
</tr>
<tr>
<td>TUB2 qPCR-F</td>
<td>ACTGTTCTCAAGGTTCCAGGTTC</td>
<td>Niu 2009 [71]</td>
</tr>
<tr>
<td>TUB2 qPCR-R</td>
<td>ACCGAGAAAGGTAAGCATATCGA</td>
<td>Niu 2009 [71]</td>
</tr>
</tbody>
</table>

Table 5. Primers used in this work. Gateway tags are indicated in **bold**.
2.5 REFERENCES


11. Toke DA, McClintick ML, Carman GM: Mutagenesis of the phosphatase sequence motif in diacylglycerol pyrophosphate phosphatase from *Saccharomyces cerevisiae*. *Biochemistry* 1999, **38**:14606-14613.


CHAPTER THREE
GENETIC, MOLECULAR AND BIOCHEMICAL CHARACTERIZATION
OF PUTATIVE PAP GENES

ABSTRACT

Despite recent advancements in the field, the identity of the phosphatidic acid phosphatase (PAP) involved in oil synthesis in Arabidopsis still remains elusive. The PAP enzymatic function is known to be required for membrane lipid and triacylglycerol (TAG) synthesis, but the gene encoding that function is still unknown. Here, I use genetic and biochemical approaches to determine whether the Arabidopsis gene \( \text{LPP}\beta \) (At4g22550) could be the main PAP involved in TAG synthesis. Attempts to knock down expression of \( \text{LPP}\beta \) using amiRNA and ihpRNA were largely unsuccessful, and attempts at a guided mutation using the CRISPR/Cas9 system have not yet succeeded. \( \text{LPP}\beta \) could also be redundant with other enzymes exhibiting PAP activity in the ER, so multiple mutants of \( \text{spp1, lpap1, pah1,} \) and \( \text{pah2} \) were pursued and tested for reductions in seed oil. None of the mutants exhibited phenotypes beyond what has been documented for \( \text{pah1pah2} \), indicating that \( \text{SPP1} \) and \( \text{LPAP1} \) are not redundant with \( \text{PAH1} \) or \( \text{PAH2} \) and are not involved in oil synthesis. In complementation assays with the yeast \( \Delta \text{pah1} \) mutant and in PAP radioactivity assays, \( \text{LPP}\beta \) showed modest activity for dipalmitoyl phosphatidic acid. Like other genes involved in oil synthesis, \( \text{LPP}\beta \) is not transcriptionally activated by \( \text{WRI1} \), despite the presence of an AW box in its promoter. At the conclusion of this study, \( \text{LPP}\beta \) demonstrates characteristics consistent with a PAP involved in oil synthesis, but still requires more study before placing it in the oil synthesis pathway.
3.1. INTRODUCTION

3.1.1. PAP ACTIVITY IN OIL SYNTHESIS

Phosphatidic acid phosphatase (PAP) enzymes catalyze the dephosphorylation of phosphatidic acid (PA) to diacylglycerol (DAG) in a hydrolysis reaction that releases inorganic phosphate. The substrate of the PAP reaction, PA, is used for the synthesis of the phospholipid phosphatidylserine, and is the precursor for CDP-DAG, which is used to make other phospholipids. The product of the PAP reaction, DAG, is utilized to make PC, which is the main ER membrane lipid, the precursor for the acyl-editing cycle (the Lands cycle), and the precursor for a separate pool of DAG which can be exported to the plastid or used to make TAG. PA and DAG are also signaling molecules in addition to their roles as intermediates in lipid synthesis. Therefore, PAP activity is an essential step in the pathway to TAG, and it may also regulate the synthesis of phospholipids, TAG, and signaling molecules.

In yeast and mammals, the PAP enzyme involved in oil synthesis has been shown to not only affect oil accumulation, but also affect processes such as membrane homeostasis [1,2], lipid droplet formation [3], development of adipose tissue [4,5], and regulation of lipid metabolism genes [6]. (See Chapter 2, Sections 2.1.1 and 2.1.2, for more details on PAP activity in yeast and mammals.) In vegetative tissues of Arabidopsis, PAH1 and PAH2 function redundantly in ER membrane homeostasis and lipid remodeling during phosphate stress [7,8]. PAH1 and PAH2 also redundantly affect oil synthesis, but a double knockout still only results in a 15% decrease in oil [7]. This indicates that another PAP(s) is involved in producing the remaining oil, and the pah1pah2 mutant provides a convenient background to determine other PAP genes that act redundantly with PAH1 and PAH2 to affect oil in the seed.
In Chapter 2, I screened the Arabidopsis genome for possible PAPs with homology to known yeast, mammalian, and plant PAPs. In addition to PAH1 and PAH2, there are 11 genes in the Arabidopsis genome which are homologous to known PAPs. Of these genes, LPPβ alone displayed characteristics which were consistent with the main PAP involved in oil synthesis. It is expressed in seed tissue, and the protein product is an integral membrane protein that localizes to the endoplasmic reticulum. Additionally, LPAP1 and SPP1 have been implicated in other functions, but they also displayed characteristics indicating that they could act redundantly with the main PAP involved in oil synthesis. LPAP1 does show some PAP activity in vitro, but its main activity is a lysophosphatidic acid (LPA) phosphatase [9]. However, its possible roles either in lipid synthesis, providing the intermediate monoacyl glycerol for membrane or storage lipid synthesis, or in attenuating the signaling functions of LPA, have not been characterized. Finally, SPP1 displays activity as a long chain base phosphate phosphatase, and has a role in ABA-mediated dehydration stress response [10]. However, it was not assayed for PAP activity, and its possible involvement in oil synthesis remained unexplored. LPAP1 and SPP1 will be tested to determine if they act redundantly with LPPβ, or with PAH1 and PAH2 in oil synthesis. (See Chapter 2 for extensive background on PAPs, and the methodology used to determine PAPs that would be studied in this work.)

PAP activity is generally thought to be essential for TAG synthesis, and since TAG is an essential energy source for developing embryos and viable pollen [11], a complete knockout of all PAPs involved in TAG synthesis may be lethal. In the case of a non-lethal mutant, it is expected that seed TAG will decrease, as is seen in other mutants in TAG biosynthesis, such as the dgat1 mutants AS11 and ABX45 [12,13]. The seeds of the AS11 and ABX45 lines show a 25% and 45% decrease in total oil, respectively, but are still able to germinate and produce viable plants.
The activity could also be redundantly encoded in the genome, making multiple mutants necessary to determine which PAP(s) are involved in oil synthesis.

### 3.1.2. ΔPAH1 Yeast Mutants

Yeast (*Saccharomyces cerevisiae*) has four PAPs: *APP1, LPP1, DPP1*, and *PAH1*. *APP1, LPP1, and DPP1* are not specific for a particular lipid substrate, and are involved in regulation and endocytosis. *PAH1* is the only PAP involved in *de novo* lipid synthesis and oil synthesis. While the mutants Δ*app1, Δlpp1*, and Δ*dpp1* do not show significant changes in TAG, Δ*pah1* mutants show reduced DAG and TAG, as well as a temperature-dependent growth phenotype: they are unable to grow at 37°C unless complemented with a functional PAP [14]. This phenotype has provided a convenient assay for determining if genes display PAP activity [15].

In this work, I describe my attempts to knock down expression of *LPPβ* in order to determine if it is involved in oil synthesis either alone or in conjunction with several other genes, and to determine if it has PAP activity. *LPAP1* and *SPP1* show conclusively that they do not function redundantly with *PAH1* or *PAH2* in oil synthesis, and are probably involved in signaling pathways as previously suggested. *LPPβ* displays PAP activity when expressed in yeast, and its subcellular localization and tissue expression are consistent with a PAP involved in oil synthesis. Continuing work is needed, however, to fully determine its biological role.
3.2. MATERIALS AND METHODS

3.2.1. PLANT MATERIALS AND GROWTH CONDITIONS

For oil quantification sets, seeds were germinated on ½ MS medium with 1% (w/v) sucrose and 2% agar in continuous light (100-150 umol m$^{-2}$ s$^{-1}$ light) at 22°C. Fourteen day old seedlings were transferred to soil and grown in a growth chamber under the same conditions, or transferred to a greenhouse with 16 h day 8 h night, 21-24°C day and 17-20°C nights and light intensity varying from 100-200 umol m$^{-2}$ s$^{-1}$ light. For each test, a minimum of 21 plants were planted per line and lines were randomized across flats.

3.2.2 IDENTIFICATION OF T-DNA INSERTIONAL MUTANTS

One homozygous T-DNA insertional mutant for $LPP\beta$ (At4g22550) was available and was ordered from the Ohio State University (SALK_058344C; $lpp\beta$-1). Insertional mutants for $SPP1$ (At3g58490) and $LPAP1$ (At3g03520) were also ordered from The Ohio State University (SALK_027084 for $spp1$-2 and SALK_065482 for $lpap1$-1) and the $pah1pah2$ mutant was requested from the Ohta lab to investigate any possible redundancy between $LPP\beta$ and those genes. Homozygous mutants of $lpp\beta$-1, $spp1$-2, and $lpap1$-1 were analyzed for full-length transcript. RNA was extracted from leaf tissue using the Trizol procedure, DNA removed with the DNA-free Kit, and cDNA synthesized using the SuperScript III kit. Full length transcript was amplified using the LBF4/LBR1 for $LPP\beta$, LDF4/LDR3 for $SPP1$, or UP1F3/UP1R6 for $LPAP1$. Genomic DNA was sequenced to confirm exact insert location.

3.2.3. OVERCOMING FUNCTIONAL REDUNDANCY: GENERATING MULTIPLE MUTANTS

Confirmed homozygous null lines were crossed in the following combinations: $pah1pah2 \times spp1$, $pah1pah2 \times lpap1$, $spp1 \times lpap1$. This generated $pah1pah2spp1$, $pah1pah2lpap1$, and $spp1lpap1$. The triple mutants were then crossed, $pah1pah2spp1 \times pah1pah2lpap1$, to create
\textit{pah1pah2spp1lpap1}. \textit{PAH1} and \textit{LPAP1} are only 6.8 centimorgans apart on Chromosome 3, so F1 plants were screened for the \textit{pah1pah2} small phenotype first \cite{8}, and then genotyped for the \textit{lpap1} and \textit{spp1} mutations.

### 3.2.4. Mutant Line Analysis: Generating a \textit{LPP\beta} Mutant

Due to the unavailability of null insertional mutants, RNA silencing and CRISPR/Cas9 directed mutagenesis were pursued. Artificial micro RNA (amiRNA) constructs were designed using Detlef Weigel’s website (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi, \cite{16}) to specifically target the \textit{LPP\beta} gene. The amiRNA has a sequence of “TAATCCCTAAAAAGCACCCG” and targets the region targeted shown schematically in Figure 1. PCR using KOD HotStart Taq (Novagen), the template plasmid pRS300, and the primers Listed in Table 1 (amiRNA-A-G, I miR-s, II miR-a, III miR*-s, IV miR*a, and amiRNA-B) was used to generate the amiRNA as described on the website. It was sequenced, then cloned into the pOEA vector \cite{17} at the SwaI site using T4 DNA ligase to produce pKC06. This construct includes a Phaseolin promoter driving strong seed-specific expression of the amiRNA and DsRed as a visual marker \cite{18}.

Intron-containing hairpin RNA (ihpRNA) constructs were created by amplifying a 200 bp segment of LB using primers hp1F-G and hp1R (see Table X) and cloning into pENTR. The sequenced gene was then subcloned into the pB7GWIWG2(II) \cite{19} plasmid using the Gateway LR Reaction (Invitrogen), which inserted the gene in both directions on either side of an intron. Unfortunately, the pB7GWIWG2(II) plasmid uses a 35S CaMV promoter, which has little activity in seeds; therefore the hairpin-intron-hairpin construct was cut out of pB7GWIWG2(II) using SwaI (New England Biolabs) and ligated into the pOEA-DsRed vector at the SwaI site using T4 DNA Ligase (New England Biolabs) to produce pKC05.
Both pKC05 and pKC06 were checked for correct orientation, then transformed into Agrobacterium (GV3101) and the floral dip method (Clough and Bent, 1998) was used to transform WT Columbia-0, *spp1-2*, *lpap1-1*, and *pah1pah2* plants.

When pKC05 and pKC06 failed to generate LPPβ knockdowns, the CRISPR/Cas9 guided mutation system was employed. Three guide RNAs targeting the 5' end of the gene were designed and constructed by overlapping PCR according to Li et al. [20], producing promoter-gRNA-terminator cassettes. Restriction sites were added by PCR using the U6-XhoI-F1 and 3'-PacI-R2 primers. These were digested by PacI and XhoI and cloned into pFGC-pcoCas9 (a kind gift from Jen Sheen, AddGene #52256) to create pKC35 for the A cut site, pKC36 for B, and pKC37 for C. The B cassette was also added to pFGC-DsRed, a version of pFGC created by Jeremy Jewell in our lab where the Cas9 enzyme appears to be nonfunctional (personal communication), but allows for red/brown screening of the seeds using the DsRed fluorophore; this generated pKC28. Combinations of A/B and B/C were transformed into agrobacterium and into all eight backgrounds as described above: WT, *spp1*, *lpap1*, *spp1 lpap1*, *pah1pah2*, *pah1pah2spp1*, *pah1pah2lpap1*, and *pah1pah2spp1lpap1*.

Additionally, pK7WGF2::hCas9 (a kind gift from Sophien Kamoun, Addgene plasmid #46965 [21]) was digested with AatII and the C cut site gRNA cassette mentioned above was ligated into the plasmid. The resulting pKC41 plasmid was transformed into WT Arabidopsis using the floral dip method. Finally, a combination of the original pK7WGF2 plasmid and pKC35 were transformed into the WT background using floral dip. Kanamycin resistance (and/or silencing) already present in the SALK lines prevented these two methods from being used on any of the mutant lines.
CRISPR/Cas9 T1 plants were identified by red fluorescence (pKC28), Basta resistance (pKC35-pKC37), Kanamycin resistance (pKC41), or a combination of Basta and Kanamycin (pK7WGF2 and pKC35). The LB gene was amplified from T2 plants using LBF1/LBR1 primers, and first analyzed for a 200 bp deletion in the gene for those transformed with A/B or B/C “double cut” gRNA combinations. T2 mutants were identified by the Guide-It Mutation Detection Kit (Clontech Laboratories, Cat #631443; using LBF1/LBR1 primers), and interesting amplicons were cloned into pCR using the ZeroBlunt kit (Invitrogen), then sequenced to determine the exact mutation(s) in each line.

3.2.5. Seed Tissue Expression of LPP\(\beta\) in Wild-Type and Multiple Mutants

Developing seed tissue was isolated according to Bates et al (2013) and stored at -80°C. Samples were ground under liquid nitrogen using a handheld pestle spinner in a microcentrifuge tube, and RNA was extracted using a modified version of the Onate-Sanchez protocol [22]. After extracting with phenol, the aqueous layer was further extracted with another 500 uL of chloroform and separated using a pre-spun Phase-Lock Gel column (5Prime). Complimentary DNA was prepared using Quanta’s qScript DNA SuperMix cDNA synthesis kit. Quantitative RT-PCR was performed on the MX3005P qPCR system (Stratagene) using the Quanta SYBR FastMix kit. Amplicons of roughly 200 bp were amplified, using LB qPCR F3 and LB qPCR R2. Data from LPP\(\beta\) specific primers was normalized to the AtTUB2 gene, followed by calculation of the relative fold change (=1/(2^exp-control)).

3.2.6. Total Fatty Acid Content of Seeds

To quantify fatty acid content of seeds in the multiple knockout lines and in the LPP\(\beta\) RNAi lines, all plants used in each test were grown together, and 20-30 individual plants from each line were randomized across all flats. Seeds were harvested at maturity, and dried under silica
gel for a minimum of 48 hours. Twenty to 40 mg of seed was weighed out and counted, then fatty acids were transmethylated according to Li et al. [23], extracted, and run on a polar GC column. Chromatographs were analyzed using ChemStation software (Rev B.03.02[341])

3.2.7. COMPLEMENTATION OF THE YEAST ΔPAH1 MUTANT

The coding sequences of LPPβ, SPP1, LPAP1, and AtPAH1 were amplified from silique cDNA using KOD Hotstart Polymerase (Toyobo) and the LBF1G/LBR1, LDF1G/LDR6, UP1F1G/UP1R3 and PAH1FW/PAH1 RV primer pairs, respectively. ScPAH1 was amplified from yeast genomic DNA using ScPAHF1G/ScPAHR1 primers. Blunt-ended products were cloned into pENTR using the Gateway pENTR/D-TOPO Cloning Kit and were sequenced. All five genes were transferred into the destination vector using Gateway LR Clonase II (Invitrogen) and checked for orientation. The first of two destination vectors, pAG425GPD-ccdB (a kind gift from Susan Lindquist, Addgene plasmid # 14154), contains the strong, constitutive GPD promoter for expression of the gene of interest, and provides the LEU2 gene as a selectable marker.

pAG425GPD-ccdB was also modified to include the ScPAH1 promoter instead of the strong, constitutive GPD promoter. It was digested with SacI and SpeI, and the 8.8 kb fragment was purified; the ScPAH1 promoter was amplified from yeast genomic DNA using the PPF1 and PPR1 primers. These were then assembled using the Gibson Assembly kit (Invitrogen) and the PPCF1, PPCR1, PPCF1*, and PPCR1* primers, to create the pKC07 plasmid. The six genes above were also transformed into pKC07 using the LR Clonase II and checked for orientation.

Both the pAG425GPD and the pKC07 versions of the six genes were transformed into the Δpah1 and Δpah1Δlpp1Δdpp1 yeast mutants [14] using the One-Step transformation method described by Chen et al. [24]. All of the experiments described, except where noted, use the Δpah1 mutant.
Complementation assays were conducted using freshly transformed cells. Cultures (2 mL) of each line were grown for 2 days, then a hemocytometer was used to count cells and determine exact density of the solution. Ten, $10^0$, $10^2$, $10^3$, $10^4$, and $10^5$ cells were spotted onto media in duplicate. Standard Leucine dropout media was used for the plates that would be placed in the 30°C incubator, while YPD plates were used for those going in the 37°C incubator. Plates were grown 2 to 7 days more, and pictures from plates with visible colonies were analyzed for growth phenotype.

3.2.8. Enzyme Assays: Yeast Microsomal PAP Activity Assays

The $\Delta pah1$ yeast mutant expressing genes under control of the PAH1 promoter from the previous section were grown together 2 days at 30°C in 100 mL of standard dropout media without Leucine. Cells were collected by centrifugation at 3500 rcf at 4°C, then washed in water and collected. Pellets were resuspended in 0.5 mL cold lysis buffer (20 mM Tris HCl, 1 mM EDTA, 5% glycerol, 10 mM MgCl2, 1 mM DTT, 300 mM NH4SO4) and cells were disrupted by vortexing with glass beads (10 x 30s). Supernatant was centrifuged at 10,000 g for 20 min at 4°C to remove intact organelles, and the resulting supernatant was centrifuged at 100,000 g for 60 min to isolate the microsomal fraction. The microsomal pellet was resuspended in 200 uL assay buffer (50 mM Tris HCl, 50 mM NaCl, 5 % glycerol, 2 mM MgCl2), and protein quantified using the Bradford Assay.

To assay for PAP activity, 20 ug total protein was added to 100 uL substrate solution (20 mM Tris-HCl, pH 7.0, 0.1% (w/v) Triton X-100, and 45 nCi L-3-PA 1,2 dipalmitoyl) and enough assay buffer was added to make 200 uL total. The reaction proceeded for 60 min, then was quenched by vigorous vortexing with 2:1 chloroform:methanol. Products were separated using 0.88% NaCl, and the organic phase removed, then back-extracted with 2:1 chloroform:methanol.
In most cases, the combined fractions were cloudy, so all samples were extracted once more with NaCl solution. The organic phase was dried down under a stream of nitrogen at RT, and resuspended in 2:1 chloroform:methanol.

Each sample was spotted on a silica gel TLC plate was developed halfway in 35:25:4:14:2 of chloroform:methanol:acetic acid:acetone:water. After drying, the plate was developed fully in 50:50:1 petroleum ether:ethyl ether:acetic acid. For samples quantified by GC, the plate was stained with 0.05% primulin in 8:2 acetone:water and marked with a soft pencil; for radioactivity the spots were quantified using the Typhoon FLA7000 laser scanner system and ImageQuantTL software (GE Healthcare); for unquantified samples the plate was stained with iodine vapor.

**3.2.9. Enzyme Assays: E. coli Expression and PAP Activity Assays**

The coding sequences of \( LPP\beta \), \( SPP1 \), and \( LPAP1 \) were amplified from their pENTR plasmids (see 2.7 above) using Ligation Independent Cloning (LIC) compatible primers: LBF1-Lic/LBR1-NsLic, LDF1-Lic/LdR6-NSLic, and UP1F1-Lic/UP1R3-NsLic, respectively. The E. coli expression vector pET-GFP-LIC (AddGene #29772) was a gift from Scott Gradia. The vector was linearized by digestion with EcoRV. Vector and inserts were treated with T4 DNA polymerase in the presence of dCTP in the case of pET-GFP-LIC or in the presence of dGTP in the case of the inserts, \( LPP\beta \), \( SPP1 \), and \( LPAP1 \). Inserts were annealed into LIC vectors in the presence of EDTA and transformed into E. coli. This cloning resulted in a C-terminal protein fusion of the gene of interest (\( LPP\beta \), \( SPP1 \), or \( LPAP1 \)) with eGFP [25], driven by the inducible T7 bacterial promoter. Sequenced plasmids were then transformed into Rosetta 2(DE3)pLysS cells for protein expression.

For protein expression, saturated cultures were used to inoculate 100-mL LB cultures. Cultures were allowed to grow at 37°C to an \( \text{OD}_{600} \) of 0.4, and then were induced with 25 uL of 0.4 M IPTG and grown for three hours at 37°C. Cells were harvested by centrifugation at 4300 x
g for 20 min at 4°C, and cell pellets were lysed with the addition of 1mg/mL lysozyme in lysis buffer (25 mM TrisHCl pH 7.0, 150 mM NaCl, 0.1% (w/v) Triton X-100, 1 unit/20 uL DNaseI). Fractions of the crude lysate were run on 10% BisTris SDS-PAGE gels for Western Blotting. Proteins were transferred to a nitrocellulose membrane and were probed with 1:5000 rabbit anti-GFP primary antibody (ab290) and 1:20,000 goat anti-rabbit secondary antibody conjugated to Horse Radish Peroxidase (ab97080). Blots were visualized using ECL Western Blotting Substrate (Pierce).

Protein was quantified using the Bradford Assay, and 20 ug total protein was used for PAP activity assay. The assay and TLC plate were run as noted above for the yeast microsomal assay (see 2.8 above), except for the composition of the assay buffer (20 mM Tris-HCl, pH 7.0, 150 mM NaCl, 15% (v/v) glycerol, 0.1% (w/v) Triton X-100).

### 3.2.10 WRII CONTROL OF LPPβ

Segregating T2 seeds expressing WRII under control of the strong seed-specific Phaseolin promoter were obtained from Neil Adhikari (unpublished data). Red (containing transgene) and brown (isogenic segregants) seeds were planted on soil and stratified for 4 days at 4°C. Developing seeds from five segregating mutant plants and four isogenic segregants were harvested according to Bates et al [26] at 7-9 days after flowering. RNA was extracted using a modified protocol from Onate-Sanchez. After extracting with phenol, the aqueous layer was further extracted with another 500 uL of chloroform and separated with the help of a pre-spun Phase-lock Gel column (5Prime). Complimentary DNA was prepared using Quanta’s qScript DNA SuperMix cDNA synthesis kit, and LPPβ expression was quantified as described above. WRII expression was also quantified in relation to TUB2 using qWRIII(F(3) and qWRIII(R(3) primers from Neil Adhikari (personal communication).
3.3. RESULTS

3.3.1. MUTANT LINE ANALYSIS: IDENTIFYING SINGLE KO’S

Putative T-DNA insertional mutant lines were obtained from ABRC: lppβ-1 for LPPβ (At4g22550), spp1-2 for SPP1 (At3g58490), and lpap1-1 for LPAP1 (At3g03520). The insertion location of each line was sequenced, and their relative locations can be seen in Figure 1. The insertions in both spp1-2 and lpap1-1 were in exons, but the insertion in lppβ-1 was upstream of the 5’UTR, indicating it may not be a true knockout. RT-PCR was pursued to determine if the T-DNA insertion lines encoded null mutations. Amplification of the Actin8 gene from two plants from each background showed similar amounts of cDNA was loaded on each sample, and RT control shows no genomic contamination. The lppβ1-1 line, a “confirmed” homozygous line, showed amplification at similar levels as WT in both of the plants tested, indicating that the insertion in the promoter does not affect expression levels. Both the spp1-2 and lpap1-1 mutants showed expression in the WT line but no amplification in the mutant lines, indicating that they are most likely null mutations.

3.3.2. MUTANT LINE ANALYSIS: CHARACTERIZING LPPβ RNAI MUTANTS

Since I could not isolate a null mutant of LPPβ from the available T-DNA lines, I attempted to reduce LPPβ expression using both an amiRNA and an ihpRNA approach. Second generation (T2) segregating seeds were separated into red (expressing the amiRNA or ihpRNA transgene) and brown (WT isogenic segregants), and total seed fatty acid levels were quantified in the WT, spp1-2, lpap1-1, and pah1pah2 backgrounds (Figure 2). While it is well established that Arabidopsis seed oil varies widely between plants [23], using LPPβ-knockdown and segregating WT seeds off the same plant compensates for that variability. Out of the roughly 80 independent
lines (~20 per background) analyzed, only seven showed a statistically significant change in oil between the knockdown and WT segregants.

Recognizing the relatively low number of RNAi lines with reduced oil, I set out to confirm that the RNAi lines actually showed reduced expression of LPPβ. Unfortunately, qPCR showed that none of the lines tested showed a reduction in expression beyond 50% of WT; instead, many of the lines had increased LPPβ expression (Figure 3). Basal levels of LPPβ expression were increased in the lpap1 and spp1 mutant, but the RNAi constructs did not decrease these levels much below WT when introduced to the mutants. Thus, the RNAi constructs did not appear to work, and are not a viable option to study the effects of LPPβ on oil synthesis.

3.3.3. Mutant Line Analysis: Investigating Redundancy

To investigate redundancy between the identified genes and prepare for eventual introduction of an lppβ mutant, crosses were made between lpap1, spp1-2, and pah1pah2. All of the genes were unlinked except LPAP1 and PAH1, which are located 6.98 cM apart (based on 1 cM equivalent to 300,000 nucleotides) on chromosome three. In the lpap1 x pah1pah2 cross, the F2 generation plants were screened first for the pah1pah2 visual phenotype and then genotyped for lpap1, and mutants were isolated which were homozygous mutant at all three loci. Since the other crosses did not include linked genes, spp1 pah1 pah2, spp1 lpap1, and spp1 lpap1 pah1 pah2 were all isolated in the F2 generation.

As I noted in the initial studies of the LPPβ RNAi lines, some of the mutants appeared to have increased expression of LPPβ. I investigated this further, by analyzing LPPβ expression using qPCR in all of the multiple mutant lines that I had generated. Indeed, all seven lines showed increased LPPβ expression relative to WT (Figure 4). This indicates that LPPβ likely exhibits
PAP activity and could be compensating for the mutations, especially in the pah1pah2 line where a known PAP activity has been knocked out.

I had hoped to introduce an LPPβ knockdown into all of the combinations of spp1, lpap1, and pah1pah2, so I prepared for these experiments by characterizing the backgrounds for changes in seed oil and fatty acid composition. Trials were grown in both the greenhouse and growth chambers, but the data from the two growth conditions was largely similar. The visible growth phenotype of all plants was very similar to WT, or to pah1pah2 in that background (the pah1pah2 phenotype includes shorter plants; smaller, greener, rounder leaves; and shorter siliques [7,8]). Seed oil was investigated as well, and Figure 5 shows a representative sample of seed oil measurements, from growth chamber-grown plants. In all cases, the mutants in the WT background were similar to WT (Figure 5A), and those in the pah1pah2 background showed the characteristic pah1pah2 phenotype (a ~15% reduction in seed oil; Figure 5B). There was also no statistically significant change in fatty acid composition of the seeds in the WT background (data not shown), and most of the changes in fatty acid composition in the pah1pah2 background were consistent with the published pah1pah2 phenotype, including a decrease in 18:2 and 20:1 (Figure 5C) [7]. The increase in 20:0 and 22:1 in the pah1pah2 background has not been examined before, but all four mutants in the pah1pah2 background showed a very similar phenotype. The only irregularity with published data was seen in 18:3, which Eastmond et al. reported as decreasing in the double mutant, while my studies indicate it increases [7]. This could be due to growth conditions, as all four lines in the pah1pah2 background show the same increase. There were also minor changes in the triple and quadruple mutants in 18:2 and 18:3 relative to WT and pah1pah2: the addition of the spp1 and lpap1 mutations seemed to recapitulate the WT phenotype. Additionally, pah1pah2spp1lpap1 showed a lower 20:1 level than either pah1pah2 or WT.
However, these are relatively minor changes, and are not indicative of any redundancy between the genes tested. Thus, the SPP1, LPAP1, PAH1, and PAH2 genes do not exhibit redundancy with each other beyond that which is characterized for PAH1 and PAH2.

### 3.3.4. Enzyme Assays: Complementation of the Yeast \( \Delta paH1 \) Mutant

In yeast, \( ScPAH1 \) is the main PAP involved in oil synthesis, and the null mutant \( \Delta paH1 \) has reduced oil levels and displays a temperature sensitive phenotype. Both the single mutant and the triple mutant \( \Delta paH1\Delta lpP1\Delta dpP1 \) have been used to characterize PAPs, including AtPAH1 and AtPAH2 [8,14,15], by providing a complementation assay to confirm PAP activity. Genes with PAP activity which complement the mutant allow growth at 37°C. The Arabidopsis LPPβ protein was expressed in both yeast mutants, under control of the strong, constitutive \( ADH \) promoter, as well as under control of the native \( ScPAH1 \) promoter. The four expression conditions provided comparable assay results, but the \( \Delta paH1 \) single mutant grew more consistently. As shown in Figure 6, LPPβ usually grew to similar levels as AtPAH1 and ScPAH1, and grew more than the negative control. LPAP1 grew to similar levels as LPPβ and AtPAH1, but SPP1 grew noticeably less. This indicates that LPPβ, LPAP1, and SPP1 all display some PAP activity, although LPPβ and LPAP1 are better able to complement the mutant. Additionally, in some assays, the negative control also grew, indicating that the assay is not consistent. More studies are needed to confirm or refute the PAP activity of LPPβ.

### 3.3.5. Enzyme Assays: Microsomal Yeast and E. coli PAP Activity Assays

PAP assays using \( ^{14}C-PA \) (dipalmitoyl) were undertaken. The \( \Delta paH1 \) yeast expressing LPPβ under control of the strong, constitutive \( ADH \) promoter from the complementation assays were used. The inconsistent growth phenotypes affected these assays as well, which is why there is no PAP assay data for SPP1 – too few cultures grew for analysis. However, I was able to assay
LPPβ and LPAP1 for activity, and LPPβ showed an average of 4 pmol/min/ug protein, while LPAP1 showed virtually no activity (see Figure 7A). The yeast expressing PAH1 showed an average activity of 32 pmol/min/ug protein, indicating that LPPβ is not as active as PAH1 under these assay conditions. These data do indicate, however, that LPPβ has PAP activity.

LPPβ, LPAP1, and SPP1 were also expressed in E. coli as C-terminal GFP fusions. This allowed confirmation that the proteins were expressed, by Western Blotting with Anti-GFP polyclonal antibodies. Western Blots showed that, indeed, the fusions were expressed, although the GFP tag was cleaved from the LPPβ and SPP1 proteins (data not shown). However, the enzyme assays showed no PAP activity at all (see Figure 7A), possibly indicating that these proteins undergo post translational modifications which are not available in E. coli.

While neither the complementation assays nor the PAP radioactivity assays were conclusive alone, together they indicate that LPPβ does exhibit PAP activity. Further studies to determine the lipid and fatty acid specificities of the protein are needed.

### 3.3.6 WRII CONTROL OF LPPβ

The WRII transcription factor has long been associated with oil synthesis, and the wri1-1 mutant has a compromised glycolysis pathway making the developing embryo unable to effectively convert sugar precursors into the fatty acids required for TAG synthesis, and leaving the mutant with an 80% reduction in seed TAG [27]. WRII has been shown to control transcript levels of enzymes involved in the generation of TAG precursors, but not the TAG assembly pathway itself [27-30]. It acts by binding to the AW Box domain in the promoter of its targets and directly activates them. LPPβ has an AW Box sequence roughly 117 bp upstream of the translation start site, indicating that like other genes with the AW Box, it may be regulated by WRII (see Figure 8A). To determine if this is the case, developing seeds were harvested from a segregating
population of plants overexpressing \textit{WRII}. Expression levels of \textit{WRII} were increased 4-6 fold in overexpressor plants compared to their isogenic segregants, but \textit{LPP}\textsubscript{β} expression decreased about 20\% (Figure 8B). Although I was unable to measure oil in these lines, other lines overexpressing the same construct of \textit{WRII} which had only 2.5 or 3 fold increase in expression still showed a 19\% and 15\% increase in oil, respectively (Neil Adhikari, personal communication). This indicates that the 4-6 fold increase seen in \textit{WRII} is enough to affect oil synthesis, and should have had an effect on \textit{LPP}\textsubscript{β} if it was a target. Since \textit{LPP}\textsubscript{β} expression in \textit{WRII} overexpressors does not show a significant change from WT, it appears that it is not regulated by \textit{WRII}, despite the presence of the AW Box in its promoter.
3.4 DISCUSSION

3.4.1 LPPβ SHOWS MODEST PAP ACTIVITY

The PAP activity assays and yeast complementation assays indicate that LPPβ displays PAP activity and that LPAP1 showed less activity. LPAP1 was published to show low PAP activity, but much stronger activity towards the substrate lysophosphatidic acid (LPA) [9]. In the complementation assays, it appeared to complement the mutant, but in my radioactivity tests, LPAP1 averaged nearly zero PAP activity. This supports the conclusion that it acts as an LPA phosphatase with only minor PAP activity. In radioactivity assays, LPPβ showed roughly one tenth the activity of PAH1; this indicates that it is a PAP, but the low activity could be due to fatty acid substrate specificity, assay or buffer conditions, or lipid substrate specificity. Coupled with the complementation assays which actually showed similar levels of complementation to AtPAH1 and ScPAH1, this presents solid evidence that LPPβ does indeed encode PAP activity. It is also consistent with unreviewed reports (e.g. not published in a peer reviewed journal) that the protein is a PAP [31].

In order to attempt to identify other possible reactions of the enzyme, the TLC separation protocol allowed for separation of nonpolar lipids (TAG, DAG, free fatty acids), as well as polar lipids (including PA, PC, PE, PI; see Figure 7B). This allowed tracking of 14C PA not only into DAG, but also into side reactions that use PA as a substrate, such as PS. There were no radioactive side products seen in any of the assays conducted with PA, indicating that DAG is the main product when dipalmitoyl PA is used. More analyses are needed to determine the specificity of LPPβ for other molecular species of PAP and to other lipids in vitro.
3.4.2 Attempts to Knock Down LPPβ Expression Have Failed

There is only one T-DNA line available for LPPβ and it is not a null mutant, which perhaps is not surprising given the small size of the gene (642 bp). However, the amiRNA and ihpRNA data were interesting in that many showed an increase in the transcript level instead of a decrease. Additionally, a CRISPR/Cas9-induced mutation in LPPβ has not yet been identified at the time of this publication. Over 100 lines with two gRNA cassettes have been screened, and none have the 200-bp deletion which would have resulted from both of gRNA’s successfully guiding Cas9 to a cut site in the same primary transgenic, despite reports that in stably transformed lines, 20-100% of lines have resulted in a double cut (deletion) [32-34].

One reason that three separate attempts to knock down expression have failed could be that a null mutation in LPPβ is lethal. It is indeed possible that all three failed due to the technical limitations of each method, especially since CRISPR is a new, unproven technology and RNAi techniques are sometimes limited by endogenous feedback mechanisms. But lethal mutations are some of the most difficult to characterize, because the phenotype is a lack of mutated progeny. A CRISPR/Cas9 system that is effectively mutating an essential gene and causing lethality is very difficult to distinguish from a system that is not working at all. Until a mutant with a significant decrease in can be isolated for LPPβ, perhaps in the form of a heterozygous mutation, the lack of a genetic test will prevent our full understanding of the role of LPPβ in the plant.

3.4.3 SPP1 and LPAP1 Are Not Involved in Oil Synthesis

PAH1 and PAH2 are the only PAPs in Arabidopsis to date to show any effect on oil synthesis. They act as a convenient background to determine which other PAPs act redundantly in oil synthesis; while the main PAP involved in TAG synthesis may show a strong phenotype alone, there is also the possibility that it will act redundantly with PAH1 and PAH2 or other genes, making
a multiple mutant necessary to determine its involvement. *LPAP1* has shown modest PAP activity, both in my experiments and in published reports [9], indicating that it could be redundant with *PAH1* and *PAH2* or the main PAP involved in oil synthesis. Additionally, *SPP1* showed minor complementation of the *Δpah1* yeast mutant, and it displayed enough interesting characteristics to warrant testing its redundancy with the other putative PAPs. Neither had been assayed for a decrease in seed oil in the mutant, so a possible involvement in oil synthesis, even if it didn’t encode a PAP, could not be ruled out.

However, none of the combinations of mutants (*spp1, lpap1, spp1lpap1, pah1pah2spp1, pah1pah2lpap1, or pah1pah2spp1lpap1*) showed a total seed oil phenotype (or a growth phenotype) beyond what has already been published for *pah1pah2*. This indicates several things. First, neither *SPP1* nor *LPAP1* are involved in oil synthesis alone. Second, since they don’t show an additional phenotype in combination with *pah1pah2*, they are probably not redundant with *PAH1* or *PAH2* in any of their known phenotypes, including oil synthesis. This clearly shows that *SPP1* and *LPAP1* can be conclusively dismissed from the conversation of which gene encodes the main PAP involved in oil synthesis.

### 3.4.4 LPPβ IS NOT TRANSCRIPTIONALLY REGULATED BY WRI1

Even though the promoter of *LPPβ* contains an AW Box indicating it may be regulated by *WRI1*, transcript levels of *LPPβ* do not increase in *WRI1* overexpressors (Figure 8). The lack of regulation at the transcript level of *LPPβ* is consistent with the fact that the other known TAG synthesis genes *GPAT9, LPAAT4, DGAT1*, and *PDAT1* are also not regulated transcriptionally by *WRI1* [30]. It may seem contradictory that *WRI1* controls the amount of precursors available to TAG synthesis, and eventually affects the amount of oil produced but does not directly target the oil synthesis pathway. However, it could be that TAG synthesis is not the rate-limiting step, and
that the increased substrate flux dictates the eventual levels of oil. Since most heterozygous mutations in oil synthesis and other pathways show a WT phenotype, I expect that most enzymes in TAG synthesis are in excess and able to handle the increased substrate load (see Chapter 4 sections 4.3.6 and 4.4.2, regarding GPAT9 as the rate-limiting step of oil synthesis).
3.5 Conclusion

PAP activity is necessary for lipid synthesis in the plant, but much remains to be understood about the gene and protein responsible for the step. I was able to confirm that \textit{SPP1} and \textit{LPAP1} are not redundant with \textit{PAH1} and \textit{PAH2} in oil synthesis, and their functions are likely confined to those already published. \textit{LPPβ} exhibits at least modest PAP activity in multiple assays, indicating that it does indeed encode PAP activity. Since I was unable to isolate a null mutant, its involvement in oil synthesis cannot be confirmed or rejected. Thus, a genetic test to determine its role in oil synthesis would clarify the physiological role of \textit{LPPβ} in Arabidopsis.
Figure 1. RT-PCR of SALK Insertional Mutants.
(A), (B), and (C) Gene schematic of the LPPβ, SPP1, and LPAP1 genes, respectively. The T-DNA insertion is denoted with an empty triangle. The two regions of LPPβ targeted for RNA interference are noted with green arrows; regions targeted by the CRISPR/Cas9 guided mutation system are indicated with blue arrows.
(D) Actin8 RT-PCR control. Similar amounts of cDNA were loaded in all samples. Numbers represent the line tested; C: Cauline leaf on WT plant #1, L: rosette leaf on WT plant #2.
(E) RT-PCR analysis of lppβ T-DNA mutants indicate that it is expressed, but spp1-2 and lpap1-1 are both null mutants.
Figure 2: Oil Levels in LPPβ RNAi lines
Paired red and black bars represent seeds off the same plant, from one independent transformation event. Standard error bars represent standard error of three samples of 30 seeds each. Red bars, seeds containing the LPPβ (LB) RNAi construct; Black bars, isogenic segregants. “hpi”, ihpRNA construct; “ami”, amiRNA construct. * - statistically significant decrease in oil in KD seeds. (A) WT background; (B) $lhap1$ background; (C) spp1 background; (D) $pah1pah2$ background.
Figure 3: LPPβ expression in RNAi Plants
Expression was normalized to the Tubulin2 gene. Error bars represent standard error of three biological replicates. Grey line represents WT level of expression for comparison.
Figure 4. *LPPβ* Expression in Multiple Mutant Lines. Error bars represent standard error of three biological replicates. Grey line represents WT level of expression for comparison.
Figure 5: Seed Oil and Fatty Acid Content of Multiple Mutants.
Data is from plants grown in a growth chamber, but is largely similar to those grown in a greenhouse. Error bars represent standard error of 15-85 plants. (A) Oil levels in the seed, WT background mutants. (B) Oil levels in the seed, \textit{pah1pah2} background mutants plus WT for comparison. (C) Fatty acid composition of seed oil, a representative chart (growth chamber grown). PL: \textit{lpap1 pah1pah2}; PS: \textit{spp1 pah1pah2}; PSL: \textit{spp1 lpap1 pah1pah2}
Figure 6. PAP Complementation Assays. A representative set of plates, using ∆pah1∆lpp1∆dpp1 yeast and expressing genes under control of the PAH1 promoter. (A) Growth of complemented yeast at 30°C. (B) Growth of complemented yeast at 37°C.
Figure 7. PAP Activity Assays.
(A) PAP activity in yeast microsomes and E. coli crude extract, normalized to negative control (GUS for Yeast and GFP for E. coli) and background subtracted. (B) Iodine-stained TLC plate of PAP activity assay showing separation of both polar and neutral lipids. Shown is an experiment using yeast microsomes. BHT, butylated hydroxytoluene, (an antioxidant used to protect the fatty acids from oxidative degradation).
<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>AW Box Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-PKβ1</td>
<td>-2</td>
<td>TGGTTAGTTTGATACGAA</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>GCCTTCGTAAGTTCGACC</td>
</tr>
<tr>
<td>BCCP2</td>
<td>-12</td>
<td>TTCTTCGTAAGCATGCAGA</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>TTCTTCGCTTTTCATCGTC</td>
</tr>
<tr>
<td>KAS1</td>
<td>-54</td>
<td>TTCTCGGAATAATTCGCT</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>AACATCGTACTTATCCCG</td>
</tr>
<tr>
<td>ENR1</td>
<td>2</td>
<td>AGCTTCGATGAGATCGAG</td>
</tr>
<tr>
<td>ACP1</td>
<td>-890</td>
<td>CTCTTTCACACTTCGCC</td>
</tr>
<tr>
<td>FATA</td>
<td>-27</td>
<td>GTCTTCGTCCTCATGTC</td>
</tr>
<tr>
<td></td>
<td>-12</td>
<td>GTCTGTCACAAATACGAA</td>
</tr>
<tr>
<td>SUS2</td>
<td>-238</td>
<td>CACTTAGAGCGTATCGGA</td>
</tr>
<tr>
<td>LPPβ</td>
<td>-117</td>
<td>CTCTTAGAAGCGGTCGTC</td>
</tr>
</tbody>
</table>

Figure 8. LPPβ is not regulated by WRII at the transcript level. (A) The promoter of LPPβ contains an AW box, a known WRII-binding domain found in genes activated by WRII. Location is relative to ATG start codon. **Underlined:** Activation by WRII was verified in Maeo et al. [35]. **Bold:** Activation verified in To et al. [30]. **Red text:** Required for fatty acid synthesis. **Blue text:** Involved in glycolysis. Blue highlighting: conserved motif of the AW Box domain. (B) LPPβ transcript level is not increased in WRII overexpression plants. Values are the means and standard error of four or five individual plants. Grey line: WT level of expression of both WRII and LPPβ.
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBF1</td>
<td>ATGTCTCGTCCGTCGACGACG</td>
<td>This Work</td>
</tr>
<tr>
<td>LBF1G</td>
<td>CACCATGTCTCGTCCGTCGACGACG</td>
<td>This Work</td>
</tr>
<tr>
<td>LBF4</td>
<td>GACAGAGATCTGTTACGACGAGTTATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LBR1</td>
<td>ACCTCCCAAAGATCATTTCA</td>
<td>This Work</td>
</tr>
<tr>
<td>LB qPCR F3</td>
<td>TGCTGAAGCTTTCCATGACTG</td>
<td>This Work</td>
</tr>
<tr>
<td>LB qPCR R2</td>
<td>CTCGAGAAATCAGAGAGCAAGAAGAGAGACG</td>
<td>This Work</td>
</tr>
<tr>
<td>LD F1</td>
<td>ATGAAGAGATATTCTAGAAGATCAATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LD F1G</td>
<td>ATGAAGAGATATTCTAGAAGATCAATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LD F4</td>
<td>AAATCTCAACAATCTGTCGAAATGCTCTCTC</td>
<td>This Work</td>
</tr>
<tr>
<td>LDR3</td>
<td>GGCACTGATTTATATATACACAAATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LDR6</td>
<td>CTCAAGTGAACGAGAAGAGATAGGAAGAGTGAAGG</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1 F1</td>
<td>ATGGTGAGGAAGAGGTAGCTCTCTG</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1 F1G</td>
<td>ATGGTGAGGAAGAGGTAGCTCTCTG</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1 R3</td>
<td>TCAAATTCACAAATCAGAAGAGAAGAGTGAAGG</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1 R6</td>
<td>CAATTTCTCAACCGCTCTTTCA</td>
<td>This Work</td>
</tr>
<tr>
<td>PAH1FW</td>
<td>GCGGCCGCATGATGTGTTGTTGAGAGGAGGAGGAGG</td>
<td>[8]</td>
</tr>
<tr>
<td>PAH1RV</td>
<td>ACGGCTACCATTGAGAGGAGGAGGAGGAGGAGG</td>
<td>[8]</td>
</tr>
<tr>
<td>ScPAH1FWG</td>
<td>CACCATGGTGGAGGAAGAGGAGGAGGAGGAGG</td>
<td>This Work</td>
</tr>
<tr>
<td>ScPAHR1</td>
<td>TTAACTTTCCGAGATTCAATCTCTG</td>
<td>This Work</td>
</tr>
<tr>
<td>amiRNA-A-G</td>
<td>CACCCTGCAAGCGATAGGTTGAGGAGGAGG</td>
<td>[16], This Work</td>
</tr>
<tr>
<td>I miR-s</td>
<td>GAATATCTCCCCCTAAAAAGGACACCCTGTCTCTCTTTGTA</td>
<td>[16], This Work</td>
</tr>
<tr>
<td>II miR-a</td>
<td>GACGGGTGGCTTTTTTTTATGATAGGATAACATAGA</td>
<td>[16], This Work</td>
</tr>
<tr>
<td>III miR*s</td>
<td>GACGGATGGCTTTTTTTTTATGATAGGATAACATAGA</td>
<td>[16], This Work</td>
</tr>
<tr>
<td>IV miR*a</td>
<td>GAATATCTCCCCCTAAAAAGGACACCCTGTCTCTCTTTGTA</td>
<td>[16], This Work</td>
</tr>
<tr>
<td>amiRNA-B</td>
<td>GCGGGATAACATTTTCACACAGGAAACAG</td>
<td>[16], This Work</td>
</tr>
<tr>
<td>hp1F-G</td>
<td>CACCCTGGTGCTAGATCTGACTCTG</td>
<td>This Work</td>
</tr>
<tr>
<td>hp1R</td>
<td>CACCAGAGCTCAGACTGAGA</td>
<td>This Work</td>
</tr>
<tr>
<td>U6-XhoI-F1</td>
<td>CGTGACATCGAGAGAAATCTCCAAAATCCG</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>3’-PacI-R2</td>
<td>CGTGATTAATATTAATATTGCCAATTTTATCAGA</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>R1-A</td>
<td>TCTTAGCGGTGGACCGCCCGGCACAACTACTCCATTCATATATA</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>F2-A</td>
<td>GCGCCGCTCCACCGCTAAGAGTTTAAGAGCTAGA</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>R1-B</td>
<td>CAATATCGCTTTCTCTCTGTAATCATCTACTCTGTCT</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>F2-B</td>
<td>ACGAGGAAAGACGGTAATGGAGGAGTTTAGAGCTA</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>R1-C</td>
<td>CCAGGATTAGCTCGAGGATATTAGCAGGAGTGAAGAGAGATAGC</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>F2-C</td>
<td>ATCTCACGCGATTTCCACACAGGGTTTAGAGCTAGA</td>
<td>[20], This Work</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>PPC F1</td>
<td>CTAAAGGGAAACAAAAAGCTGGAGGAGCGTCgeggtacctagagtcagaacacaggc</td>
<td>This Work</td>
</tr>
<tr>
<td>PPC R1</td>
<td>TGATGGGGCATCCACTAGTTCTAGAATCCGACATATGgccccagtaacacggatgtggtc</td>
<td>This Work</td>
</tr>
<tr>
<td>PPC F1*</td>
<td>GGTCTGTTGAGTTGGACTCTAGGTACCGGACGTCCTCCATTTAG</td>
<td>This Work</td>
</tr>
<tr>
<td>PPC R1*</td>
<td>GACACATCGGTGGATTATAGAATTCCGGACATATGCGGATTCTAGATGTGGATCCCCCATCA</td>
<td>This Work</td>
</tr>
<tr>
<td>LBF1-Lic</td>
<td>TTTAAGAAGGAGATATAGATCATGTCTCCGACGAGC</td>
<td>This Work</td>
</tr>
<tr>
<td>LBR1-NsLic</td>
<td>GTTGGAGGATGAGAGGATCCCTCCAAAGATCATTTCATCA</td>
<td>This Work</td>
</tr>
<tr>
<td>LDF1-Lic</td>
<td>TTTAAGAAGGAGATATAGATCATGAAGAAGATTCTAGAAAATAGATATG</td>
<td>This Work</td>
</tr>
<tr>
<td>LDR6-NsLic</td>
<td>GGGGAGGATGAGGAGGATCCCTCCAAAGATACGTAAAGTTAACGTAATAGGGAGGAGAATTTCATCA</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1 F1-Lic</td>
<td>TTTAAGAAGGAGATATAGATCATGTGGAGGAGAAGCAGTCGAGCT</td>
<td>This Work</td>
</tr>
<tr>
<td>UP1 R3-NsLic</td>
<td>GTTGGAGGATGAGGAGGATCCCTTCATACAAATCCAAACGGAGAATA</td>
<td>This Work</td>
</tr>
<tr>
<td>P1F1</td>
<td>GTGCTCAGTTTGACAATCTAAATGG</td>
<td>This Work</td>
</tr>
<tr>
<td>P1R1</td>
<td>TAGATTCGATCTCCAAGAGTACG</td>
<td>This Work</td>
</tr>
<tr>
<td>P2F1</td>
<td>CTGTTGTTATCTCTCTTGATGTGTC</td>
<td>This Work</td>
</tr>
<tr>
<td>PAH2Rv</td>
<td>CTGCAGTCACATAAGCGATGGAGGAGGAGGAGGAGGAGGAGGAACATAGATGTA</td>
<td>Nakamura et al 2009</td>
</tr>
<tr>
<td>pROK2 deep</td>
<td>TTTAAGGTTCCGAGTTAGTGCTTTA</td>
<td>This Work</td>
</tr>
<tr>
<td>qWR1 (3)F</td>
<td>GACCGGACACCACATGTTGGAAT</td>
<td>Neil Adhikari</td>
</tr>
<tr>
<td>qWR1 (3)R</td>
<td>GGCAGGAGAAGCCCAAT</td>
<td>Neil Adhikari</td>
</tr>
</tbody>
</table>

Table 1. Primers used in this work. **Blue Bold**: Gateway cloning tags. **Red Bold**: amiRNA targeting **LPPβ**. **Green Bold**: LIC cloning tags. **Orange Bold**: gRNA targeting **LPPβ** in CRISPR/Cas9 system. **Purple Bold**: Restriction Site.
3.6 References


CHAPTER FOUR
IDENTIFICATION OF ARABIDOPSIS GPAT9 (AT5G60620) AS AN ESSENTIAL GENE INVOLVED IN TRIACYLGLYCEROL BIOSYNTHESIS

ABSTRACT

The first step in the biosynthesis of nearly all plant membrane phospholipids and storage triacylglycerols is catalyzed by a glycerol-3-phosphate acyltransferase (GPAT). The requirement for an endoplasmic reticulum (ER) localized GPAT for both of these critical metabolic pathways was recognized more than 60 years ago. However, identification of the gene(s) encoding this GPAT activity has remained elusive. Here we present the results of a series of in vivo, in vitro, and in silico experiments designed to assign this essential function to AtGPAT9. This gene has been highly conserved throughout evolution, and is largely present as a single copy in most plants, features consistent with essential housekeeping functions. A knockout mutant of AtGPAT9 demonstrates both male and female gametophytic lethality phenotypes, consistent with the role in essential membrane lipid synthesis. Significant expression of AtGPAT9 in developing seed is required for wild-type levels of triacylglycerol accumulation, and transcript level is directly correlated to the level of microsomal GPAT enzymatic activity in seeds. Finally, the AtGPAT9 protein interacts with other enzymes involved in ER glycerolipid biosynthesis, suggesting the possibility of ER-localized lipid biosynthetic complexes. Together these results suggest that GPAT9 is the ER-localized GPAT enzyme responsible for plant membrane lipid and oil biosynthesis.
**Preface**

This chapter of my dissertation is a manuscript accepted for publication in Plant Physiology and includes data from Jay Shockey\(^1\), Anushobha Regmi\(^2\), Kimberly Cotton\(^3\), Neil Adhikari\(^3\), John Browse\(^3\), and Philip Bates\(^1\†). I was responsible for the data obtained in the qrt1 mutant background (crossing, tetrad analysis, pollen germination) and assisting in manuscript preparation. This data is found in Table 4 (“Pollen germination and tube growth of \(GPAT9/gpat9-2\) within the \(qrt1\) background”) and Figure 4 (“Reduced Pollen Size and Pollen Tube Growth of \(GPAT9/gpat9-2\) Pollen in the \(qrt1\) Background”).

4.1 Introduction

Glycerol-3-phosphate acyltransferase (GPAT, EC 2.3.1.15) was first characterized biochemically in extracts from animal and plant tissues in a series of reports beginning more than 60 years ago [1-3]. GPAT transfers an acyl moiety from either acyl-CoA or acyl-acyl carrier protein (ACP) to the \(sn-1\) position of glycerol-3-phosphate to produce 1-acyl-2-lyso-glycerol-3-phosphate (or lysophosphatidic acid). It catalyzes the first step in the synthesis of various types of glycerolipids including membrane lipids and triacylglycerol (TAG) in all living organisms, and thus plays an extremely important role in basic cellular metabolism. In addition, the enzymes which esterify fatty acids (FAs) to the glycerol backbone of TAG control the TAG FA composition, and thus the functional value of plant oils. Acyl selective isozymes for the last step in TAG assembly have demonstrated to be crucial for enhancing the control of seed oil content.

---

\(^1\) These co-authors contributed equally to this work.
\(^1\) United States Department of Agriculture – Agricultural Research Service, Southern Regional Research Center, New Orleans, LA 70124
\(^2\) Department of Chemistry and Biochemistry, The University of Southern Mississippi, Hattiesburg, MS 39406
\(^3\) Institute of Biological Chemistry, Washington State University, Pullman, WA 99163
\(^†\) Corresponding author. Philip.Bates@usm.edu
through bioengineering [4]. Therefore, our ability to produce designer plant oils that will meet the nutritional or industrial needs of the future will require identifying the genes which encode all of the acyltransferase steps in TAG production, including the essential first step catalyzed by GPAT.

Early *in vitro* biochemical characterization of GPAT activities from various plant tissues demonstrated multiple sub-cellular locations for GPAT activity including endoplasmic reticulum (ER) [3,5,6], plastid [7], and mitochondria [8]. These, and other gene-independent biochemical analyses of plant lipid metabolic activity have in part contributed to the “two-pathway” hypothesis for plant glycerolipid synthesis that has stood the test of time and has been reviewed extensively [9-11]. Plants contain two parallel lipid biosynthetic pathways each containing distinct GPAT enzymes to produce glycerolipids. The ER localized “eukaryotic pathway” utilizes a membrane-bound acyl-CoA dependent GPAT for synthesis of lysophosphatidic acid (LPA). LPA is subsequently acylated to produce phosphatidic acid (PA) by an acyl-CoA dependent lysophosphatidic acid acyltransferase (LPAAT). PA, or its dephosphorylated diacylglycerol (DAG), provide the glycerolipid backbones for ER localized phospholipid or triacylglycerol (TAG) synthesis. The three consecutive acyl-CoA dependent acylations of the glycerol backbone to produce TAG in the ER is commonly called the Kennedy pathway [2,12].

In contrast to the eukaryotic glycerolipid assembly pathway, the plastid-localized “prokaryotic pathway” utilizes a soluble acyl-ACP dependent GPAT, and an acyl-ACP dependent LPAAT to produce PA for phosphatidylglycerol synthesis in plastids of all plants, and the DAG for galactolipid synthesis in some plants. The prokaryotic pathway GPAT has been characterized extensively through both biochemical and molecular genetic approaches demonstrating species-specific acyl selectivities which control the composition of prokaryotic glycerolipids [13-17]. In addition to prokaryotic galactolipid synthesis in the plastid, plants also utilize DAG derived from
the eukaryotic pathway synthesized membrane lipid phosphatidylcholine (PC) for galactolipid synthesis. The PC-derived DAG is imported into the plastid for galactolipid production, and contains a distinctive eukaryotic FA composition [18]. Therefore the ER-localized GPAT of the eukaryotic pathway is vital for lipid assembly within both the ER and the plastid.

The essential nature of the ER GPAT for membrane lipid biosynthesis, and its role in controlling one-third of the FA composition within economically valuable plant oils, has inspired many studies focused on characterizing the ER GPAT enzyme. Microsomal extracts from a wide range of plant tissues have demonstrated a variety of in vitro ER GPAT activities utilizing various common and unusual FAs as substrates [19-24]. However, dissection of the range of potential effects of the ER-localized GPAT on seed oil content and FA composition has still remained unclear, awaiting the identification and molecular characterization of the requisite gene(s) encoding the ER localized GPAT.

Zheng et al. (2003) was first to classify a family of genes as ER and mitochondrial localized GPATs in plants. This eight gene family was identified based on homology to the acyltransferase domains of GPATs from yeast, and most members of this group demonstrated GPAT activity when heterologously expressed. Genetic analysis of Arabidopsis mutants defective in the first of these genes, designated \textit{AtGPAT1}, indicated altered tapetal differentiation and frequent abortion of microspores impacting male fertility, and suggesting that \textit{AtGPAT1} may be involved in essential membrane lipid production [25]. However, more recent biochemical and genetic characterization has revealed that \textit{AtGPAT1-AtGPAT8} are a land plant-specific family of \textit{sn}-2 GPATs involved in cutin and suberin synthesis [26-29]. These GPATs predominantly utilize acyl groups not found in membranes or TAG (e.g. \textit{ω}-hydroxy and \textit{α,ω}-dicarboxylic FAs) to produce the \textit{sn}-2 monoacylglycerol precursors utilized during extracellular polymerization of both the cutin and
suberin barriers of land plants [30]. Therefore, the \textit{GPAT1-GPAT8} family is likely not involved in membrane lipid and triacylglycerol biosynthesis required in all living organisms.

Another gene, designated \textit{AtGPAT9}, previously has been identified as a candidate for the ER membrane and oil biosynthetic GPAT activity based on substantial sequence identity to mouse and human \textit{GPAT3} [31,32]. The expression profiles and biochemical properties of \textit{mGPAT3} and \textit{hGPAT3} strongly support a role for these enzymes in TAG biosynthesis within lipid-rich mammalian tissues. However, definitive determination of the roles of \textit{AtGPAT9} in plant lipid metabolism has remained elusive. Complete understanding of plant lipid biosynthesis, and the ability to predictably engineer designer plant oils, awaits identification of all the component enzymes (and associated genes), and determination of how each one functions within the lipid metabolic network to produce the thousands of different possible molecular species of membrane and storage lipids.

The molecular identification of an ER-localized GPAT involved in membrane lipid and/or TAG biosynthesis in plants has not been accomplished until now. Here we present the results of a series of experiments designed to address the role of \textit{AtGPAT9} in lipid biosynthesis. We present strong evidence that: (A) \textit{AtGPAT9} is a highly conserved, single copy, and essential gene which when knocked out is homozygous lethal and causes both male and female gametophyte lethalties; (B) significant expression of \textit{AtGPAT9} is required for normal levels of microsomal GPAT enzymatic activity in seeds, and is required for wild-type levels of TAG accumulation \textit{in vivo}; (C) \textit{AtGPAT9} protein interacts with other enzymes involved in ER glycerolipid biosynthesis, suggesting the possibility of ER-localized lipid biosynthetic complexes. Together these results suggest that GPAT9 is the ER-localized GPAT enzyme responsible for plant membrane lipid and TAG biosynthesis.
4.2 MATERIALS AND METHODS

4.2.1 PHYLOGENETIC ANALYSES

Plant GPAT9 and LPEAT protein sequences were identified from the species-specific searches of the Phytozone version 10.3 server (http://phytozone.jgi.doe.gov/pz/portal.html#). Phylogenetic comparisons were carried out using alignment (.aln) files generated using the default settings provided in CLUSTALX (v1.8) [33]. Neighbor joining dendograms were created in TreeView program [34,35], using the Boxshade tool at the ExPASy Bioinformatics Resource Portal (http://www.ch.embnet.org/software/BOX_form.html).

4.2.2 GENOTYPING OF T-DNA MUTANTS

Mutant genotyping utilized gene and T-DNA specific primers (Supplemental Table S1), and GoTaq Green (Promega, www.promega.com) with manufacturers recommended protocol. For insertion site sequencing the left border-gene junction was amplified by PCR with GoTaq® Flexi (Promega), gel-purified by QIAquick Gel Extraction Kit (QIAGEN and sequenced from both the left border primer and a gene-specific primer in the flanking sequence by Eurofins (http://www.eurofinsgenomics.com/).

4.2.3 GENE EXPRESSION ANALYSIS

Leaf RNA was extracted with RNeasy Plant Mini Kit (QIAGEN, www.qiagen.com) and treated with RNase-free DNase (QIAGEN) using on-column DNase digestion. RNA was quantified using a Nanophotometer (Implen Inc). cDNA was synthesized using SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies). RT-PCR utilized standard conditions as per the GoTaq® Flexi (Promega) protocol with 30 cycles of amplification. Transcript levels were analyzed by quantitative PCR (qPCR) using Platinum SYBR Green qPCR SuperMix-UDG (Life Technologies) and the Stratagene Mx3005P Quantitative PCR system (Agilent
Technologies Inc). For developing seed transcript analysis, frozen developing seeds (9-10 DAF) were collected from liquid N₂ frozen siliques [36]. Approximately 50-100 mg of frozen seeds were mechanically pulverized to a fine powder with steel beads (TissueLyser LT, QIAGEN), RNA extracted [37], and DNA removed (DNA-Free RNA Kit™, Zymo Research, www.zymoresearch.com/). cDNA synthesis and quantification were performed as described above. Normalized Relative Quantity (NRQ) calculated [38] against TIP41-like [39].

4.2.4 MATERIALS, PLANTS, AND GROWTH CONDITIONS

Unless specified all chemicals are from Fisher Scientific (https://www.fishersci.com), and solvents are HPLC grade or higher. *Arabidopsis thaliana* lines gpat9-1, Salk_052947C [40] and qrt1-4 CS25041 [41] were obtained from ABRC (https://abrc.osu.edu/), and gpat9-2 GABI_867A06 [42] were obtained from Gabi-kat (https://www.gabi-kat.de/). Control plants were the Arabidopsis Col-0 ecotype. Seeds were surface sterilized (30% ethanol, 10% bleach, 0.1% SDS) for 5 min, rinsed 5 times with sterile water and plated onto germination media (2.5 mM MES pH 5.7, 1% sucrose, 1x Murashige and Skoog Plant Salt Mix, (MP Biomedicals, http://www.mpbio.com/), 0.8% agar). Seeds from GPAT9/gpat9-2 plants were always germinated on the plate media containing an additional 5.25 mg/L sulfadiazine (Sigma, www.sigmaaldrich.com/) to select for the heterozygotes. Plated seeds were stratified for 4 days at 4 °C, prior moving to a growth chamber. After 10 days on the plate seedlings were transferred to soil. Plants were grown in growth chambers under continuous white light ~130-170 μmol photons m⁻² sec⁻¹, at 22-24 °C. Silique ages were determined by trimming each plant to one main shoot and counting the new open flowers/siliques produced each day.
4.2.5 DNA MANIPULATION AND PLASMID CONSTRUCTION

Binary complementation plasmids were generated as follows. The phaseolin promoter from cloning vector pK8 [43] was removed by KpnI and NotI digestion and replaced with an 1594 bp portion of the 5’ upstream region of the AtGPAT9 gene, resulting in the synthesis of cloning vector pK51. The open reading frames of AtGPAT9 and tung GPAT9 (VfGPAT9) were cloned into the NotI and SacII sites of pK51, resulting in plant shuttle plasmids pB447 and pB448, respectively. The AscI fragments bearing the respective AtGPAT9 promoter:GPAT9 ORF:35S terminator cassettes from pB447 and pB448 were purified by gel electrophoresis and ligated into the AscI site of the DsRed-selectable binary vector pB110 [43], resulting in the final plant transformation binary plasmids pE434 and pE437. Binary plasmids were transformed into Agrobacterium tumefaciens strain GV3101 via electroporation and selection on solid media containing 50 µg/ml each of kanamycin and gentomycin. Individual A. tumefaciens colonies were inoculated into liquid media containing the same antibiotics and cultured overnight at 28 °C prior to plant transformation by floral dip, as described previously [44].

4.2.6 MUTANT COMPLEMENTATION

Approximately 100,000 T1 seeds from each of the three transformations were sown on flats of soil for approximately 10 days, then heavily misted with a solution of 500 mg/L sulfadiazine containing 0.03% of the surfactant Silwet L-77, as described in Thomson et al. [45]. Plants that survived the sulfadiazine application and displayed red fluorescence were transferred to pots of untreated soil and grown to maturity. At the end of this experiment, nine, twenty-five, and thirty individual T1 transgenic plants were identified for B110 control, E434, and E437, respectively. Representative lines from each transgenic genotype were propagated to the T3/T4 generations, with sulfadiazine selection, until plants producing homozygous red fluorescent seeds were identified.
Batches of seeds from these plants, representing empty vector control B110 and *GPAT9* overexpressor E434 and E437 lines, were surface-sterilized and sown on sulfadiazine agar plates for complementation testing.

### 4.2.7 *GPAT9/gpat9-2* Gametophyte Viability Analysis

To determine the rate of *gpat9-2* ovule abortion, aborted ovules were counted in opened half silique from Col-0 and *GPAT9/gpat9-2*, and imaged under a Leica DM2000 microscope equipped with a DFC295 camera. Alexander staining of aborted pollen, and Nitro Blue Tetrazolium (NBT) staining for pollen viability was done by previously optimized methods [46,47].

### 4.2.8 Analysis of Qrt1-4 *GPAT9/gpat9-2* Pollen

*GPAT9/gpat9-2* was crossed with *qrt1-4* homozygous mutant as the pollen donor. F1 seeds were germinated on sulfadiazine to identify *gpat9-2* heterozygotes. F1 plants were selfed, and F2 seeds again selected on sulfadiazine. Pollen from F2 plants was examined visually for the tetrad pollen phenotype to identify those homozygous for the *qrt1-4* mutation, and genotype was confirmed with PCR. Pollen tube germination was measured by dusting pollen from *qrt1* *GPAT9/gpat9-2* and *qrt1* plants onto one of two different optimized pollen germination media [48,49], except that the pH of Boavida and McCormick media was adjusted with 1M KOH instead of NaOH. Pollen was then prehydrated and allowed to germinate for 8 hours at 22°C as described previously [48]. Images of germinated quartets were captured using Leica Application Suite (version 4.2.0) software using a Leica DM2000 microscope equipped with a DFC295 camera. Pollen size was measured using ImageJ software (National Institutes of Health; http://rsb.info.nih.gov/ij). To analyze the small pollen phenotype, tetrads were analyzed only if all four pollen grains were in the plane of focus. The longest apparent diameter (a) and the diameter
perpendicular to the longest diameter (b) of each pollen grain of 100 tetrads was measured, and the apparent 2-D area of the pollen grain was calculated using the formula for an ellipse, \( A = \pi \frac{ab}{2} \).

### 4.2.9 Seed and Leaf Lipid Composition and Quantification

Whole seed (or leaf) lipids were converted to fatty acid methyl esters (FAMEs) in 5% (v/v) sulfuric acid in methanol, to which was added 0.2 ml toluene containing 20 µg tri-17:0 TAG (Nucheck Prep Inc., www.nu-chekprep.com/) as internal standard, for 1.5 h at 85°C. FAMEs were quantified on an Agilent gas chromatograph with flame ionization detection on a wax column (EC wax; 30 m X 0.53 mm i.d. X 1.20 µm; Alltech).

### 4.2.10 Split Ubiquitin-Based Membrane Yeast Two-Hybrid Assay

Protein:protein interactions between AtGPAT9 and other lipid metabolic enzymes, including those representing other steps in the Kennedy pathway, were characterized using the DUALmembrane split-ubiquitin system (Dualsystems Biotech, Schlieren, Switzerland), essentially as described in Gidda et al. [50]. Plasmids were constructed for AtGPAT9, AtLPAAT2 [51], AtDGAT1 [52] and AtLPCAT2 [53]. Primarily, only bait fusion proteins containing Cub-LexA fused to the C-terminus of proteins, and prey fusion proteins containing NubG fused to the N-terminus of proteins were used in this studies to avoid false negative results, as explained in Gidda et al. [50]. Basic analysis for interactions was conducted via dilution assays on solid agar media lacking leucine, histidine, tryptophan and adenine, using multiple serial 1:5 dilutions of cell cultures starting from an OD\textsubscript{600nm} value of 0.5. The strength of specific bait-prey combinations was measured by quantitative \( \beta \)-galactosidase activity from cell lysates using the \( \beta \)-gal assay kit from Pierce Protein Research Products (Thermo Scientific).
4.2.11 Artificial MicroRNA Knockdown of AtGPAT9 Expression

Artificial microRNAs were designed using Web microRNA Designer (http://wmd3.weigelworld.org). AT5G60620.1 (GPAT9) was input as the target gene into the form on the “Designer” section of WMD3 and compared against Arabidopsis thaliana cDNA TAIR8 to get recommendations of amiRNAs. The minimum number of targets were 1, and accepted off-site targets 0. We selected nucleotide region +191-211 (5’-TAGATGTCTAGCAAATCGCGC-3’) for cloning. Using the GPAT9-specific amiRNA sequences, oligonucleotide sequences were generated by WMD3 for site-directed mutagenesis to introduce the sequences of interest in the microRNA precursor gene MIR319a. PCR was conducted using these oligos and pRS300 (MIR319a) as template and following instructions described on WMD3. Using Gateway cloning, the final PCR products were cloned into the pENTR vector and eventually introduced into the pDS-Red-PHAS binary vector under control of the P. vulgaris phaseolin promoter [54]. Both constructs were transformed into Agrobacterium tumefaciens strain GV3101 which was used for transforming Col-0 WT plants.

4.2.12 GPAT Enzyme Activity Assays

GPAT assays were performed with microsomes collected from 9-11 DAF developing seeds, or whole siliques, from wild-type and amiRNA knockdown lines. Microsomes were prepared as described previously [55]. In brief, approximately 100 whole siliques, or 0.1 ml volume of developing seeds dissected from siliques, were homogenized in 0.1 M potassium phosphate buffer pH 7.2, 1% BSA, 1000 units of catalase/mL, 0.33 M sucrose and 1x Thermo Scientific Halt™ Protease Inhibitor Cocktail, EDTA-free (100X) with a Kinematica AG Polytron® PT-MR 2100 on ice. The homogenate was filtered through two layers of buffer soaked Miracloth (EMD Millipore Corporation), and centrifuged at 15,000g for 10 min at 4°C. The
supernatant was centrifuged at 105,000g for 90 min at 4°C. The pellet was rinsed with phosphate buffer pH 7.2 then resuspended in 0.1 ml of phosphate buffer pH 7.2 containing 1000 units of catalase. Microsomes were quantified using Thermo Scientific Pierce™ BCA Protein Assay Kit. Approximately, 0.3-0.6 mg total protein of isolated microsomes were utilized for GPAT assays and performed in 0.1 ml of 0.1 M Tris, pH 7.2, 4 mM MgCl₂, 1% BSA and 1 mM DTT [29], with 3.55 nmol [¹⁴C]-glycerol 3-phosphate ([¹⁴C]G3P, 161mCi/mmol, www.perkinelmer.com) as acyl acceptor and 25 nmol palmitoyl-CoA (Sigma, http://www.sigmaaldrich.com/) as acyl donor at 24°C with constant mixing (1250 rpm) for 15 minutes. The reaction was terminated by adding 0.12 ml of 0.15 M acetic acid with 400 nmol unlabeled G3P and 1.2 ml CHCl₃:MeOH (1:2). Approximately, 25 µg of carrier LPA and PA in 0.4 ml CHCl₃ were added as a carrier and lipids extracted [56]. In brief, the organic layer was collected after phase separation, followed by two back-extractions of the aqueous phase with CHCl₃. The combined chloroform extracts were rinsed with H₂O:MeOH (1:1) to remove any remaining labeled G3P, and evaporated under N₂. The lipids were resuspended in 0.1 ml CHCl₃:MeOH (9:1), total radioactivity in 10 µl was quantified by liquid scintillation counting on a Beckman Coulter™ LS 6500 Multi-Purpose Scintillation Counter, and the remaining extract was loaded onto Merck 20x20 cm silica gel 60 plates developed with solvent system CHCl₃:MeOH:acetic acid:H₂O, 75:15:10:3.5 (v/v/v/v). Radioactivity on TLC plates was quantified with a GE Typhoon FLA 7000, and ImageQuant™ TL Image Analysis Software v8.1.

4.2.13 SOFTWARE AND STATISTICS

Graphs and statistical analysis indicated in each figure were produced with GraphPad Prism (http://www.graphpad.com/).
4.2.14 ACCESSION NUMBERS

Sequence data from this article can be found in The Arabidopsis Information Resource (http://www.arabidopsis.org/) under the following accession numbers: AtGPAT9 (At5g60620), AtLPEAT1 (At1g80950), and AtLPEAT2 (At2g45670). The Genbank accession numbers for RcGPAT9 and VfGPAT9 are ACB30546 and FJ479751, respectively. All other gene identifiers are as described in the species-specific data subsets in the Phytozome database (version 10.3, http://phytozome.jgi.doe.gov/pz/portal.html#). All identifiers are listed individually in Supplemental Figure S2.
4.3 RESULTS

4.3.1 GPAT9 PROTEIN SEQUENCES HAVE BEEN STRONGLY CONSERVED ACROSS PLANT EVOLUTION

 Genome evolution often gives rise to large, complex gene families [57]. Duplicated genes can either be maintained and achieve new function, or lost, causing one of the genes in question to revert to ‘singleton’ status, which often occurs with genes that encode essential functions [58,59]. A single GPAT responsible for the majority of extra-plastidial membrane lipid and TAG biosynthesis would fit this essential housekeeping role, and is supported by digital northern analyses of \textit{AtGPAT9} [32,60], which suggests a ubiquitous expression pattern. From this perspective, we analyzed the evolution of plant \textit{GPAT9} genes, by comparing the predicted protein sequences of GPAT9s from representative species from each major taxonomic grouping of plants. These included \textit{Arabidopsis thaliana} (Brassicaceae), a bryophyte moss (\textit{Physcomitrella patens}), a lycophyte (\textit{Selaginella moellendorffii}), a core eudicot (\textit{Aquilegia coerulea}), a grass (\textit{Brachypodium distachyon}), and representatives from the dicot families Pentapetalae, Malvidae, and Fabidae (\textit{Solanum tuberosum}, \textit{Salix purpurea}, \textit{Gossypium raimondii}, and \textit{Cucumis sativus}). The divergence and speciation events separating these plants cover at least 400 million years of evolution. As shown in Figure 1, GPAT9 proteins from this diverse group of plants have maintained a remarkable level of sequence identity, with approximately 55% identity and 65% similarity overall, with several individual cross-species pairs at >80% identity and >90% similarity.

 To add perspective to these findings, GPAT9 sequences from each of these plants were compared to the corresponding LPEAT1 and LPEAT2 proteins from each species. GPATs and lysophosphatidylethanolamine acyltransferases (LPEATs) utilize the same acyl donor and similar
acyl acceptors (in the case of LPEAT1/2, lysophospholipids including LPA, lysophosphatidylcholine, and lysophosphatidylethanolamine), and the sequences of GPAT9, LPEAT1, and LPEAT2 are homologous themselves [61], making phylogenetic comparisons between them useful. As members of a larger gene family derived at least in part from gene duplication events, LPEAT1 and LPEAT2 display more sequence divergence and are much less likely to encode essential housekeeping functions. LPEAT1 and LPEAT2 sequences from several plant species were aligned and analyzed in a fashion similar to the GPATs (Supplemental Figures S1 and S2, respectively). The input protein sequences for the two LPEAT groups derive from a much less diverse range of species (including sequences from only three monocots and nine dicots) than those in the GPAT9 group shown in Figure 1, yet the sequences within the LPEAT1 and LPEAT2 groups are significantly more diverged, sharing only 34% identity/47% similarity and 31% identity/43% similarity, respectively (Supplementary Figures S1 and S2). The contrast between the relative rates of evolution amongst these three families of genes can be seen directly in Figure 2. GPAT9, LPEAT1, and LPEAT2 protein sequences from the three monocot species (Zea mays, Setaria italica, and Oryza sativa) and nine dicot species (Arabidopsis thaliana, Citrus clementina, Glycine max, Prunus persica, Populus trichocarpa, Ricinus communis, Solanum tuberosum, Theobroma cacao, and Vitis vinifera) were aligned and compared phylogenetically. Each of the three gene families form distinct clades, containing smaller subclades specific to monocots and dicots. However, in all cases the branch lengths for the GPAT9 family are significantly shorter than that of the members of either LPE acyltransferase clade, indicating less evolutionary drift of protein coding sequences in the GPAT9s relative to either the LPEAT1 or LPEAT2 gene families. Significantly higher constraint on sequence divergence is consistent with the hypothesis that plant GPAT9s encode an essential housekeeping function (Figure 2).
4.3.2 A *gpat9* T-DNA Knockout Is Homozygous Lethal

The homology of *AtGPAT9* to GPATs in lipid-rich mammalian tissues alludes to a possible important role in TAG synthesis, but does not directly indicate its function within lipid metabolism. Therefore, we initially set out to investigate if T-DNA insertional mutants of *GPAT9* display an altered glycerolipid phenotype. The Arabidopsis Information Resource (TAIR, https://www.arabidopsis.org/) indicates only two available T-DNA insertion lines targeted to exons in the *GPAT9* locus, and both were obtained. The *gpat9-1* line was a “confirmed” homozygous Salk line (Salk_052947C) putatively containing a T-DNA insertion the first exon [40]. Initial plantings of *gpat9-1* indicated no obvious growth phenotypes, and we confirmed the presence of a homozygous T-DNA insertion near the 5' end of *GPAT9* by PCR of genomic DNA with T-DNA and gene-specific primers (primer sequences are shown in Supplemental Table S1).

The *gpat9-2* (GABI_867A06) line [42] indicated a T-DNA insertion in the fourth exon, and contains a sulfadiazine resistance selectable marker. Germination of *gpat9-2* seeds on sulfadiazine-containing media over several generations produced only heterozygous *GPAT9/gpat9-2* plants as confirmed by segregation of resistant and susceptible seedlings, and by PCR. Once transferred to soil there was no obvious difference in growth of *GPAT9/gpat9-2* plants from wild-type. The quantitation of seed germination and seedling viability on sulfadiazine (Table 1) indicated that at each generation ~6-26% of seeds from *GPAT9/gpat9-2* plants do not germinate, and over 84% of germinated seedlings do not survive on sulfadiazine, and thus are wild-type. All sulfadiazine-resistant seedlings continue to produce sulfadiazine-susceptible offspring, indicating the parental lines were heterozygous *GPAT9/gpat9-2*. Together, the inability to isolate a homozygous line, and the non-Mendelian segregation of sulfadiazine resistance suggests the *gpat9-2* T-DNA insertion line is homozygous lethal.
The contradictory phenotypes of the *gpat9-1* and *gpat9-2* insertion lines necessitated further characterization of each mutation. We identified the actual T-DNA insertion locations by sequencing of the T-DNA insertion regions. PCR products were obtained with T-DNA left border primers paired with both upstream and downstream gene-specific GPAT9 primers, using genomic DNA from both *gpat9-1* and *GPAT9/gpat9-2* plants; the results indicated the presence of multiple linked inverted T-DNAs at the *GPAT9* insertion site in each line. The *gpat9-1* T-DNA insertions were located 46 bp upstream of the *GPAT9* coding sequence start site, and the *gpat9-2* T-DNA insertions were located at the end of exon 4 (Supplemental Figure S3 A). RT-PCR of *GPAT9* mRNA from leaves of *gpat9-1* indicated that full length coding sequence was expressed at similar levels to wild-type (Supplemental Figure S3 B). Quantitative RT-PCR of *GPAT9* mRNA from leaves of *GPAT9/gpat9-2* demonstrated that *GPAT9* expression was reduced to ~50% of wild-type (Supplemental Figure S3 C). Given the minor effect on transcription of *GPAT9* in *gpat9-1*, we concentrated on the sole exon-targeted mutant, *gpat9-2* (GABI_867A06) for the remainder of our studies.

4.3.3 The *GPAT9-2* T-DNA Mutation Produces Pollen Lethality and Partial Female Gametophyte Lethality Phenotypes

The expected Mendelian segregation ratios from a self-fertilized heterozygous parent that produces embryonic lethal mutants is ~67% heterozygous and ~33% wild-type. Single gametophyte lethality mutants are expected to produce 50% heterozygous, and 50% wild-type offspring. That *GPAT9/gpat9-2* plants produced <30% heterozygous seeds and >70% wild-type (Table 1) suggests the *gpat9-2* mutation may have reduced viability of both gametophytes. Therefore, we analyzed the transmittance of the *gpat9-2* mutation (by sulfadiazine resistance)
through reciprocal crosses with wild-type (Table 2). When GPAT9/gpat9-2 flowers were used as the pollen donor with Col-0 pistils as the pollen acceptor no sulfadiazine resistant seedlings were obtained. However, when Col-0 was the pollen donor with GPAT9/gpat9-2 pistils, ~14-22% of the germinated seedlings were resistant to sulfadiazine, similar to the transmittance of sulfadiazine resistance from selfed GPAT9/gpat9-2 in Table 1. These results suggest that gpat9-2 pollen is not viable, and that most gpat9-2 female gametophytes do not survive until fertilization with pollen containing wild-type GPAT9.

In support of a partial female gametophyte lethality phenotype of gpat9-2, aborted ovules were observed in siliques of GPAT9/gpat9-2 but not in Col-0 (Figure 3). Quantitation of the developing seeds and aborted ovules indicated that GPAT9/gpat9-2 siliques contained ~28-43% aborted ovules, with an average of 35% (Table 3). Siliques from heterozygous plants should produce wild-type and mutant ovules at a ratio of 50:50. If 35 of the 50 gpat9-2 ovules abort, the segregation of gpat9-2:WT ovules is 15:50, or ~23% gpat9-2. Considering that no pollen from a selfed GPAT9/gpat9-2 plant can transmit the gpat9-2 mutation (Table 2), then all viable ovules will be fertilized with wild-type pollen, producing ~23% heterozygotes and ~77% wild-type seeds. Therefore the proportion of aborted ovules in GPAT9/gpat9-2 siliques also supports the segregation results obtained from selfing GPAT9/gpat9-2 (Table 1) and the reciprocal crosses (Table 2).

4.3.4 THE GPAT9-2 POLLEN ARE SMALLER AND DO NOT PRODUCE POLLEN TUBES

The reciprocal crosses (Table 2) suggested a male gametophyte defect. To investigate pollen lethality we utilized histochemical staining for pollen abortion and pollen viability with Alexander [46] and NBT [47] staining, respectively (Supplemental Figure S4). To our surprise we
did not detect any quantitiative differences in histochemical staining from pollen of GPAT9/gpat9-2 plants compared to that of wild-type. TAG production in developing pollen is required for fertilization [62]. Therefore, we hypothesized that the gpat9-2 mutation may produce pollen with viable cells, however if TAG production is reduced due to the gpat9-2 mutation it may limit pollen tube growth and the transmittance of the gpat9-2 genotype. To investigate this hypothesis, we crossed the GPAT9/gpat9-2 into the qrt1-4 background, where microspores from a single meiosis event remain attached throughout development [63,64]. The qrt1 mutant background facilitates analysis of pollen defects because a heterozygous parent will produce tetrad pollen containing two affected and two unaffected pollen grains. Tetrads from qrt1 GPAT9/gpat9-2 displayed two larger and two smaller pollen grains (Figure 4A-B), consistent with the segregation of the mutant gpat9-2 allele producing aberrant pollen. While all the pollen from qrt1 parent was close to the same size and followed a Gaussian distribution, the pollen from qrt1 GPAT9/gpat9-2 plants showed a bimodal distribution, with the small pollen roughly half the size of the larger pollen (Figure 4C).

To determine if gpat9-2 pollen can produce pollen tubes, we analyzed tetrad pollen germination (Table 4). In the qrt1 background, 5.5% of ~4800 tetrads produced 3 or 4 pollen tubes (Figure 4A, Table 4). However, when over 6800 tetrads from the qrt1 GPAT9/gpat9-2 background were analyzed, no tetrads with 3 or 4 pollen tubes were observed (Table 4, Figure 4B). These results indicate that the gpat9-2 pollen grains do not produce pollen tubes, and supports the lack of transmittance of gpat9-2 through pollen in the reciprocal crosses (Table 2).
4.3.5 Functional Plant GPAT9 Genes Complement the Gametophytic Lethality Phenotypes of the Heterozygous GPAT9/gpat9-2 Mutant

Wild-type copies of GPAT9 ORFs from Arabidopsis and tung tree (Vernicia fordii) driven by the AtGPAT9 promoter were used to test for complementation of the GPAT9/gpat9-2 phenotypes. We used the AtGPAT9 promoter in an attempt to match the tissue- and cell-type specific transgenic expression of GPAT9s as closely as possible to that of the native gene, instead of using other, more well-known constitutive promoters (e.g. cauliflower mosaic virus 35S) which does not express well in pollen or developing seeds [65,66] and would therefore be unlikely to complement the gpat9-2 phenotypes. Transgenic plants containing the complementation constructs were selected using cassava vein mosaic virus (CVMV) promoter-driven expression of the DsRed fluorescent protein as a selectable marker [67,68].

A population of sulfadiazine-resistant GPAT9/gpat9-2 seedlings were transformed with Agrobacterium bearing either empty binary plasmid (B110), or the complementation plasmids E434 or E437, which contained the ORFs for AtGPAT9 or VfGPAT9, respectively. Very few red seeds from the empty vector B110 line were sulfadiazine resistant (approximately 4%, Figure 5A, D) similar to the segregation analysis of GPAT9/gpat9-2 in Table 1. However, the homozygous red fluorescent seeds from the E434 (Figure 5B, 5E) and E437 (Figure 5C, 5F) lines effectively germinated and established photosynthetic competency on sulfadiazine agar media, at about 60-70% of all seeds sown. Therefore, the complementation constructs clearly function to complement the rare occurrence of sulfadiazine-resistant GPAT9/gpat9-2 progeny (Table 1). This result confirms that the AtGPAT9 gene is essential and responsible for the observed segregation and gametophytic lethality phenotypes in mutant plants. Additionally, GPAT9 genes from multiple plant species can complement the Arabidopsis mutant lethality phenotype.
4.3.6 Seed-specific Knockdown of GPAT9 Reduces Oil Content and Alters Fatty Acid Composition

Our analysis of GPAT9/gpat9-2 suggests that AtGPAT9 is an essential gene and a homozygous gpat9 knockout cannot be obtained. Analysis of lipid content in heterozygous GPAT9/gpat9-2 plants indicated no change from wild-type in leaves (Supplemental Figure S5A), or seeds (Supplemental Figure S5B-C). The latter was as expected since most seeds produced from GPAT9/gpat9-2 are wild-type (Table 1). Therefore, a single copy of GPAT9 is sufficient for vegetative growth, and a different approach must be taken to investigate the role of GPAT9 in lipid metabolism. A strong, constitutive knockdown of GPAT9 expression would be expected to severely reduce plant growth, complicating characterization of a lipid phenotype associated with aberrant GPAT9 expression. Therefore, we chose to investigate the role of GPAT9 in TAG biosynthesis in seeds (which has high flux through ER GPAT activity compared to other tissues) by creating seed-specific GPAT9 knockdown Arabidopsis lines. An artificial microRNA (amiRNA) construct for GPAT9 was expressed under the seed specific Phaseolus vulgaris phaseolin promoter [54], with the fluorescent protein DsRed as a selectable marker [68]. Heterozygous transformed (red) T1 seeds were planted, and the corresponding segregating T2 seeds were harvested from 20 individual T1 transformants. Red T2 seeds were separated from the segregating untransformed brown seeds, and each set analyzed for oil content (Figure 6). Oil content varies considerably between individual Arabidopsis plants [69], and the co-segregating non transformed seed acts as a plant specific wild-type control for seed oil levels. Red seeds from each individual transformed line had a 15-75% reduction in total lipid relative to the plant-specific brown seeds (Figure 6A). In the lines with the largest reduction in oil content, the red seeds were
smaller and displayed a wrinkled phenotype compared to the corresponding brown seeds (Figure 4 B-C). A wrinkled seed phenotype is typical for Arabidopsis seeds with very low oil content [70].

Eight \textit{GPAT9} amiRNA lines with low oil content were identified to have single T-DNA inserts based on 3:1 segregation ratios of red and brown seeds, and were propagated further. Homozygous red T\textsuperscript{3} seeds were analyzed for seed oil content and composition (Figure 7). Oil content was reduced by 26-44\% (Figure 7A). Seed lipids from all eight amiRNA lines also contained altered fatty acid composition (Supplemental Figure S6). The FA composition of two \textit{GPAT9} amiRNA knockdown lines chosen for further experiments (#2 and #12) is demonstrated in Figure 7B. Significant changes in all lines included an average increase in 18:3 from 19.8\% to 28.9\%, and in 20:3 from 0.5\% to 0.9\%. Correspondingly, 20:1 was decreased from an average of 20.6\% to 13.1\%, and 18:2 was decreased from 26.6\% to 24.3\%. Analysis of \textit{GPAT9} gene expression by qPCR in developing T\textsubscript{4} seeds from lines #2 and #12 indicated that \textit{GPAT9} transcript is reduced >88\% from wild-type in each line (Supplemental Figure S7). Together these results indicate that reduced \textit{GPAT9} expression impacts the amount and composition of TAG in seeds.

4.3.7 \textbf{SEEDS FROM GPAT9 KNOCKDOWN LINES HAVE REDUCED GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE ACTIVITY}

The sequence of AtGPAT9 closely resembles mammalian GPATs that contribute to TAG synthesis [31]. However, AtGPAT9 displays high sequence homology to multiple classes of plant lysophospholipid acyltransferases (including LPEAT1 and LPEAT2 which also use LPA as a substrate, Figure 2). Additionally, plant GPAT9s in general more closely match plant LPAATs in protein length (~375-390 amino acid residues for both enzyme classes) [71] than they do the land plant specific \textit{sn}-2 GPAT family (500-585 residues). Reduction of either GPAT or LPAAT activity in the amiRNA knockdown lines has the possibility to affect seed oil content. Therefore to
determine which enzymatic activity is affected in the knockdown lines, we performed in vitro GPAT and LPAAT assays from microsomes isolated from developing siliques of wild-type and homozygous T4 GPAT9 amiRNA lines (Figures 8 and 9). Figure 8A demonstrates that silique microsome GPAT assays utilizing [\(^{14}\)C]G3P and palmitoyl-CoA as substrates produce LPA and PA as the major products. For both wild-type and amiRNA lines, [\(^{14}\)C]G3P distribution between LPA and PA was approximately 3:7 LPA/PA at 5-7 days after flowering (DAF), and 7:3 LPA/PA at 9-11 DAF (Fig. 8A). However, the total amount of [\(^{14}\)C]G3P incorporated into lipids differed between the wild-type and knockdown lines. [\(^{14}\)C]G3P is incorporated into PA only after LPA has been formed therefore the quantification of both LPA and PA represents total GPAT activity. Total GPAT activity was reduced ~19-49% at the two different stages of silique development (Fig. 8B-C). To confirm that the reduced GPAT activity measured from whole silique microsomes is due to the seed specific knockdown of ER GPAT activity in GPAT9 amiRNA lines, we also performed GPAT assays on microsomes isolated from developing seeds dissected out of 9-11 DAF siliques (Supplemental Figure S8). Similar to the whole silique results, the amiRNA lines had ~25-55% reduction in seed microsome GPAT activity from wild-type.

LPAAT assays were performed with the same set of whole silique microsomes as for GPAT assays (Figure 8) utilizing LPA and [\(^{14}\)C]oleoyl-CoA as substrates. The [\(^{14}\)C]oleoyl-CoA can also be used by other acyltransferases and endogenous lipid acceptors found within the microsomes. However, since only LPA was added as an exogenous acyl acceptor, PA was the major product of the assay in all samples (Figure 9A-B). No differences in LPAAT activity between wild-type and the GPAT9 amiRNA lines were detected (Figure 9C-D). The reduction of GPAT enzymatic activity in seeds of GPAT9 amiRNA knockdown lines (Figures 8, S8), and the corresponding lack of a change in LPAAT activity (Figure 9), together supports the conclusion
that the reduced oil accumulation in these transgenic lines (Figures 6, 7) is due to a reduction in seed GPAT activity, specifically caused by the \textit{GPAT9} mRNA knockdown.

\textbf{4.3.8 \textit{AtGPAT9} Physically Interacts With Other Enzymes in the Kennedy Pathway and Acyl Editing Cycle}

Enzymes that make up various biochemical pathways, including some lipid biosynthetic pathways, often form multi-component complexes [72,73]. Such complexes allow for rapid and efficient transfer of metabolites to downstream enzymes in the pathway. Given the possibility of metabolic networking between the enzymes of the Kennedy pathway to assemble glycerolipids, and the enzymes of acyl editing which provides polyunsaturated fatty acids for lipid assembly [11,18,74], we sought to explore potential protein:protein interactions between \textit{AtGPAT9} and other potential partners using the split-ubiquitin yeast two hybrid (Y2H) system [75]. The split-ubiquitin system eliminates the need for importation of target enzymes and proteins into the nucleus of the yeast cell as in traditional Y2H systems [76], which is especially problematic for integral membrane proteins such as GPAT9 and other ER-localized acyltransferases. This system previously has been used to search for interacting partners of glycerolipid synthesis enzymes from tung tree [50]. As shown in Figure 10, \textit{AtGPAT9} interacted with itself, \textit{AtLPAAT2}, and \textit{AtLPCAT2}, but not with \textit{AtDGAT1}, the dominant TAG biosynthetic DGAT isozyme in Arabidopsis [52]. \textit{AtDGAT1} did weakly interact with both \textit{AtLPAAT2} and \textit{AtLPCAT2} however, as shown in Supplementary Figure S9. The homomeric and heteromeric interactions between \textit{AtGPAT9} and \textit{AtLPAAT2} or \textit{AtLPCAT2} were ~8-20-fold stronger than the corresponding interactions with \textit{AtDGAT1}. Together these results suggest that \textit{AtGPAT9} likely interacts \textit{in vivo} with \textit{AtLPAAT2}, the next step in glycerolipid assembly after GPAT; and with \textit{AtLPCAT2}, the
main enzyme involved in acyl editing which provides polyunsaturated FAs to the acyl-CoA pool for incorporation into de novo glycerolipid synthesis [74,77-81].
4.4 DISCUSSION

The genetic dissection of the Kennedy pathway with the ultimate goal for complete biotechnological control plant oil synthesis began in earnest more than 15 years ago, with the identification of the genes that encode what are now known as \textit{AtDGAT1} [52,82] and \textit{AtDGAT2} [83]. Meaningful progress has been made towards identification and functional characterization of these and many other enzymes that contribute to plant membrane phospholipid and storage triacylglycerol biosynthesis in the years since. However, despite such progress on many fronts, definitive isolation of the gene or genes that encode for the initial G3P acylation reaction that feeds both the ER localized membrane lipid and TAG biosynthesis pathways has eluded the plant lipid biotechnology community. Evidence is presented here that firmly supports the assignment of \textit{AtGPAT9} to that role.

4.4.1 \textbf{AtGPAT9 is the GPAT involved in TAG biosynthesis}

The results presented here establish that AtGPAT9 is a single copy and essential gene, demonstrated by both male and female gametophyte lethality phenotypes of the \textit{gpat9-2} mutant, and the inability to obtain a homozygous mutant (Figures 1-5, Tables 1-4). The reduction in oil content of seed specific \textit{GPAT9} knockdowns (Figures 6, 7), and the corresponding reduction in GPAT activity (Figures 8), but not LPAAT activity (Figure 9), demonstrates that AtGPAT9 is a glycerol-3-phosphate acyltransferase involved in seed oil accumulation. This result could not be demonstrated with any of the genes in the \textit{AtGPAT1-AtGPAT8} family, which further supports their \textit{vastly different} roles within plant metabolism. Since \textit{GPAT9} is essential for gametophyte function, the GPAT activity involved in seed TAG biosynthesis is likely also required for ER membrane lipid synthesis in other tissues.
4.4.2 \textit{AtGPAT9} Knockdown Seed Oil Phenotype, and Protein Interactors Helps to Further Define the Plant Oil Synthesis Metabolic Network

Protein:protein interaction analysis (Figure 10) provided insights that strongly suggest the proper placement of \textit{AtGPAT9} at an intersection between the ‘Kennedy’ \textit{de novo} glycerolipid biosynthetic pathway, and the acyl editing cycle, supporting current models of the Arabidopsis lipid biosynthetic network (Figure 11A) [11,18,74]. In the current study, \textit{AtGPAT9} was found to interact with itself, \textit{AtLPAAT2}, and \textit{AtLPCAT2}, but not \textit{AtDGAT1}. How each interaction fits into current models of TAG biosynthesis is discussed below.

Self-interaction of \textit{AtGPAT9} is consistent with discovery of oligomerization of yeast-expressed \textit{Erysimum asperum} plastidial GPAT [84]. The self-associating properties of \textit{EaGPAT} may suggest that it undergoes self-allosteric regulation, since many allosteric enzymes exhibit quaternary structure. Future studies will address possible correlations between homomeric assembly of \textit{AtGPAT9} subunits and allosteric regulation of GPAT activity. The interaction between \textit{AtGPAT9} and \textit{AtLPAAT2} (the activity expected to occur immediately downstream of GPAT in the lipid biosynthetic pathway) suggests metabolic channeling between Kennedy pathway enzymes within lipid biosynthesis (Figure 11), and supports previous studies that found lipid biosynthetic enzymes co-localized to specific microdomains of the ER [50,85].

\textit{LPCAT} works in both the forward and reverse directions [80,81]. The interaction of \textit{AtGPAT9} with \textit{AtLPCAT2} fits with current metabolic models from \textit{in vivo} lipid flux analysis which indicates that the acyl groups removed from PC by acyl editing are the major source of acyl-CoA for \textit{de novo} glycerolipid synthesis by Kennedy pathway enzymes [11,86-88]. The combined loss of \textit{AtLPCAT1} and \textit{AtLPCAT2} stops acyl editing in Arabidopsis seeds, and causes a dramatic shift in acyl flux such that newly synthesized FAs are directly incorporated into \textit{de novo}
glycerolipid synthesis through GPAT and LPAAT rather than direct incorporation into PC through acyl editing as in the wild-type [77,78]. The interaction of AtGPAT9 and AtLPCAT2 demonstrated here supports that nascent acyl groups destined for AtLPCAT2-mediated acyl editing in wild-type can be readily shifted to AtGPAT9 within the lpcat1/2 double mutant for enhanced de novo glycerolipid synthesis.

The lack of interaction between AtGPAT9 and AtDGAT1 importantly fits with current models of TAG biosynthesis from in vivo labeling studies that indicate a kinetic separation of de novo glycerolipid synthesis and TAG synthesis. [14C]glycerol kinetic labeling in Arabidopsis and soybean indicates that DAG synthesized de novo from the GPAT, LPAAT, and PA phosphatase activities of the Kennedy pathway is not directly utilized for TAG synthesis. Instead de novo DAG is utilized to synthesize PC, and DAG for TAG synthesis is later derived from PC [87,89]. This flux of DAG through PC, combined with the acyl editing cycle, enhances the residence time of acyl groups in PC for production of polyunsaturated FAs [11,86]. In Arabidopsis the fluxes of de novo DAG into PC, and PC-derived DAG out of PC are predominantly controlled by phosphatidylecholine:diacylglycerol cholinephosphotransferase (PDCT) [90]. Since PDCT activity does not involve the net synthesis or turnover of PC, only a headgroup exchange reaction, a reduction in glycerol flux into the TAG biosynthetic pathway by reduced GPAT activity should not affect the rate of PDCT action (Figure 11).

AtLPCAT2 interacted with both AtGPAT9 and AtDGAT1 (Figures 10, S9). LPCAT activity has been demonstrated within multiple cellular localizations [91,92]. Together these results suggest that there may be multiple sites for acyl editing within the cell (Figure 11), and strongly suggest that both the first and last glycerol acylation reactions may be fed at least in part with acyl-CoA derived directly from PC by AtLPCAT2. These results are also consistent with
recent co-expression and biochemical analyses using flax (*Linum usitatissimum*) DGAT1 and LPCAT [79]. These authors showed that LPCAT likely mediates direct channeling of PC-derived acyl-CoA to DGAT1, thus accounting for the elevated polyunsaturated FA content in flax oil, while also helping to overcome the thermodynamically unfavorable reverse reaction of LPCAT [79]. The results demonstrated here (Figures 10, S9) strongly suggest that a high degree of channeling also occurs in Arabidopsis (and likely many other plants as well), with at least two entry points for PC-derived acyl-CoA into glycerolipid assembly (Figure 11).

Finally, a consideration of TAG biosynthesis as a metabolic network involving multiple highly active exchange reactions in/out of PC that are independent of total acyl flux through the network (Figure 11, large black arrows), can be used to explain how reduced GPAT activity leads to increased levels of polyunsaturated FAs in the seed oil of *GPAT9* amiRNA knockdown lines (Figure 6B, S6). Acyl editing involves a cycle of PC deacylation and LPC reacylation by LPCAT enzymes in which the rate of acyl exchange on/off PC can be up to 15 times the rate of *de novo* glycerolipid synthesis [74,87,88]. Likewise PDCT activity can be independent of the total rate of DAG and PC synthesis and turnover, leading to rapid exchange of DAG in/out of PC. The reduction of GPAT activity in *GPAT9* amiRNA lines slows the total rate of glycerol flux (thus DAG flux) through the lipid metabolic network into oil (Figure 11B, smaller purple arrows). However, the flux through the acyl and headgroup exchange reactions with PC are independent of the total rate glycerolipid biosynthesis, allowing FAs and DAGs to continually cycle in/out of PC, enhancing the total residence time in PC prior to incorporation into TAG. The longer time an acyl group spends within the PC pool allows for greater access to the desaturases, and the higher the probability that oleate will be fully desaturated to linolenate prior to incorporation into TAG [11,86]. Therefore, a slower rate of TAG assembly, combined with no change in the exchange
reactions explains the significant increase in 18:3 content of TAG in the GPAT9 amiRNA knockdown lines.

4.4.3 Strict Maintenance of Single Copy GPAT9 Genes in Plants Reinforces Their Essentiality

Plant genomes have been subjected to enormous evolutionary pressures over time, including whole- or partial-genome duplications (GD) [93], and small-scale duplication events [57] which often result in sets of duplicated genes that help an organism to meet new metabolic requirements. Some duplicated genes are retained and ultimately achieve specialized function via changes in temporal or tissue/organ-specific gene expression profile, subcellular targeting, enzyme substrate specificity, etc. [59]. However, duplicated genes from genome duplication events can revert back to singleton status due to pressures from factors such as functional redundancy, epigenetic silencing, and chromosomal instability [58]. Many such ‘singleton’ genes have been found to encode for essential housekeeping functions [58,59], and the same appears to be true for AtGPAT9.

The various functional analyses shown here reveal that AtGPAT9 (and likely, by extension, GPAT9 genes from other plants in general), is not genetically redundant (Figures 1-5, Tables 1-4), unlike DGATs, LPAATs, and several other classes of plant lipid metabolic enzymes [18]. Most plants contain a single copy of GPAT9, or at least a single copy per diploid genome, and appear to have been subject to selection pressures that maintain the respective GPAT9s genes as singletons [58,59]. Searches across a wide array of sequenced plant genomes, covering many of the major branch points in plant evolution, found a very high number of single-copy GPAT9 genes (Figure 1). Pima cotton (Gossypium raimondii) is the only one of nine diverse species shown in Figure 1 that contained two copies of GPAT9, as per the data available in Phytozome 10.3. The ancestor to
G. raimondii and other Gossypium sp. underwent a cotton-specific whole genome duplication event ~16-17 MYA [94]; the existing tetraploid status of cotton, and the relative recency of this duplication event may not yet have allowed for enough selection pressure, or provided enough evolutionary time for resolution of the ultimate fate of the second copy of GraGPAT9.

GPAT9s are also highly conserved across broader sections of the tree of life, suggesting a role in essential cellular metabolism. The homology between AtGPAT9 and mouse GPAT3 (which contains minimal sequence identity to any of the other eight known Arabidopsis extra-plastidial GPATs) was one of the initial indicators that AtGPAT9 might be an actual sn-1 glycerol-3-phosphate acyltransferase [31,32]. Also unlike the larger family of sn-2 GPATs, AtGPAT9 is highly conserved across Animalia in general. AtGPAT9 is closely related to representative proteins from hundreds of animal species, the closest of which is shown in Supplementary Figure S10. A GPAT from the orca whale (Orcinus orca) shares 39% identity and 58% similarity with the amino acid sequence of AtGPAT9 over a 360 amino acid region of the protein (nearly the full length of AtGPAT9). In contrast, the top Animalia BLASTP hit for AtGPAT1 is an uncharacterized protein from moonfish (Xiphophorus maculates) which is only 27% identical and 42% similar, over a 256 residue portion of the protein (less than half of the AtGPAT1 protein). Conversely, O. orca does not possess a GPAT1, while X. maculates GPAT9 still retains 36% identity and 54% amino acid similarity to AtGPAT9 (Supplementary Figure S10).

Many of the genes in the ‘duplication-resistant’ and ‘mostly single-copy’ categories described by De Smet et al. [95] encode essential housekeeping functions. That AtGPAT9 is essential was proven by our inability to recover a homozygous gpat9-2 mutant plant, even after germination of heterozygous mutant seeds on solid media containing vitamins and sugars. The abnormally low recovery of sulfadiazine-resistant progeny from heterozygous gpat9-2 parents also
indicates that AtGPAT9 mutants are gametophytic lethal. The incomplete gametophytic lethality of the gpat9-2 female gamete may arise from minimally sufficient transfer of residual GPAT9 protein and GPAT9 mRNA between cells during cell division to maintain viability of a few eggs until fertilization with wild-type pollen, as has been demonstrated for other genes [96]. The complementation of the low penetrance of the GPAT9/gpat9-2 heterozygous phenotype with both transgenic Arabidopsis and tung tree GPAT9 constructs (Figure 5 and Supplementary Figure S5) proved that the gpat9-2 mutation was indeed responsible for the observed mutant phenotypes, and that somewhat distantly related plant GPAT9s can functionally complement for one another. This latter result may be key to finally controlling the FA composition at each position within the TAG backbone by isolating acyl selective GPAT enzymes from different species.
4.5 CONCLUSIONS

Taken together, the results presented here confirm that \textit{AtGPAT9} is an ancient gene that is essential in Arabidopsis, and likely other plants as well. Experimental evidence suggests that GPAT9 fulfills an indispensable role in catalysis of the first step in the synthesis of storage and membrane lipids required for life. This role may help to explain why \textit{GPAT9} genes have been maintained throughout most of the course of evolution of life on earth and why the sequences of GPAT9s have evolved so conservatively. Finally, the characterization of GPAT9 at the intersection of \textit{de novo} glycerolipid synthesis and acyl editing, and our demonstration that a distantly related plant GPAT9 can replace the essential \textit{AtGPAT9}, suggests that bioengineering strategies around GPAT9 may be valuable for production of the designer vegetable oils of the future.
Figure 1. Sequence Comparison of Selected GPAT9 Enzymes from Plants.

Amino acid sequences from Aquilegia coerulea, Arabidopsis thaliana, Brachypodium distachyon, Cucumis sativus, Gossypium raimondii, Physcomitrella patens, Salix purpurea, Selaginella moellendorffii, and Solanum tuberosum were aligned using the ClustalX algorithm [33]. Amino acids identical in all ten proteins are shaded in black, similar residues are shaded in grey. Name abbreviations indicate the first letter of the genus, and the first two letters of the species.
Protein sequences were aligned using ClustalX (version 1.8.1, [33]). An unrooted phylogenetic tree was created from the alignment, using TREEVIEW (version 1.6.6, http://taxonomy.zoology.gla.ac.uk/rod/rod.html, Page 1996). The branch lengths are proportional to the degree of divergence, with the scale of “0.1” representing 10% change.
Figure 3. Aborted Ovules in GPAT9/gpat9-2 Siliques

White arrows indicate aborted ovules in GPAT9-2/gpat9-2 siliques. The Col-0 silique did not contain aborted ovules.
Table 1. Non-Mendelian segregation of GPAT9/gpat9-2 under sulfadiazine selection

<table>
<thead>
<tr>
<th>Generation</th>
<th>Seeds sown</th>
<th>Germinated seeds</th>
<th>Sulfadiazine resistant</th>
<th>% resistant of germinated</th>
<th>% not germinated</th>
<th>F4 ave. ± S.D.</th>
<th>F5 ave. ± S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F3</td>
<td>n.d.</td>
<td>425</td>
<td>66</td>
<td>15.5</td>
<td>n.d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4 (F3-1)</td>
<td>376</td>
<td>276</td>
<td>35</td>
<td>12.7</td>
<td>26.6</td>
<td>11.4 ± 2.3</td>
<td>23.1 ± 2.5</td>
</tr>
<tr>
<td>F4 (F3-2)</td>
<td>126</td>
<td>100</td>
<td>8</td>
<td>8.0</td>
<td>20.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4 (F3-3)</td>
<td>285</td>
<td>221</td>
<td>26</td>
<td>11.8</td>
<td>22.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4 (F3-4)</td>
<td>327</td>
<td>253</td>
<td>33</td>
<td>13.0</td>
<td>22.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5 (F3-1-1)</td>
<td>255</td>
<td>228</td>
<td>26</td>
<td>11.4</td>
<td>10.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5 (F3-2-1)</td>
<td>207</td>
<td>194</td>
<td>17</td>
<td>8.8</td>
<td>6.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5 (F3-3-1)</td>
<td>242</td>
<td>223</td>
<td>19</td>
<td>8.5</td>
<td>7.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5 (F3-4-1)</td>
<td>223</td>
<td>203</td>
<td>26</td>
<td>12.8</td>
<td>9.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5 ave. ± S.D.</td>
<td></td>
<td></td>
<td>10.4 ± 2.1</td>
<td>8.4 ± 1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Transmittance of sulfadiazine resistance from reciprocal crosses of GPAT9/gpat9-2 and Col-0

<table>
<thead>
<tr>
<th>Cross</th>
<th>Male gametophyte</th>
<th>Female gametophyte</th>
<th>Germinated seedlings</th>
<th>Sulfadiazine resistant</th>
<th>% resistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Col-0</td>
<td>GPAT9-2/gpat9-2</td>
<td>90</td>
<td>20</td>
<td>22.2</td>
</tr>
<tr>
<td>1-2</td>
<td>GPAT9-2/gpat9-2</td>
<td>Col-0</td>
<td>113</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2-2</td>
<td>Col-0</td>
<td>GPAT9-2/gpat9-2</td>
<td>139</td>
<td>20</td>
<td>14.4</td>
</tr>
<tr>
<td>2-2</td>
<td>GPAT9-2/gpat9-2</td>
<td>Col-0</td>
<td>242</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3. Ovule viability within *GPAT9/gpat9-2* half siliques

<table>
<thead>
<tr>
<th>Plant – Silique</th>
<th>Developing seed</th>
<th>Aborted ovules</th>
<th>% aborted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>21</td>
<td>9</td>
<td>30.0</td>
</tr>
<tr>
<td>1-2</td>
<td>21</td>
<td>8</td>
<td>27.6</td>
</tr>
<tr>
<td>1-3</td>
<td>18</td>
<td>11</td>
<td>37.9</td>
</tr>
<tr>
<td>1-4</td>
<td>20</td>
<td>11</td>
<td>35.5</td>
</tr>
<tr>
<td>1-5</td>
<td>20</td>
<td>14</td>
<td>41.2</td>
</tr>
<tr>
<td>2-1</td>
<td>22</td>
<td>14</td>
<td>38.9</td>
</tr>
<tr>
<td>2-2</td>
<td>20</td>
<td>15</td>
<td>42.9</td>
</tr>
<tr>
<td>2-3</td>
<td>23</td>
<td>10</td>
<td>30.3</td>
</tr>
<tr>
<td>2-4</td>
<td>20</td>
<td>9</td>
<td>31.0</td>
</tr>
<tr>
<td>2-5</td>
<td>21</td>
<td>13</td>
<td>38.2</td>
</tr>
</tbody>
</table>

Ave. ± S.D. 20.6 ± 1.3 11.4 ± 2.5 35.4 ± 5.3
Figure 4. Reduced Pollen Size and Pollen Tube Growth of GPAT9-2/gpat9-2 Pollen in the qrt1 Background.

(A) Tetrads pollen in the qrt1 mutant can germinate to produce up to four pollen tubes, one from each attached pollen grain.

(B) Tetrads pollen in the qrt1 GPAT9-2/gpat9-2 line does not produce more than two pollen tubes per tetrad.

(C) Size of individual tetrad pollen grains from qrt1 and qrt1 GPAT9-2/gpat9-2 plants. Pollen grain length and width was measured to calculate the apparent surfact area of each pollen grain.
Table 4. Pollen germination and tube growth of GPAT9/gapt9-2 within the qrt1 background.

<table>
<thead>
<tr>
<th>qrt1</th>
<th>Pollen tetrads</th>
<th>3-4 tubes/tetrad</th>
<th>Pollen tetrads</th>
<th>3-4 tubes/tetrad</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>1b</td>
<td>1285</td>
<td>52</td>
<td>4</td>
<td>1256</td>
</tr>
<tr>
<td>2b</td>
<td>694</td>
<td>15</td>
<td>2.2</td>
<td>1380</td>
</tr>
<tr>
<td>3c</td>
<td>991</td>
<td>52</td>
<td>5.2</td>
<td>1830</td>
</tr>
<tr>
<td>4c</td>
<td>608</td>
<td>108</td>
<td>17.8</td>
<td>836</td>
</tr>
<tr>
<td>5c</td>
<td>1282</td>
<td>41</td>
<td>3.2</td>
<td>1543</td>
</tr>
</tbody>
</table>

a Predicted tube growth based on percent growth in qrt1 during that experiment. Germination medias: b Zhu et al., [49]; c Boavida & McCormick [48].
Figure 5. Genetic Complementation of gpat9-2 Mutants with Transgenic Plant GPAT9.

Hemizygous gpat9-2 plants were transformed with different binary vectors, with or without functional full-length plant ORFs driven by the AtGPAT9 promoter. T4 generation seeds from DsRed-fluorescent and sulfadiazine-resistant representative lines from: negative control B110 (A, D); AtGPAT9prom::AtGPAT9 (pE434, B, E); or AtGPAT9prom::VfGPAT9 (pE437, C, F) were sown on solid agar media containing sulfadiazine. Wide-angle views of the agar plates are shown in the top panels, close views of the same plates are shown in the bottom panels. Note the high proportion of stunted and aborted seedlings on the negative control plate compared to the high number of healthy, complemented seedlings established on both the E434 and E437 plates.
Figure 6. Segregating Red and Brown T	extsubscript{2} GPAT9-amiRNA Seed Oil Content and Size.

(A) Oil content of segregating brown (not transformed, black bar) and red (transformed, red bar) T	extsubscript{2} seed from 20 individual T	extsubscript{1} plant lines. Twenty red or brown seeds were used per analysis. (B) Line 7 seed under white light.

(C) Same seed as (B) under green light with red filter.

(B-C) The red GPAT9 amiRNA seed are smaller than the brown untransformed seed.
**Figure 7.** Oil Quantity and Composition of Homozygous T₃ GPAT9 amiRNA Lines.

(A) Distribution of whole seed FAME content. Each sample contained 50 seeds from an individual plant. Red bar indicates average.

(B) Fatty acid composition of seeds from part A. Lines Col-0, and amiRNA knockdowns 12 and 2. Average ± S.D.
Figure 8. Developing Silique Microsome GPAT Activity.

GPAT assays with microsomes isolated from whole developing siliques 5-7 and 9-11 days after flowering (DAF), from Col-0 and GPAT9 amiRNA knockdown lines 2 and 12. The assay utilized 3.55 nmol \([^{14}\text{C}]\)glycerol 3-phosphate and 25 nmol palmitoyl-CoA, for 15 min at 24 °C.

(A) Phosphor image of TLC plate indicating that \([^{14}\text{C}]\)LPA and \([^{14}\text{C}]\)PA are the major products of whole silique microsome GPAT assays utilizing palmitoyl-CoA as an acyl donor.

(B, C) Quantification of total products from GPAT activity in whole silique microsomes.
Figure 9. Silique Microsome LPAAT Assays.

LPAAT assays with microsomes isolated from whole developing siliques of 5-7 (left side) and 9-11 (right side) days after flowering (DAF), from Col-0 and GPAT9 amiRNA knockdown lines 2 and 12. The assay utilized the same conditions as the GPAT assays in the manuscript text except 3.33 nmol $^{14}$Coleoyl-CoA was the acyl donor and 25 nmol 18:1-LPA was used as acyl acceptor for 15 min at 24°C. The $^{14}$Coleoyl-CoA can also be used by other acyltransferases and endogenous lipid acceptors found within the microsomes. However PA was the major product of the assay in all samples.

(A-B) Phosphor image of TLC separation of products.

(C-D) Quantitation of PA within each lane of the TLC image.
Coding sequences for each of two different membrane proteins of interest are ligated in-frame to either the C-terminal half of ubiquitin (Cub)-LexA transcription factor protein fusion or the N-terminal half of ubiquitin (Nub). Nub may be represented as either native polypeptide sequence (NubI), or a mutant form of Nub containing an isoleucine/glycine conversion point mutation (NubG). NubI strongly interacts with Cub. NubG has very weak affinity for Cub, and must be brought into close proximity to Cub to allow for interaction of the two halves of ubiquitin, release of the LexA transcription factor and finally, activation of the reporter genes (histidine and adenine prototrophic markers and β-galactosidase, for quantitative analysis).

(A-B) Prototrophic growth assay of yeast strains containing various combinations of AtGPAT9 bait plasmid co-expressed with potential prey NubG-acyltransferase plasmids. Serial dilutions of cells expressing different bait-prey combinations were plated on non-selective (A) or selective (B) media conditions.

(C) Quantitative measurement of β-galactosidase activity from cell lysates of the strains used in the serial dilution assays.

Figure 10. Testing of Protein:Protein Interactions Between AtGPAT9 and Other Lipid Acyltransferases.
Figure 11. Models of GPAT9-dependent Glycerol Flux, and Spatial Organization of TAG Biosynthesis.

(A) Wild-type TAG biosynthesis is dependent on the flux of G3P through GPAT9 for the initial incorporation of the glycerol backbone into glycerolipids. GPAT9, LPAAT2, and LPCAT2 are localized together for efficient flux of acyl groups out of PC into de novo glycerolipid synthesis. De novo DAG (DAG1) is utilized to produce PC, and eventually PC-derived DAG (DAG2) is incorporated into TAG. DGAT1 is spatially separated from GPAT9 and de novo DAG, and is associated with PC-derived DAG. The amount of 18:2/18:3 in TAG is dependent on the residence time of acyl groups in PC for desaturation vs the flux out of PC for incorporation into TAG.

(B) In GPAT9 amiRNA lines the flux of G3P into glycerolipid synthesis is reduced (smaller purple arrows), lowering total TAG accumulation. However, the rate of acyl exchange, and DAG exchange into/out of PC is not changed (large black arrows). Therefore, the residence time of acyl groups in PC for desaturation increases, leading to higher overall levels of polyunsaturated fatty acids in the knockdown lines.

(A-B) The differences in arrow width between A and B indicate changes in flux.
4.6 REFERENCES


6. Ichihara Ki: sn-Glycerol-3-phosphate acyltransferase in a particulate fraction from maturing safflower seeds. *Archives of biochemistry and biophysics* 1984, **232**:685-698.


13. Frentzen M, Heinz E, McKeon TA, Stumpf PK: Specificities and Selectivities of Glycerol-3-Phosphate Acyltransferase and Monoacylglycerol-3-Phosphate Acyltransferase


84. Chen X, Miles R, Snyder C, Truksa M, Zhang J, Shah S, Weselake RJ: **Possible allostery and oligomerization of recombinant plastidial sn-glycerol-3-phosphate acyltransferase.** *Archives of biochemistry and biophysics* 2014, **554**:55-64.

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


89. Bates PD, Browse J: **The pathway of triacylglycerol synthesis through phosphatidylcholine in Arabidopsis produces a bottleneck for the accumulation of unusual fatty acids in transgenic seeds.** *The Plant Journal* 2011, **68**:387-399.


4.7 SUPPLEMENTAL FIGURES AND METHODS

Supplementary Figure S1. Sequence Alignment and Identity/Similarity Shading of Plant LPEAT1 Proteins. Phylogenetic tree of this alignment makes up part of Figure 2 in the main body of the text.
<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TcLPEAT1</td>
<td>383</td>
<td>QS---</td>
<td></td>
</tr>
<tr>
<td>VvLPEAT1</td>
<td>373</td>
<td>QR---</td>
<td></td>
</tr>
<tr>
<td>GmLPEAT1</td>
<td>374</td>
<td>KDE</td>
<td></td>
</tr>
<tr>
<td>PpLPEAT1</td>
<td>396</td>
<td>KDD</td>
<td></td>
</tr>
<tr>
<td>CcLPEAT1</td>
<td>393</td>
<td>KDD</td>
<td></td>
</tr>
<tr>
<td>PtLPEAT1</td>
<td>399</td>
<td>KDD</td>
<td></td>
</tr>
<tr>
<td>RcLPEAT1</td>
<td></td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>StLPEAT1</td>
<td>392</td>
<td>QQ---</td>
<td></td>
</tr>
<tr>
<td>AtLPEAT1</td>
<td>395</td>
<td>KEE</td>
<td></td>
</tr>
<tr>
<td>OsLPEAT1</td>
<td>401</td>
<td>KDD</td>
<td></td>
</tr>
<tr>
<td>SiLPEAT1</td>
<td>391</td>
<td>KDD</td>
<td></td>
</tr>
<tr>
<td>ZmLPEAT1</td>
<td>370</td>
<td>KDD</td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Figure S2. Sequence Alignment and Identity/Similarity Shading of Various Plant LPEAT2 Proteins. Phylogenetic tree depiction of this alignment makes up part of Figure 2 in the main body of the text.

PpLPEAT2 42 S-----------------NSTHHHR--NPAGSGD--FTVPGSTTADPGFRL
VvLPEAT2 45 T-----------------KSDHTOPESNDPEFSGAG--FSVPGETVDPERNPD
PtLPEAT2 36 N-----------------NHKPNNTSSRNPPEFGSGD--LSVPAPSTLDPERNDP
RcLPEAT2 36 H-----------------NSNSSSNG-LRNPPEFGSDF--LSVPSPSTVPERNPD
CcLPEAT2 49 N-----------------NHDTRYHNPNPWFSGD--LSVPGPNATDLNP
TcLPEAT2 43 SSLQTGIPQNHQNYGNGNQGQSSRNPPEFGSGD--FSVPAPTTIDPERNPD
AtLPEAT2 43 DP-----------------RVSRGFEFHLNPQFSESE--PPVLGPTVDPERNPD
GmLPEAT2 34 PRN-----------------PRNPVRGTDDDDDSLSPSSSTLDPERRTP
StLPEAT2 50 D-----------------HLGTHISEVDPNPAPAGNRR--FDMPGTSVDPERNPD
SlLPEAT2 42 AA-----------------AVSVCDGGG-DPAPFSEDRTPF----QDRGSPPAPFRGP
ZmLPEAT2 44 AA-----------------AVSVCDGGG-DPAPFSEDRTPF----RDRGSPPAPFRGP
OsLPEAT2 42 SP-----------------TVCGDGGDGGDFPAPFSEDRAW--WSPRGVSPPAPFRGP

PpLPEAT2 78 EIRD--LEYELILGIICPLALRVLFTASLIDVAVKLAIDDWGK--------KPPMP
VvLPEAT2 83 KDIP--LEYELILGIIIPLAIARVLFTLCPVAALKADGNGK--------QPPMP
PtLPEAT2 75 DJEL-LYLELIIIVCLPAALRVLFTLCVALKADGNGK--------HPPMP
RcLPEAT2 74 KIEL--LYPELIIIVCLPAALRVLFTACVALKADGNGK--------NPPMP
CcLPEAT2 88 HVVS--VEEFYIVCPVLARVLFTFCLVALKADGNGK--------QPPMP
TcLPEAT2 98 FVSE--VEEVIIVCLPAALRVLFTFCVALKADGNGK--------QPPMP
AtLPEAT2 83 GVSF--LPEAVLICPLPAALRVLFTASACVALKADGNGK--------EPPMP
GmLPEAT2 68 AIEL--LEYAATTVCLPAALRVLFTALCPVAATKLAIDGWGK--------EPPMP
StLPEAT2 89 RESE--LEYELIIIIVCLPAALRVLFTLCVALKLAIDGWGK--------SPPMP
SlLPEAT2 82 ANGPGVAWATLILAPAAVRLFPLAATPAVAIARSPQDRP--REGACPMP
ZmLPEAT2 94 ANGPGVAWATLILAPAAVRLFPLAATPAVAIARSPQDRP--RIVGACPMP
OsLPEAT2 95 GWNL--LEYELIIIIVCLPAALRVLFTLCVALKLAIDGWGK--------MPPMP
Figure S3. Characterization of GPAT9 T-DNA Insertions

(A) Representation of GPAT9 T-DNA insertion locations. For both lines T-DNA left border primers will generate amplicons with gene specific primers flanking either side of the insertion suggesting there are multiple insertions with inverse orientation. The gpat9-1 insertions are 46 bp upstream of the coding sequence. The gpat9-2 insertion is at the end of the 4th exon. Pictured is just the first four exons out of 12 in AtGPAT9.

(B) RT-PCR of full length GPAT9 coding sequence from leaves of homozygous gpat9-1.

(C) qPCR results for GPAT9 wild-type (Col-0) and GPAT9/gpat9-2 leaves.

Methods:

(A) For insertion site sequencing the left border-gene junction was amplified by PCR with GoTaq® Flexi (Promega), gel purified by QIAquick Gel Extraction Kit (QIAGEN) and sequenced from both the left border primer and a flanking gene-specific primer by Eurofins (http://www.eurofinsgenomics.com/).

(B-C) Leaf RNA extraction with RNeasy Plant Mini Kit (QIAGEN) and treated with RNase-free DNase (Qiagen) and on-column DNase digestion. RNA was quantified using a Nanophotometer (Implen Inc). cDNA was synthesized using SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies). RT-PCR utilized standard conditions as per GoTaq® Flexi (Promega) protocol with 30 cycles of amplification. For qPCR transcript levels were analyzed by quantitative PCR (qPCR) using Platinum SYBR Green qPCR SuperMix-UDG (Life Technologies) and the Stratagene Mx3005P Quantitative PCR system (Agilent Technologies Inc). Transcript was normalized to the TIP41-like gene [1].

(A-C), the sequences for primers utilized are shown in supplemental table S1.
Figure S4. Analysis of Viability in Wild-Type and GPAT9/gpat9-2 Pollen.

(A-B) Alexander staining [2] for pollen abortion. Purple = non-aborted pollen, green pollen (black arrows) = aborted pollen. Pollen from wild-type (A), and GPAT9/gpat9-2 (B) plants.

(C-D) Nitro blue tetrazolium staining for pollen viability [3]. Viable pollen stain black. Anthers from wild-type (C), and GPAT9/gpat9-2 (D) plants.
Figure S5. Lipid Characterization of Wild-Type and GPAT9-2/gpat9-2 Plants.

(A) The relative fatty acid composition of leaves. Average and standard deviation. \( n = 5 \) for wild-type (Col-0), and 13 for GPAT9-2/gpat9-2.

(B) Total seed fatty acid content from seeds of plants grown at the same time under 24 hr light ~150-170 \( \mu \text{mol/m}^2/\text{sec} \) white light. 30 seeds per FAME analysis from a single plant. \( n = 19-20 \) individual plants per line.

(C) The average fatty acid composition of seeds from part C. Bars are average and standard deviation.
Supplemental Figure S6. Seed Fatty Acid Composition of \textit{GPAT9} amiRNA Knockdown Lines.

Fatty acid compositions are from seed samples analyzed for total oil content in Figure 7A of the main text. Average ± S.D.
**Supplemental Figure S7.** Expression of GPAT9 in Developing Seeds of Wild-Type and GPAT9 amiRNA Knockdown Lines.

*GPAT9* transcript was normalized to TIP41-like [1]. Two biological replicates for Col-0, and three biological replicates for each amiRNA knockdown line. Each biological replicate is the average of three technical replicates.

**Methods:**

Frozen developing seeds (9-11 DAF) were separated from silique tissue by silique popping method [4]. ~50 µL volume of frozen seeds per biological replicates were mechanically pulverized to a fine powder with steel beads (TissueLyser LT, QIAGEN). RNA was extracted and used for cDNA synthesis using SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies). RT-PCR utilized standard conditions as per Maxima SYBR Green/ROX qPCR Master Mix (Thermo) and BioRad CFX96™ Real-Time System and C1000™ Thermal Cycler with 45 cycles of amplification. The sequences for primers utilized are shown in supplemental table S1.
Supplemental Figure S8. Developing Seed Microsome GPAT Assays of Wild-Type and amiRNA Knockdown Lines.

GPAT assays with developing 9-11 DAF seed microsomes utilizing 3.55 nmol [14C]glycerol 3-phosphate and 25 nmol palmitoyl-CoA, for 15 min at 24°C.

(A) Phosphor image of TLC plate indicating that [14C]LPA was the major product of developing seed microsome activity with palmitoyl-CoA.

(B) Percent of wild-type (Col-0) GPAT activity in GPAT9 amiRNA knockdown lines 2 and 12. Mean and S.D. of two replicate assays.
Supplementary Figure S9. Testing of Protein:Protein Interactions Between AtLPAAT2, AtDGAT1, and AtLPCAT2.

Testing of Protein:Protein Interactions Between AtLPAAT2, AtDGAT1, and AtLPCAT2. Cub-LexA bait fusions and Nub prey fusions were produced as described in Figure 4 and in the main text. (A-B) Prototrophic growth assay of yeast strains containing various combinations of Arabidopsis acyltransferase bait plasmids co-expressed with potential NubG-fusion acyltransferase prey plasmids. Serial dilutions of cells expressing different bait-prey combinations were plated on non-selective (A) or selective (B) media. (C) Quantitative measurement of β-galactosidase activity from cell lysates of the strains used in the serial dilution assays. Cultures used for serial dilutions shown in A. and B. contained bait and prey combinations as numbered in C.
Supplementary Figure S10. Sequence Identity Comparisons Between AtGPAT1 and AtGPAT9 to the Animalia Subset of the NCBI Protein Database.

The top hits to AtGPAT9 and AtGPAT1 are shown first, followed by the best hits from within the individual species that provided the top hits to the alternate AtGPAT sequence. See text for additional details.

>AtGPAT9
MSSTAGRLVTSKSELDDLHPNIEDYLPSSGSINEPRGKLSLRLDLDISPTLTEAAGAIVDSSFTRCFKSNPPEPNWY
NIYLFLPLYCFGIVVRYCILFPLRCFTLAFLGWGFFILSFIPVNLALKQGDRLRK\KI\RLVLEMICFFVAVSNTGVKYY
HGP\PSIRPKQVYVANHTSMIDFIVLQE\MAF\AVIM\RKHGP\VGLQQSTILES\GCI\NF\RSEAKDREIVAKKL\RHD\V
QV\ADS\NP\LLIFPEGTCV\NYTV\MFKGAF\LCTVCP\AIAKN\KFV\DF\N\RSRQ\Q\S\TM\HLQL\MT\N\AVCEV
WYLEP\Q\TIP\GETGIEF\E\R\MDISLR\RL\LKKV\PD\GYL\Y\RS\SPK\H\E\RK\Q\Q\S\AE\LIRL\EEK

vs Metazoa/Animalia:
PREDICTED: glycerol-3-phosphate acyltransferase 3 [Orcinus orca]
Sequence ID: ref|XP_004282203.1|
Length: 438
Number of Matches: 1
Gene-associated gene details
Map Viewer-aligned genomic context
Range 1: 75 to 429
Alignment statistics for match #1
Score Expect Method Identities Positives Gaps Frame
259 2e- Compositional 139/360(39%) 211/360(58%) 17/360(4%)
Features:
Query 24  DYLPSGS\SINEPRGK-LSLRDLDDISPTLTEAA\GAI\VDSSFTRCFKSNPPEPNW----N 78
  D P ++ RG+ L D+ S EA IV+D T+ F S WN N
  DESPEMKGLSGLRGRDFE45DVFYFSKKGLEA---IVEDEVTQ\PSSEELV\N\LTRIN 131
Query 79 I-\YLF\-\LYC\CG\VV\RCIL\LPRC\TLAFGWGFFILSFIPV\NLALKQGD\RLK+ 131
  Y+ P ++ GV+Y+L+ + L + LR +
  VNFQ\YS\PKLTM\VV\GL\V\R\Y+V\CVL\L\LVR+TLAFG\ISF\L\V\G\LQ\L-ESSL\RNR 189
Query 132 IER\VLVLE\MC\SSF\F\V\A\SW\T\G\V\K\H\P\R\P\R\P\K\Q\V\Y\V\A\H\T\S\M\ID\F\V\L\E\Q\T\F\A\V\IM\Q+
  + ++ \V C + + G + YH + + + V\N\HTS ID + + L + A++ Q
  LSELVHLT\CC\R\C\RLG\ST\GH\Y\N\K\R\Y\P\K\G\IC\V\A\N\HT\S\P\D\VL\LTD\G\C\Y\A\MV\Q 249
Query 192 KHPGWG\L\QQ\S\T\L\E\SV\G\CI\W\N\ER\SE\K\DK\R\D\V\Q\G\A\D\S\P\L\L\IF\PE\G\TCV\NV\N G+G+Q +++S +WF RSE KDR +V K+L R+H+ P+LIFPEGTC+NN
  VHGHGLMG\I\MQ\AM\RL\F\M\D\K\R\L\K\R\E\H\I\AD\KK\L\L\P\L\L\IF\PE\G\TC\NN 309
Query 252 NY\TM\FKGAF\L\CT\V\CP\AIAKN\KFV\DF\N\RSRQ\Q\S\TM\H\L\Q\L\MT\SA\V\C\E\V\WYL +MF\KK+FE+ T+ P+AI\YN F DAF\N S K + +L+MTS\WA+V+C+VYW+ TSV\MKF\K\G\S\F\E\G\T\G\T\Y\P\AI\YN\P\Q\F\G\DF\N\SSK\KN\M\V\Y\L\R\M\TS\WA\V\CD\V\YNW 369
Query 310 EPQ\TIP\GET\GIE\E\F\R\MDISLR\RL\LKKV\PD\GYL\Y\RS\SPK\H\E\RK\Q\Q\S\AE\LIRL\EEK P T GE +++FA RV+ I+++ GL ++P\D\G\K+ ++ E +Q+++++ I+
  PPM\TRE\EG\D\A\V\Q\FA\N\V\K\SA\AI\IQ\G\G\T\L\E\P\D\G\G\L\K\RA\V\K\DT\F\E\E\QQ\KN\YS\K\MV\G 429
AtGPAT1

MVPPELLVILAEMVYLLAKSCYRAARKLRGYGFQLKNLLSLSKTQSLTQSLHNNHQHONQNLQDSLDPLFPSLTKYQELLKRNACSVSDDHYRDTFFCDCIDGVLLRQHSSKHHTFFFPYFMLVAEGGSIIRAILLLLSCSWLWTLQETKLRLVSLFITFSLGRLVXDMNVSRLVPLKFLENLNIQYWIARTEYKSVTVTPLQVLFEXREHLNADVIGTKGQVIEKVMRKFYTGLASGSFVLKHKSAEDYFFDSKKKPALGSSSSPQH6FISICKEAYFWNEEEMS

KNNALPREYPKPLIFHDGRALFPTPLATLAMFILWPgLAVFRISVGFLPYHVAHFALASMGVRIFTKTHN

NNGRPEKGNVLVYLVCHNRTLDPFVLFLTSSLGKPLTAVTYSLSKFSFIEFLKTVSXLRKDREKDEAMQRLLSGDLVVCPEGTTTCREPYLRFSPFAELTED1IVPVAVDARSVMFYGTTASGLKCLKDPIFLMNFPVYCLEIKLKLKPENMTCAAGKSSFEVANFIQGELARCLFGFCNNLIRRRDKYLVNHAGKVR

vs Metazoa/Animalia:
PREDICTED: ancient ubiquitous protein 1-like isoform X1 [Xiphophorus maculatus]

Sequence ID: ref|XP_005811174.1| Length: 415 Number of Matches: 1

Gene-associated gene details
Map Viewer-aligned genomic context
Range 1: 20 to 262

<table>
<thead>
<tr>
<th>Score</th>
<th>Expect Method</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>59.3</td>
<td>3e-</td>
<td>69/256(27%)</td>
<td>109/256(42%)</td>
<td>31/256(12%)</td>
<td></td>
</tr>
</tbody>
</table>

Features:
Query 336  
LATLAMFILWPgLAVFRISVGV------------------------FLPHVANFIASMGVRIFTKTH 383  
+A L + I + P + G L + R I + G V F + V E G + + S + G + + + 
Sbjct 20  
VALLLLLIYSPVGVCLMLMRIGFHVHSSCAIPDSFVRVRVFVRMSSVLGMHVRQR- 77

Query 384  
NLNNGRPEKGNVLVYLVCHNRTLDP--VFLTTSLGKPLTAVTYSLSKFSFIEFLKTVS 441  
N R R V C N H T D + L T P E + + + S 
Sbjct 78  
---NPRSRDKNTK-LCVCNHTEFDHHINILLTPFNTQPLEGSTGFMCMWARFMMEHSAS 133

Query 442  
LKRDKREKDEAMQRLLSGD----LVVCPEGTTTCREPYLRFSPFAELTED1IVPVAVDA 497  
++ G E + + + QU S L V E T T L L + F S L T E I P VA + 
Sbjct 134  
---SQEAVGESLQRYCSADGAPPLLVFEEDTTNARGLKLKSSWPFTLTESIQPVALRV 190

Query 498  
RVSMFYGTASGLKCLKDPIFLMNFPVYCLEIKLKLKPENMTCAAGKSSFEVANFIQGEL 557  
+ T D + + P V V + L + + G + + + E A N + Q L 
Sbjct 191  
SRPILLSLSTEVSSWLLLWMTFFSPCTVYHVSWLPPVSQ---DGESTQEFANKVQELL 246

Query 558  
ARVLGFECTNLIRDDK 573  
A LG + T + T + D K 
Sbjct 247  
AIELGLVSTKITKADK 262
**Supplemental Table S1.** Primers utilized for PCR, RT-PCT, & qPCR.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Primer Name</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genotyping gpat9-1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Salk_052947C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene specific GPAT9 salk F1</td>
<td>ACTTTCTTTGGAGAGTCGCTGT</td>
<td></td>
</tr>
<tr>
<td>Gene specific GPAT9 salk R1</td>
<td>CATCAACAATGGCACCAGCA</td>
<td></td>
</tr>
<tr>
<td>T-DNA left border LBb1.3</td>
<td>ATTTTGCCGATTTCGGAAC</td>
<td></td>
</tr>
<tr>
<td><strong>Genotyping gpat9-2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GABI_867A06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene specific GPAT9 GK F1</td>
<td>GCAGGCCTGATTTTGCTAGAC</td>
<td></td>
</tr>
<tr>
<td>Gene specific GPAT9 GK R1</td>
<td>AATGCGGTCATCTGCTCCAA</td>
<td></td>
</tr>
<tr>
<td>T-DNA left border o8409</td>
<td>ATATTGACCATCATACTCATTC</td>
<td></td>
</tr>
<tr>
<td><strong>Genotyping qrt1-4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CS25041)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene specific Qrt F2</td>
<td>TTTGCCTCTCAAGGTTTTC</td>
<td></td>
</tr>
<tr>
<td>Gene specific Qrt R1</td>
<td>AATGAGGCTATGCGGGAGA</td>
<td></td>
</tr>
<tr>
<td>T-DNA left border LBb1.3</td>
<td>ATTTTGCCGATTTCGGAAC</td>
<td></td>
</tr>
<tr>
<td><strong>GPAT9 RT-PCR (Fig. S1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPAT9 ORF start</td>
<td>ATGAGCAGTACGGCAGGGAGGC</td>
<td></td>
</tr>
<tr>
<td>GPAT9 ORF end</td>
<td>TGCAAAAGTGAGTTATGTTATGAGA</td>
<td></td>
</tr>
<tr>
<td><strong>qPCR of GPAT9 (Fig. S1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPAT9LP1046</td>
<td>AGAGAGGTCAGAGACATGA</td>
<td></td>
</tr>
<tr>
<td>G9 qPCR R 1172</td>
<td>CGACTCTGCAGAAACCTGAGT</td>
<td></td>
</tr>
<tr>
<td>qTIP41-like829F</td>
<td>ACATTTCAGTCTCTATCTGCGAAAGG</td>
<td></td>
</tr>
<tr>
<td>qTIP41-like946R</td>
<td>GGATCTTCAGTTTCTGTGTGCATG</td>
<td></td>
</tr>
</tbody>
</table>

173
Supplemental Table S2. Gene identifiers for GPAT9, LPEAT1, and LPEAT2 sequences used in Figures 1 and 2 in the main text. AGI locus identifiers for the three Arabidopsis genes are described at The Arabidopsis Information Resource (TAIR) homepage (https://www.arabidopsis.org/). The Genbank accession numbers are shown for tung tree GPAT9 (*Vernicia fordii*, Gidda et al., 2009) and *Ricinus communis* GPAT9 [originally annotated as a putative LPAAT, Burgal et al., (2008)]. All other identifiers are as listed for each species in the Phytozome 10.3 database (http://phytozome.jgi.doe.gov/pz/portal.html#).

*: not analyzed in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>GPAT9</th>
<th>LPEAT1</th>
<th>LPEAT2</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>At5g60620</td>
<td>At1g80950</td>
<td>At2g45670</td>
</tr>
<tr>
<td><em>Vernicia fordii</em></td>
<td>FJ479751</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Physcomitrella patens</em></td>
<td>Phpat.011G104300</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Selaginella moellendorfii</em></td>
<td>152980</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Aquilegia coerulea</em></td>
<td>Aquca_047_00038.1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Brachypodium distachyon</em></td>
<td>Bradi1g25790</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Solanium tuberosum</em></td>
<td>PGSC0003DMT400032021</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Salix purpurea</em></td>
<td>SapurV1A.0039s0700</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Gossypium raimondii</em></td>
<td>Gorai.012G158300 and Gorai.005G258500</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Cucumis sativus</em></td>
<td>Cucsa.099550</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>GRMZM2G165681</td>
<td>GRMZM2G096010</td>
<td>GRMZM2G116243</td>
</tr>
<tr>
<td><em>Setaria italica</em></td>
<td>Si030220m</td>
<td>Si022261m</td>
<td>Si006175m</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>LOC_Os07g34730</td>
<td>LOC_Os05g28960</td>
<td>LOC_Os06g49790</td>
</tr>
<tr>
<td><em>Citrus clementina</em></td>
<td>Ciclev10028668m</td>
<td>Ciclev10001396m</td>
<td>Ciclev10019536m</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>Glyma.09G119200</td>
<td>Glyma.04G033600</td>
<td>Glyma.03G019200</td>
</tr>
<tr>
<td><em>Prunus persica</em></td>
<td>ppa007262m</td>
<td>ppa006689m</td>
<td>ppa003784m</td>
</tr>
<tr>
<td><em>Populus trichocarpa</em></td>
<td>Potri.004G183300</td>
<td>Potri.014G042200</td>
<td>Potri.014G074300</td>
</tr>
<tr>
<td><em>Ricinus communis</em></td>
<td>ACB30546</td>
<td>30174.m008937</td>
<td>30170.m014002</td>
</tr>
<tr>
<td><em>Theobroma cacao</em></td>
<td>Thecc1EG006479t1</td>
<td>Thecc1EG033997t1</td>
<td>Thecc1EG005455t1</td>
</tr>
<tr>
<td><em>Vitis vinifera</em></td>
<td>GSVIVT01023935001</td>
<td>GSVIVT01009661001</td>
<td>GSVIVT01027580001</td>
</tr>
</tbody>
</table>
4.8 Supplemental References


CHAPTER FIVE:
CONCLUSIONS AND PERSPECTIVES

Triacylglycerol (TAG) is one of the most energy-rich molecules made by plants, and it serves as an important source of calories and essential fatty acids in the human diet. TAG also serves many non-biological roles for humans, and the long hydrocarbons derived from the fatty acids in TAG are of particular interest for fuels, chemical feedstocks, and additives to many consumer products. The bulk of oil produced by most plants, including Arabidopsis, is made in the seed, where is serves the plant as a source of energy for the germinating seed or an attractant for seed dispersers.

The importance of oils both for plant biology and for human uses has led to over 60 years of study, and the Kennedy Pathway of TAG synthesis has driven our knowledge of fat and oil synthesis ever since. It describes how glycerol-3-phosphate, derived from the glycolysis intermediate dihydroxyacetone phosphate, is acylated at the \( sn-1 \) and then \( sn-2 \) positions to produce first lysophosphatidic acid (LPA) and then phosphatidic acid (PA). The dephosphorylation of phosphatidic acid produces the nonpolar lipid diacylglycerol (DAG), which is then acylated at the \( sn-3 \) position to produce TAG. The acyl groups are provided to the enzymes as acyl-CoA, exported from the plastid after fatty acid synthesis or cleaved from phosphatidyl choline (PC) during the acyl editing cycle. Today we know that the Kennedy Pathway acts slightly differently in Arabidopsis than it does in mammals and some plants. Flux to TAG actually goes through PC, the main membrane lipid in the ER and the site of acyl editing.

The fatty acid chains esterified to the three positions of the glycerol backbone are the main contributor to the properties of the oil. For example, a high proportion of saturated fatty acids
encourages a solid state at room temperature, while unsaturated fatty acids lead to a more liquid state. Monounsaturated and omega-3 fatty acids are desirable in food oils because of the health benefits to humans, while polyunsaturates such as α-linolenic acid can be used as drying oils in paints and varnishes. These fatty acids are modified in the acyl editing cycle by elongation, desaturation, or specialized modifications (e.g. hydroxylation, conjugation, etc.), and then esterified to the glycerol backbone using acyltransferases. The fatty acid composition of the oil (and therefore its properties) is dependent on the overall composition of the fatty acids in the acyl-CoA pool as well as the specificity of the acyltransferases.

The importance of acyltransferase specificity has been demonstrated in the attempts to produce ricinoleic acid in Arabidopsis. Castor bean (Ricinus communis) accumulates up to 90% of its fatty acids in seed oil as hydroxylated fatty acids (HFAs) such as ricinoleic acid, which are useful industrially in the manufacturing of products such as lubricants, paints, plastics, and soaps. The castor plant is a poor agronomic crop, and so attempts have been made to produce the fatty acid in other oilseed plants, starting with Arabidopsis. When the FAH12 enzyme, responsible for generating ricinoleic acid, was heterologously expressed in Arabidopsis, only 17% of the fatty acids in the seed oil were hydroxy-fatty acids [1]. However, when castor isoforms of diacylglycerol acyltransferase (DGAT, RcDGAT2), and phospholipid:diacylglycerol acyltransferase (PDAT, RcPDAT1A), were expressed in the transgenic Arabidopsis, HFA levels rose to 20-30% and 27%, respectively [2,3]. Additionally, while the oil of castor accumulates up to 90% HFA, there is almost no HFA in the membranes of the seed. A castor phospholipase A2, RcsPLA2α, appears to be involved in specifically cleaving HFA from PC [4], removing it from membrane lipids and encouraging incorporation into TAG. These examples indicate that acyl-selective enzymes do control the fatty acid content, and therefore the properties, of seed oil.
As our society attempts to become more sustainable and rely less on non-renewable resources, we have been working to produce plant oils that meet the needs of a growing population. We aim to produce food oils that are healthier, with higher levels of monounsaturated and omega-3 fatty acids and lower levels of saturated fatty acids, and both without the need for chemical hydrogenation which generates unhealthy trans fats. We are working to produce industrially- and commercially-useful oils that could be used in the generation of plastics, coatings, and dyes, in a renewable, domestically grown crop, while preserving land for food crop growth. Finally, while trying to replace all petroleum products with plant-grown fuels would be impossible due to the scale needed, certain niche markets of fuel or fuel additives could benefit from production in plants. But before any of these goals can become reality, we must understand the steps generating oil. Identifying and characterizing the genes and enzymes involved in the biosynthesis of TAG is the first step to being able to generate healthier, more renewable oils for food, fuel, and chemical feedstocks.

**GPAT9 is an Essential Gene and Shows GPAT Activity**

The first step of lipid biosynthesis is catalyzed by a glycerol-3-phosphate acyltransferase (GPAT), which transfers an acyl group from acyl-CoA to the sn-1 position of glycerol-3-phosphate. While eight GPAT homologs are involved in suberin or cutin synthesis, the identity of the main GPAT involved in membrane lipid and oil synthesis has not yet been determined. GPAT9 is homologous to known mammalian and yeast GPATs which are involved in generating TAG, and indeed it appears to be highly conserved throughout evolution, a hallmark of genes with essential housekeeping function. The gpat9 mutant must be maintained as a heterozygote, and it shows no transmission of the mutant allele through the male germ line and reduced transmission in the female germ line. This lethality is consistent with an essential role in producing the
membranes and oil necessary for sustaining life. In amiRNA knockdown lines, only lines with high GPAT9 transcript levels accumulated WT oil levels, and transcript level was directly related to the level of GPAT activity in developing seeds. This indicates that GPAT9 encodes GPAT activity, and is the main GPAT responsible for oil accumulation in the seed. The GPAT9 protein also interacts with other ER glycerolipid biosynthetic enzymes, indicating that the proteins form complexes to funnel substrate into glycerolipids. Together, these results indicate that GPAT9 is the main GPAT involved in ER lipid biosynthesis, and it plays an essential role in the generation of membrane and storage lipids.

**LPPβ MAY BE A PAP INVOLVED IN OIL SYNTHESIS**

The phosphatidic acid phosphatase (PAP) step is the third step in the TAG synthesis pathway, and catalyzes the hydrolysis of inorganic phosphate from phosphatidic acid (PA) to produce DAG. Several PAPs have been identified in the prokaryotic pathway of lipid synthesis in chloroplasts, and several others have been implicated in signaling pathways; however the main PAP involved in oil synthesis in the ER remains elusive. After analyzing all the genes in Arabidopsis for homologs to known PAPs, only one homolog displayed characteristics consistent with the PAP involved in oil synthesis. LPPβ is expressed in developing seed tissue and is localized to the ER, probably as an integral membrane protein. Attempts to decrease LPPβ transcript using amiRNA and ihpRNA resulted in increased or no change in transcript level, and attempts to mutate the gene using the CRISPR/Cas9 system have so far been ineffective. However, LPPβ shows PAP activity when expressed in yeast. The pah1pah2 knockout line showed increased LPPβ transcript level, indicating that it may act redundantly with PAH1 and PAH2 in oil synthesis. More studies, particularly genetic analysis of mutants, are still needed to confirm its role in the plant.
**SPP1 and LPAP1 Are Not Involved in Oil Synthesis**

In Chapter 2, I discussed my rationale in determining which PAP homolog(s) could possibly be involved in oil synthesis. Of the 13 genes in Arabidopsis homologous to known PAPs, most were set aside due to involvement in signaling pathways or the plastidial pathway of lipid synthesis. While \( LPβ \) was the only PAP that could potentially be the main PAP involved in oil synthesis, \( SPP1 \) and \( LPAP1 \) also showed enough characteristics to warrant determining if they were involved in oil synthesis through redundancy. Single mutants showed no oil phenotype, and multiple mutants with each other and \( pah1pah2 \) also showed no oil phenotype, and only minor changes in fatty acid composition. This indicates that they are not involved in oil synthesis, and is consistent with their placement in the signaling pathway for ABA-mediated dehydration stress response in the case of \( SPP1 \) [5], and in the case of \( LPAP1 \), as a lysophosphatidic acid phosphatase which is highly expressed during stress conditions [6]. Since it is unusual for PAPs which are involved in signaling to also be involved in \textit{de novo} lipid synthesis, it is not unexpected to find that neither gene is involved in oil synthesis.

**The State of Knowledge of TAG Synthesis Genes**

As described in Chapter Four, GPAT9 is the main GPAT involved in oil synthesis, and the first step of oil and membrane lipid synthesis. It provides LPA for the lysophosphatidic acid acyltransferase (LPAAT) enzyme. There are five LPAATs in Arabidopsis, but \( LPAT2 \) encodes the ubiquitous, ER-localized LPAAT [7]. It has a female-gametophyte lethality phenotype, due to the fact that the pollen is complemented by the \( LPAT3 \) isoform, which is mainly expressed in that tissue. This lethality phenotype suggests an essential function similar to what is seen in \textit{GPAT9}. Recombinant \( LPAT2 \) and \( LPAT3 \) both showed LPAAT activity \textit{in vitro}, but \( LPAT4 \) and \( LPAT5 \)
did not. This indicates that LPAT2 is the main LPAAT involved in membrane lipid and oil synthesis, and LPAT3 is an active LPAAT in the pollen.

The next enzyme in the pathway is PAP, discussed in Chapters Two and Three. There are 13 PAPs in Arabidopsis that are homologous to known PAPs, and all have been dismissed from consideration as the PAP involved in oil synthesis, except LPPβ, which encodes PAP activity and is expressed in seed tissue. However, work remains to characterize mutants and place it in the oil synthesis pathway, and there remains the possibility that a PAP with no homology to known PAPs could be encoding this step.

After the PAP reaction, the oil synthesis pathway in Arabidopsis diverges from the traditional Kennedy pathway, and DAG used for TAG synthesis is actually derived from PC [8]. There are two distinct DAG pools: a de novo pool resulting from PAP activity and one derived from PC. Thus, de novo DAG is converted to PC probably through headgroup exchange catalyzed by the ROD1 gene encoding a phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT) enzyme [9], or possibly through the action of CDP-choline:DAG cholinephosphotransferase (CPT) encoded by AAAPT1 and/or AAAPT2 [10]. Conversion of PC back to DAG could be achieved by PDCT or CPT. Additionally, a non-specific phospholipase C (nsPLC;), or a combination of a phospholipase D (PLDζ) and a PAP could encode the conversion of PC to DAG; however, PLC and PLD enzymes are normally found in signaling pathways and not de novo lipid synthesis, and of the 6 nsPLC’s and 2 PLDζ’s in Arabidopsis, none have been implicated in oil synthesis. Thus, the understanding of which enzyme contributes most to this flux remains to be explored, but genes have been identified for both of the enzymatic steps in generating PC then DAG for oil synthesis.
This second DAG pool can be converted to TAG by one of two enzymatic functions: the acylation of DAG by diacylglycerol acyltransferase (DGAT) using acyl-CoA as a substrate, or the acylation of DAG using PC as the substrate by phosphatidylcholine:diacylglycerol acyltransferase (PDAT). A double homozygous mutant of \textit{pdat1} and \textit{dgat1} is lethal in Arabidopsis, indicating that these may be the main isoforms, but combinations of other DGATs and PDATs have yet to be studied [11]. There are two more DGATs, \textit{DGAT2} and \textit{DGAT3}: \textit{DGAT2} has even higher activity than \textit{DGAT1} does, and different substrate specificity, indicating that it could still contribute strongly to TAG synthesis [12]. The role of \textit{DGAT3} is less clear, but it appears to be involved in oil synthesis in sucrose-rescued seedlings [13]. \textit{PDAT2} has not been specifically characterized, but its expression is strongest in developing seed (compared to \textit{PDAT1} which is expressed constitutively), indicating that it may also be involved in oil synthesis.

This also highlights the growing body of data that indicates that plant oil synthesis does not contain a single Kennedy pathway generating oil as is seen in mammals. Instead the first portion of the pathway (GPAT, LPAAT, PAP) are likely involved in \textit{de novo} lipid synthesis and provide substrate to the PC pool; DAG is then derived from the PC pool specifically for oil synthesis, and possibly for export to the plastid. This is supported not only by flux analysis indicating two pools of DAG, but also by interaction data indicating that GPAT and DGAT do not interact, but enzymes of the membrane synthesis (GPAT, LPAAT, and LPCAT, which provides substrate), do interact. This is not to say that the early steps are not involved in oil synthesis; indeed, decreased levels of GPAT9 show reduced oil. However, the idea of a linear pathway from G3P to oil appears to be evolving in Arabidopsis to a pathway of membrane lipid synthesis followed by the generation of TAG.
Many of the genes involved in oil synthesis in Arabidopsis have now been identified. GPAT, LPAAT, DGAT, and PDAT enzymatic activities in oil synthesis have all been assigned to at least one gene. The PAP step has one gene with several promising indicators of its function. The identification of the genes involved in this pathway will allow further characterization and manipulation of the pathway, which will eventually lead to more valuable, healthier, more sustainable plant oils.
Figure 1: TAG biosynthesis in the eukaryotic pathway of Arabidopsis. Recent data suggests that GPAT, LPAAT, and LPCAT proteins interact, and DGAT and LPCAT interact, but GPAT and DGAT do not interact, indicating that the pathway can be roughly partitioned into two portions of the ER lipid synthesis pathway.
REFERENCES


