CRISPR-Cas9 mediated gene editing to enhance cover crops

by

Dexter White

B.S., Kansas State University, 2023

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
Dr. Timothy P. Durrett

Copyright

Abstract

Cover crops provide environmental benefits and a potential source of sustainable oil, but their uncertain economic viability remains a challenge. Specialty oils, valuable even at lower production volumes, are often derived from species with poor agronomic traits. *Camelina sativa* L. Crantz (camelina) and *Thlaspi arvense* L. (pennycress) are oilseed cover crops amenable to metabolic engineering and genetic transformation. Increasing the value of the oil produced in these crops through metabolic engineering addresses a key barrier to implementation and could encourage higher cover-cropping rates among farmers. Efficiently redirecting carbon flux towards engineered triacylglycerols (TAGs) requires minimizing competition from endogenous TAG biosynthesis pathways. We hypothesize that disrupting native *Diacylglycerol O-acyltransferase 1* (DGAT1) and *Phospholipid:diacylglycerol acyltransferase 1* (PDAT1), the terminal enzymes in TAG biosynthesis, would enhance the accumulation of specialty TAGs produced by transgenic acyltransferases. Eliminating endogenous DGAT1 and PDAT1 activity effectively channels diacylglycerol (DAG) towards transgenic TAG-synthesizing enzymes, boosting specialty TAG production. Previous work demonstrated the potential of RNAi-mediated *DGAT1* knockdown to increase specialty-TAG production (Alkotami et al. 2024). However, RNAi-based gene silencing can exhibit variable knockdown efficiency and residual enzyme activity remains. Using CRISPR-Cas9, we generated *dgat1* and *pdat1* knockout lines in both camelina and pennycress. CRISPR-Cas9 knockouts offer a genetically stable solution for long-term metabolic engineering by eliminating enzyme activity. These results demonstrate a scalable and stable genetic engineering strategy to enhance the commercial potential of cover crops by tailoring seed oil composition for diverse industrial applications.

Table of Contents

List of Abbreviations

ASP – Amplicon Size Polymorphism
CaMV 35S – Cauliflower Mosaic Virus 35S
Cas9 – CRISPR associated protein 9
CDS – Coding DNA Sequence
CRC – Cruciferin C
CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa – CRISPR activation
CRISPRi – CRISPR interference
DAG – Diacylglycerol
dCas9 – deactivated Cas9
ddPCR – droplet digital PCR
DGAT1 – Diacylglycerol O-Acyltransferase 1
DGAT3 – Diacylglycerol O-Acyltransferase 3
dsDNA – double-stranded DNA
DsRed – *Discosoma* Red
EaDAcT – *Euonymous alatus* Diacylglycerol Acetyl Transferase
EfDAcT – *Euonymous fortunei* Diacylglycerol Acetyl Transferase
EMS – Ethyl methanesulfonate
ER – Endoplasmic Reticulum
FAD2 – Fatty-Acid Desaturase 2
FAE1 – Fatty-Acid Elongase1
FAME- Fatty Acid Methyl Ester
GC – Gas Chromatography
GDSL1 – Glycine Aspartate Serine Leucine 1
GFP – Green Fluorescent Protein
gRNA – guide RNA
HDR – Homology Directed Repair
MAG – Monoacylglycerol
MFCA – Medium Chain Fatty Acid
MFE – Minimum-Free Energy
NGS – Next Generation Sequencing
NHEJ – Non-Homologous End Joining
PAM – Protospacer Adjacent Motif
PC – Phosphatidylcholine
PDAT1 – Phospholipid:diacylglycerol acyltransferase 1
PUFA – Polyunsaturated Fatty Acid
RNAi – RNA interference
RNP – Ribonucleoprotein Complexes

SAF – Sustainable Aviation Fuel
sgRNA – synthetic guide RNA
SOM – Soil Organic Matter
ssDNA – single stranded DNA
SV – Structural Variant
TAG – Triacylglycerol
TT8 – Transparent Testa 8
VLCFA – Very-Long Chain Fatty Acid

List of Figures

List of Tables

Acknowledgements

I am grateful for the guidance from my advisor, committee members, and mentors: Dr. Timothy Durrett, Dr. Kathrin Schrick, Dr. Ruth Welti, and Dr. Lawrence Davis.

I would like to thank my lab associates for their help and support: Alisha, Dr. Dechassa Duressa, Shelby Innes, Dave Manning, Dr. Nicholas Neumann, Pamela Tamura, Aytuğ Ulutaş, P.A.D.B. Vinusha Wickramasinghe, Tingyuan Xiao, and especially Dr. Linah Alkotami.

I would like to express my gratitude to the lab assistants who contributed to my projects: RJ Felice, Emma Gaebler, Andrea Kuhlman, Kamryn Mendoza, Gracie Munk, Taylor Pringle, Grant Snider and Nicholas Sokol.

I would like to thank my friend Shahla Mohammedi for her help and support. I would also like to thank my friends Aissata Diaby and Becky and Kenny Taylor for their support.

I am grateful for the support from my family members: Peggy and Ralph Beidelman, Mary and Fred Graca, Jean and Steve Rawe, and Alex, Zach, and Marita Thomas. I am thankful for the support from my siblings: Dr. Cassidy White and Jacob White. I would especially like to thank my parents, Susan and John White, for their immense support.

Chapter 1: Introduction

1.1: Cover Crops

Cover crops, when strategically implemented within arable land during fallow periods, confer multifaceted benefits to soil beyond conventional off-season conditioning, including enhanced biophysical properties and ecosystem function (Lamichhane and Alletto, 2025). Their rise as a pivotal component of sustainable agricultural paradigms stems from their capacity to deliver a spectrum of ecological, environmental, and economic advantages. Key attributes of cover-cropping systems include the temporal niche they occupy between primary crop cycles, thereby avoiding direct competition for land resources essential for food and commodity crop production. This temporal complementarity facilitates enhanced land-use efficiency and contributes to the preservation of natural habitats susceptible to agricultural conversion. Furthermore, cover crops demonstrably enhance soil structural stability (Gentsch et al., 2024), augment soil organic matter (SOM) accrual (Gentsch et al., 2024), a critical determinant of soil fertility and carbon sequestration, and mitigate diffuse nutrient pollution via the reduction of surface runoff, thereby preventing eutrophication in nearby aquatic ecosystems (Weyers et al., 2024). They also provide crucial overwintering and foraging habitats for pollinator populations and other beneficial insects (Bowers, 2021). Notably, the emerging application of cover crop biomass as a feedstock for sustainable aviation fuel (SAF) production presents a promising pathway towards the long-term decarbonization of the aviation sector (Taheripour et al., 2022).

While several cover crop species with potential for dual cropping, providing both soil benefits and harvestable seed oil and protein, are now cultivated in the U.S., adoption remains limited, with only 4.7% of arable land under cover cropping systems (United States Department of Agriculture, 2024). Despite the numerous agronomic and environmental advantages of cover

cropping, inherent limitations warrant consideration. One primary challenge is the potential for preceding cover crops to negatively impact subsequent cash crop yields through nutrient depletion. Furthermore, while generally requiring fewer inputs than conventional monocultures, the necessity for even minimal inputs can contribute to overall production costs. Consequently, the economic viability of integrating dual-purpose cover crops remains a significant impediment to widespread adoption, particularly when the market value of harvested products and non-market environmental benefits do not adequately offset implementation costs for producers. Future advancements in domestication and targeted genome editing offer promising avenues to enhance the economic proposition of these systems by increasing the value of seed oils, reducing the accumulation of undesirable compounds such as glucosinolates, and ultimately incentivizing broader farmer adoption (Ott et al., 2019).

Camelina sativa (L.) Crantz, commonly known as false flax or camelina, is an emerging oilseed crop species with favorable agronomic traits that position it as an attractive component of modern cover-cropping systems. As a member of the Brassicaceae family and a close relative of the model species *Arabidopsis thaliana* (diverged approximately 8 million years ago; Beilstein et al., 2008), *Camelina sativa* benefits from a substantial wealth of genetic and genomic resources. The allohexaploid nature of *Camelina sativa* (2n = 6x = 40) resulted from relatively recent hybridization events involving three distinct diploid ancestors. Genetic evidence indicates contributions from the known diploid species *Camelina neglecta* and *Camelina hispida*, along with a third, unidentified progenitor species closely related to *Camelina neglecta* (Brock et al., 2022; Chaudhary et al., 2020), collectively constitute the three subgenomes of *Camelina sativa* (Table 1.1) (Mandacova et al., 2019). The final polyploidization event leading to *Camelina sativa* is estimated to have occurred around 65,000 years ago (Brock et al., 2024). Similar allopolyploid

crops of agricultural significance include wheat (*Triticum*), sugarcane (*Saccharum*), upland cotton (*Gossypium hirsutum*), Pima cotton (*Gossypium barbadense*), oats (*Avena sativa*), tobacco (*Nicotiana tabacum*), coffee (*Coffea arabica*), and peanut (*Arachis hypogaea*).

Native to Europe and with a history of cultivation spanning thousands of years (Murphy, 2016), *Camelina sativa* exhibits advantageous traits such as low fertilizer input requirements (Berti et al., 2019), drought tolerance (Wittenberg et al., 2019), and the availability of both winter and spring varieties (Quezada & Cherian, 2012). While this study focuses on the spring *Camelina sativa* cultivar, these ecotypes enhance the versatility of the crop, allowing for its successful integration into diverse cropping systems, either through fall planting for overwintering and early summer harvest (winter varieties) or through early spring sowing for a single-season cycle (spring varieties). The key difference between the winter and spring varieties is the vernalization requirement of the seeds to induce germination, winter varieties necessitate a sustained period of cold temperatures to trigger reproductive development, whereas spring varieties are vernalization-independent (Ott et al., 2023). Traits such as cold hardiness arose from thousands of years of cultivation in varied climatic regions (Vollmann and Eynck, 2015). These characteristics render *Camelina sativa* particularly well-suited for cultivation in northern and southern latitudes across the Great Plains regions.

The seed yield of *Camelina sativa* exhibits a quantifiable range of approximately 1.5 to 3.0 metric tons per hectare (Mg ha^{-1}), a phenotypic trait significantly influenced by a complex interplay of environmental variables, including macro- and microclimatic conditions, soil factors such as nutrient availability and soil physicochemical properties, and inherent genotypic variation (Vollmann et al., 2007; Berti et al., 2011; Berti et al., 2016; Masella et al., 2014). Under agronomically optimized cultivation regimes, yields exceeding 3.0 Mg ha^{-1} have been

documented (Vollmann and Eynck, 2015), underscoring the species' yield potential under ideal management practices. The mature seed of *Camelina sativa* accumulates a substantial oil content, ranging from 32.1% to 45.7% of its dry weight (Strasil, 1997), representing a significant reservoir of triacylglycerols (TAGs). Notably, the fatty acid profile of *Camelina sativa* seed oil is characterized by a high proportion of polyunsaturated fatty acids (PUFAs), particularly linoleic acid (C18:2ω-6) and α-linolenic acid (C18:3ω-3) (Abramovic and Abram, 2005), exhibiting a compositional divergence from other Brassicaceae species. The elevated levels of polyunsaturated acyl chains within the TAGs directly impact the oil's physicochemical properties, resulting in reduced viscosity and melting point, which can influence its suitability for industrial applications, such as lubricants (Quinchia et al., 2010), transformer oils (Sbravati et al., 2014), hydraulic fluids (Carnes, 2004), heat-transfer oils (Kenda et al., 2017), and food and cosmetics (Burnett et al., 2017) applications. Furthermore, the increased degree of unsaturation correlates with decreased energy density and diminished oxidative stability, factors of critical importance for storage and utilization.

Thlaspi arvense L., commonly known as field pennycress, represents a recently domesticated oilseed species within the Brassicaceae family, currently undergoing intensive breeding programs and targeted genome editing to optimize its agronomic performance (McGinn et al., 2019; Chopra et al., 2019; Gautam et al., 2025). Sharing familial ties with *Camelina sativa* as a member of the Brassicaceae and exhibiting a close evolutionary relationship with the model organism *Arabidopsis thaliana*, *Thlaspi arvense* presents both similarities and key distinctions. While exhibiting comparable plant morphology and seed oil yield potential to *Camelina sativa*, *Thlaspi arvense* possesses a diploid genome (2n = 2x = 14; Dorn et al., 2015). This comparatively less complex genomic architecture renders *Thlaspi arvense* a more tractable

system for precision genome engineering. The relative simplicity of a diploid genome facilitates accelerated trait improvement via gene modification, demanding significantly reduced temporal and labor investment from researchers compared to the allohexaploid *Camelina sativa*.

The mature seed of *Thlaspi arvense* exhibits a distinctive fatty acid profile characterized by a significant accumulation of monounsaturated very-long-chain fatty acids (VLCFAs), notably eicosenoic acid (C20:1) and erucic acid (C22:1) (Moser et al., 2009). The total extractable seed oil content in *Thlaspi arvense* constitutes approximately 36% of the seed dry weight. Similar to *Camelina sativa*, *Thlaspi arvense* demonstrates a substantial capacity for protein biosynthesis within its seeds, with protein and free amino acid content collectively constituting 25% to 30% of the dry seed biomass (Berti et al., 2016). However, a key differentiating biochemical characteristic of *Thlaspi arvense* is the elevated endogenous concentration of glucosinolates, primarily sinigrin, which are secondary metabolites prevalent across numerous Brassicaceae species that serve as herbivore deterrents (Vaughn et al., 2006). The presence of high glucosinolate levels in the seed meal renders it toxic to chickens and pigs at elevated inclusion rates in feed formulations (Tripathi and Mishra, 2007). In contrast, *Camelina sativa* naturally accumulates considerably lower concentrations of these compounds. To mitigate this limitation in *Thlaspi arvense*, targeted gene editing approaches have been successfully implemented to downregulate glucosinolate biosynthesis (Williams, 2020), representing a crucial step towards enhancing the nutritional value and broader applicability of its protein co-product.

Both *Camelina sativa* and *Thlaspi arvense* are amenable to *Agrobacterium tumefaciens*-mediated genetic transformation via the established vacuum infiltration floral dip method (Lu and Kang, 2008; McGinn et al., 2019). This facilitates the relatively straightforward introduction and stable integration of the CRISPR-Cas9 gene editing system into their genomes. The

efficiency of targeted mutagenesis in plant CRISPR experiments is typically high, reported at approximately 65% in *Brassica napus* and exhibiting a range of 27% to 96% across various species and experimental designs (Yang et al., 2017). Critically, once the desired genomic modification has been achieved, the introduced transgene (encoding the Cas9 nuclease and guide RNA) can be segregated away through subsequent generations of selfing. This process yields a heritable, targeted alteration to the plant's endogenous genes, resulting in a “gene-edited” line that, notably, falls outside the regulatory definition of a “genetically modified organism” in many jurisdictions (Aftab et al., 2023).

The strategic application of genetic engineering offers a powerful avenue for the targeted enhancement of traits and the diversification of product streams derived from cover crop species. By augmenting the economic value of biomass and seed-derived compounds in species such as *Camelina sativa* and *Thlaspi arvense*, the concept of "cash cover crops" gains traction, enabling producers to realize direct economic returns while simultaneously delivering crucial agroecological benefits and contributing to the sustainable production of biofuels and bio-based oils (Quinn et al., 2019). Elevating the market value and expanding the utility of these specialized seed oils through genetic modifications represents a key incentive for broader farmer adoption of cover-cropping systems, fostering a synergistic relationship between agricultural productivity and environmental stewardship.

1.2: Fatty acid biosynthesis in plants

Post-harvest processing of major oilseed crops, such as soybean and maize, typically involves mechanical disruption to facilitate the extraction of oil and proteinaceous components from the seed matrix. The recovered crude oil fraction undergoes subsequent refining processes to yield edible oils, industrial lubricants, or feedstocks for biofuel production, while the residual

protein-rich meal serves as a significant input for the animal feed industry. Within the intricate network of plant metabolism, the manipulation of TAG biosynthesis and accumulation represents a particularly promising target for genetic engineering strategies aimed at enhancing crop value. TAGs constitute the predominant lipid class in plant seeds, often exceeding 90% of the total oil content, with minor quantities of other acyl lipids including structural membrane phospholipids, the TAG biosynthetic intermediates diacylglycerol (DAG) and monoacylglycerol (MAG), sterols, and free fatty acids also present. Functionally, TAG serves as the primary reservoir of stored energy within plant cells and, specifically in seeds, provides the essential metabolic fuel to support germination and seedling establishment. Beyond their role in energy storage, acylglycerol molecules, including DAG, also participate in intracellular signaling pathways, acting as second messengers or precursors for the synthesis of signaling lipids.

The physicochemical properties of seed oils are fundamentally determined by the specific composition of fatty acids (FAs) esterified to the glycerol backbone of TAG molecules and their positional distribution within the TAG structure. The primary metabolic route for TAG biosynthesis in developing seeds is the *de novo* Kennedy Pathway (Chapman and Ohlrogge, 2012), a series of critical enzymatic reactions localized to the endoplasmic reticulum (ER) membrane. The terminal, committed step in this pathway is catalyzed by DIACYLGLYCEROL O-ACYLTRANSFERASE 1 (DGAT1), an integral membrane protein of the endoplasmic reticulum (ER) (Figure 1.1). DGAT1 utilizes DAG, itself synthesized at the ER membrane, and activated fatty acids (Acyl-CoAs) drawn from the cytosolic acyl-CoA pool as substrates. This enzyme catalyzes the transfer of a fatty acyl moiety to the *sn*-3 hydroxyl group of DAG, resulting in the formation of TAG, which initially accumulates within the hydrophobic core of the ER membrane bilayer before coalescing into discrete oil bodies. An alternative, albeit less

dominant, pathway known as the PC-derived pathway, involves PHOSPHOLIPID:DIACYLGLYCEROL ACYLTRANSFERASE 1 (PDAT1) (Dahlqvist et al., 2000; Ståhl et al., 2004). PDAT1 catalyzes the acyl-CoA-independent transfer of a fatty acyl group from the *sn*-2 position of membrane phospholipids, typically phosphatidylcholine (PC), to DAG, yielding TAG.

The functional roles of these orthologous acyltransferases have been extensively investigated through the characterization of T-DNA insertion mutant lines in the model species *Arabidopsis thaliana*. Notably, loss-of-function *pdat1* mutants exhibited no discernible alterations in plant morphology or seed oil accumulation (Mhaske et al., 2005). In contrast, *dgat1* knockout lines displayed a phenotype characterized by delayed seed germination and a significantly modified fatty acid profile, specifically a reduction in the molar percentage of VLCFAs coupled with a concomitant increase in PUFAs, particularly α-linolenic acid (C18:3ω-3) (Katavic et al., 1995; Zou et al., 1999). The essential and partially redundant roles of DGAT1 and PDAT1 in maintaining reproductive viability were underscored by the observation that *Arabidopsis thaliana dgat1 pdat1* double knockout mutants exhibited lethality due to the production of non-viable pollen (Zhang et al., 2009). The introduction of transgenic TAG-biosynthesizing enzymes, when engineered to include an intron derived from *Arabidopsis thaliana DGAT1*, has been demonstrated to rescue the pollen lethality phenotype in these double knockout lines (McGuire et al., 2024), highlighting the potential for specific gene regulatory elements to modulate transgene functionality and complement endogenous gene loss.

1.3: CRISPR-Cas9 gene editing

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated protein 9 (Cas9) system is a revolutionary genome editing technology adapted from

bacterial adaptive immune systems, where it functions as a defense mechanism against bacteriophage infections (Jinek et al., 2012). In CRISPR-mediated gene knockout experiments, researchers design and synthesize a single guide RNA (sgRNA), a critical component of the Cas9 ribonucleoprotein complex. This sgRNA directs the Cas9 endonuclease to a specific genomic locus (protospacer or target site) through complementary base-pairing with the target DNA sequence immediately upstream of a requisite protospacer adjacent motif (PAM) (Shan et al., 2013). Upon recognition and binding to the target site, the Cas9 enzyme catalyzes a precise double-stranded DNA (dsDNA) cleavage event. Following this targeted DNA scission, the Cas9 complex disengages, and the cellular DNA repair machinery is activated. The resulting dsDNA break can be resolved primarily through two distinct pathways: homology-directed repair (HDR), a high-fidelity mechanism that can precisely repair or introduce specific nucleotide sequences using a homologous DNA template, and non-homologous end joining (NHEJ), an error-prone pathway that directly ligates the blunt DNA ends, frequently introducing insertions or deletions (indels) at the repair junction. These NHEJ-induced mutations often result in frameshifts or premature stop codons within the coding sequence of a gene, effectively leading to its functional knockout. The widespread presence of PAM recognition sequences (5'-NGG-3' for *Streptococcus pyogenes* Cas9, a commonly employed variant also used in this study) throughout diverse genomes enables the CRISPR-Cas9 system to theoretically target nearly any gene within virtually any species.

The inherent simplicity and remarkable versatility of the CRISPR-Cas9 system have rapidly established it as the preeminent tool for targeted genome editing across diverse biological systems. Beyond its foundational application in gene knockout studies, the CRISPR platform has spurred the continuous development of a diverse and expanding array of sophisticated

applications. These include the utilization of catalytically deactivated Cas9 (dCas9) for targeted transcriptional regulation through CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi) (Qi et al., 2021), the employment of Cas9 nickase variants engineered to generate single-stranded DNA breaks for precise base editing and targeted mutagenesis (Le Cong et al., 2013), strategies for precise gene insertion and knock-in (Mali et al., 2013), and the more recently developed CRISPR prime editing technology, which enables targeted single-nucleotide substitutions, insertions, and deletions without requiring double-strand breaks or donor DNA templates (Anzalone et al., 2019).

In plant biology, the CRISPR-Cas9 system, typically encoded within plant-codon-optimized expression vectors constructed using bacterial cloning techniques (Xing et al., 2014), is delivered into plant cells via *Agrobacterium tumefaciens*-mediated transformation, often employing the efficient floral dip method for dicotyledonous species (Lu and Kang, 2008). However, alternative delivery methods such as biolistics (particle bombardment) (Lacroix and Citovsky, 2020), protoplast transformation (Wu et al., 2020), and viral vectors are also employed, particularly for monocots or when stable integration via *Agrobacterium* is inefficient. CRISPR-Cas9 components can be introduced as plasmid DNA for stable integration, mRNA for transient expression (Zhang et al., 2016), or as pre-assembled ribonucleoprotein complexes (RNPs) for potentially transgene-free editing (Liang et al., 2018). Expression of the Cas9 protein and sgRNA is driven by various promoters, including constitutive, inducible, and tissue-specific options, often alongside selectable marker genes to facilitate the identification of transformed cells during plant regeneration. While nuclear genome editing is the most prevalent application, researchers have modified the CRISPR-Cas9 system to target chloroplast and mitochondrial genomes, requiring the development of specific delivery strategies and targeting sequences (Kang et al.,

2021). Notably, both *Camelina sativa* and *Thlaspi arvense* are amenable to the relatively simple and well-established *Agrobacterium*-mediated floral dip method of plant genetic transformation, facilitating CRISPR-Cas9-based genome editing for both stable and transient applications in these species.

Previous studies have successfully employed CRISPR-Cas9 to achieve complete gene knockouts in *Camelina sativa* for genes modulating seed oil content, composition, and other traits. For instance, a complete knockout of *FATTY ACID ELONGASE1* (*FAE1*), encoding an enzyme responsible for VLCFA biosynthesis, was achieved in an EMS-mutagenized background using a single sgRNA targeting all three homeologs (Ozseyhan et al., 2018). The resulting small deletions were confirmed and quantified using homeolog-specific primers followed by Sanger sequencing. The transcription factor-encoding gene *TRANSPARENT TESTA 8* (*TT8*), involved in flavonoid biosynthesis (where overexpression inhibits fatty acid accumulation) and seed coat development, was also targeted. CRISPR-Cas9-induced mutations in *Camelina sativa TT8* resulted in easily identifiable golden seeds (compared to wild-type light-brown seeds) and a slight increase in overall seed oil content (Cai et al., 2024). In this study, the mutations were confirmed through PCR and sequencing of the target sites in the *TT8* homeologs. For the complete knockout of *FATTY ACID DESATURASE 2* (*FAD2*), an enzyme critical for PUFA biosynthesis, reported by Jiang et al. (2017), Illumina Amplicon-Seq was employed to detect mutations in all three subgenomes. This next-generation sequencing (NGS) approach involves the targeted amplification of the genomic region of interest (the amplicon) using PCR, followed by high-throughput sequencing using the Illumina platform. Amplicon-Seq allows for deep sequencing of the targeted region, enabling the identification of even low-frequency mutations and providing detailed information about the spectrum of indels generated by CRISPR-Cas9

within each homeolog. The resulting sequence data is then bioinformatically analyzed to identify and quantify the mutations present in the treated samples compared to the wild type, providing a comprehensive assessment of the editing efficiency and the types of mutations induced across the *FAD2* homeologs. In contrast, droplet digital PCR (ddPCR), as demonstrated by Lyzenga et al. (2019) in their study of *CRUCIFERIN C* (*CRC*), offers a highly sensitive and precise method for detecting and quantifying CRISPR-induced indel mutations, even at low frequencies and without the need for sequence-specific probes. In ddPCR, the PCR reaction is partitioned into thousands of individual nanoliter-sized droplets, allowing for the digital counting of target DNA molecules in each droplet after endpoint PCR. This provides absolute quantification of the mutant alleles, making it particularly useful for confirming mutations across multiple homeologs, as seen in the *CRC* study where indel mutations were quantified in all three copies. Another complete gene knockout in *DIACYLGLYCEROL O-ACYLTRANSFERASE 3* (*DGAT3*) was generated with CRISPR-Cas9, and mutations were initially screened and detected via restriction enzyme-catalyzed digestion followed by Sanger sequencing (Lee et al., 2022). The use of restriction enzyme-catalyzed digestion for mutation detection relies on the design of the CRISPR target site such that the induced indel mutation (insertion or deletion) either creates or abolishes a recognition site for a specific restriction enzyme. Following PCR amplification of the target region, the PCR products from wild-type and CRISPR-treated plants are subjected to digestion with the chosen restriction enzyme. If a mutation has occurred at the target site in a treated plant, the digestion pattern will differ from that of the wild type. For example, a deletion might remove the restriction site, resulting in a larger undigested PCR product. Conversely, an insertion might create a new restriction site, leading to smaller DNA fragments after digestion. This method allows for a relatively quick and cost-effective initial screening of mutant lines, and the resulting

altered DNA fragment sizes can be visualized using gel electrophoresis. Putative mutants identified through restriction enzyme digest are then typically confirmed and the specific nature of the mutation (indel size and location) determined by Sanger sequencing.

Moving beyond single gene disruption, metabolic engineering approaches in plant science increasingly leverage multiplex genome editing strategies. These sophisticated techniques involve the simultaneous suppression, overexpression, or precise editing of multiple endogenous genes, often coupled with the introduction of transgenes or the enhancement of native gene expression within a single plant line to achieve complex improvements in agronomic traits or the production of high-value biochemicals. In the context of seed oil modification, transgenic acyltransferases hold significant potential to enhance the accumulation of specialty TAGs in oilseed platforms like *Camelina sativa* and *Thlaspi arvense*. However, the efficiency of these introduced enzymes can be limited by competition for shared substrates with endogenous acyltransferases. To overcome this constraint and facilitate the efficient production of diverse specialty TAGs, the generation of knockout lines in key endogenous acyltransferase genes is proposed as a strategy to create versatile "platform crop lines." These engineered platforms, devoid of competing endogenous enzymatic activities, could then serve as optimized recipients for the targeted introduction of various specialty-TAG synthesizing enzymes. By eliminating substrate competition, this approach is anticipated to substantially enhance the efficiency of the introduced transgenes in directing TAG biosynthesis, while the availability of well-characterized platform lines would significantly reduce the time and resources required to generate viable production systems for a spectrum of tailored oil profiles.

The allohexaploid genome of *Camelina sativa* presents a substantial hurdle for targeted gene editing strategies. Due to its evolutionary history, up to six copies of each target gene reside

across three distinct but related homeologous chromosomes, each typically exhibiting a diploid inheritance pattern. Consequently, achieving a complete functional knockout of a gene necessitates the introduction of multiple independent CRISPR-Cas9-induced mutations across all homeologous copies, followed by extensive breeding efforts to consolidate these mutations within a single homozygous plant line. Furthermore, the high degree of sequence conservation among *Camelina sativa* homeologous genes, often exceeding 95% identity (Kagale et al., 2014), significantly complicates the precise discrimination of individual gene copies using conventional molecular techniques. For instance, the design and application of homeolog-specific PCR primers – primers engineered to selectively amplify a target sequence within one of the three subgenomes – and subsequent Sanger sequencing can face limitations in unequivocally distinguishing and verifying mutations in each of the highly similar gene copies.

Conventional methods for detecting CRISPR-Cas9-induced mutations in plants often necessitate screening extensive populations, ranging from tens to hundreds of individuals, using laborious and relatively costly gene sequencing techniques. Furthermore, these methods do not invariably guarantee the identification of complete gene knockouts, particularly in polyploid species. To streamline mutation detection and circumvent the need for extensive sequencing, innovative cloning constructs incorporating multiple target sites (typically three) within each target gene were employed. The induction of two simultaneous double-stranded DNA breaks by the CRISPR-Cas9 complex within a single gene region can trigger large DNA excision mutations, a process mediated by endogenous cellular NHEJ repair. These substantial truncation mutations can be readily detected using standard PCR and gel electrophoresis, obviating the requirement for sequencing individual genes from each plant and enabling the relatively high-throughput screening of large sample sizes. The Cas9 transgene was successfully segregated

away in all complete and partial gene knockout lines characterized within this thesis. While this dual-target site excision strategy facilitates the identification of mutant DNA via the observation of truncated amplicons on PCR gels, it inherently lacks the resolution to determine mutation zygosity or the specific subgenome harboring the alteration in the allohexaploid *Camelina sativa*. In contrast, the generation of gene knockout lines using CRISPR-Cas9 is significantly more facile in the diploid *Thlaspi arvense*, owing to its reduced genomic complexity. Furthermore, this strategy fails to detect small indel mutations in both *Camelina sativa* and *Thlaspi arvense* unless used in conjunction with DNA sequencing.

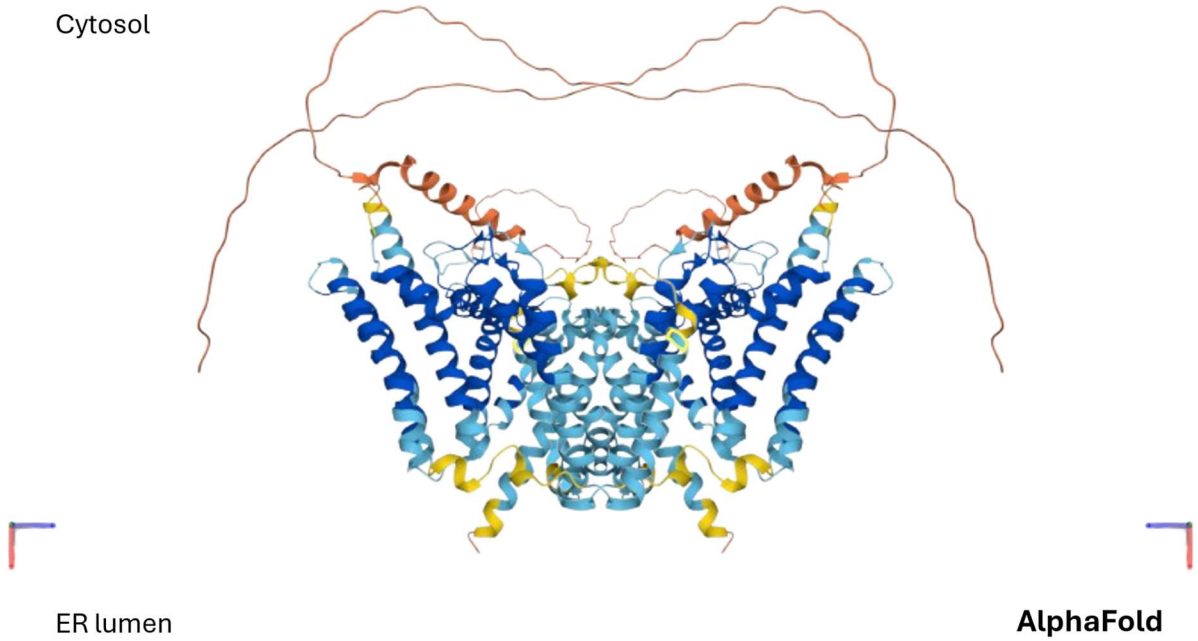

Figure 1.1: Predicted 3D structure of CsDGAT1. Forming a homodimer embedded in the ER membrane, each monomer contains nine alpha helices and an N-terminal intrinsically disordered region (IDR).

	Homeologous chromosomes	Subgenome	Ancestral species
DGAT1	Chromosome 01	Subgenome 2	*Camelina neglecta-like*
	Chromosome 15	Subgenome 3	*Camelina hispada*
	Chromsome 19	Subgenome 1	*Camelina neglecta*
PDAT1	Chromsome 08	Subgenome 1	*Camelina neglecta*
	Chromosome 13	Subgenome 2	*Camelina neglecta-like*
	Chromosome 20	Subgenome 3	*Camelina hispada*

Table 1.1: Nomenclature of homeologous chromosomes in *Camelina sativa.* This table shows the relationships between the three ancestral species composing the allohexaploid genome of *Camelina sativa* (Brock et al., 2022; Chaudhary et al., 2020; Mandacova et al., 2019; Brock et al., 2024). Homeologous chromosomes are referred to by chromosome number in this project. The color highlight indicates which ancestral species contributed each homeologous chromosome to the *Camelina sativa* genome.

Chapter 2: Engineering Platform Crop Lines

2.1: Introduction

Augmenting the economic value of seed oil derived from cover crops directly mitigates a primary financial impediment to their widespread adoption, potentially incentivizing increased cover-cropping rates among agricultural producers. Metabolic engineering initiatives in plant systems employ diverse strategies aimed at the rational optimization of biochemical product profiles. In the context of cover crops, these projects offer the transformative potential to fundamentally alter seed oil composition by targeted modulation of endogenous metabolic pathways via precise genome editing and the introduction of heterologous transgenes. The *de novo* synthesis of specialty TAGs in cover crop species typically necessitates the incorporation of a transgenic acyltransferase that catalyzes the terminal, rate-limiting step in the biosynthesis of the desired specialty TAG.

A significant challenge in developing sustainable plant-based feedstocks for industrial applications is that TAGs with desirable specialty fatty acid profiles are often biosynthesized in plant species possessing inherently suboptimal agronomic traits for large-scale cultivation. These species, while accumulating valuable lipids, may exhibit limitations in seed yield, disease resistance, or adaptability to diverse growing conditions. A compelling example of this is the pursuit of medium-chain fatty acids (MCFAs) as a bio-based alternative to conventional C9–C16 petroleum-derived jet fuels. Several *Cuphea* species indigenous to the United States demonstrate a remarkable capacity for high-level MCFAs biosynthesis and accumulation in their seed oils (Kim et al., 2011). However, consistent with the general trend, the domestication of *Cuphea* for widespread agricultural production has encountered significant agronomic challenges (Berti and Gesch, 2015).

The feasibility of metabolic engineering for the production of specialty oils has been demonstrated in both *Camelina sativa* and *Thlaspi arvense*. Species within the genus *Euonymus* naturally accumulate high levels of 3-acetyl-1,2-diacylglycerols (acetyl-TAGs) in their seeds (Durrett et al., 2010). Unlike conventional TAGs, acetyl-TAGs incorporate an acetyl group at the sn-3 position instead of a long-chain fatty acid, resulting in oils with reduced viscosity and distinct physical properties (Durrett et al., 2010). Despite the industrial and nutritional potential of this unique oil, *Euonymus* species lack agronomically favorable traits, limiting their use as crop plants.

The key enzyme responsible for acetyl-TAG biosynthesis was first identified and functionally characterized in *Euonymus alatus* (EaDAcT) (Durrett et al., 2010). When *EaDAcT* was expressed in *Arabidopsis thaliana*, transgenic seeds accumulated up to 45 mol% acetyl-TAG. This level increased to 65 mol% when *EaDAcT* expression was combined with a knockout mutation in *DGAT1*, a competing TAG biosynthesis gene. In *Camelina sativa*, co-expression of *EaDAcT* and *DGAT1* RNAi suppression resulted in up to 85 mol% acetyl-TAG in the highest-performing line (Liu et al., 2015).

Subsequently, an orthologous enzyme, diacylglycerol acetyltransferase from *Euonymus fortunei* (EfDAcT), was identified and shown to have higher enzymatic activity in vivo (Tran et al., 2017). Replacing *EaDAcT* with *EfDAcT* in transgenic *Camelina sativa* increased acetyl-TAG accumulation by an additional 20 mol%. When paired with *DGAT1* RNAi suppression, *EfDAcT* expression resulted in up to 90 mol% acetyl-TAG in the highest-producing lines (Alkotami et al., 2021), nearly reaching the endogenous levels of *Euonymous* species.

To build upon previous metabolic engineering efforts in *Camelina sativa* and *Thlaspi arvense*, expression vectors containing *EfDAcT* along with RNAi constructs to suppress *DGAT1*

were introduced into transgenic plant lines harboring knockout mutations in *FAE1*. This gene encodes a key enzyme responsible for synthesizing VLCFA by elongating fatty acyl-CoA chains Endogenous VLCFA levels were found to be higher in *Thlaspi arvense* compared to *Camelina sativa*, and in the *fae1* mutant background, acetyl-TAG accumulation reached as high as 98 mol% in the top-performing line (Alkotami et al., 2024). By inhibiting VLCFA biosynthesis, this strategy redirected fatty acid flux toward acetyl-TAG production, demonstrating the effectiveness of multiplexed genetic engineering approaches in optimizing the accumulation of high-value lipid products.

Building on this concept, this project aims to use CRISPR-Cas9 to knock out two additional endogenous enzymes, *DGAT1* and *PDAT1*, in *Camelina sativa* and *Thlaspi arvense*. These enzymes compete for DAG substrates that could otherwise be utilized by transgenically expressed enzymes like *EfDAcT*. By eliminating this competition, the goal is to further enhance the efficiency of specialty TAG biosynthesis and to better understand the metabolic bottlenecks that limit oil trait optimization in engineered oilseed crops.

To efficiently redirect carbon flux towards engineered TAG biosynthesis, minimizing substrate competition from endogenous TAG biosynthetic pathways is crucial. We hypothesized that the targeted disruption of *DGAT1* and *PDAT1*, terminal enzymes in the native TAG assembly pathway that compete for DAG and acyl-CoA substrates with heterologous acyltransferases, will significantly enhance the accumulation of desired specialty TAGs (Figure 2.1). While prior research has indicated the potential of RNA interference (RNAi)-mediated *DGAT1* downregulation to augment specialty-TAG production, RNAi-based gene silencing can exhibit variable and often incomplete knockdown, is susceptible to off-target effects, and demonstrates limited heritable stability due to potential gene silencing suppression mechanisms (Dougherty

and Parks, 1995). In contrast, CRISPR-Cas9-mediated gene knockouts offer a robust and genetically stable solution for long-term metabolic engineering by achieving permanent abrogation of target enzyme activity. Therefore, this research aimed to establish a scalable and heritable genetic engineering strategy in agronomically amenable species, specifically *Camelina sativa* and *Thlaspi arvense*, to enhance their commercial potential by tailoring seed oil composition for diverse industrial applications. We anticipated that the complete elimination of endogenous *DGAT1* and *PDAT1* activity will effectively channel the available DAG pool towards the introduced, specialty-TAG synthesizing enzymes, thereby substantially increasing the yield and purity of the desired lipid production.

2.2: Methods

2.2.1 Construction of plasmids

The sgRNAs for the CRISPR-Cas9 vector targeting genes in *Camelina sativa* and *Thlaspi arvense* were designed with the aid of CHOPCHOP (Labun et al., 2019; Labun et al., 2016; Montague et al., 2014). To induce large-DNA excision mutations detectable as amplicon size polymorphisms (ASPs) in PCR gels, multiple CRISPR-Cas9 target sites were implemented within each targeted gene. The insert DNA for the intermediate vector was designed according to the method outlined in Xing et al. (2014). Guide RNA (gRNA) sequences targeting specific genes were cloned using Overlap Extension PCR for *Camelina sativa* targets and synthesized by Twist Biosciences for *Thlaspi arvense* targets. The insert DNA was cloned into the expression vector using the Golden Gate method (Engler et al., 2008). The expression vector, pEC666Red3, was provided by the Sedbrook Lab at Illinois State University. The Cas9 protein was expressed under the egg cell-specific EC1.2 promoter, and the *Discosoma* Red (DsRed) selection marker protein (Jach et al., 2001) was expressed with the Cauliflower Mosaic Virus 35S (CaMV 35S)

constitutive promoter. The expression vectors for each target gene were transformed into *Agrobacterium tumefaciens* strain GV3101.

2.2.2 Plant transformation and propagation

Plants were transformed using the vacuum infiltration floral dip method described in Lu and Kang, 2008. The presence of the DsRed selection marker protein allowed for the identification of transgenic plants, indicated by red-light fluorescence in the seed coat under green-light excitation. In all transgenic lines, successfully transformed seeds, up to a maximum of 16, were planted on soil. Confirmation of transgene integration involved PCR amplification of a Cas9 gene segment using purified DNA from plant leaf tissue, with the amplicon's presence verified by gel electrophoresis. Plant leaf tissue was clipped from plants at approximately 1-3 weeks of age. DNA was purified from leaf tissue following an isopropanol-precipitation extraction method (Li et al., 2013). An approximate 3:1 segregation ratio of DsRed-positive to DsRed-negative in seeds indicated probable single-insertion of the transgene in independent insertion events. The approximate 3:1 ratio was used in subsequent generations to identify plants heterozygous for the transgene.

2.2.3 Mutant identification

In the T2 generation, *Camelina sativa* plants exhibiting putative large-DNA excision mutations, detectable as ASPs via PCR, were selected for propagation, maintaining both Cas9-positive and Cas9-negative genotypes. To increase throughput for screening these ASPs in subsequent generations, Cas9-positive T2 seeds were sown at high density (approximately 10 seeds per pot). Leaf tissue was harvested from each plant, pooled, and genomic DNA extracted for initial mutant screening by PCR and gel electrophoresis. Pots exhibiting evidence of an ASP indicative of a large-DNA excision mutation in the pooled DNA were subjected to a second

round of sampling, where leaf tissue from each individual plant was collected, and DNA extracted and analyzed independently (Figure 2.2). Plants that presented large-DNA excision mutations in lines segregating for the Cas9 transgene (heterozygous) were advanced to subsequent generations to facilitate the induction of mutations across all homoeologs of the target genes.

For amplification of a segment of *PDAT1* in *Camelina sativa*, forward primers specific to each of the three subgenomes were implemented to sequence and characterize mutations. In contrast, a single universal forward primer was paired with two reverse primers for *CsDGAT1* mutation detection. The reverse primer designated "Rev1" specifically annealed to the deleted chromosomal DNA region identified in chromosome 01 of the T2 mutant. Amplicons generated from plants homozygous for this chromosome 01 deletion corresponded to genomic DNA in chromosome 15 and chromosome 19. Screening gels were used for detection of new mutations in T3 and beyond. A homoeolog-specific forward primer was successfully designed for chromosome 15 DNA sequencing. Due to the high level of DNA sequence homology between chromosome 15 and chromosome 19, attempts to design a homeolog-specific primer for chromosome 19 were unsuccessful. Mutation zygosity was determined through the use of homoeolog-specific primers, when available, followed by Sanger sequencing with services provided by Azenta Biosciences and Eurofins Genomics. As an alternative method for zygosity assessment, the presence or absence of specific ASPs in screening gels characterized following Mendelian inheritance. Homozygosity of a mutation in a parent plant was inferred when specific ASPs presented in populations of ten or more progeny without absences.

To facilitate the identification, visualization, and analysis of large-DNA excision mutations, a Python script was developed utilizing Biopython (Cock et al., 2009). This script

aligned each sample sequence to each of the three homoeologous chromosomes, returning the alignment with the highest score to infer the subgenome of origin for each mutation. The scoring algorithm was optimized to prioritize the detection of large-DNA excisions while applying higher penalties to SNPs and smaller indels to aid in distinguishing between highly homologous subgenomes.

2.2.4 *Thlaspi arvense* cloning and screening

In *Thlaspi arvense* experiments, the cloning vector was engineered with two CRISPR gRNA target sites per gene, in contrast to the three utilized for *Camelina sativa*. Apart from this modification, cloning vector design and plant genetic transformation procedures mirrored those employed for *Camelina sativa*. The *Thlaspi arvense* lines used in this study harbored a pre-existing knockout mutation in the *TRANSPARENT TESTA 8* (*TT8*) gene and were provided by the Sedbrook Lab at Illinois State University. All T1 generation plants were advanced to the T2 generation, regardless of whether genotyping gels indicated a large-DNA excision mutation or not. T2 plants that screened negative for the Cas9 transgene were subjected to DNA sequencing to identify mutations in *DGAT1* in the *Thlaspi arvense tt8* background. To induce additional *DGAT1* mutant alleles in the *Thlaspi arvense tt8* background and to obtain Cas9 negative plants through segregation, Cas9-positive plants were advanced to the T3 generation. The custom Python script for sequence analysis was used to identify and characterize mutations.

2.2.5 Fatty acid methyl esters (FAMEs) analysis

Seed lipids were extracted following mechanical grinding using a Polytron homogenizer while immersed in 1 mL toluene. FAMEs were subsequently generated via an acid-catalyzed transmethylation protocol adapted from the method established by Li et al., 2006. To facilitate quantitative analysis, 100 µg each of two internal standards, tripentadecanoin (C15:0) and

triheptadecanoin (C17:0), were added to each sample prior to transmethylation. The resulting FAMEs were then resolved and quantified using gas chromatography (GC) according to the method described in Aznar-Moreno and Durett, 2017.

2.2.6 Germination assays

Seeds utilized for germination assays underwent surface sterilization via sequential washes with 70% ethanol and a 50% bleach solution supplemented with 0.05% Tween 20. Sterilized seeds were then imbibed on Petri dishes containing either 1% agar or 1% agar supplemented with 1% sucrose. Immediately following imbibition, plates were transferred to growth chambers maintained at 21 °C under a 16-hour light/8-hour dark photoperiod. Radicle emergence (defined as the visual protrusion of the radicle) and cotyledon emergence (defined as the visible separation and expansion of cotyledons) were tabulated daily for a period of seven days at consistent time points.

2.3: Results

2.3.1 Mutants of DGAT1 characterized in *Camelina sativa*

A *Camelina sativa* plant line homozygous for induced mutations across all three *DGAT1* homoeologous subgenomes was generated and characterized (plant designated CsDGAT1.8.R13A.B33.5, Cas9-negative progeny) (Figure 2.3). Sequence analysis and protein prediction assessment revealed that all three independent mutations in this line were in-frame deletions (Figure 2.4). Each homeologous copy of DGAT1 retained partial structure based on the predicted amino acid sequence. Segments between one and 105 amino acids were exscinded from the mutant protein product in each homeolog, with all occurring in transmembrane domain 1. One homeolog lost transmembrane domain 2 and another lost most of the intrinsically disordered region of the final protein product.

Out of 174 T2 generation *Camelina sativa DGAT1* CRISPR-edited plants across 12 independent lines, only a single plant line exhibited a heritable large-DNA excision mutation that produced viable seeds. This mutation, mapped to chromosome 01 at a genomic location distant from the designed CRISPR-Cas9 target sites, involved a 423 base pair deletion. PCR amplification of the mutant allele displayed biased amplification, likely attributable to stable secondary DNA structures, which predicted within the single-stranded DNA (ssDNA) of the wild-type amplicon (Figure 2.5). These secondary structures could potentially hinder primer annealing and polymerase efficiency. This initial mutant line was propagated to the T3 and T4 generations in lines segregating for the Cas9 transgene to facilitate the induction of subsequent mutations in the remaining homoeologous chromosomes. In a specific T4 plant designated CsDGAT1.8.R13A.B33.5, a novel large-DNA excision was detected on chromosome 19 using a reverse primer ("Rev1") specifically designed to anneal within the excised region of the previously identified chromosome 01 mutation. This large-DNA excision resulted in the deletion of 413 base pairs, predicting an in-frame omission of a substantial portion of the intrinsically disordered region within this homoeolog's protein product. The third and final mutation, discovered in chromosome 15 in the same plant, was identified using a homoeolog-specific primer for sequencing. This in-frame 3 base pair deletion is predicted to result in the loss of Alanine-134, a well conserved amino acid residue comprising transmembrane domain 1 of the predicted DGAT1 protein.

Phenotypic assessment revealed no significant alterations in plant morphology or seed germination. However, fatty acid profile analysis demonstrated statistically significant changes, specifically an increase in α-linolenic acid (C18:3ω-3) and a concomitant reduction in eicosenoic acid (C20:1) (Figure 2.6), mirroring the characteristic fatty acid profile observed in *Arabidopsis*

thaliana dgat1 mutants (Katavic et al., 1995). Nevertheless, the magnitude of these fatty acid alterations was significantly attenuated compared to those observed in *Arabidopsis thaliana* and *Thlaspi arvense tt8 dgat1* mutants.

A notable phenotype was observed in a T2 generation plant (designated CsDGAT1.8.B3B) that germinated but subsequently failed to produce viable seeds. This plant exhibited a pattern of developmental decline, leading to mortality approximately 30 days post-germination, with visible signs of physiological stress preceding its demise. Molecular analysis of this line revealed a large-DNA excision mutation on chromosome 15, spanning the genomic interval between CRISPR-Cas9 target sites 1 and 3. This frameshift mutation is predicted to result in a severely truncated protein product due to the introduction of a premature stop codon in this *DGAT1* homoeolog. A new large-DNA excision mutation in chromosome 15 was recently discovered in a sibling, designated CsDGAT1.8.R13B.R1, of the potentially lethal mutant plant. The excision results in the deletion of 317 bp CDS, three bp less that the potentially lethal mutant. The predicted amino acid sequence shows a premature stop codon encoded. This plant shows no signs of growth or reproductive abnormalities.

2.3.2 Mutant of DGAT1 characterized in *Thlaspi arvense*

A *Thlaspi arvense dgat1* mutant in the *tt8* background was identified and characterized, revealing significant alterations in seed composition and seedling germination and establishment. A 13 base pair frameshift mutation located at CRISPR target site 1 within the *Thlaspi arvense DGAT1* gene introduced a premature stop codon in the predicted primary amino acid sequence, likely resulting in a non-functional DGAT1 protein product (Figure 2.7). The zygosity of mutants was determined by using Sanger sequencing electropherogram analysis. Seeds from homozygous mutant lines exhibited a statistically significant reduction in total oil content and a slight decrease

in seed weight. Seedlings derived from the *dgat1* mutant displayed severely delayed germination and establishment (Figure 2.8), with only approximately 20% of mutant seeds developing into viable seedlings with emerged cotyledons by day 19 of the germination assay, the point at which plants were transferred to soil. A subset of mutant seeds germinated but failed to achieve cotyledon emergence from the seed coat, leading to seedling death.

The seed fatty acid profile was significantly altered compared to the wild-type, displaying substantially elevated levels of α-linolenic acid (C18:3ω-3) with a concomitant decrease in the very-long-chain fatty acid eicosenoic acid (C20:1) (Figure 2.9), consistent with the fatty acid profile of *dgat1* mutants in *Arabidopsis thaliana* (Katavic et al., 1995). Once established, surviving mutant plants exhibited a seemingly normal growth trajectory, displaying no overt phenotypic or morphological differences from the wild-type, apart from the initial developmental delay.

2.3.3 Mutants of PDAT1 characterized in *Camelina sativa*

A large-DNA excision mutation from a T4 generation plant was identified in a mutant screening gel and selected for further characterization. Gel electrophoresis screening confirmed the successful segregation of the Cas9 transgene out from this specific line (CsPDAT1.C.R1.BR25.B1sg). To characterize the mutations present, homoeolog-specific primers designed for each of the three *PDAT1* homoeologs were utilized to obtain sequencing data (Figure 2.10). Sanger sequencing revealed two small insertion mutations, undetectable by the large-DNA excision screening gel, located on chromosome 08 and chromosome 20. The observed large-DNA excision mutation was mapped to chromosome 13. Sequence analysis confirmed that each of these mutations resulted in a frameshift, leading to a premature stop codon and predicted loss of functional protein product. The zygosity of the small indel mutations

was determined by the presence of distinct, non-overlapping peaks in the Sanger sequencing electropherograms at the observed mutation sites, and these genotypes were further validated in subsequent progeny. Zygosity analysis of the large-DNA excision, performed by genotyping 11 progeny of the mutant plant, indicated homozygosity for the mutation in all tested individuals. By the T5 generation, no plants homozygous for large-DNA excision mutations across all three homoeologous chromosomes were generated, as assessed by the PCR amplification and gel electrophoresis method.

Three independent mutant lines harboring alterations in the *Camelina sativa PDAT1* gene were characterized in this study. All three selected lines were confirmed to be devoid of the Cas9 transgene. A line representing a complete knockout across all three homeologous subgenomes (*Cspdat1-/-/-*) was phenotypically compared to two partial knockout lines retaining at least one functional gene copy (*Cspdat1+/-/-*, wild-type allele on chromosome 08; and *Cspdat1-/+/-*, wild-type allele on chromosome 13) (Figure 2.11). Consistent with observations in the *Arabidopsis thaliana pdat1* mutant (Mhaske et al., 2005), no phenotypic or morphological differences were discernible in any of the characterized *CsPDAT1* mutant lines when compared to the wild-type control. Three fatty acids showed small but statistically significant differences in the complete knockout line when compared to the wild type (Figure 2.12).

2.3.4 Mutants of PDAT1 identified in *Thlaspi arvense*

Two distinct mutations in the *PDAT1* gene were identified in separate, independent *Thlaspi arvense* lines. The first mutation, detected via large-DNA excision screening gels and confirmed by Sanger sequencing, comprised a substantial genomic deletion spanning the intergenic region between the two CRISPR-Cas9 target sites. Homozygous mutant plants for this deletion were identified in the T2 generation by the absence of a wild-type band in the mutant

screening gels (Figure 2.13). Sequence analysis revealed that this mutation induced a codon frameshift, leading to the introduction of a premature stop codon and a predicted non-functional PDAT1 protein product. The second mutation was identified through bioinformatic analysis of sequencing data using the custom Python script (Figure 2.14). This alteration consisted of a complete DNA inversion of the genomic sequence within the *PDAT1* gene, encompassing the region between the defined target sites, with no net loss of nucleotide bases. Despite the preservation of gene length, this inversion resulted in a predicted premature stop codon within the coding sequence, suggesting a loss of functional protein. Notably, neither of these *pdat1* mutant lines exhibited any discernible morphological or germination phenotypes that were distinct from the wild-type control.

In the T1 generation, three independent *Thlaspi arvense* plant lines exhibited large-DNA excision mutations detectable as ASPs on screening gels. Notably, two of these mutant lines displayed reduced viability, with individual plants succumbing prior to reproductive maturity. These plants exhibited developmental abnormalities reminiscent of the lethal *Camelina sativa dgat1* mutant observed in this study, which underwent mortality approximately 30 days post-germination. Furthermore, a general growth reduction was observed across all CRISPR-edited T1 generation *Thlaspi arvense* plants at 30 days post-germination, as evidenced by a reduction in plant height compared to concurrently grown wild-type controls.

2.4: Discussion

2.4.1 Mutant screening and detection method

The novel mutant screening methodology, relying on the CRISPR-Cas9 system to create large-DNA excisions that could be detected as ASPs using PCR and gel electrophoresis, did not yield the anticipated results. Contrary to our hypothesis, the expected three unique large-DNA

excisions were not observed in any plant samples across all targeted genes. Specifically, no large-DNA excisions were detected in the T1 generation of *Camelina sativa*. The estimated mutation efficiency in the T2 generation of *Camelina sativa*, determined by the frequency of detectable mutations among screened plants (Table 2.1), suggests that the rate of mutant generation is likely influenced by factors such as individual target site efficacy, transgene integration locus within the plant chromosome, and Cas9 nuclease activity. While a high efficiency rate was not anticipated, the absence of the predicted three unique large-DNA excisions by the T5 generation or earlier indicates a significant limitation of this approach. Furthermore, a lower frequency of new large-DNA excisions was observed beyond the T3 generation. Consequently, this initial screening method proved insufficient for the efficient identification of complete gene knockouts in *Camelina sativa* and necessitated augmentation with conventional sequencing-based techniques. Despite its relative simplicity and lower per-sample cost, the protracted timeframe required to potentially achieve complete knockouts by the T5 generation without readily detectable markers renders this strategy less feasible compared to alternative methodologies.

While the large-DNA excision screening method did not prove effective for the direct isolation of complete gene knockouts, it may serve as a valuable pre-screening tool to enrich for plant lines harboring active Cas9 activity. The identification of even a single large-DNA excision mutation within a lineage indicates successful transgene expression and nuclease activity, potentially increasing the likelihood of detecting smaller indel mutations, characteristic of NHEJ repair, upon subsequent targeted sequencing. Implementing this PCR-based screening method in the T1 and T2 generations could substantially reduce the number of plant samples requiring resource-intensive sequencing to identify lines with promising, albeit incomplete, gene modifications.

Several challenges were encountered during Sanger sequencing of mutant amplicons. Firstly, a CRISPR-Cas9-induced mutation unaligned with gRNA target sites in *CsDGAT1* resulted in the deletion of the genomic region complementary to the reverse primer binding site, precluding its detection via ASP screening on PCR DNA gels. This mutation was only discovered through the utilization of an alternative reverse primer that annealed farther from the target sites. Subsequently, amplicons generated with this alternative primer demonstrated PCR amplification bias and proved recalcitrant to high-quality Sanger sequencing, likely due to the formation of stable DNA secondary structures. Secondly, the occurrence of two distinct large-DNA excision mutations of near identical base pair size would have resulted in comigrating mutant bands on gel electrophoresis, rendering accurate sequencing of the individual mutations problematic due to overlapping signals.

2.4.2 Phenotypes

The absence of discernible phenotypic alterations in *pdat1* knockout lines of both *Camelina sativa* and *Thlaspi arvense* under standard growth conditions was anticipated. Firstly, the lack of mutant phenotypes matches the results in *Arabidopsis thaliana* knockouts in *pdat1* (Mhaske et al., 2005). Secondly, the primary *de novo* TAG biosynthesis pathway, which utilizes DGAT1 to catalyze the terminal esterification step, remained functionally intact. PDAT1, in contrast, employs phospholipid substrates for TAG synthesis and plays a significant role in membrane lipid remodeling. Consequently, we hypothesize that these *pdat1* mutant lines may exhibit deleterious phenotypes when subjected to various environmental stresses that necessitate membrane adaptation, such as heat stress and cold stress. The slight statistically significant differences in fatty acids observed in the *pdat1* knockout lines in *Camelina sativa* may indicate a small difference in substrate affinity for DGAT1.

The *dgat1* complete gene knockout mutants in *Camelina sativa* and *Thlaspi arvense* exhibited seed oil phenotypes consistent with previously characterized *dgat1* mutants in *Arabidopsis thaliana* (Katavic et al., 1995). These shared phenotypes included a significantly altered fatty acid profile, characterized by elevated α-linolenic acid (C18:3ω-3) and reduced eicosenoic acid (C20:1), alongside a reduction in overall seed oil content. However, a stark divergence was observed in early developmental phenotypes: the *Thlaspi arvense dgat1* mutant displayed a substantial increase in non-viable seeds and a pronounced delay in emergence and seedling establishment among viable progeny, whereas the homozygous *Camelina sativa dgat1* triple mutant showed none of these detrimental effects.

Considering the wild-type germination and establishment phenotypes observed in the *Camelina sativa dgat1* triple mutant, we infer that the CRISPR-Cas9-induced mutations likely resulted in a partial, rather than complete, loss of endogenous DGAT1 function. Specifically, the in-frame deletion of Alanine-134 within the DGAT1 isoform encoded by chromosome 15 is hypothesized to retain partial or even full catalytic activity. Although the mutations in the *DGAT1* isoforms encoded by chromosome 01 and chromosome 19 may impair protein dimerization or substrate binding, the presence of a putatively functional DGAT1 isoform from chromosome 15 appears sufficient to maintain a threshold level of total DGAT1 enzymatic activity and TAG biosynthesis necessary for normal early plant development and viability.

These findings indicate that *Camelina sativa* possesses a notable resilience to DGAT1 perturbation, likely necessitating a complete functional null across all homoeologous copies to elicit significant developmental or viability-related phenotypic consequences. Nevertheless, the conserved alteration in fatty acid profiles observed in *dgat1* mutants of *Arabidopsis thaliana* and *Camelina sativa* suggests a discernible additive effect of *DGAT1* homoeolog copy number on

lipid biosynthesis, the precise nature of which warrants further investigation. This partial attenuation of endogenous DGAT1 activity, achieved without compromising plant germination and establishment, may represent a valuable strategy for metabolic engineering applications. By reducing native DGAT1 activity and thereby diminishing substrate competition for introduced transgenic acyltransferases, while maintaining a basal level of endogenous TAG synthesis sufficient to prevent deleterious pleiotropic effects, this approach could optimize the accumulation of target specialty TAGs.

All three of the mutations characterized in the *Camelina sativa dgat1* knockout line produced in-frame deletions, which were predicted to result in truncated protein products that could potentially retain some functionality. In the previously mentioned T2 plant with the observed lethal phenotype, the CRISPR-Cas9-induced mutation caused a frameshift and encoded a premature stop codon in a *DGAT1* homoeolog on chromosome 15. The individual plant harboring this specific mutation germinated normally and the seedling became well-established, but soon after it began showing signs of ailment, succumbing approximately 30 days post-germination, despite its sibling plants displaying normal development. Given the presence of five remaining functional *DGAT1* gene copies, the mechanism underlying this lethality remains unclear and warrants further investigation.

One potential mechanism for the observed lethality, despite the presence of multiple *DGAT1* homoeologs, is dosage sensitivity leading to functional haploinsufficiency. While multiple gene copies exist, the loss of function in a highly expressed or developmentally critical homeolog on chromosome 15 might have reduced total DGAT1 activity below a crucial physiological threshold. However, this explanation is rendered less probable by the observation of the recently identified deletion in chromosome 15 encoding a premature stop codon. This

mutant isoform did not elicit detrimental effects on plant growth or reproductive maturation. Alternatively, the truncated protein resulting from the premature stop codon could exert a dominant-negative effect by forming non-functional dimers with wild-type DGAT1 monomers or by competitively binding to substrates or localization machinery, thereby impairing the activity of the remaining functional enzymes. While plausible, this scenario is also less likely given the presence of five other putatively functional *DGAT1* copies, which would presumably necessitate the widespread formation of dysfunctional dimers to manifest such a severe phenotype. Furthermore, the potentially lethal mutant on chromosome 15 might have resulted in cis-regulatory element disruption or position effects, altering the expression of neighboring essential genes. Although a possibility, this is also less probable given the continued presence of putatively functional gene copies within the genome. Finally, insertional mutagenesis resulting from *Agrobacterium*-mediated transformation could have disrupted an essential gene or regulatory element at the transgene integration site; however, this remains a possibility that could not be assessed through transgene segregation analysis in this specific lethal line due to its failure to produce viable offspring.

It is also conceivable that DGAT1 possesses uncharacterized functions beyond its established role in TAG biosynthesis, functions that exhibit sensitivity to even partial functional impairment across the homoeologous genome. Furthermore, the CRISPR-Cas9 editing event at the *DGAT1* locus on chromosome 15 could have induced epigenetic modifications within the surrounding genomic region. These modifications might not only indirectly impact the expression of neighboring genes critical for plant viability at that specific locus but could also, potentially through trans-chromosomal interactions or mobile epigenetic factors, spread to influence the epigenetic state and expression of the other *DGAT1* homeologs located on

chromosomes 01 and 19, given their high sequence homology. Such genome-wide or homeolog-specific epigenetic dysregulation could contribute to the observed lethal phenotype. Finally, while deemed improbable, the possibility of synthetic lethality arising from an interaction between the *dgat1* mutation and other subtle, uncharacterized genetic factors unique to this specific plant line cannot be definitively excluded. Investigating this hypothesis would necessitate comprehensive whole-genome sequencing and comparative genomic analysis.

While the conserved seed oil and fatty acid profile phenotypes across *Arabidopsis thaliana*, *Camelina sativa*, and *Thlaspi arvense dgat1* mutants point to a consistent role in lipid metabolism, the divergent developmental phenotypes highlight the complexity of gene function in polyploid systems. Further research, including detailed genetic and expression analysis of *Camelina* mutant lines, is necessary to elucidate the precise mechanisms underlying the functions of the DGAT1 enzyme.

2.5: Figures

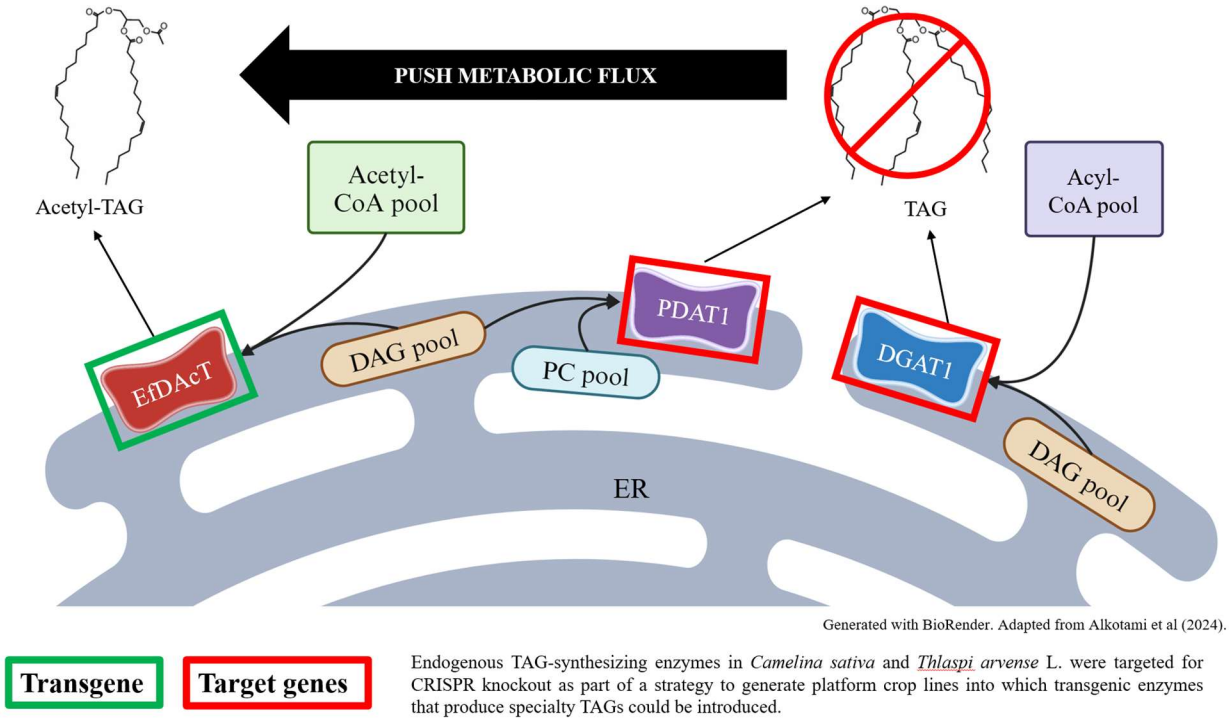

Figure 2.1: Metabolic engineering strategy of DGAT1 and PDAT1 CRISPR knockouts. This diagram illustrates the targeted enzymes, DGAT1 and PDAT1 in the ER membrane, and intracellular sources of substrate pools and products. The elimination of TAG-biosynthesizing enzymes DGAT1 and PDAT1 is hypothesized to reduce competition to transgenic enzymes and improve their efficiency.

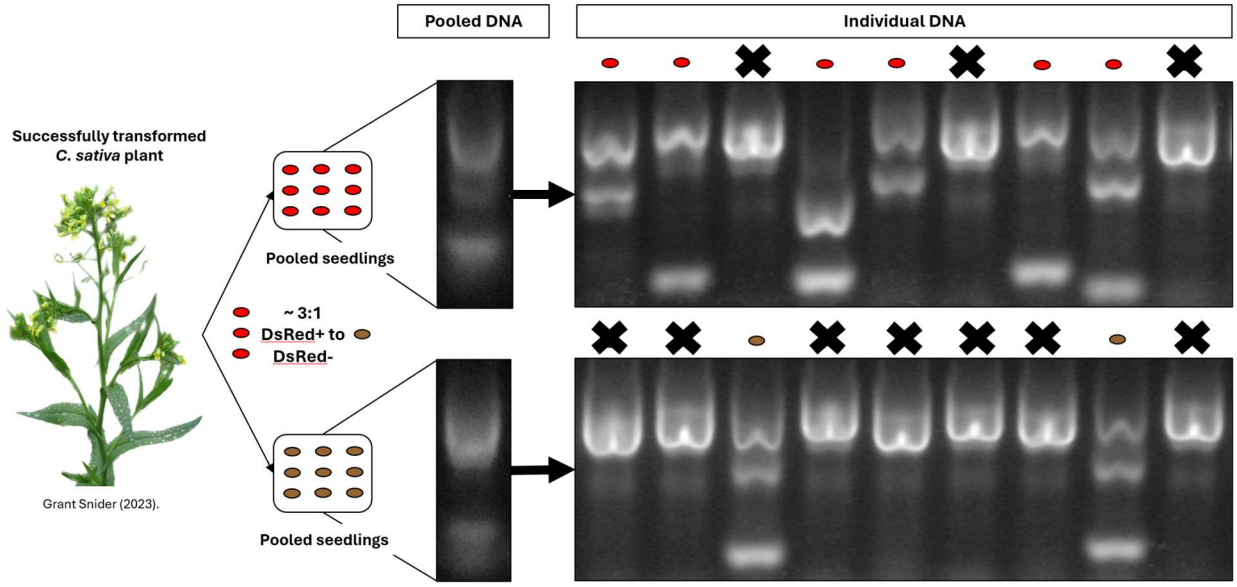

Figure 2.2: Diagram of the high-throughput mutant screening and detection method. Genomic DNA was extracted from pooled leaf tissue samples, with each pool comprising tissue from up to ten individual T2 generation plants. Universal primers flanking multiple CRISPR-Cas9 target sites within the gene of interest were employed for PCR amplification, generating a product encompassing the targeted region. The presence of ASPs on gel electrophoresis of these pooled samples indicated potential large-DNA excision mutations. Subsequently, individual T2 plants contributing to ASP-positive pools were resampled, and their genomic DNA was subjected to independent PCR analysis to identify the specific plant(s) harboring the mutation. These ASPs originated from large-DNA excision events mediated by the activity of CRISPR-Cas9 at multiple target sites, followed by NHEJ repair of the resultant DNA termini. Plants exhibiting detectable ASPs, indicative of a mutation, were selected for propagation and further characterization, while plants displaying the wild-type amplicon profile were discarded (as indicated by an X).

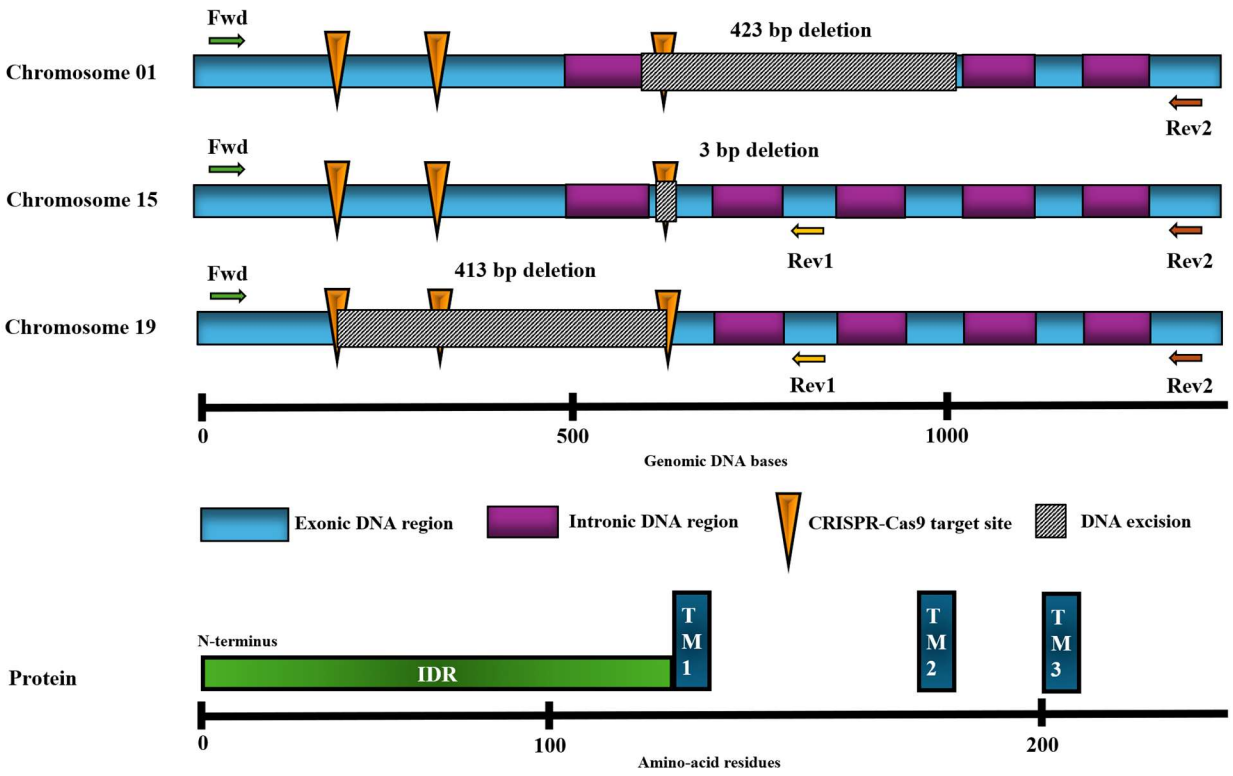

Figure 2.3: *Camelina sativa dgat1-/-/-* genomic DNA mutations diagram in triple knockout line. A graphic summary detailing the homozygous DNA mutations in the characterized *Cs dgat1* triple mutant. Schematics from each homeologous chromosome represent genomic DNA encoding the corresponding genes. The 5'-3' orientation of the non-template strand is shown. The position of primers indicates the genomic regions that were amplified in PCR reactions. The protein schematic indicates domains encoded by the genomic DNA aligned with deletions.

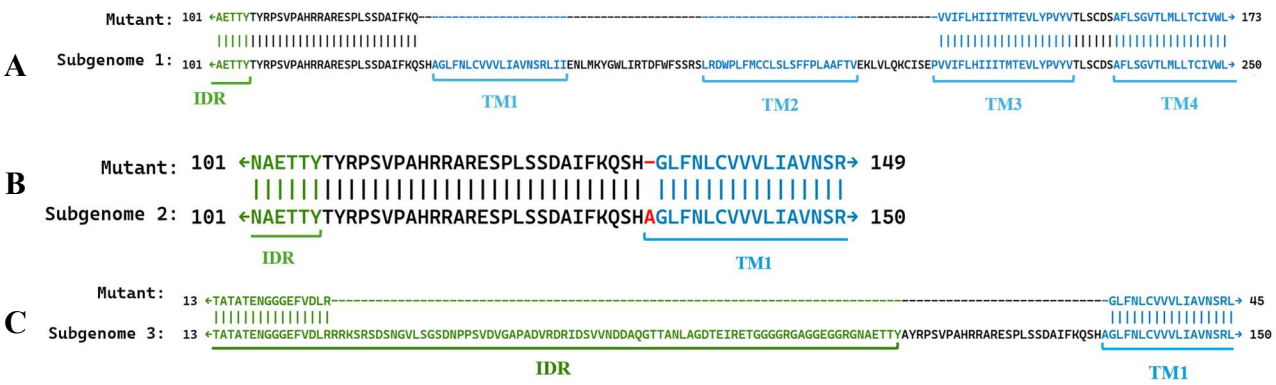

Figure 2.4: ***Camelina sativa dgat1-/-/-*** **amino acid sequences in triple knockout line.** Mutations in all three homeologous chromosomes were in-frame. Primary sequence alignments show deleted amino acid residues and motifs of the protein product. **A.** Large-DNA excision mutation in chromosome 01 resulted in the entire deletion of transmembrane domain 1 and transmembrane domain 2 and deletion of the first residue comprising transmembrane domain 3. **B.** The three base pair deletion in chromosome 15 resulted in the deletion of Alanine-134 (red), the first amino-acid residue composing transmembrane domain 1. **C.** The large-DNA excision in chromosome 20 resulted in the deletion of most of the N-terminus intrinsically disordered region and the first amino acid residue composing transmembrane domain 1.

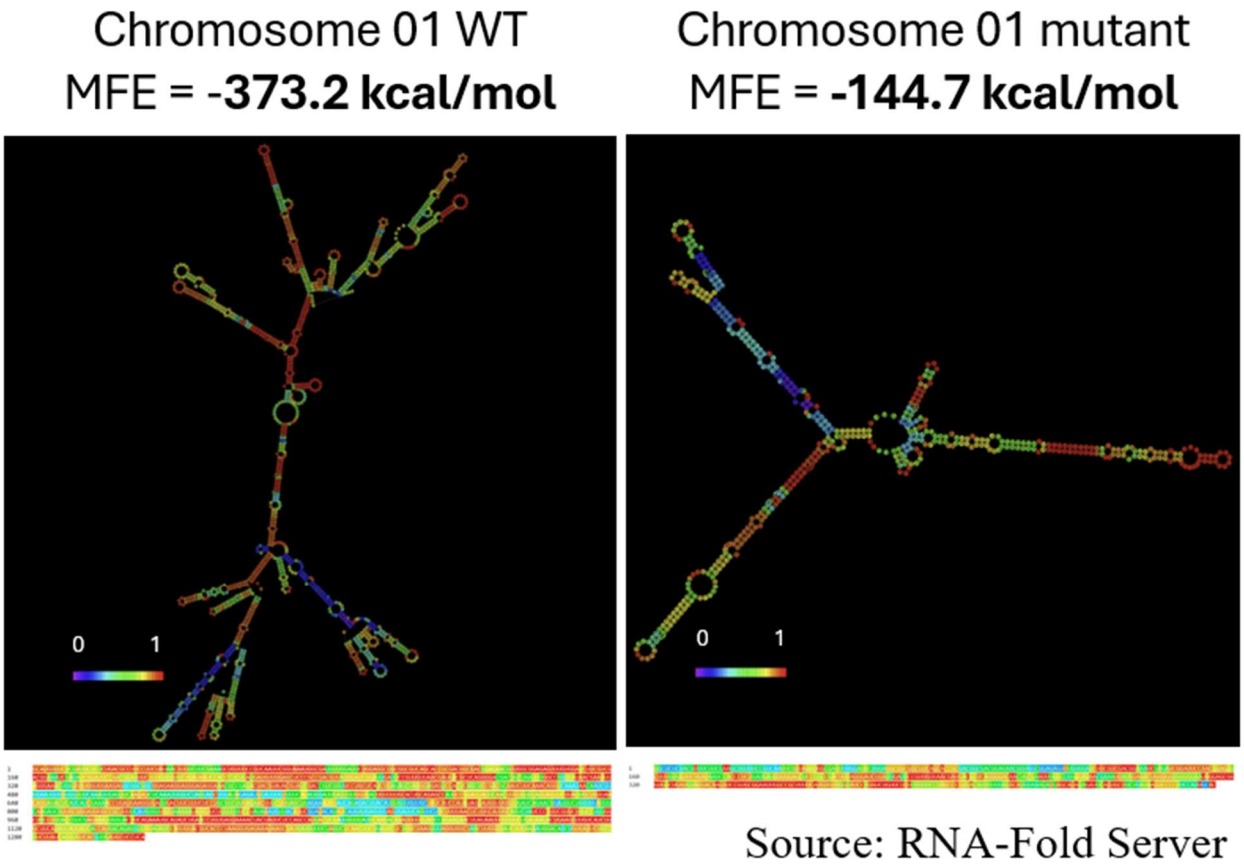

Figure 2.5: Predicted secondary structures in *CsDGAT1* amplicons. Predicted self-dimerization of nucleotides within the wild-type amplicon (left) and the chromosome 01 mutant amplicon (right) shows the minimum-free energy (MFE) values, with more negative values indicating a more stable structure, estimated by RNA-Fold Server (Gruber et al., 2008).

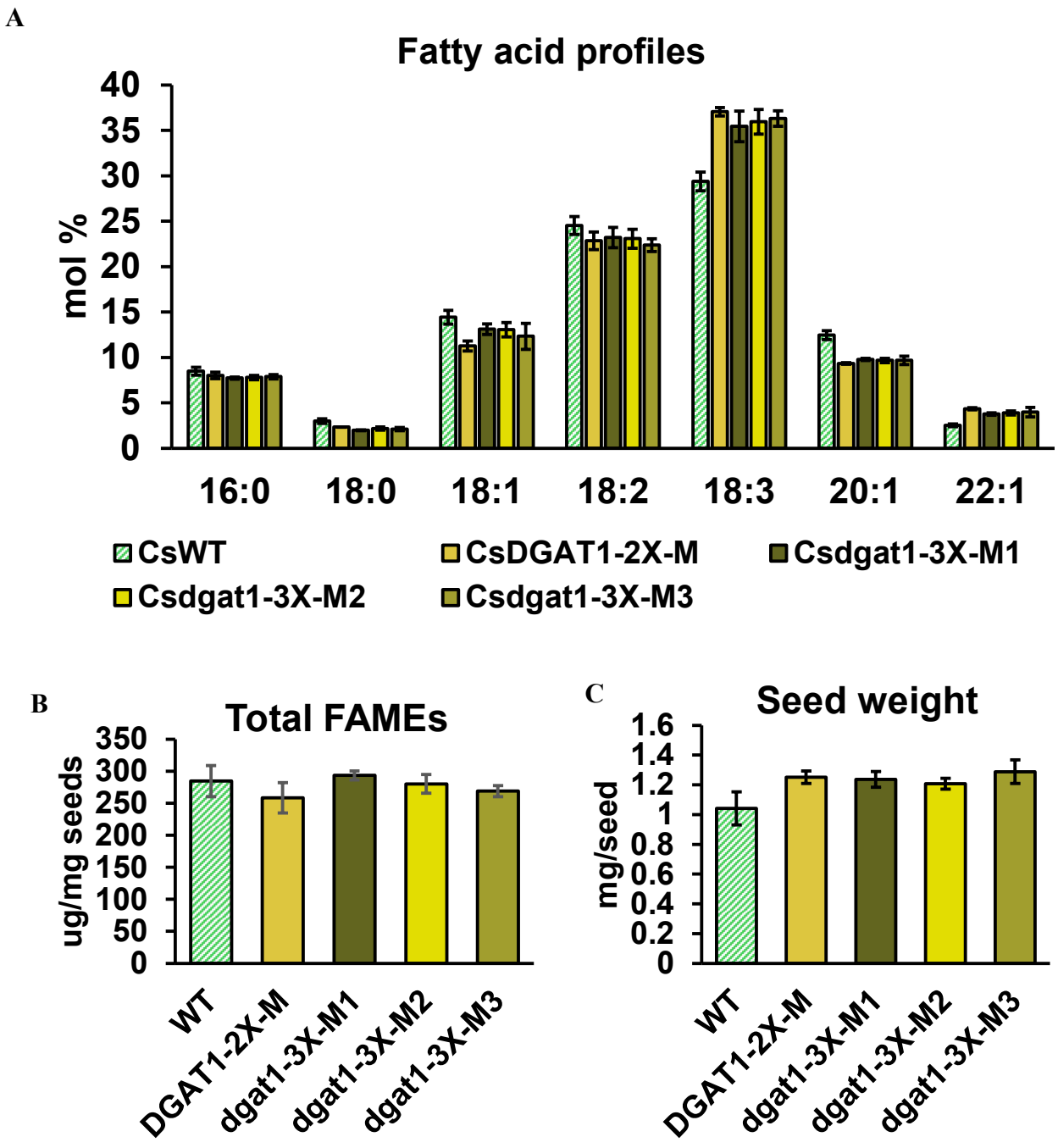

Figure 2.6: Double-knockout and triple-knockout seed oil and fatty acid profiles in *Camelina sativa DGAT1* lines. A., B. and C. Three biological replicates of each line were averaged. Three unique lines (M1, M2, M3) homozygous (inherited) for the large-DNA excision mutation in chromosome 19 and homozygous for mutations in chromosome 01 and chromosome 15 (3X – all homeologs) were compared to one line wild-type homozygous in chromosome 19 and homozygous for mutations in chromosome 01 and chromosome 15 (labeled 2X-M).

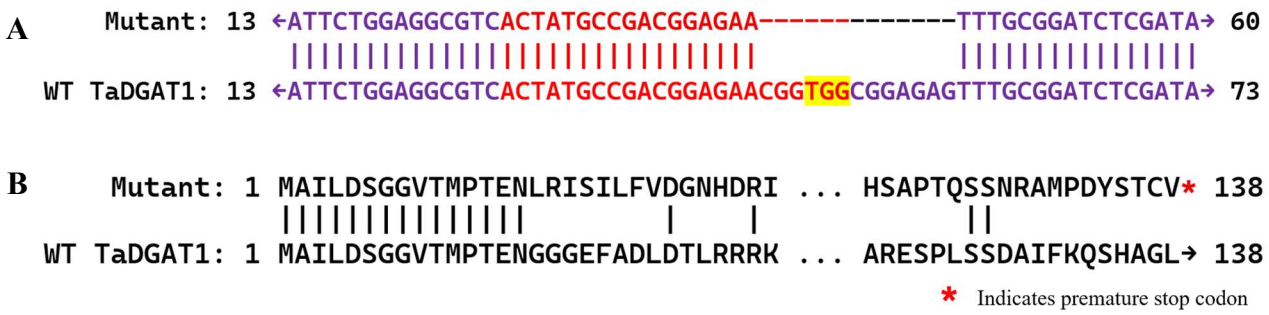

Figure 2.7: *Thlaspi arvense tt8 dgat1* mutant genomic DNA. A. The sequence of a 13 bp deletion that occurred near target site 1 is shown. Purple shading indicates exonic DNA regions. Red shading indicates the Cas9 target site with the PAM site highlighted yellow. **B.** The resulting frameshift mutation resulted in a predicted truncated protein rendered by a premature stop codon.

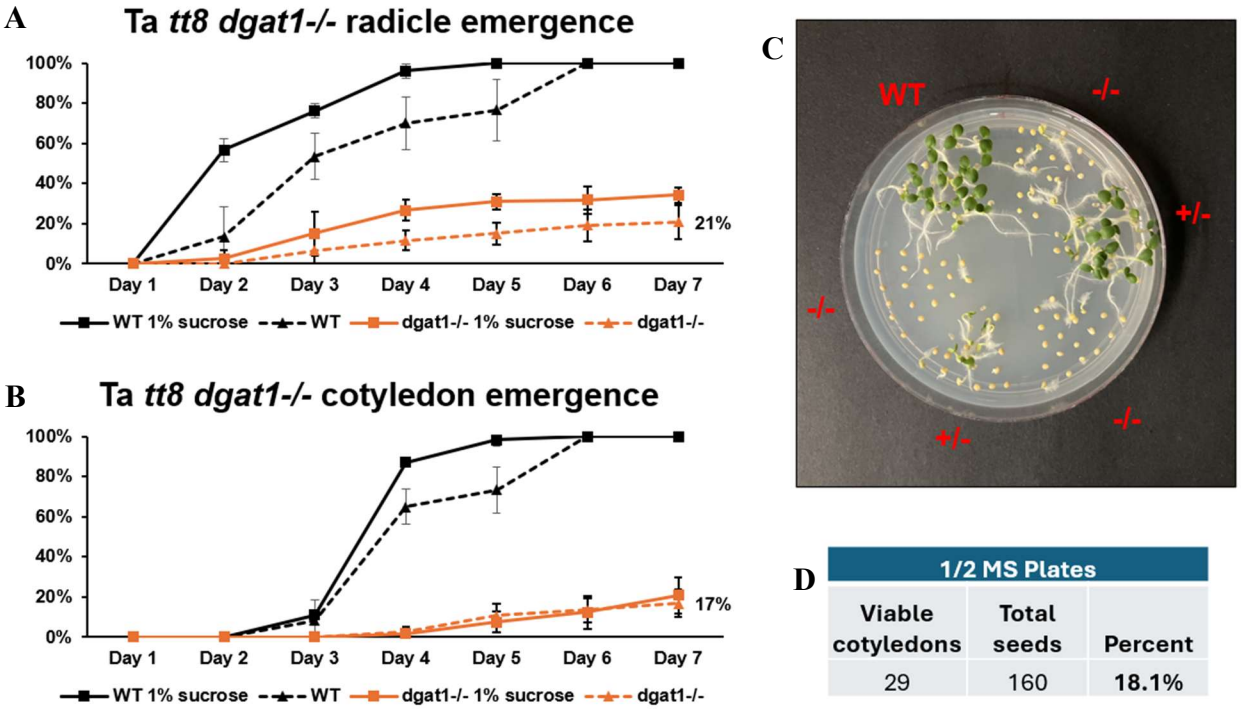

Figure 2.8: *Thlaspi arvense tt8 dgat1* germination phenotypes. A. Radicle emergence was tabulated each day for a period of seven days from wild-type and *DGAT1* mutant lines. Visible radicle emerged from seed coat constituted emergence. **B.** Cotyledon leaf visualization and separation constituted emergence. **C**. This 1% MS petri plate shows seeds from wild-type and mutant genotypes on day 5 of germination experiment. **D.** Table shows 18.1% of seeds in the mutant background established viable cotyledons 19 days after plating.

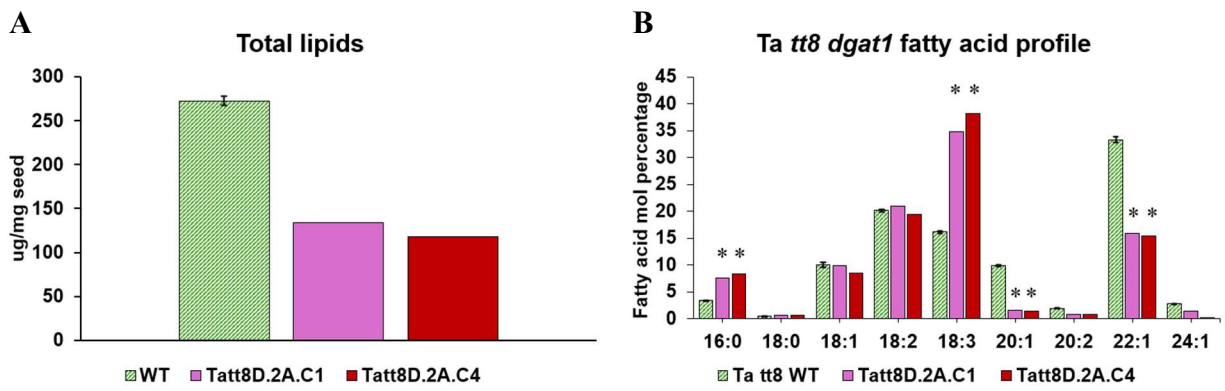

Figure 2.9: ***Thlaspi arvense tt8 dgat1*** **seed phenotypes. A. and B.** Total seed lipid content and fatty acid profiles are compared from three averaged wild-type replicates to two individual *dgat1* knockout mutants. Significant differences in specific fatty acids in the knockout lines are denoted by an asterisk.

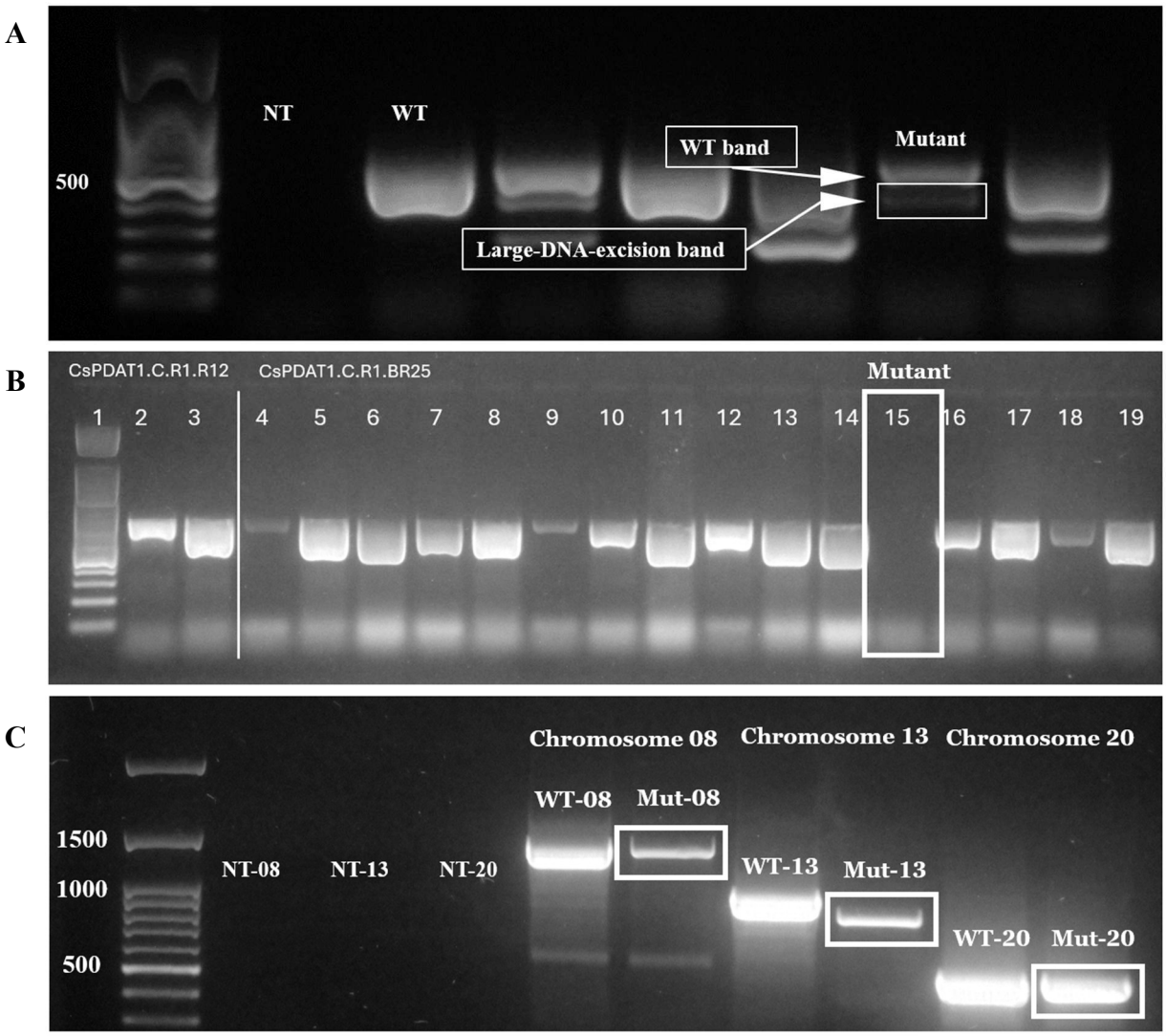

Figure 2.10: ***Camelina sativa pdat1*** **mutant identification. A.** Mutant screening gel shows ASP in lane labeled "mutant" indicating a large-DNA excision mutation in this T4 plant. **B.** Cas9 screening gel indicated the absence of the transgene in the mutant plant. **C.** Homeolog-specific primers amplified a region of *CsPDAT1* on each homeologous chromosome for Sanger sequencing. The single amplicon band in chromosome 13 in the mutant (Mut-13) runs approximately 100 base pairs below the wild-type. Non-template (NT) controls lacking template DNA were run alongside wild-type (WT) controls and samples.

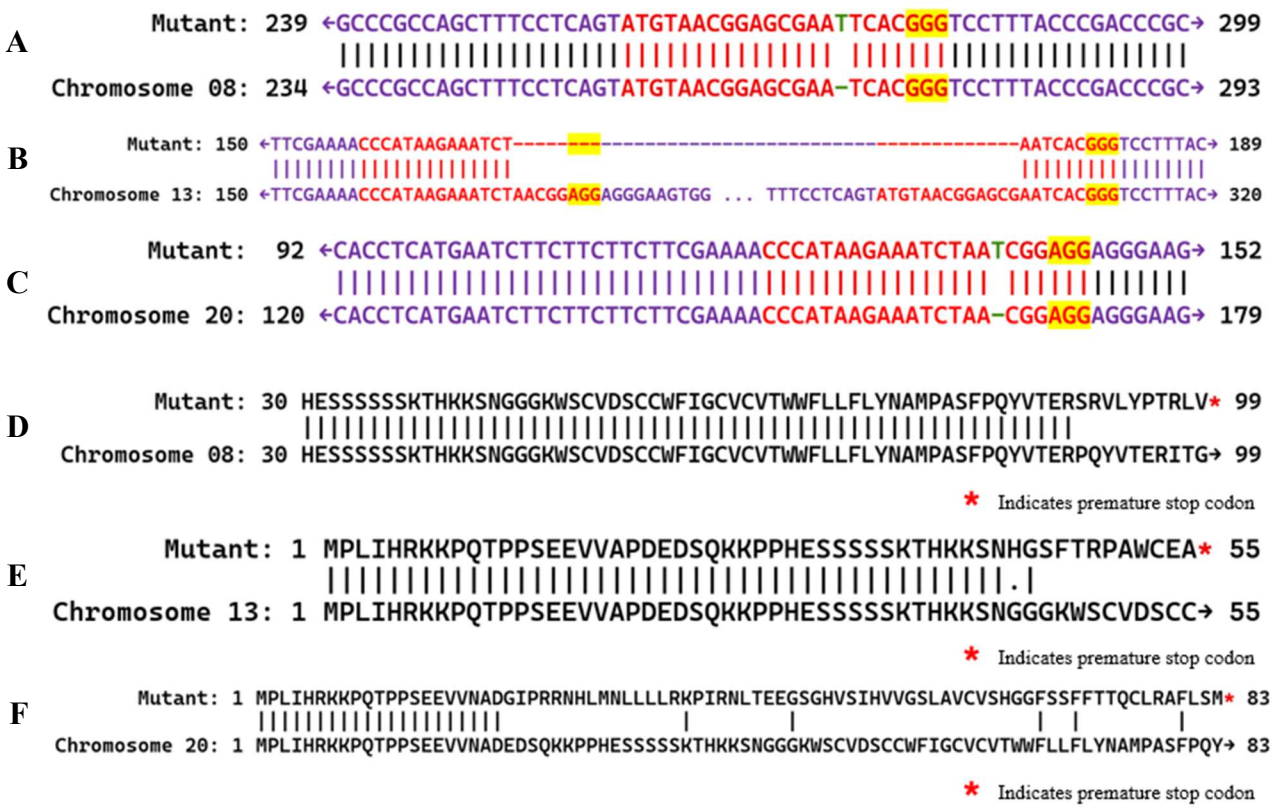

Figure 2.11: Genomic DNA of *Camelina sativa pdat1-/-/-* mutant and predicted primary sequences in all three subgenomes. Exonic DNA regions are shaded purple. Cas9 target sites are shaded red with PAM sites highlighted yellow. **A.** A one base pair insertion occurred in target site 2 in chromosome 08. **B.** A large-DNA excision mutation occurred in chromosome 08 between target sites 1 and 2. **C.** A one base pair insertion occurred in target site 1 in chromosome 20. **D.** The one bp insertion in chromosome 08 resulted in a premature stop codon in the primary sequence of the mutant protein. **E.** The large-DNA excision mutation in chromosome 13 B resulted in a premature stop codon encoded in the mutant protein's primary sequence. **F.** The mutation in chromosome 20 resulted in a premature stop codon in the protein's primary sequence.

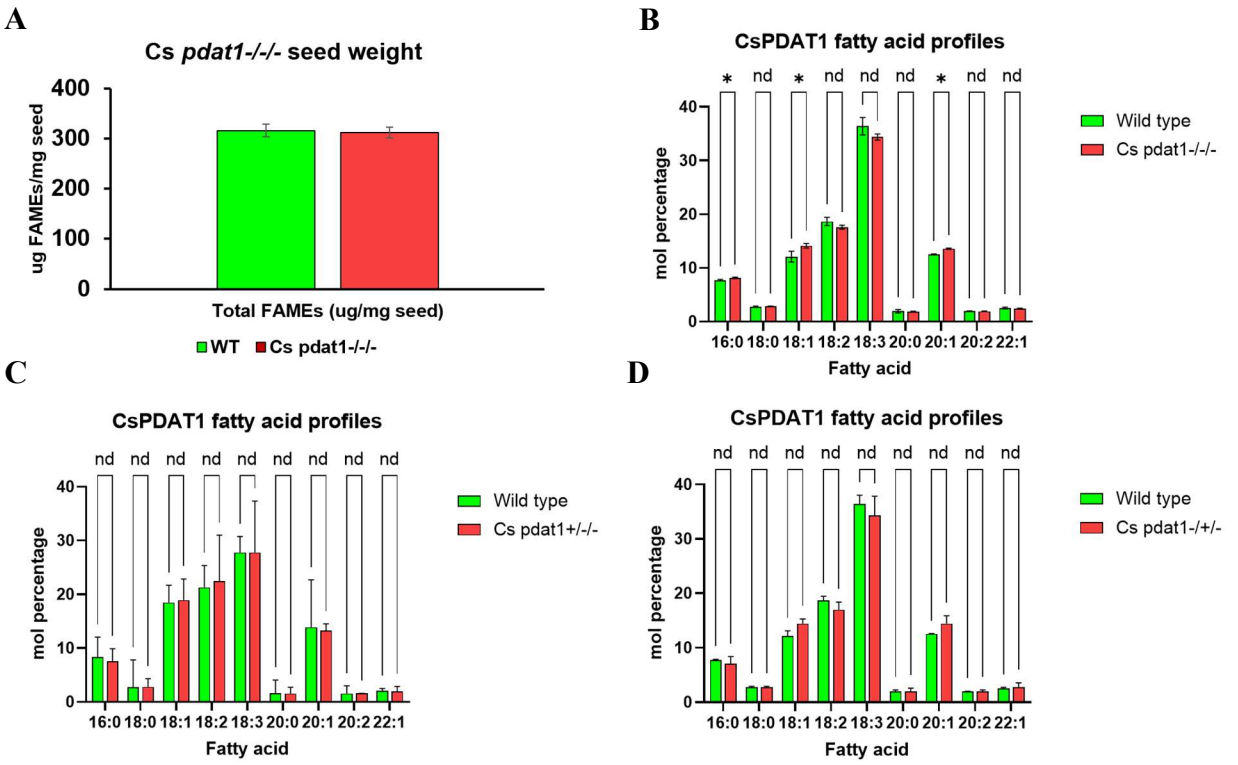

Figure 2.12: Seed oil and fatty acid profiles of *Camelina sativa pdat1* mutants. A. Total FAMEs, used to measure total seed oil content, showed no significant difference between the wild-type and the complete-knockout mutant. Four biological replicates of wild-type plants were compared to eleven biological replicates of *Cs pdat1-/-/-* grown at the same time in the same conditions. **B.** Fatty acid composition was measured in four biological replicates of wild-type *Camelina sativa* and eleven biological replicates of *Cs pdat1-/-/-* grown together. Three fatty acids (C16:0, C18:1 and C20:1) presented slight but statistically significant differences. **C.** Fatty acid profile of one wild-type plant was compared to two individual biological replicates of *Cs pdat1+/-/-* (wild-type in chromosome 08) grown together. **D.** Three biological replicates of wild-type *Camelina sativa* were compared to three individual biological replicates of *Cs pdat1-/+/-* (wild-type in chromosome 13) grown at the same time.

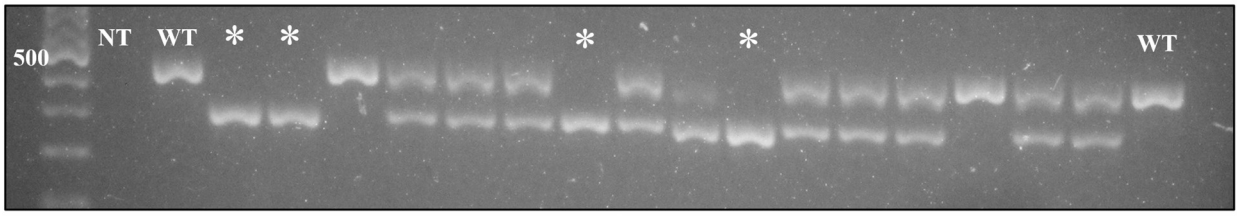

Figure 2.13: *Thlaspi arvense tt8 pdat1* mutant zygosity screening gel. A segment of the *DGAT1* gene was amplified. Wells denoted with an asterisk indicate plants lacking the wild-type allele in *DGAT1* based on the results of this ASP screening gel.

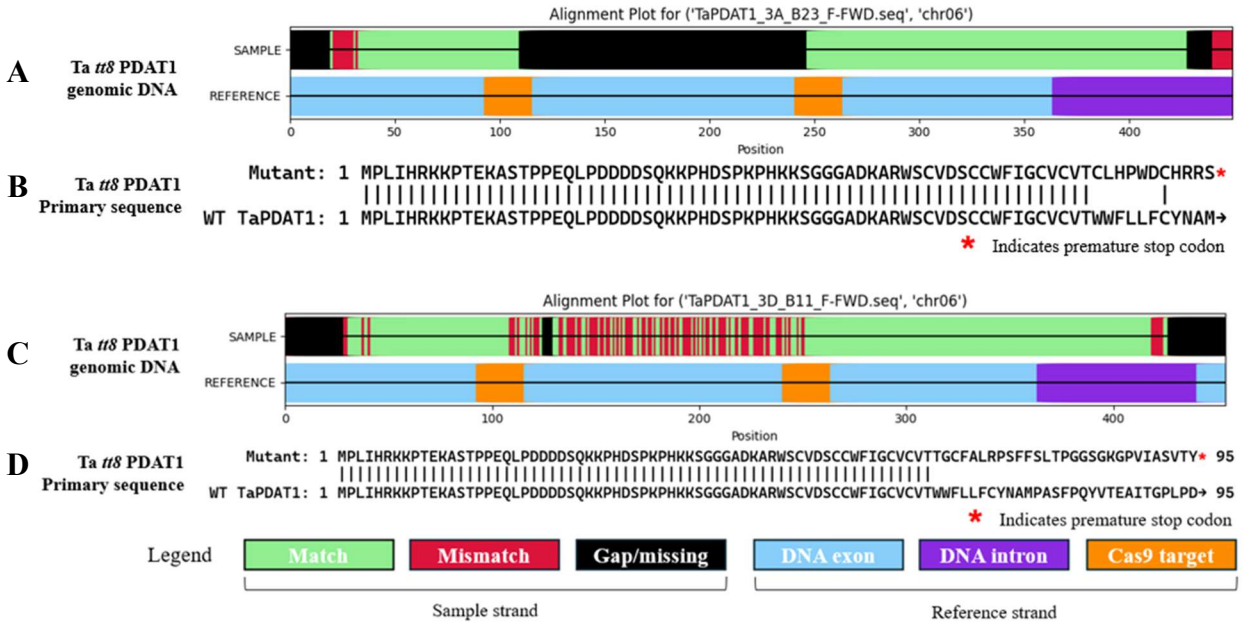

Figure 2.14: *Thlaspi arvense tt8 pdat1* mutant identification. A custom Python script was utilized to align and annotate sample and reference strands to visualize mutations allowing for simple detection. **A.** The mutant detection script identified a large-DNA excision mutation between target sites in one independent plant line. **B.** The mutation results in a premature stop codon in the mutant protein product. **C.** A second mutation was identified by the mutant detection script. Sequencing determined the mutation was caused by a DNA inversion. **D.** A premature stop codon occurred in the resulting mutant protein product.

2.6: Tables

	T1 screened	T1 mutants	T1 efficiency	T2 screened	T2 mutants	T2 efficiency
CsDGAT1	12	0	0%	171	2	**1.17%**
CsPDAT1	16	0	0%	60	3	**5.00%**
CsOBL1	30	0	0%	234	35	**14.96%**
CsSDP1	15	0	0%	226	82	**36.28%**
CsGDSL1	12	0	0%	145	15	**10.34%**

Table 2.1: T2 mutant efficiency rates of various CRISPR-Cas9 gene target constructs. ASPs observed in mutant screening gels constitute positive mutant identification. This calculation does not include mutations undetectable by PCR screening gels, including small indels and large inversions.

Chapter 3: Intronic-DNA CRISPR structural variants

3.1: Introduction

Although CRISPR-Cas systems represent the most precise genome editing technology available, their use can inadvertently induce large genetic alterations known as structural variants (SVs), which are defined as genomic rearrangements typically involving DNA segments larger than 50 base pairs, encompassing deletions, insertions, inversions, duplications, or translocations (van Belzen et al., 2021). SVs can arise naturally through various mutagenic processes, and they are responsible for driving genetic diversity and contributing to species differentiation and genome evolution. However, they have also been specifically associated with unintended off-target activity of CRISPR-Cas systems. These unintended genomic changes, occurring at genomic sites other than the intended pairing site of gRNAs, are collectively termed "off-target effects." While off-target events most commonly occur at sites exhibiting sequence similarity to the guide RNA, they have also been identified, albeit less frequently, at genomic locations significantly distant from the intended cleavage site, including separate chromosomes. These unintended SVs can disrupt genes, alter gene regulation, and potentially cause adverse cellular outcomes. Such off-target events pose significant challenges to achieving the level of precision required for therapeutic and research applications.

The most common source of off-target effects is non-specific binding, which occurs when genomic DNA shares a high percentage of sequence homology with the intended target site encoded by the gRNA (Fu et al., 2013). Computational tools that aid the design of sgRNAs can predict off-target cut sites to reduce the occurrence of these errors (Montague et al., 2014; Doench et al., 2016). Engineering of Cas9 enzymes for greater accuracy in target site recognition has further mitigated this type of error (Slaymaker et al., 2015). However, much less frequently,

other errors are observed in Cas9 experiments unrelated to non-specific binding. Large deletions (Kosicki et al., 2018), chromothripsis (Leibowitz et al., 2021), and other large SVs (Höijer et al., 2022; Hwang et al., 2024) have been observed in CRISPR-Cas9 experiments. A large deletion was observed in *CsDGAT1* CRISPR lines in chromosome 01 in this experiment that did not correspond to target sites (Chapter 2, Figure 2.3, chromosome 01). The mechanisms underlying these uncommon SVs induced by CRISPR-Cas9 are not well understood.

The occurrence of aberrant SVs in CRISPR-based genome editing experiments is challenging to predict, as the mutations induced are neither consistent nor reproducible. Although Cas9 evolved naturally in bacteria, its deployment in mammalian and plant genomes involves complex protein-DNA interactions that remain incompletely understood. These unpredictable off-target effects typically pose minimal concern in plant genetic engineering, where screening processes routinely evaluate hundreds of individuals; plants exhibiting significant aberrant SVs are naturally eliminated during selection and propagation. However, the precision required in therapeutic gene-editing applications demands higher accuracy. To reduce unintended genomic alterations, engineered Cas9 variants, such as nickases designed to induce single-strand DNA break (Shen et al., 2014), and base editors, which facilitate precise base modifications without double-strand breaks (Komor et al., 2016), are commonly employed. Nevertheless, even these refined methods have been reported to induce unintended structural variants (Wu et al., 2024), underscoring the necessity for continued improvement in editing fidelity.

To accurately elucidate the mechanisms underlying aberrant SVs induced by CRISPR-Cas9, the resulting mutations must be reproducible across independent events. In *Camelina sativa* CRISPR lines targeting the esterase/lipase GDSL1 (Glycine Aspartate Serine Leucine 1) (Table 3.1), a striking example of such reproducibility was observed: four discrete intronic

regions, ranging from 73 to 97 base pairs in length, were precisely excised in three independent T3 lines (Figure 3.1). DNA sequence analysis confirmed that these deletions occurred at identical positions in each line. Notably, these same intronic regions were intact in the corresponding T1 generation, indicating that the deletions were not inherited but arose independently in each T3 line. This convergence of identical structural variants across independent transformation events strongly suggests a reproducible, Cas9-mediated mutational mechanism, rather than random genomic instability.

Elucidating the molecular mechanisms underlying aberrant SVs induced by CRISPR-Cas9 holds broad implications, not only for plant biology, but also for the advancement of human gene therapies. A deeper understanding of how the Cas9 complex interacts with DNA-associated proteins in eukaryotic systems, including those of humans and other mammals, will be critical for improving the precision, safety, and therapeutic potential of genome-editing technologies. Although this study lacked the time and resources to pursue a mechanistic investigation of the observed phenomenon, we report this reproducible and unexpected finding in the hope that it will prompt further research into the causes and consequences of CRISPR-induced SVs.

3.2: Methods

3.2.1 Construction of plasmids

The same CRISPR-Cas9 cloning methods used in Chapter 2 were applied in this experiment. CRISPR target sites were designed to specifically target the *Camelina sativa GDSL1* gene in three locations universal to each homeolog. Detailed methods can be found in Chapter 2, Section 2.1.

3.2.2 Plant transformation and propagation

The same plant transformation and propagation protocol described in Chapter 2 was applied here. Detailed methods can be found in Chapter 2, Section 2.2.

3.3.3 Mutant identification

Plants were screened for mutations by the detection of ASPs as described in detail in Chapter 2, section 2.3. In these lines, complete mutants were identified by a single DNA band in electrophoresis gels running lower than wild-type DNA bands. PCR primers were designed to universally amplify the same portion of the *Camelina sativa GDSL1* gene in all three homeologous chromosomes. After DNA sequencing revealed the excision of two intronic DNA regions in a sample, a new reverse primer was designed to flank and amplify all four proximate intronic DNA regions in the *GDSL1* genomic DNA.

3.3: Results

Mutant screening gels for *Camelina sativa* GDSL1 at the T3 generation revealed that large DNA excision mutations had occurred in all three subgenomes in many of the samples (Figure 3.2). In these cases, the absence of the wild-type amplicon band indicated that all three homeologous copies of the target gene had been successfully mutated, suggesting efficient and simultaneous editing across the hexaploid genome.

Sequencing of the mutant amplicons revealed large DNA excision events; however, the observed cut sites did not correspond directly to the predicted guide RNA target sites. Instead, two intronic regions of the *GDSL1* gene were found to be precisely excised, while the intervening exonic regions remained intact. Specifically, two distinct deletions were observed: one corresponding exactly to intronic region 1 and the other to intronic region 2. Notably, the exonic region situated between these introns, which contained CRISPR target site 3, was

preserved in the genomic DNA of the mutant plants. Based on the reference genome, two additional intronic DNA regions were identified near the deleted ones. To test for their deletion, a new reverse primer was designed to span all four introns. Sequencing using this primer confirmed that all four intronic regions were precisely deleted in the mutant samples (Figure 3.3).

The precise deletion of four intronic DNA regions within all six homeologous copies of the *GDSL1* gene in *Camelina sativa* was confirmed by sequencing in at least three independent lines, each derived from separate *Agrobacterium*-mediated transformation events. Remarkably, these deletions were identical in size and position across all three lines, indicating a reproducible SV induced by Cas9 activity. While additional T3 lines displayed mutant genotyping profiles consistent with these deletions, sequencing was not performed on all samples; thus, the true number of independently edited lines harboring the same mutation is likely underestimated. In the T1 generation, mutant screening gels showed no evidence of large DNA excision, as experimental samples exhibited amplicon sizes identical to wild-type controls.

The *GDSL1* gene was sequenced in multiple *Camelina sativa* CRISPR-Cas9 lines targeting genes other than *GDSL1*. In all cases, the four intronic DNA regions remained intact, and no deletions were observed, indicating that the excision of these introns is specific to CRISPR activity directed at *GDSL1*. To investigate whether these intronic regions possess any features that might predispose them to deletion, each was queried using BLASTn against the full NCBI nucleotide database without restriction to plant species. However, no conserved motifs, sequence similarities, or other notable characteristics were identified in any of the four regions. These results suggest that the observed deletions are not explained by known sequence-based

features and may result from unknown interactions between Cas9 activity and local chromatin or structural context.

3.4: Discussion

At the T3 generation, the *GDSL1* CRISPR lines were the only lines among the five gene targets in this study to show complete absence of the wild-type amplicon in mutant screening gels. In these apparent full knockouts, each sample consistently displayed a single mutant DNA band on the gel, suggesting that large-DNA excisions had occurred in all six gene copies of *GDSL1* among all three homeologous chromosomes. The uniform size of the mutant band across numerous samples implied that the same CRISPR target sites were utilized by the Cas9 complex in each mutation. This highly consistent and unexpected result first signaled the presence of an aberration specific to the *GDSL1* target, prompting further investigation into the nature and reproducibility of the underlying structural variant.

By the T3 generation, over half of the analyzed plants lacked the wild-type amplicon band, suggesting that large intronic deletions had become prevalent in a significant portion of the mutant population. The prevalence of the intronic-DNA deletions suggests that this deletion had occurred in the previous generation (T2) as well and was passed in part by inheritance. Polymorphic amplicon bands consistent with the expected intronic deletions emerged in T2 plants, though no samples showed the absence of the wild-type band. Although these bands matched the predicted size of the deletion allele, they were not sequenced and therefore remain putative. Whether or not all six gene copies were mutated in a single generation remains unknown at this time.

The Cas9 complex initiates genome scanning by searching for protospacer adjacent motif (PAM) sites along chromosomal DNA. Upon PAM recognition, the gRNA attempts to hybridize

with the complementary DNA strand. If binding fails, the complex resumes scanning. The intronic regions of *GDSL1* are excised only when CRISPR-Cas9 specifically targets the *GDSL1* locus. These regions remain intact when other genes are targeted (Figure 3.4). This finding strongly suggests that recognition of the target site by the Cas9-gRNA complex, triggers the excision of these intronic regions. The preservation of the exonic DNA flanking the deleted introns further supports the hypothesis that this excision mechanism is activated after target site recognition, but prior to, or in lieu of, dsDNA cleavage. A wild-type sample was sequenced to rule out reference errors (Figure 3.5).

The four excised intronic DNA regions did not display any identifiable transposable element motifs based on sequence analysis. However, in plants where these introns were excised, large-DNA excision mutations were also observed at the same alleles across independent homeologs. This suggests that in some cases, the Cas9 complex may have generated a double-stranded break before the cellular machinery responsible for the intron deletions could act. We hypothesize that the excised intronic regions represent previously uncharacterized Class II (DNA-based) transposable elements, and that DNA-associated proteins normally involved in suppressing their mobilization may be disrupted by Cas9 target site recognition or local chromatin remodeling. We further speculate that Cas9 activity induces stress or structural instability in the genome, indirectly activating a transposition-like mechanism responsible for the precise excision of these regions. The molecular basis of this mechanism remains unknown and warrants further investigation.

Understanding atypical outcomes of CRISPR activity, such as this one, is critical to improving the precision and safety of genome editing technologies. Notably, homologous intronic regions are present in the orthologous *GDSL1* gene of the model species *Arabidopsis*

thaliana, offering a genetically tractable system to investigate the underlying molecular drivers of this phenomenon (Figure 3.6).

3.5: Figures

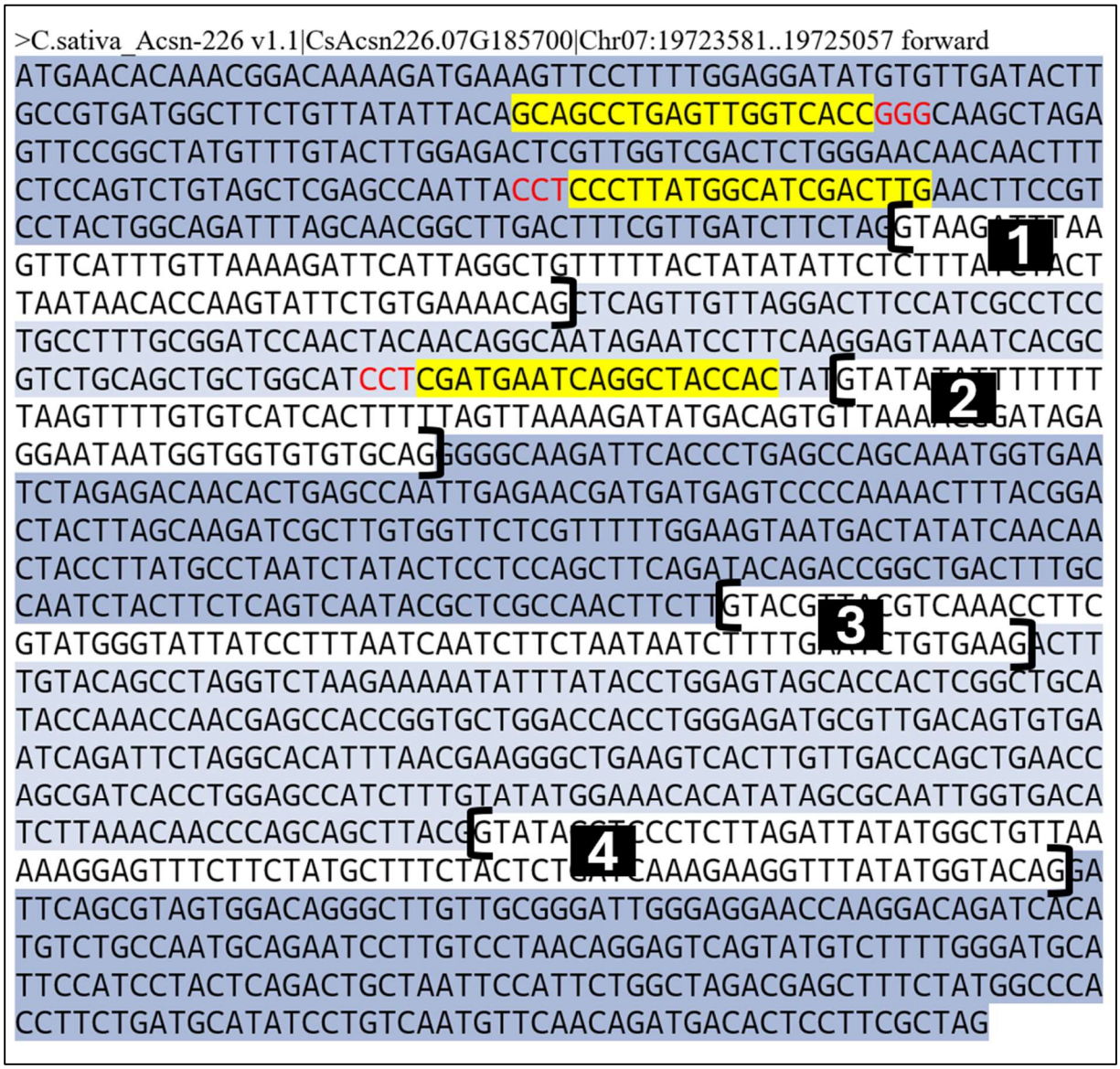

Figure 3.1: Genomic DNA of GDSL1 gene in *Camelina sativa* chromosome 07. Numbering shows four proximate intronic-DNA regions with the *GDSL1* gene in chromosome 07 of *Camelina sativa*. The other two genes on homeologous chromosomes contain homologous intronic-DNA regions. The Cas9 target sites are highlighted in yellow and the PAM sites are denoted in red.

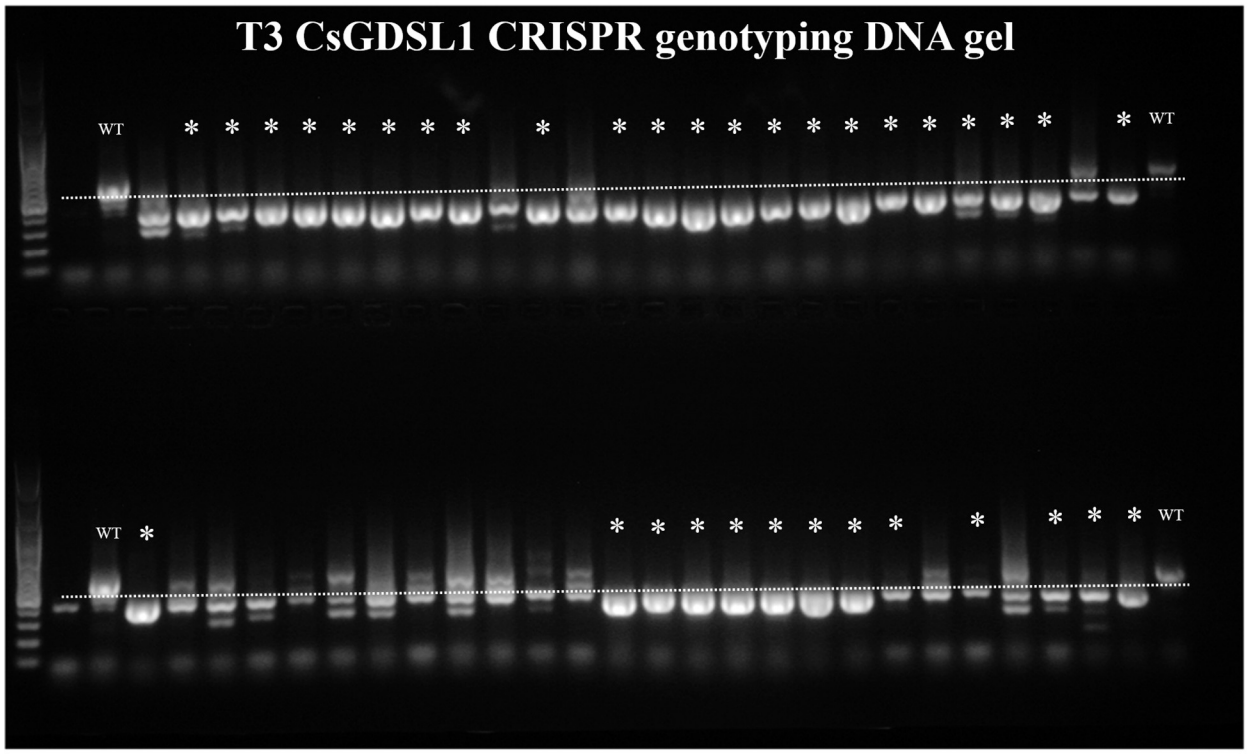

Figure 3.2: Majority of T3 plants in *Camelina sativa GDSL1* CRISPR line lack wild-type DNA band. DNA electrophoresis gel shows wild-type control samples at the left and right ends of the gel flanking CRISPR-mutant samples. A dashed line indicates the expected size of the wild-type DNA band. The lack of a wild-type band is indicated by an asterisk, evidenced by a single DNA band running lower than the wild-type DNA band.

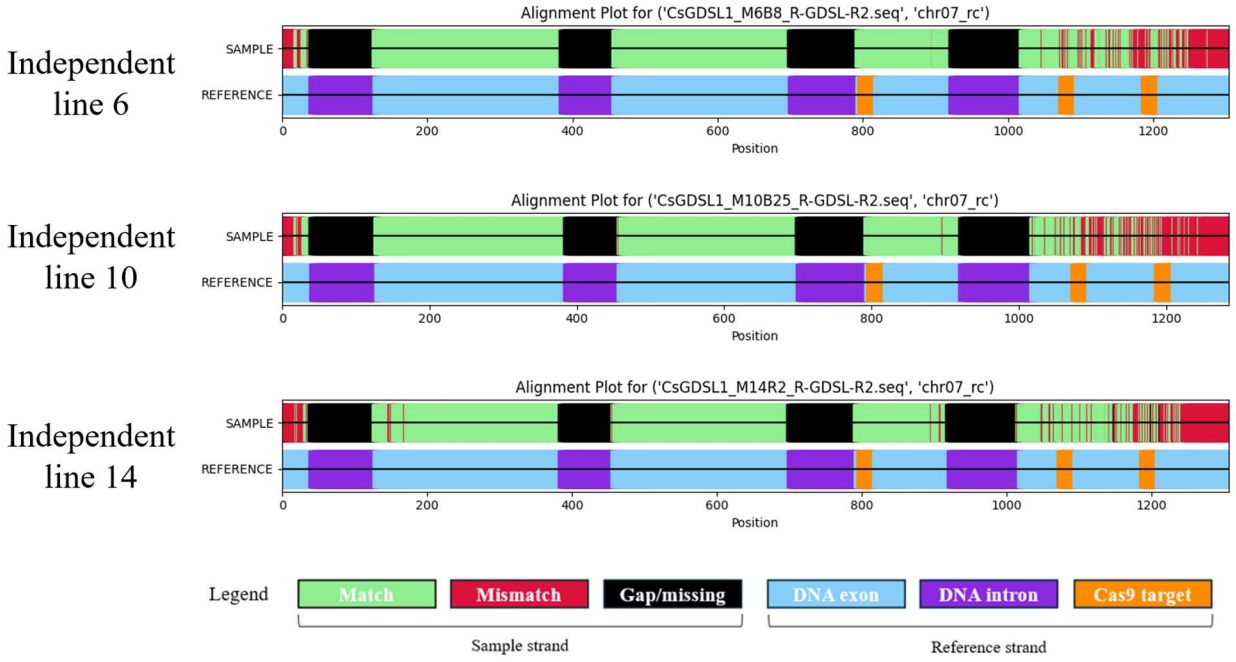

Figure 3.3: Mutant mapping to reference genome. Python script generated alignments based on Sanger sequencing data. These reverse reads flanking all four intronic regions show that DNA excisions (black) map precisely to intronic-DNA regions (purple) on the reference strand in samples taken at T3 from three independent lines. The red mismatch bars at the end of the sequences result from a decline in read quality, which is a common limitation as Sanger sequencing approaches its maximum length.

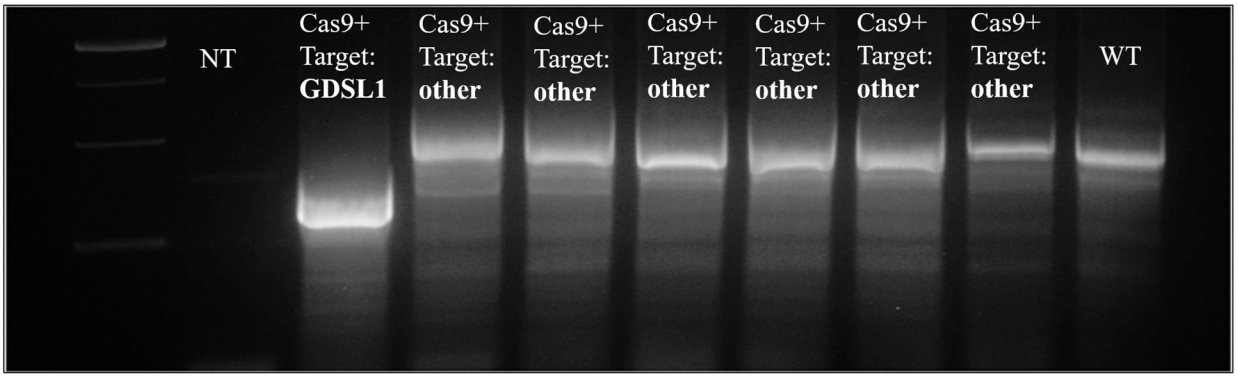

Figure 3.4: *GDSL1* screening in CRISPR-Cas9 lines. The truncated DNA band is present in the CRISPR line targeting *GDSL1,* indicating large-DNA excisions in this line. CRISPR lines targeting other genes show the wild-type band profile in the *GDSL1* gene.

Figure 3.5: Multiple sequence alignment of wild-type amplicon to reference genome. The Sanger sequencing read of the wild-type *GDSL1* gene (top line) in *Camelina sativa* compared to the reference sequence of all three homeologs shows the first two intronic DNA regions of interest remaining intact.

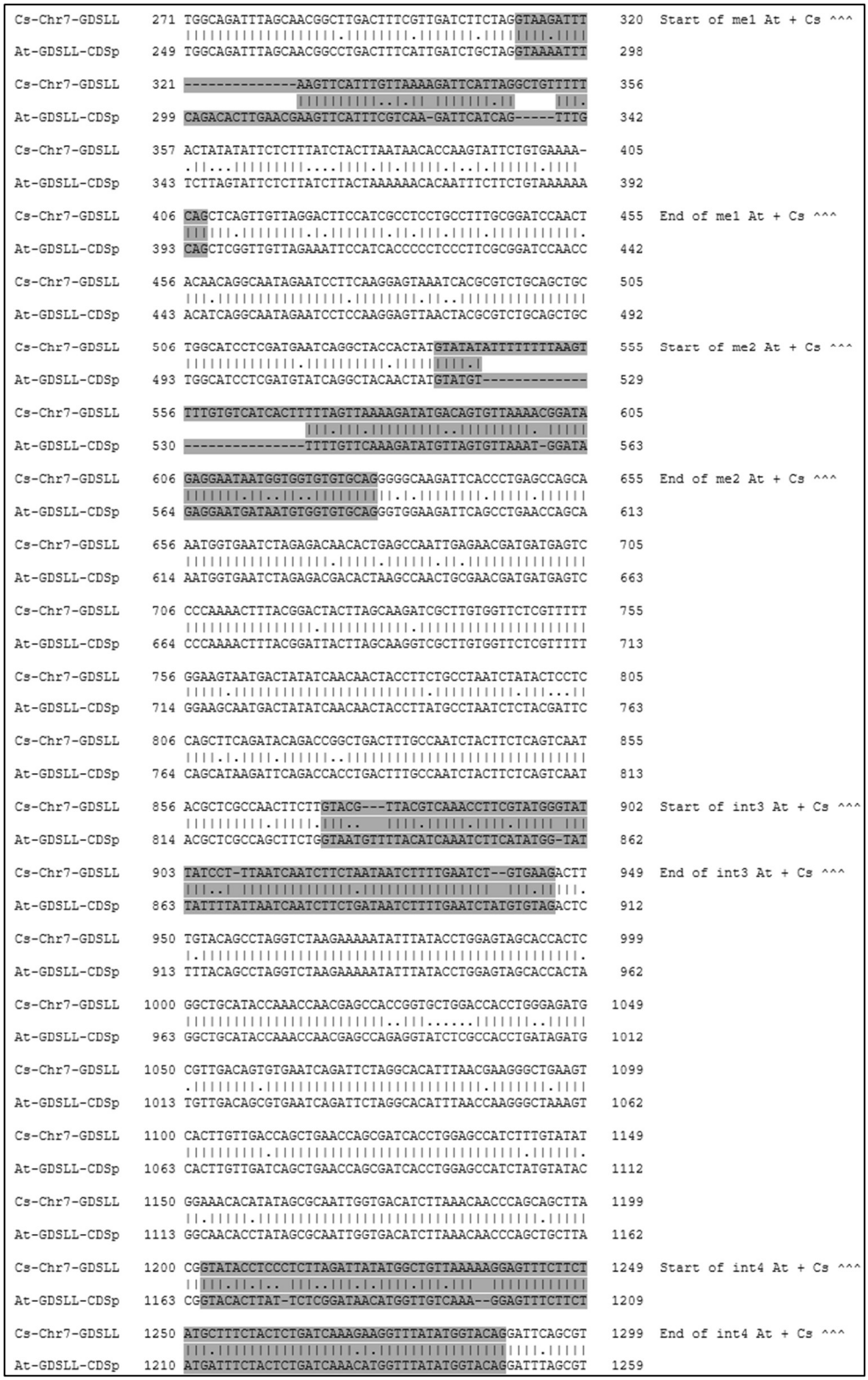

Figure 3.6: Orthologous gene in *Arabidopsis thaliana.* Alignment of orthologous *GDSL1* genes in *Arabidopsis thaliana* and *Camelina sativa* chromosome 07. Four analogous intronic-DNA regions are highlighted gray.

Homeologous gene 1	**CsAcsn226.07G185700**
Homeologous gene 2	**CsAcsn226.09G225800**
Homeologous gene 3	**CsAcsn226.16G181400**
Universal target site 1	**5-GCAGCCTGAGTTGGTCACCGGG-3**
Universal target site 2	**5-CCTCCCTTATGGCATCGACTTG-3**
Universal target site 3	**5-CCTCGATGAATCAGGCTACCAC-3**

Table 3.1: Gene identity and CRISPR-Cas9 target sites. *GDSL1* genome assembly identification of *GDSL1* in *Camelina sativa.* Genomic DNA sequence of gRNA-Cas9 target sites within the gene featuring red text indicating the PAM site.

References

Abramović, H., & Abram, V. (2005). Properties of *Camelina sativa* oil. *Food Technology and Biotechnology, 43*(1), 63–70.

Aftab, A., Amer, J., & Nayla, M. (2023). GMOs or non-GMOs? The CRISPR conundrum. *Frontiers in Plant Science, 14*, 1232938. https://doi.org/10.3389/fpls.2023.1232938

Alkotami, L., White, D. J., Schuler, K. M., & Durrett, T. P. (2024). Targeted engineering of camelina and pennycress seeds for ultrahigh accumulation of acetyl-TAG. *Proceedings of the National Academy of Sciences, 121*(47), e2412542121. https://doi.org/10.1073/pnas.2412542121

Alkotami, L., Kornacki, C., Campbell, S., McIntosh, G., Wilson, C., Tran, T. N. T., & Durrett, T. P. (2021). Expression of a high-activity diacylglycerol acetyltransferase results in enhanced synthesis of acetyl-TAG in camelina seed oil. *The Plant Journal, 106*(5), 1346–1357. https://doi.org/10.1111/tpj.15210

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature, 576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Aznar-Moreno, J. A., & Durrett, T. P. (2017). Simultaneous targeting of multiple gene homeologs to alter seed oil production in *Camelina sativa*. *Plant and Cell Physiology, 58*(7), 1260–1267. https://doi.org/10.1093/pcp/pcx058

Bates, P. D., Stymne, S., & Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Current Opinion in Plant Biology, 16*(3), 358–364. https://doi.org/10.1016/j.pbi.2013.02.015

Beilstein, M. A., Al-Shehbaz, I. A., Mathews, S., & Kellogg, E. A. (2008). Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: Tribes and trichomes revisited. *American Journal of Botany, 95*(10), 1307–1327. https://doi.org/10.3732/ajb.0800065

Berti, M., Gesch, R., Eynck, C., Anderson, J., & Cermak, S. (2016). Camelina uses, genetics, genomics, production, and management**.** *Industrial Crops and Products, 94*, 690–710. https://doi.org/10.1016/j.indcrop.2016.09.034

Berti, M. T., & Gesch, R. W. (2015). Cuphea production and management. In V. M. V. Cruz & D. A. Dierig (Eds.), *Industrial Crops and Products* (Vol. 9, pp. 313–336). Springer. https://doi.org/10.1007/978-1-4939-1447-0_13

Berti, M., Wilckens, R., Fischer, S., Solis, A., & Johnson, B. (2011). Seeding date influence on camelina seed yield, yield components, and oil content in Chile. *Industrial Crops and Products, 34*(2), 1358–1365. https://doi.org/10.1016/j.indcrop.2010.12.008

Bowers, C., Toews, M. D., & Schmidt, J. M. (2021). Winter cover crops shape early-season predator communities and trophic interactions. Ecosphere, 12(7). https://doi.org/10.1002/ecs2.3635

Brock, J. R., Bird, K. A., Platts, A. E., Gomez-Cano, F., Gupta, S. K., Palos, K., Railey, C. E., Teresi, S. J., Lee, Y. S., Magallanes-Lundback, M., Pawlowski, E. G., Nelson, A. D. L., Grotewold, E., & Edger, P. P. (2024). Exploring genetic diversity, population structure, and subgenome differences in the allopolyploid *Camelina sativa*: Implications for future breeding and research studies. *Horticulture Research, 11*(11). https://doi.org/10.1093/hr/uhae247

Brock, J. R., Mandáková, T., McKain, M., Lysak, M. A., & Olsen, K. M. (2022). Chloroplast phylogenomics in Camelina (Brassicaceae) reveals multiple origins of polyploid species and the maternal lineage of *C. sativa*. *Horticulture Research, 9***.** https://doi.org/10.1093/hr/uhab050

Brock, J. R., Ritchey, M. M., & Olsen, K. M. (2022). Molecular and archaeological evidence on the geographical origin of domestication for Camelina sativa. *American Journal of Botany, 109*(7), 1177–1190. https://doi.org/10.1002/ajb2.16027

Burnett, C. L., Fiume, M. M., Bergfeld, W. F., Belsito, D. V., Hill, R. A., Klaassen, C. D., Liebler, D., Marks, J. G., Shank, R. C., Slaga, T. J., Snyder, P. W., & Andersen, F. A. (2017). Safety assessment of plant-derived fatty acid oils. *International Journal of Toxicology, 36*(3_suppl), 51S–129S. https://doi.org/10.1177/1091581817740569

Carnes, K. (2004). Offroad hydraulic fluids: Beyond biodegradability. *Tribology & Lubrication Technology, 60*(9), 32–40. https://er.lib.k-state.edu/login?url=https://www.proquest.com/scholarly-journals/offroad-hydraulic-fluids-beyond-biodegradability/docview/226955778/se-2

Chapman, K. D., & Ohlrogge, J. B. (2012). Compartmentation of triacylglycerol accumulation in plants. *Journal of Biological Chemistry, 287*(4), 2288–2294. https://doi.org/10.1074/jbc.R111.290072

Chaudhary, R., Koh, C. S., Kagale, S., Tang, L., Wu, S. W., Lv, Z., Mason, A. S., Sharpe, A. G., Diederichsen, A., & Parkin, I. A. P. (2020). Assessing diversity in the Camelina genus provides insights into the genome structure of Camelina sativa. *G3: Genes|Genomes|Genetics, 10*(4), 1297–1308. https://doi.org/10.1534/g3.119.400957

Chen, M., Xuan, L., Wang, Z., Zhou, L., Li, Z., Du, X., Ali, E., Zhang, G., & Jiang, L. (2014). TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several

seed development regulators in Arabidopsis. *Plant Physiology, 165*(2), 905–916. https://doi.org/10.1104/pp.114.235507

Chopra, R., Johnson, E. B., Emenecker, R., Cahoon, E. B., Lyons, J., Kliebenstein, D. J., Greenham, K., Ort, D. R., & Sedbrook, J. C. (2020). Identification and stacking of crucial traits required for the domestication of pennycress. *Nature Food, 1*, 84–91. https://doi.org/10.1038/s43016-019-0007-z

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics, 25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., & Stymne, S. (2000). Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proceedings of the National Academy of Sciences of the United States of America, 97*(12), 6487–6492. https://doi.org/10.1073/pnas.120067297

Ding, L. N., Li, M., Guo, X. J., Tang, M. Q., Cao, J., Wang, Z., Liu, R., Zhu, K. M., Guo, L., Liu, S. Y., & Tan, X. L. (2020). Arabidopsis GDSL1 overexpression enhances rapeseed Sclerotinia sclerotiorum resistance and the functional identification of its homolog in Brassica napus. *Plant Biotechnology Journal, 18*(5), 1255–1270. https://doi.org/10.1111/pbi.13289

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. https://doi.org/10.1093/dnares/dsu045

Dougherty, W. G., & Parks, T. D. (1995). Transgene and gene suppression: Telling us something new? *Current Opinion in Cell Biology, 7*(3), 399–405. https://doi.org/10.1016/0955-0674(95)80096-4

Durrett, T. P., McClosky, D. D., Tumaney, A. W., Elzinga, D. A., Ohlrogge, J., & Pollard, M. (2010). A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in Euonymus and transgenic seeds. *Proceedings of the National Academy of Sciences of the United States of America, 107*(20), 9464–9469. https://doi.org/10.1073/pnas.0914427107

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE, 3*(11), e3647. https://doi.org/10.1371/journal.pone.0003647

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology, 31*(9), 822–826. https://doi.org/10.1038/nbt.2623

Gautam, B., Jarvis, B. A., Esfahanian, M., McGinn, M., Williams, D., Liu, S., Phippen, M. E., Heller, N. J., Wesley, T. L., Phippen, W. B., Ulmasov, T., Marks, M. D., Chopra, R., & Sedbrook, J. C. (2025). [Preprint]. *bioRxiv*. https://doi.org/10.1101/2025.03.15.643467

Gentsch, N., Peth, S., Wirth, S., & Poll, C. (2024). Cover crops improve soil structure and change organic carbon distribution in macroaggregate fractions. *Soil, 10*, 139–150. https://doi.org/10.5194/soil-10-139-2024

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA Websuite. *Nucleic Acids Research, 36*(suppl_2), W70–W74. https://doi.org/10.1093/nar/gkn188

Hwang, G., Lee, S.-H., Oh, M., Kim, S., Habib, O., Jang, H.-K., Kim, H. S., Kim, C. H., Kim, S., & Bae, S. (2024). [Preprint]. *bioRxiv*. https://doi.org/10.1101/2024.01.04.574288

Jach, G., Binot, E., Frings, S., Luxa, K., & Schell, J. (2001). Use of red fluorescent protein from *Discosoma* sp. (dsRED) as a reporter for plant gene expression. *The Plant Journal, 28*(4), 483–491. https://doi.org/10.1046/j.1365-313X.2001.01153.x

Jiang, W. Z., Henry, I. M., Lynagh, P. G., Comai, L., Cahoon, E. B., & Weeks, D. P. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, Camelina sativa, using CRISPR/Cas9 gene editing. *Plant Biotechnology Journal, 15*(6), 648–657. https://doi.org/10.1111/pbi.12663

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., Spillane, C., Robinson, S. J., Links, M. G., Clarke, C., Higgins, E. E., Huebert, T., Sharpe, A. G., & Parkin, I. A. P. (2014). The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nature Communications, 5*, 3706. https://doi.org/10.1038/ncomms4706

Kang, B. C., Bae, S. J., Lee, S., Lee, J. S., Kim, A., Lee, H., Baek, G., Seo, H., Kim, J., Kim, S. T., Song, J. T., Kim, J. S., & Kim, S. (2021). Chloroplast and mitochondrial DNA editing in plants. *Nature Plants, 7*(8), 899–905. https://doi.org/10.1038/s41477-021-00943-9

Katavic, V., Reed, D. W., Taylor, D. C., Giblin, E. M., Barton, D. L., Zou, J., MacKenzie, S. L., Covello, P. S., & Kunst, L. (1995). Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in *Arabidopsis thaliana* affecting diacylglycerol acyltransferase activity. *Plant Physiology, 108*(1), 399–409. https://doi.org/10.1104/pp.108.1.399

Kim, K., Gesch, R. W., Cermak, S. C., Phippen, W. B., Berti, M. T., Johnson, B. L., & Marek, L. (2011). Cuphea growth, yield, and oil characteristics as influenced by climate and soil

environments across the Upper Midwest USA. *Industrial Crops and Products, 33*(1), 99–107. https://doi.org/10.1016/j.indcrop.2010.09.002

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature, 533*(7603), 420–424. https://doi.org/10.1038/nature17946

Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., & Valen, E. (2016). CHOPCHOP v2: A web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Research, 44*(W1), W272–W276. https://doi.org/10.1093/nar/gkw398

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. (2019). CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research, 47*(W1), W171–W174. https://doi.org/10.1093/nar/gkz365

Lacroix, B., & Citovsky, V. (2020). Biolistic approach for transient gene expression studies in plants. *Methods in Molecular Biology, 2124*, 125–139. https://doi.org/10.1007/978-1-0716-0356-7_6

Lai, C. P., Huang, L. M., Chen, L. O., Chan, M. T., & Shaw, J. F. (2017). Genome-wide analysis of GDSL-type esterases/lipases in *Arabidopsis*. *Plant Molecular Biology, 95*(1–2), 181–197. https://doi.org/10.1007/s11103-017-0648-y

Lamichhane, J. R., & Alletto, L. (2022). Ecosystem services of cover crops: A research roadmap. *Trends in Plant Science, 27*(8), 758–768. https://doi.org/10.1016/j.tplants.2022.03.014

Le Cong, Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science, 339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Lee, K.-R., Kim, Y. M., Yeo, Y., Kim, S., & Suh, M. C. (2022). Camelina cytosol-localized diacylglycerol acyltransferase 3 contributes to the accumulation of seed storage oils. *Industrial Crops and Products, 189*, 115808. https://doi.org/10.1016/j.indcrop.2022.115808

Leibowitz, M. L., Papathanasiou, S., Doerfler, P. A., Blaine, L. J., Sun, L., Yao, Y., Zhang, C. Z., & Pellman, D. (2021). Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nature Genetics, 53*(6), 895–905. https://doi.org/10.1038/s41588-021-00838-7

Li, H., Li, J., Cong, X. H., Duan, Y. B., Li, L., Wei, P. C., Lu, X. Z., & Yang, J. B. (2013). A high-throughput, high-quality plant genomic DNA extraction protocol. *Genetics and Molecular Research, 12*(4), 4526–4539. https://doi.org/10.4238/2013.October.15.1

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J.-L., & Gao, C. (2018). Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 in vitro transcripts or

ribonucleoproteins. *Nature Protocols, 13*(3), 413–430. https://doi.org/10.1038/nprot.2017.145

Lu, C., & Kang, J. (2008). Generation of transgenic plants of a potential oilseed crop Camelina sativa by Agrobacterium-mediated transformation. *Plant Cell Reports, 27*(2), 273–278. https://doi.org/10.1007/s00299-007-0454-0

Lyzenga, W. J., Harrington, M., Bekkaoui, D., Wigness, M., Hegedus, D. D., & Rozwadowski, K. (2019). CRISPR/Cas9 editing of three CRUCIFERIN C homeologues alters the seed protein profile in *Camelina sativa*. *BMC Plant Biology, 19*, 292. https://doi.org/10.1186/s12870-019-1873-0

Mali, P., Esvelt, K. M., & Church, G. M. (2013). Cas9 as a versatile tool for engineering biology. *Nature Methods, 10*(10), 957–963. https://doi.org/10.1038/nmeth.2649

Mandáková, T., Pouch, M., Brock, J. R., Al-Shehbaz, I. A., & Lysak, M. A. (2019). Origin and evolution of diploid and allopolyploid Camelina genomes were accompanied by chromosome shattering. The Plant Cell, 31(11), 2596–2612. https://doi.org/10.1105/tpc.19.00366

Masella, P., Martinelli, T., & Galasso, I. (2014). Agronomic evaluation and phenotypic plasticity of Camelina sativa growing in Lombardia, Italy. *Crop and Pasture Science, 65*(5), 453–460. https://doi.org/10.1071/CP14025

McGinn, M., Phippen, W. B., Chopra, R., Bansal, S., Jarvis, B. A., Phippen, M. E., Dorn, K. M., Esfahanian, M., Nazarenus, T. J., Cahoon, E. B., Durrett, T. P., Marks, M. D., & Sedbrook, J. C. (2019). Molecular tools enabling pennycress (*Thlaspi arvense*) as a model plant and oilseed cash cover crop. *Plant Biotechnology Journal, 17*(5), 776–788. https://doi.org/10.1111/pbi.13014

McGuire, S. T., Shockey, J., & Bates, P. D. (2025). The first intron and promoter of Arabidopsis DIACYLGLYCEROL ACYLTRANSFERASE 1 exert synergistic effects on pollen and embryo lipid accumulation. *New Phytologist, 245*(1), 263–281. https://doi.org/10.1111/nph.20244

Mhaske, V., Beldjilali, K., Ohlrogge, J., & Pollard, M. (2005). Isolation and characterization of an *Arabidopsis thaliana* knockout line for phospholipid:diacylglycerol transacylase gene (At5g13640). *Plant Physiology and Biochemistry, 43*(4), 413–417. https://doi.org/10.1016/j.plaphy.2005.02.001

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research, 42*(W1), W401–W407. https://doi.org/10.1093/nar/gku410

Moser, B. R. (2012). Biodiesel from alternative oilseed feedstocks: Camelina and field pennycress. *Biofuels, 3*(2), 193–209. https://doi.org/10.4155/bfs.12.6

Moser, B. R., Shah, S. N., Winkler-Moser, J. K., Vaughn, S. F., & Evangelista, R. L. (2009). Composition and physical properties of cress (*Lepidium sativum* L.) and field pennycress (*Thlaspi arvense* L.) oils. *Industrial Crops and Products, 30*(2), 199–205. https://doi.org/10.1016/j.indcrop.2009.03.007

Müller, A. O., & Ischebeck, T. (2018). Characterization of the enzymatic activity and physiological function of the lipid droplet-associated triacylglycerol lipase AtOBL1. *New Phytologist, 217*(3), 1062–1076. https://doi.org/10.1111/nph.14902

Ott, M. A., Eberle, C. A., Thom, M. D., Archer, D. W., Forcella, F., Gesch, R. W., & Wyse, D. L. (2019). Economics and agronomics of relay-cropping pennycress and camelina with soybean in Minnesota. *Agronomy Journal, 111*(3), 1281–1292. https://doi.org/10.2134/agronj2018.04.0277

Ott, M., Chopra, R., Frels, K., Brusa, A., Gjesvold, E. S., Marks, M. D., & Anderson, J. A. Facultative winter accession of *Camelina sativa* (L. Crantz) with early maturity contributes to understanding of the role of *FLOWERING LOCUS C* in camelina flowering (2022). [Preprint]. *bioRxiv*. https://doi.org/10.1101/2022.05.30.494064

Ozseyhan, M. E., Kang, J., Mu, X., & Lu, C. (2018). Mutagenesis of the FAE1 genes significantly changes fatty acid composition in seeds of *Camelina sativa*. *Plant Physiology and Biochemistry, 123*, 1–7. https://doi.org/10.1016/j.plaphy.2017.11.021

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell, 152*(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Quinchia, L. A., Delgado, M. A., Valencia, C., Franco, J. M., & Gallegos, C. (2010). Viscosity modification of different vegetable oils with EVA copolymer for lubricant applications. *Industrial Crops and Products, 32*(3), 607–612. https://doi.org/10.1016/j.indcrop.2010.07.011

Quinn, D., Poffenbarger, H., & Lee, C. (2019). Cover crop benefits and challenges in Kentucky. *University of Kentucky Cooperative Extension Service.* https://publications.ca.uky.edu/sites/publications.ca.uky.edu/files/AGR240.pdf

Sbravati, A., Delfiacco, G., & Bingenheimer, D. (2014). A practical approach to designing more cost efficient transformers: A total cost of ownership perspective between mineral oil and Envirotemp™ FR3™ fluid filled transformers (Version 1.0, W1050). *Cargill, Incorporated – Envirotemp™ Dielectric Fluids.*

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J. L., & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology, 31*(8), 686–688. https://doi.org/10.1038/nbt.2650

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., & Skarnes, W. C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods, 11*(4), 399–402. https://doi.org/10.1038/nmeth.2857

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science, 351*(6268), 84–88. https://doi.org/10.1126/science.aad5227

Ståhl, U., Carlsson, A. S., Lenman, M., Dahlqvist, A., Huang, B., Banaś, W., Banaś, A., & Stymne, S. (2004). Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from *Arabidopsis*. *Plant Physiology, 135*(3), 1324–1335. https://doi.org/10.1104/pp.104.044354

Taheripour, F., Sajedinia, E., & Karami, O. (2022). Oilseed cover crops for sustainable aviation fuels production and reduction in greenhouse gas emissions through land use savings. *Frontiers in Energy Research, 9*. https://doi.org/10.3389/fenrg.2021.782274

Thomson, G., Dickinson, L., & Jacob, Y. (2024). Genomic consequences associated with Agrobacterium-mediated transformation of plants. *The Plant Journal, 117*(2), 342–363. https://doi.org/10.1111/tpj.16496

Toxicological profile for JP-5, JP-8, and JET A fuels. (2017). Agency for Toxic Substances and Disease Registry (US). https://www.ncbi.nlm.nih.gov/books/NBK592027/

Tripathi, M. K., & Mishra, A. S. (2007). Glucosinolates in animal nutrition: A review. *Animal Feed Science and Technology, 132*(1–2), 1–40. https://doi.org/10.1016/j.anifeedsci.2006.03.003

United States Department of Agriculture, National Agricultural Statistics Service. (2024). *2022 Census of Agriculture*. U.S. Government Printing Office. https://www.nass.usda.gov/Publications/AgCensus/2022/index.php

van Belzen, I. A. E. M., Schönhuth, A., Kemmeren, P., & Hehir-Kwa, J. Y. (2021). Structural variant detection in cancer genomes: Computational challenges and perspectives for precision oncology. *NPJ Precision Oncology, 5*(1), 15. https://doi.org/10.1038/s41698-021-00155-6

Vaughn, S. F., Palmquist, D. E., Duval, S. M., & Berhow, M. A. (2006). Herbicidal activity of glucosinolate-containing seedmeals. *Weed Science, 54*(4), 743–748. https://doi.org/10.1614/WS-06-007R.1

Vollmann, J., & Eynck, C. (2015). Camelina as a sustainable oilseed crop: Contributions of plant breeding and genetic engineering. *Biotechnology Journal, 10*(4), 525–535. https://doi.org/10.1002/biot.201400200

Vollmann, J., Moritz, T., Kargl, C., Baumgartner, S., & Wagentristl, H. (2007). Agronomic evaluation of camelina genotypes selected for seed quality characteristics. *Industrial Crops and Products, 26*(3), 270–277. https://doi.org/10.1016/j.indcrop.2007.03.017

Weyers, S. L., Thompson, N. M., Johnson, J. M., Douglas, C. L., & Liebig, M. A. (2021). Surface runoff and nutrient dynamics in cover crop–soybean systems in the Upper Midwest. *Journal of Environmental Quality, 50*(1), 158–171. https://doi.org/10.1002/jeq2.20158

Williams, D. (2020). Elucidating the role of the high aliphatic glucosinolate (Hag) genes in pennycress (*Thlaspi arvense* L.) glucosinolate production (Master's thesis, Illinois State University). Illinois State University Institutional Repository. https://doi.org/10.30707/ETD2020.1606247535.297019ab

Wu, L., Jiang, S., Shi, M., Song, J., Wang, S., Tang, J., Ding, Y., Liu, Y., Xu, W., Han, Y., & Zhou, J. (2024). Adenine base editors induce off-target structure variations in mouse embryos and primary human T cells. *Genome Biology, 25*, 291. https://doi.org/10.1186/s13059-024-03434-0

Wu, S., Zhu, H., Liu, J., Yang, Q., Shao, X., Bi, F., Hu, X., & Li, Y. (2020). Establishment of a PEG-mediated protoplast transformation system based on DNA and CRISPR/Cas9 ribonucleoprotein complexes for banana. *BMC Plant Biology, 20*, 425. https://doi.org/10.1186/s12870-020-02609-8

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., Wang, X. C., & Chen, Q. J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology, 14*, 327. https://doi.org/10.1186/s12870-014-0327-y

Xu, J. C., Anders, S., Francis, T., Zhang, M., Hoffman, T., Giblin, M. E., & Taylor, D. C. (2012). Triacylglycerol synthesis by PDAT1 in the absence of DGAT1 activity is dependent on re-acylation of LPC by LPCAT2. *BMC Plant Biology, 12*, 120. https://doi.org/10.1186/1471-2229-12-120

Yang, H., Wu, J. J., Tang, T., Liu, K. D., Dai, C., & Zhang, Y. (2017). CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in *Brassica napus*. *Scientific Reports, 7*, 7489. https://doi.org/10.1038/s41598-017-07871-9

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. https://doi.org/10.1105/tpc.109.071795

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J. L., & Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature Communications, 7*, 12617. https://doi.org/10.1038/ncomms12617

Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G., & Taylor, D. C. (1999). The *Arabidopsis thaliana* TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. *The Plant Journal, 19*(6), 645–653. https://doi.org/10.1046/j.1365-313x.1999.00555.x