Identifying the bottlenecks limiting medium-chain fatty acid accumulation in transgenic pennycress seeds

by
Fnu Alisha

B.Tech., Amity University, India, 2018
M.Tech., Harcourt Butler Technical University, India, 2020

A THESIS

submitted in partial fulfillment of the requirements for the degree
MASTER OF SCIENCE

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences
KANSAS STATE UNIVERSITY
Manhattan, Kansas
2025

Approved by:
Major Professor
Timothy Durrett

Copyright

Abstract

Pennycress (*Thlaspi arvense*) is an emerging oilseed biofuel crop within the Brassicaceae family, closely related to *Arabidopsis thaliana*. It holds significant potential as a platform for the production of biotechnologically derived compounds, such as medium-chain fatty acids (MCFAs). These MCFAs are saturated hydrocarbon chain molecules with 8 to 14 carbon atoms. These MCFAs are abundant in plants like coconut and species of *Cuphea*, including *Cuphea viscosissima* (*Cv*) and *Cuphea var. avigera pulcherrima* (*Cpu*), which can comprise up to 94 mol% of the oil content. To enhance MCFA production, pennycress was genetically engineered to express an MCFA-specific thioesterase (*CvFatB1*) and two MCFA-specific acyltransferases (*CvLPAT2* and *CpuDGAT1*), resulting in the accumulation of C10:0 fatty acids at 7 mol%. Using systems biology approaches, this study aims to identify the bottlenecks in C10:0 production in transgenic pennycress. Fatty acid composition analysis reveals increased C16:0 levels in transgenic plants compared to wild-type (WT) plants, with higher total fatty acid content throughout seed development. Lipidomic analysis indicates the production of altered C10:0-containing triacylglycerols (TAGs) and acyl steryl glucosides (ASGs) in transgenic plants, particularly in later stages of seed development. Moreover, the levels of TAG containing multiple medium-chain fatty acids decline as seed development progresses. Transcriptomic analysis shows no significant changes in gene expression for most genes associated with fatty acid and TAG synthesis, degradation, and peroxisomal β-oxidation pathways. Notably, transcriptomic data also highlight several upregulated genes in transgenic seeds related to stress responses, including heat shock proteins (HSPs) and genes associated with hypoxia response (HUPs).

Table of Contents

List of Figures vii

List of Tables ix

Acknowledgements x

Chapter 1- Introduction and Literature Review 1

1.1 Triacylglycerol and Its Significance 1

1.2 Fatty acids and their Classification 2

1.3 Medium-chain fatty acids 3

1.4 Triacylglycerol Synthesis 4

1.4.1 *De novo* FA synthesis 4

1.4.1.1 Plastidial acetyl-CoA carboxylase (ACCase) 4

1.4.1.2 FA synthase (FAS) complex and FA Export 5

1.4.1.3 Acyl-editing 6

1.4.1.4 De novo DAG (Kennedy pathway) 8

1.4.2 GPAT 8

1.4.3 LPAT 9

1.4.4 PAP 10

1.4.5 DGAT 11

1.5 Specialized enzymes engineered to synthesize medium-chain fatty acids 12

Chapter 2- System-Level Evaluation of a Transgenic Pennycress Line for Oilseed Enhancement 18

2.1 Introduction 18

2.2 Materials and Methods 19

Ds-Red Fluorescence Screening and Germination 19

Fatty acid composition analysis 20
Plant material and growth conditions 21
Flower Tagging 21
Germination and seedling emergence 21
Cryogenic seed preparation and grinding 22
Lyophilization Treatment of Developing Seeds 22
2.3 Results 23
Selection of transgenic lines for comprehensive systems biology profiling experiments 23
No morphological changes in transgenic pennycress siliculae and seeds 25
Lower total fatty acid content in transgenic pennycress seeds 30
Significantly elevated C16:0 levels in transgenic pennycress seeds 31
2.4 Discussion **33**
Chapter 3- Transcriptome and Lipidome Dynamics in Transgenic Pennycress Seeds 34
3.1 Introduction 34
3.2 Materials and Methods 35
Lipid analysis 35
RNA extraction 36
RNA-seq data filtering and analysis 36
Plot generation and statistical analysis 37
3.3 Results 38
Transgenic seeds possess a distinct lipid profile 38
Accumulation of altered C10:0 TAGs and some of the ASGs in transgenic seeds 40

TAG molecular species containing multiple C10:0 peaks during the seed development.....44
Transcriptomics reveals similar gene expression between transgenic and WT46
Significant changes in the gene expression pattern in transgenic seeds.....48
The top differentially expressed upregulated genes across the seed development 51
3.4 Discussion 54
Chapter 4- Conclusion and Future Work.....58
Conclusion 58
Future Work 60

List of Figures

Figure 2.1 Fatty acid composition of transgenic pennycress lines engineered to produce MCFAs.....24

Figure 2.2 No morphological changes were observed in siliculae and seeds of transgenic pennycress. 26

Figure 2.3 Transgenic pennycress seeds germinate at the same rate as wild-type (WT) seeds.....27

Figure 2.4 Variation in the seed weight across the developmental stages. 29

Figure 2.5 Altered fatty acid accumulation in pennycress seeds. 30

Figure 2.6 Transgenic pennycress lines possess increased levels of C16:0. Individual fatty acid profiles between transgenic pennycress FLD-9B3 and Wild-type (WT). 32

Figure 3.1 Transgenic seeds possess distinct lipid profiles from wild-type across seed development. 39

Figure 3.2 Lipid Accumulation Dynamics of transgenic seeds accumulating altered 10-carbon TAGs. 41

Figure 3.3 Transgenic FLD-9B3 possesses significantly elevated TG 16:0 across the seed development.....42

Figure 3.4 Transgenic FLD-9B3 accumulates 10-carbon (a.) and 8-carbon (b.) altered Acyl Sterol Glycosides (ASGs) across the seed development.....43

Figure 3.5 TAG 10:0_x:x lipid species containing multiple medium chains decrease relative to those with no or at least one MCFA throughout the seed development.45

Figure 3.6 2D-PCA plot demonstrates a similar gene expression pattern between wild-type and transgenic seeds. 47

Figure 3.7 Identification of Differentially expressed genes (DEGs) across the seed development. 48

Figure 3.8 Expression profile of lipid-related genes during seed development. 51

Figure 3.9 Top 10 differentially expressed genes (DEGs) in developing transgenic seeds. 53

Figure 4. 1 Proposed remodeling of endogenous pennycress lipid metabolism to enhance MCFA accumulation…….61

List of Tables

Table 2.1 Homozygous Ds Red and C10:0 Fatty Acid content in germinating characterized T_3 seedlings. 25

Acknowledgements

I would like to express my deepest gratitude to my major professor, Dr. Timothy Durrett, for his unwavering support, guidance, and valuable insights throughout the development and completion of this research. His expertise and encouragement played a crucial role in shaping this thesis.

I am equally grateful to my committee members, Dr. Kathrin Schrick and Dr. Ruth Welti, for their invaluable time, constructive comments, and useful suggestions on ways to improve my research project.

I would like to express my sincere appreciation to Pamela Tamura for her huge assistance in the cryogenic seed preparation and grinding process, followed by their distribution for multi-omics, which involved considerable work. I am also thankful to both Dexter White and Pamela Tamura for their significant contributions during the flower tagging and siliculae harvesting experiments.

I want to express my sincere gratitude to Dr. Ruth Welti, Mary Roth, and Pamela Tamura at the Kansas Lipidomics Research Center, Manhattan, KS, for their expert analytical support and contributions to the lipidomics analysis conducted in this study. I am also thankful to Dr. Edgar Cahoon and Eleana Kang at the Center for Plant Science Innovation, University of Nebraska-Lincoln, NE and Dr. Trupti Joshi and Yen On Chan at the Department of Biomedical Informatics, Biostatistics and Medical Epidemiology (BBME), University of Missouri, Columbia, for their valuable assistance and expertise in carrying out the transcriptomics analysis. Their involvement was integral to the generation of high-quality data and the overall success of this research.

The research was supported by the U.S. Department of Energy, DOE Award # DE-SC0023142, which made this work possible.

Lastly, I wish to express my appreciation to my peers and colleagues for their encouragement throughout this journey.

Dedication

I dedicate this work to my family and friends, whose unwavering support and encouragement have been a constant source of strength and inspiration in my pursuit of higher education.

Chapter 1- Introduction and Literature Review

1.1 Triacylglycerol and Its Significance

Triacylglycerol (TAG) comprises a glycerol backbone in which all three hydroxyl groups are esterified to fatty acid chains. It represents the major component of common vegetable oils, typically accounting for more than 90% of their composition. It is a highly reduced molecule that serves as an important energy reserve in eukaryotes. These are non-membrane storage lipids that mainly accumulate in the developing seeds, pollen grains, and in numerous fruit species, but are also present in small amounts in different vegetative tissues, including leaves, flower petals, stems, and roots (Xu & Shanklin, 2016). Different plant species have varying amounts of oil; it can be as little as 1% or as much as 70% of the seed's total dry weight (Ohlrogge & Browse, 1995). These storage lipids are of great nutritional and nutraceutical importance and are the major components of common edible vegetable oils for both human consumption and various industrial applications. Seeds' TAGs contain similar fatty acyl chains present in membrane lipids, including palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and alpha-linolenic acid (C18:3), which are often termed common fatty acids (Millar et al., 2000).

TAGs are assembled in seed tissues, such as the embryo and/or endosperm, in the endoplasmic reticulum (ER) during seed development (Chapman & Ohlrogge, 2012). TAGs build up in the spaces between the ER membrane's leaflets, forming lipid droplets (LDs) or oil bodies to develop into separate organelles within the cytosol. They are located in the center of LDs, where they are surrounded by a phospholipid and LD-associated protein monolayer (Martin & Parton, 2006). TAGs have been employed in various activities such as lipid homeostasis, flower and pollen development, cellular energy balance, seedling, and post-germination organ development (Yang

& Benning, 2018). The process of biosynthesis of TAG molecules involves complex enzymatic reactions and regulatory pathways that can be manipulated to either increase their production in modify TAG composition or both. This has great potential to yield high-value oils with improved functional qualities and production of oils rich in medium chain fatty acids (MCFAs), that meet a variety of needs in current industrial and nutritional settings.

1.2 Fatty acids and their Classification

Fatty acids (FAs) consist of a hydrocarbon chain with a carboxylic acid functional group at its one terminal end. FAs can be classified as saturated FAs (SFAs, with no double bonds), monounsaturated FAs (MUFAs, with one double bond), and polyunsaturated FAs (PUFAs, with two or more double bonds) based on the number of double bonds. Based on the carbon chain length, FAs can be distinguished by their carbon chain length, ranging from short-chain (SCFAs, with 6 carbons or less), medium-chain (MCFAs, with 8-14 carbons), long-chain (LCFAs, with 16-18 carbons) and very long chain (VLCFAs, with 20 carbons or more).

A simplified nomenclature of FAs is represented by specifying the number of carbons in the chain, followed by the number of double bonds, separated by a colon. Based on the position of double bonds, FAs can be further classified as omega-c (ω-c or n-c), with their distance (c) relative to the methyl end, and delta-c (Δ-c), denoting the distance (c) relative to the carboxyl end. For example, Linoleic acid (C18:2 Δ9,12), is a polyunsaturated fatty acid with 18-carbon atoms and two double bonds at positions 9 and 12 from the carboxyl end, whereas these double bonds are at ω-6 and ω-9 positions from the methyl end, termed as (C18:2 ω-6,9), respectively.

A diverse array of naturally occurring unusual FAs exists, accumulating in non-membrane lipids, including seed storage lipids of various native plant species (Badami & Patil, 1980b). These unusual FAs typically consist of fatty acid chain lengths (from C8 to C14, and from C20 to C22),

double bonds in unusual positions, or contain functional groups such as acetyl, hydroxyl, epoxy, and cyclopropane, that can make up to 90% of FAs in the oils of local species (Cahoon & Li-Beisson, 2020; Dyer *et al*., 2008). These unusual FAs containing oils, especially medium chain length category, hold significance as raw materials and have the potential to serve as an excellent feedstock for different industrial applications such as the production of cosmetics, detergents, and lubricants (Avato & Tava, 2022; Dyer *et al*., 2008).

1.3 Medium-chain fatty acids

Medium-chain fatty acids (MCFAs) are carboxylic acids with carbon chain lengths ranging from C8 to C14. These saturated fatty acids are naturally found in coconut (*Cocos nucifera*) and various species of *Cuphea*, including *Cuphea viscosissima* (Cv), *Cuphea avigera* var. *pulcherrima* (Cpu), and *Cuphea hookeriana* (*Ch*). These *Cuphea* species, which thrive in warm temperate to tropical climates, naturally accumulate high levels of MCFAs, particularly C8 and C10, which together constitute more than 90% of the total fatty acids in their seed oils (Graham & Knapp, 1989). The seeds of *Cuphea* species predominantly produce short and medium-chain fatty acids, such as caprylic (C8:0), capric (C10:0), lauric (C12:0), and myristic (C14:0) acids. For instance, *C. viscosissima* contains 9.1% C8:0, 75.5% C10:0, 3.0% C12:0, 1.3% C14:0, with the remaining 11.1% consisting of other fatty acids (Kleiman, 1990). In North America, the most prevalent fatty acid in *Cuphea* seed oil is C12:0, accounting for 50% of the oil composition, while in other regions, C10:0 is the most common (41%), followed by C12:0 (27%). Other fatty acids are less common and collectively make up 32% of the oil composition in North American *Cuphea* species (Graham et al., 2016).

However, these plant species are unfit to grow on agricultural land because of the lack of specialized cultivation equipment and the unfavorable climate (Gesch et al., 2014). Therefore,

three seed-specific enzymes, including *CvFatB1*, *CvLPAT2*, and *CpuDGAT1 which* are specifically involved in the synthesis, incorporation, and accumulation of these MCFAs, isolated from various *Cuphea* species, were introduced into pennycress.

These MCFAs have a broad range of applications, including food (Aoyama et al., 2007; Babayan, 1981; Xiang et al., 2019), feed, the production of different products such as detergents (Knaut & Ritchler, 1985), surfactants, aviation fuels, renewable sources for antibiotics, antimicrobial and flavoring intermediates, cosmetics, etc (Steinbusch et al., 2011).

1.4 Triacylglycerol Synthesis

TAG synthesis is a multi-step process that takes place in different subcellular compartments. The primary processes involve the production of FAs in plastids, their modification in the endoplasmic reticulum (ER), and the assembly of FAs onto the glycerol backbone of TAG in the ER.

1.4.1 *De novo* FA synthesis

1.4.1.1 Plastidial acetyl-CoA carboxylase (ACCase)

De novo fatty acid synthesis is a process that occurs in the plastids of all plant cells and employs the concerted activity of two enzymes, acetyl-CoA carboxylase and fatty acid synthase (Ohlrogge & Browse, 1995). ATP-dependent acetyl-CoA carboxylase (ACCase) is responsible for catalyzing the first committed step of fatty acid biosynthesis by converting acetyl-CoA and bicarbonate into malonyl-CoA (Alberts & Vagelos, 1968; Sasaki & Nagano, 2004). A plastid ACCase is a heteromeric complex that is composed of four subunits consisting of biotin carboxylase (BC), biotin carboxyl carrier protein (BCCP), and α- and β carboxyltransferase (α-CT and β-CT) (Konishi et al., 1996; Salie & Thelen, 2016; Sasaki *et al*., 1993). HtACCase activity is a key regulatory point in regulating the carbon flux into de novo fatty acid biosynthesis. This involves a

two-step reaction. First, the biotin-carboxylase domain (BC) catalyzes the carboxylation of the biotin cofactor covalently bound to BCCP, the central biotin carboxyl carrier domain. Second, the carboxyl group transfers from carboxy-biotin to acetyl-CoA to produce malonyl-CoA, with the help of the carboxyltransferase (CT) domain.

1.4.1.2 FA synthase (FAS) complex and FA Export

The next step involved in fatty acid synthesis is the conversion of malonyl-CoA into malonyl-acyl carrier protein (malonyl-ACP) through the action of malonyl-CoA: ACP malonyl transferase (MCAT/MCAMT) (Jung *et al*., 2019; Lessire & Stumpe, 1983). Malonyl-ACP then participates in a condensation reaction with acetyl-CoA as an initial substrate; the reaction is catalyzed by β-ketoacyl synthase III (KAS III), for the production of β-ketoacyl ACP (Clough *et al*., 1992; Jaworski *et al*., 1989; Wu & Xue, 2010). The condensation reaction initiates the four enzymatic reactions that constitute the fatty acid synthase complex (FAS) cycle. The process continues with the reduction of β-ketoacyl ACP to β-hydroxy acyl-ACP in the presence of the reducing agent NADPH (Chen et al., 2022) by β-ketoacyl ACP reductase (KAR), followed by dehydration into trans-2-enoyl ACP, by the action of β-hydroxyl-ACP dehydratase (HD) (Ohlrogge & Browse, 1995). The second reduction step leads to the formation of the elongated butyryl-ACP in the presence of enoyl-ACP reductase (ENR). This resultant newly synthesized acyl-ACP chain, along with malonyl-ACP, re-enters the FAS cycle. There are six additional cycles, with each one having malonyl-ACP providing two carbon units that result in the gradual elongation of butyryl-ACP to form C16:0-ACP or over seven cycles to form C18:0-ACP. These condensation reactions from butyryl-ACP to C16:0-ACP production are catalyzed by β-ketoacyl ACP synthase I (KAS I), and from C16:0-ACP to produce C18:0-ACP is enzymatically catalyzed by β-ketoacyl ACP synthase

II (KAS II). The resulting C18:0-ACP is desaturated to form C18:1-ACP by stromal stearoyl-ACP desaturase (SAD) (Kachroo et al., 2007).

Some fatty acyl-ACPs, such as C16:0-ACP and C18:1-ACP, are employed in the production of glycerolipids in the plastid, while the rest, including C18:0-ACP, are converted to free fatty acids by acyl-ACP thioesterases (FATA and FATB), for export from the plastid in the ER (Mayer & Shanklin, 2007). It has been observed that FATB enzymes have acyl specificity towards C16:0 > C18:1 > C18:0 as their substrate, whereas FATA enzymes typically show acyl specificity for C18:1 > C18:0 > C16:0 as their substrate (Salas & Ohlrogge, 2002). Plants that synthesize medium chain fatty acids (MCFAs), such as coconut (*Cocos nucifera*), date palm (*Elaeis guineensis Jacq*.), and *Cuphea spp.* release their fatty acids from the FAS complex early to produce acyl-ACPs with carbon chain lengths containing C8:0 to C14:0 carbons (Graham, 1989). FFAs generated at the inner envelope of plastid membranes are transported by Fatty Acid Export (FAX1,2 & 4) proteins (Li et al., 2015). Later on, long-chain acyl-CoA synthetase (LACS9, 1, and 4) (Jessen et al., 2015; Zhao et al., 2021; Shockey et al., 2002) converts FFAs to CoA esters in the outer envelope of the plastid membrane; the CoA esters are exported from the plastid to the cytoplasm, resulting in the formation of an acyl-CoA pool.

1.4.1.3 Acyl-editing

Isotopic tracking investigations have demonstrated that more than half of the newly synthesized C18:1-CoA is incorporated into the membrane lipid phosphatidylcholine once it is exported to the ER (Bates et al., 2013). Acyl editing, a cycle in which the fatty acids on phosphatidylcholine (PC) are exchanged with those in the acyl-CoA pool without net PC synthesis or degradation. It involves modifying C18:1 via desaturation to PUFA or other lipid hydrolysis modifications (Bates *et al*., 2009; Bates *et al*., 2012; Zhou *et al*., 2020). This editing involves the action of

lysphosphatidylcholine acyl transferase 1 & 2 (LPCAT1/2), which transfers the acyl group from fatty-CoAs to the sn-2 position of lyso PC (LPC), resulting in the formation of fatty-acyl PC. Fatty acid desaturases modify fatty acyl-PC. FATTY ACID DESATURASE2 (FAD2), a Δ12-desaturase, converts C18:1-PC to C18:2-PC, followed by desaturation facilitated by FATTY ACID DESATURASE3 (FAD3), a Δ15-desaturase, to convert C18:2-PC form C18:3-PC (Arondel et al., 1992; Okuley et al., 1994).

In addition, 18:1-phosphatidylcholine (PC), which undergoes acyl editing and modification, can also be synthesized from de novo diacylglycerol (DAG) through several pathways. One such pathway involves the enzyme phosphatidylcholine: diacylglycerol choline phosphotransferase (PDCT), which facilitates the exchange of phosphocholine between PC and DAG (Lu et al., 2009). Another pathway utilizes CDP-choline: diacylglycerol choline phosphotransferase (CPT), which generates PC from DAG (Slack et al., 1983; Slack et al., 1985). Additionally, lipase-based reactions, including the action of phospholipase D with phosphatidic acid phosphohydrolase (PAP) or phospholipase C, contribute to this interconversion. The enzymatic regulation of PDCT controls the equilibrium between PC and DAG, ensuring no net accumulation of PC. Furthermore, acyl-editing processes result in the transfer of approximately two-thirds of polyunsaturated fatty acids (PUFAs) from PC to triacylglycerol (TAG), as observed in the model plant *Arabidopsis* (Bates et al., 2012).

Given that PC acts as the site for acyl editing and FA modification in the ER, PC-derived DAG may have a different molecular composition from de novo DAG (Bates et al., 2009). The initial substrate for TAG synthesis in the ER is diacylglycerol (DAG), which can be generated either de novo via the Kennedy pathway or derived from PC (Li Beisson et al., 2013 & Ohlrogge & Browse, 1995).

1.4.1.4 *De novo* DAG (Kennedy pathway)

The Kennedy pathway begins with the acylation of glycerol-3-phosphate (G3P), which utilizes the acyl-CoA pool (Weiss et al., 1960). Acyl-CoA: glycerol-3-phosphate acyltransferase (GPAT) initiates the transfer of the acyl group from the acyl-CoA pool to G3P to the *sn*-1 position to form lysophosphatidic acid (LPA). Acyl-CoA: lysophosphatidic acid acyltransferase (LPAT) enzymatically transfers an acyl group from the acyl-CoA pool to the *sn*-2 position, generating phosphatidic acid (PA). Finally, the removal of the phosphate group by phosphatidic acid phosphatase (PAP) from the *sn*-3 position results in the formation of de-novo DAG. Finally, this de novo DAG is either utilized in the formation of *de novo* TAG directly or to produce PC and phosphatidylethanolamine (PE) (Lung & Weselake, 2006; Lu et al., 2009).

1.4.2 GPAT

GPAT catalyzes the first acylation at the backbone of glycerolipids, which is employed in the synthesis of storage lipids, membrane lipids, and extracellular lipids utilizing acyl-CoA, acyl-ACP, or fatty-acid derivative (Li-Beisson et al., 2013). In Arabidopsis, acyl-CoA: glycerol-3-phosphate acyltransferase 9 (GPAT9) initiates the first committed step of acylation in the Kennedy pathway by transferring an acyl group from acyl-CoA to the *sn-1* position of G3P, generating lysophosphatidic acid (LPA) (Shockey et al., 2016). Recent research has demonstrated that *AtGPAT9*, which is a membrane-binding protein located in the ER, exhibits *sn*-1 acyltransferase activity with a preference for acyl-CoA as its substrate. It was identified as the only GPAT employed in the eukaryotic intracellular lipids (membrane lipids) and storage TAG in Arabidopsis (Lv et al., 2020). Overexpression of BnaGPAT9 in *Brassica napus* seeds led to increased oil accumulation and enhanced galactolipid biosynthesis.

In Arabidopsis, overexpressing BnaGPAT9 resulted in larger seeds with increased seed weight, seed area, and seed oil content, whereas its downregulation caused significant reductions in these traits (Singer et al., 2016). A complete loss of GPAT9 led to the lethality of male and female gametophytes in Arabidopsis (Shockey et al., 2016). GPAT9 suppression led to reduced levels of seed oil and total polar lipids (Shockey et al., 2016; Singer et al., 2016).

1.4.3 LPAT

LPAT is a second enzyme involved in the Kennedy pathway, responsible for sequentially transferring acyl chains from acyl-CoAs to the *sn*-2 position and which uses lysophosphatidic acid (LPA) to form phosphatidic acid (PA) (Körbes et al., 2015). It is a large and ancient gene family that belongs to the MBOAT (membrane-bound O-acyltransferase) superfamily, found in both prokaryotes and eukaryotes. LPAT traditionally has strong selectivity and specificity towards unsaturated C18:0 acyl groups in oilseed crops and can distinguish between acyl groups with longer and shorter chain lengths. Five LPAT genes have been characterized in the model plant *Arabidopsis thaliana*: LPAT1, LPAT2, LPAT3, LPAT4, and LPAT5, among which *AtLPAT1* encodes a plastidial isoform, while the other four are cytoplasmic genes (Kim and Huang, 2005). These LPAT genes play significant roles in TAG biosynthesis, plant growth, and development (Kim and Huang, 2005; Angkawijaya et al., 2017 & 2019). Overexpression of LPAT1 in the microalga *Chlamydomonas reinhardtii* resulted in a considerable increase in oil content under nitrogen-deficiency conditions (Yamaoka et al., 2016). Overexpression of the seed-specific *AhLPAT2*, a specific LPAT isolated from the peanut (*Arachis hypogaea*), led to a 32.2% increase in oil content in Arabidopsis (Chen et al., 2015). Kim et al. (2015) showed the accumulation of medium-chain fatty acids at the sn-2 position of TAG due to the co-expression of *CpuLPATB* and *CvLPAT2* in Camelina. Another LPAT gene, named LPATB, was first cloned and characterized

from *Limnanthes douglasii* (Brown et al., 1995; Hanke et al.,1995), *Cocos nucifera* (Knutzon et al., 1995), and *Ricinus communis* (Arroyo-Caro et al., 2013). The LPATB gene is expressed in seeds and has substrate specificity for unusual acyl groups, including C12:0, which is naturally abundant in coconut (Knutzon et al., 1995), and C22:1, and found in meadowfoam (*Limnanthes alba*) (Cao et al., 1990; Hanke et al., 1995; Lassner et al., 1995).

1.4.4 PAP

Phosphatidic acid phosphatase (PAP) catalyzes the Mg^{2+}-dependent dephosphorylation of phosphatidic acid (PA) to form diacylglycerol (DAG), thereby playing a crucial role in regulating the metabolic flux between various glycerolipid classes within the endoplasmic reticulum (Carman et al., 2018; Pearce & Slabas, 1998). Two Mg^{2}-dependent phosphatidate phosphohydrolases, PAH1 and PAH2, in *Arabidopsis thaliana*. These enzymes are central to both lipid biosynthesis and homeostasis, acting as key mediators in the eukaryotic pathway of membrane lipid metabolism, particularly in response to phosphorus (Pi) starvation (Eastmond et al., 2010 & Nakamura et al., 2009). Disruption of both PAH1 and PAH2 in the *pah1 pah2-1* mutant results in significant alterations to lipid composition, characterized by an increase in phospholipid levels and expansion of the ER membrane, accompanied by a decrease in galactolipid synthesis derived from the eukaryotic biosynthetic pathway (Nakamura et al., 2009). This phenotype correlates with the increased expression of several genes involved in phospholipid biosynthesis, suggesting that PAH1 and PAH2 function as transcriptional repressors, potentially by modulating the accumulation of phosphatidic acid (PA), a known lipid signaling molecule in plants (Munnik, 2001; Wang et al., 2006).

Despite their involvement in production of diacylglycerol (DAG), a precursor for triacylglycerol (TAG) synthesis, the *pah1 pah2-1* mutant does not exhibit a significant reduction in seed TAG

content, likely due to compensation by other PA phosphohydrolases (Pearce and Slabas, 1998; Pierrugues et al., 2001; Nakamura et al., 2007). These observations suggest that PAH1 and PAH2 are critical in lipid biosynthesis, lipid homeostasis, and membrane biogenesis.

Further investigation revealed that PAH1 and PAH2 are integral to the eukaryotic pathway of membrane lipid metabolism by hydrolyzing PA to diacylglycerol (DAG), which serves as a substrate for the biosynthesis of galactolipids, such as MGDG and DGDG, in the inner envelope of chloroplasts (Nakamura et al., 2009). This process is crucial for membrane remodeling, particularly during Pi starvation, when plants shift from phospholipids to galactolipids to enhance their capacity to cope with nutrient deprivation. In the *pah1 pah2* double mutant, this pathway is severely impaired, leading to a significant reduction in galactolipid levels and an accumulation of phospholipids under normal growth conditions. These effects are exacerbated during Pi starvation, further emphasizing the pivotal role of PAH1 and PAH2 in membrane lipid dynamics under stress conditions (Cruz-Ramírez et al., 2004).

1.4.5 DGAT

DGAT is an enzyme that transfers acyl groups from acyl-CoA to the *sn*-3 position of DAG to produce TAG, expressed widely in various organisms, including bacteria, algae, plants, and mammals (Chen *et al*., 2022). There are three classes of DGAT enzymes, namely DGAT1, DGAT2, and DGAT3. Membrane-bound DGAT1 and DGAT2 are structurally distinct integral membrane proteins with variations in substrate selectivity, pattern of tissue expression, and membrane ER localization (Shockey *et al*., 2006). DGAT3 is a soluble enzyme first identified in peanuts (Saha *et al*., 2006). In Arabidopsis, DGAT1 is mainly expressed in developing seeds and pollen, with less expression seen in vegetative tissues, and on the other hand, DGAT2 is primarily

expressed in vegetative tissues but has less expression in the late stages of seed development (Zhou *et al*., 2013). Mutations in Arabidopsis DGAT1 resulted in the reduction of seed oil content by 20% with an increase in C18:3 FA and a decrease in C20:1 and C18:1 FAs, with similar results observed during DGAT1 suppression in camelina (Katavic *et al*., 1995; Zhang *et al*., 2009). Overall, DGAT1 is the major enzyme responsible for seed TAG synthesis. Research evidence suggested that DGAT1 in the model plant Arabidopsis and various other plant DGAT2s function in distinct TAG-synthesizing metabolons, each one of them engaging different PC-derived DAG pools for TAG synthesis (Bates, 2022).

At least three metabolically distinct pools of DAG are involved in the lipid metabolic network flux in developing Arabidopsis and other plants, according to studies involving in vivo isotopic trace studies of acyl or glycerol flux. There are two important DAG pools known as de-novo DAG and initially produced PC-derived DAG, both representing small and quickly turned-over pools. The third largest and more slowly turned over DAG pool derived from the PC-derived DAG pool is found, possibly connected to the oil body seen in both Arabidopsis and soybean (Bates et al., 2009; Regmi et al., 2020; Bates et al., 2022).

1.5 Specialized enzymes engineered to synthesize medium-chain fatty acids

A specific medium-chain acyl-acyl carrier protein thioesterase was identified in the developing oilseeds of California Bay (*Umbellularia californica)* (Pollard et al., 1991). This plant naturally accumulates C10:0 and C12:0 as major fatty acyl groups. When transformed in Arabidopsis, overexpression of this acyl-ACP thioesterase, designated as *UcFatB1,* exhibited a 70-fold increase in C12:0-ACP thioesterase activity, which led to 23.5% C12:0 in transgenic seeds (Voelker et al., 1992). Additionally, when these specialized thioesterases were expressed in transgenic crucifers

such as *Brassica rapa*, a significant portion of the storage oil was diverted toward the synthesis of the medium-chain fatty acids. When *B. napus* plants were transformed with *UcFatB1* driven by the *CaMV 35S* promoter, medium-chain fatty acids were found in seeds, not leaves or roots (Eccleston et al., 1996). Despite high lauroyl-ACP thioesterase activity in leaves, C12:0 accumulation was minimal, likely due to rapid degradation by β-oxidation.

Subsequently, acyl-ACP thioesterase cDNAs were identified from the maturing seeds of *C. hookeriana*, designated as *Ch FatB1* and *Ch FatB2* (Jones et al., 1995; Dehesh et al., 1996). Both differ in their amino-acid sequence and their substrate specificity. In vitro enzyme assay results showed that when these enzymes were expressed in *E. coli*, *Ch FatB1* is active on C14:0-, C16:0-, C18:0-, and C18:1-ACP, with the strongest preference for C16:0-ACP, whereas *Ch FatB2* is active on C8:0- and C10:0-ACP, with a stronger preference for C10:0-ACP. The seed-specific expression of *Ch FatB1* in transgenic canola led to the production of C16:0-enriched oil, whereas, on the other hand, *Ch FatB2* resulted in substantial accumulation of C8:0 and C10:0 in the seed oil.

Two novel thioesterases were isolated from the developing seeds of *C. palustris*, which were subsequently designated as *CpFatB1* and *CpFatB2* (Dehesh et al., 1996). Analysis of mature seed oil from *C. palustris* revealed that the seed composition predominantly comprises C8:0 (20%) and C14:0 (64%) fatty acids. When assessed using an *E. coli* expression system in conjunction with enzyme activity assays, it indicated that *CpFatB1* exhibited high specificity for C8:0 and C10:0 fatty acids, while *CpFatB2* showed a stronger preference for C14:0 and C16:0 fatty acids. Additionally, *CpFatB1* displayed approximately 10-fold lower enzymatic activity compared to *CpFatB2*. These findings suggest a correlation between enzyme activity, specificity, and the triacylglycerol (TAG) composition in the seeds of *C. palustris*. Three seed-specific and structurally

divergent thioesterases (TEs) from typical FatB enzymes, which release C16:0 from ACPs, were identified: *CpuFatB3, CvFatB1*, and *CpuFatB4*. When these genes were expressed in camelina, the expression of *CvFatB1* and *CpuFatB3* led to the accumulation of capric acid (C10:0) in camelina oil, while *CpuFatB4* expression resulted in C14:0 accumulation and an increase in C16:0 levels. Additionally, the introduction of coconut lysophosphatidic acid acyltransferase (*CnLPAT*) increased C12:0 and C14:0 levels at the *sn*-2 TAG position, but did not affect C10:0 levels (Kim et al., 2015b).

Another set of experiments includes the isolation of three additional TEs from the developing seeds of *C. viscosissima*, designated as *CvFatB1, CvFatB2, and CvFatB3*. When expressed in E. coli to evaluate substrate specificity and fatty acid composition, findings indicate that all three TEs exhibit distinct substrate specificities, showing that *CvFatB1* produced C8:0 (2-fold of C10:0), whereas it yielded approximately four times more C10:0 when sourced from *C. viscosissima* seed oil. *CvFatB1* demonstrated substrate specificity for both C8:0 and C10:0, *CvFatB2* showed a preference for substrates C14:0 and C16:0, while *CvFatB3* displayed a higher specificity towards C14:0 (Jing et al., 2011).

It was noted that the saturated fatty acids are rare in the *sn*-2 position in TAG in seed oils. Therefore, a medium-chain fatty acid-specific LPAT was isolated from the endosperm of immature *C. nucifera* (coconut) (Knutzon et al., 1995). Its confirmation was expressed in an *E. coli* system, which demonstrates that the expressed coconut LPAT showed high MCFA substrate activity (C10:0- C14:0), with the highest activity for C12:0. Later, the functionality of this enzyme was demonstrated through its transgenic expression in *B. napus* (canola), which led to efficient laurate (C12:0) deposition at the *sn*-2 position of TAG, resulting in the accumulation of trilaurin (Knutzon et al., 1999). Previously discovered C12:0-ACP-specific thioesterase isolated from *Umbellularia*

californica (California bay laurel) was expressed in developing seeds of *B. napus*, which led to oil production containing up to 50% laurate (Voelker et al., 1996). Thus, the introduction of the coconut protein into BTE oilseed canola lines above 50 mol% further increased total laurate levels, as mentioned in (Knutzon et al., 1999).

Several LPAT cDNAs were identified, including *CvLPAT2* isolated from *Cuphea viscosissima*, a seed-specific class A LPAT2 isolated from *Cuphea avigera var. pulcherrima*, designated as *CpuLPAT2a,* and another seed-specific LPAT from bacterial-type class B from *C. avigera var. pulcherrima* as *CpuLPATB*. When co-expressed in *Camelina sativa* with Cuphea FatB acyl-ACP thioesterases, *CvLPAT2* and *CpuLPAT2a* resulted in C10:0 accumulation at the *sn*-2 position in TAG, and *CpuLPATB* resulted in C14:0 accumulation at the *sn*-2 TAG position, but not C10:0 (Kim et al., 2015a).

The transcriptome of developing *C. avigera* var. *pulcherrima* seeds was analyzed to identify candidate genes encoding diacylglycerol acyltransferases (DGATs) potentially involved in medium-chain fatty acid (MCFA) metabolism (Kim et al., 2015b). This analysis led to the identification of one DGAT1-like gene, designated *CpuDGAT1*, and three DGAT2-like genes, named *CpuDGAT2_A*, *CpuDGAT2_B*, and *CpuDGAT2_C*. Reverse transcription PCR (RT-PCR) analysis revealed that *CpuDGAT1* is specifically expressed in seeds, whereas *CpuDGAT2_A*, *CpuDGAT2_B*, and *CpuDGAT2_C* exhibit constitutive expression across various tissues of *C. avigera* var. *pulcherrima* (Iskandarov et al., 2017).

The expression of *CvFatB1* and *CpuDGAT1* in *Camelina* seeds resulted in a significant increase in C10:0 levels, ranging from 4 to 13.5 mol%, nearly double the levels observed with the expression of *CvFatB1* alone. When *CvFatB1*, *CvLPAT2*, and *CpuDGAT1* were coexpressed, C10:0 levels were further enhanced, reaching 12 to 21.5 mol%, approximately three times higher

than those observed with *CvFatB1* expression alone. In particular, the *CvFatB1+CpuDGAT1* line accumulated 14.5 mol% C10:0 in seed triacylglycerols (TAGs), while coexpression with *CvLPAT2* elevated this value to 23.7 mol%. Furthermore, the *CvFatB1+CvLPAT2+CpuDGAT1* line exhibited 2.8 mol% C8:0 and 38.4 mol% C10:0 at the sn-2 position of TAGs (Iskandarov et al., 2017).

Similarly, engineered transgenic pennycress (*Thlaspi arvense*) lines were developed using gene combinations of *CvFatB1 + CvLPAT2* and *CvFatB1 + CvLPAT2 + CpuDGAT1* to enhance the accumulation of medium-chain fatty acids (MCFAs). Under the control of the soybean *Glycinin1* promoter, the *CvFatB1+CvLPAT2* lines produced triacylglycerols (TAGs) containing 7.5 to 16.6 mol% MCFAs, with C10:0 as the predominant fatty acid (up to 10 mol%), alongside a significant increase in C16:0 levels (5.6 to 12.7 mol%), compared to the wild type (3.1 to 3.9 mol%). In lines expressing *CvFatB1+CvLPAT2+CpuDGAT1*, the levels of C8:0 and C10:0 increased to 5.3 mol% and 15.5 mol%, respectively, resulting in total MCFA content ranging from 6.7 to 26.9 mol%. When combined with C16:0, the total fatty acid content ranged from 13.5 to 38.5 mol%. These results demonstrate the successful genetic engineering of fatty acid profiles in both *Camelina sativa* and *Thlaspi arvense* through targeted gene expression (Esfahanian et al., 2021).

These studies demonstrate that it is possible to engineer the production of medium-chain fatty acids (MCFAs) in the triacylglycerols (TAGs) and acetyl-TAGs of *Thlaspi arvense* (pennycress) seeds. However, the plant species from which these MCFA-synthesizing enzymes were isolated, like *Cuphea* species, are naturally capable of accumulating over 90 mol% MCFAs. One possibility for pennycress is that these plants efficiently direct MCFAs into TAGs, preventing their incorporation into cellular membranes where they could disrupt membrane function. Further

research is needed to explore this mechanism and gain a better understanding of how efficiently MCFAs are incorporated into TAGs in pennycress, as unincorporated or free MCFAs can be toxic to plant cells and tissues (Wang et al., 2015; Yang and Benning, 2018). Another limitation of this process is the significant reduction in total seed oil content observed in engineered pennycress lines, which could limit the feasibility of using these oils. Therefore, systems biology analyses will be essential to identify potential bottlenecks that restrict the production of MCFAs, particularly C10:0, in engineered pennycress, especially for high-value markets such as consumer products and other industrial applications.

Chapter 2- System-Level Evaluation of a Transgenic Pennycress Line for Oilseed Enhancement

2.1 Introduction

Pennycress (*Thlaspi arvense L.*), a member of the Brassicaceae family, is an emerging oilseed crop with significant potential for domestication due to its ability to produce large quantities of seeds rich in triacylglycerols (TAGs) (Sedbrook et al., 2014). The seeds typically contain about 33% TAGs by dry weight, primarily made up of fatty acids (FAs), including erucic acid (C22:1; ~33%), oleic acid (C18:1; ~13%), linoleic acid (C18:2; ~20%), and linolenic acid (C18:3; ~15%) (Moser et al., 2009). Pennycress has a relatively small diploid genome (Dorn et al., 2015) and can be genetically transformed using an *A. tumefaciens* floral dip method (McGinn et al., 2019). Wild-type (WT) pennycress seeds do not naturally accumulate medium-chain fatty acids (MCFAs), defined as saturated fatty acids with hydrocarbon chain lengths ranging from C8:0 to C14:0 (Drenth et al., 2015). MCFAs have widespread industrial and consumer applications, including in cosmetics, surfactants, soaps, detergents, and lubricants (Knaut and Richtler, 1985).

To broaden the genetic diversity of pennycress and enable the production of novel fatty acids that could enhance the value of pennycress oil, a biotechnological approach was adopted. This approach involved introducing genes from non-agronomic plant species like *Cuphea viscosissima (Cv)* and *Cuphea avigera var. pulcherrima (Cpu*), which produce a significant amount of MCFAs, particularly C10:0, which comprises ≥90 mol% of their seed oils. The key genes responsible for MCFA biosynthesis, incorporation, and accumulation into TAGs have been well-characterized. For instance, the FatB gene from *Cuphea viscosissima (CvFatB1)* facilitates the generation of MCFAs, such as C10:0, which are then incorporated into seed TAGs. The

CvLPAT2 enzyme preferentially incorporates MCFA at the *sn*-2 position of lysophosphatidic acid (LPA) (Kim et al., 2015a), while DGAT, isolated from *Cuphea avigera var. pulcherrima (CpuDGAT1)*, acylates the *sn*-3 position of diacylglycerol (DAG), thereby completing TAG synthesis (Iskandarov et al., 2017).

When *CvFatB1* and *CvLPAT2* were co-expressed in pennycress under the control of the soybean Glycinin-1 promoter, the resulting transgenic plants produced TAGs containing 7.5 to 16.6 mol% MCFAs, with C10:0 accumulating up to ~10 mol%. Additionally, the co-expression of *CpuDGAT1* with *CvFatB1* and *CvLPAT2* further increased C10:0 levels to approximately 15.5 mol%, with total MCFAs ranging from 6.7 to 26.9 mol% (Esfahanian et al., 2021). However, this C10:0 accumulation is considerably lower than the ≥90 mol% C10:0 observed in *Cuphea* species, despite the successful incorporation of the relevant genes.

The primary objective of this project is to conduct comprehensive systems biology experiments to identify potential bottlenecks that may limit the accumulation of C10:0 in transgenic pennycress seeds. Furthermore, the project aims to explore strategies for mitigating or eliminating these bottlenecks, with the ultimate goal of enhancing C10:0 accumulation in the transgenic seeds.

2.2 Materials and Methods

Ds-Red Fluorescence Screening and Germination

Independent transgenic pennycress lines expressing FatB1 from *Cuphea viscosissima*, LPAT2 from *Cuphea viscosissima*, and DGAT1 from *Cuphea avigera var. pulcherrima* were developed. These lines, including the combination *CvFatB1* + *CvLPAT2* + *CpuDGAT1*, collectively termed FLD, were obtained from Dr. John C. Sedbrook's laboratory at Illinois State University.

DsRed, a red fluorescent protein used as a selectable marker in transgenic seeds, was screened in 15 seeds from each FLD line. Wild-type pennycress (WT) seeds served as controls. All seeds were germinated on damp filter paper in glass trays, covered with PVC film for incubation. This experiment was conducted in a growth chamber under a 16-hour light/8-hour dark cycle, with a light intensity of 250 μmol m^2s^1, a temperature of 21ºC, and a relative humidity of 50%. Fluorescence was verified using a Dual Fluorescent Protein Flashlight (Night Sea, Electron Microscopy Sciences) and a pair of red filter glasses (UV400) in a dark room.

Fatty acid composition analysis

The fatty acid composition and accumulation of individual fatty acids in seed triacylglycerols (TAGs) were determined using seed fatty acid transmethylation, as described by Li et al. (2006). For each sample, 20 mature seeds were combined with 1 mL of toluene and briefly homogenized using a Polytron (*PT 2500E*). Triheptadecanoic acid (Nu Chek Prep, Inc., MN) of 100 μg was added as an internal standard, followed by the addition of 1 mL of 5% H2SO4 (v/v) in methanol and 0.2% butylated hydroxytoluene (BHT). The samples were vortexed for approximately 5 seconds and then heated at 90-95°C for 60 minutes in a dry heat block. After cooling to room temperature, 1.5 mL of 0.9% potassium chloride (KCl) and 1 mL of hexane were added to each sample. The samples were vortexed briefly and then centrifuged at 2500 rpm for approximately 5 minutes. The organic phase was carefully transferred to a clean glass tube, and approximately 1 mL was subsequently placed into small glass chromatography vials, which were sealed with caps. The resulting fatty acid methyl esters were then analyzed by gas chromatography, following the method outlined in Aznar-Moreno & Durrett (2017).

Plant material and growth conditions

The transgenic FLD-9B3 and the wild-type (WT) pennycress were germinated in a growth chamber under controlled conditions (n=6 biological replicates), with a 16-hour light and 8-hour dark photoperiod. The growth conditions were maintained with a light intensity of 250 µmol m^2s^1, a temperature of 21ºC, and a relative humidity of 50%. Before planting, seeds were stored overnight at -20ºC in 1.5 mL Eppendorf tubes. For planting, a soil mixture consisting of Berger BM6 Optimal Porosity all-purpose soil mix, perlite, and vermiculite (4:2:1 ratio). Osmocote 14-14-14 slow-release fertilizer was used.

Flower Tagging

Seeds representing distinct developmental stages—specifically 10, 15, 20, 25, 30, 35 days after flowering (DAF), as well as mature seeds, were collected by tagging newly emerged flowers with color-coded threads and orthodontic ligatures. Tagged siliculae were harvested at defined intervals corresponding to the designated time points. Immediately following harvest, seeds and remaining siliculae were flash-frozen in liquid nitrogen and stored at -80°C for subsequent analyses. Mature seeds, once dried, were stored at room temperature for future use.

Images of siliculae were captured using an Android phone, while seed images were acquired using a Carl Zeiss Stemi 2000-C Stereo Microscope (magnification range 6.5x-50x).

Germination and seedling emergence

Mature seeds from both wild-type (WT) and transgenic FLD-9B3 were subjected to a seed sterilization protocol. For each genotype, approximately 45 seeds were selected from six biological replicates and placed into 1.5 mL Eppendorf tubes. The seeds were initially washed with 1 mL of 70% ethanol for 1 minute. After the ethanol was removed, the seeds were treated with a 50%

bleach solution containing 0.05% Tween-20 for 10 minutes, with gentle agitation. Following this treatment, the bleach solution was discarded, and the seeds were rinsed 3 to 4 times with sterile, distilled water to ensure the complete removal of any residual bleach.

Once sterilized, the seeds were plated on 0.5X Murashige and Skoog (MS) media, without sucrose, with 15 seeds per biological replicate for each genotype. The plating procedure was performed under sterile conditions in a laminar flow hood. The plates were subsequently transferred to a growth chamber, where conditions were set to a 16-hour light/8-hour dark cycle, with a light intensity of 250 µmol m^2s^1, a temperature of 21°C, and a relative humidity of 50%. The experiment was conducted with six biological replicates per genotype, and each experimental setup was repeated in triplicate as technical replicates. Germination was assessed by the emergence of the radicle, while cotyledon emergence was determined by the appearance of the shoot.

Cryogenic seed preparation and grinding

Seeds were ground into a fine powder under liquid nitrogen. The seed tissue samples were allocated for five distinct experimental analyses: Fatty acid content analysis (used 20 seeds per sample), proteomics (~20 mg), transcriptomics (~100 mg), acyl carrier protein (ACP) analysis (~50 mg), targeted metabolomics (~30 mg), and lipidomics (~30 mg). After measurement, the samples were placed on dry ice and immediately immersed in liquid nitrogen to preserve their integrity until further processing.

Lyophilization Treatment of Developing Seeds

A total of 20 seeds from both wild-type and FLD-9B3 were collected for all six biological replicates at seven different time points. These samples were stored in 1.5 mL Eppendorf tubes, each fitted with a small, pierced hole in the cap. After the freeze-drying process, the samples were

carefully removed from the vacuum pump vessel. The pierced holes in the caps of the Eppendorf tubes were immediately sealed with Scotch tape, and the tubes were stored in a desiccator to prevent exposure to moisture before subsequent fatty acid content analysis.

2.3 Results

Selection of transgenic lines for comprehensive systems biology profiling experiments

Several independent transgenic pennycress lines were designed to overexpress *CvFatB1*, *CvLPAT2*, and *CpuDGAT1* in a three-gene combinatorial construct. These independent lines were obtained from Dr. John Sedbrook, Illinois State University. Prior to this study, these genetically engineered events had been evaluated for the amount of fatty acids (FAs) accumulation in TAGs. T_2 seeds from transgenic pennycress plants containing *CvFatB1+ CvLPAT2+ CpuDGAT1* gene combination accumulated up to 15.5 mol% C10:0. In T_3 seeds, the average fatty acid (FA) accumulation in TAGs among independent transgenic events was approximately 7.73 mol% of C10:0 (Esfahanian et al., 2021), but these T_3 generation lines were not characterized extensively. Therefore, to confirm which lines had the highest C10:0 accumulation, were pure homozygous for the introduced three transgene traits, and germinated efficiently, several characterization tests were performed on T_3-generation seedlings in pennycress to select the best homozygous transgenic FLD event.

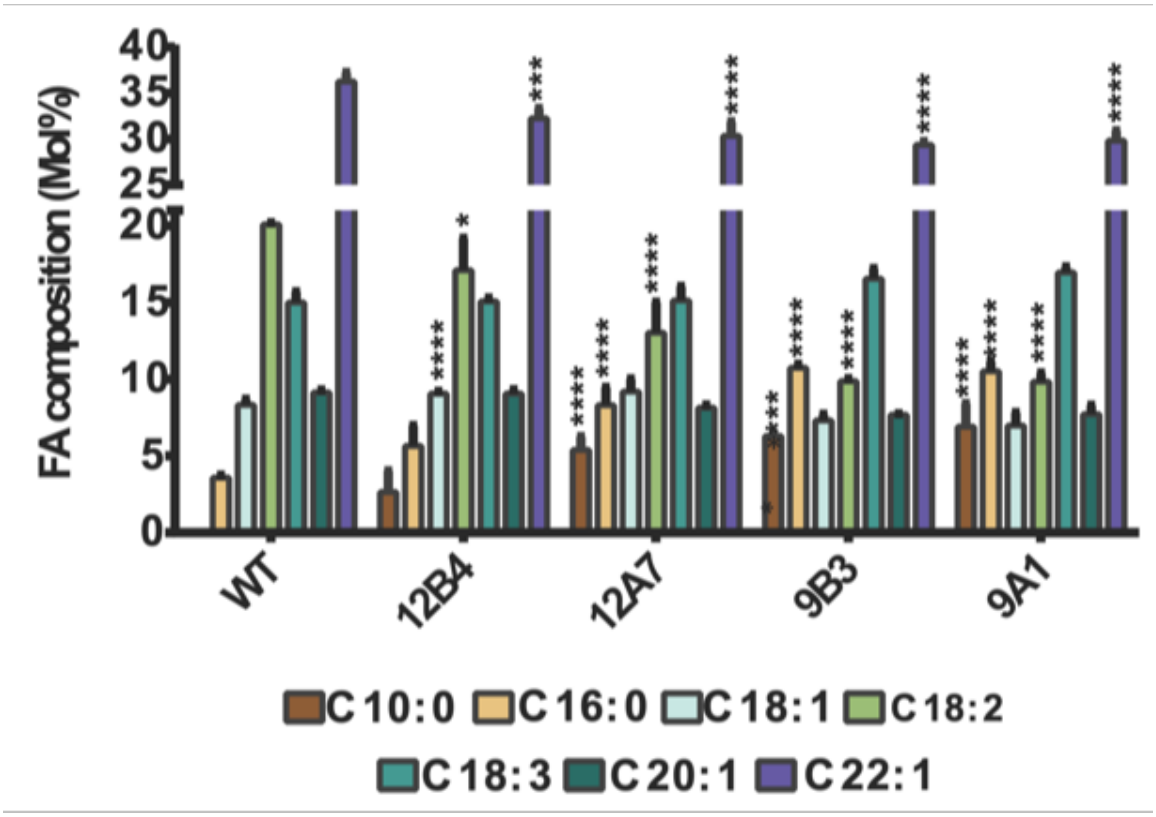

Figure 2.1 Fatty acid composition of transgenic pennycress lines engineered to produce MCFAs. Fatty acid composition of pennycress wild-type (WT) seeds and T_3 seeds for independent transgenic lines expressing *CvFatB1+CvLPAT2+CpuDGAT1* genes. Data represent mean ± standard deviation, n=3 biological replicates. Asterisks indicate statistical differences from wild-type seeds (Tukey's multiple comparisons test; **P< 0.05; ***P< 0.001; ****P<0.0001*).

Based on the screening results obtained for the T_3 seedlings, we identified line FLD-9B3 as the best transgenic pennycress line overexpressing the *CvFatB1+ CvLPAT2+ CpuDGAT1* gene combination. This line demonstrated a ~70% germination rate, was homozygous based on Ds Red fluorescence, and demonstrated stable and consistent C10:0 accumulation among three biological replicates, with an average of approximately 6.3 mol% C10:0 levels (Table 2.1). The fatty acid

profile of all the transgenic lines indicated that, in addition to the presence of C10:0, higher levels of C16:0 were present compared to WT seeds (Figure 2.1).

Genotype	Fluorescence (%)	Germination rate (%)	C10:0 fatty acid content (mol%)
FLD-9B3	100	73.33	6.3 ±0.20
FLD-12A7	100	100	5.39 ±0.86
FLD-12B4	100	73.33	2.64 ±1.39
FLD-9A1	100	93.33	6.9 ±1.52
WT	0	86.67	0

Table 2.1 Homozygous and C10:0 Fatty Acid content in germinating characterized T_3 seedlings. A total of (n=15) seeds were taken for every genotype.

No morphological changes in transgenic pennycress siliculae and seeds

To identify potential metabolic bottlenecks contributing to the lower C10:0 levels in these transgenic lines, T_3 seeds from line FLD-9B3 were planted along with the wild type (WT) T_4 seeds were collected at seven different time points. No morphological differences were observed in either the fresh siliculae or seeds of the transgenic plants when compared to the wild-type (WT) controls grown at the same time (Figure 2.2).

To further evaluate the viability of the seeds, germination rate, and cotyledon emergence were measured in T_4 seeds of transgenic pennycress line FLD-9B3 and wild type (WT). Both the transgenic FLD-9B3 and WT germinated at the same rate for both the radicle and cotyledon emergence (Figures 2.3a & b).

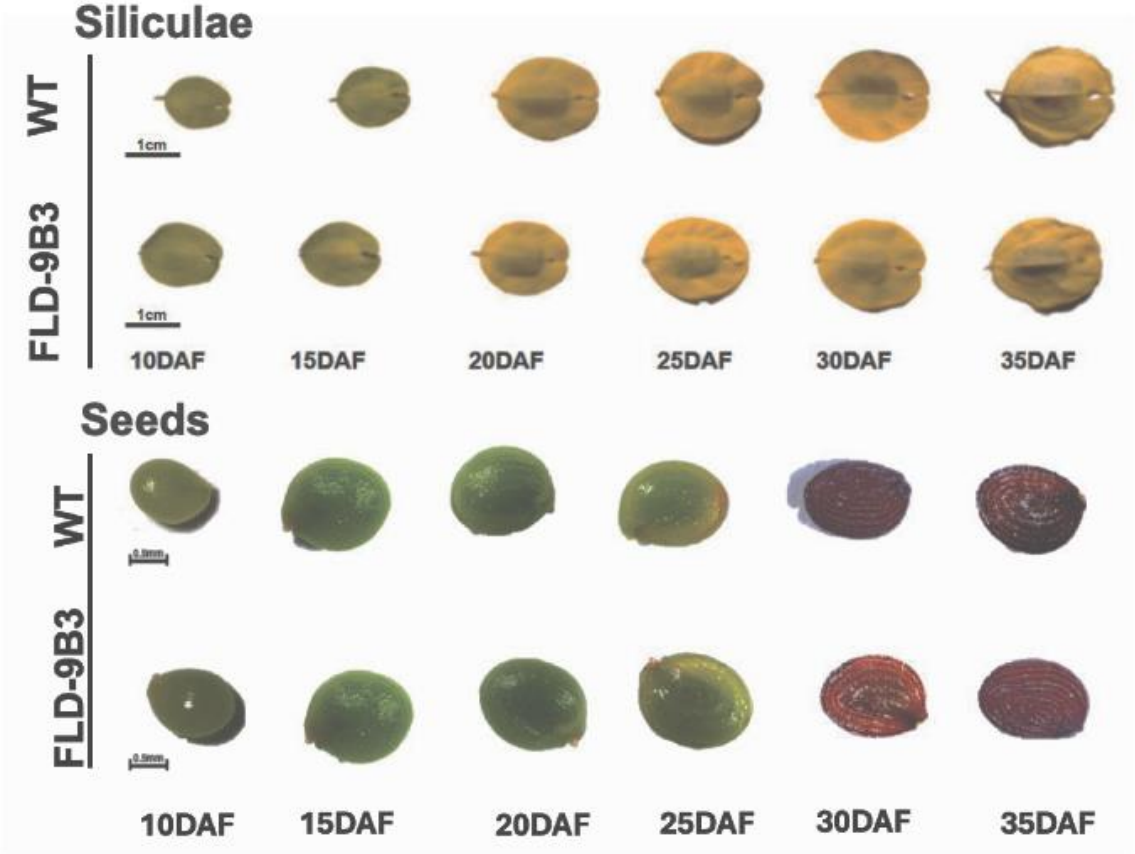

Figure 2.2 No morphological changes were observed in siliculae and seeds of transgenic pennycress. Siliculae and seeds from Wild-type (WT) and transgenic pennycress (FLD-9B3) were collected at different Days After Flowering (DAF). Bars represent 1cm for siliculae and 0.5mm for seeds.

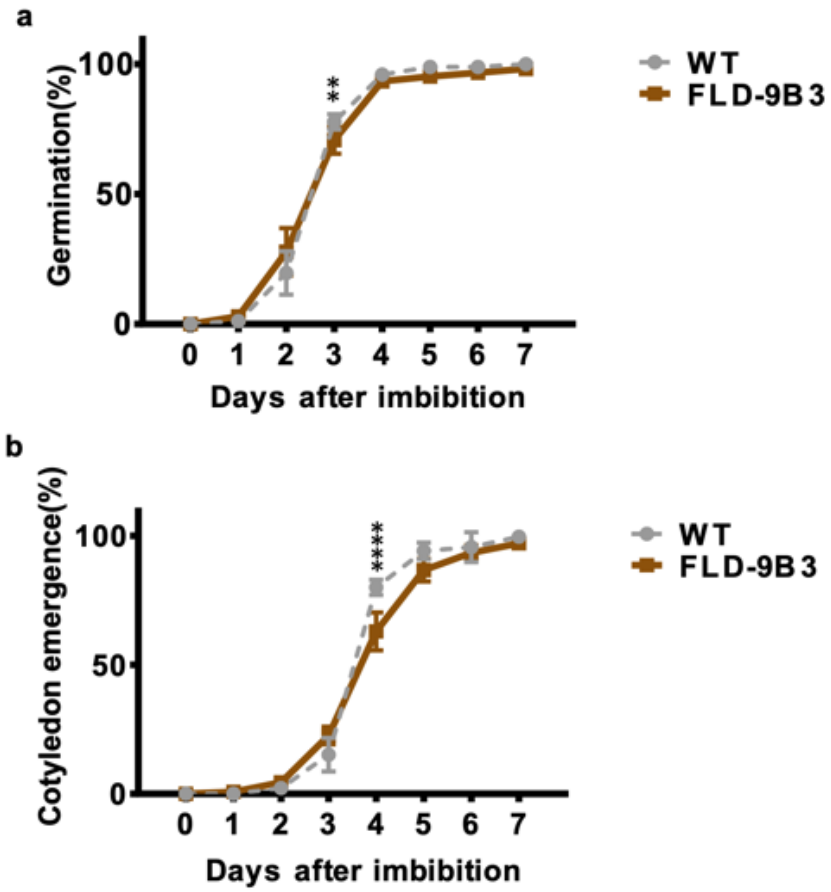

Figure 2.3 Transgenic pennycress seeds germinate at a similar rate to wild-type (WT) seeds. Germination rate (a) and Cotyledon emergence (b) at different Days After Imbibition. A total of (n=15) seeds were plated on Murashige and Skoog (MS) media, with no sucrose. Data represent mean ± standard deviation. Statistical analysis was performed using two-way ANOVA to assess the effects of genotype (FLD-9B3 vs. WT) and 7 time points, as well as their interaction. Post hoc comparisons were conducted using Sidak's multiple comparison test. Significant differences are indicated by asterisks *(**$P \leq 0.01$; ****$P \leq 0.0001$)*; n=6 biological replicates.

Comparative Analysis of Fresh and Dry Weight Dynamics During Seed Development

The fresh weight per seed initially increases during the early stages of seed development, specifically at the 10th DAF and 15th DAF (Figure 2.4a). This increase is followed by a noticeable decline in fresh weight after the 20th DAF as the seeds near maturity. Overall, the fresh weight curve demonstrates typical growth followed by a decline, likely reflecting the onset of desiccation during seed maturation. In contrast, the dry weight per seed gradually increases over time (Figure 2.4b). This steady increase is followed by a plateau from the 25th DAF onwards, reaching stability as the seeds approach full maturation. These trends were consistent across both genotypes, reflecting typical seed development dynamics regarding water loss and dry matter accumulation as the seeds mature. In conclusion, the data suggest that there are no significant differences noticed between FLD-9B3 and the WT for both the fresh weight and dry weight per seed.

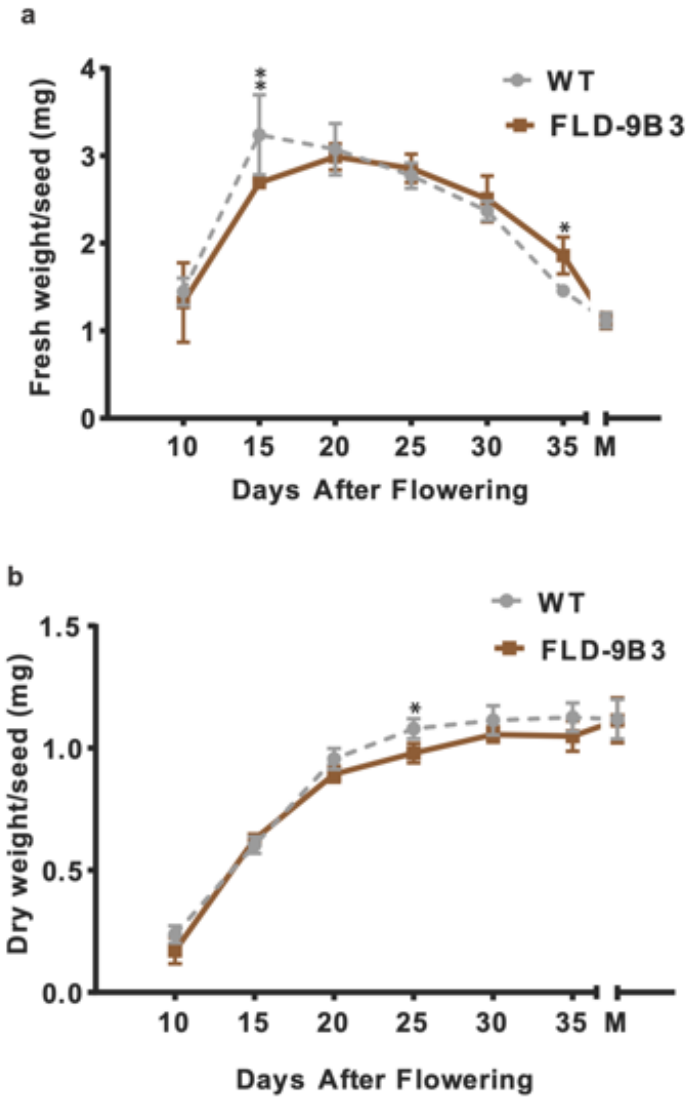

Figure 2.4. Comparative Analysis of Fresh and Dry Weight Dynamics During Seed Development. Fresh weight (a) and Dry weight (b) at different Days After Flowering (DAF) for wild-type (WT) and transgenic pennycress seeds expressing *CvFatB1+CvLPAT2+CpuDGAT1*. Values represent the mean ± standard deviation. Statistical significance was determined using two-way ANOVA followed by Sidak's multiple comparison test. Asterisks indicate significant differences between the genotypes at each developmental stage *(*P ≤ 0.05, ** P ≤ 0.01)*; n=6 biological replicates.

Lower total fatty acid content in transgenic pennycress seeds

The total fatty acid content in the transgenic pennycress line FLD-9B3 was consistently lower than that of the wild-type (WT) plants during the mid to late stages of seed development, specifically at 20, 25, 30, and 35 days after flowering (DAF), as well as at seed maturity (Figure 2.5). However, no significant differences in total fatty acid content were observed between FLD-9B3 and WT during the early stages of seed development, at 10 and 15 DAF. These results suggest that the transgenic modification primarily affects fatty acid accumulation during the mid to later stages of seed maturation (Figure 2.5).

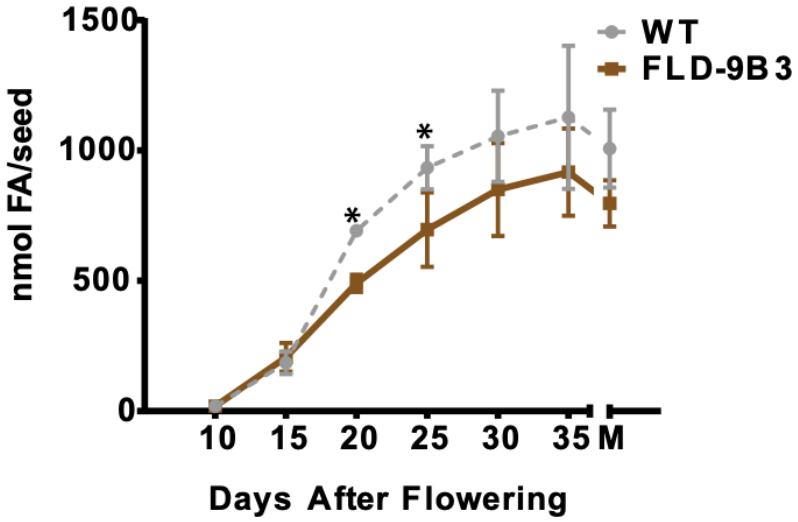

Figure 2.5 Lower total fatty acid content in transgenic FLD-9B3. Data represents the mean ± standard deviation. Data were assessed for normality using the Shapiro–Wilk test (α=0.05) and were found to be normally distributed ($p > 0.05$). Therefore, multiple unpaired two-tailed Student's *t*-tests were used to assess statistical significance across the seed development; Asterisks indicate significant differences *(*P ≤ 0.05)*; n=6 biological replicates.

Significantly elevated C16:0 levels in transgenic pennycress seeds

The analysis of fatty acid composition in the transgenic FLD-9B3 revealed distinct accumulation patterns when compared to the wild-type (WT) pennycress. The transgenic line exhibited significant accumulation of C10:0, of 24.17 mol% at 10 days after flowering (DAF), before declining progressively to 5.74 mol% at seed maturation. In contrast, the wild type showed no detectable accumulation of C10:0. While C16:0 was present in both genotypes, its accumulation was slightly higher in the wild type at 10 DAF compared to the transgenic FLD-9B3, but displayed a decline in its accumulation, with the wild type (WT) achieving 3.71 mol% at maturity, while the transgenic line remained higher at 10.63 mol% at maturity. In terms of the long and very long chain fatty acids, including those of C18:1, C18:2, C20:1, and C22:1, transgenic FLD-9B3 showed decreased accumulation compared to the wild type (WT). Additionally, fatty acids such as C18:3 showed no significant differences in transgenic FLD-9B3 compared to the wild type (WT) across all seven time points (Figure 2.6).

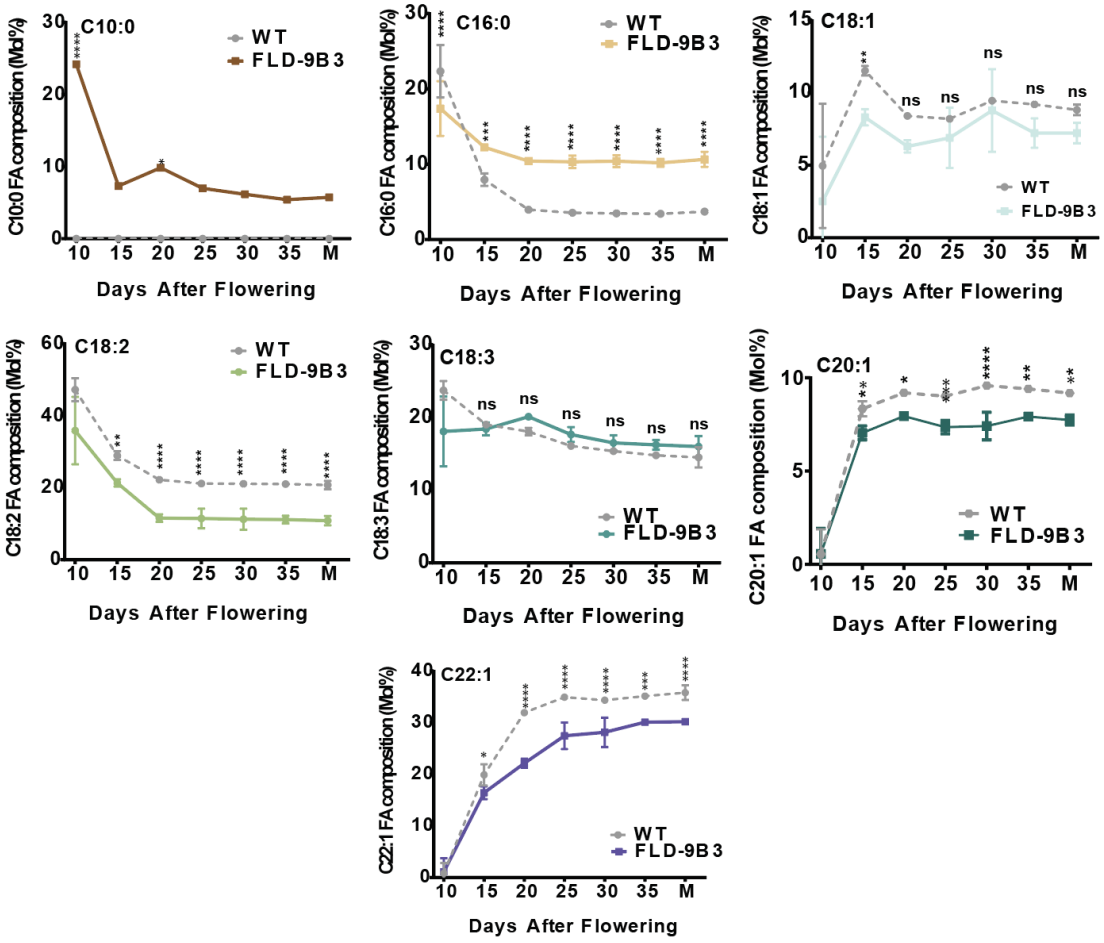

Figure 2.6 Transgenic pennycress lines possess altered fatty acid composition. Shown are individual fatty acid profiles of transgenic pennycress FLD-9B3 and Wild-type (WT). Data values are the average mean ± standard deviation. Asterisks indicate statistical differences from wild-type seeds (Sidak's multiple comparisons test; **P< 0.05; ***P< 0.001; ****P<0.0001*); n=6 biological replicates across seven seed developmental stages.

2.4 Discussion

This study involves the initial characterization of transgenic pennycress lines to produce medium-chain fatty acids (MCFAs), particularly C10:0, through overexpression of *CvFatB1*, *CvLPAT2*, and *CpuDGAT1* from *Cuphea*. The FLD-9B3 line showed the highest MCFA accumulation, peaking at 24.17 mol% C10:0 during early seed development and maintaining ~5.74 mol% at maturity levels. Despite this achievement, mature seeds displayed lower C10:0 content compared to native *Cuphea*, highlighting potential metabolic or regulatory bottlenecks limiting optimal MCFA accumulation. Notably, FLD-9B3 seeds maintained normal morphology and viability, suggesting no adverse developmental effects. The altered lipid profile was characterized by decreased total fatty acid content, elevated C16:0, and reduced LCFAs/VLCFAs, suggesting a redirection of metabolic fluxes away from MCFA production, which may hinder optimal C10:0 production. One likely explanation is metabolic competition with native LCFA and VLCFA synthesis pathways. These findings underscore the challenge of overcoming native pathway dominance in heterologous metabolic engineering. Further, Chapter 3 describes the systems biology approach we are implementing to uncover the potential limitations and explore the rational design of strategies to optimize MCFA accumulation in transgenic pennycress as a viable platform for metabolic engineering of oil seeds, enabling the production of industrially valuable fatty acids.

Chapter 3- Transcriptome and Lipidome Dynamics in Transgenic Pennycress Seeds

3.1 Introduction

Plant oils represent one of the most economically and nutritionally significant products that are primarily stored in seeds in the form of triacylglycerol (TAG) (Yang et al., 2022). The economic and nutritional value of plant oil is commonly assessed based on the content of TAGs and fatty acid composition. TAGs serve as a key energy source for germination and seedling growth, while also playing a crucial role in diet and serving as raw materials for various industrial applications such as detergents and lubricants (Yang et al., 2022). Most native plant species naturally produce and store TAGs with a set of regular fatty acids (C ≥16:0 with up to 3 *cis* double bonds) but no medium-chain fatty acids (MCFAs), which are of great industrial and economic importance. Some of the rich sources of these medium-chain fatty acids (MCFAs) are coconut (*Cocos nucifera*) and several Cuphea species, including *Cuphea viscosissima (Cv), Cuphea var. avigera pulcherrima (Cpu),* naturally accumulating ≥90 mol% of C10:0 levels. Despite the potential of these species to produce MCFAs, their poor agronomic traits limit their ability, making them unfit to grow in agricultural lands. To overcome this limitation, three seed-specific enzymes employed for synthesis, incorporation, and accumulation of MCFAs, namely *CvFatB1, CvLPAT2*, and *CpuDGAT1*, were transformed in emerging oilseed crops like Pennycress (*Thlaspi arvense L.*) to genetically engineer the production of these MCFAs (Esfahanian et al., 2021). However, the best engineered pennycress line selected was only able to yield 6.3 mol% of C10:0 levels, which was relatively extremely lower than (≥90 mol%) the C10:0 levels in *Cuphea*.

Additionally, as described in Chapter 2, transgenic lines revealed a reduction in the total fatty acid content profile and significantly elevated C16:0 levels compared to the wild-type (WT).

Therefore, to further address these bottlenecks that limit C10:0 accumulation in transgenic pennycress seeds, it is essential to gain a deep understanding of the metabolic pathways governing oil biosynthesis and regulation in plants. A comprehensive strategy involving targeted lipidomic and transcriptomic analyses of overexpressed *CvFatB1+CvLPAT2+CpuDGAT1* transgenic seeds across seven seed developmental stages was investigated. The prime objective was to gain insights into the lipid profiles of 10-carbon containing TAGs accumulation pattern across the development and further look into the gene expression levels to identify the potential targets to explore strategies for C10:0 level enhancement in these transgenic seeds.

3.2 Materials and Methods

Lipid analysis

Lipids were extracted from approximately 30 mg of freshly ground developing seed tissue using a one-step tissue extraction method (Shiva et al., 2018). Briefly, the ground tissue was placed in a 4-mL glass vial with a flat bottom made of Teflon (PTFE), and 0.4 mL of hot isopropanol containing 0.01% butylated hydroxytoluene (BHT) was added. The vial was sealed and maintained at approximately 75°C on a heat block for 15 minutes, ensuring the preservation of tissue integrity and prevention of lipolytic activity. After cooling to room temperature, 1.2 mL of extraction solvent (chloroform-methanol-water, 30:41.5:3.5, v/v/v) was added, and the mixture was shaken on an orbital shaker at 50-100 rpm for 24 hours. The resulting lipid extract was then transferred to a new 2.0 mL vial and stored at -80°C until further analysis. The prepared samples were analyzed

using a triple quadrupole mass spectrometer equipped with an electrospray ionization source (Sciex Triple Quad 6500+ System), according to methods described in Shiva et al. (2013).

RNA extraction

The isolation of RNA was followed by the removal of genomic DNA contamination from the ground seed tissue of both wild-type and transgenic pennycress, using the PureLink™ Plant RNA Reagent kit from Invitrogen, which was conducted by the laboratory of Dr. Cahoon (University of Nebraska). RNA quality was assessed with the Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA), ensuring that all samples had an RNA Integrity Number (RIN) of ≥8. Extracted RNA samples were stored at -80ºC.

RNA-seq data filtering and analysis

The RNA-seq analysis was performed by the Department of Biomedical Informatics, Biostatistics and Medical Epidemiology (BBME), University of Missouri. All the pennycress RNA-seq FASTQ files were assessed for quality using the FASTQC tool. Based on the FASTQC results, no further trimming was performed on the sequencing data. The RNA-seq data were then aligned against three different references, including the PacBio Hifi Iso-seq transcripts, the Trinity De Novo assembled transcripts, and the *Thlaspi arvense v1.1* reference genome from the Phytozome web portal, using HISAT2 separately. Gene expression quantification to compute FPKM (Fragments per Kilobase of Transcript per million mapped reads) values for each gene and each reference was carried out by Cufflinks. For differential gene expression analyses, Cuffdiff was utilized to compare gene expression across different time points of C10 pennycress, Wild-type (WT) pennycress, and between C10 and wild-type pennycress at the same points, separately for each reference. This process was extended with four additional *Cuphea* genes into each reference, which

were realigned against all the C10 pennycress samples. Similarly, Cufflinks and Cuffdiff were also utilized on the C10 pennycress data with extended references. Additionally, expression differences in C10 pennycress across various time points were quantified. Furthermore, minor corrections were applied to the summed FPKM values for each gene in both the treatment and control groups, such as adjusting any FPKM with a value of zero by adding 1×10^6 to avoid issues with log2 fold changes. These adjustments also required recalculating the log2 fold changes to prevent infinite values in the results.

Plot generation and statistical analysis

Principal component analysis (PCA), correlation heatmap, and student t-test were performed using MetaboAnalyst 6.0 (https://www.metaboanalyst.ca/) (Pang *et al.*,2024). Lipids with an FDR (adjusted *P*-value) ≤ 0.05 and fold change ≥ 2 were considered as differentially accumulated (DA). Statistical significance was determined using two-way ANOVA followed by Sidak's multiple comparison test.

3.3 Results

Transgenic seeds possess a distinct lipid profile

Lipidomic analysis was performed to comprehensively assess lipid alterations associated with the transgenic line FLD-9B3. This analysis aimed to detect and quantify a broad spectrum of lipid species across multiple classes, providing insights into lipid metabolism changes linked to the transgenic modification. To evaluate the overall variation in lipid profiles between the transgenic pennycress seeds and the wild-type (WT) during seed development, a 2D-Principal Component Analysis (PCA) was performed (Figure 3.1). The first two principal components, PC1 and PC2, accounted for 42.2% and 24.3% of the total variance. PC1 distinguishes transgenic samples in the earlier stages of seed development, with minimal separation after 20 DAF. PC2 separates transgenic and WT timepoints, but with less separation in the later stages of seed development. Both genotypes possess similar lipid profiles early in development at 10 days after flowering (DAF). However, as seed development progresses, the lipid profiles of the two genotypes diverge, indicating a distinct lipid profile towards seed maturation.

Further, triacylglycerol (TAG) species containing at least one MCFA had a strong positive loading on PC1, suggesting it is a major factor separating FLD from the WT. In contrast, Long Chain (LCFAs) and Very Long Chain Fatty Acids (VLCFAs) induced triacylglycerol (TAG) species had a strong positive loading on PC2, demonstrating accumulation of LCFA & VLCFA during seed development in both WT and transgenic seeds (Figure 3.1).

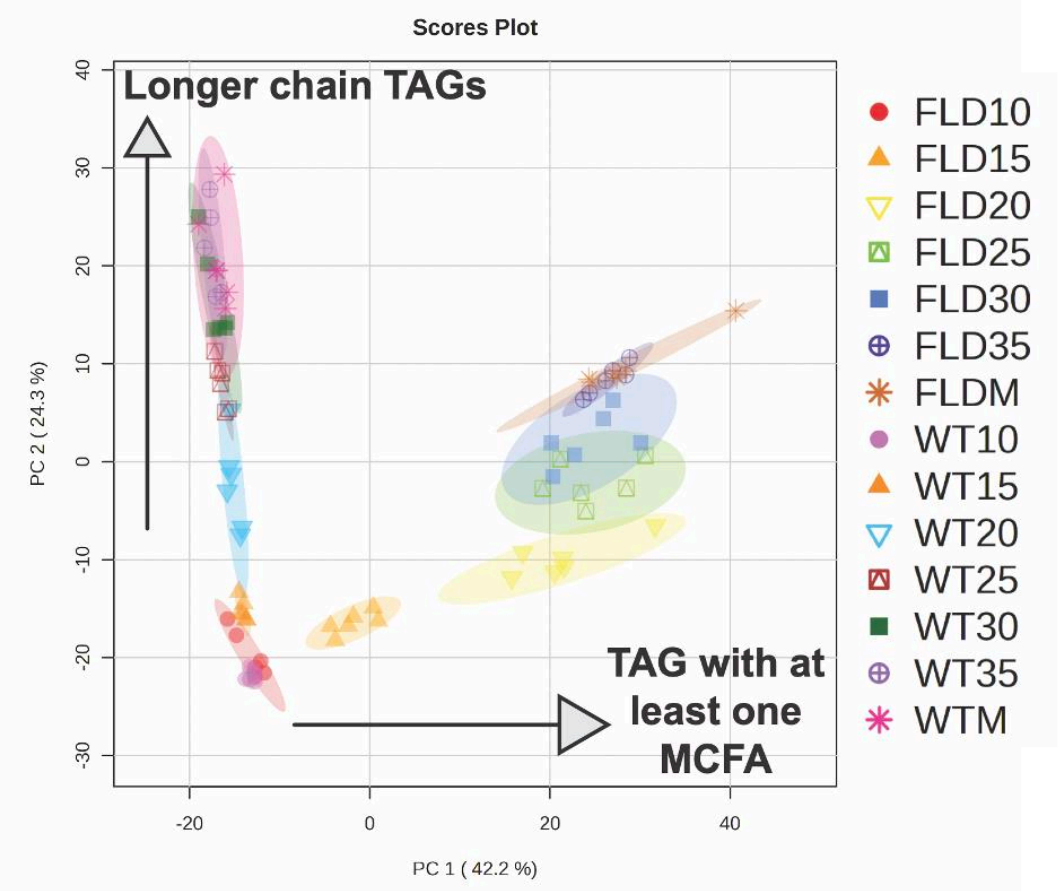

Figure 3.1 Transgenic seeds possess a distinct lipidomic profile from wild-type across seed development. 2D-PCA plot showing variation in fatty acid composition over seven time points in FLD-9B3 and wild-type (WT) plants. Each point represents a biological replicate (n = 6 per genotype per time point), and samples are color-coded by genotype and time. Data was auto-scaled before analysis. Ellipses represent 95% confidence intervals for each group. Labels: DAF, Days After Flowering; WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature.

Accumulation of altered C10:0 TAGs and some of the ASGs in transgenic seeds

Hierarchical clustering of the lipidomics data reveals a distinct lipid accumulation pattern between FLD and the wild type (Figure 3.2). Specifically, transgenic pennycress (FLD) demonstrates a notable accumulation of triacylglycerols (TAGs) containing medium-chain fatty acids (MCFAs), which include C10:0, which peaks on the 20th day after flowering (DAF) during seed development. In contrast, these MCFA-containing TAGs are absent in the wild-type (WT). Additionally, triacylglycerols (TAGs) containing long-chain (LCFAs) and very long-chain fatty acids (VLCFAs), including C18:1, C18:2, C18:3, C20:0, and C22:1, are more abundant in the wild-type, especially during the mid-late and late stages of seed development, compared to the transgenic pennycress (FLD) (Figure 3.2).

Furthermore, the accumulation of C16:0-enriched TAGs exhibits a notable increase in transgenic FLD, where C16:0 progressively increases from the 20th DAF toward maturity. This trend contrasts with very few C16:0-containing TAGs in the WT (Figure 3.3). Previously, a similar accumulation pattern for individual C16:0 fatty acid was also observed in the fatty acid content analysis (Figure 2.6).

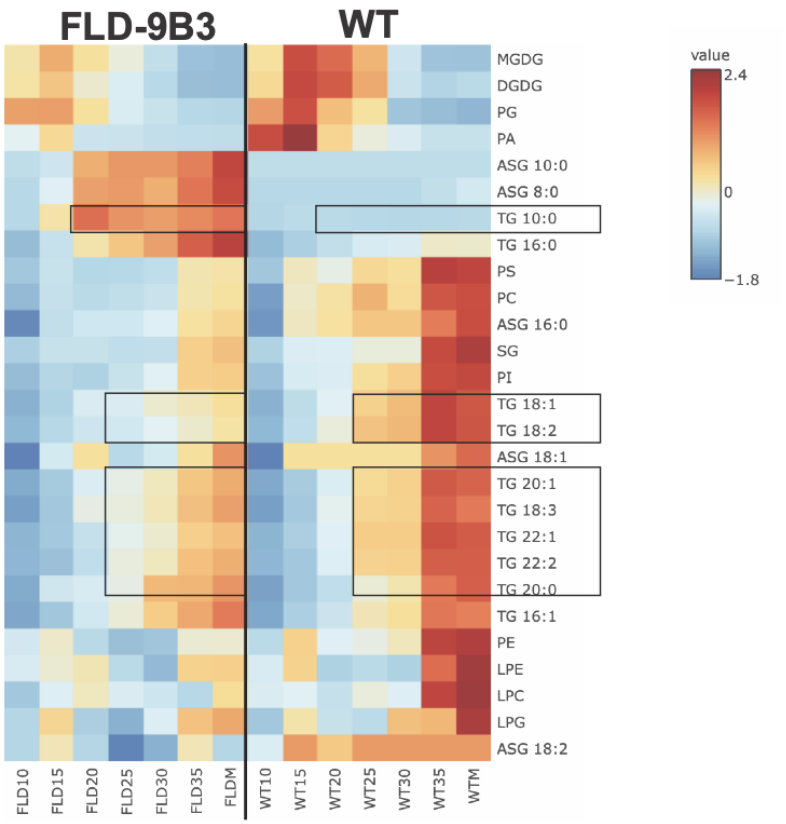

Figure 3.2 Lipid Accumulation Dynamics of transgenic seeds accumulating altered 10-carbon TAGs. Hierarchical clustering heatmap analysis illustrates long-chain TAGs and C10:0 variants. The lipid species from the two genotype groups are presented and ranked by t-tests, with normalization done based on auto-scaling. Values are measured by Euclidean distance with a Ward clustering algorithm (n=6 biological replicates). Abbreviations: WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature; MGDG, monogalactosyldiacylglycerol; DGDG, di galactosyl diacylglycerol; PG, phosphatidylglycerol; PA, phosphatidic acid; ASG x:x, x-carbon linked Acyl steryl glucosides, TG x:x, x-carbon linked triglyceride; PS, phosphatidylserine; PC, phosphatidylcholine; SG, sterol glucosides; PI, phosphatidylinositol; PE, phosphatidylethanolamine; LPE, Lyso phosphatidylethanolamine; LPC, Lyso phosphatidylcholine; LPG, Lys phosphatidylglycerol.

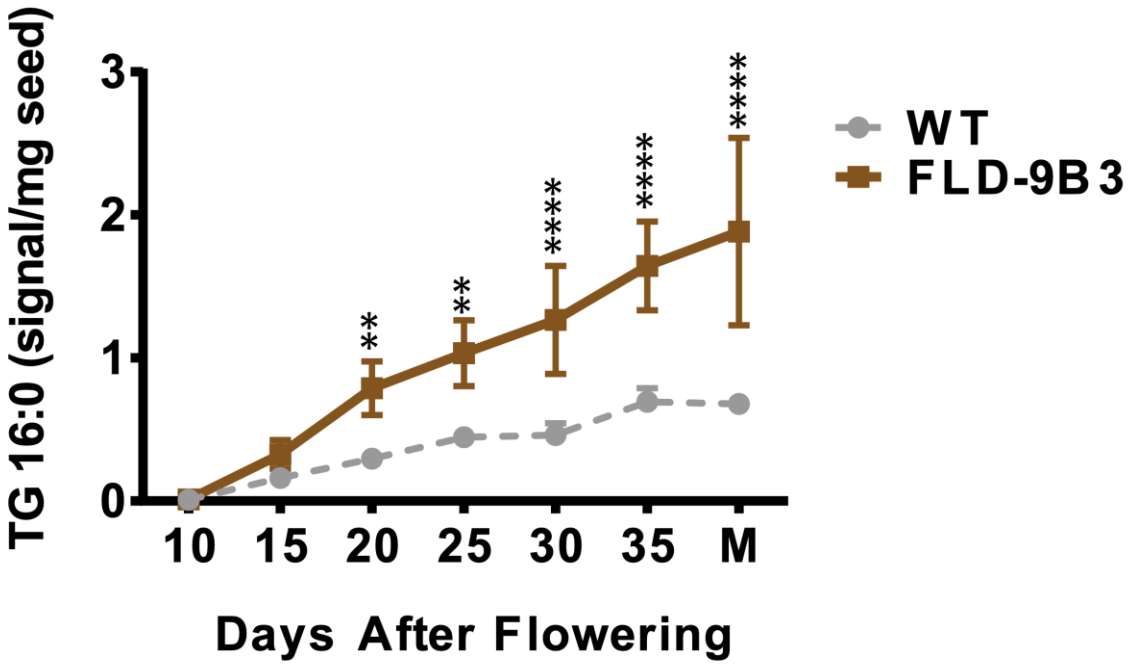

Figure 3.3 Transgenic FLD-9B3 possesses significantly elevated TG 16:0 across the seed development. Levels of TAG molecular species containing 16:0 were measured using ESI-MS. Data is expressed in signal/mg seed. Data represent mean ± standard deviation, n=3 biological replicates. Asterisks indicate statistical differences from wild-type seeds. (Sidak's multiple comparison test; $**P \leq 0.01$; $****P \leq 0.0001$); WT, Wild type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature; n=6 biological replicates.

Both short-chain (C8:0) and medium-chain (C10:0) acyl-steryl glucosides (ASGs) accumulate in transgenic pennycress (FLD-9B3). The levels of these modified ASGs increase/peak from the 20th DAF through to maturity, mirroring the accumulation pattern seen for medium-chain altered TAGs (Figure 3.4.a & b).

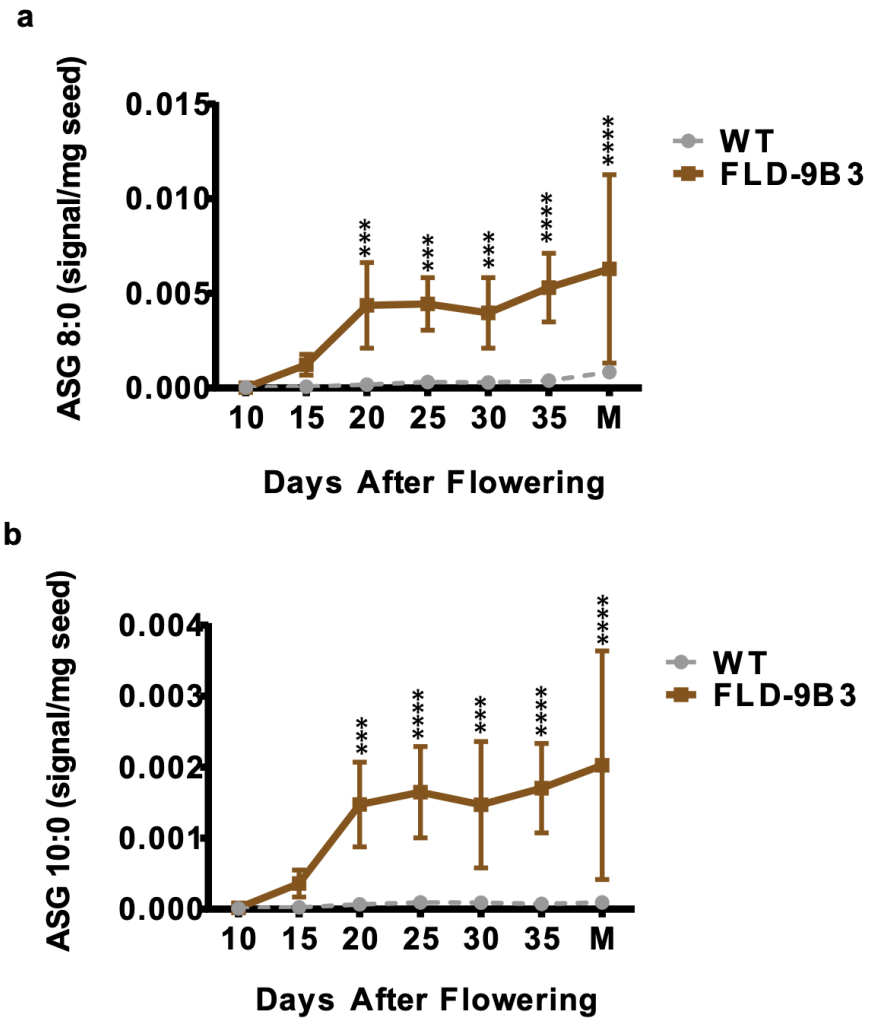

Figure 3.4 Transgenic FLD-9B3 accumulates altered Acyl Sterol Glycosides (ASGs) across the seed development. Levels of ASG 8:0 (a) and ASG 10:0 (b) were measured using ESI-MS. Data is expressed in signal/mg seed. Expressed in signal/mg seed vs. DAF. Data represent mean ± standard deviation, n=3 biological replicates. Asterisks indicate statistical differences from wild-type seeds. (Sidak's multiple comparison test; ****P*≤*0.001;* *****P*≤ *0.0001*); WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature; n=6 biological replicates.

TAG molecular species containing multiple C10:0 peaks during the seed development

TAG10:0 containing lipids with multiple medium-chain fatty acids (MCFAs), such as TAG 10:0_20:0, TG 10:0_22:0, TG 10:0_24:0, TG 10:0_26:0, TG 10:0_28:1, TG 10:0_30:1, and TG 10:0_32:1 peaks at 20 DAF and then goes down as the seeds mature. In contrast, TAG10:0, containing lipids with one C10:0, including TG 10:0_34:1, TG 10:0_36:1, TG 10:0_38:1, TAG 10:0_40:2, and TG 10:0_42:2, displays an increasing trend in accumulation from mid to late stages as the seed development progresses. In summary, these findings suggest that TG 10:0_x:x lipids with multiple medium-chain fatty acids (MCFAs) tend to decrease after peaking at 20 DAF, whereas those containing no or at least one medium-chain fatty acid (MCFA) increase as the seeds mature in transgenic development (Figure 3.5).

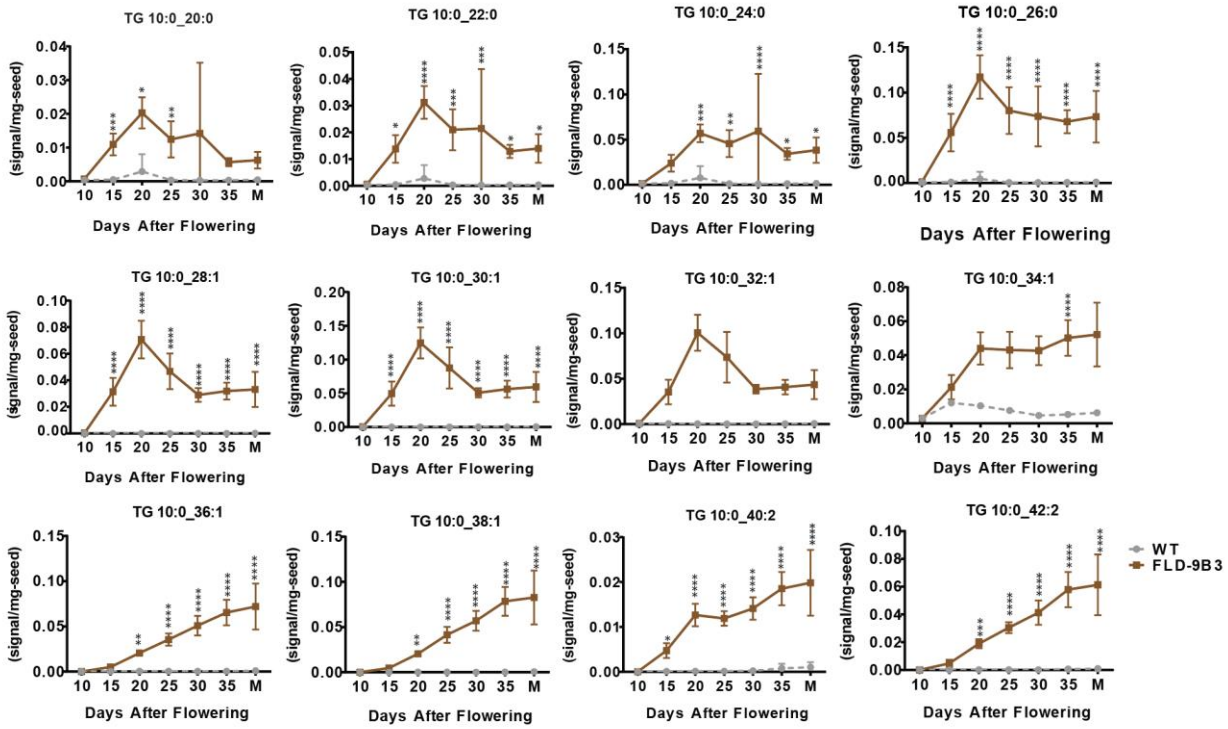

Figure 3.5 TAG 10:0 lipid species containing multiple medium chains decrease relative to those with no or at least one MCFA throughout the seed development. Expressed as signal/ mg-seed at Days After Flowering. Data is presented for two genotypes (FLD-9B3 and WT) at 7 DAFs, with values representing the mean ± standard deviation. Data represent mean ± standard deviation, n=3 biological replicates. Asterisks indicate statistical differences from wild-type seeds. (Sidak's multiple comparison test; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; ****$P \leq 0.0001$); WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature; n=6 biological replicates.

Transcriptomics reveals similar gene expression between transgenic and WT

Transcriptomic profiling was conducted to evaluate global gene expression changes in the FLD-9B3 transgenic line compared to wild-type (WT) pennycress across developmental stages. The analysis revealed broadly similar expression patterns, particularly within core lipid metabolic pathways, while highlighting specific transcriptional shifts related to stress responses and regulatory networks. Transcriptomics data revealed a consistent gene expression pattern throughout seed development in both genotypes. In contrast, lipid profiles exhibited distinct variations across all seven developmental stages. Notably, the final two stages, 35 DAF and maturity, showed closely clustered data points, indicating similar gene expression patterns at these stages. The first two principal components, PC1 and PC2, accounted for 64.6% and 21.6% of the total variance, respectively (Figure 3.6).

These findings demonstrate dynamic changes in gene expression throughout seed development, with the highest number of Differentially Expressed Genes (DEGs) observed at mid-developmental stages (15–20 DAF). The gradual decline in DEGs toward maturity suggests a convergence of transcriptional activity between FLD and WT lines, indicating that early and mid-stages are critical windows for transcriptional divergence in the transgenic seeds (Figure 3.7).

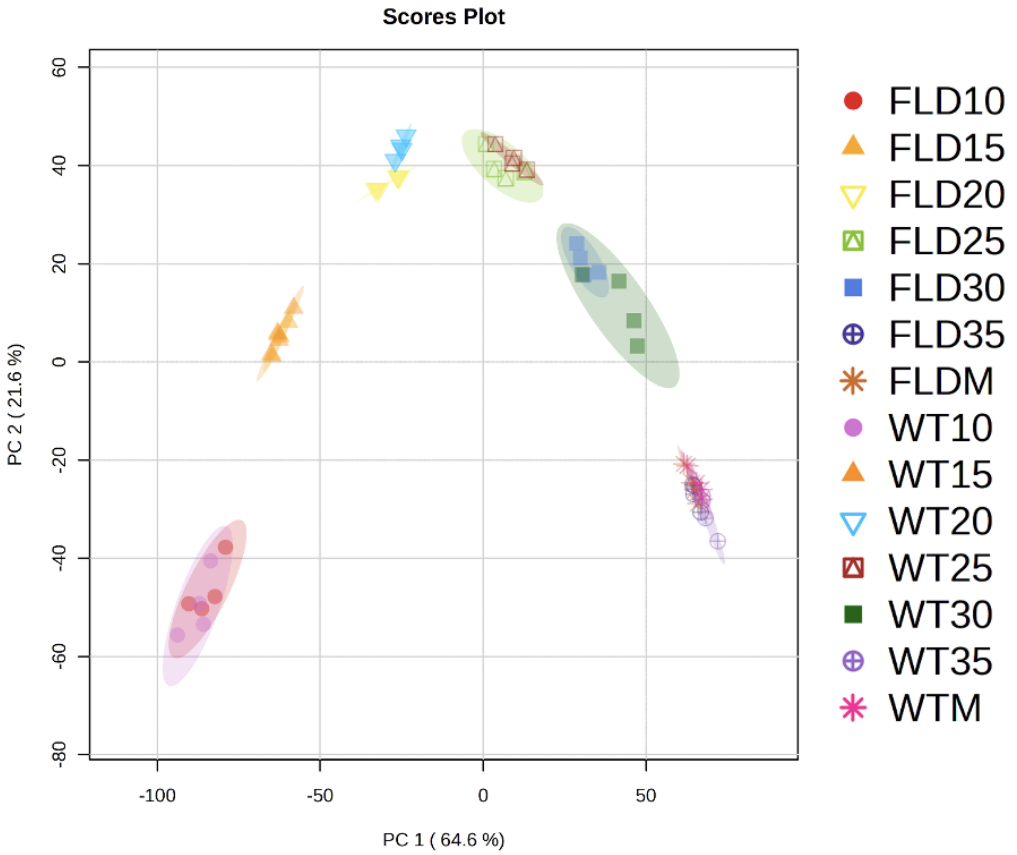

Figure 3.6 2D-PCA plot demonstrates a similar gene expression pattern in wild-type and transgenic seeds. 2D-PCA plot showing variation in fatty acid composition over seven time points in FLD-9B3 and wild-type (WT) plants. Each point represents a biological replicate (n = 6 per genotype per time point), and samples are color-coded by genotype and time. Data was auto-scaled before analysis. Ellipses represent 95% confidence intervals for each group. Labels: DAF, Days After Flowering; WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature.

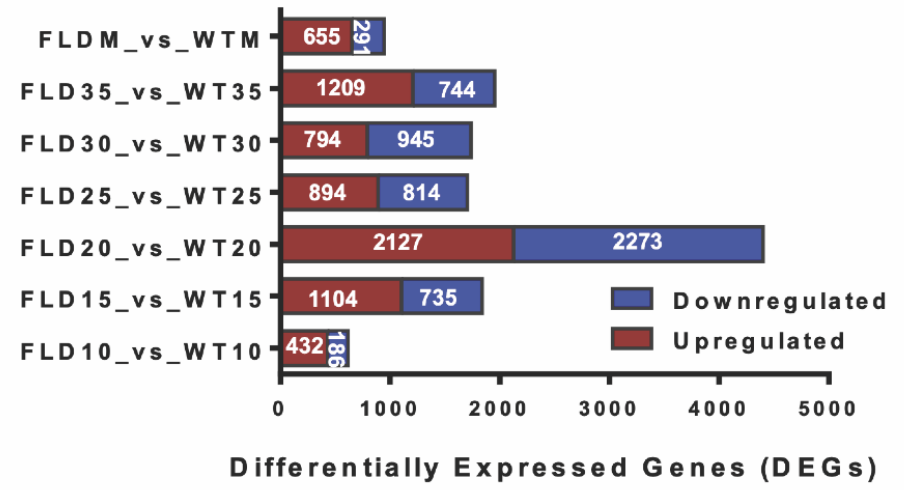

Figure 3.7 Identification of differentially expressed genes (DEGs) across the seed development. Total number of transgenic FLD-9B3 seeds DEGs relative to the wild-type (WT) at each developmental stage. False Discovery Rate (FDR) ≤ 0.05 and fold change ≥2. The x numerical value linked with (FLDx_vs_WTx) indicates the respective seed development stage—WT, Wild type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1*.

Significant changes in the gene expression pattern in transgenic seeds

We investigated key genes involved in various metabolic pathways essential for the synthesis, incorporation, and accumulation of fatty acids in transgenic pennycress seeds. These pathways include fatty acid synthesis (FAS), TAG synthesis, degradation, and peroxisomal β-oxidation pathways. Transcriptomics data show that most genes involved indicate no significant differences in the gene expression levels in FLD-9B3 compared to WT across the seed development (Figure 3.8 b-h). Further comparison showed that *Cuphea viscosissima* transgene FATB (*CvFatB1*) is expressed at approximately 100-fold higher levels compared to the endogenous FATB homolog.

(Figure 3.8a). Fatty-acyl desaturases, such as FAD2 and FAD3, show no significant differences observed between the two genotypes, FLD-9B3 and the WT, across the seed development (Figure 3.8i & j).

Transcriptomic data show that enzymes, including GPAT, LPAT, PAP, and DGAT, showed no significant differences between FLD-9B3 and WT across the seed development (Figure 3.8k-n). Additionally, comparison showed that *Cuphea viscosissima* transgene LPAT2 (*CvLPAT2*), was approximately 10-fold higher compared to the endogenous LPAT2 homolog (Figure 3.8l), whereas, when *Cuphea viscosissima* transgene DGAT1 (*CpuDGAT1*) was overexpressed in pennycress, its levels depict an approximately 100-fold higher expression compared to the endogenous DGAT1 homolog (Figure 3.8n). The expression of *WRINKLED1* (WRI1) shows no significant gene expression difference in transgenic FLD-9B3 compared to the wild-type (WT) (Figure 3.8o).

After TAG formation, its degradation occurs in lipid droplets (LDs), where lipases come into action, which hydrolyze ester bonds between glycerol and fatty acids. TAG lipases (TAGLs) play a pivotal role in breaking down TAGs into diacylglycerol (DAG) and free fatty acids (FFAs). After TAG and DAG breakdown, monoacylglycerol (MAG) is further hydrolyzed by monoacylglycerol lipase (MAGL). Both TAGLs and MAGLs show no differences in gene expression levels in the FLD line compared to WT (Figures 3.8p &q).

Similar results were observed for the peroxisomal β-oxidation pathway, where enzymes such as Acyl-CoA oxidases (ACXs) and Enoyl-CoA hydratase 2 (ECH2) both show no differences in gene expression throughout the development (Figures 3.8r &s). However, Isomerase Enoyl-CoA 2 (Isom-ECI2) shows an upregulation in the transgenic line at most time points, except at 35 DAF and maturity (Figure 3.8t).

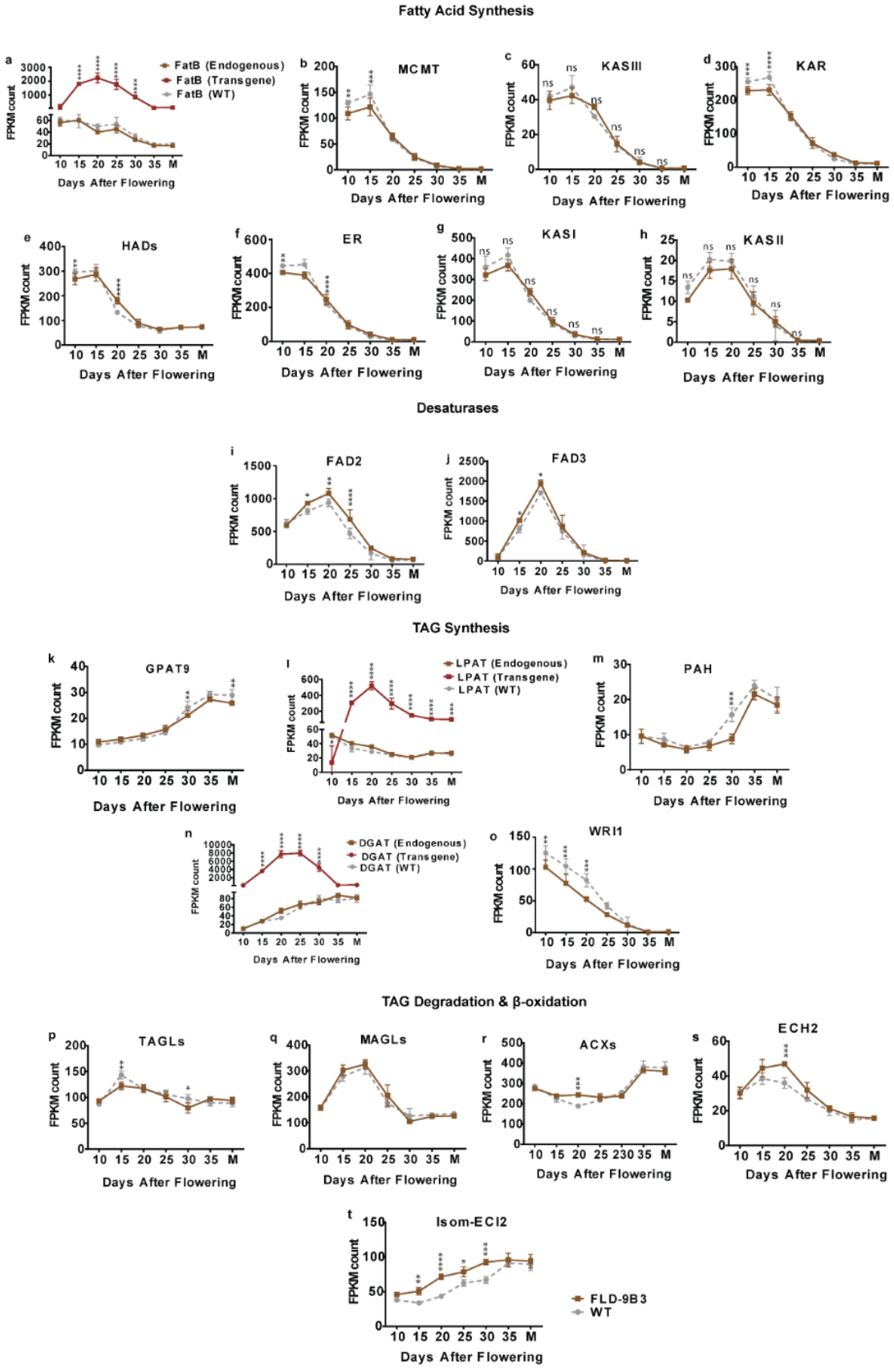
Fatty Acid Synthesis
a
FatB (Endogenous)
FatB (Transgene)
FatB (WT)
FPKM count
Days After Flowering
b MCMT
c KASIII
d KAR
e HADs
f ER
g KASI
h KASII
Desaturases
i FAD2
j FAD3
TAG Synthesis
k GPAT9
l
LPAT (Endogenous)
LPAT (Transgene)
LPAT (WT)
m PAH
n
DGAT (Endogenous)
DGAT (Transgene)
DGAT (WT)
o WRI1
TAG Degradation & β-oxidation
p TAGLs
q MAGLs
r ACXs
s ECH2
t Isom-ECI2
FLD-9B3
WT

Figure 3.8 Expression profile of lipid-related genes during seed development. The y-axis represents fragments per kilobase of transcript per million reads mapped (FPKM) values, while the x-axis represents Days After Flowering (DAF). Genes associated with Fatty Acid Synthesis & TAG synthesis, Desaturation, TAG degradation & β-oxidation are shown. For some enzymes, multiple isoforms of the same genes are summed together. Values are represented as the mean ± standard deviation. Asterisks indicate statistical differences from wild-type seeds. (Sidak's multiple comparison test; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; ****$P \leq 0.0001$); WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature; n=4 biological replicates. Abbreviations: (KASII, KAS1 & KASIII)- keto-acyl synthetases (I, II & III); FATB- fatty acyl thioesterase B; MCMT- Malonyl-CoA: ACP Malonyl transferase; KAR- 3-Ketoacyl-ACP Reductase; HAD - Hydroxyacyl-ACP Dehydratase; ER - Enoyl-ACP Reductase; FAD2 & 3- Fatty Acyl Desaturases 2 & 3; GPAT9- glycerol-3-phosphate acyltransferase 9; LPAT- lysophosphatidic acid acyltransferase; DGAT1- diacyl glycerol acyltransferase 1; WRI1- wrinkled 1 phenotype; TAGL- triacyl glycerol lipases; MAGL- monoacyl glycerol lipases; ACXs- acyl-CoA oxidases; ECH2- enoyl-CoA hydratase 2; Isom-ECI2- enoyl-CoA delta isomerase 2; FPKM, fragments per kilobase of transcript per million mapped reads count.

The top differentially expressed upregulated genes across the seed development

Further analysis of the transcriptomic data identified the top ten differentially expressed genes (DEGs) over all time points. One of the DEGs, HSP22, a small heat shock protein, exhibited significant upregulation as the seed matured. Another heat shock protein, HSP23.5, is a member of the Low Molecular Weight (LMW) HSP family. It was found to be significantly upregulated in

transgenic pennycress across all time points in FLD-9B3 compared to the WT. Additionally, HSP70-3, another cytosolic heat shock protein, was significantly upregulated at all time points except at the 10th DAF in the transgenic plants (Figure 3.9a-c).

The expression of two hypoxia-responsive unknown proteins (HUP32 and HUP54). HUP32 showed significant upregulation in the transgenic FLD-9B3 line at 10th, 15th, 20th, 25th, and 30th DAF, whereas HUP54 was significantly upregulated at 15th, 20th, 25th, 30th, and 35th DAF. No significant changes were observed at the 10th DAF and maturity (Figure 3.9d &e).

Another DEG identified was LBD41, a Lateral Organ Boundary (LOB) domain-containing transcription factor, which plays a key role in regulating seed development, dormancy, and germination (Joshi et al., 2023), was significantly upregulated at early and the mid stages of the development in FLD-9B3 compared to WT, with no significant difference observed at later developmental stages (Figure 3.9f). DJC66, a chloroplast-localized J protein from the DnaJ heat shock protein family, is involved in protein folding and stress responses. This gene was significantly upregulated from early to mid-stages, with no significant differences observed in the later seed developmental stages (Figure 3.9g). WIP4 (Wound-Induced Polypeptide 4) is a member of the wound-induced protein family, typically upregulated in response to mechanical injury or environmental stress in plants, and was significantly upregulated from early to mid-stages in FLD-9B3 compared to the WT, with no significant differences observed at later stages of the development (Figure 3.9h). Additionally, expression profiles of a hypothetical protein, which was significantly upregulated from 10th to 30th DAF, and a transmembrane protein, which was upregulated at the 25th and 30th DAF through to maturity. (Figure 3.9 i&j).

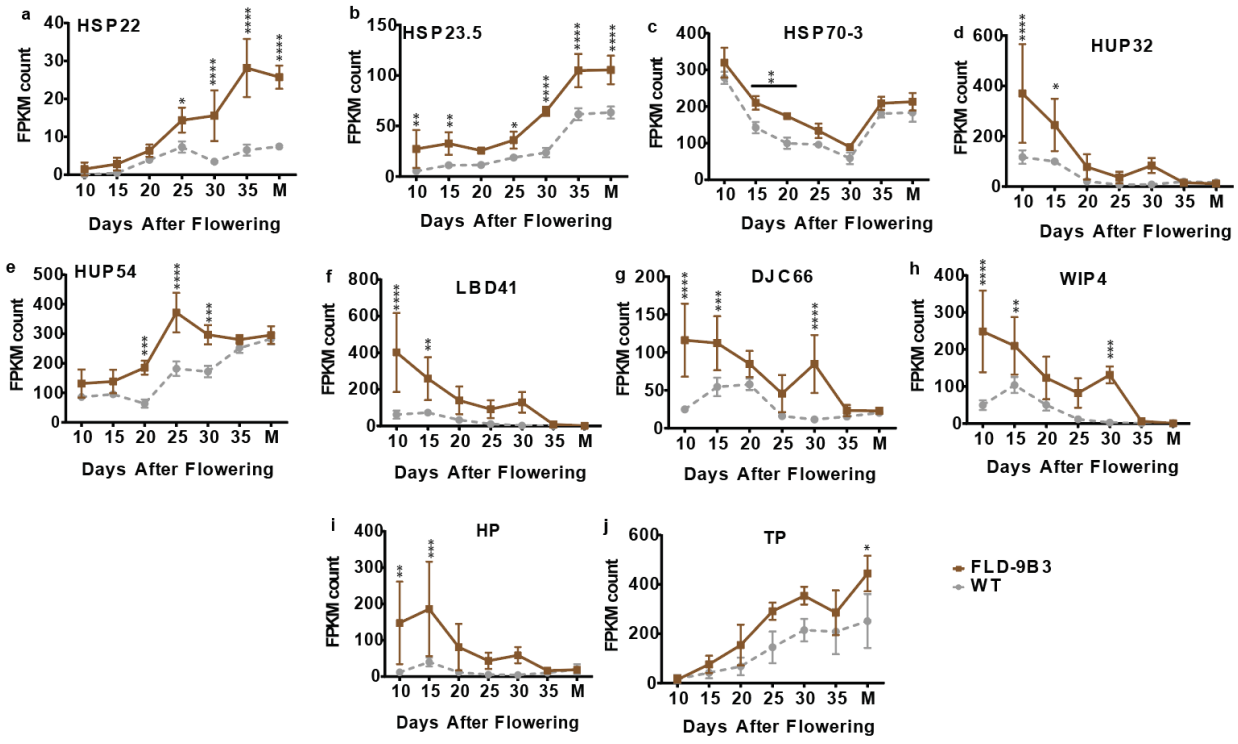

Figure 3.9 Differentially expressed genes (DEGs) in developing transgenic seeds. The y-axis represents gene expression based on RNA-Seq analysis, expressed as Fragments per Kilobase of transcript per million reads mapped (FPKM) values. The x-axis indicates days after flowering (DAF). The figures are listed as: Three heat shock proteins (HSP22, HSP23.5 & HSP70-3); (a-c); Two hypoxia response proteins (HUP32 & HUP54); (d & e); LBD41, LOB Domain-containing protein 41(f); DJC66, DNA J protein C66, a chaperone from the DnaJ-domain superfamily (g); WIP4, Wound-Induced polypeptide 4 (h); hypothetical protein (HP) (i); Transmembrane protein (TP) (j). Data represent the mean ± standard deviation. Asterisks indicate statistical differences from wild-type seeds. (Sidak's multiple comparison test; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; ****$P \leq 0.0001$); WT, Wild-type; FLD, *CvFatB1+CvLPAT2+CpuDGAT1;* M, Mature; n=4 biological replicates.

3.4 Discussion

The integrated lipidomic and transcriptomic analysis of FLD-9B3 transgenic pennycress seeds revealed distinct metabolic profiles across seed development. The lipid profile of FLD-9B3 seeds provided valuable insights into the lipid metabolic changes associated with the introduction of *Cuphea* transgenes. The results highlighted significant alterations in lipid profiles between the transgenic line and the wild-type (WT) throughout seed development, indicating that the FLD-9B3 undergoes distinct lipid biosynthesis and accumulation patterns across the development. The 2D-Principal Component Analysis (PCA) revealed a distinct lipid profile between FLD and WT, except at the early 10 DAF, with the first two principal components (PC1 and PC2) accounting for over 66% of the total variance in lipid distribution across seven time points. Additionally, the strong positive loading of triacylglycerol (TAG) species containing more than one medium-chain fatty acid (MCFAs) on PC1 indicates that these lipid species are more abundant in FLD. In contrast, the strong positive loading of TAG species containing long-chain fatty acids (LCFAs) and very long-chain fatty acids (VLCFAs) on PC2 indicates that these lipid species are more abundant in the WT, particularly during the mid and late stages of seed development.

The observation that FLD-9B3 seeds accumulate fewer LCFAs and VLCFAs suggests that the three-gene transgenic overexpression in pennycress resulted in shifting the metabolic flux balance away from LCFAs and VLCFAs enriched TAGs towards medium-chain containing TAGs (Figure 3.1). Hierarchical clustering underscored a clear divergence in TAG lipid profiles between FLD-9B3 and wild-type (WT), particularly during the mid to late developmental stages, depicting accumulating altered 10-carbon TAGs that peak at 20DAF and then the level goes down towards seed maturity in transgenic seeds. In contrast, the TAGs containing LCFAs and VLCFAs

accumulated more in the WT during the late mid to late stages of development in WT, which are not abundant in the FLD-9B3 line (Figure 3.2). The accumulation pattern of C16:0-enriched TAGs in FLD-9B3 further emphasizes the distinct lipid profile of the FLD-9B3 line. Similar observations were made in the fatty acid content analysis (Figure 2.6), where individual C16:0 fatty acid was significantly higher in FLD-9B3 compared to WT, suggesting that these three transgene modifications may influence the C16:0 level metabolism (Figure 3.3).

Furthermore, another metabolic insight observed was the accumulation of short-chain (C8:0) and medium-chain (C10:0) acyl-steryl glucosides (ASGs) in FLD-9B3. These acyl steryl glucosides (ASGs), the acylated form of steryl glucosides (SGs), consist of a sterol attached to a glucose molecule, which is further modified by an acyl group (a fatty acid attached via an ester bond). Specifically, the 6'-hydroxyl group of the glucose moiety in the steryl glucoside is esterified with a fatty acid. This acylation is carried out by steryl glycoside acyl transferase (SGAT). While the specific acyltransferase is responsible for the acyl transfer, it is still not fully identified (Wojciechowski et al., 1975) (Figure 3.4a &b).

Additionally, we observed distinct trends in the accumulation of TAG species with varying fatty acid compositions during seed development. Specifically, TAGs containing more than one C10:0 chain exhibited a peak in accumulation at 20 DAF, followed by a decline towards maturation. In contrast, TAGs with a single C10:0 fatty acid showed an increasing trend in their accumulation from mid to the late stages of seed development. This pattern indicates a progressive accumulation of TAGs with multiple MCFAs at 20 DAF, which later declines towards maturity (Figure 3.5). Transcriptomic data demonstrated overall conservation in core lipid pathway gene expression, despite the extremely high overexpression of *Cuphea* transgenes. When the *Cuphea viscosissima* transgene FATB (*CvFatB1*) was overexpressed in pennycress, its expression levels were found to

be approximately 100-fold higher than those of the endogenous FATB homolog (Figure 3.8a). Similarly, the overexpression of the *Cuphea viscosissima* transgene LPAT2 (*CvLPAT2*) in pennycress resulted in expression levels approximately 10-fold higher than the endogenous LPAT2 homolog (Figure 3.8l). Likewise, the overexpression of the *Cuphea var. avigera pulcherrima* transgene DGAT1 (*CpuDGAT1*) in pennycress led to expression levels approximately 100-fold higher than those of the endogenous DGAT1 homolog (Figure 3.8n). The gene expressions of transgenes FatB and DGAT1 are both 100-fold higher compared to the native pennycress genes, but transgene LPAT has only 10-fold higher expression due to the moderate strength oleosin promoter being used for LPAT, in contrast to the strong glycinin-1 promoter used for *CvFatB1* and *CpuDGAT1* (Esfahanian et al., 2021).

To investigate global gene expression changes in the FLD-9B3 line relative to the WT pennycress, a 2D-PCA plot highlighted broadly similar gene expression profiles between the FLD-9B3 and the WT across all the developmental stages, with data points for the later stages of the seed development, including 35 DAF and maturity, clustered closely together, with PC1 and PC2 accounted for 64.6% and 21.6% of the total variance, respectively (Figure 3.6). This is further observed through the enzymes involved in the important lipid biosynthesis, including fatty acid synthesis (FAS), TAG synthesis, TAG degradation, desaturases, and β-oxidation, suggesting that there are no significant gene expression changes observed between the transgenic FLD-9B3 and WT across the seed development (Figure 3.8). In further analysis of the transcriptomics data, we identified the top DEGs, and noticed that there was the upregulation of stress-responsive genes, including various heat shock proteins HSPs (HSP22, HSP23.5 and HSP70-3), some hypoxia-responsive unknown proteins, HUPs (HUP32 and HUP54), and transcriptional regulators like LBD41 and DJC66, and WIP4 (Wound-Induced Polypeptide 4), a protein linked to mechanical

injury and environmental stress responses, indicates a cellular stress response potentially triggered by extremely high expression of the three foreign transgenes and metabolic imbalances (Figure 3.9). As no significant differences were observed at the transcriptome level, further investigation is needed to look at the proteome level of expression. These findings collectively suggest that FLD-9B3 undergoes some physiological stress, emphasizing the importance of understanding the interplay between engineered lipid metabolism and stress adaptation in developing stable, high-oil-yielding seeds.

Chapter 4- Conclusion and Future Work

Conclusion

To enhance the production of value-added fatty acids, specifically MCFAs, in pennycress, a biotechnological approach was undertaken to introduce genes from non-model plants, such as several *Cuphea* species, including *Cuphea viscosissima* (*Cv*) and *Cuphea avigera var. pulcherrima* (*Cpu*), known to naturally produce ≥90 mol% MCFAs (Graham and Kleiman, 1992). These saturated fatty acids, typically containing 8- to 14-carbon chains, are industrially relevant due to their utility in commercial products such as cosmetics, soaps, detergents, surfactants, and lubricants (Knaut and Richtler, 1985). This research aimed to enhance the biosynthesis of industrially relevant medium-chain fatty acids (MCFAs), particularly C10:0, in pennycress (*Thlaspi arvense*) by leveraging genes from non-model *Cuphea* species known for their natural production of ≥90 mol% MCFAs. Using a biotechnological approach, transgenes *CvFatB1, CvLPAT2, and CpuDGAT1* were introduced and overexpressed in pennycress. Among the various transgenic lines characterized, the highest C10:0 accumulating line yielded approximately 6.3 mol% in mature seeds. Thus, the important question of this research is why C10:0 levels in transgenic pennycress are relatively low compared to the naturally occurring *Cuphea* species.

The main research goal is to conduct a systems biology approach to investigate and identify the bottlenecks responsible for limiting these C10:0 levels in transgenic pennycress seeds. Developmental stage-specific analysis revealed that C10:0 levels peaked at 24.17 mol% at 10 DAF, followed by a decline to ~5.74 mol% at maturity, demonstrating dynamic changes in MCFA accumulation. In contrast, WT did not accumulate any C10:0 fatty acid. An important finding was the significantly elevated C16:0 levels observed in the FLD-9B3 compared to the WT.

Additionally, the total amount of fatty acid content was reduced in transgenic seeds. Later on, another important finding showed that TAG species containing more than one C10:0 chain peaked at 20DAF and then declined through to maturity. In contrast, the TAG species containing at least one C10:0 chain progressively increased throughout development.

The lipidome and transcriptome both exhibited lipid and gene expression profiles quite different from one another across seed development when observed via a 2D-PCA plot, with a distinct lipid profile but a similar gene expression pattern seen across the seven time points during the seed development. Transcriptomics data revealed no significant gene expression changes between FLD-9B3 and WT. However, the most important finding revealed that the introduction of transgenes including *CvFatB1, CvLPAT2, and CpuDGAT1,* isolated from *Cuphea viscosissima* (*Cv*) and *Cuphea var. avigera pulcherrima* (*Cpu*), when overexpressed in pennycress, showed extremely higher gene expression levels, depicting approximately 100, 10, and 100-fold compared to the endogenous pennycress gene homologs. Moreover, the transcriptomic analysis identified several stress-related genes, particularly heat shock proteins (HSPs), hypoxia-responsive proteins (HUPs), and other transcriptional regulators, LBD41 and DJC66, along with another WIP4, which were all upregulated in FLD-9B3 plants, indicating some potential cellular stress the plants might be under due to the overexpression of all three introduced transgenes. These findings suggest that while overexpression of *Cuphea*-derived genes can successfully redirect fatty acid synthesis toward MCFA production in pennycress, metabolic flux is limited by both developmental constraints and possible stress responses triggered by transgene overexpression.

Future Work

Collectively, this study investigated the key metabolic and regulatory constraints limiting the accumulation of medium-chain fatty acids (MCFAs), particularly decanoic acid (C10:0), in transgenic pennycress engineered with a three-gene construct comprising *CvFatB1*, *CvLPAT2*, and *CpuDGAT1*. These transgenes, isolated from various *Cuphea* species, such as *Cuphea viscosissima* (Cv) and *Cuphea avigera var. pulcherrima* (Cpu), are known to naturally accumulate MCFAs at levels exceeding 90 mol%. However, when expressed in pennycress, the transgenic line FLD-9B3 accumulated approximately 6.3 mol% C10:0, which remains substantially lower than the levels observed in native *Cuphea* species. To gain a deeper understanding of this disparity, a systems biology approach was utilized to identify the underlying metabolic bottlenecks that limit C10:0 production in transgenic pennycress seeds. Building upon the results obtained in this study, future research will focus on a more detailed examination of the remaining omics analyses, including proteomics.

As the current research is constrained by relatively low C10:0 levels, future investigations could build upon these findings by utilizing RNA interference (RNAi) technology to selectively knock down the expression of three key endogenous genes in pennycress, namely FATB, LPAT2, and DGAT1, which are integral to lipid biosynthesis. RNAi constructs targeting these genes will be designed to disrupt their expression, and these constructs will be introduced into the plant system via the *Agrobacterium*-mediated transformation method. Following successful transformation, transgenic plants will be screened for RNAi expression, and the knockdown efficiency will be confirmed through qRT-PCR and Western blot analysis to assess reductions in both mRNA and protein levels. This strategy aims to more effectively redirect metabolic flux towards medium-chain fatty acid (MCFA) production, potentially increasing C10:0 levels to

match those found in *Cuphea* species (≥90 mol%), while concurrently minimizing the accumulation of long-chain fatty acids (LCFAs) and very long-chain fatty acids (VLCFAs).

Previous studies have shown that the complete knockout of LPAT2 (*lpat2-1*) results in lethality, causing embryo death. However, suppression of LPAT2 led to decreased phospholipid levels in both leaves and flowers, smaller seed sizes, and no significant effect on total seed triacylglycerol (TAG) content, although the fatty acid composition of TAGs was altered, with reduced levels of C18:1 in the suppressed lines (Barroga et al., 2022). In contrast, an Arabidopsis mutant with a T-DNA insertion in the FATB gene exhibited reduced palmitate (C16:0) content in glycolipids across various plant tissues, particularly a 56% reduction in seeds. This mutation resulted in slower growth rates and deformed seeds with low viability (Bonaventure et al., 2003). Additionally, knockout of DGAT1 in Arabidopsis does not result in lethality but significantly impairs seed oil accumulation by approximately 20-40% (Zhang et al., 2009).

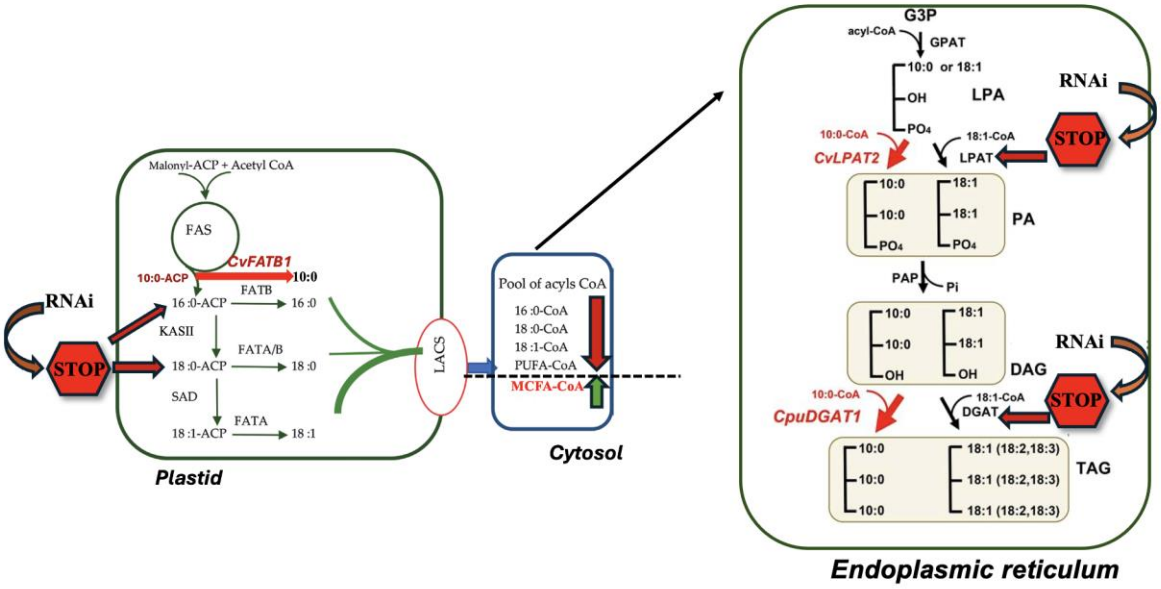

Figure 4. 1 Proposed remodeling of endogenous pennycress lipid metabolism to enhance MCFA accumulation.

An additional strategy under consideration is to express the transgenes *CvFatB1, CvLPAT2, and CpuDGAT1* in pennycress under the control of weaker seed-specific promoters, replacing the current seed-specific soybean Glycinin-1 promoter, which was initially employed to mitigate potential toxicity in vegetative tissues (Esfahanian et al., 2021). This modification is intended to facilitate more refined regulation of transgenic expression in pennycress, with the expectation that it will prevent the upregulation of stress-related genes and other transcriptional regulators associated with stress responses. Ultimately, this approach aims to enhance C10:0 accumulation in the transgenic seeds.

In parallel, phenotypic assessments including seed germination rate, cotyledon emergence, seed size, and weight will be conducted to ensure that the genetic modifications do not adversely impact essential seed traits.

References

- **Alberts, A. W., & Vagelos, P. R. (1968).** Acetyl-CoA carboxylase. I. Requirement for two protein fractions. *Proceedings of the National Academy of Sciences*, *59*(2), 561-568.
- **Angkawijaya, A. E., Nguyen, V. C., & Nakamura, Y. (2017).** Enhanced root growth in phosphate-starved Arabidopsis by stimulating de novo phospholipid biosynthesis through the overexpression of LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE 2 (LPAT2). *Plant, Cell & Environment*, *40*(9), 1807-1818.
- **Angkawijaya, A. E., Nguyen, V. C., & Nakamura, Y. (2019).** LYSOPHOSPHATIDIC ACID ACYLTRANSFERASES 4 and 5 are involved in glycerolipid metabolism and nitrogen starvation response in Arabidopsis. *New Phytologist*, *224*(1), 336-351.
- **Aoyama, T., Nosaka, N., & Kasai, M. (2007).** Research on the nutritional characteristics of medium-chain fatty acids. *The Journal of Medical Investigation*, *54*(3, 4), 385-388.
- **Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. M., & Somerville, C. R. (1992).** Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. *Science*, *258*(5086), 1353-1355.
- **Arroyo-Caro, J. M., Chileh, T., Kazachkov, M., Zou, J., Alonso, D. L., & García-Maroto, F. (2013).** The multigene family of lysophosphatidate acyltransferase (LPAT)-related enzymes in Ricinus communis. Cloning and molecular characterization of two LPAT genes that are expressed in castor seeds. *Plant Science*, *199*, 29-40.
- **Avato, P., & Tava, A. (2022).** Rare fatty acids and lipids in plant oilseeds: Occurrence and bioactivity. *Phytochemistry Reviews*, *21*(2), 401-428.

- **Aznar-Moreno, J. A., & Durrett, T. P. (2017).** Simultaneous targeting of multiple gene homologs to alter seed oil production in Camelina sativa. *Plant and Cell Physiology*, *58*(7), 1260-1267.
- **Babayan, V. K. (1981).** Medium chain length fatty acid esters and their medical and nutritional applications. *Journal of the American Oil Chemists' Society*, *58*(1), 49A-51A.
- **Badami, R. C., & Patil, K. B. (1980).** Structure and occurrence of unusual fatty acids in minor seed oils. *Progress in lipid research*, *19*(3-4), 119-153.
- **Banaś, W., Sanchez Garcia, A., Banaś, A., & Stymne, S. (2013).** Activities of acyl-CoA: diacylglycerol acyltransferase (DGAT) and phospholipid: diacylglycerol acyltransferase (PDAT) in microsomal preparations of developing sunflower and safflower seeds. *Planta*, *237*, 1627-1636.
- **Barroga, N. A. M., & Nakamura, Y. (2022).** LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE 2 (LPAT2) is required for de novo glycerolipid biosynthesis, growth, and development in vegetative and reproductive tissues of Arabidopsis. *The Plant Journal*, *112*(3), 709-721.
- **Bates, P. D., Durrett, T. P., Ohlrogge, J. B., & Pollard, M. (2009).** Analysis of acyl fluxes through multiple pathways of triacylglycerol synthesis in developing soybean embryos. *Plant Physiology*, *150*(1), 55-72.
- **Bates, P. D., Fatihi, A., Snapp, A. R., Carlsson, A. S., Browse, J., & Lu, C. (2012).** Acyl editing and headgroup exchange are the major mechanisms that direct polyunsaturated fatty acid flux into triacylglycerols. *Plant physiology*, *160*(3), 1530-1539.

- **Bates, P. D., Stymne, S., & Ohlrogge, J. (2013).** Biochemical pathways in seed oil synthesis. *Current opinion in plant biology*, *16*(3), 358-364.
- **Bonaventure, G., Salas, J. J., Pollard, M. R., & Ohlrogge, J. B. (2003).** Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. *The Plant Cell*, *15*(4), 1020-1033.
- **Brown, A. P., Carnaby, S., Brough, C., Brazier, M., & Slabas, A. R. (2002).** Limnanthes douglasii lysophosphatidic acid acyltransferases: immunological quantification, acyl selectivity and functional replacement of the Escherichia coli plsC gene. *Biochemical Journal*, *364*(3), 795-805.
- **Cahoon, E. B., & Li-Beisson, Y. (2020).** Plant unusual fatty acids: learning from the less common. *Current opinion in plant biology*, *55*, 66-73.
- **Cao, Y. Z., Oo, K. C., & Huang, A. H. (1990).** Lysophosphatidate acyltransferase in the microsomes from maturing seeds of meadowfoam (Limnanthes alba). *Plant physiology*, *94*(3), 1199-1206.
- **Carman, G. M., & Han, G. S. (2018).** Phosphatidate phosphatase regulates membrane phospholipid synthesis via phosphatidylserine synthase. *Advances in biological regulation*, *67*, 49-58.
- **Chapman, K. D., & Ohlrogge, J. B. (2012).** Compartmentation of triacylglycerol accumulation in plants. *Journal of Biological Chemistry*, *287*(4), 2288-2294.
- **Chen, S., Lei, Y., Xu, X., Huang, J., Jiang, H., Wang, J., ... & Li, Y. (2015).** The peanut (Arachis hypogaea L.) gene AhLPAT2 increases the lipid content of transgenic Arabidopsis seeds. *PLoS One*, *10*(8), e0136170.

- **Chen, A. (2022).** *Enzymology of Acyl Carrier Protein-Dependent Synthases in Escherichia coli*. University of California, San Diego.
- **Clough, R. C., Matthis, A. L., Barnum, S. R., & Jaworski, J. G. (1992).** Purification and characterization of 3-ketoacyl-acyl carrier protein synthase III from spinach. A condensing enzyme utilizing acetyl-CoA A to initiate fatty acid synthesis. *Journal of Biological Chemistry*, *267*(29), 20992-20998.
- **Cruz-Ramírez, A., Oropeza-Aburto, A., Razo-Hernández, F., Ramírez-Chávez, E., & Herrera-Estrella, L. (2006).** Phospholipase DZ2 plays an important role in extraplastidic galactolipid biosynthesis and phosphate recycling in Arabidopsis roots. *Proceedings of the National Academy of Sciences*, *103*(17), 6765-6770.
- **Dehesh, K., Jones, A., Knutzon, D. S., & Voelker, T. A. (1996).** Production of high levels of 8: 0 and 10: 0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from Cuphea hookeriana. *The Plant Journal*, *9*(2), 167-172.
- **Dorn, K. M., Fankhauser, J. D., Wyse, D. L., & Marks, M. D. (2015).** A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *Dna research*, *22*(2), 121-131.
- **Drenth, A. C., Olsen, D. B., & Denef, K. (2015).** Fuel property quantification of triglyceride blends with an emphasis on industrial oilseeds, camelina, carinata, and pennycress. *Fuel*, *153*, 19-30.
- **Dyer, J. M., Stymne, S., Green, A. G., & Carlsson, A. S. (2008).** High-value oils from plants. *The Plant Journal*, *54*(4), 640-655.
- **Eastmond, P. J., Quettier, A. L., Kroon, J. T., Craddock, C., Adams, N., & Slabas, A. R. (2010).** PHOSPHATIDIC ACID PHOSPHOHYDROLASE 1 and 2 regulate

phospholipid synthesis at the endoplasmic reticulum in Arabidopsis. *The Plant Cell*, *22*(8), 2796-2811.

- **Esfahanian, M., Nazarenus, T. J., Freund, M. M., McIntosh, G., Phippen, W. B., Phippen, M. E., ... & Sedbrook, J. C. (2021).** Generating pennycress (*Thlaspi arvense*) seed triacylglycerols and acetyl-triacylglycerols containing medium-chain fatty acids. *Frontiers in Energy Research*, *9*, 620118.
- **Gesch, R. W., Isbell, T. A., Oblath, E. A., Allen, B. L., Archer, D. W., Brown, J., ... & Vigil, M. F. (2015).** Comparison of several Brassica species in the north central US for potential jet fuel feedstock. *Industrial crops and products*, *75*, 2-7.
- **Graham, S. A., & Knapp, S. J. (1989).** Cuphea: A new plant source of medium-chain fatty acids. *Critical Reviews in Food Science & Nutrition*, *28*(2), 139-173.
- **Graham, S. A., & Kleiman, R. (1992).** Composition of seed oils in some Latin American Cuphea (*Lythraceae*). *Industrial Crops and Products*, *1*(1), 31-34.
- **Graham, S. A., José, G. P. C., Murad, A. M., Rech, E. L., Cavalcanti, T. B., & Inglis, P. W. (2016).** Patterns of fatty acid composition in seed oils of Cuphea, with new records from Brazil and Mexico. *Industrial Crops and Products*, *87*, 379-391.
- **Hanke, C., Wolter, F. P., Coleman, J., Peterek, G., & Frentzen, M. (1995).** A plant acyltransferase involved in triacylglycerol biosynthesis complements an Escherichia coli sn-1-acylglycerol-3-phosphate acyltransferase mutant. *European journal of biochemistry*, *232*(3), 806-810.
- **Iskandarov, U., Silva, J. E., Kim, H. J., Andersson, M., Cahoon, R. E., Mockaitis, K., & Cahoon, E. B. (2017).** A specialized diacylglycerol acyltransferase contributes to the

extreme medium-chain fatty acid content of Cuphea seed oil. *Plant physiology*, *174*(1), 97-109.

- **Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandeau-Nelson, M. D., & Reilly, P. J. (2011).** Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochemistry*, *12*, 1-16.
- **Jones, A., Davies, H. M., & Voelker, T. A. (1995).** Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *The Plant Cell*, *7*(3), 359-371.
- **Joshi, S., Hill, K., Chakrabarti, M., & Perry, S. E. (2023).** Regulatory mechanisms of the LBD40 transcription factor in Arabidopsis thaliana somatic embryogenesis. *Plant Direct*, *7*(12), e547.
- **Jung, H. R., Jeon, J. M., Yi, D. H., Song, H. S., Yang, S. Y., Choi, T. R., ... & Yang, Y. H. (2019).** Poly (3-hydroxybutyrate-co-3-hydroxyvalerate-co-3-hydroxyhexanoate) terpolymer production from volatile fatty acids using engineered Ralstonia eutropha. *International Journal of Biological Macromolecules*, *138*, 370-378.
- **Jaworski, J. G., Clough, R. C., & Barnum, S. R. (1989).** A Cerulenin-Insensitive Short-Chain 3-Ketoacyl-Acyl Carrier Protein Synthase in Spinacia oleracea Leaves. Plant Physiol, 90(1), 41-44.
- **Jessen, D., Roth, C., Wiermer, M., & Fulda, M. (2015).** Two activities of long-chain acyl-coenzyme A synthetase are involved in lipid trafficking between the endoplasmic reticulum and the plastid in Arabidopsis. *Plant Physiology*, *167*(2), 351-366.

- **Kachroo, A., Shanklin, J., Whittle, E., Lapchyk, L., Hildebrand, D., & Kachroo, P. (2007).** The Arabidopsis stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant molecular biology*, *63*, 257-271.
- **Katavic, V., Reed, D. W., Taylor, D. C., Giblin, E. M., Barton, D. L., Zou, J., ... & Kunst, L. (1995).** Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity. *Plant physiology*, *108*(1), 399-409.
- **Kim, H. U., Li, Y., & Huang, A. H. (2005).** Ubiquitous and endoplasmic reticulum–located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in Arabidopsis. *The Plant Cell*, *17*(4), 1073-1089.
- **Kim, H. J., Silva, J. E., Iskandarov, U., Andersson, M., Cahoon, R. E., Mockaitis, K., & Cahoon, E. B. (2015).** Structurally divergent lysophosphatidic acid acyltransferases with high selectivity for saturated medium-chain fatty acids from Cuphea seeds. *The Plant Journal*, *84*(5), 1021-1033.
- **Kim, H. J., Silva, J. E., Vu, H. S., Mockaitis, K., Nam, J. W., & Cahoon, E. B. (2015).** Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. *Journal of Experimental Botany*, *66*(14), 4251-4265.
- **Konishi, T., Shinohara, K., Yamada, K., & Sasaki, Y. (1996).** Acetyl-CoA carboxylase in higher plants: Most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. *Plant and cell physiology*, *37*(2), 117-122.

- **Körbes, A. P., Kulcheski, F. R., Margis, R., Margis-Pinheiro, M., & Turchetto-Zolet, A. C. (2016).** Molecular evolution of the lysophosphatidic acid acyltransferase (LPAAT) gene family. *Molecular Phylogenetics and Evolution*, *96*, 55-69.
- **Knaut, J., & Richtler, H. J. (1985).** Trends in industrial uses of palm and lauric oils. *Journal of the American Oil Chemists' Society*, *62*(2), 317-327.
- **Knutzon, D. S., Lardizabal, K. D., Nelsen, J. S., Bleibaum, J. L., Davies, H. M., & Metz, J. G. (1995).** Cloning of a coconut endosperm cDNA encoding a 1-acyl-sn-glycerol-3-phosphate acyltransferase that accepts medium-chain-length substrates. *Plant physiology*, *109*(3), 999-1006.
- **Knutzon, D. S., Hayes, T. R., Wyrick, A., Xiong, H., Maelor Davies, H., & Voelker, T. A. (1999).** Lysophosphatidic acid acyltransferase from coconut endosperm mediates the insertion of laurate at the sn-2 position of triacylglycerols in lauric rapeseed oil and can increase total laurate levels. *Plant physiology*, *120*(3), 739-746.
- **Kleiman, R. (1990).** Chemistry of new industrial oilseed crops.
- **Lassner, M. W., Levering, C. K., Davies, H. M., & Knutzon, D. S. (1995).** Lysophosphatidic acid acyltransferase from meadowfoam mediates insertion of erucic acid at the sn-2 position of triacylglycerol in transgenic rapeseed oil. *Plant physiology*, *109*(4), 1389-1394.
- **Lessire, R., & Stumpe, P. K. (1983).** Nature of the fatty acid synthetase systems in parenchymal and epidermal cells of Allium porrum L. leaves. *Plant physiology*, *73*(3), 614-618.

- **Li, Y., Beisson, F., Pollard, M., & Ohlrogge, J. (2006).** Oil content of Arabidopsis seeds: the influence of seed anatomy, light, and plant-to-plant variation. *Phytochemistry*, *67*(9), 904-915.
- **Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., ... & Ohlrogge, J. (2013).** Acyl-lipid metabolism. *The Arabidopsis book/American Society of Plant Biologists*, *11*, e0161.
- **Li, N., Gügel, I. L., Giavalisco, P., Zeisler, V., Schreiber, L., Soll, J., & Philippar, K. (2015).** FAX1, a novel membrane protein mediating plastid fatty acid export. *PLoS biology*, *13*(2), e1002053.
- **Lu, C., Xin, Z., Ren, Z., Miquel, M., & Browse, J. (2009).** An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. *Proceedings of the National Academy of Sciences*, *106*(44), 18837-18842.
- **Lung, S. C., & Weselake, R. J. (2006).** Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. *Lipids*, *41*(12), 1073-1088.
- **Lv, Y., Zhang, X., Luo, L., Yang, H., Li, P., Zhang, K., ... & Wan, Y. (2020).** Characterization of glycerol-3-phosphate acyltransferase 9 (AhGPAT9) genes, their allelic polymorphism, and association with oil content in peanut (Arachis hypogaea L.). *Scientific Reports*, *10*(1), 14648.
- **Mayer, K. M., & Shanklin, J. (2007).** Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *BMC Plant Biology*, *7*, 1-11.

- **McGinn, M., Phippen, W. B., Chopra, R., Bansal, S., Jarvis, B. A., Phippen, M. E., ... & Sedbrook, J. C. (2019).** Molecular tools enabling pennycress (Thlaspi arvense) as a model plant and oilseed cash cover crop. *Plant biotechnology journal*, *17*(4), 776-788.
- **Millar, A. A., Smith, M. A., & Kunst, L. (2000).** All fatty acids are not equal: discrimination in plant membrane lipids. *Trends in plant science*, *5*(3), 95-101.
- **Martin, S., & Parton, R. G. (2006).** Lipid droplets: a unified view of a dynamic organelle. *Nature reviews Molecular cell biology*, *7*(5), 373-378.
- **Munnik, T. (2001).** Phosphatidic acid: an emerging plant lipid second messenger. *Trends in plant science*, *6*(5), 227-233.
- **Nakamura, Y., Tsuchiya, M., & Ohta, H. (2007).** Plastidic phosphatidic acid phosphatases are identified in a distinct subfamily of lipid phosphate phosphatases with prokaryotic origin. *Journal of Biological Chemistry*, *282*(39), 29013-29021.
- **Nakamura, Y., Koizumi, R., Shui, G., Shimojima, M., Wenk, M. R., Ito, T., & Ohta, H. (2009).** Arabidopsis lipins mediate the eukaryotic pathway of lipid metabolism and cope critically with phosphate starvation. *Proceedings of the national academy of sciences*, *106*(49), 20978-20983.
- **Ohlrogge, J., & Browse, J. (1995).** Lipid biosynthesis. *The plant cell*, *7*(7), 957.
- **Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., & Browse, J. (1994).** Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *The Plant Cell*, *6*(1), 147-158.
- **Pang, Z., Lu, Y., Zhou, G., Hui, F., Xu, L., Viau, C., ... & Xia, J. (2024).** MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Research*, *52*(W1), W398-W406.

- **Pearce, M. L., & Slabas, A. R. (1998).** Phosphatidate phosphatase from avocado (Persea americana)–purification, substrate specificity and possible metabolic implications for the Kennedy pathway and cell signalling in plants. *The Plant Journal*, *14*(5), 555-564.
- **Pierrugues, O., Brutesco, C., Oshiro, J., Gouy, M., Deveaux, Y., Carman, G. M., ... & Kazmaier, M. (2001).** Lipid Phosphate Phosphatases in Arabidopsis: Regulation of the AtLPP1 gene in response to stress. *Journal of Biological Chemistry*, *276*(23), 20300-20308.
- **Pollard, M. R., Anderson, L., Fan, C., Hawkins, D. J., & Davies, H. M. (1991).** A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of Umbellularia californica. *Archives of biochemistry and biophysics*, *284*(2), 306-312.
- **Regmi, A., Shockey, J., Kotapati, H. K., & Bates, P. D. (2020).** Oil-producing metabolons containing DGAT1 use separate substrate pools from those containing DGAT2 or PDAT. *Plant Physiology*, *184*(2), 720-737.
- **Salas, J. J., & Ohlrogge, J. B. (2002).** Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. *Archives of biochemistry and biophysics*, *403*(1), 25-34.
- **Salie, M. J., & Thelen, J. J. (2016).** Regulation and structure of the heteromeric acetyl-CoA carboxylase. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, *1861*(9), 1207-1213.
- **Sasaki, Y., Konishi, T., & Nagano, Y. (1995).** The compartmentalization of acetyl-coenzyme carboxylase in plants. *Plant physiology*, *108*(2), 445.

- **Sasaki, Y., & Nagano, Y. (2004).** Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Bioscience, biotechnology, and biochemistry*, *68*(6), 1175-1184.
- **Sedbrook, J. C., Phippen, W. B., & Marks, M. D. (2014).** New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (Thlaspi arvense L.). *Plant Science*, *227*, 122-132.
- **Shiva, S., Vu, H. S., Roth, M. R., Zhou, Z., Marepally, S. R., Nune, D. S., ... & Welti, R. (2013).** Lipidomic analysis of plant membrane lipids by direct infusion tandem mass spectrometry. *Plant lipid signaling protocols*, 79-91.
- **Shiva, S., Enninful, R., Roth, M. R., Tamura, P., Jagadish, K., & Welti, R. (2018).** An efficient modified method for plant leaf lipid extraction results in improved recovery of phosphatidic acid. *Plant Methods*, *14*, 1-8.
- **Shockey, J., Regmi, A., Cotton, K., Adhikari, N., Browse, J., & Bates, P. D. (2016).** Identification of Arabidopsis GPAT9 (At5g60620) as an essential gene involved in triacylglycerol biosynthesis. *Plant physiology*, *170*(1), 163-179.
- **Shockey, J. M., Fulda, M. S., & Browse, J. A. (2002).** Arabidopsis contains nine long-chain acyl-coenzyme A synthetase genes that participate in fatty acid and glycerolipid metabolism. *Plant physiology*, *129*(4), 1710-1722.
- **Singer, S. D., Chen, G., Mietkiewska, E., Tomasi, P., Jayawardhane, K., Dyer, J. M., & Weselake, R. J. (2016).** Arabidopsis GPAT9 contributes to the synthesis of intracellular glycerolipids but not surface lipids. *Journal of Experimental Botany*, *67*(15), 4627-4638.
- **Slack, C. R., Campbell, L. C., Browse, J. A., & Roughan, P. G. (1983).** Some evidence for the reversibility of the choline phosphotransferase-catalyzed reaction in developing

linseed cotyledons in vivo. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, *754*(1), 10-20.

- **Slack, C. R., Roughan, P. G., Browse, J. A., & Gardiner, S. E. (1985).** Some properties of choline phosphotransferase from developing safflower cotyledons. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, *833*(3), 438-448.
- **Steinbusch, K. J., Hamelers, H. V., Plugge, C. M., & Buisman, C. J. (2011).** Biological formation of caproate and caprylate from acetate: fuel and chemical production from low-grade biomass. *Energy & Environmental Science*, *4*(1), 216-224.
- **Voelker, T. A., Worrell, A. C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D. J., ... & Davies, H. M. (1992).** Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science*, *257*(5066), 72-74.
- **Wang, X., Devaiah, S. P., Zhang, W., & Welti, R. (2006).** Signaling functions of phosphatidic acid. *Progress in lipid research*, *45*(3), 250-278.
- **Wang, J., Wang, X., Li, J., Chen, Y., Yang, W., & Zhang, L. (2015).** Effects of dietary coconut oil as a medium-chain fatty acid source on performance, carcass composition, and serum lipids in male broilers. *Asian-Australasian journal of animal sciences*, *28*(2), 223.
- **Weiss, S. B., Kennedy, E. P., & Kiyasu, J. Y. (1960).** The enzymatic synthesis of triglycerides. *The Journal of Biological Chemistry*, *235*, 40-44.

- **Wojciechowski, Z. A., & Zimowski, J. (1975).** Acyl composition and biosynthesis of acylated steryl glucosides in Calendula officinalis. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, *398*(1), 111-117.
- **Wu, G. Z., & Xue, H. W. (2010).** Arabidopsis β-ketoacyl-acyl carrier protein synthase I is crucial for fatty acid synthesis and plays a role in chloroplast division and embryo development. *The Plant Cell*, *22*(11), 3726-3744.
- **Xiang, M. S., Tan, J. K., & Macia, L. (2019).** Fatty acids, gut bacteria, and immune cell function. *The molecular nutrition of fats*, 151-164.
- **Xu, C., & Shanklin, J. (2016).** Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. *Annual review of plant biology*, *67*(1), 179-206.
- **Yamaoka, Y., Achard, D., Jang, S., Legéret, B., Kamisuki, S., Ko, D., ... & Lee, Y. (2016).** Identification of a Chlamydomonas plastidial 2-lysophosphatidic acid acyltransferase and its use to engineer microalgae with increased oil content. *Plant biotechnology journal*, *14*(11), 2158-2167.
- **Yang, Y., & Benning, C. (2018).** Functions of triacylglycerols during plant development and stress. *Current Opinion in Biotechnology*, *49*, 191-198.
- **Yang, F., Zhang, B., Chen, B., Yang, N., Wang, R., Zhang, X., & Li, G. (2022).** A lipidomic approach for profiling and distinguishing seed oils of Hibiscus manihot L., flaxseed, and sunflower oil. *Journal of the American Oil Chemists' Society*, *99*(3), 203-213.
- **Zhang, M., Fan, J., Taylor, D. C., & Ohlrogge, J. B. (2009).** DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis

and are essential for normal pollen and seed development. *The Plant Cell*, *21*(12), 3885-3901.

- **Zhao, H., Kosma, D. K., & Lü, S. (2021).** Functional role of long-chain acyl-CoA synthetases in plant development and stress responses. *Frontiers in Plant Science*, *12*, 640996.
- **Zhou, X. R., Shrestha, P., Yin, F., Petrie, J. R., & Singh, S. P. (2013).** AtDGAT2 is a functional acyl-CoA: diacylglycerol acyltransferase and displays different acyl-CoA substrate preferences than AtDGAT1. *FEBS letters*, *587*(15), 2371-2376.
- **Zhou, X. R., Bhandari, S., Johnson, B. S., Kotapati, H. K., Allen, D. K., Vanhercke, T., & Bates, P. D. (2020).** Reorganization of acyl flux through the lipid metabolic network in oil-accumulating tobacco leaves. *Plant Physiology*, *182*(2), 739-755.