Tailoring Camelina Seed Fatty Acid Composition via Fast Neutron Mutagenesis

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Food and Bioproduct Sciences
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

By

Xinjie Liu

2022

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying, publication, or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head

Department of Food and Bioproduct Sciences

University of Saskatchewan

Saskatoon, Saskatchewan, S7N 5A8

Canada

or

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan, S7N 5C9

Canada

ABSTRACT

Despite its minor status as an oilseed, *Camelina sativa* is a crop with the potential to reach greater levels of production in Canada due to its unique seed oil composition, relatively low input requirements, and availability of both spring and winter varieties. The limited genetic diversity of the species presents a challenge for camelina improvement through breeding. The aim of this study was to induce genetic variations in camelina's genome through Fast Neutron (FN) irradiation applied to seeds in a dose gradient ranging from 7.1 to 49.5 Gy. A range of mutations should be caused following FN irradiation including deletions, point mutations, and rearrangements due to DNA damage and subsequent repair. In germination assays of FN-irradiated seeds, FN irradiation had no impact on germination, but higher doses of irradiation resulted in plants with a slower maturity. A FN mutagenesis population was established from irradiated camelina seeds for phenotypic and genotypic characterization. Male sterility and toothed leaf margin of irradiated camelina were observed based on visual observations. Seed fatty acids were analyzed and lineages with altered fatty acid composition, such as high C18:3 and low C20:1, were identified. To study novel genotypes, genotyping by sequencing was applied. An approach was designed and implemented that utilized Illumina sequence technology for generating large amounts of genome sequence data, in conjunction with Oxford Nanopore Technologies (ONT) sequencing for validation. One mutant lineage FN49-38 was characterized whose total very long chain fatty acid composition was 19.6%, a relatively lower percentage than of wild type (23.9%). This lineage carried a genomic DNA deletion of approximately 1.2kb spanning the 5' UTR (untranslated region) and the front half of the *FAE1B*-ORF (open reading frame) detected through Illumina sequencing, which was validated by ONT sequencing. Furthermore, the accuracy and authenticity of the deletion were confirmed by Sanger sequencing of the cloned mutant gene amplified by PCR (polymerase chain reaction) using specific primers flanking the mutated region. The ONT data was used to review and correct the current annotation

of the *FAE1* gene on chromosomes 11 and 12. A backcrossing experiment was conducted and confirmed that the deletion was linked to the reduced very long chain fatty acid composition phenotype.

ACKNOWLEDGEMENTS

My sincere gratitude goes to my supervisor Dr. Mark Smith and my co-supervisor Dr. Xiao Qiu for giving me this exceptional opportunity to be enrolled in the master's program at the University of Saskatchewan. The way they taught me to think scientifically and positively, and work professionally and with courage, this I truly appreciated and valued. My thanks go to them for their advice regarding how I could improve my laboratory work ability as well as my academic writing ability. Additionally, without the support and encouragement they provided, it is highly unlikely that I would be able to do my presentation well and enjoy the time at the AOCS conference in 2021. As an additional thank you, I would like to thank Dr. Darren Korber and Dr. Yongfeng Ai for their time and their encouragement in assisting me. I would like to extend my recognition and gratitude to my colleagues, Mrs. Helen Lui and Mr. Daniel Hupka, who provided useful advice, which gave me the opportunity to successfully complete my research. I am also grateful for the funding provided through the Agriculture and Agri-Food Canada Camelina Genomic research program. And last but not least, I have to thank my parents for their support, because without them, I probably wouldn't be so carefree to enter the field of scientific research and pursue a master's degree.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Camelina sativa, as a re-emerging oil crop, is currently being studied extensively. Known as gold-of-pleasure and false flax, this annual oilseed plant is a member of the Brassicaceae family. It has been cultivated since at least one thousand years ago in Northern Europe, Russia, and the Near East (Knorzer, 1978). Camelina is regarded as a low-input crop (Putnam et al., 1993) due to its lower fertilizer requirements than other oilseed crops (Nguyen et al., 2013). This plant grows well in temperate climates and is tolerant of cold temperatures during germination (Hunter & Roth, 2010). Incredibly, it can be spread at low seeding rates on frozen soil, and it can survive the freeze-thaw cycles after emergence in late winter and early spring (Budin et al., 1995). Contrary to other crops such as soybeans, canola, and corn, camelina is particularly suited to production on marginal lands and semi-arid zones (Nguyen et al. 2013, Ibrahim and El Habbasha 2015). While soybeans do not perform well in some northern regions of North America (Budin et al., 1995), camelina can thrive in a wide range of soil conditions, such as the Canadian prairies, where both precipitation and soil fertility are limited and the growing season is short (Li & Mupondwa, 2016). In addition, camelina has a moderate level of resistance to insects and weeds (Putnam et al., 1993). Considering the low cost of production of camelina, and its ease of cultivation in an environment with limited nutrients and fertilizers, this crop is appealing to farmers.

The seeds of camelina contain about 40% oil (Canada Health, 2012). The fatty acid composition of camelina seed oil is predominantly composed of oleic, linoleic, and α-linolenic acids. It is reported that oleic acid accounts for 15%, linoleic acid contributes for 25%, and α-linolenic acid occupies up to 40% (Yuan & Li, 2020) of total fatty acid by weight. With its distinctive fatty acid composition, camelina has potential to be a robust oil crop that could be bred into different improved varieties according to its target market. Although the plant has a

relatively low level of genetic diversity, mutations are imperative to improve it. The purpose of this study is to create a large mutant population using Fast Neutron irradiation (FN), which causes large deletion in the DNA sequence. Since the loci of mutations caused by FN irradiation are not predictable, we attempted to identify mutations that affect the expression of genes related to the lipid biosynthesis pathway through screening. Screening procedures were divided into two categories in accordance with the design of the study. The first category of screening is the forward genetic approach, where fatty acid profiles were determined by gas chromatography (GC) to identify potential mutant lines, and then bioinformatics analysis of the sequencing data generated by the sequencing techniques was performed for validation. The second category of screening is the reverse genetic approach, which utilizes PCR in combination with a pooling strategy or genotyping by sequencing (GBS) for identifying novel genotypes and phenotyping these mutant lines afterwards. The forward genetic approach was carried out in this study, where an emphasis was placed on detecting large deletions, which were likely the result of FN mutation. The development of the long-read sequencing technique improved its efficiency and accuracy (Dumschoot et al., 2020). Therefore, GBS approaches including Oxford Nanopore Technologies (ONT) in conjunction with Illumina sequencing (short-read sequencing technology) was applied in this project with the purpose of exploring whole genome information and validating the assembly data derived from the short-read sequencing data. In essence, the proposed design can effectively verify whether the deletion found in the short-read sequencing data is indeed a deletion and not a false-positive detection. In order to prove our proposed methodology is practicable and effective, we applied the traditional method as described by Li et al. (2001), which utilizes PCR with the tactic of discrimination of amplification. By this method, deletions could be confirmed by PCR validation. Additionally, to confirm that the genetic information obtained from Illumina sequencing and ONT were accurate, the mutated regions were cloned and sent for Sanger sequencing for confirmation since the accuracy of Sanger

sequencing data is almost 100%. Briefly, the PCR results and Sanger sequencing results of the cloned target gene were sufficient to demonstrate that the deletion is present.

Backcrossing is a method for combining wild-type and mutant genotypes to create heterozygote. Since both parents' gametes are randomly incorporated into one fertilized egg to form a zygote, based on Mendel's law of segregation, if the wild type and mutant lines are homozygous, their progeny should be heterozygous. To determine genotype, allele specific PCR analysis would be used. If the backcrossing progenies have been proved to be mutant homozygotes, and they produce seeds containing predicted fatty acid composition detected via GC, the linkage between the target mutated locus and its associated fatty acid phenotype can be confirmed. This approach is intended to demonstrate that an altered fatty acid composition is linked to a specific mutation and not the result of another mutation that has not yet been identified.

Overall, this project aims to apply FN mutagenesis to establish a FN library and identify useful phenotypes and genotypes for improving camelina diversity. We investigated the composition of fatty acids in camelina seeds to discover a mutant line having a higher C18:3 and a reduced VLCFA (very long chain fatty acid) ratio that could be beneficial to human health.

2. Literature review

2.1 Camelina as a minor oilseed crop

C. sativa is an annual oilseed crop from the family Brassicaceae and is also known as gold-of-pleasure, German sesame, Siberian oilseed, or false flax (Figure 2.1). Camelina has been cultivated for more than one thousand years in Northern Europe, Russia, and the Near East (Knorzer, 1978), and it was believed that it was originally native to southeast European and southwest Asian regions (Radatz & Hondelmann, 1981, Larsson, 2013), though it was discovered for the first time in Scandinavia during the Early Bronze Age (1800 BCE (Before the Common Era)) (Karg, 2012). The seeds of camelina have been found in a storage vessel dating from 700 to 900 BCE in Eastern Turkey (Dönmez & Belli, 2007) and in Romania during the transition from the Copper Age to Bronze Age (Toncea, 2014), providing evidence that it was probably cultivated for oil. Afterward, it became more prevalent during the Late Bronze Age and pre-Roman Iron Age (100 CE (Common Era)–250 BCE) (Larsson, 2013). While its cultivation decreased in importance during the Medieval Age (Karg, 2012), it persisted sporadically thereafter.

Figure 2.1 Photographs of *Camelina sativa* growing in field conditions. A: research plots of *Camelina sativa*, Saskatoon, Canada, August 2022; B: *Camelina sativa* inflorescences.

Camelina is considered a low input crop (Putnam et al., 1993) as cultivation requires less fertilizer and chemical pest and disease control than other oilseed crops (Nguyen et al., 2013). Camelina grows well in a temperate climate and has a short growing season of 85 to 100 days. It can tolerate low temperatures during germination enabling early seeding and crop establishment (Hunter & Roth, 2010, Ehrensing & Guy, 2008). Camelina is not susceptible to damage by flea beetles, a major pest of canola (*Brassica napus*) and shows resistance to many other insect pests (Soroka et al., 2015). The crop also shows good resistance to many fungal diseases of Brassica crops, although it is susceptible to clubroot (*Plasmodiophora brassicae*), white rust (*Albugo candida*) and aster yellows (*Phytoplasma asteris*) (Seguin-Swartz et al., 2009). Compared to other crops like soybean, canola, and corn, camelina is well adapted to production on marginal land and semi-arid areas (Nguyen et al., 2013, Ibrahim & El Habbasha, 2015). Camelina can survive under many different climates and soil environments, like the Canadian Prairies where is limited rainfall and low soil fertility (Li & Mupondwa, 2016). In this regard, *C. sativa* is easy to grow in an environment constrained by limited nutrients and moisture and the low cost of production is attractive to farmers. Furthermore, like the oilseed crop species *Brassica napus* (canola), camelina has both winter and spring biotypes, however the greater cold tolerance of camelina allows winter types to be successfully grown in North America and Canada (Walia et al., 2021). Camelina is therefore unique in being the only winter oilseed crop that can be grown on the Canadian Prairies.

Camelina seeds are relatively small, with a 1000-seed weight of 0.6 to 1.6g but are rich in oil (30-48% dry weight (DW)) and protein (27-33% DW) with composition influenced by growing conditions and genotype (Vollmann et al., 2005, Gugel & Falk, 2006, Vollmann et al., 2007, Pecchia, 2014). The seed oil of plants is generally composed of triacylglycerol (TAG), a molecule containing three fatty acids esterified to a glycerol backbone and is used by plants to store energy (Athenstaedt & Daum, 2006). The composition of fatty acids in TAGs of seed oils influences its physicochemical characteristics as well as its application (Yuan & Li, 2020). The

fatty acid composition of camelina seed oil is shown in Table 2.1. Camelina oil contains more than 90% unsaturated fatty acids with a high amount of polyunsaturated fatty acids (PUFAs) (Zubr, 2003, Moser, 2010). The essential fatty acids α-linolenic acid (18:3) and linoleic acid (18:2) account for around 38% and 17% of total fatty acids respectively. The amount of erucic acid, which has been reported to be detrimental to human health (Bremer & Norum, 1982), is in the range of 2.3-2.9% (Moser, 2010), a level currently below the maximum amount (5%) permitted in edible oil, as per B.09.022 of the Canada Food and Drug Regulations. Although the high oil content of camelina's seeds makes it suitable for cultivation as an oil crop, the relatively high amount of PUFAs is not always desirable for non-food applications. On the food side, camelina contains a comparatively high amount of α-linolenic acid (C18:3), which is desirable. This fatty acid is beneficial to human health and is recommended for daily consumption to aid in lowering blood cholesterol levels (Karvonen et al., 2002). For non-food applications such as jet fuel or biofuel, saturated fatty acids and very-long-chain unsaturated fatty acids are not desired in methyl-ester-based biofuel. They have comparatively high melting points which could lead to solidification and precipitation under low-temperature condition which can lead problems such as the clogging of fuel filters in the cold winters of Canada. In addition, a high PUFA content is not always desired in methyl-ester-based biofuel either. In contrast to monounsaturated fatty acids (MUFAs), PUFAs reduce biofuel stability influencing the cetane number (CN), which is the crucial parameter referring to the ignition quality of biofuel (Qu et al., 2012). The presence of easily oxidized bis-allylic methylene group between double bonds significantly reduces oxidative stability. For example, linoleate, with one bis-allylic methylene group, is 40 times more reactive than oleate (Frankel, 2012). Therefore, a high ratio of MUFAs to PUFAs in biofuel is desired (Moser & Vaughn, 2010). Biodiesel can flow through an engine fuel system without clogging if the liquid contains a high amount of alkyl monoester holding a high kinematic viscosity (Moser & Vaughn, 2010), which means a higher ratio of MUFAs to PUFAs is necessary for biofuel to work properly in the vehicle's engine. In camelina oil MUFAs account for around 23-33% of the

total content of fatty acids compared to around 60% for PUFAs (Vollmann et al., 2007). Studies have shown that despite the high PUFA content, kinematic viscosities of B5 biofuel blend (containing 5% of camelina derived fatty acid methyl esters) is 2.38 mm2s-1, which is within the parameters set for use as a fuel (Campbell et al., 2013). Therefore, the high oil content and fatty acid composition of camelina seeds suggest that the crop can be a successful candidate for producing biofuel.

Table 2.1 The major fatty acids in camelina oil (adapted from Moser, 2010).

Fatty acid	Content %
C16:0 Palmitic	5.3-6.8
C18:0 Stearic	2.5-2.7
C18:1$^{\Delta 9}$ Oleic	12.6-18.6
C18:2$^{\Delta 9,12}$ Linoleic	14.3-19.6
C18:3$^{\Delta 9,12,15}$ alpha-linolenic	32.6-38.4
C20:0 Arachidic	1.2-1.5
C20:1$^{\Delta 11}$ Eicosenoic	12.4-16.8
C20:2 $^{\Delta 11,14}$ Eicosadienoic	1.3-1.9
C20:3 Eicosatrienoic	0.8-1.7
C22:0 Behenic	0.2-0.3
C22:1 $^{\Delta 13}$ Erucic	2.3-2.9

In general, oils with a high proportion of unsaturated fatty acids are more susceptible to oxidation. Camelina oil, however, contains a high level of natural antioxidants, like phenolics and tocopherols (vitamin E). These natural antioxidants greatly enhance oxidative stability compared to other types of oils like flaxseed oil which also contains a high level of unsaturated fatty acid (Budin et al., 1995, Abramovic & Abram, 2005).

The high PUFA content of camelina oil is often used to position it as a healthy oil for human consumption. With some exceptions PUFAs are generally classified as one of 2 types omega-3 (n-3) and omega-6 (n-6), where the first C=C double bond is located either on the third or sixth carbon atom counting from the methyl-end of the molecule, respectively (Calder, 2006). A large

number of studies have reported the health effects of these two types of fatty acids, of which omega-3 is anti-inflammatory and omega-6 is pro-inflammatory (Zhuang et al., 2018). This distinction is based on their pathway of modification in the body and conversion to their eicosanoid derivatives such as leukotrienes and prostaglandins (Innes & Calder, 2018, Calder, 2020). In our daily diet, based on the study of Simopoulos (2008), people are recommended to balance the consumption of omega-6 and omega-3, in a ratio of approximately 1/1. Excessive consumption of omega-6 will have adverse health impacts, such as obesity, inflammation, cardiovascular disease or even cancer (Vazquez et al., 2019). Unfortunately, the amount of omega-6 in the modern western diet is considered out of balance, with omega-6 to omega-3 ratios closer to 16:1 (Simopoulos, 2006). In these circumstances, it is advisable to increase the proportion of omega-3 in the daily diet in order to maintain a healthy ratio between omega-6 and omega-3. There are many kinds of omega-6 and omega-3. Among them, the C18 fatty acids linoleic acid and α-linolenic acid are the most important and essential fatty acids in a typical diet, and act as substrates for synthesis of longer chain derivatives (Mišurcová et al., 2011). It is important to note that linoleic acid and α-linolenic acid are indispensable since they are precursors to arachidonic acid (AA), eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). The activities of AA in the body are related to flexibility, fluidity, and selective permeability of membranes (Pompéia et al., 2000, Brash, 2001). DHA and EPA appear to be related to the prevention of cardiovascular disease and are major components of brain lipids (Dawczynski et al., 2010). Hence, the health benefits that can be derived from linoleic acid and α-linolenic acid cannot be ignored. These two fatty acids cannot be biosynthesized in the human body and are therefore called essential fatty acids for human beings (Kaur et al., 2012).

Normally, the major fatty acids in cooking vegetable oils are eighteen carbon chain, with between 0 and 3 double bonds (Orsavova et al., 2015). Accordingly, cooking vegetable oils could be a good source of the essential fatty acids. There is a lot of edible oil in the marketplace that claim to be PUFA rich, like canola oil. Canola oil, however, contains a higher amount of

linoleic acid (19%) than α-linolenic acid (9%) (Canola Council, 2022), which makes a poor source of omega-3 fatty acid. Camelina oil has a α-linolenic acid content of approximately 38% and a linoleic acid content of around 17% (Health Canada, 2012). As α-linolenic acid is more than twice the amount of linoleic acid camelina seed oil has potential to be developed as an alternative omega-3 sources for the daily diet.

In addition to C18 fatty acids, camelina oil contains VLCFAs like gondoic acid (C20:1) and erucic acid (C22:1) at approximately 16.1% and 3.4% of the total camelina oil fatty acid, respectively (Government, 2012, Berti et al., 2016). Although they are omega-9 fatty acids, they have distinct effects on health compared to oleic acid (C18:1). Erucic acid has been shown to be detrimental in animal studies and is associated with triacylglycerol accumulation in the heart causing negative effects on health (Vetter et al., 2020), so the maximum amount of this fatty acid in vegetable oil must be lower than 5%, as per B.09.022 of the Canada Food and Drug Regulations. Gondoic acid is a rare fatty acid in vegetable oils (Ewing et al., 2019) and there are no studies addressing the effect of this fatty acid on human health.

Camelina seeds are rich in protein and contain other components. After oil extraction, the residuum, which is also known as camelina meal, contains about 40-45% protein (Boyle et al., 2018, Li et al., 2014) together with fiber, such as mucilage (6.7%), lignin (7.4%), and crude fiber (15%) (Ibrahim & El Habbasha, 2015). Based on transcriptome analysis of developing camelina seeds, cruciferin (11 isoforms were identified) and napin constitute the main component of their seed storage proteins (SSPs). The remaining meal protein is composed of vicilin, oil body proteins (OBPs), and LEA (late embryogenesis abundant) proteins (Perera et al., 2022). There are many similarities between the SSPs present in camelina meal and those found in *Brassica napus* (canola/rapeseed). After isolation, glutelin occupied the major proportion in camelina meal, which was 64.64%, followed by globulin and albumin, which were 17.67% and 10.54%, respectively (Li et al., 2014). With further analysis, the amino acid composition of camelina meal proteins has been determined. According to Perera et al. (2022), camelina meal contains all the

essential amino acids (EAA), with 408 mg of EEA per gram of protein exceeding the standard requirement for adults recommended by World Health Organization. Camelina protein contains a high percentage of S-containing amino acids, is soluble at acidic pH, and is structurally stable at low pH and high temperature, which is similar to other economically relevant crucifers (Perera et al., 2022).

Besides being used as a food source for humans, camelina is also used as a source of food for animals. Using feed rations in which fish oil was fully substituted with camelina oil or fish meal partially substituted with camelina meal (10%) for farmed Atlantic salmon (*Salmo salar*), a study was conducted to determine whether there was an impact on tissue lipids and sensory quality. Salmon consuming diets containing camelina displayed a significantly higher total lipid content (22% w/w) than salmon consuming diets containing fish oil (14% w/w). Diets containing 100% fish oil and 100% camelina oil had no effect on the sensory characteristics of salmon fillets. Similarly, Hixson and co-workers (2016) fed Atlantic salmon and rainbow trout with different amounts of camelina meal (0–240 g/kg for salmon, 0–210 g/kg for trout). As a result, the level of C18:3 was elevated dramatically in the muscle tissue of fish fed with a higher level of camelina meal. Additionally, EPA levels were significantly higher in salmon fed the 24% camelina meal diet compared to salmon fed the 0% and 8% camelina meal diets. As a result, the level of C18:3 in trout fed without the camelina meal diet was significantly lower than the level of C18:3 in trout fed with the camelina meal diet. By comparison, trout growth was not affected by different levels of camelina meal. As well, Hixson et al. (2016) fed Atlantic cod (*Gadus morhua*) a range of levels of camelina meal (0% to 40%). The Atlantic cod fed with increased camelina meal contained decreased amounts of 16:0, 16:1ω7, and 22:1ω9 in the muscle tissue, whereas levels of 18:3ω3, 20:1ω9, and PUFA increased. It is also possible to add camelina oil or meal to the diet of broiler chickens in addition to adding it to fish diets. In comparison to the control diet, the camelina oil diet reduced the fraction of abdominal fat in the carcass, while the

breast meat's total protein content and n-3 PUFA, especially α-linolenic acid, increased (Pietras & Orczewska-Dudek, 2013).

Entering the 21st century, the continued acceleration of urbanization and the rapid growth of the global population have exacerbated the gap between the demand for agricultural land and the demand for urban land. The influence caused by the increasing urban expansion in the world is profound, which has been shown to lead to the loss of agricultural lands (D'Amour et al., 2017) because over one fifth of the expanded urban land was formerly agricultural land (Güneralp et al., 2020). To alleviate this problem, scientists from all over the world have conducted research on many cash crops, such as oil crops, to improve their production efficiency and achieve higher production on the remaining agricultural land base. Over the past three decades, soybeans have been studied extensively. With the emergence and rapid development of transgenic technology, sophisticated genetically modified soybeans that possessed high yield and healthy characteristics have been sold in many countries. The mere improvement of soybeans, however, has not matched the contemporary requirement of food. Accordingly, many other oil crops, like canola, rapeseed, peanuts and camelina, have been studied by scientists to increase the world's vegetable oil production. In comparison to soybean seeds, which are the predominant seeds for oil production worldwide (Miklaszewska et al., 2021), *C. sativa* seeds contain significantly greater amounts of oil (Clemente & Cahoon, 2009). In addition to its agronomic advantages, camelina possesses potential to be developed as a new source of oil in the near future due to its agronomic attributes.

2.2 Camelina genetics and genomic resources

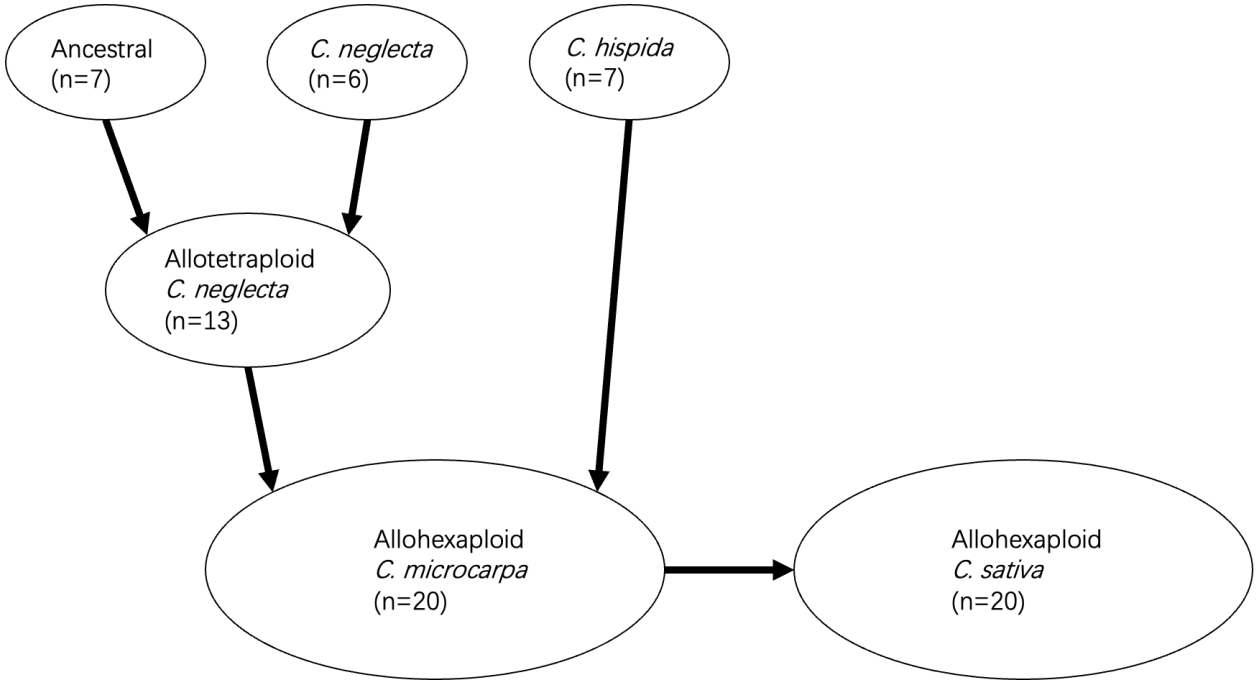

Figure 2.2 A simplified pathway of genome evolution of *C. sativa*. (Adapted from Mandáková et al., 2019)

C. sativa has a hexaploid genome with 20 chromosomes in a haploid set and an estimated genome size of 750 to 785Mb (Hutcheon et al., 2010, Kagale et al., 2014, Chaudhary et al., 2020). The reason for the relatively large genome size compared to other Brassicaceae species is that *C. sativa* is thought to have evolved from three species with lower chromosome numbers, one n=6 and two n=7 (Kagale et al., 2014). The origin of these subgenomes has been investigated by a number of research groups and is complicated by the diverse ploidy of species within the genus *Camelina* (Brock et al., 2018, Mandakova et al., 2019, Chaudhary et al., 2020). Current hypotheses are that the *C. sativa* subgenome 3 is likely derived from the diploid *C. hispida* (2n=14) with subgenomes 1 and 2 from a tetraploid *C. neglecta* (2n = 26) (Mandakova et al., 2019) or tetraploid *C. microcarpa* (2n= 26) (Chaudhary et al., 2020). Camelina has a sequenced genome (Kagale et al., 2014) and about 82% of the estimated genome is assembled,

which involves 641.45 Mb of sequence data. There are 89,418 non-redundant genes predicted (Kagale et al., 2014). Based on comparison to lipid biosynthetic pathways in arabidopsis (Li-Beisson et al., 2013), over 80% of the 736 non-redundant genes associated with the management of diverse procedures in acyl-lipid biosynthesis, accumulation, and degradation are retained as three copies in *C. sativa*. For instance, both *fatty acid desaturase 2* (*FAD2*) and *fatty acid elongase 1* (*FAE1*), which are single copy genes in diploid *Arabidopsis thaliana*, are present as three copies in *C. sativa*, and all copies of these two genes are expressed during the seed development period (Hutcheon et al., 2010). Kagale et al. (2014) suggested that this entire-genome triplication event may have resulted in more complicated regulatory mechanisms in acyl-lipid metabolism's regulatory mechanisms in this species. Because most of the genes involved in acyl-lipid metabolism have three copies whose function is likely identical, camelina has a high degree of gene redundancy.

In addition to a sequenced genome, many other genetic resources exist for *C. sativa*. This includes extensive transcriptome data sets. For example, using 454 pyrosequencing short-read sequencing and Sanger sequencing, Nguyen et al. (2013) sequenced camelina seeds at 15-20 days after pollination, which resulted in over 60,000 transcripts, including assembled contigs and singletons, being identified. Furthermore, camelina's developmental transcriptome has also been determined. To establish and sequence RNAseq libraries from 12 different tissue samples, a variety of developmental stages were obtained: early growth (germinating seeds and cotyledons), vegetative growth (young leaves, roots, stems, and senescing leaves), flower development (buds and flowers), and seed development (early seed, early mid seed, late mid seed, and late seed) (Kagale et al., 2016). According to their findings, approximately 88% of total annotated gene models were functional in at least one of the 12 tissue types, but around 12% were not expressed in any of these tissues or were detected below background levels. This research provides a detailed representation of the genes expressed during the developmental stage of seed/embryos,

and due to the ease of access (www.bar.utoronto.ca), it can be used for future studies intended to improve the quality of camelina seeds.

In plants and animals, microRNAs (miRNAs) are responsible for regulating gene expression and are utilized for gene expression modification (Pritchard et al., 2012). These miRNAs are typically 20-24 nucleotide non-coding RNAs, and they can bind to either the untranslated regions (UTR) or the coding sequences (CDS) of complementary mRNAs, thus inhibiting their translation or degrading (Bartel, 2009). In 2015, Poudel et al. (2015) attempted to identify miRNAs in the camelina genome. According to their study, 34,346,751 reads ranging from 15 to 27 nucleotides in length were obtained using Illumina technology, with most reads measuring 21 and 24 nucleotides in length. The study identified 133 miRNA targets that are associated with lipid metabolic pathways. Additionally, the Csa-miR166a, the most prevalent miRNA, targets genes encoding an acyl-CoA oxidase which is implicated in the first step of β-oxidation of fatty acids. They concluded that miRNAs are imperative for the development of seeds in order to minimize the rate of lipid breakdown (Poudel et al., 2015).

Camelina can be transformed by Agrobacterium-mediated infiltration under vacuum (Lu & Kang, 2008), which has been widely applied to arabidopsis research. It is possible to produce transgenic camelina lines in six to eight weeks using this method as opposed to soybean transformation, which typically takes six to ten months due to genes delivered to cells via biolistic or Agrobacterium-based approaches followed by tissue culture for a substantial period (Wang et al., 2012). Furthermore, an Agrobacterium-mediated floral dip without vacuum was successfully conducted on *C. sativa* (Liu et al., 2012). According to their research, the transgenic lines were derived in 9 days by dipping all buds and flowers in the Agrobacterium solution, covering them with plastic film, and then placing them in the dark for 24 hours. A second dip and covering process was conducted 1 week later to create transgenic lines. In this respect, camelina surpasses other oil crops in terms of generation time and ease of transformation (Nguyen et al., 2013). Concurring with ease of transformation and well-established genome

information, camelina is well positioned for improvement using modern biotechnology techniques.

2.3 Pathways of oil biosynthesis in camelina

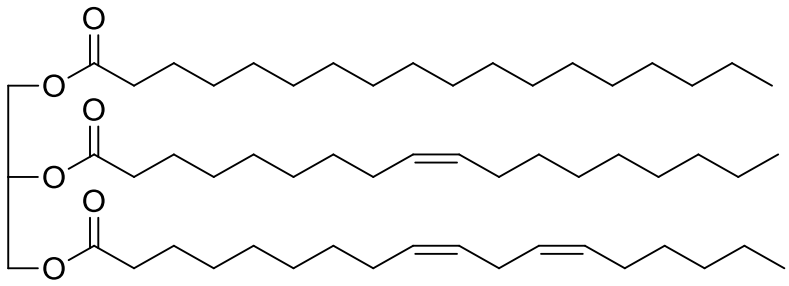

Figure 2.3 The general structure of triacylglycerol (TAG), showing three fatty acids joined by ester bonds to a glycerol "backbone".

TAGs present in oilseeds (Figure 2.3) are natural organic molecules with a high energy density and serve as an essential component of plant physiology (Athenstaedt & Daum, 2006). TAG is composed of three fatty acids esterified to a glycerol backbone. It is synthesized during seed development and reaches its maximum level when the seed is mature (Lung & Weselake, 2006). Fatty acids are the fundamental component of triacylglycerol and are synthesized *de novo* from acetyl-coenzyme A (CoA) in plastids catalyzed by acetyl-CoA carboxylase (ACC) and fatty acid synthase (FAS) (He et al., 2020). Fatty acids are then exported from the plastid for desaturation and elongation and subsequent assembly into TAG in the endoplasmic reticulum (ER) membrane. In seeds, the synthesized TAG molecules are stored in structures within the cell called oleosomes or oil bodies, which have a matrix of TAGs surrounded by a monolayer of phospholipids and proteins such as oleosins (Miquel et al., 2017, Shimada et al., 2018). The fatty acid composition of seed TAGs largely determines the properties and potential used of the oil.

In plants, *de novo* fatty acid biosynthesis occurs in plastids (Figure 2.4) catalyzed by the activity of type II fatty acid synthase, which comprises several enzymes and enzyme complexes such as the acetyl-CoA carboxylase (ACCase) and fatty acid synthase complex (FAS) (Troncoso-Ponce et al., 2016). Almost all seeds transport carbon to fatty acid synthesis through glycolysis, with hexose and/or trisaccharide serving as the principal carbohydrates entering the plastid (Bates et al., 2013). Acetyl-CoA and bicarbonate as the substrates for the formation of malonyl-CoA is the confirmed first step of fatty acid biosynthesis catalyzed by ACCase. The malonyl group is then transferred from CoA and binds to the acyl carrier protein (ACP) to form malonyl-ACP through the activity of malonyl-CoA: acyl carrier protein S-malonyltransferase (MCAMT). The FAS complex, which is composed of three different KAS, is responsible for elongating the carbon chains. KAS isoform III catalyzes the first condensation reaction between acetyl-CoA and malonyl-ACP, resulting in the formation of a four-carbon product. The KAS isoform I catalyzes condensations from four to sixteen carbon chains. After a series of condensation reactions, 16:0-ACP is produced from malonyl-ACP (Miklaszewska et al., 2021). This 16:0-ACP can be desaturated to 16:1-ACP, but notably elongated to 18:0-ACP. Another enzyme, KAS isoform II, catalyzes this elongation of C16:0-ACP to C18:0-ACP. The remaining FAS reactions undergo a reduction of 3-ketyoacyl-ACP intermediates to saturated acyl-ACP before a subsequent cycle of fatty acid synthesis begins (Ohlrogge & Jaworshi, 1997), which is catalyzed by 3-ketyoacyl-ACP reductase (KAR), 3-hydroxyacyl-ACP dehydrase (HAD), and enoyl-ACP reductase (ENR) sequentially. FAS complex releases only C16 and C18 fatty acids, which can be used as substrates for the elongation reactions that lead to VLCFA (Baud, 2018).

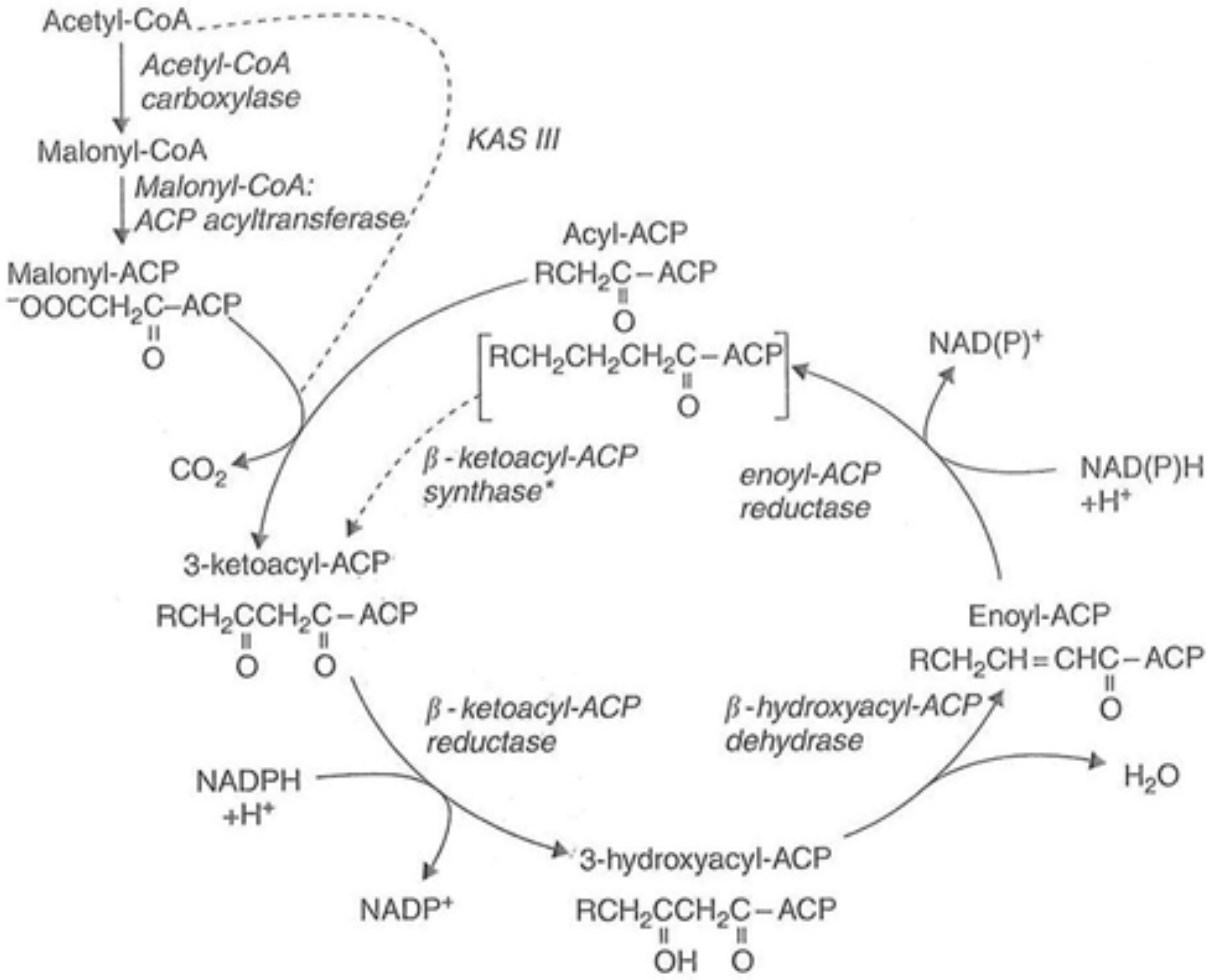

Figure 2.4 The scheme of *de novo* fatty acid biosynthesis in plastids. Adapted from Plant Fatty Acid Synthesis. (n.d.). Retrieved September 27, 2022, from https://lipidlibrary.aocs.org/chemistry/physics/plant-lipid/plant-fatty-acid-synthesis

After *de novo* fatty acid synthesis, the newly formed fatty acids (FFAs) are exported from the plastid to the ER for TAG synthesis. Acyl-ACP thioesterases (FAT) located at the inner plastid envelope membrane (IE) are responsible for the hydrolysis from acyl-ACP to FFAs, thus terminating *de-novo* fatty acid biosynthesis. There are two types of FAT, type A and type B, and type A of FAT (FATA) mainly uses 18:1-ACP as a substrate and type B of FAT (FATB) mainly uses 16:0-ACP and 18:0-ACP as the substrates (Salas & Ohlrogge, 2002, Bonaventure et al., 2003). Hence, malfunction of FATB proteins is related to the lower saturated fatty acids

production in oil seeds of camelina (Ozyseyhan et al., 2018). Furthermore, fatty acid export 1 (FAX1) protein located in the IE of the plastid plays an important role as a FFAs transporter in arabidopsis leaves and flowers (Li et al., 2015). After crossing the plastid inner and outer envelopes, the free fatty acids are converted to acyl-CoA via the activity of a long-chain acyl-CoA synthase (LACS) (Miklaszewska et al., 2021). Afterward, acyl-CoAs are thought to be transferred to the cytosolic face of the ER where phosphatidylcholine (PC) synthesis occurs (Botella et al., 2016). Additionally, ATP-binding cassette transporter subfamily A (ABCA) proteins are believed to act as a fatty acid/acyl-CoA transporters, and may regulate the transport of fatty acids into the ER (Kim et al., 2013). Cai et al. (2021) stated that the overexpression of arabidopsis *ABCA9* can increase the oil production in camelina seeds.

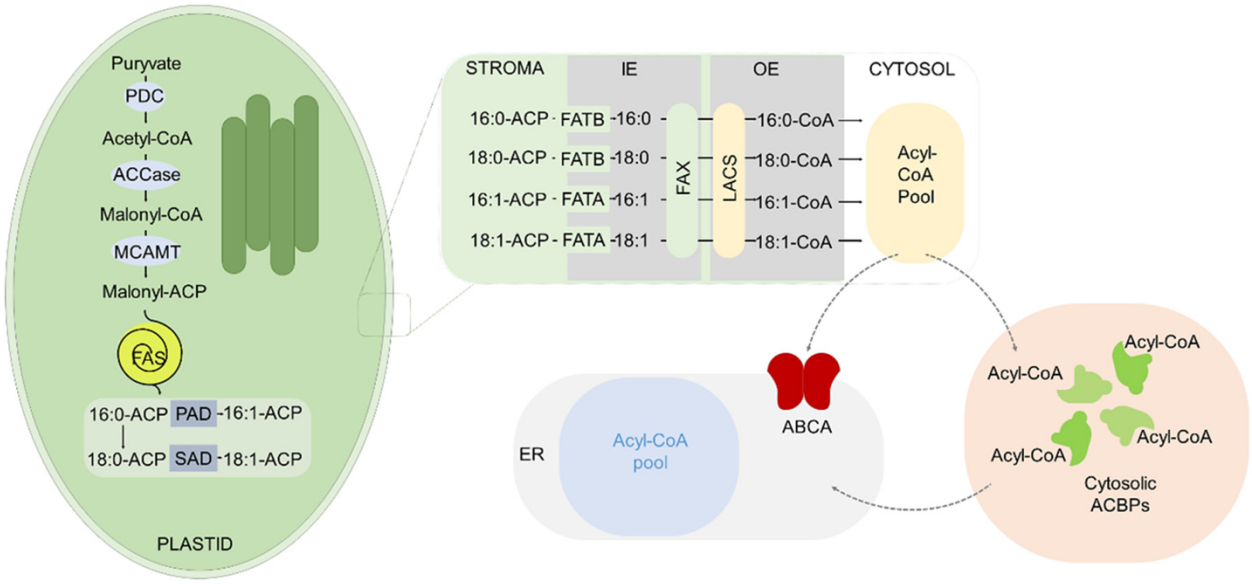

Figure 2.5 General overview of fatty acid biosynthesis and export from the plastid in oilseed cells. Fatty acid synthesis takes place in plastids where malonyl-ACP is used as a substrate. FAs are synthesized by the FAS complex in a series of condensation and elongation reactions which lead to formation of FA conjugates with ACP proteins. Their channeling to cytosol includes hydrolysis by thioesterases (FAT) to free fatty acids at the inner plastid envelope membrane (IE), followed by their transfer through the membrane mediated by FAX proteins. Once they reach the outer plastid envelope membrane (OE) they undergo vectorial acylation catalyzed by LACS enzymes. The resulting acyl-CoAs are transferred to the cytosol. The cytosolic pool of acyl-CoAs directly interacts with ACBP proteins, which serve as major transporters of acyl-CoAs from the cytosol to the ER. In the ER, acyl-CoAs are used as acyl donors for TAGs synthesis pathways. Abbreviations: ABCA: ATP binding cassette (ABC) transporter subfamily A; ACBPs: acyl-CoA binding proteins; ACCase: acetyl-CoA carboxylase; ACP: acyl carrier protein; CoA: coenzyme A; FAs: fatty acids; FAS: fatty acid synthase complex; FATA/B: fatty acyl-ACP thioesterase; ER: endoplasmic reticulum; FAX: fatty acid export protein; IE: inner plastid envelope; LACS: long chain acyl-CoA synthases; MCAMT: malonyl-CoA: acyl carrier protein S-malonyltransferase; OE: outer plastid envelope; PAD: palmitoyl-ACP desaturase; PDC: pyruvate dehydrogenase complex; SAD: stearoyl-ACP desaturase. (Adapted from Miklaszewska et al., 2021)

In seeds, TAG synthesis takes place in the cytosol catalysed by enzymes located in the endoplasmic reticulum membrane. Studies conducted with arabidopsis and other oilseeds have uncovered two primary routes for TAG assembly (Figure 2.6): the acyl-CoA-dependent pathway (Kennedy pathway), and the acyl-CoA-independent pathway (Bates et al., 2013, Miklaszewska et al., 2021). The Kennedy pathway (Gibellini & Smith, 2010) is relatively straightforward, and involves four significant reactions: the first reaction is the formation of lysophosphatidic acid

(LPA) from glycerol-3-phosphate (G3P) by transfer of a fatty acid from acyl-CoA to the *sn*-1 position, catalyzed by glycerol-3-phosphate acyltransferase (GPAT); secondly, lysophosphatidic acid acyltransferase (LPAAT) catalyze the formation of phosphatidic acid (PA) from LPA by incorporating an acyl-CoA on the *sn*-2 position of LPA; third, the phosphate group is then removed from PA catalyzed by phosphatidic acid phosphatase (PAP) to form de novo diacylglycerol (DAG); the last reaction uses DAG as a substrate and a fatty acid from acyl-CoA is esterified to the *sn*-3 position catalyzed by diacylglycerol acyltransferase (DGAT) (Miklaszewska et al., 2021). The other pathway of TAG biosynthesis is an acyl-CoA independent pathway comprised of three sub pathways, acyl editing, producing TAG directly with PC and DAG, and synthesizing TAG from a PC-derived DAG substrate (Bates et al., 2012). The first sub pathway, acyl editing, is a cycle exchanging fatty acid from the acyl-CoA pool in or out of PC. This includes a cycle of PC-deacylation and lysophosphatidylcholine (LPC)-reacylation. Lysophosphatidylcholine acyltransferase (LPCAT) is believed to catalyze both the deacylation and reacylation. LPCAT possesses forward actions and reverse actions, but it is more likely to catalyze the forward activity of the cycle (Bates et al., 2012). For the deacylation, in which free fatty acids are released from PC, the activity of phospholipase A2 (PLA2) is involved (Miklaszewska et al., 2021). The second sub pathway produces TAG directly from PC and DAG, in which the enzyme phospholipid: diacylglycerol acyltransferase (PDAT) plays a critical role. In this sub pathway, an acyl chain is transferred from the *sn*-2 position of PC to the *sn*-3 position of DAG to synthesize TAG and lyso-PC (Dahlqvist et al., 2000). In addition to this reaction, there is another reaction contained in the Kennedy pathway responsible for the attachment of acyl chains onto the *sn*-3 position, and DGAT catalyzes this reaction. The last sub pathway is synthesizing TAG from a PC-derived DAG substrate, which is accomplished accompanied by three types of enzymes corresponding to three alternative pathways. They are CDP-choline: diacylglycerol cholinephosphotransferase (CPT), and phospholipase C (PLC), and phosphatidylcholine: 1,2-*sn*-diacylglycerol choline phosphotransferase (PDCT) (Bates et al.,

2013). It is through the activity of CPT that newly synthesized DAGs are incorporated into PC, and then through the action of PLC that PC-derived DAGs are produced (Bates, 2016). The phosphocholine headgroup is transferred between DAG and PC through the activity of PDCT (Lu et al., 2009).

The composition of fatty acids in an oil is primarily determined by acyltransferases involved in the Kennedy pathway. A wide range of moieties, such as saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids, are utilized by ER-bound GPAT (Christie et al., 1991; Brokerhoff & Yurkowski, 1966; Lisa & Holcapek, 2008), indicating GPAT may not possess a strict substrate preference in vivo. The availability of acyl-CoA may be more critical in determining the acyl composition at the position of *sn*-1 of TAG (Voelker & Kinney, 2001). Of the three enzymes involved in the Kennedy pathway, microsomal LPAAT has the highest substrate selectivity (Laurent & Huang, 1992). Polyunsaturated fatty acids are more preferred by this enzyme (Snyder et al., 2009). Furthermore, an experiment conducted by Sørensen et al. (2005) demonstrated that in flax, LPAAT prefers 18:2$^{cis\Delta 9,12}$-CoA > 18:1$^{cis\Delta 9}$-CoA > 18:3$^{cis\Delta 9,12,15}$-CoA. VLCFA-CoA may, however, not be preferred by LPAAT, for example, this enzyme does not normally use erucoyl-CoA (22:1$^{cis\Delta 13}$) as substrate in *B. napus* (Bernerth & Frentzen, 1990; Taylor et al., 1991). The substrate specificity of DGAT is lower than that of the other two acyltransferases (GPAT and LPAAT) involved in the Kennedy pathway (Perry & Harwood, 1993). Accordingly, the substrate preferences of DGAT have been extensively studied (Chen et al., 2022), as well as its potential to increase seed oil content through overexpression (Snyder et al., 2009).

Synthesized TAG molecules are encapsulated in lipid droplets (LDs), which serve as a neutral lipid core surrounded by a phospholipid monolayer containing structural proteins, such as oleosins, caleosins and steroleosins (Zienkiewicz & Zienkiewicz, 2020). Oleosins, which comprised cytosolic-facing N- and C- termini and a huge hydrophobic domain, are responsible for the mediation of LDs stability (Siloto et al., 2006). There are 16 oleosin-encoding genes

identified in the arabidopsis genome, in which 5 of them are expressed in seeds only (Kim et al., 2002). The caleosin proteins contain a hydrophilic domain at the N-terminus which contains a calcium binding motif, a hydrophobic region at the center containing a proline residue which anchors the protein to the LD membrane, and a hydrophilic region at the C-terminus which harbours several phosphorylation sites (Hanano et al., 2006, Purkrtova et al., 2007). There are 8 genes in the arabidopsis genome related to encode this protein (Næsted et al., 2000). Steroleosins contains a sterol-binding dehydrogenase/reductase domain located at its C-terminus, and its hydrophobic N-terminus, which contains a proline knob motif, anchors it to the LDs membrane (Lin & Tzen, 2004, d'Andréa et al., 2007). After the synthesis of LD, storage lipids accumulate between the two leaflets of the ER bilayer and generate lenses-like structures (Miklaszewska et al., 2021). In addition to co-translationally inserting oleosins into ER cells, nascent small LDs fuse together and become mature LDs that separate from the ER (Chapman & Ohlrogge, 2012, Huang, 2018).

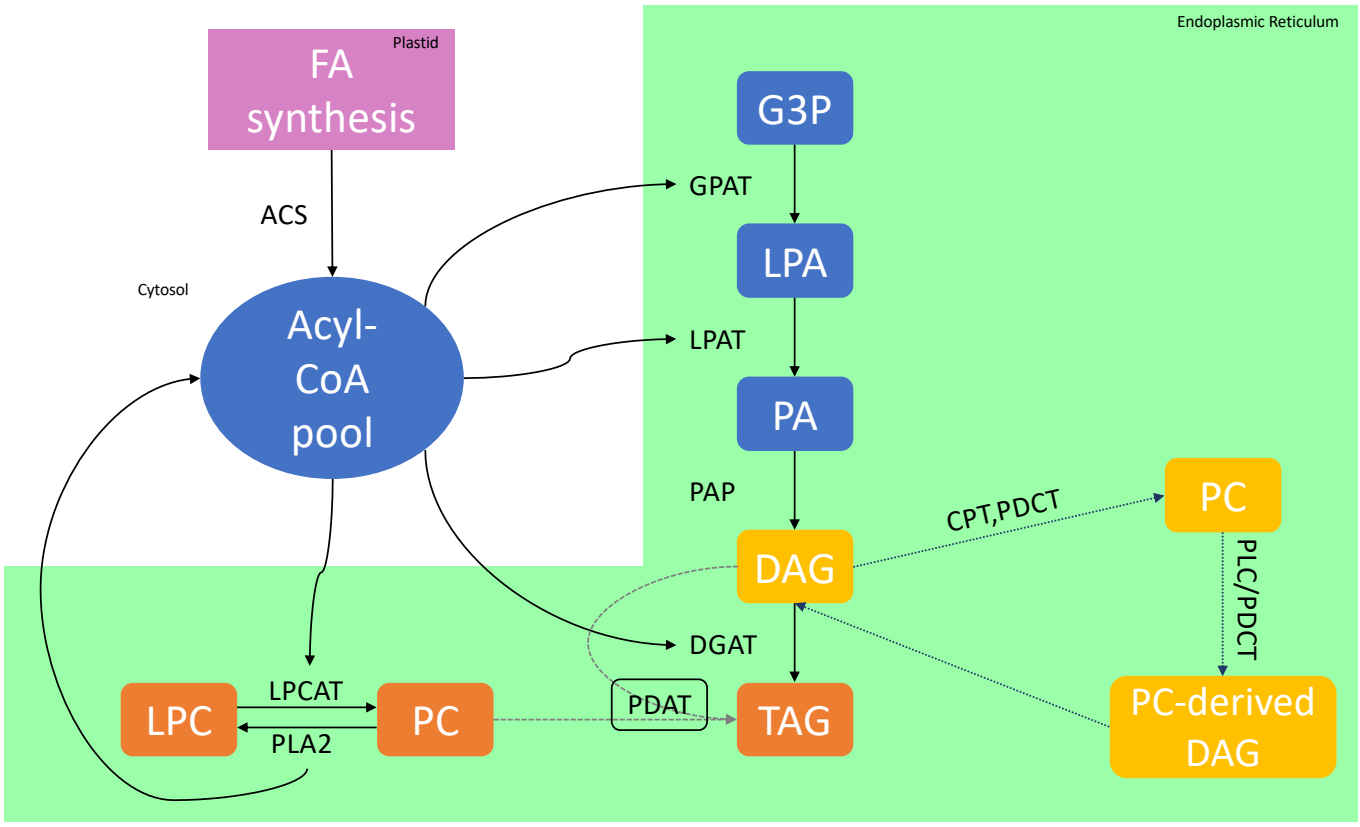

Figure 2.6 Schematic presentation of triacylglycerol (TAG) biosynthesis in developing seeds of oleaginous plants. Acyl-CoAs produced from the plastid. They are substrates in glycerol transesterification reactions of the Kennedy pathway. In addition to the Kennedy pathway, TAG can be produced in other three pathways: acyl editing (the cycles of diacylation/reacylation of PC/LPC catalyzed by PLA2 and LPCAT), producing TAG directly with PC and DAG (PDAT plays crucial in this reaction), and synthesizing TAG from a PC-derived DAG substrate (CPT, PLC, and PDCT involve in this pathway). Abbreviations: acyl-CoA synthetase (ACS); sn-1,2-diacylglycerol:cholinephosphotransferase (CPT); sn-1,2-diacylglycerol (DAG); acyl-CoA:diacylglycerol acyltransferase (DGAT); sn-glycerol-3-phosphate (G3P); acyl-CoA:sn-glycerol-3-phosphate acyltransferase (GPAT); acyl-CoA:lysophosphatidic acid acyltransferase (LPAT); lysophosphatidylcholine (LPC); acyl-CoA:lyso-phosphatidylcholine acyltransferase (LPCAT); phosphatidic acid phosphatase (PAP); phosphatidic acid (PA); phospholipid:diacylglycerol acyltransferase (PDAT); phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT); phospholipase A2 (PLA2); phospholipase C (PLC); triacylglcerol (TAG).

2.3.1 Fatty acids desaturation and elongation

After export from the plastid, newly synthesized fatty acids can be further modified by desaturation and elongation to generate PUFAs and VLCFAs. The common PUFAs linoleic acid (C18:2) and α-linolenic acid (C18:3) are synthesized by the sequential introduction of double bonds into oleate catalyzed by membrane bound desaturase enzymes (Baud, 2018). Linoleic acid is synthesized from oleoyl-CoA incorporated into the membrane phospholipid phosphatidylcholine (PC) in the ER catalyzed by the Δ12 fatty acid desaturase 2 (FAD2) enzyme (Kang et al., 2011, Zeng et al., 2017). The Δ15 fatty acid desaturase 3 (FAD3) can further desaturate linoleoyl-PC to produce α-linolenic acid (O'Neill et al., 2011). These two desaturases (FAD2/FAD3) are membrane-bound proteins residing in the ER (Los & Murata, 1998). In order to form a double bond, oxygen and two electrons, supplied by NADH, are required (Shanklin & Cahoon, 1998). Electrons are transferred to a flavoprotein cytochrome b5 reductase, which releases one electron at a time to a carrier protein. Afterward, two carrier proteins, which are cytochrome b5, sequentially transfer electrons to the desaturase (Smith et al., 1990, Shanklin & Cahoon, 1998).

VLCFAs are fatty acids possessing a carbon chain with more than 18 carbon units, and they are synthesized from C18 acyl-CoA by the endoplasmic reticulum (ER)-associated fatty acid elongase (FAE) complex (Haslam & Kunst, 2013). Four reactions are carried out to add two carbon units in each cycle. First, malonyl-CoA and C18 acyl-CoA are condensed by a β-ketoacyl-CoA synthase (KCS) (Chen et al., 2011). The product from the first reaction was 3-ketoacyl-CoA, and it is reduced by the second reaction accomplished by a β-ketoacyl-CoA reductase (KCR) (Beaudoin et al. 2009). The third reaction is dehydrating 3-hydroxyacyl-CoA from the second reaction to 2-enoyl-CoA by a β-hydroxyacyl-CoA dehydratase (HCD). In the last reaction, the acyl-CoA product from the third reaction is reduced to the elongated acyl-CoA (Salas et al., 2005) via activity of enoyl-CoA reductase (ECR) (Zheng et al., 2005). For

monounsaturated fatty acids, like oleic acid (C18:1), it is synthesized via desaturation of C18:0-ACP by a stromal Δ9 stearoyl-ACP desaturase (Baud, 2018). Afterward, these molecules are exported to the ER for further elongation reactions to produce gondoic acid (C20:1), erucic acid (C22:1), and so on (Baud, 2018).

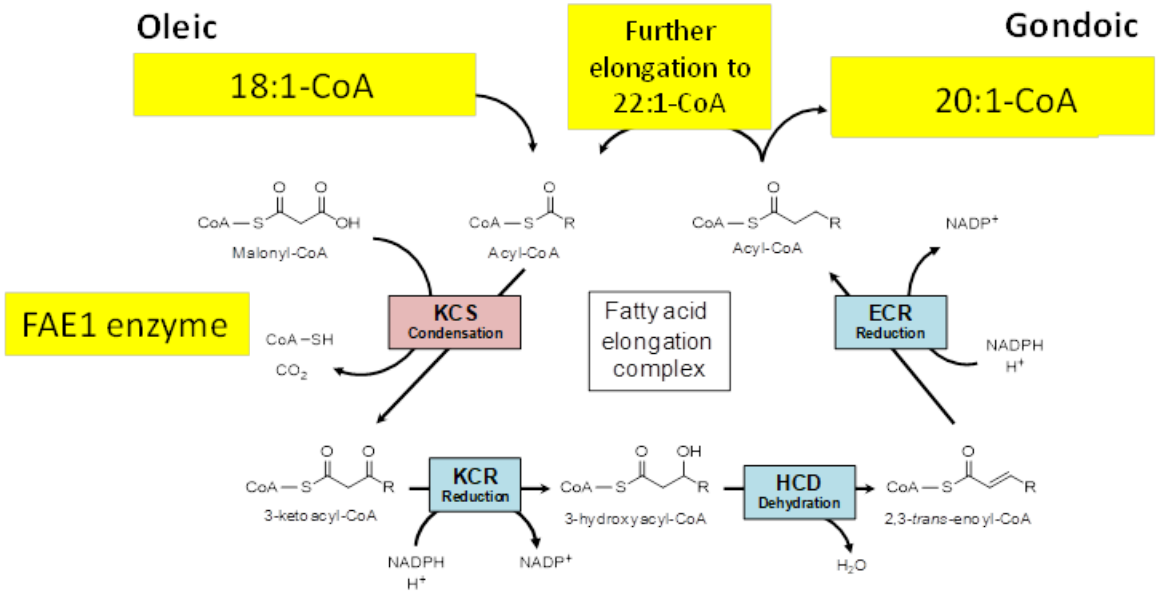

Figure 2.7 General overview of the elongation of oleic acid to gondoic acid occurring on the endoplasmic reticulum (ER) mediated by fatty acid elongase (FAE) complex. KCS: β-ketoacyl-CoA synthase; KCR: β-ketoacyl-CoA reductase; HCD: β-hydroxyacyl-CoA dehydratase; ECR: enoyl-CoA reductase. FAE: fatty acid elongase. (Adapted from Sarvas et al., 2021)

2.4 Modifying the oil profile of camelina

Camelina seed oil contains a mixture of fatty acids varying in chain length and degree of unsaturation. This composition is not ideal for food use where VLCFAs are undesirable, or for industrial and non-fuel use where more uniform composition is generally favoured. Due to its high oil content and amenability to genetic engineering camelina is however considered to be a

versatile platform crop that could be modified for multiple applications based on different market demands (Bansal & Durrett, 2016, Malik et al., 2018, Yuan and Li, 2020). For example, camelina plants capable of producing mid-oleic acid (~40%) seed oils were recently generated by stacking target mutant alleles of genes involved in seed fatty acid desaturation and elongation (Neumann et al., 2021). In their study, they used chemical mutagenesis with EMS (ethyl methanesulfonate) to generate a mutant population and GC was applied to determine the oil profile of seeds starting at the M_3 generation. By analogy to the fatty acid biosynthesis in *A. thaliana*, they identified and studied the mutations that happened in fatty acid elongase1 *(FAE1),* fatty acid desaturase2 *(FAD2),* and fatty acid desaturase3 *(FAD3)*. They transferred the mutant DNA sequence into yeast to understand the influence on the corresponding enzyme function caused by those mutations. According to the fatty acid composition obtained from GC, they concluded that *FAE1* mutant could lead to increased C18:1 and decreased C20:1, *FAD2* mutant could result in enhanced C18:1and reduced C18:2, and *FAD3* mutant could contribute to more C18:2 and less C18:3 than the wild type (Neumann et al., 2021). In their study, they revealed that the genetic redundancy resulting from camelina's hexaploid genome could lower the effects caused via a mutation in one of the three homologs (Neumann et al., 2021). As a final step, *fae1/fad2/fad3* mutant lines were crossed to achieve the mid-oleic oil profile. In a further example, CRISPR/Cas9 mediated gene editing was recently used to target all three copies of the camelina *FAE1* genes to generate an oil with very low VLCFA content (Ozseyhan et al., 2018).

Gondoic acid accounts for around one eighth of total oil content in wild type camelina seed oil and has no known healthy effects on people. It is a desirable target to be removed, allowing for increased C18 fatty acid content. Camelina synthesizes gondoic acid in a pathway similar to other species of plants which is through the elongation of pre-formed acyl-CoA substrates by addition of a C2 units derived from malonyl-CoA. Chain elongation occurs in the cytosol and a four-enzyme fatty acid elongase complex (*FAE*) is responsible for this reaction (Sarvas et al. 2021). So far, there is no natural *FAE1* mutant, this gene would be the first target if alteration of

MUFAs is expected. Hence, in 2018, Ozseyhan and his group successfully used EMS mutagen to create a *FAE1B* gene mutant, which led to the amount of VLCFAs decrease over 60% in camelina seed (Ozseyhan et al., 2018).

The CRISPR/Cas9 approach has also been used to reduce the PUFA content of camelina oil through targeting of *FAD2* genes (Jiang et al., 2017). This work resulted in significant enhancement of oleic acid content and reduction of linoleic and linolenic acid, which decreased from ~16% to <4% and from ~35% to <10%, respectively (Jiang et al., 2017).

To data there are no reports of FN mutagenesis applied to camelina to modify its seed oil's fatty acid composition. This project therefore focusses on the use of FN mutagenesis to achieve the improvement of camelina seed oil composition, with reduction of VLCFAs as the primary target. From a nutritional point of view, edible oils should contain a slight amount of VLCFAs, but an abundance of oleic acid, linoleic acid, or α-linolenic acid. VLCFAs such as erucic acid, which has been found to be detrimental to human health (Knutsen et al., 2016), and gondoic acid, which has not been determined whether to be beneficial or harmful to human health, are desired to be eliminated. In contrast, α-linolenic acid known as essential fatty acid, which can lower the risk of ischemic stroke (Simon et al., 1995), is wanted. In this regard, the objective of the project is to decrease VLCFAs and increase C18:3. Aside from altering the fatty acid composition, FN mutagenesis can also enhance camelina's genetic diversity.

2.5 Mutagenesis in crop breeding

2.5.1 Comparison of conventional mutagenesis approaches

Mutagenesis has a long history of use in plant breeding and is a non-targeted, non-GMO approach to generating genetic variation in a species. In the past, induced mutagenesis has been successfully applied to the improvement of soybean oil quality (Patil et al., 2009), the

development of a high-resistant starch for diabetic patients from rice (Chen et al., 2006), and the improvement of groundnut seed quality for increased yields and a better oil and protein profile (Hamid et al., 2006). Induced mutagenesis can be achieved in plants in a variety of ways, including the use of physical mutagens, such as fast neutrons, X-rays, gamma radiations, and ultraviolet rays, or chemical mutagens, such as hydrogen fluoride, N-methyl-N-nitrosourea, methyl methanesulfonate, sodium azide, or EMS (Chaudhary et al., 2019). After treatment with a mutagen, plants with desired changes are identified through forward or reverse genetic approaches. As mutagenesis generally results in the introduction of multiple mutations in a single plant, backcrossing is then required to remove unwanted mutations. The technique is generally inexpensive, but large populations of plants are required, and screening can be slow and labour intensive.

Generally, there are two categories in conventional mutagenesis, which are chemical mutagenesis and physical mutagenesis. Chemical mutagenesis usually uses some chemical mutagen to induce mutation such as single nucleotide polymorphisms (SNPs) (Sikora et al., 2011). EMS, which is a commonly used chemical mutagen, is an alkylating agent that can induce base transitions in DNA, often at a high frequency (Talebi et al., 2012, Chen et al., 2013). A number of outcomes can result from base transition/substitution if it occurs in the coding region of a gene, these include missense, which results in the substitution of a different amino acid in the encoded protein, nonsense, which causes a codon to be changed to a stop codon, or null mutation, in which the substitution does not affect amino acid production, thus resulting in no change in gene expression or function. The benefits of using EMS are hydrolysable toxicity and ease of use (Pathirana, 2011). After treating with EMS, the novel mutation can be detected via a variety of means such as next-generation sequencing and targeting induced local lesion in genomes (TILLING) (Henikoff et al., 2004, Espina et al., 2018).

In comparison with chemical mutagenesis, physical mutagenesis has distinct advantages and characteristics. Physical mutagenesis includes a number of different approaches, such as the use

of ionizing radiation (IR). One of the main characteristics of mutations resulting from IR is that the technique can induce point mutations, but also deletions, insertions, and rearrangement of chromosomal DNA. Deletions size may range from a single base to a few million bases (Mb) (Sakai et al., 1987). Gamma rays and fast neutrons (FN) are representative IR approaches. FN mutagenesis is a method of producing mutations by bombarding or irradiating cells with neutrons (Kumawat et al., 2019). It can result in the deletion of DNA sequences ranging from 1 base pair to 18 million base pairs with 1 base pair to 4 kilobases deletions being most frequent (Li et al., 2016). Incorrect DNA repair in the plant may also result in chromosome rearrangement (Gilchrist & Haughn, 2010). The advantage of FN mutagenesis over gamma-ray is that its relative biological effectiveness is higher than that of gamma-ray, since FN mutagenesis leads to a high proportion of double-strand breaks that are not easily repaired (Tanaka et al., 1999). Comprehensively, the contrast between this process and chemical mutagenesis is striking. Therefore, as a form of irradiation mutagenesis, FN mutagenesis has been widely utilized in plants to induce large structural variants in their genomes (Wyant et al., 2020). However, large structural variants such as deletion, insertion, and chromosome rearrangement are a double-edged sword. On one side, due to the large alteration of the DNA sequence lethality can be a problem, especially in diploid species that lack gene redundancy. On the other side, genomic deletions, compared to the SNPs generated by chemical mutagenesis, are more likely to result in gene silencing and the generation of novel functional phenotypes. As a hexaploid organism with genetic redundancy, camelina is amenable to FN mutagenesis, and this mutagenesis approach is a promising technique for the creation of mutants for seed oil improvement. Since FN irradiation can cause numerous mutations in a genome, different mutations occurring at different loci may affect different genes expression, thereby presenting multiple novel phenotypes after the application of FN irradiation. Despite this, it is likely that one mutant line containing a desired mutation will also contain a number of other non-related mutations, which will need to be purified through further backcrossing. Accordingly, purification of the mutated genome to obtain

a stable and clear genotype is necessary after the observation of a desirable phenotype in the FN population.

2.5.2 Screening of mutagenesis populations to identify traits and genes of interest

Since conventional mutagenesis is a non-targeted approach, mutations can occur anywhere in the genome, with the probability of generating a particular mutation depending on the mutagen, mutation frequency and the size of the target genome. A large population of plants is generally required and these need to be screened to identify lines with the phenotype of interest or mutations in a target gene. Screening can be classified into forward genetic approaches and reverse genetic approaches. For forward genetics, plants are screened for phenotype, then the genetic basis of the phenotype is established. Reverse genetics begins with the identification of mutation and the subsequent determination of phenotype. Both approaches are now heavily dependent on DNA sequencing and bioinformatics. Moreover, a reverse genetic approach can be accomplished in two ways, one by using traditional methods such as PCR screening, and the other by utilizing bioinformatics to examine the sequence data generated by next-generation sequencing techniques to achieve the same goal.

With a forward genetic approach it is possible to identify a line of interest by using the phenotype of the mutant line. For example, seed fatty acid composition can be determined for lines in a mutant population to identify lines with altered composition, such as the reduction of C20:1 content in camelina. In the case of a large population, this screening approach is time-consuming and labour-intensive and requires high-throughput methods. The genetic basis of the phenotype is then determined through techniques such as mapping, or more recently GBS. If a desired mutation is not expected to result in an easily detected novel phenotype, the forward genetic approach should be changed to a reverse genetic approach.

An example of a reverse genetic PCR based screening technique is TILLING which was a popular method to screen for mutation caused by single nucleotide substitution (Figure 2.8). This method utilizes the mechanism that heteroduplexes formed during a PCR amplification process can be cleaved by CEL 1, a single strand-specific DNA endonuclease. The technique utilized pools of DNA collected from multiple plants but requires knowledge of the genome sequence of the target plant as gene specific primers must be generated. Resolving the cleavage products of a CEL 1 digest on a polyacrylamide gel will reveal two bands for the mutants and one band for the wild-type lines indicating that identification of the mutant fragments in a pool mixed with mutants and wild-type fragments can be achieved (McCallum et al., 2000).

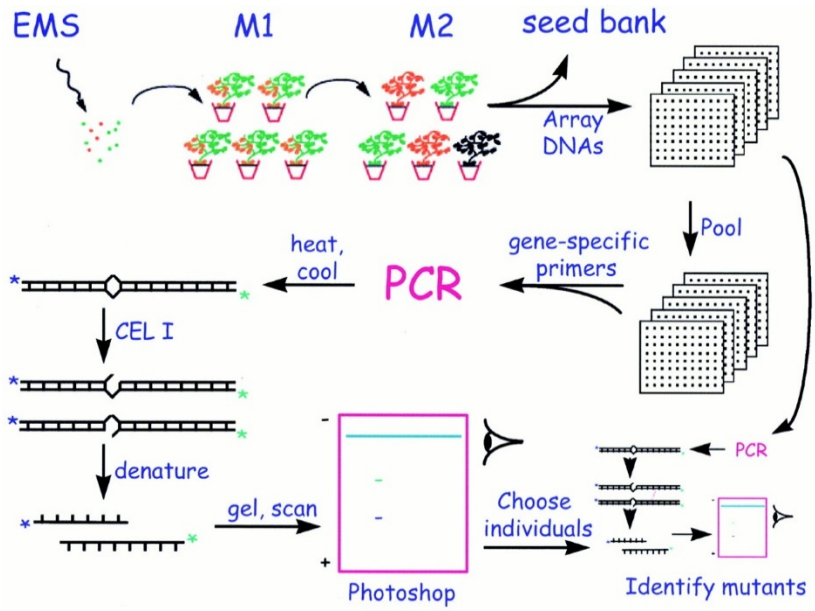

Figure 2.8 A general scheme of the TILLING approach. (Adapted from Colbert et al., 2001).

An alternative, and ingenious reverse genetics system to screen and determine the deletions in arabidopsis caused by FN mutagenesis was designed, which applied PCR with a pooling strategy to screen for mutations in target genes in a population of FN mutants (Li et al., 2001). The DNA fragments generated in lines carrying a deletion were shorter than the wild-type DNA

fragment, so simple electrophoresis can distinguish the mutants, which have shorter fragments that move further on the agarose gel than for the wild-type. This strategy is also known as discrimination by amplification. A smart design in Li et al.'s study (2001) was the pooling strategy in a way to organize the samples into pools of progressively more complex DNA samples from all mutant lines (Figure 2.9). This method can dramatically reduce the labour of PCR, which can determine the mutation line within several rounds of PCR among thousands of tested lines.

With the massive reduction in DNA sequencing costs, reverse genetic approaches based on high throughput sequencing and bioinformatics are now common. An example is GBS, which can be used to detect either SNPs or structural variants like insertion and deletion. GBS entails the sequencing of genomic DNA from the target lines and prior knowledge of a genome sequence is not necessary. By comparing sequence data between individuals, or to a reference sequence SNPs and other polymorphisms can be detected. These can then be associated to a trait and used to develop perfect markers for marker assisted selection in breeding programs (Deschamps et al., 2012). In other words, the GBS approach combines marker discovery and genotyping to establish a one-step method for analyzing model or non-model species' genetic diversity on a genome-wide scale (Poland & Rife, 2012, Peterson et al., 2014). GBS methods will be described in more detail below.

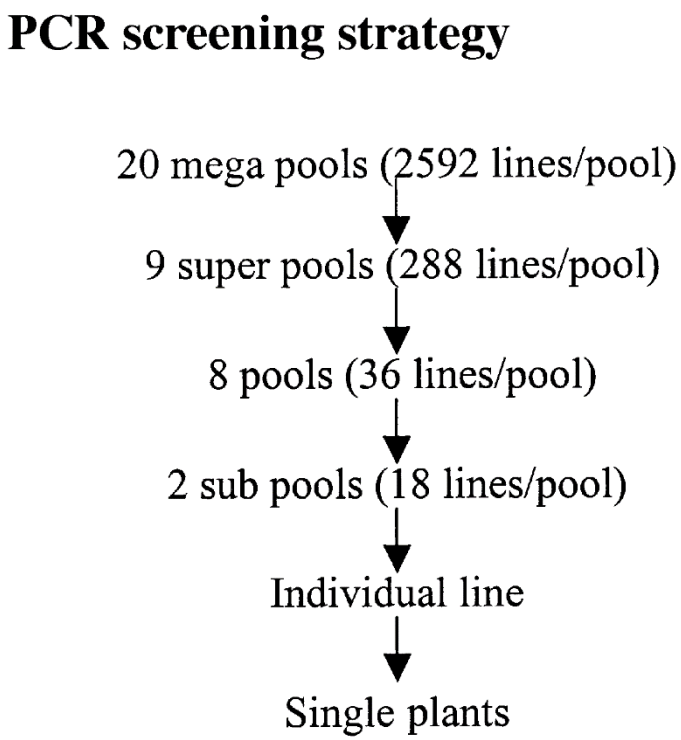

Figure 2.9 Flow chart for the PCR pooling strategy used in Li et al.'s study (2001). From initial PCR on mega pools to identifying single deletion mutant plants. The first round of PCR was applied on the 20 mega pools. If one of them was detected mutation, second round of PCR was applied on the 9 super pools that involved by that mega pool. Similarly, once one of the 9 super pools was detected mutation, third round of PCR was done on the 8 pools that contained by that super pool. If one of them was detected mutation, fourth round of PCR was applied on the 2 sub pools that contained by that pool. Finally, if one of the 2 sub pools was detected mutation, it means at least one line in that sub pool carried the mutation. The final round of PCR would be used on the 18 lines involved in that sub pool to find out which individual line had the mutation.

The hexaploid genome with three subgenomes could be both an aid and a hindrance to the improvement of camelina seed oil by a mutagenesis approach. The large structural variants like

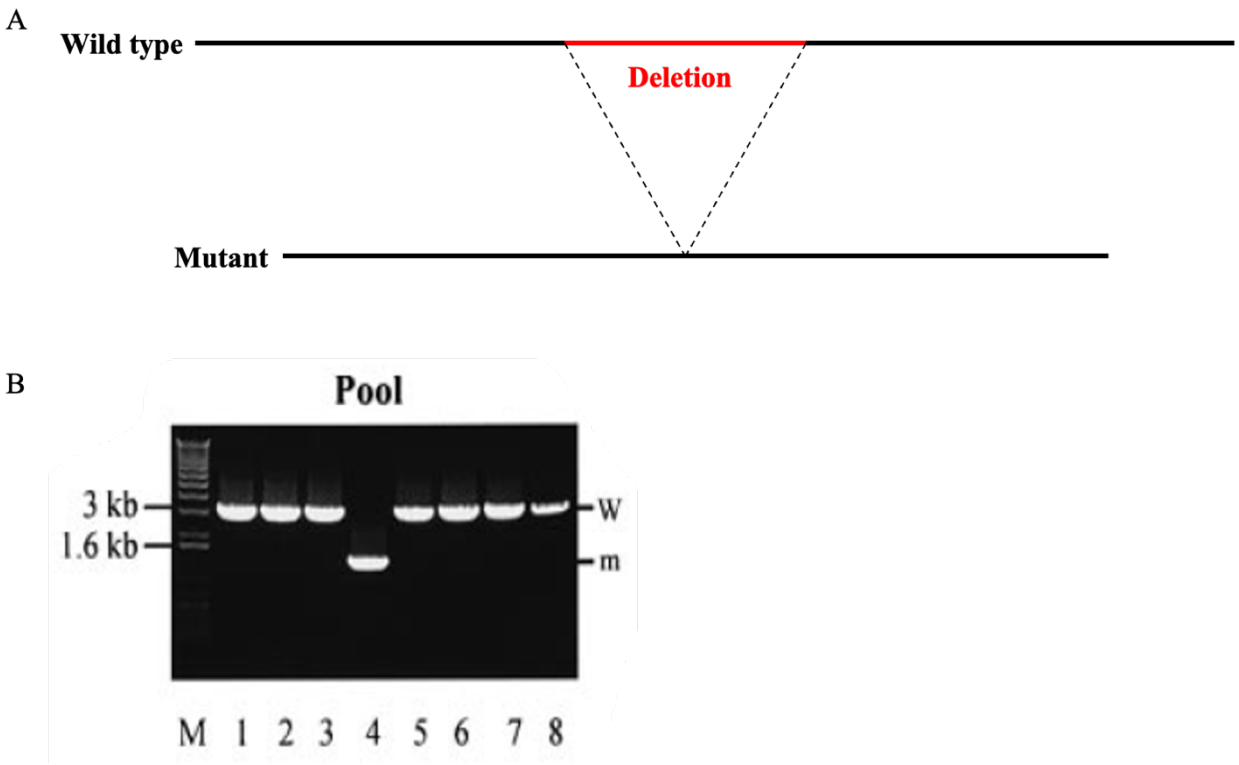

Figure 2.10 a) Display of the divergence of a WT fragment and a mutant fragment. b) PCR display of discriminate amplification. (Adapted from Li et al., 2001)

large indels and the rearrangement, which were traditionally regarded as negative due to causing severe genetic disorders (Gorkovskiy & Verstrepen, 2021), can be tolerated in camelina's genome due to the genetic redundancy, in which the malfunction of one gene can be effectively compensated for by the remaining two genes (Nowak et al. 1997). This characteristic of the camelina genome enriches the achievable mutagenesis methods. However, the large and complex genome of camelina has its merits and drawbacks. Genetic redundancy is an obstacle where an entire change of phenotype is desired. For example, if a decreased seed oil content is the goal, mutations on one copy of the target gene could merely lead to subtle effects on the reduction of seed oil amount. For another example, if gondoic acid is the target to be removed from camelina seed oil, a deletion of one of the three corresponding genes is unlikely to eliminate the fatty acids because the rest of the two responsible genes could partially compensate for the function, which means the final seed oil still contains a lower amount of gondoic acid but it will not be completely eliminated. Hence, more work should be done on genetic engineering to induce

mutations on all three copies of the target gene so that completely alter the phenotype. Furthermore, when only one or two copies of a gene are required to be modified, genetic redundancy raises challenges due to the high sequence similar among three copies of the gene, which makes it difficult to target a specific sub gene. Accordingly, the task of genetic engineering on this kind of hexaploid genome structure is complex and challenging.

2.5.3 Next-generation sequencing technologies and GBS

The sequencing of DNA has advanced considerably since the technique was developed in the 1970s. First generation sequencing techniques include Maxim-Gilbert methods and the more popular Sanger chain termination methods. Development of next generation sequencing (NGS) techniques resulted in dramatic increases in sequencing throughput and reduction in the cost of sequencing. Sanger sequencing as the typical first-generation sequencing, first published in 1977 (Sanger et al., 1977). The dsDNA is denatured at first to form ssDNA that is annealed to primers, and a mixture of dNTPs (arginine (A), tyrosine (T), cytosine (C), and guanine (G) deoxynucleotides) or chain-terminating dideoxynucleotide triphosphates (ddNTP) (Gomes and Korf, 2018). When the ddNTP attaches to the DNA template, the extension is stopped, and the fluorescent marker combined on ddNTP is released. Each ddNTP (ddATP, ddTTp, ddCTP, ddGTP) is bound with distinct fluorescent colour, so the corresponding nucleotide can be identified. Sanger sequencing has been used over 4 decades and still owns merit over short-read sequencing technologies, which can generate reads over 500 nucleotides and has high accuracy, approximately 99.99% (Shendure & Ji, 2008). Nowadays, the application of this sequencing technique is to sequence the regions that are poorly resolved by NGS (Hagemann, 2015).

NGS techniques are often divided into second generation, or short read methods, and third generation, or long read single molecule sequencing. For short-read sequencing, Illumina

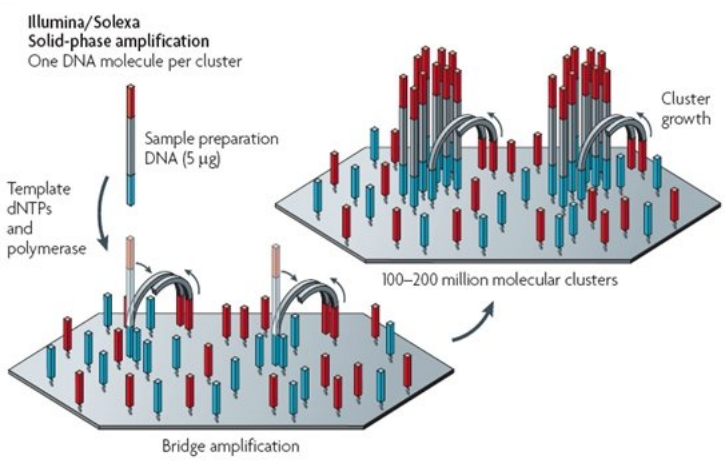

Figure 2.11 Overview of bridge amplification in Illumina sequencing technique. Initial priming and extending of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. (Adapted from Metzker et al., 2010)

sequencing technology is the most widely used. It applies a sequencing by synthesis (SBS) method, which also has been used in several distinct approaches such as Ion Torrent, 454 Pyrosequencing and so on (Slatko et al., 2018). In the SBS method, four fluorescent labelled nucleotides are used to sequence the tens of millions of clusters on a flow cell surface simultaneously. Deoxynucleoside triphosphate (dNTP) with one of four labelled nitrogenous bases (A, T, C, G) serves as a terminator to stop the polymerization, is incorporated to the cluster then the fluorescent dye is detected to identify the base, and it is necessary to wash away any remaining unincorporated nucleotides as well as to remove the terminating group and fluorescent dye before starting the next cycle (Metzker, 2010). The distinction of Illumina is that it utilizes a technique named “bridge amplification” (Figure 2.11), in which the DNA molecules (around 500 bp), carrying proper adapters at each end, are ligated on oligonucleotide sequences complementary to an adapter that fixed on a glass slide (Slatko et al., 2018). Then the DNA molecules formed a bridge shape act as substrates for repeated amplification synthesis reactions to generate clusters (Illumina sequencing technology, n.d.). The sequencing pattern is pair-end, reading around 150 to 300 bp from each end (Sequencing platform, n.d.), so the total length of a read is up to 600 bp (Amarasinghe et al., 2020). There are various platforms developed with different data generation capability, ranging from 1.2 Gb output in 19 hours to 360 Gb output in

48 hours (Sequencing platform, n.d.). The short DNA sequence reads produced by these techniques require extensive bioinformatics processing to assemble them into the contigs and scaffolds required for genome sequencing. Compared to Sanger sequencing, the first-generation sequencing technology, Illumina sequencing has higher discovery power and higher scalability (Myllykangas and Ji, 2010). In 2009, Illumina short-read sequencing combined with Sanger sequencing was first reported to be used to assemble the genome of cucumber (Huang et al. 2009), and this remarkable application of Illumina sequencing was adopted by scientists (Michael and Jackson, 2013). The long-read sequencing technologies, also known as third-generation sequencing, have become more and more popular over the last few years. Technologies for long-read sequencing produce reads typically greater than 10 kb (Pollard et al., 2018). One major advantage of third generation long-read sequencing is improving the accuracy of assembling genomic regions comprised of highly repetitive segments because this new sequencing technology can generate reads spanning several kilobases (Dumschoot et al., 2020). Currently, a wide range of applications in genomics are being explored based on long-read technologies, and early limitations in accuracy and throughput are being overcome (Amarasinghe et al., 2020). Oxford Nanopore Sequencing Technologies (ONT) is representative of third generation sequencing technology, based on a single molecule, real-time sequencing commercially released in 2014 (Van Dijk et al., 2018). According to Wang et al. (2021), this technique employs nanopores embedded in an electrical resistant membrane. A constant voltage is applied to an electrolytic solution to generate an ionic current through the nanopore to drive negatively charged single-stranded DNA and RNA molecules through the nanopore from the negatively charged '*cis*' side to the positively charged 'trans' side of the nanopore. The motor proteins possess helicase activity, which enables DNA or RNA–DNA duplexes to be unwound into single-stranded molecules through a nanopore. They also regulate the speed of translocation by ratcheting the nucleic acid molecule through the nanopore in a stepwise manner. The ionic current changes during translocation correspond to the nucleotide sequence in the sensing region,

which allows for real-time sequencing of single molecules based on the changes in the ionic current. (Figure 2.12) (Wang et al., 2021). In other words, inferring the sequence of bases can be done by examining the specific patterns of current variation because different nucleotides confer different resistances to stretch in the pore (Amarasinghe et al., 2020). Given this feature, ONT is capable of sequencing DNA fragments above 300 kb reads, and even nearly 1 Mb reads (Midha et al., 2019). The longest recorded read length for ONT is 2.3 Mb (Payne et al., 2019). In addition to its powerful sequencing ability, ONT also offers portable devices like MinION, which has been used in Nigeria for metagenomic sequencing of Lassa virus outbreak (Kafetzopoulou et al., 2019). As the extensively used sequencing technologies, Illumina sequencing and ONT sequencing have their unique advantages and disadvantages. Illumina sequencing utilizes a SBS approach to sequence, whose average error rate of each nucleotide is 0.1% (Fox et al., 2014), so it can produce high accurate data. For ONT, it possesses higher raw error rates ranging from 10% to 20% (Zeng et al., 2022) higher than that of SGS (<2%) (Nagarajan & Pop, 2013). Nevertheless, different algorithms are available for error correcting ONT data, which can reduce its error rate to approximately 1% (Sahlin et al., 2021). ONT can be utilized to generate ultra-long reads can easily cover more mutations, especially large structural variants. Accordingly, it is appropriate to the generation of high-quality assemblies of plant genomes with highly repetitive segments (Dumschoot et al., 2020). Therefore, errors caused by short read sequencing in *de novo* assembly, which may be caused by imperfect computer algorithm or interference from the existing of repetitive fragments in the original sequence, can be well validated by the ultra-long reads' information produced by ONT.

Since the advancement of NGS technology, generation of sequence information has been improved dramatically and has become a cost-effective method. The technological advances have altered the notion of variant discovery and genotyping in plant species with complicated genomes or limited public resources available (Deschamps et al., 2012). A comparatively neoteric approach appeared around a decade ago, namely GBS. A well-developed GBS approach

was first proposed in 2011, which applied to maize (IBM) and barley (Oregon Wolfe Barley) (Elshire et al., 2011). GBS accompanied largely by Illumina short-read sequencing offers a powerful and cost-effective approach for genotyping of various crops. The GBS approach generally consist of five steps: 1. generating a collection (library) of DNA fragments representing the full genome, or specific parts of the target genome; 2. barcoding the digested DNA fragments with indexed adaptors; 3. sequencing the barcoded DNA fragments with NGS technique; 4. determining the genetic variants via a bioinformatics analysis; and 5. performing a genetic analysis of sequenced samples based on variation matrix (Baral et al., 2018). In the first step, libraries are often prepared by fragmenting genomic DNA using restriction enzymes. For this it is better to choose restriction enzymes (REs) which do not frequently cut in the major repetitive region of the genome, and form "sticky ends" within the RE recognition site (Elshire et al., 2011). Elshire et al. pointed out that choosing the RE that can produce an overhang constituting at least two nucleotides is useful in enhancing the efficiency of ligation between an adapter and a treated DNA fragment (2011). After the genome is digested by selected REs, the restricted DNA fragments are ligated to adapters containing specific nucleotide sequences referred to as barcodes. The redundant adapters are removed from the DNA sample, which is already modified, through a size exclusion column. Primers are added with the DNA sample and PCR is used to expand the fragment pool. DNA fragments after PCR from the library are used for DNA sequencing (Elshire et al., 2011). The use of barcodes and pooling means that multiple libraries can be sequenced at the same time in high-capacity systems. Alternative methods for library construction include mechanical fragmentation (Hess et al., 2020) and would replace the use of REs for fragmentation. Acoustic waves are used to generate cavitation bubbles, which then collapse to cause high local liquid velocities and break DNA strands, as part of the sonic and ultrasonic method for fragment lengths (Hess et al., 2020). This method may be applicable for short fragments, such as 350 to 550 base pairs required by Illumina sequencing (Truseq DNA

Nano, n.d.), whereas hydrodynamic fragmentation may be preferred when the desired fragment length is longer (Hess et al., 2020).

There are some reports of the successful use of the GBS approach. According to Li et al.'s research (2018), the GBS method accompanied by Illumina sequencing was successfully applied to analyze northern wheatgrass's (*Elymus lanceolatus ssp. Lanceolatus*) genetic diversity. Also, the GBS approach was useful in oat's diploid and polyploid genomes study (Yan et al., 2016, Bekele et al., 2018, Al-Hajaj et al., 2018). Therefore, GBS accompanied by Illumina sequencing technology is an accessible way to reveal the plant's genetic information and identify genetic variation, and it has the potential to be applied to other plants like camelina for genotyping. In addition to use of Illumina short-read sequencing (second-generation short-read sequencing) in the GBS approach, the long-read sequencing approach which also known as third-generation sequencing can be applied as well. The benefit of third generation long-read sequencing is improving the accuracy of assembling the fragments comprised of highly repetitive segments because this new sequencing technology can generate reads spanning several kilobases (Dumschoot et al., 2020). Recently, the cost of long-read sequencing technologies has dropped, and their accuracy is improved (Dumschoot et al., 2020), which gives it potential to be used in combination in GBS approach to validate the sequence information attained from Illumina sequencing. In this project, we used a GBS approach with Illumina sequencing at first to discover a camelina FN mutant's genotype, then we used ONT sequencing to generate long reads that cover the mutant region to validate the data produced by Illumina sequencing. Moreover, we used PCR to amplify the mutant fragment and applied Sanger sequencing to verify the results. In short, we have designed an approach integrating GBS with short-read sequencing technique and long-read sequencing technique to discover and characterize the novel genotype and validate the facticity of the sequencing data via bioinformatics analysis. We also decided to applied PCR with a discrimination by amplification strategy to verify the deletion found in the bioinformatics analysis.

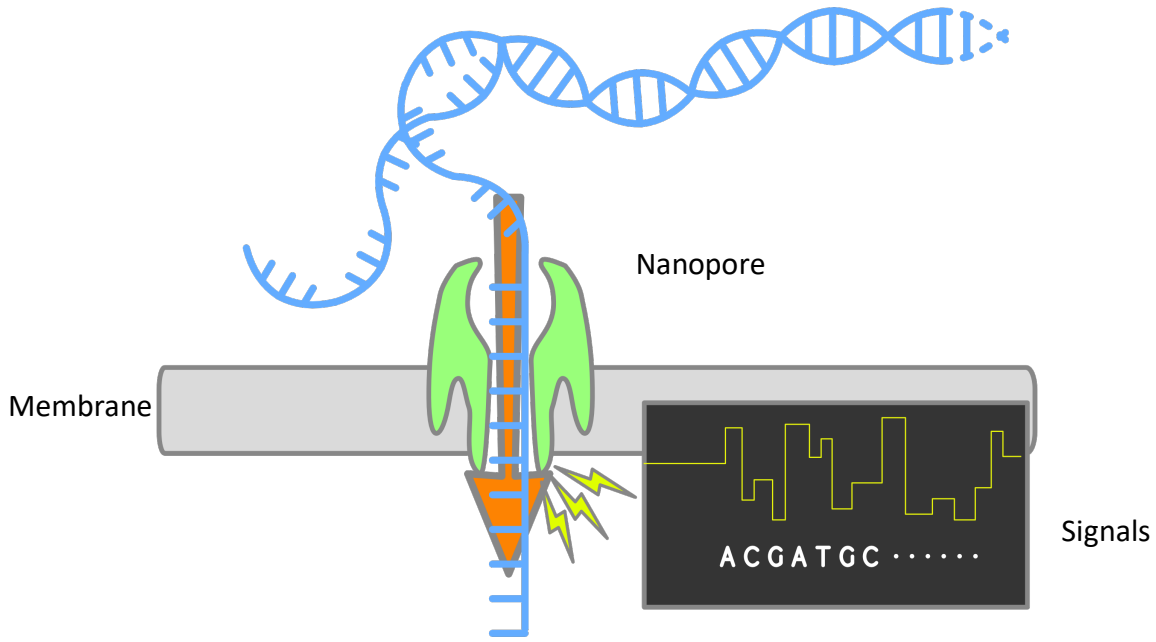

Figure 2.12 Scheme of nanopore sequencing and corresponding current signals. Adapted from https://upload.wikimedia.org/wikipedia/commons/6/6b/202001_nanopore_sequencing.svg.

3. Materials and methods

3.1 Creation and establishment of the *C. sativa* fast neutron mutant (FN) population

3.1.1 Irradiation and seed germination test of *C. sativa*

To obtain the mutated seed, wild-type camelina seeds of the elite line 13CS0583 (Midas^{TM}) supplied by Dr. Christina Eynck AAFC Saskatoon, were sent to UC-Davis, McClellan Nuclear Research Centre, California USA for irradiation. Eight Ziploc bags, each containing 100 grams of seeds were prepared. There were seven bags of seeds for irradiation and one control (not irradiated). Three different irradiation doses were applied: 7, 14, and 28 Gy (Gray). To produce varied irradiation treatments, each batch was irradiated at different doses one time or more than once (Table 3.1). A total of 7 batches were irradiated, and the final dosages are as follows: 7.1, 13.8, 20.9, 28.6, 35.7, 42.4, and 49.5 Gy. The irradiation was completed on July 16th, 2018. The seeds after irradiation treatment were identified as M_1 seeds, and they were packaged in labelled paper bags and stored at room temperature in the laboratory. The moisture content of the seeds sent for irradiation was 5.3%.

Table 3.1 Generation of camelina seeds irradiated by varied irradiation dose.

Dosage							
7 Gy	A (7)		C (7)		E (7)		G (7)
14 Gy		B (14)	C (7+14=21)			F (14)	G (7+14)
28 Gy				D (28)	E (7+28=35)	F (14+28=42)	G (7+14+28=49)

Note. A: the seed bag irradiated once at 7 Gy once. B: the seed bag irradiated once at 14 Gy once. C: the seed bag irradiated 7 Gy once and 14 Gy once, its final dose was around 21 Gy. D: the seed bag irradiated once at 28 Gy. E: the seed bag irradiated at 7 Gy once and 28 Gy once, its final dose was around 35 Gy. F: the seed bag irradiated at 14 Gy once and 28 Gy once, its final dose was 42 Gy. G: the seed bag irradiated at 7 Gy once, 14 Gy once, and 28

Gy once, its final dose was around 49 Gy. The actual final irradiation dose for each bag: A:7.1 Gy; B:13.8 Gy; C:20.9 Gy; D:28.6 Gy; E: 35.7 Gy; F:42.4 Gy; G:49.5 Gy.

Germination tests on the M_1 seeds were conducted over the course of the project to assess the effect of FN irradiation. For the germination tests, Petri dishes with 3 layers of Wattman filter paper were moistened with distilled water and excess water was removed. Fifty M_1 seeds from each irradiation treatment dosage (0, 7.1, 13.8, 20.9, 28.6, 35.7, 42.4, 49.5 Gy) were evenly distributed on the surface of the filter paper then the lid was replaced, and the edges sealed with parafilm to prevent the filter paper drying out. The seeds treated with 0 Gy were used as the control group. The plates were placed in an incubator at 21 ˚C for three days in a dark environment. The observational time point for recording the number of germinated seeds was 24, 48, and 72 hours after setting up the plates. The seeds with a white tip of the radicle pricking out the seed coat are considered as germinated (Figure 3.1). Because camelina seed is tiny, use of a 5x magnifying glass was found to be helpful while scoring germinated seeds. All groups with seeds from each irradiation treatment dosage including the control group were performed in triplicate (3 plates of 50 seeds each per dosage). The germination percentage of each group was the mean of its three repeated experiments. The data was analyzed by simple linear regression test. A total of four germination tests were conducted with Petri dishes, one year apart, with the first test being completed in July 2019.

Additionally, a plant establishment test was conducted on the field. Establishment is the stage at which the germinating seed has turned into a growing plant. This experiment was conducted on two field plots, resulting in two repeat tests. Within each plot, 20 rows, each 10’ long, were planted, with seeds from the same group being planted in double, adjacent rows. The plots were organized as follows: Midas, 0 Gy, 7.1 Gy, 13.8 Gy, 20.9 Gy, 28.6 Gy, 35.7 Gy, 42.4 Gy, 49.5 Gy, and another Midas. Midas groups on the sides acted as guard row to protect the testing variety from shading effect (Keng & Hall, 1987). Additionally, they could serve as a control group for the 0 Gy group. A difference between the Midas group and the 0 Gy group

(control group) was that the Midas seeds used were fresh, harvested Summer 2021, whereas the 0 Gy seeds were the same batch as the irradiated M_1 seeds. For each row, 150 seeds were planted giving a total of 300 seeds for each treatment group. Planting date was 25th May 2022 and plants were observed periodically over the summer. Plant establishment was measured once on 12th July 2022, when plants were at the early flower stage. Plants were counted at ground level with a single stem indicating that one seed had germinated and turned into a growing plant.

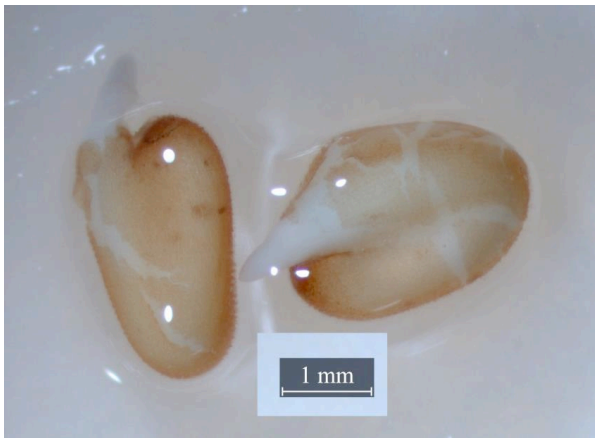

Figure 3.1 Germinating camelina seed showing the white tip of the radicle emerging from the seed coat.

3.1.2 Establishment of a camelina FN population, phenotype characterization, and nomenclature

The M_1 seeds treated with 42 and 49 Gy were cultivated in a controlled environment greenhouse. The temperature of the greenhouse was 18-22 ˚C from 8:00 to 20:00, and 16-20 ˚C for the rest of day. The humidity was set at 75%. The photoperiod was from 4:00 to 20:00. M_1 seeds were planted in hydrated soil Sunshine 3 or PROMIX HP in 12.7 cm (5 inches) diameter round pots, then transferred to the greenhouse for growing. One seed was planted in each pot, and the M_1 plants were allowed to self-pollinate. Then M_2 seeds were harvested from individual

M_1 plant. This planting cycle repeated for generating M_3 seeds. Establishment of the mutant population would make the original lines in the population trackable in phenotyping or genotyping analysis. Any visual phenotypes were observed and recorded photographically during growth. The seed fatty acid composition phenotypes were determined starting from M_3 seeds. The planting cycle would be resumed from M_3 seeds if the lines carried a noteworthy phenotype. Seeds from individual plants were collected in small envelopes with clear labels and stored at room temperature.

Figure 3.2 Scheme of the establishment of FN population.

For nomenclature, each mutant lineage was given a name beginning with FN. Corresponding irradiation dose was written following the "FN". FN42/FN49 was the name assigned to all mutant lineages in this project since only the seeds irradiated with 42 and 49 Gy were planted for the purpose of generating next-generation seed and plant materials. Each lineage was assigned a distinct number and was linked by the symbol "-". Accordingly, a lineage would be called FN42/FN49-x. In the case of the M_2 generation, a number would be appended to its lineage name. FN49-38-1, for example, indicated that it was of the M_2 generation and belongs to the FN49-38 lineage. Seeds and plants of subsequent generations would be identified by the same nomenclature, for example a plant from the next generation of FN49-38-1 would be FN49-38-1-1.

There were two crossing events during this project with lines treated as lineages. One was from Male sterile line x Midas, and the other was from Midas x *fae1b* mutant. Their respective designations were "BLK" and "FNCX". Names were assigned according to the same logic as others. Accordingly, "BLK1-8-1" indicates that it was a member of the BLK1-8 lineage, and it belongs to the F_2 generation. As well, the designation "FNCX16-1-1" indicated that it belongs to the FNCX16-1 lineage and that it was of the F_2 generation.

In this project, "M" represents the generations of FN mutant lines, and "F" represents the generations of backcrossing lines.

3.1.3 Determination of seed fatty acid composition via gas chromatography (GC)

The determination of seed fatty acid composition by gas chromatography began with M_3 seeds. Fatty acid methyl esters (FAMEs) were derived from total seed lipids by acid-catalyzed (methanolic H_2SO_4) transmethylation were prepared for the GC analysis. Batches of 30 seeds were weighed and transferred into one 16x100 mm Pyrex screw-top methylation tube, and 1.5% (v/v) sulphuric acid in methanol (1.5 ml) and toluene (0.4 ml) were added. Fifty microlitres of internal standard 10 mg/ml triheptadecanoin in toluene was added to each tube and tubes were securely closed. After a 2-hour incubation at 90 ˚C, tubes were allowed to cool, then 1 ml 0.9% NaCl solution and 1 ml hexane were added into each tube. The tube was vortexed briefly, then centrifuged for 1 minute at 1000 rpm at room temperature in a Thermo Legend XTR centrifuge to separate phases. Two hundred microlitres of the hexane phase were transferred from each tube to a labelled GC vial with a glass insert. The vials of FAMEs from each sample's seeds were loaded to GC autosampler for subsequent analysis by GC-FID or stored in the fridge at 4 ˚C to wait for loading. The prepared FAMEs were separated by GC using an Agilent 6890N GC (Agilent Technologies, Santa Clara, CA, USA) equipped with a DB-23 capillary column (0.25

mm x 30 m, 0.25 μm thickness; J&W, Folsom, CA, USA) and a flame ionization detector (FID) as described previously (Sun et al., 2017).

3.1.4 Determination of the alteration of fatty acid composition during seed development stage

To monitor changes of fatty acid composition in camelina seed during seed development, 4 different time points were set for seed collection, which were 22 days, 30 days, 38 days, and 46 days after flowering. Midas, FN49-38-1-3-12, and FN49-38-1-11 were selected for this experiment. Plants were planted according to the procedure described above and three individuals were planted for each line. The flowers just beginning to bloom were tagged with four different colour tapes representing four different observation times. On a single plant, there were thirty flowers tagged with each colour tape. Pods marked with tape were removed from the plant at the corresponding observation time. Seeds were removed from the pod with a tweezer until 200 seeds had been collected from each plant. The 200 seeds of a single plant were put into a clear tube with 1.5ml of sulphuric acid (1.5% v/v) in methanol and then methylated for subsequent GC analyses.

3.1.5 Pollen viability test

A pollen viability test based on Li's protocol (Li, 2011) was used to determine fertility of pollens grains in the flowers of mutant lines. Fluorescein diacetate (FDA) sucrose solution was prepared by adding 50 μl of 0.2% w/v FDA in acetone (2 mg FDA/ml of acetone) to 1 ml of 17% sucrose in water and this was used to stain pollen-containing flowers that are about to bloom.

One 30 µl drop of the solution was placed on a slide (25x75x1 mm, VWR Micro Slides) and dabbed several pollen-containing anthers into the drop. Fluorescence images were collected using a Zeiss Stereo Lumar Fluorescence Binocular Stereoscope equipped with an HBO 100 Mercury fluorescence light source, 1.5x NeoLumar objectives and AxioCam 512c and 512m cameras, using a Lumar 47 HE filter set (489047-9901-000: excitation BP 436/25 nm DMR 25, beamsplitter FT 455 HE, emission BP 480/40 nm DMR 25). Images were processed using ZEN 2 pro software (www.Zeiss.com). The final magnification was 80x.

3.2 Genotyping of mutant lines

3.2.1 Extraction of genomic DNA from plants

Genomic DNA from the young leaf tissue was prepared using a CTAB (cetyl trimethyl ammonium bromide) DNA extraction approach modified based on Shukla et al. article (2016). The CTAB extraction buffer was prepared as follows: 100 ml CTAB buffer, 5 ml of 2M Tris-HCl pH 8, 4 ml of 0.5M EDTA (ethylenediaminetetraacetic acid) pH 8, 28 ml of 5M NaCl, and 63 ml of sterile water were mixed with 2 grams of CTAB powder with a stirring bar on a heated plate. After the powder was fully dissolved, 1 ml of 1M DTT (dithiothreitol) was added into 100 ml CTAB buffer.

About 2 grams of fresh leaf tissue of young plants were put in each labelled 2 ml Eppendorf tube placed on ice. The collection tubes were sealed with parafilm with a small hole on top, and they were lyophilized for at least 48 hours. After lyophilization, one glass bead was put into each tube and the tube was placed on a milling machine (Qiagen Schwingmühle TissueLyser II, Hilden, Germany) at 20 revolution/second for 5 minutes or until all leaf tissue was milled into powder. Subsequently 500 µl of CTAB buffer prepared ahead was added to each tube and tubes

were mixed to ensure the powder was dissolved entirely. The tubes were incubated at 65 ˚C for 2 hours, then put in -20 ˚C for 10 minutes. After the cold treatment, 500 µl of chloroform was added and the tubes were shaken at room temperature for 5 minutes followed by 10-minute centrifugation at 6000 rpm (Sigma 4-16S benchtop centrifuge, Germany). The liquid was divided into two layers and the supernatant was transferred into a new tube that already contained 250 µl of isopropanol. The tubes were inverted to mix and left in a dark environment overnight. Centrifugation at 5600 rpm for 20 minutes was applied on the next day. The supernatant was poured off and 500 µl of 70% ethanol was added to the pellet to wash away the salts. After another 20-minute centrifugation at 5600 rpm, the pellet was kept, and liquid removed. The tube with pellet was placed on bench to dry. When the pellet was dry, 50 µl of TE was added and left in the same incubator for 30 minutes to resuspend the pellet. After the resuspension, the isolation of genomic DNA was completed, and it was stored at -20 ˚C for future experiments.

3.2.2 DNA sequence

3.2.2.1 Cloning and sequencing of *fae1b*

According to the results from bioinformatics analysis, FN49-38-1 line carried a clear deletion with no scattered fragment in the gap. Its offspring FN49-38-1-2 was chosen to provide template DNA for the cloning of the mutant *fae1b* gene. For cloning, the first step was applying PCR to amplify the target region. The primer set (B523 + B538, Table 3.4) used was same as the one used in PCR screening step Section 3.2.4.3. Initial denaturation of the PCR was performed at 94 °C for 2 minutes, followed by 30 cycles of 30 s at 94 °C for denaturing, 30 s at 55 °C for annealing, and 3 min at 72 °C for an extension. An additional 10 minutes at 72 °C was spent for an extra extension period. The machine was held at 4 °C until the samples were taken out for

further experiments. Invitrogen's Taq DNA polymerase recombinant was used as the working enzymes, which generates 3'-A overhangs on PCR products (Taq DNA polymerase, n.d.). After gel electrophoresis to determine size, the DNA bands were extracted from the gel by using a QIAquick® Gel Extraction Kit (Qiagen, Hilden, Germany). For ligation into a plasmid vector, 4 µl of extracted DNA was added to 1 µl of salt solution (1.2M NaCl; 0.06M $MgCl_2$) and 1 µl of vector (pCR™4-TOPO™) and placed at 4 ˚C fridge overnight for ligation. The ligation mixture was transformed into the One Shot™ TOP10 chemically competent *E. coli* (Invitrogen, Carlsbad, CA) and kept on ice for 30 minutes. A heat shock at 50 ˚C for 30 seconds was applied followed by adding 200 µl of S.O.C medium. The mixture was incubated at 37 ˚C for 1 hour. The mixture was subsequently transferred onto sterile LB agar plates containing carbenicillin (100 µg/ml) by the streak plate method. The plates were clearly labeled and were incubated overnight at 37 ˚C. For each mutant line, two colonies were randomly picked up and mix with 2 ml LB solution contained carbenicillin (100 µg/ml) in three 17x100 mm, in individual 14 ml sterile culture tubes The clear labeled culture tubes covered with sterile cap were placed in a shaker at 200 rpm at 37 ˚C for further incubation. Then, plasmid DNA extraction was accomplished using a QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany). The purified plasmids DNA were sent to Eurofins Genomics (Eurofins Genomic DNA sequencing services, Suit #5700 100 King Street West Toronto, ON M5X 1C7) for sequencing. The technique they applied was Sanger sequencing technology. Two pairs of primers were applied for this sequencing, which were T3/T7 designed particularly for TOPO vector provide by Eurofins, and B542/B560 designed internal primer.

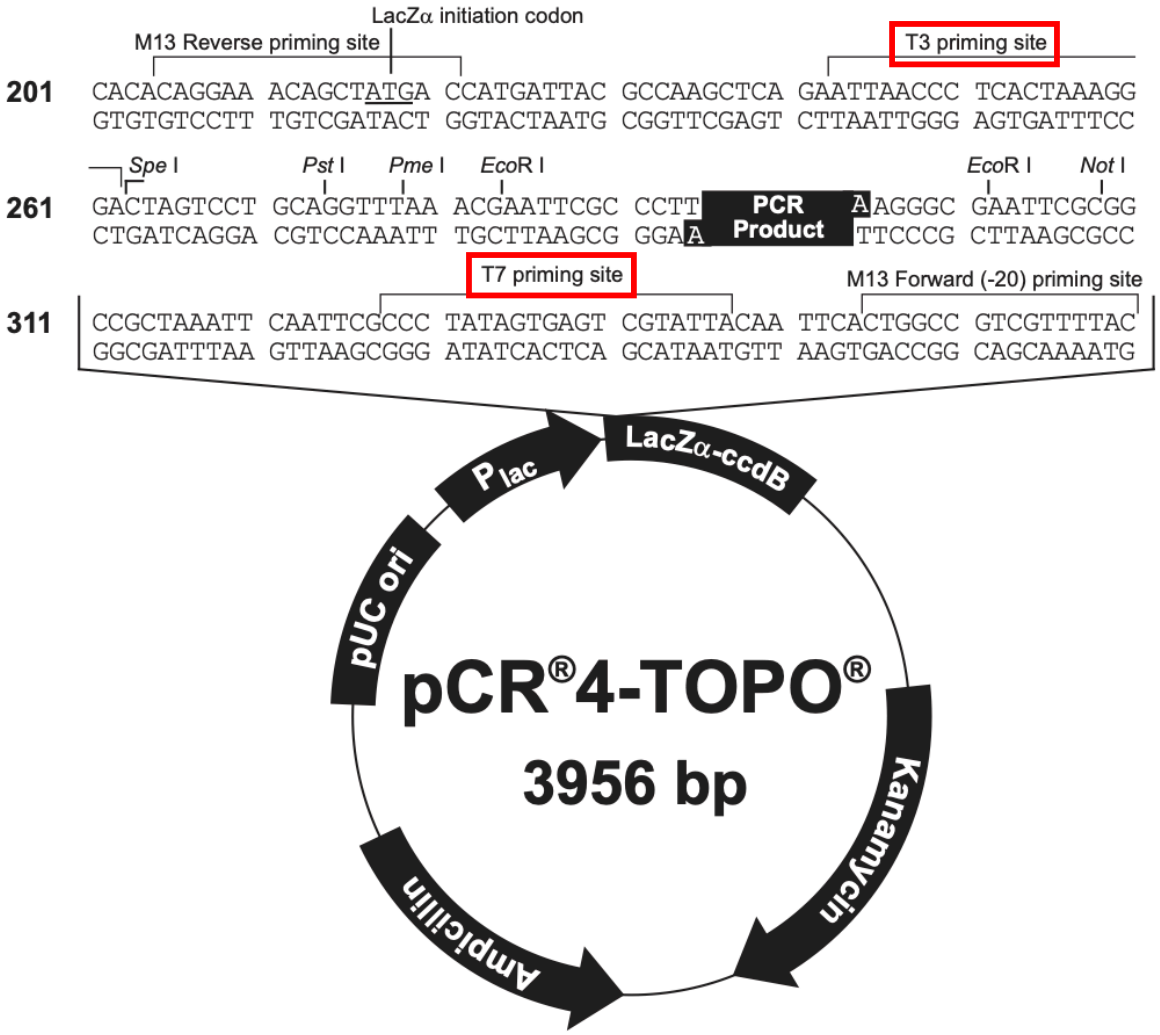

Figure 3.3 Physical map of pCR®4-TOPO® vector and the base information of T3/T7. Adapted from https://tools.thermofisher.com/content/sfs/vectors/pcr4topo_map.pdf. T3/T7 primer set was used for the Sanger sequencing.

Table 3.2 Information of designed internal primers for Sanger sequencing.

Primer name	Primer sequence
B542	CACATGTATCGTATTGAGCACCACTC
B560	CCTTTTGCTTCTATGTATGCCAATTCATACC

3.2.2.2 DNA sequencing through next-generation sequencing technologies

Genomic DNA sequencing was achieved by two sequencing technologies, Illumina sequencing and Oxford Nanopore (long read) sequencing. For Illumina sequencing, a Riptide™ High Throughput Rapid DNA Library Prep (96-well plate format) kit was used to make the library following the manufacturer (Twist Bioscience HQ, South San Francisco, CA) protocol as given in Appendix A. The genomic DNA for library construction was prepared by CTBA method described in Section 3.2.1. The prepared library was sent to the McGill University and Genome Quebec Innovation Centre for sequencing. A total of 105 camelina lineages were selected randomly.

For Oxford Nanopore sequencing, the genomic DNA was isolated from Nuclei following the protocol provided by Omics and Precision Agriculture Laboratory (OPAL) at the Global Institute for Food Security (GIFS), University of Saskatchewan (Appendix B). The prepared library was sent to OPAL for sequencing, which sequenced on a single flow cell using an ONT PromethION sequencing system. The selected mutant line was FN49-38-1-3-12 due to its low C20:1 phenotype.

3.2.3 Bioinformatics analysis

The bioinformatics analysis of all sequencing data was conducted on the Biocluster server (Agriculture and Agri-food Canada Linux server), which runs on the Linux operating system.

3.2.3.1 Processing and mapping of Illumina sequencing data

The raw Illumina DNA sequence data was paired end and was provided in the FASTQ format (nucleotide sequence together with corresponding quality scores). Each sequence included the identification barcode allowing it to be assigned to an individual biological sample. The first step of bioinformatics analysis was demultiplexing which is to group the reads according to the barcode. A FASTQ format file would be created by grouping all reads with the same barcode, which would be achieved via the DemuxFastqs pipeline included in the fgbio tools. During demultiplexing, all adapter sequences were removed simultaneously by the pipeline, which was done to prevent any disturbance in subsequent analyses. The quality control (QC) of the reads was examined using FASTQC and low-quality bases whose Phread quality score were lower than 20 would be discarded by Trimmomatic pipeline.

Following the QC step, the Burrows-Wheeler alignment tool (BWA-MEM pipeline) was applied for mapping. This pipeline would try to match each read to the provided reference genome. This step took the FASTQ format files as input and output the SAM files (Sequence Alignment Maps) for each line. To optimize storage space, the SAM files were converted into BAM (Binary Alignment Map) files. When generating a genome, for example, Midas, contigs were constructed by *de novo* assembly of contigs via the Megahit pipeline. The mapping files generated from BWA-MEM pipeline was used as input file during the assembly process. The ragtag pipeline was used to correct contigs and build scaffolds guided by the reference genome of *C. sativa* DH55. Ultimately, a FASTA file containing polished scaffolds covering the entire Midas genome was generated. The established Midas genome was used as a reference for comparative analysis with genome sequences of mutant lines.

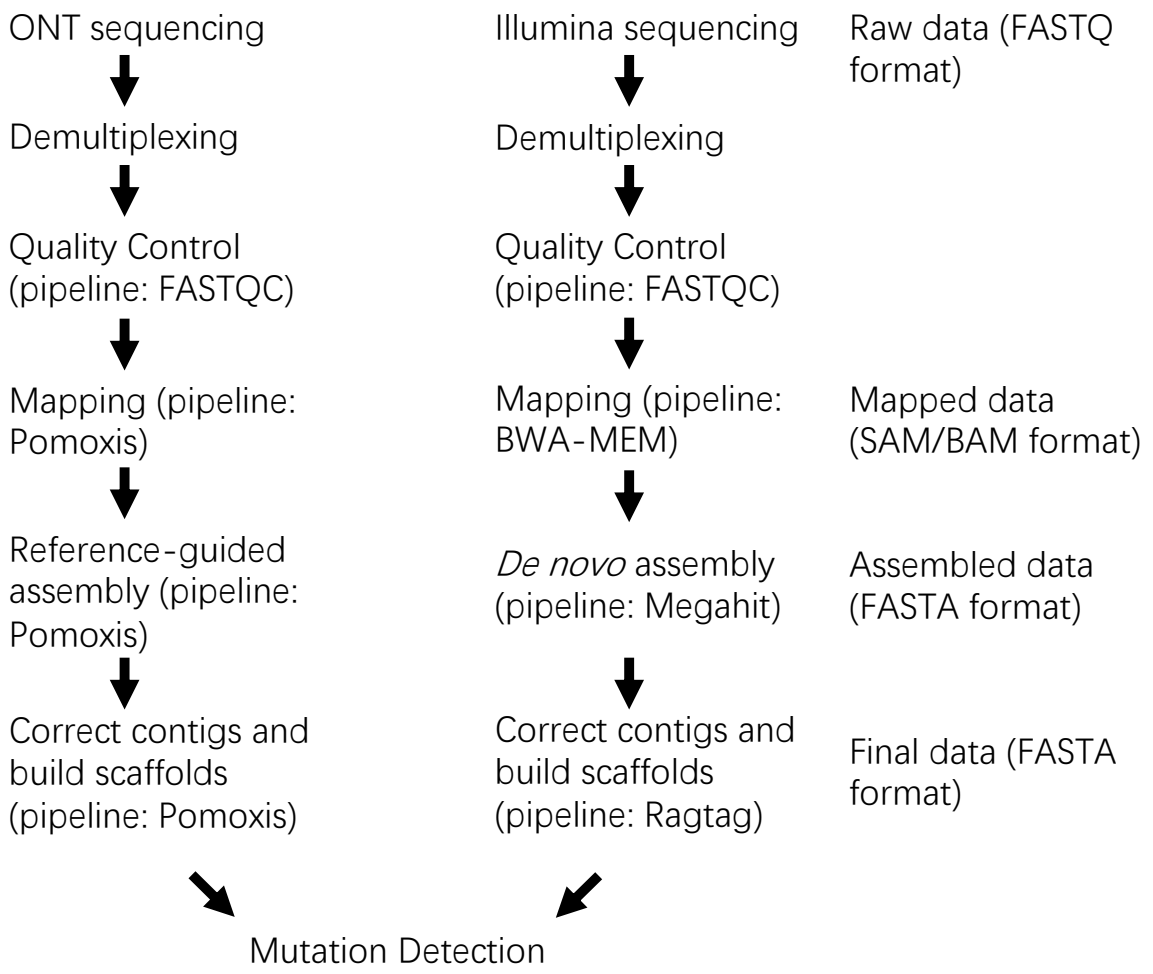

Figure 3.4 Scheme of the processing of sequenced data from Illumina sequencing and ONT sequencing. For Illumina's data, different pipelines were applied. For ONT's data, pipeline Pomoxis, which integrated various software, was used for mapping, assembly, and final polish step.

3.2.3.2 ONT data processing

Similarly, the raw sequencing files generated by ONT (Oxford Nanopore Technologies, Oxford, UK) were also in FASTQ format. It is essentially the same process to process Illumina data and ONT data. Mapping, assembling, and polishing was conducted by Pomoxis software. It was developed by the ONT company, which incorporates various pipeline applications. Because

single mutant line's library was established for ONT sequencing, the demultiplexing step was no required while polishing the ONT sequencing data in this project. If more than one sample built in one library, the demultiplexing will still be necessary.

3.2.3.3 Approaches of mutation detection

Following mapping, sequencing data were converted into BAM format, and after assembly, FASTA format. It is possible to begin the mutation detection process after the mapping stage or after the assembly stage, depending on the final objective. Using the GATK4 HaplotypeCaller pipeline, VCF (Variant Call Format) files were created and visualized with the Integrative Genomics Viewer (IGV), which is a web application (https://igv.org/app/), to obtain a picture depicting where and what mutations appear throughout the entire genome. Detecting mutations in specific regions of the genome was achieved by using assembly data as the second method of mutation detection. The region was extracted from the assembling data acquired with Samtools. DNASTAR software was used to align the extracted sequencing reads with the same region obtained from the reference genome. The differences of aligned sequences between the mutant line and the reference would identify mutation. A specific region was required in advance for retrieving genes from sequencing data.

3.2.3.4 Extraction of specific single read from ONT data

A single ONT read covering the entire target region was extracted as a reference for validating the assembling result obtained from short-read sequence data. The single reads falling within the target region were filtered through Samtools. To determine the ID of the single read

that optimally matches the target region, the BLAST Linux version pipeline was applied. Using command code, the next step was to extract the specific read from the ONT data based on the ID that was previously revealed. A specific read extracted from the short-read data could be used as a reference for guiding short-read assembly or for validating the contigs that have been assembled from the short-read data.

3.2.4 PCR applications

3.2.4.1 Primer design

Primers were designed using Midas, DH55, and a few sequenced mutant lines as references. The MegAlign Pro^{TM} (DNASTAR®) software was used to align the designed primers with the reference so that the match could be evaluated. Most lipid-metabolism-related genes consist of three copies, which means that the specific primer for one of the three genes can only be created in the region where there are a few base differences. In general, the primers were 25-40 bp in length. The GC ratio of used primers was approximately 50%. The repeat region was not considered during design. Through Oligo Calculator (Kibbe, 2007), it was ensured that no self-complementary sequences, hairpin formation sequences, or self-annealing sites were present in the designed primers. Moreover, BLASTn (basic local alignment search tool for nucleotides) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to determine whether the primer would have a chance of bonding to another region of the camelina genome. When using BLASTn, the database selected was "standard databases", the organism chosen was "*Camelina sativa* (taxid:90675)", and the program selection for BLASTn was “highly similar sequences (megablast)”. If search results were not restricted to the gene or region of interest, the primer design would be rejected.

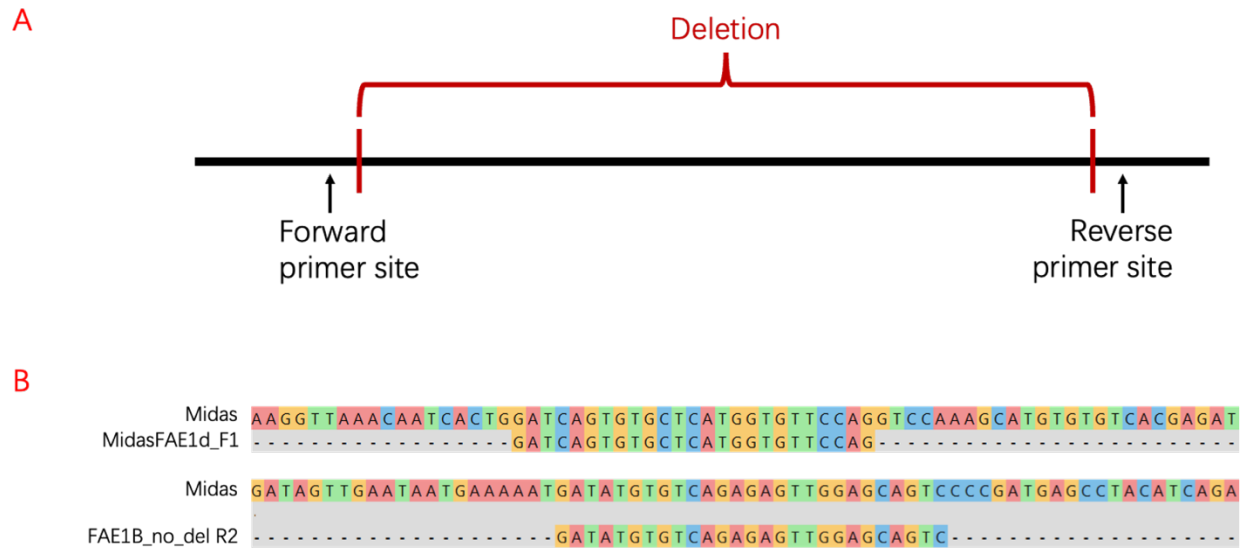

Figure 3.5 Scheme of primer design. A: Scheme of primer flanking site. The deletion is inside the primer amplifying region. B: Base alignment information of designed forward primer MidasFAE1d_F1 and designed reverse primer FAE1B_no_del R2. Midas was used as reference to guide the primer design. Forward primer site was 941 bp ahead coding region. Reverse primer site was 526 bp after coding region. The dot lines represent missing bases.

3.2.4.2 PCR confirmation of the deletion

The primers for PCR amplification of target genes in certain mutant lines were designed based on genome sequence of the wild type (Midas) genome information generated by Illumina sequencing technology. PCR amplification was conducted on the DNA isolated from target mutant plant's young leaf with two specific primers for the target gene (*FAE1B*) to determine whether deletion on the gene occurred in the line. The PCR program was started by a denaturation period at 94 ˚C for 2 minutes. The main period was 30 cycles of 94 ˚C for 30 seconds, 50 ˚C for 30 seconds, and 72 ˚C for 3 minutes. Then an extra extension period at 72 ˚C for 10 minutes was added. Finally, the temperature held at 12 ˚C. The working enzyme used was *Taq* DNA polymerase recombinant purchased from Invitrogen company. The PCR products were

run on a 1% TAE agarose gel stained by SYBR™ Safe DNA Gel Stain and visualized using blue DNA loading dye. DNA size was marked using 1 kb plus DNA ladder.

Table 3.3 Information of primers used for the confirmation of the deletion in *fae1b*.

Primer name	Primer sequence
FAE1B_del F2 (B542)	GAGAGGAGTTTTCTCCTTCCGAAGTGATTTA
FAE1B_no_del R2 (B538)	GACTGCTCCAACTCTCTGACACATATC

3.2.4.3 *FAE1B* mutant screening via PCR

To discover which mutant line carried the *fae1b* deletion in an efficient way, screening accompanied with pooling strategy through PCR among mutant lines was applied. A simple pooling practice was applied. According to the GC results, the mutant lines contained gondoic acid amount lower than 11% were selected. Their DNA extracted from CTAB experiment were pooled into 8 pools (Table 4.6), which 6 pools (pool 1-5 and pool 8) contained DNA from 8 individual lines, 1 pool (pool 6) contained DNA from 7 individual lines, and 1 pool (pool 7) contained DNA from 6 individual lines. Additionally, 2 test pools that mixed with various mutant lines with no deletion and only 1 mutant line with the deletion were established. Test pool 1 was 10:1, which means the pool contained 10 mutant lines with no deletion and one mutant line had deletion. Test pool 2 was 20:1.

Similarly, primers were deigned according to the wild type (Midas) genome information generated by Illumina sequencing technology. The whole progress was almost identical as the PCR confirmation step. The only difference was the annealing temperature was increased to 55 ˚C.

Table 3.4 Information of primers used for the PCR screening of *FAE1B* mutant.

Primer name	Primer sequence
MidasFAE1d_F1 (B523)	GATCAGTGTGCTCATGGTGTTCCAG
FAE1B_no_del R2 (B538)	GACTGCTCCAACTCTCTGACACATATC

3.3 Backcrossing

FN49-38-4 and FN49-38-1-3, two mutant lines from the same lineage FN49-38 carrying *fae1b* deletion, were backcrossed to the wild type (Midas), wherein the mutant line acted as the female (pollen acceptor) and the Midas acted as the male (pollen donor). Buds that were just about to flower were selected for emasculation. Tweezers, disinfected with alcohol, were used to make symmetrical gentle incisions on the corolla to assist in opening it for emasculation. A single flower had six anthers, along with their filaments, which were completely removed with tweezers. To ensure that the carpel, petals, and calyx were not damaged by the tweezers during emasculation, careful handling of the tweezers was essential. As soon as the flower has just bloomed, the stamen covered with pollen was picked off from the Midas and placed in direct contact with the stigma of mutant line to form a fertilized ovum. Afterwards, the petals were softly closed so that the bud would return to its original form. Each bud was wrapped in a small paper bag in which a clear label was attached. The plants grew in the greenhouse under the same environmental conditions as other populations of FN after backcrossing treatment. Mature backcrossed seeds, referred to as F_1 seeds, were collected and planted for the next generation, namely F_1. The fatty acid profile of F_2 seeds from F_1 plants was determined using the GC method described previously. A few F_2 seeds were planted for the purposes of determining the type of genotype as well as for generating F_3 seeds. A PCR method was used to determine the plant's

genotype, using the same primers set (B523+B538) used for the PCR screening depicted previously. Initial denaturation of the PCR was performed at 94 °C for 2 minutes, followed by 30 cycles of 30 s at 94 °C for denaturing, 30 s at 55 °C for annealing, and 3 min at 72 °C for an extension. An additional 10 minutes at 72 °C was spent for an extra extension period. The machine was held at 4 °C until the samples were taken out for analysis. We used Invitrogen's *Taq* DNA polymerase recombinant and Platinum *Taq* DNA polymerase high fidelity as the working enzymes. The GC analysis was conducted on those F_3 seeds as well to study whether the genotype was associated with the seed fatty acid composition.

4. Results

4.1 Establishment of the *C. sativa* Fast Neutron (FN) mutagenesis population

The first objective of the project was to build a population of plants for further study. For this, seeds from the irradiated seed batches (M_1) were planted to generate M_2 seed and establish lineages as described in Materials and Methods. This was continued throughout the duration of the project. Seed viability was assessed and visual and seed fatty acid composition phenotyping was conducted.

4.1.1 Viability and aging of irradiated seed

To determine the effect of irradiation of seed viability and to monitor seed aging, four germination tests were completed, each one year apart, commencing in July 2019. Germination percentage was calculated from the number of germinated seeds / the total number of seeds (Figure 4.1). The control seeds (0 Gy) were from the same batch as the irradiated seed and traveled to and from the site of irradiation together with the experimental material.

In germination test 1, the control group had the highest germination percentage of 67% after 24 hours of incubation. The germination percentages for the groups treated with 7.1, 13.8, 20.9, 28.6, 35.7, and 42.4 Gy were 65%, 45.5%, 52%, 49.5%, 45.5%, 44.5%, and 41.5%, respectively. A decline in germination percentage was observed with increasing irradiation dose, and the group treated with 49.5 Gy recorded the lowest germination percentage of 41.5%. After 48 hours of incubation, germination percentages of all groups exceeded 90%. Additionally, in germination tests 2 and 4, a similar observation was made in which the germination percentage of irradiated

seeds at 24 hours was generally lower than the control with the lowest germination rate corresponding to the highest irradiation dosage (Figure 4.1). For the 24 hour time point in test 3, scoring germination was interrupted. Due to a delay of hours between scoring low and high dosage plates, the data for this time period was not considered reliable. For test 1 and 3 there was no significant differences between samples at 48 and 72 hour time points (test 3), as their *p* values were greater than 0.05 (Table 4.1). In test 4, the higher dosage seeds showed reduced germination at the 48 hour time point, but not at 72 hours. Based on the germination tests, it was concluded that the germination percentage decreased as the irradiation dose increased over the first 24 hours. In all groups, nearly 100% germination was observed after 72 hours, suggesting that higher FN irradiation may slow down germination, but may not be lethal.

To assess plant vigour in the M_1 plants from irradiated seed a field trial was conducted in summer 2022. Two plots were seeded, each with double rows of control and irradiated seeds as described in materials and methods. Plots were seeded on May 25th 2022 and observed periodically over the summer. Photographs of plot 2 at two time point (July 12th 2022 and August 12th 2022) are shown in Figure 4.2. Plant establishment was scored on July 12th 2022, when plants were entering the flowering stage, by counting individual plants in each row (Figure 4.2 C). The plant establishment of the 0 Gy lines was higher than the irradiated lines in both plots, whose percentage of establishment both were 18.3%. All irradiated lines gave a lower percentage of plant establishment, especially the 42.4 Gy line, which was 7.3% for plot 1 and 10.7% for plot 2. For the 20.9 Gy line, it had the highest percentage of plant establishement in both plot, whose average was 14.5%. Overall, there was no clear correlation between establishment and irradiation dosage. Furthermore, the irradiation dosage appeared to affect plant vigour. As shown in Figure 4.2, plants that received higher dosages appeared less vigorous and flowered later. Plants from the 49.5 and 42.4 Gy rows, however, appeared to be more susceptible to pests (particularly root maggots (*Delia* spp)), although no formal measurements were made. As shown by germination test 4, aged seeds experienced a lower plant establishment percentage

than the fresh seeds. Irradiated seeds, regardless of the irradiation dosage, experienced a lower plant establishment percentage than the untreated seeds. Additionally, the plants that received higher doses of irradiation developed more slowly than those that received lower doses. In comparison to Midas, the control group and 7.1 Gy group exhibited similar growth rates. The detailed data of germination tests and the field test was displayed in appendix C.

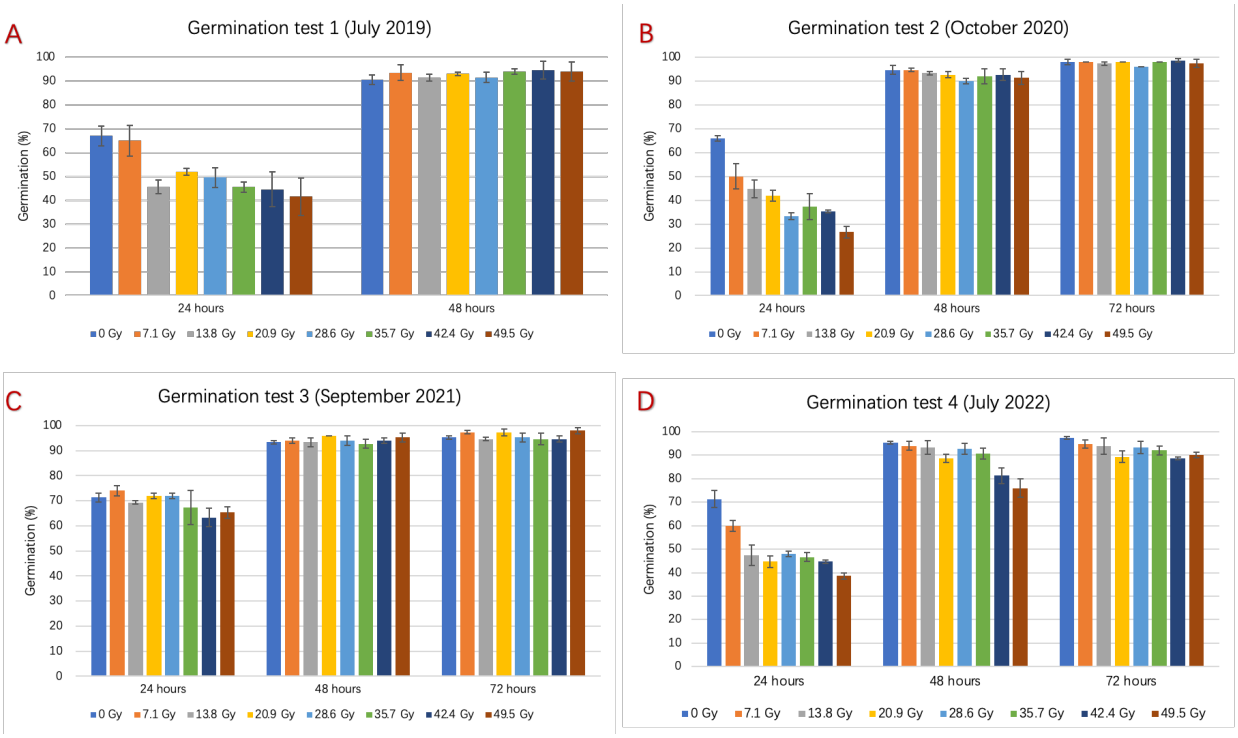

Figure 4.1 Percent germination of irradiated camelina seeds. Germination of control (non-irradiated) and irradiated seed samples were assessed by plating group of 50 seeds on moist filter paper with incubation in darkness at 21˚C. The number of germinated seeds on each plate was recorded at intervals of 24 and 48 for test1, and 24, 48, and 72 hours for tests 2-4. Average germination percentage and standard deviation were calculated for 4 replicates (test 1) and three replicates (tests 2-4). A: germination test 1; B: germination test 2; C: germination test 3; D: germination test 4.

Table 4.1 Simple linear regression analyses of applied dose of FN irradiation and average germination percentage.

	Average germination percentage					
	24 hours		48 hours		72 hours	
	β (SE)	p value	β (SE)	p value	β (SE)	p value
Test 1						
Dose	-0.86 (0.21)	0.007	0.71 (0.29)	0.05		
Test 2						
Dose	-0.91 (0.17)	0.001	-0.74 (0.28)	0.04	-0.05 (0.41)	0.90
Test 3						
Dose	0.80* (0.25)	0.02	0.27 (0.39)	0.52	0.07 (0.41)	0.87
Test 4						
Dose	-0.84 (0.22)	0.008	-0.86 (0.21)	0.006	-0.81 (0.24)	0.01

Note. This linear regression analysis was conducted by R function *lm* (R Core Team, 2020). The significance threshold is 0.05. P-value < 0.05 means the characteristic that the germination percentage decreased as irradiation dose increased was not due to chance. The data was standardized before conducting regression analysis. *The data for the 24 hour time point of test 3 was not considered reliable because scoring germination of the seed was interrupted during this period.

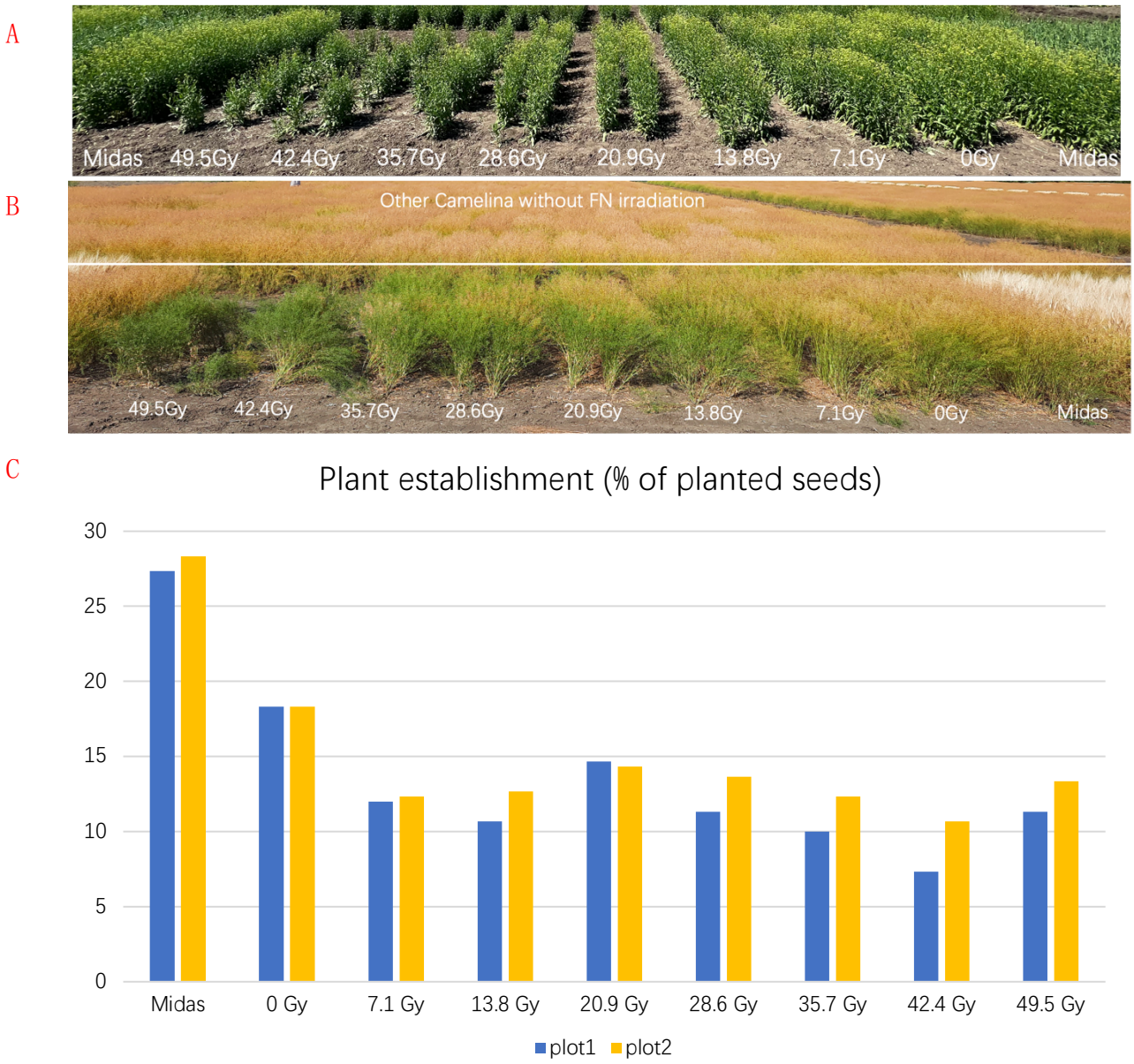

Figure 4.2 Field trial of FN irradiated camelina lines, Summer 2022.

A field test was conducted by planting irradiated (M_1) and control (Midas) seeds at the AAFC Lowe Road Research Farm, Saskatoon Saskatchewan. Plots were seeded on 25th May 2022. For each treatment, seeds were planted as a double row. Plot 2 was photographed on (A) 12th July 2022 and (B) 12th August 2022. Plant establishment was determined by counting plants on 12th July 2022. Percentage establishment ((number of plants/number of seeds planted) x 100) is shown in C for plots 1 and 2.

4.1.2 Organization of FN population

Over the past 4 years, the FN populations were established. The M_1 seeds were planted to generate M_2 seeds. Each M_1 plant gave rise to an independent lineage. The cultivation of some M_2 seeds from each lineage resulted in the production of M_3 seeds. It was considered that M_3 seeds obtained from different M_2 individuals belonging to the same lineage had the same ancestor. The different M_2 individuals were considered to be siblings under the same lineage, and their seeds, namely M_3 seeds, were collected and labeled according to the nomenclature described previously. M_3 seeds could be used to produce seeds for subsequent generations if necessary. There have been 194 lineages established to date, with 113 lineages coming from FN42, 61 lineages coming from FN49.

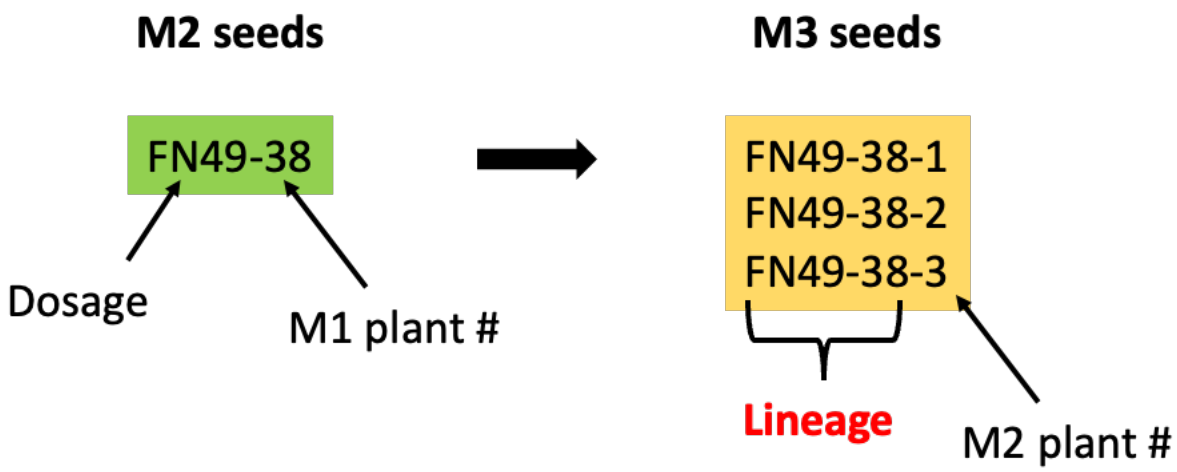

Figure 4.3 A naming scheme for a line of FN individuals and their lineage.

4.1.3 Observation of visual phenotypes

All plants grown for development of the FN population were observed for visual phenotypes. Although most lines did not look very different to plants from non-irradiated Midas seed a small number of plants showed distinct differences (Figure 4.6). Two potential phenotypes observed at the M_2 generation were investigated to determine if the phenotype could be inherited, which would suggest a genetic component that could be the basis of further study. Lineage FN49-33 appeared to have a reduced male fertility trait and lineage FN49-336 appeared to show a difference in leaf shape.

4.1.3.1 Preliminary characterization of a male fertility mutation

The M_1 plant FN49-33 only produced a small number of M_2 seeds. When these seeds were planted, it was observed that two plants FN49-33-6 and FN49-33-7 were producing small empty pods. The anthers of these plants were found to be devoid of pollen (Figure 4.4). To rescue the trait, pollen from Midas plants was transferred to flowers of FN49-33-6 to generate F_1 seeds (the crossing event was referred to as BLK1 (where BLK=blank, i.e., no pollen)). A total of ten F_1 seeds were planted and allowed to self-pollinate to produce F_2 seeds. F_2 seeds from 2 lines, BLK1-4 and BLK1-8 were planted and 32 of the resulting plants were monitored for aberrant pollen phenotypes. To determine the amount of pollen produced by each line, three flowers were randomly selected from a single plant and examined under a microscope. The production of pollen appeared to be variable, so a numbering system was devised to rank pollen abundance from 0 (no pollen) to 4 (abundant pollen equivalent to levels in the Midas anther) (Table 4.2). Scores were assigned to flowers using this system. Among the 17 F_2 plants from the BLK1-4 line, 6 plants produced flowers that all showed normal levels of pollen (score 4) whereas the

other plants had multiple flowers with reduced pollen abundance (Table 4.3). No plants produced flowers without visible pollen. Among the 15 plants from the BLK1-8 line, no plants produced flowers that all had normal pollen levels. For three of the plants, 2 out of the 3 flowers examined did not produce visible pollen. The results indicated that the reduced pollen phenotype appeared to be inheritable and that the genotypes of lines BLK1-4 and BLK1-8 were likely different.

Table 4.2 Pollen level scoring system.

Rating score system	Description
0	No pollen
1	Only a few pollen grains visible
2	Approximately half (50%) of Midas levels
3	Reduced pollen, but above 50% of Midas levels
4	Abundant pollen equivalent to Midas levels

Table 4.3 Pollen scores for flowers from F_2 plant resulting from the crossing of FN49-33-6 (female) with Midas (male).

Plant	Flower 1	Flower 2	Flower 3	Average
Midas	4	4	4	4.0
BLK1-4-1	4	4	4	4.0
BLK1-4-2	4	4	2	3.3
BLK1-4-4	2	3	2	2.3
BLK1-4-6	2	3	2	2.3
BLK1-4-8	4	4	4	4.0
BLK1-4-9	4	4	4	4.0
BLK1-4-11	1	2	2	1.7
BLK1-4-12	3	2	3	2.7
BLK1-4-13	2	2	2	2.0
BLK1-4-14	2	2	1	1.7
BLK1-4-16	2	2	1	1.7
BLK1-4-17	3	4	4	3.7
BLK1-4-18	3	2	2	2.3
BLK1-4-21	4	4	4	4.0
BLK1-4-22	2	2	3	2.3

BLK1-4-23	4	4	4	4.0
BLK1-4-24	3	3	4	3.3
BLK1-8-1	0	1	0	0.3
BLK1-8-2	4	3	3	3.3
BLK1-8-3	2	2	2	2.0
BLK1-8-4	2	3	3	2.7
BLK1-8-6	0	2	0	0.7
BLK1-8-7	3	2	3	2.7
BLK1-8-12	4	4	2	3.3
BLK1-8-13	4	3	4	3.7
BLK1-8-15	2	2	3	2.3
BLK1-8-16	2	2	3	2.3
BLK1-8-17	3	2	3	2.7
BLK1-8-18	2	1	2	1.7
BLK1-8-21	1	0	0	0.3
BLK1-8-22	2	2	2	2.0
BLK1-8-23	2	4	3	3.0

Note. F_2 seeds from lines BLK1-4 and BLK1-8 were planted. For 17 plants from line BLK1-4 and 15 plants of line BLK1-8, three flowers were chosen and scored for pollen abundance.

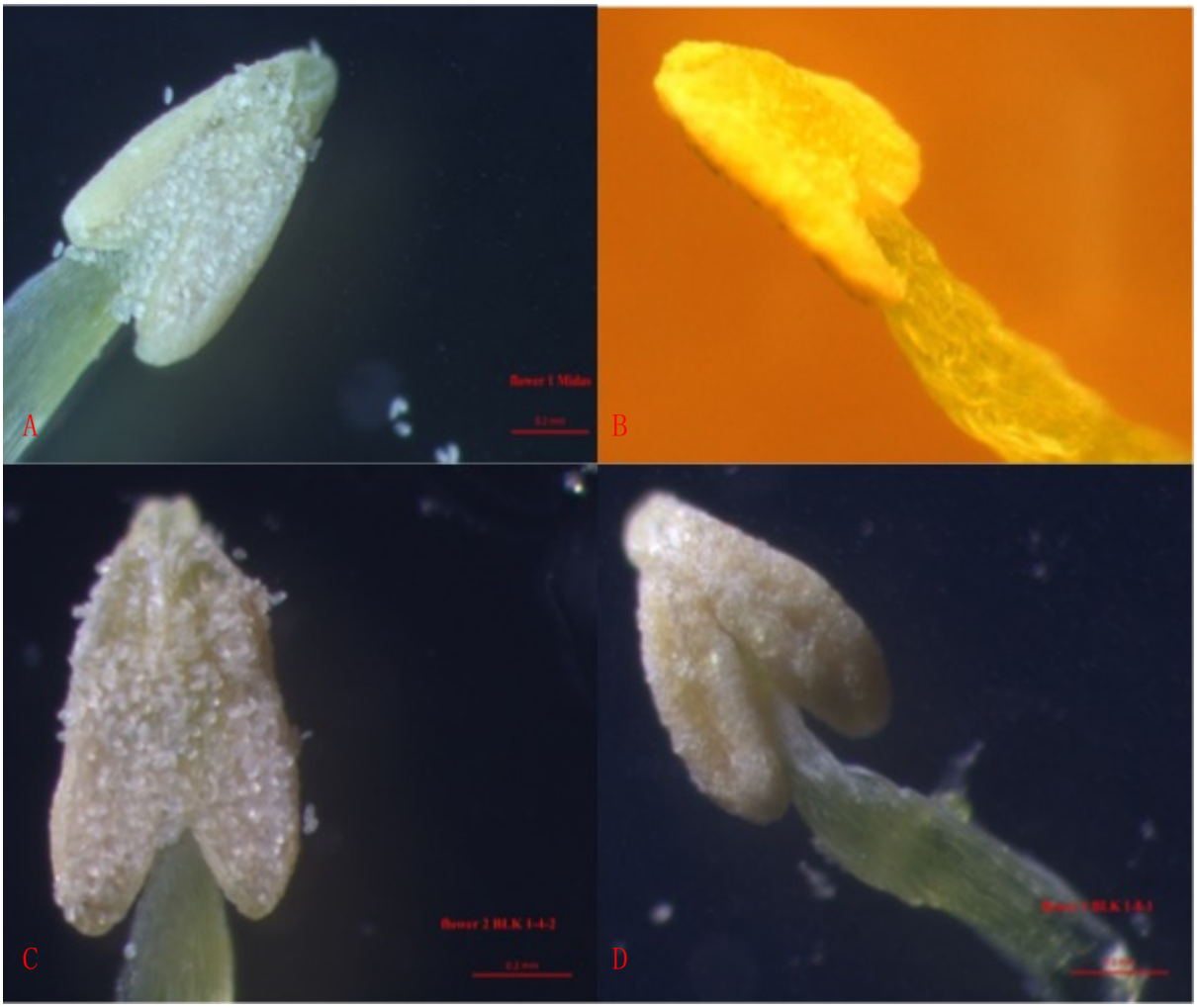

Figure 4.4 Representative anthers from Midas and plants of the FN49-33 lineage. A, anther from Midas; B, anther from the M2 plant FN49-33-7 (without pollen); C, anther from F2 plant BLK 1-4-2 (abundant pollen); D, anther from F2 plant BLK 1-8-1 (no pollen).

To determine whether pollen viability was compromised in these plants, viability was tested by fluorescein diacetate staining. The examination was conducted under the Zeiss Stereoscope microscope without a filter (Lumar 47 HE filter set: excitation BP 436/25 nm DMR 25, beamsplitter FT 455 HE, emission BP 480/40 nm DMR 25) to visualize all pollen, and with the filter to visualize viable pollen. The fluorescein diacetate stain was hydrolyzed by a non-specific intracellular esterase in viable cells with an active metabolism, producing green fluorescein,

which would be retained and accumulated in a cell, causing them to fluoresce green (Kwolek-Mirek & Zadrag-Tecza, 2014). The limitations of the microscope in our laboratory prevented the use of an appropriate filter set. As a result of limitations imposed by the equipment, an option that is relatively suitable was chosen. In this regard, the results obtained in this experiment were considered to be the pilot experiment for future assays when the correct filter set (excitation BP 470/40 nm, emission BP 525/50 nm) becomes available. Comparison of the image with the filter and the image without the filter (Figure 4.5) in terms of the number of pollens would provide viability rates. As the figure 4.5 exhibited, all pollen grains were visible under the filter. Accordingly, the conclusion could be that after crossing, most of progenies were producing less pollen than Midas, but their pollen grains were viable. Nonetheless, the generated results could be improved after buying an appropriate filter set, and the conclusion is required to be verified by further experiments with proper setting microscope. As a male sterility trait may be of use in a hybrid breeding program, this line deserves further investigation.

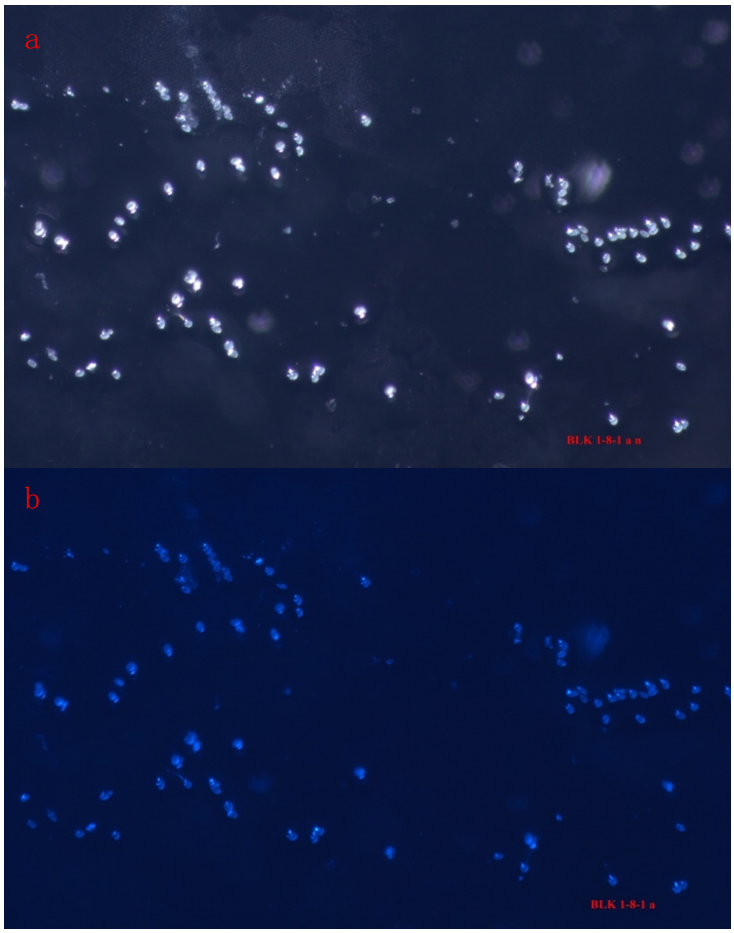

Figure 4.5 Comparison of the pollen with the filter (b) and with no filter (a) of BLK1-8-1. a: The pollen observed with normal mode (no filter applied between the objective lens and the ocular lens) of Zeiss microscope after treating with FDA. b: The pollen observed by Zeiss microscope with a blue florescent filter between the objective lens and the ocular lens. The objective lens was 0.8x and ocular lens were 25x. The final magnification of both figures was 80x. The applied filter was supplied 47 HE filter set whose excitation is 436 nm and emission is 480 nm. BLK1-8-1: backcrossing offspring that was rescued from the male sterile plant.

4.1.3.2 Preliminary characterization of a leaf morphology mutation

A greenhouse was used to grow mutant lines and visual phenotypic data was collected during the growing period. FN49-336 displayed phenotypic changes in the leaf shape during the

growing period, with leaves appearing more toothed than those on Midas plants (Figure 4.6 B). Further generations of this lineage were established to determine whether this altered phenotype was inheritable. In comparison to the leaves of Midas shown in Figure 4.6 B, the FN49-336 lineage leaves were toothed rather than elongated elliptic as observed with Midas. As a result of FN49-336-2's progeny (30 individuals in total), three different phenotypes concerning toothed shape were observed (Figure 4.6 A). There were three M_3 individuals that produced leaves that resembled those of the wild type. The rest of the individuals grew toothed leaves. In addition, some plants exhibited a strong extent of toothed leaves, which were more apparent, while others had a moderate extent. Considering these findings, the trait of FN49-336-2 appeared to be segregating in the M_3 generation. The cultivation of the M_4 generation was undertaken to verify that this visual phenotype may be passed on to its descendants. As a result, 4 individuals were planted for each of the lines FN49-336-2-10 and FN49-336-2-16. Offspring of these two lines produced toothed leaves similar to the parent line in Figure 4.6, which demonstrated that the toothed leaf shape of lineage FN49-336 can be inherited. Additionally, all FN49-336-2-10 individuals produced moderately toothed leaves, while all FN49-336-2-16 individuals produced strongly toothed leaves. Further evidence of the segregation that occurred during the M_3 generation was presented by this observation. The M_4 generation FN49-336-2-16 was highly likely to be homozygous.

Figure 4.6 Toothed leaves of FN49-336-2 and leaves of Midas. A: Three different degrees of toothed leaves were found in the M_3 generation of FN49-336-2. Plants in the picture were 20 days old. B: FN49-336-2-46 put close to Midas for comparison. Plants in the picture were 24 days old.

4.1.4 Screening for variability in seed fatty acid composition

To identify lines with altered seed fatty acid profiles, seed fatty acid composition was determined by GC-FAMES for pools of 30 seeds from individual plants. To increase the chance of detecting inherited traits, M_3 seeds were the first generation examined. While preparing seed

for analysis many lines were observed with small, light (<20 mg/30 seeds) or abnormally shaped seeds.

As poor growing condition caused by environmental factors can impact the fatty acid composition profoundly, GC results for these lines were not considered reliable and were therefore excluded from the data analysis.

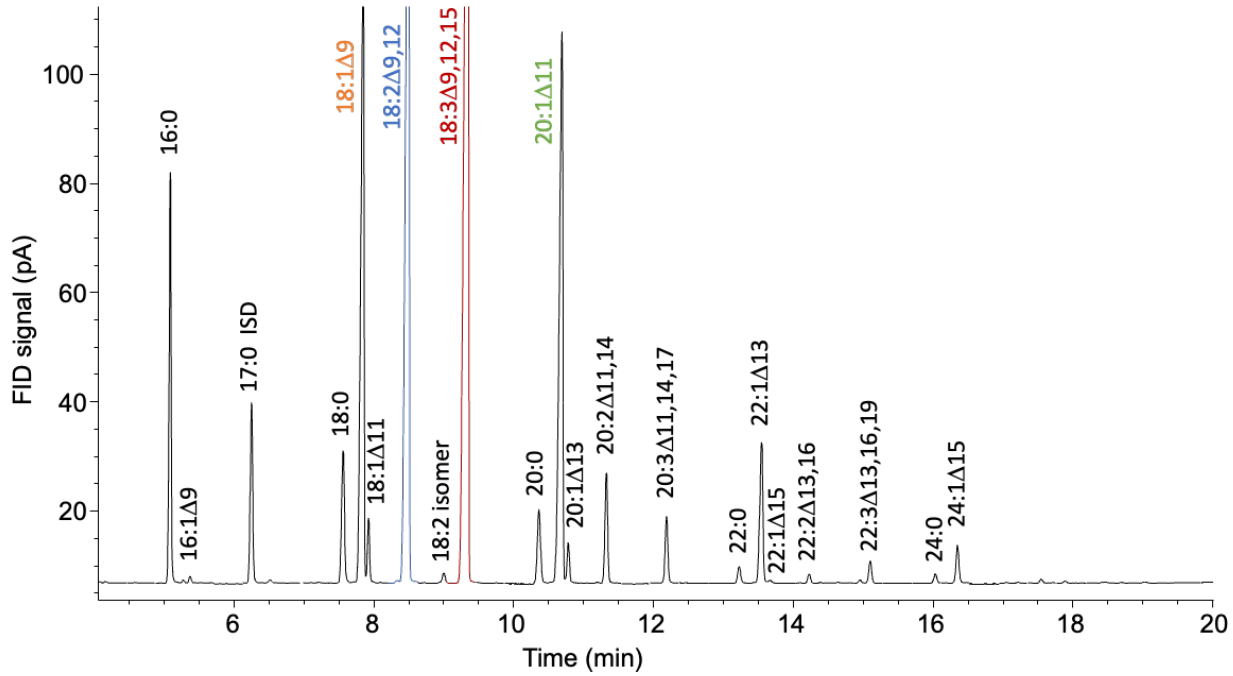

Figure 4.7 GC chromatogram of fatty acid in Midas seed. The fatty acid composition was determined by GC of FAMEs derived from total seed lipids by acid-catalyzed (methanolic H_2SO_4) transmethylation. FID: flame ionization detector.

So far, seed fatty acid composition for a total of over 700 individuals has been determined via the GC approach. Multiple Midas (wild type) lines were also analyzed by GC as the reference. A chromatograph showing the FAMES from a typical Midas line is shown in Figure 4.7. The range of seed fatty acid composition observed in the mutant lines was wide (Table 4.4). Although most of mutant lines showed a fatty acid profile like that of the ancestor Midas. A few

mutant lines showed distinct fatty acid compositions with 12 lineages showing reduced C20:1 content, as shown in Figure 4.8.

Table 0.4 Range of seed fatty acid composition and oil content of the FN population and Midas.

Type of Fatty acid	Mutant lines range (%)	Midas range (%)
C16:0	5.0-18.6	6.0-6.7
C16:1	0.1-4.2	0.11-0.13
C18:0	0.1-11.5	2.5-2.6
C18:1	1.1-18.0	9.9-13.5
C18:2	11.5-59.0	17.9-20.4
C18:3	21.9-45.2	31.9-34.9
C20:0	1.1-4.4	1.6-1.8
C20:1	4.9-16.0	13.0-13.4
C20:2	1.0-3.3	2.0-2.3
C20:3	0.5-2.6	1.3-1.8
C22:0	0.3-0.9	0.3-0.4
C22:1	1.6-6.1	2.9-3.9
C22:2	0.1-0.7	0.2-0.3
C22:3	0.3-1.6	0.5-0.7
C24:0	0.1-0.8	0.2-0.3
C24:1	0.2-1.5	0.7-0.8
Oil content	4.9-49.3	27.0-40.0

Mutant line FN49-438 line contained an increased amount of C18:3 as well as a reduced content of C20:1. M_3 seeds from three individual M_2 plants of this lineage were analyzed by GC (detailed fatty acid composition is shown in appendix D). The mean of their C18:3 content was 39.6%. FN49-438-1 contained the highest amount of C18:3, which was 41.2% of the total fatty acids as compared to 33.4% in Midas. Its sibling lines FN49-438-2 and -4 had a relatively high content of this omega-3 fatty acid as well, which were 39.1% and 38.4%, respectively. Although the C18:3 was enhanced in this mutant line, its C18:1 and C18:2 content was merely slightly lower than that of Midas, which was 11.7% in this line compared to 12.1% in Midas and 18.9%

in this line compared to 19.5% in Midas, respectively. The average amount of C20:1 over these three lines was 9.8%, which was less than Midas's 13.3%. The highest amount of C20:1 was found in FN49-438-1 at 9.9%, followed by FN49-438-2 and -4 at 9.7%.

Another line with divergent fatty acid composition was FN49-287 (detailed fatty acid composition is shown in appendix D). Its progenies FN49-287-3 and -4 contained a higher level of C18:2, which were at 34.4% and 33.5%, respectively, as comparted to 19.5% in Midas. Their content of C18:3 was 21.9% and 22.8%, respectively. For C20:1 in these two lines, it was 8.0% in FN49-287-3 and was 8.2% in FN49-287-4. In addition, the content of C18:1 in these two lines was 12.4% and 12.0%, respectively. Interestingly, the other two progenies FN49-287-1 and -2 had slightly lower content of 18:2 than Midas, which were 17.4% and 18.9%, respectively. In the meanwhile, their C18:3 amount was 35.5% and 34.0%, respectively. Their content of C18:1 and C20:1 was 10.4% and 9.2% in FN49-287-1 and -2, respectively, and was 14.3% and 13.6% in FN49-287-1 and -2, respectively. It was found that the seed weight of FN49-287-3 and -4 was merely 11.0 mg/30 seeds and 10.0 mg/30 seeds, respectively. Therefore, their GC results may not be reliable. These lineages may be incorrectly characterized by a lower level of C18:3 than C18:2 but are worth further investigation as potential *fad3* mutants.

FN49-38 line had a comparatively lower amounts of total VLCFA, whose average of all offspring (M_3+M_4+M_5) was 19.6%, compart to 23.9% in Midas. Additionally, this lineage possessed a lower content of C20:1 whose average of all offspring was around 10.0%, compared to 13.3% in Midas (detailed fatty acid composition is shown in appendix D). Significantly, this mutant line was not the only one that had lower C20:1 content. Mutant lineages FN49-38, FN49-72, FN49-213, FN49-312, FN49-438, FN49-620, FN42-5, FN42-50, FN42-81, FN42-97, FN42-119, and FN42-353 were all determined to have lower C20:1 content than Midas (Figure 4.8). The sample size of each group was 9, 2, 3, 7, 3, 2, 2, 4, 4, 3, 4, and 3, respectively. Furthermore, even after determining the fatty acid composition of FN49-38's M_3, M_4, and M_5 seeds, the C20:1 content was consistently low, with respective mean amount of 9.9%, 9.3%, and 10.4% (detailed

fatty acid composition is shown in Figure d2). This finding suggested an inherited characteristic of low C20:1.

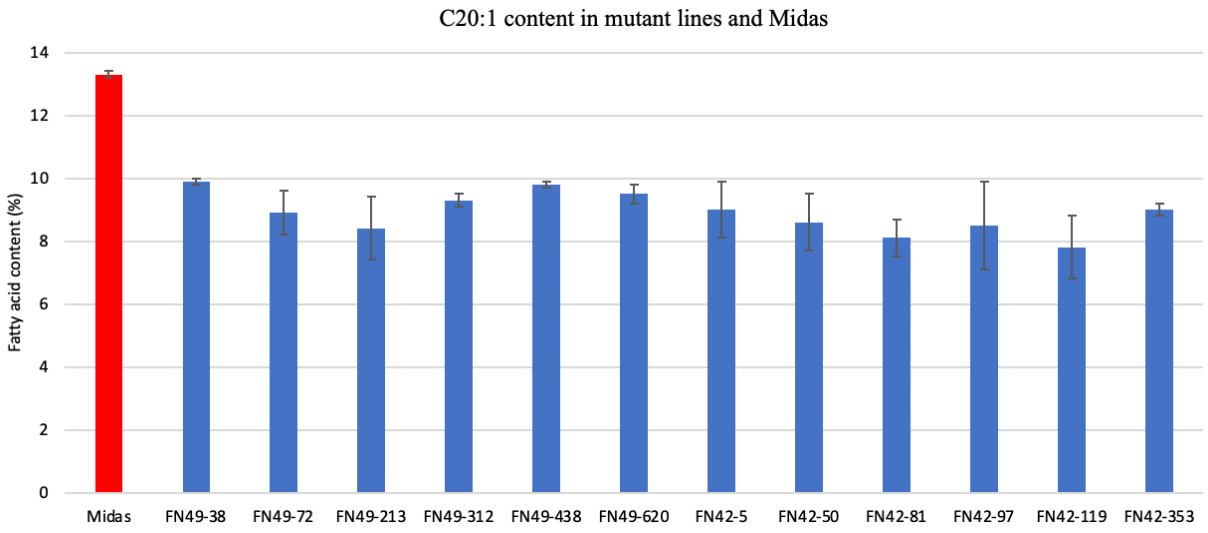

Figure 4.8 Gondoic acid (C20:1) in FN mutant lines and Midas (wild type). Mutant line (sample size): FN49-38 (9); FN49-72 (2); FN49-213 (3); FN49-312 (7); FN49-438 (3); FN49-620 (2); FN42-5 (2); FN42-50 (4); FN42-81 (4); FN42-97 (3); FN42-119 (4); FN42-353 (3). The data was displayed in percentage. The error bar was the standard error. Calculating the mean value was done using all the individuals from each lineage.

Erucic acid content in the FN49-70 lineage was higher than the Midas (detailed fatty acid composition is shown in appendix D). As shown in Table 4.5, the mean value of C22:1 in FN49-70 was 4.5% compared to 3.4% in Midas. Given the increased C22:1, the content of C18:1 was decreased and was lower than that of Midas, which the mutant line and Midas were 8.1% and 12.1%, respectively. The content of C18:2 and C18:3 in this mutant line was slightly different than the Midas, which was 16.1% compared to 19.5%, and 37.1% compared to 33.4%, respectively. This subtle alteration was observed in C20:1 amount as well, which was 12.6% in FN49-70 and was 13.3% in Midas. This lineage had a higher total VLCFA than Midas, which was 27.3% versus 23.9%.

Table 0.5 The mean value of percentage of C18:1, C18:2, C18:3, C20:1, and C22:1 in FN49-70 lineage.

Name of line	C18:1 (%)	C18:2 (%)	C18:3 (%)	C20:1 (%)	C22:1 (%)	Total VLCFA (%)
FN49-70	8.1±0.8	16.1±0.4	37.1±0.6	12.6±0.4	4.5±0.2	27.3±0.3
Midas	12.1±0.7	19.5±0.4	33.4±0.4	13.3±0.1	3.4±0.2	23.9±0.2

Note. Data were summarized as mean ± standard error. Sample size of FN49-70 was 11. Sample size of Midas was 7.

4.1.5 Selection of lines for further study

Based on the GC data of M_3 seed fatty acid composition, several lines with reduced C20:1 content was selected for further study. M_3 seeds were planted and M_4 seeds collected and subjected to seed fatty acid composition analysis to validate the potential reduced C20:1 phenotype. The first GC results were obtained from M_4 of FN49-38. The average C20:1 content among 29 samples of FN49-38's M_4 seeds was 9.2%. As compared to 9.9% among 9 samples of M_3 seeds, this lineage had a lower level of C20:1, and this trait was heritable. Furthermore, the C20:1 content of their M_5 seeds, which averaged 10.7% over 11 samples, has also been demonstrated to support this conclusion. Despite the increase, the value remained lower than that of Midas, which was 13.3%. Accordingly, the FN49-38 lineage was selected for a genotyping study to identify the cause of the altered seed fatty acid composition. Considering that *FAE1* is one of the most prominent factor controlling C20:1, this gene was the target gene in the subsequent study.

4.2 Pilot study on Genotyping by Sequencing (GBS)

As a pilot study for GBS, genomic DNA was isolated from 96 camelina plants and libraries prepared using the Riptide protocol (appendix A) for paired end sequencing using the Illumina sequencing platform. The information of selected 96 lines are listed in appendix E. Four Midas samples were included in 96 samples and were used for a pilot study to understand the quality of the sequences and to resolve a suitable bioinformatic approach to process the data. Since higher depth of sequencing data can lead to better accuracy of assembly, four Midas files were concatenated. The total number of reads contained in the integrated BAM file was 210,151,363, and the number of mapped reads was 208,247,910, whose coverage of the Midas genome reached 99.1%.

4.2.1 Processing of Illumina sequencing data

A total of 96 samples were included in the first GBS data set, including 4 samples of Midas, DNA from various M_1 and M_2 plants from the FN population and EMS lines associated with a separate project. The sequence data were processed using bioinformatics analysis with the overall objective of identifying SNPs and indels within the lines. The first step of data processing was conducted using the Midas sequencing data which was extracted after demultiplexing the raw data. The objective was to assemble a Midas genome sequence using the published *C. sativa* DH55 genome as a reference sequence. A raw sequencing file containing both reads 1 and 2 was received as paired-end data in the original FASTQ file. When the data was paired-end, both mapping pipeline BWA-MEM (Li, 2013) and *de novo* assembly pipeline Megahit required two input files, one for read 1 and the other for read 2, so one simple code was used “cat SAMPLE.fastq | grep ‘^@.*/1$’ -A 3 –no-group-separator > SAMPLE_r1.fastq” and “cat

SAMPLE.fastq | grep '^@.*/2$' -A 3 –no-group-separator > SAMPLE_r2.fastq" to split one FASTQ file of paired-end reads into two separate files. Having separated the original sequencing file, read 1 and read 2 were used as input files for the pipeline. For the mapping process, a reference file of *C. sativa* DH55 was used and the command "bwa mem ref.fa read1.fq read2.fq > aln.sam" was used. As the output mapping file was unsorted, Samtools was used to convert it to BAM format and sort it. This was accomplished by using the commands "samtools view -bS aln.sam > aln.bam" and "samtools sort aln.bam -o aln.sorted.bam". Mapping produced a file containing the location information for each read yet overlapped reads would not be merged to form contigs. For the *de novo* assembly process, there was no reference file required. Default values were set for all parameters. Contigs would be built from the paired end reads, however, location information was not provided for each contig. In view of the differences between the mapping file and the assembly file, they could be used in a variety of ways. As the pilot study aimed to construct an entire Midas genome, the assembly file was used for the following steps. Because the contigs were generated without location information, they needed to be matched to a reference. This goal was accomplished by employing the pipeline ragtag. The location information was revealed after the scaffolding has been constructed. The reference genome for this experiment was *C. sativa* DH55, and the functional commands were "ragtag.py correct ref.fasta contigs.fasta" and "ragtag.py scaffold ref.fasta corrected_contigs.fasta". The default values for all parameters were left unchanged. Once the pipeline ragtag was executed, its output file contained the location information for the scaffolds that had been constructed. The scaffold file contained 266,453 reads and 188,286 of these reads were mapped to the reference. Therefore, it was used as a reference for bioinformatic analysis of mutant lines in the subsequent data analysis. In conclusion, using Midas sequencing data, a scaffold file covering most of the Midas genome was generated. After the mapping files and genome assembly for Midas was completed, the data was used to extract genome sequence for specific target genes.

In some cases, downstream bioinformatics analyses only required contigs within a specific region. Due to the excessive amount of nonrelevant information, it would be difficult to handle the entire genome file in this scenario, so the mapping file was used for subsequent steps rather than the assembly file. Since the mapping file already contained information regarding the location of each read, Samtools was used to extract the reads that were located in the target region. It was executed with the command "samtools view -b -h input.bam "target region (ex. Chr10:100000-120000)" > output.bam". The extracted reads were then used to construct contigs by software called SeqMan Ultra (www.dnastar.com/software/lasergene/seqman-ultra/) under reference-guided mode. As SeqMan Ultra requires FASTA files as inputs, the extracted reads file was converted from BAM to FASTA by the command "samtools fasta input.bam > output.fasta". The reference used in the current process was the same as the previous process, which used *C. sativa* DH55. The reference would be changed to Midas data produced from this bioinformatics pilot study for the mutant lines. This pilot study generated contigs of target genes related to TAG biosynthesis including *FAE1*, *FAD2/3*, *DGAT1/2*, *PDAT1/2*, *ROD1* and *LPCAT2*.

4.3 Genetic characterization of lines with altered seed fatty acid composition

4.3.1 Screening for mutations in *FAE1* genes

The primary targets for investigation into the genetic basis of the reduced C20:1 phenotype in lineage FN49-38 were the *FAE1* genes. Prior to screening for mutations in the *FAE1* genes in the FN49-38 lineage, the published *C. sativa* genome sequence was examined to identify the 3 genes. Preliminary analysis suggested the presence of assembly or annotation errors in the published genome.

4.3.1.1 Correcting the annotation of *FAE1* genes

C. sativa shares a high degree of similarity with *A. thaliana*, which is widely accepted as a model plant. In arabidopsis, C20:1 is produced from C18:1 through a 2 carbon elongation catalyzed by the elongase complex with FAE1 as the condensing enzyme. *FAE1* belongs to the 3-ketoacyl-CoA gene family. There are 21 members of this gene family in arabidopsis (Joubes et al., 2008) and *FAE1* is referred to as *KCS18*. In the arabidopsis genome, two KCS genes, *KCS18* and *KCS17* are located adjacently as a tandem pair on chromosome 4. These genes do not contain introns. Because camelina is an allohexaploid, there should be three copies of the *CsFAE1* (*CsKCS18*) gene. Synteny with arabidopsis suggests that *CsFAE1* genes may also be arranged in tandem pairs with *CsKCS17*.

Annotations on NCBI data retrieved recently indicated that two versions of *FAE* genes (18-like and 17-like) are adjacent on chromosomes 10 and 12, in a pattern similar to arabidopsis, which supports the hypothesis. As an exception, the gene spanning from 3110775 to 3112479 on chromosome 11 is annotated as 3-ketoacyl-CoA synthase 18-like (NCBI GeneID: *LOC104721342*), whereas the gene spanning from 3107019 to 3108726 is also annotated as 3-ketoacyl-CoA synthase 18-like (NCBI GeneID: *LOC104721341*). There is almost no difference in the sequence information between these two genes. The same discrepancy is observed in camelina genome assembly available from the Crucifer Genome Initiative (www.cruciferseq.ca). Here 3 genes are annotated as 3-ketoacyl-CoA synthase 18-like, *Csa10g007610*, *Csa11g007400* and *Csa12g009060* (Figure 4.9). On chromosome 10 the FAE1 gene is accompanied by a *KCS17* homologue in a tandem pair. On chromosomes 11 and 12, genes annotated as *FAE1* appear unusually large, contain introns and are not accompanied by *KCS17* homologues (Figure 4.9). For example, *Csa11g007400* is one gene which contains two exons (Figure 4.10). Those two exons share identical base information and annotated as coding sequence of 3-ketoacyl-CoA

synthase 18-like *FAE* gene. It appears that based on the hypothesis, the annotation and the base information regarding the *FAE* genes on chromosome 11 and 12 could be incorrect.

To resolve this issue, long-read sequencing data obtained from ONT sequencing was used to determine whether the current annotation for *FAE* on chromosome 11 is accurate. Using the *CsFAE1A* template as a search guide, one single 10684 bp long read was extracted from ONT data sequenced from FN49-38-1-3-12 to cover the target region. In the subsequent alignment process, this long read served as a reference. FASTA files of two 3-ketoacyl-CoA synthase 18-like genes on chromosome 11 obtained from NCBI were aligned to the reference. As a guide in the alignment project, NCBI's coding sequence region for the 3-ketoacyl-CoA synthase 17-like *FAE* gene on chromosomes 10 and 12 was used to determine where the gene is actually located on chromosome 11. In accordance with the alignment result, *LOC104721341* and *LOC104721342* were nearly 100% identical and aligned to the same region of the reference, whereas the coding sequences of 3-ketoacyl-CoA synthase 17-like *FAE* genes from chromosomes 10 and 12 were aligned to another region of the reference. The consequence of this finding indicated that there is another region where the 3-ketoacyl-CoA synthase 17-like *FAE* gene should be located, as opposed to the current region on NCBI. Since the 3-ketoacyl-CoA synthase 17-like *FAE* gene is located in another region, the base information on NCBI appears to be incorrect. Using our ONT data, the 3-ketoacyl-CoA synthase 17-like FAE gene was 3582 bp away from the 3-ketoacyl-CoA synthase 18-like *FAE* gene from the 5' direction, rather than 2239 bp according to the NCBI database.

The same approach was used to verify the accuracy of the annotation provided by the Crucifer Genome Initiative for the *FAE* gene on chromosome 12. In their analysis, gene *Csa12g009060* was identified as a single *FAE* gene containing three exons. However, according to our hypothesis, there should be two *KCS* genes located in that region corresponding to *CsKCS18* (*CsFAE1*) and *CsKCS17*. For the purpose of identifying where the three exons are located, the three exons were aligned with the extracted ONT single long read. Consequently, exon 1 was in the region where 3-ketoacyl-CoA synthase 18-like *FAE* was located. The exon 2 and 3 were in a region where the 3-ketoacyl-CoA synthase 17-like was found, and these two

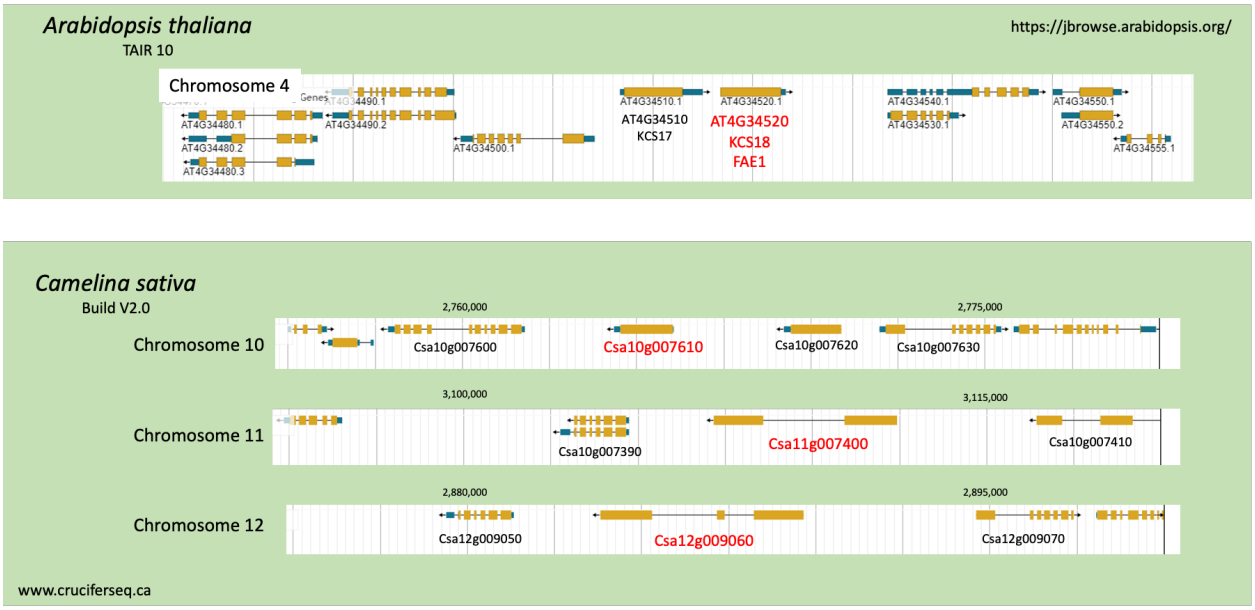

Figure 4.9 Present reported *FAE1* gene predictions in the arabidopsis genome and camelina genome by Crucifer Genome Initiative (www.cruciferseq.ca).

sequences overlapped (Figure 4.10). Considering this result, it is evident that the pattern and annotation reported by Crucifer Genome Initiative are inaccurate. Rather, the actual pattern should be consistent with the hypothesis that two versions of the *FAE* gene are located adjacently on chromosome 12, like that found in arabidopsis.

In view of these results, the arrangement of the *FAE* gene on chromosomes 10, 11, and 12 of camelina resembles arabidopsis to be composed of tandem pairs of KCS genes, one 3-ketoacyl-

CoA synthase 18-like, and the other 3-ketoacyl-CoA synthase 17-like. In accordance with this arrangement, the *FAE* gene annotation on chromosomes 11 and 12 should be revised.

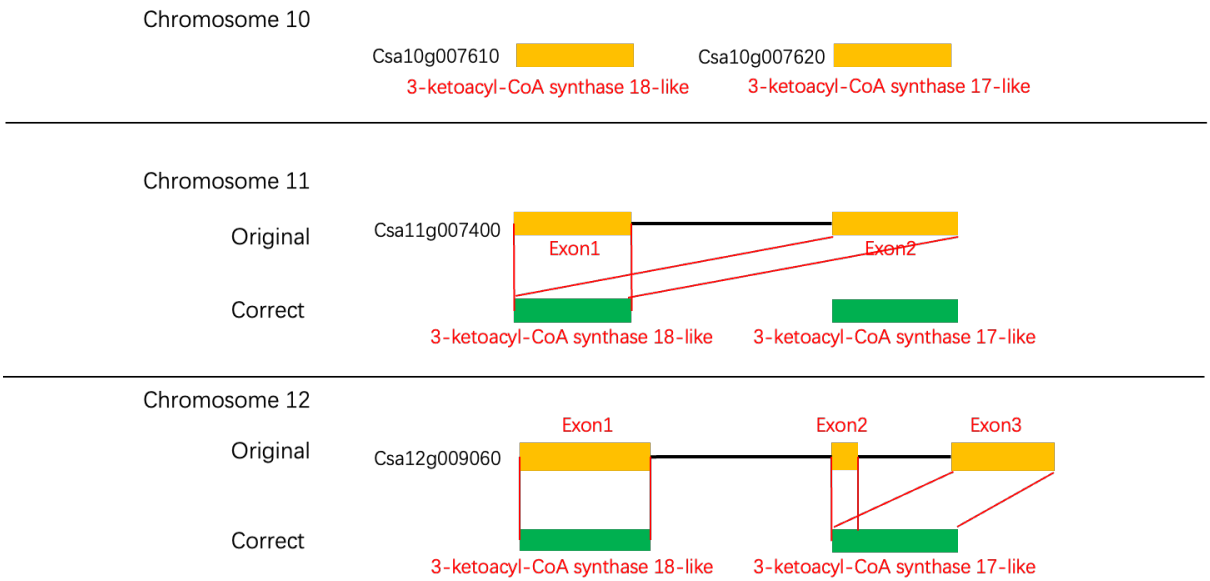

Figure 4.10 Scheme of original pattern and annotation of *FAE* and the pattern and annotation after correction. Chromosome 10: two versions of FAE gene displayed in correct pattern with proper annotation. Chromosome 11: the original pattern of FAE gene was shown in one whole gene annotated with two exons. After correction (labeled in green), exon1 is 3-ketoacyl-CoA synthase 18-like FAE gene, and exon2 is 18-like FAE gene as well. The 3-ketoacyl-CoA synthase 17-like FAE gene on chromosome 11 was added based on our ONT data. Chromosome 12: the original pattern of FAE gene was shown in one whole gene annotated with three exons. After correction (labeled in green), exon1 is 3-ketoacyl-CoA synthase 18-like FAE gene, and exoon2 and exon3 shared identical base information were combined into one 3-ketoacyl-CoA synthase 17-like FAE gene. Original FAE gene organization was retrieved from http://cruciferseq.ca/.

4.3.2.2 Analysis of GBS data set 1 and identification of a *fae1b* deletion

The FN49-38 lineage produced plants with lower C20:1 level in their seed oil (9.5% - 10.2%) than the Midas (13.3%). It was hypothesized that this alteration of seed oil phenotype could be due to the decreased expression of a fatty acid elongase1 (*FAE1*) gene. The sequence

data set contained genomic sequence information for 8 M_2 plants of the FN49-38 lineage. Therefore, the *FAE1* gene sequences in the mutant lines was extracted and compared to that of the Midas reference. In camelina, three *CsFAE1* genes were found in three different chromosomes (chromosome 10, 11, and 12) (Hutcheon et al., 2010). Comparison of these *CsFAE1* sequences from these sibling lines with those in Midas from the Illumina sequencing showed that there were deletions in the *CsFAE1* gene on chromosome 10 in these mutant lines. The major deletion was about 1,220 bp in length spanning from the 5' UTR to the front half of the ORF of *FAE1B* including the start codon (ATG) in FN49-38-1, FN49-38-2, FN49-38-5, and FN49-38-6 lines. However, the deletions were not completely identical in all FN49-38 offspring lines. As shown in the Figure 4.11, the deletion was shorter in FN49-38-3, FN49-38-7, and FN49-38-9 with a slightly different starting site. The deletion was not consecutive in those mutant line as well. The depth of the sequencing data was displayed in Figure 4.12. The average depth of the target region that covering the deletion (chromosome 10:2,760,000-2,770,000) among these 8 samples was 7-fold (number of reads overlapping with one position). The FN49-38-3 had the best sequencing depth, which was around 12 reads, and FN49-38-8 had the lowest sequencing depth, which was only around 3-fold. Contigs generated from sequencing data with a depth below 4-fold may be less accurate (Jiang et al. 2019).

To confirm the deletion identified from the Illumina sequencing results, genomic DNA from FN49-38-1-3-12 and FN49-38-4-7 were sequenced by the ONT. The depth of the target region was 171x and 4.2x, respectively. The sequence alignment of *FAE1* genes on chromosome 10 from the mutant lines and reference exhibited a similar pattern, i.e., a deletion of about 1,220 bp spanning from the 5' UTR to the front half of the ORF of the gene (Figure 4.13). This result validated the occurrence of the deletion in the *FAE1B* in FN49-38 mutant lineage.

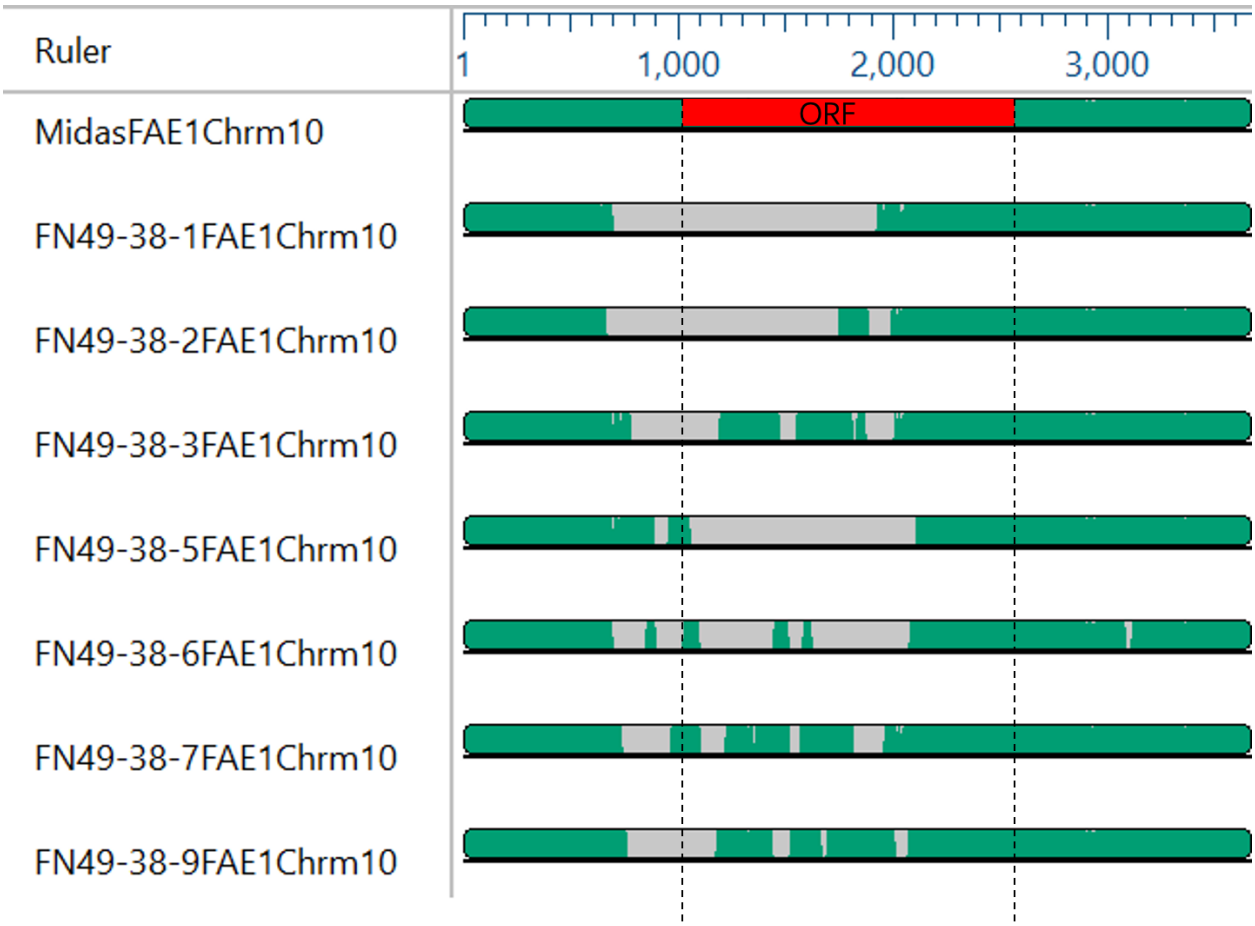

Figure 4.11 Alignment of *FAE1* gene on chromosome 10 of Midas (wild type) and FN49-38-1, -2, -3, -5, -6, -7, -9. The green region indicated the DNA sequence, where was aligned to each other, and grey region represented the missing sequences. The red region represented the ORF of *FAE1B*. The algorithm of alignment applied was MAAFT. The figure was generated by software DNASTAR. The green regions in the middle of the grey part means the deletion was not identical and consecutive in those mutant lines.

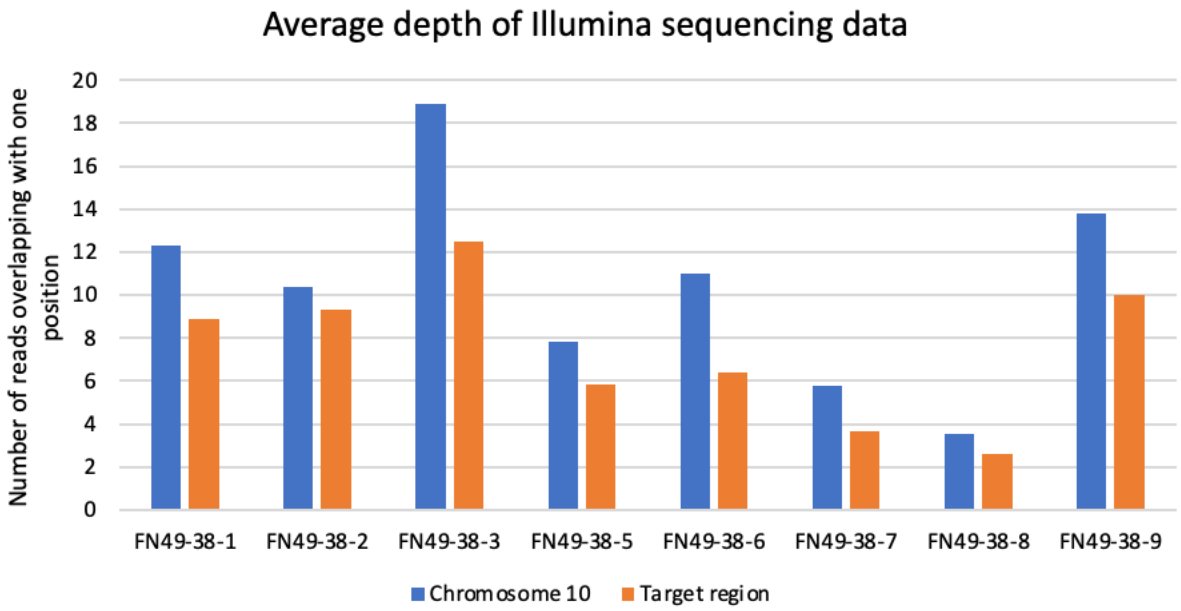

Figure 4.12 Average depth of Illumina sequencing data of FN49-38 siblings. Blue bar: average depth of reads in chromosome10. Orange bar: average depth of target region. The target region was 2,760,000-2,770,000 bp from the beginning of chromosome 10, which involved the *FAE1B*. The depth was calculated by Samtools. Depth lower than 4 can cause higher errors during assembly.

Figure 4.13 Sequence alignments with bases information of identical region of *FAE1B* genes from Midas (wild type), FN49-38-1 (sequenced by Illumina technology), and FN49-38-1-3-12 (sequenced by ONT). This alignment was accomplished through EMBL-EBI Multiple Sequence Alignment (MSA) (https://www.ebi.ac.uk/Tools/msa/). The algorithm of alignment applied was MAAFT. The base information in this figure was covered from part of the 5' UTR of *FAE1B* and part of the ORF of *FAE1B*. The base pair fragments observed in the deletion region may not reflect those bases that were aligned exactly there, and the alignment algorithm likely generated this misrepresentation. These bases should be located at either end of the deletion and should be mismatched to the Midas sequence and coding region. Also, they could be the result of incorrect DNA repair.

4.4 Characterization of a *fae1b* deletion line

4.4.1 Confirmation of the *fae1b* deletion by cloning and sequencing *fae1b* genes

For cloning, three lines were selected, FN49-38-1-2, -3 and -11, in which FN49-38-1-3 and -11 were homozygous mutants (Figure 4.16), and FN49-38-1-2 was highly possible to be mutant homozygous. The *FAE1B* as the target region was amplified by PCR and extracted for cloning, which has been depicted in Section 3.2.2.1. Two colonies for each line were picked for Sanger sequencing, resulting in a total of 10 colonies sequenced. The sequencing was implemented from both ends using two primer sets, which were an initial sequencing primer set T3/T7 designed specifically for the TOPO4 vector, and an internal primer set B542/B560 designed to fill in the gaps left by the T3/T7 primer set. Only colony number two from the FN49-38-1-2 generated the expected result with a contig length of 1751 bp, which was in accordance with the sequence information of the mutant *fae1b* gene carrying the deletion (Figure 4.14). The rest of the colonies sent for Sanger sequencing still had gaps of various sizes that cannot be filled by the sequenced data. As a result, this experiment confirmed the base information that had been generated from both Illumina and ONT sequencing around the deletion region for *FAE1B*.

Figure 4.14 Sequence alignments with bases information of identical regions of *FAE1B* genes from Midas (wild type), FN49-38-1-3-12 (sequenced by ONT), and cloned *FAE1B* from FN49-38-1-2. This alignment was accomplished through EMBL-EBI Multiple Sequence Alignment (MSA) (https://www.ebi.ac.uk/Tools/msa/). The algorithm of alignment applied was MAAFT. The base information in this figure was covered from part of the 5' UTR of *FAE1B* and part of the ORF of *FAE1B*. The base pair fragments observed in the deletion region may not reflect those bases that were aligned exactly there, and the alignment algorithm likely generated this misrepresentation. These bases should be located at either end of the deletion and should be mismatched to the Midas sequence and coding region. Also, they could be the result of incorrect DNA repair.

4.4.2 Study of fatty acid composition during seed development

FN49-38 lines contained a reduced amount of gondoic acid (C20:1 n-9). Accordingly, the offspring of FN49-38-1-3-12 and FN49-38-1-11 (mutant homozygous), and Midas were used to monitor the fatty acid composition during seed development. Figure 4.15 shows the changes in fatty acid composition of Midas, FN49-38-1-3-12 lines, and FN39-38-1-11 lines at 4 developmental stages, which were 22, 30, 38, and 46 days after flowering. At the 22nd day, a

small amount of gondoic acid was produced in all the lines including Midas. For the mutant lines, gondoic acid accounted for around 4.0% on average, which was slightly lower than the wild type (4.5% on average). At the 30th day and 38th day, the level of gondoic acid in the mutant lines was less than that of the Midas. At the last developmental stage, 46th day, gondoic acid in Midas reached an average 11.5% as compared to 9.5% and 10.5% in FN49-38-1-3-12 lines and FN49-38-1-11 lines, respectively.

Another observation was the changes in the amount of C18:3 during the seed development. On the 22nd day, the content of C18:3 of all lines was 36.6% on average. This amount was achieved at 40.1% on the 30th day. But on the 38th day, there was merely 33.7% on average, and the amount dropped to 33.1% on average on the 46th day. The overall shape of the proportion of C18:3 is an increasing trend followed by a downward trend, and the alteration of Midas's data reflects this trend well.

Since C18:2 is the substrate of C18:3, this could explain why the trend was opposite that of C18:3. On the 22nd day, the mean of the proportion of C18:2 was 26.8% of all measured lines. But it dropped to 20.6% after 8 days. The number was increased to 22.7% on the on the 38th day, and it seemed to become steady after 38 days because the proportion was 21.6% on average on the 46th day.

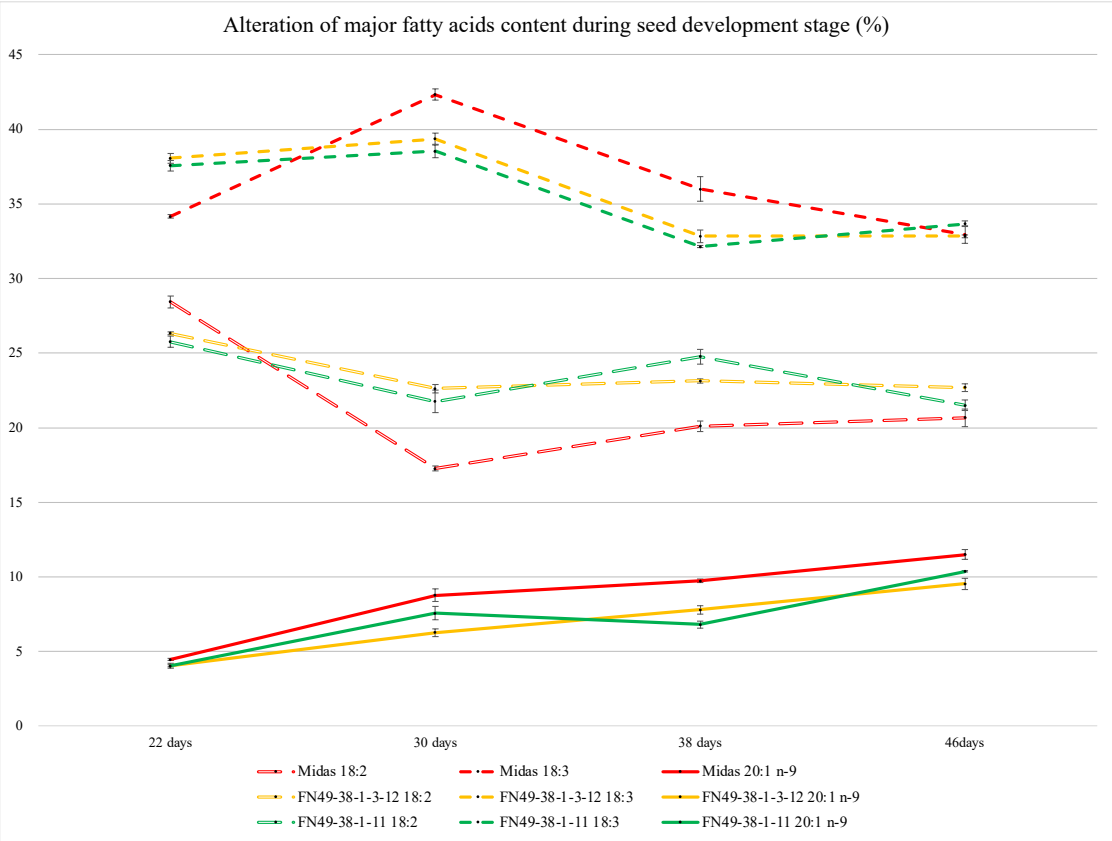

Figure 4.15 Comparison of the % content of the major fatty acids during seed development in Midas (wild type), FN49-38-1-3-12-, and FN49-38-1-11- lines. The observation days were 22, 30, 38, and 46 days after flowering. Only linoleic acid, α-linolenic acid, and gondoic acid composition were exhibited in this figure. The fatty acid composition was determined by the GC on FAMEs derived from total seed lipids by acid-catalyzed transmethylation. The data was shown in percentage of total seed fatty acids.

4.4.3 Development of PCR primers for detection of deletion lines

The objective of this work was to develop primers to enable plants to be screened by PCR for presence of the *fae1b* deletion allele, and to genotype them as homozygous, heterozygous, or null lines.

4.4.3.1 Design and optimization of primers and PCR conditions

FN49-38-1-11 and its sibling lines were selected for PCR genotyping to examine whether the deletion occurred at the region of the *FAE1* gene. The primer design was completed after the mutated gene was aligned with the Midas genome. It was expected that the forward primer would be located before the deletion region and the reverse primer would be located behind the deletion region to completely cover the deletion region. Given that the deletion spanned between the 5'UTR and the first half of the *FAE1* ORF, the desired primer set would cover the whole ORF, as well as some of the 5'UTR. Furthermore, it was anticipated that the primer set would bind only to *FAE1* on chromosome 10. During several attempts, a primer set consisting of the forward primer B542 (CACATGTATCGTATTGAGCACCACTC), which was 1402 bases 5' upstream of the FAE coding sequencing region, and the reverse primer B538 (GACTGCTCCAACTCTCTGACACATATC), which was 500 bases after the coding sequencing region, was found to be effective on the FN49-38 lineage. This primer set required an annealing temperature of 50 ˚C and an extension time of 3 minutes. Figure 4.16 shows the PCR products produced after 30 working cycles under the optimal condition. A fragment with about 3 kb was amplified from wild type (Midas) while a fragment with about 2 kb was amplified from all four sibling lines from FN49-38 including FN49-38-1-11. This result indicated that about 1.2 kb was deleted from the *FAE1* gene in the mutant lines. This observation confirmed the sequencing data that a deletion indeed occurred in the *FAE1* gene of FN49-38- lines and enabled genotyping of the plants as described in section 4.4.4.

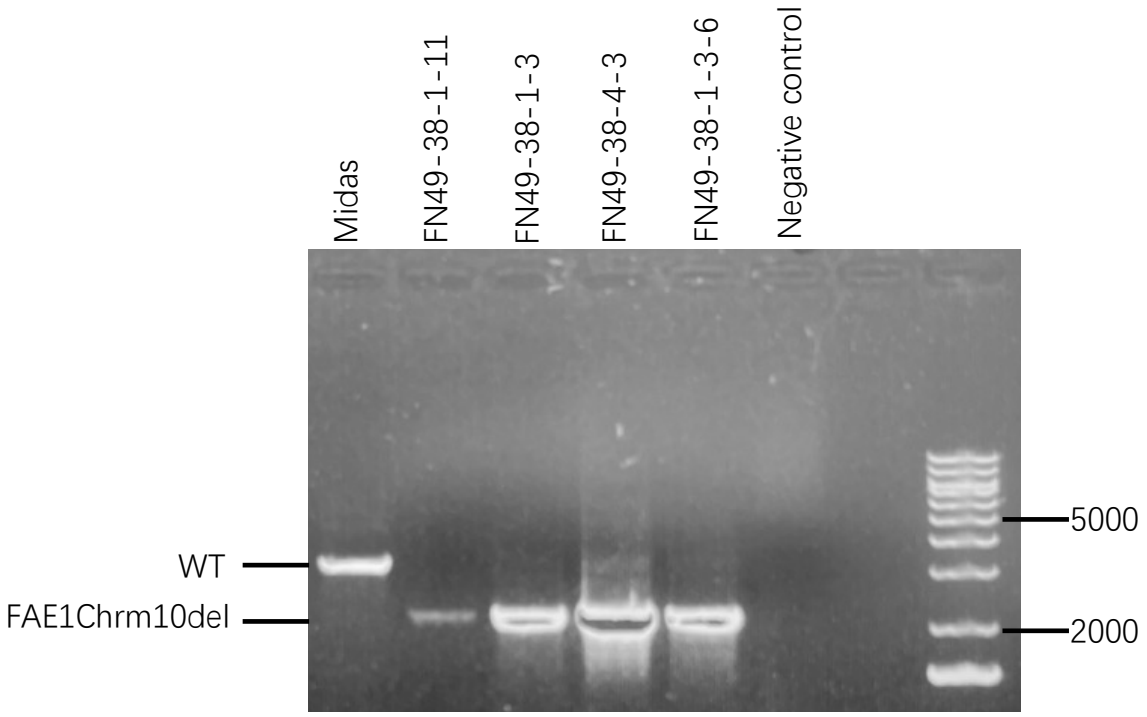

Figure 4.16 A gel image of *CsFAE1* amplification in Midas and mutant lines. PCR was conducted using genomic DNA as temple. A pair of primers flanking the deletion from the 5' UTR to the first half of the *FAE1B*-ORF was used. The PCR products were analyzed by 1% agarose gel electrophoresis followed by SYBR Safe DNA Gel Stain. Lane 1, the wild type (Midas); lane 2, FN49-38-1-11; lane 3, FN49-38-1-3; lane 4, FN49-38-4-3; lane 5, FN49-38-1-3-6; lane 6, negative control (no DNA template); lane 8, 1 kb+ lambda DNA (marker). WT: wild type. FAE1Chrm10del: FAE1 chromosome 10 deletion.

4.4.4 Confirmation of linkage of reduced VLCFA trait and the *fae1b* chromosomal deletion

Plants of camelina lineage FN49-38 showed a phenotype of reduced C20:1 in their seed oil. Through sequencing, a deletion in the *FAE1B* gene was detected. To confirm that the phenotype was linked to the deletion genotype, a backcrossing study was initiated, as described in Section 3.3, and phenotype and genotype tracked through F_1 and F_2 generations.

After backcrossing, a total of 19 backcrossing lineages were generated. The backcrossing lines were labeled from FNCX2 to FNCX 20 with the mutant lines as pollen acceptors and Midas (wild type) as pollen donor. Among them, FNCX5 (FN49-38-4-5 x Midas), FNCX 9 (FN49-38-4-7 x Midas), FNCX13 (FN49-38-1-3-7 x Midas), and FNCX16 (FN49-38-1-3-10 x Midas) were chosen for further analysis. As shown in Figure 4.16, plant FN49-38-1-3 was shown to be a homozygous mutant by PCR. As progeny of this plants were used for the FNCX13 and FNCX16 crosses, true F_1 seeds from these can be considered to have a Bb genotype (where B and b refer to the *FAE1B* and mutant *fae1b* alleles respectively). F_1 seeds from the 4 chosen FNCX events were planted and DNA extracted from 3 F_1 plants to assess genotype by PCR. As expected, plants FNCX16-10, FNCX16-11 and FNCX16-12 showed 2 bands using primers specific for the *fae1b* deletion (Figure 4.19) indicating that these plans were hemizygous (Bb). Samples of F_2 seeds collected from all plants were analyzed by GC for seed fatty acid composition. For this, 30 seeds were pooled from each plant. Due to poor growing conditions, some plants produced poor quality seeds, and these were excluded for the survey. A total of 23 F_2 seed samples were analysed (detailed fatty acid composition is shown in appendix D). The average of C20:1 content (%) of individuals from each crossing event is shown in Figure 4.17 and is compared to Midas. Reduced C20:1 levels were observed with the average of the 7 F_2 seed samples analysed from the FNCX16 cross for example being 10.7% ± 0.4% compared to 13.3% ± 0.12% for Midas. For a more detailed analysis of segregating lines, F_2 seed from plants FNCX16-10, FNCX16-11 and FNCX16-12 were planted and resulting plants were genotyped by PCR. As shown in Figure 4.20, segregation was occurring with WT plants (e.g., FNCX16-11-2, WT PCR product only), and lines heterozygous (e.g., FNCX16-11-4, WT and deletion PCR product bands) and homozygous (e.g., FNCX16-11-1, only deletion PCR product band) for the *fae1b* (b) deletion allele being observed.

A GC analysis was conducted on F_3 seeds produced from FNCX16-11 and FNCX16-10. The detailed fatty acid composition is shown in appendix D. For the lineage FNCX16-11, the results

of FNCX16-11-4, -5 and -7 were excluded from the analysis due to poor seed quality. The amount of C20:1 in FNCX16-11-1, -3, and -6, which were mutant homozygous, was 11.0%, 11.0%, and 10.4%, respectively, which was lower than the amounts in FNCX16-11-2, and -8, which were wild-type homozygous. Based on the comparison of the fatty acid compositions among wild type, heterozygous and mutant homozygous individuals of the lineage FNCX16-10, as shown in Figure 4.18, it was found that wild type had the highest concentrations of C20:1 and C22:1 while the lowest concentration of C18:3 was found in the wild type, which was 13.4%, 3.8%, and 32.8%, respectively. The homozygous mutant contained the lowest amounts of C20:1 and C22:1, but the highest amounts of C18:3, which were respectively 10.5%, 2.4%, and 37.6%. These three fatty acids were present in heterozygotes at levels between wild-type and mutant.

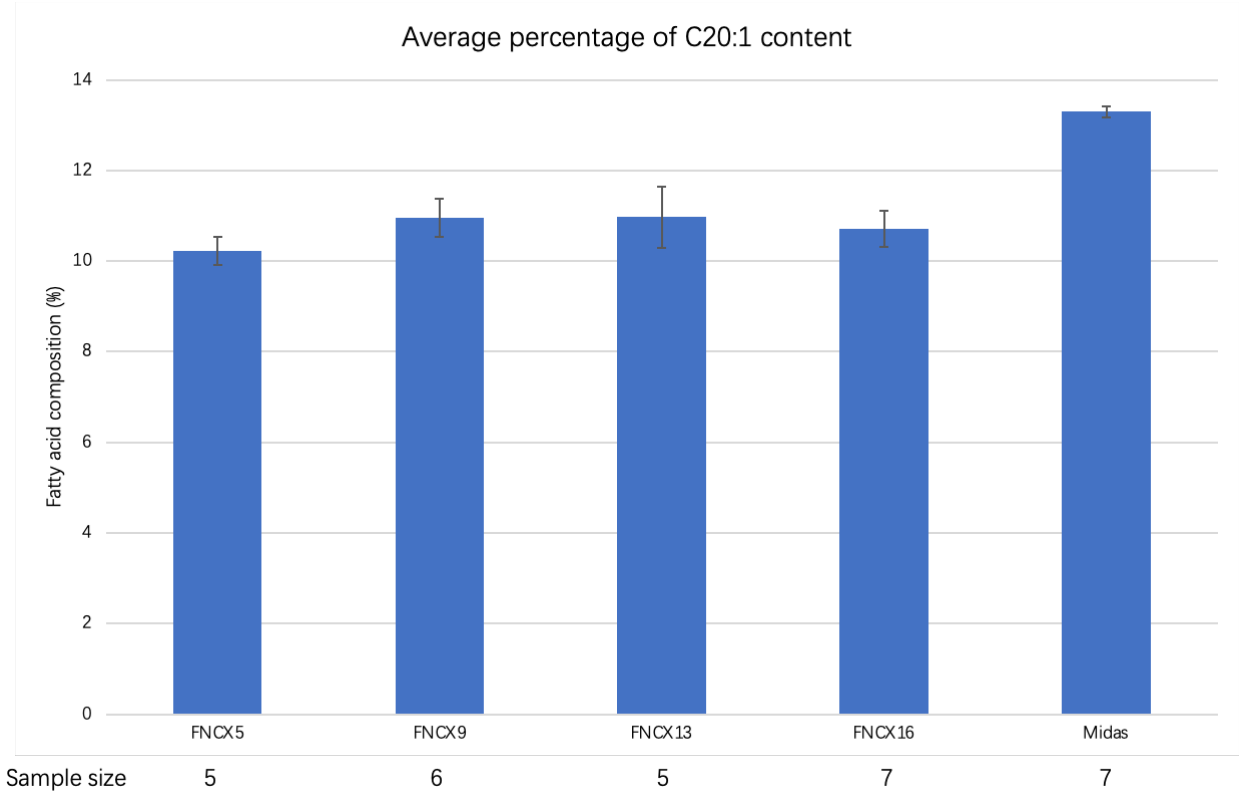

Figure 4.17 Average percentage of gondoic acid (C20:1) content in backcrossing F_2 seeds and Midas seeds.

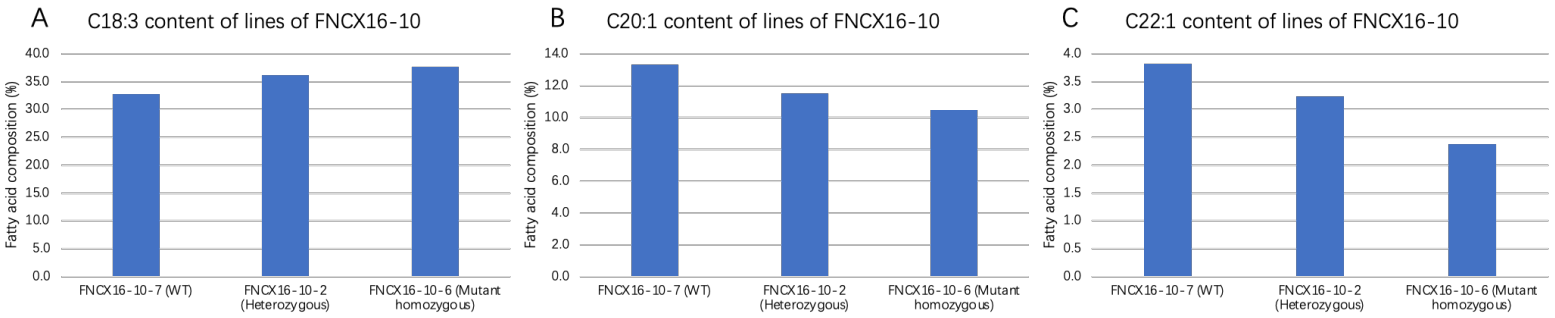

Figure 4.18 Comparison of fatty acid composition obtained from F_3 seeds of FNCX16-10 via GC among wild type, heterozygous, and mutant homozygous. A: C18:3 content determined from FNCX16-10-2, -6, and -7. B: C20:1 content determined from FNCX16-10-2, -6, and -7. C: C22:1 content determined from FNCX16-10-2, -6, and -7.

Overall, it appeared from the genotype results and GC results of FNCX16's progenies that the mutant homozygotes were associated with the reduced C20:1, as the GC results of the F_3 seeds agreed with the genotype of F_2 plants (Figure 4.18). The linkage between the deletion and the reduced C20:1 phenotype was subsequently identified.

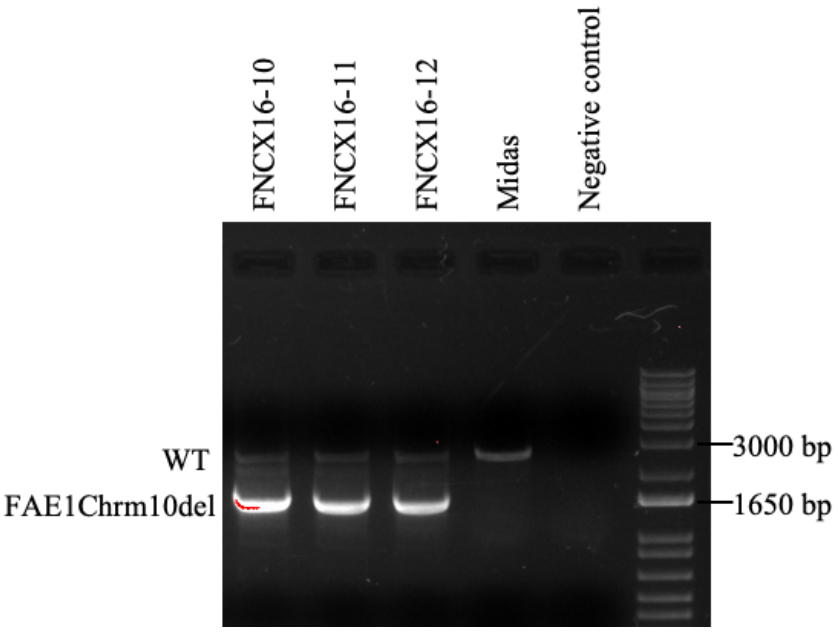

Figure 4.19 A gel image of *FAE1B* amplification from F_1 plants FNCX16-10, -11, -12, and Midas. PCR was conducted using genomic DNA from young leaf of F1 plants as template. A pair of primers flanking the deletion from the 5' UTR to the first half of the FAE1B-ORF was used. The PCR reaction total was 20 µl for each tube. The PCR products were analyzed by 1% agarose gel electrophoresis followed by SYBR Safe DNA Gel Stain. Lane 1 to lane 3 was FNCX16-11-10 to FNCX16-11-12; lane 4, Midas; lane 5, negative control (no DNA template added); lane 6, 1kb+ lambda DNA (marker). Twenty microlitre of each group was used for the electrophoresis. WT: wild type. FAE1Chrm10del: FAE1 chromosome 10 deletion.

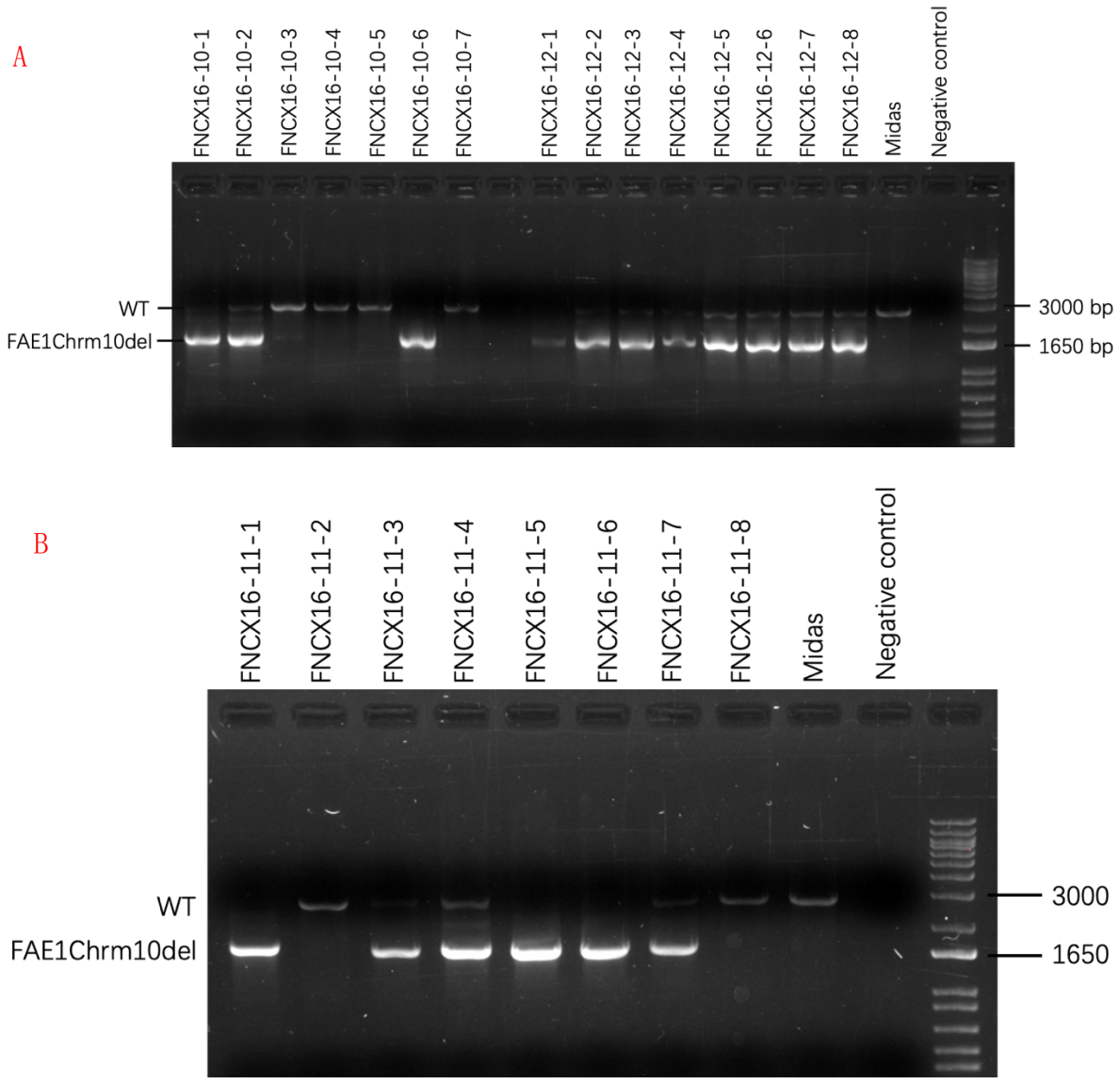

Figure 4.20 Gel images of *CsFAE1* amplification in FNCX16 offspring and Midas. PCR was conducted using genomic DNA from young leaf of F_2 plants as template. A pair of primers flanking the deletion from the 5' UTR to the first half of the FAE1B-ORF was used. The PCR reaction total was 20 µl for each tube. The PCR products were analyzed by 1% agarose gel electrophoresis followed by SYBR Safe DNA Gel Stain. Twenty microlitre of each group was used for the electrophoresis. WT: wild type. FAE1Chrm10del: FAE1 chromosome 10 deletion. A: lane 1 to lane 7 was FNCX16-10-1 to FNCX16-10-7; lane 9 to lane 16 was FNCX16-12-1 to FNCX16-12-8; lane 17, Midas; lane 18, negative control (no DNA template added); lane19, 1kb+ lambda DNA (marker) B: lane 1 to lane 8 was FNCX16-11-1 to FNCX16-11-8; lane 9, Midas; lane 10, negative control (no DNA template added); lane 11, 1kb+ lambda DNA (marker). Same Midas's DNA template used in A and B.

4.4.5 Potential of primers for screening DNA pools

Phenotyping the FN population uncovered several lineages that showed reduced levels of C20:1 in their seed oil (Section 4.1.4). As the genetic basis of these differences was unknown it was decided to attempt PCR screening to determine if any of these additional lines also carried deletions in the *FAE1B* gene. With the hope of producing a stronger WT band, a revised primer set was created with the forward primer changed to B523 (GATCAGTGTGCTCATGGTGTTCCAG) and the reverse primer remaining at B538 (GACTGCTCCAACTCTCTGACACATATC). Compared to primer B542 (CACATGTATCGTATTGAGCACCACTC), this primer B523 was 917 bases ahead of the FAE coding sequence, which was closer to the coding sequence than primer B542. Therefore, this revised primer set could produce WT PCR product with 2984 bp, which should be 525 bp smaller than WT with forward primer B542. The PCR conditions were changed in response to the change of the forward primer. The ideal PCR conditions for the revised primer set were 55° C for annealing, 3 minutes for extension, and 30 cycles of amplification.

Several mutant lineages with lower C20:1 content were selected and DNA from multiple plants was pooled to reduce the number of reactions required for screening, and to test the utility of this approach for potentially broader screening of the population. The information of chosen lines was shown in Table 4.6 A straightforward pooling strategy was applied, in which 8 pools were built, and 6 pools contained 8 individual lines, 1 pool contained 7 individual lines, and 1 pool contained 6 individual lines. The Figure 4.22 shows the result after PCR amplification and gel electrophoresis. Pool 1 and Pool 3 gave two bands. Their lower bands were about 1.6 kb, and their top bands were around 2.8 kb, which was identical to the band of Midas (wild type). This observation indicated that there were mutant lines potentially carrying a *fae1b* deletion either as heterozygotes or mixed with the mutant lines in the pool did not have mutations. For the rest of pools, pool 2, pool 4, pool 5, pool 6, pool 7, and pool 8, all mutant lines in these pools contained

deletions on the region related to *FAE1B*, because the PCR products of these pools merely gave one lower band. The difference in band size was approximately 1.2 kb represented the size of the deletion. This band was a similar size in all pools even though the lines composing the pools were not the same. Moreover, the apparent deletion was similar in size to that found in lineage FN49-38. As Illumina sequencing data was available from plants of the FN49-312 lineage this was examined and revealed the presence of a potential deletion like that of lineage FN49-38 (Figure 4.11 and Figure 4.23).

The complexity of DNA pooling can affect the efficiency of PCR screening. To further explore DNA pooling and PCR for the detection of deletion lines, 2 test pools were built for measuring pool depth and complexity. For the 10:1 test pool, genomic DNA from 10 FN lines of the FN49-52 lineage (FN49-52-5-1 to -10) was used. Similarly, the 20:1 test pool contained 20 FN lines with no deletion (FN49-52-5-1 to -10 plus from FN49-35-10-1 to -10). The FN49-52 and FN49-35 lineages were plants with a wild type seed fatty acid composition. Each pool also contained DNA from FN42-353-1, a line that gave a band when screened for PCR for a potential deletion in the *FAE1B* gene. Based on the result of two test pools, both the PCR products generated from 10:1 and 20:1 gave two bands. The band size was same as anticipation, which the top band was around 2.8 kb and bottom band was around 1.6 kb. This observation was consistent with the result of the 8 PCR screening pools. In conclusion, the results of two test pools indicated that even though the complexity of the sequencing was 20:1, the specific primer set was able to properly amplify the sequence containing the deletion in the pool that contained multiple DNA templates.

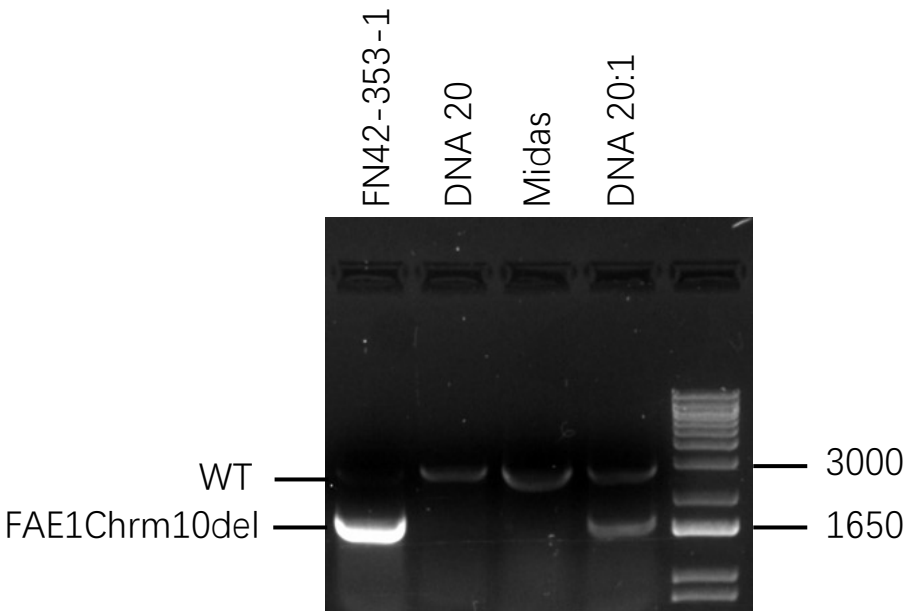

Figure 4.21 A gel image of PCR *FAE1* amplification in Midas, FN42-353-1, and DNA in test pool 20:1. PCR conducted using genomic DNA as temple. A pair of primers flanking the deletion from the 5' UTR to the first half of the FAE1B-ORF was used. The PCR reaction total was 20 µl for each tube. The PCR products were analyzed by 1% agarose gel electrophoresis followed by SYBR Safe DNA Gel Stain. Lane 1, FN42-353-1; lane 2, DNA 20 (the pool that contained 20 mutant lines with no deletion); lane 3, Midas (wild type); lane 4, DNA 20:1 (the pool that contained 20 mutant lines with no deletion plus one mutant line carrying a potential deletion); lane 5, 1kb+ lambda DNA (marker). Twenty microlitre of each group was used for the electrophoresis. WT: wild type. FAE1Chrm10del: FAE1 chromosome 10 deletion. The 10:1 complexity was not shown in this figure. Forward primer: B523 (GATCAGTGTGCTCATGGTGTTCCAG); reverse primer: B538 (GACTGCTCCAACTCTCTGACACATATC).

Table 0.6 Information of mutant lines grouped in 8 different DNA pools for screening deletion around *FAE1B* region.

Individual line*	Pool 1	Pool 2	Pool 3	Pool 4	Pool 5	Pool 6	Pool 7	Pool 8
1	FN49-72-3	FN42-97-2	FN42-119-1	FN49-620-8	FN42-50-4	FN42-97-3	FN49-312-7-1	FN49-312-1-1
2	FN49-72-5	FN42-97-5	FN42-119-2	FN49-620-9	FN49-438-1	FN42-97-4	FN49-312-7-2	FN49-312-1-2
3	FN42-119-4	FN49-375-8-4	FN42-119-3	FN42-353-1	FN49-438-2	FN42-5-3	FN49-312-7-3	FN49-312-1-3
4	FN42-81-2	FN49-375-8-6	FN42-119-5	FN42-353-2	FN49-506-1	FN42-5-4	FN49-312-7-4	FN49-312-1-4
5	FN42-81-5	FN49-438-3	FN49-213-2	FN42-353-4	FN49-438-4	FN42-81-1	FN49-312-7-5	FN49-312-1-5
6	FN42-5-1	FN49-353-5	FN49-213-3	FN49-72-2	FN49-506-3	FN42-81-3	FN49-312-7-6	FN49-312-1-6
7	FN42-5-2	FN49-213-1	FN49-213-4	FN42-50-1	FN49-506-4	FN42-81-4		FN49-312-1-7
8	FN42-5-5	FN42-353-2	FN49-72-4	FN42-50-2	FN42-97-1			FN49-312-1-8

* Mutant lines were selected based on GC results. It was only the lines with a C20:1 content of less than 11% that were selected.

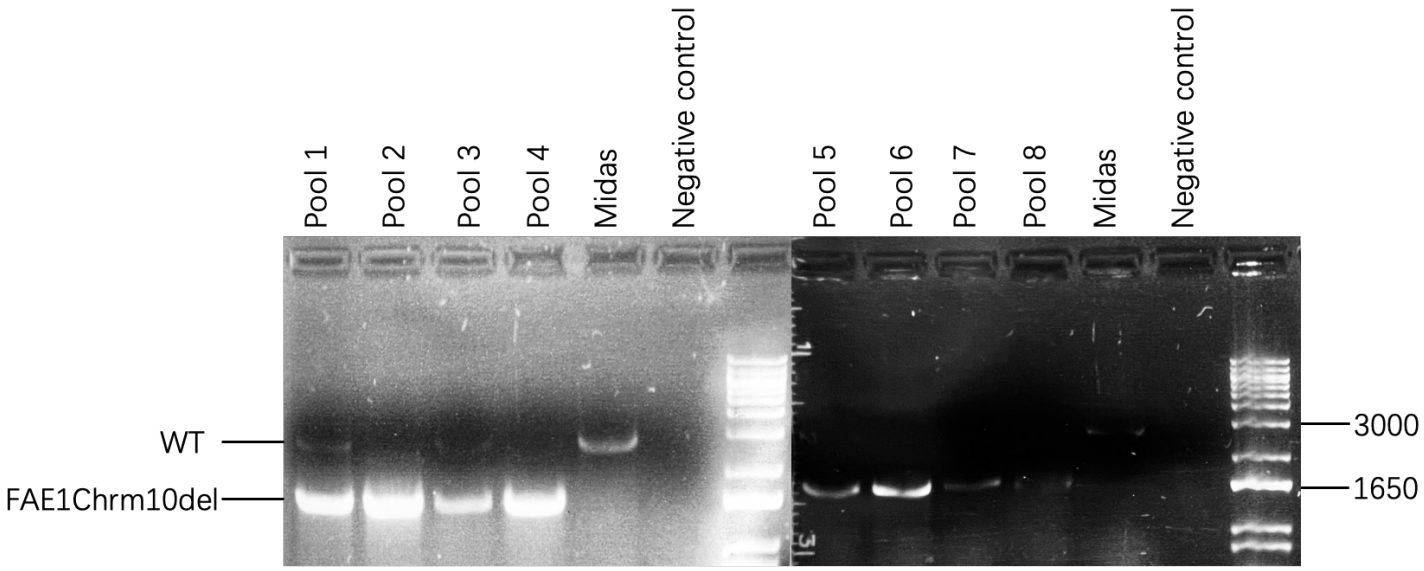

Figure 4.22 Gel images of PCR *FAE1* amplification in Midas and 8 DNA pools contained mutant lines. PCR conducted using genomic DNA as temple. A pair of primers flanking the deletion from the 5' UTR to the first half of the FAE1B-ORF was used. The PCR reaction total was 20 µl for each tube. The PCR products were analyzed by 1% agarose gel electrophoresis followed by SYBR Safe DNA Gel Stain. All PCR products were generated in same single run. Twenty microlitre of each line including Midas was used for the electrophoresis. WT: wild type. FAE1Chrm10del: FAE1 chromosome 10 deletion. Forward primer: B523 (GATCAGTGTGCTCATGGTGTTCCAG); reverse primer: B538 (GACTGCTCCAACTCTCTGACACATATC).

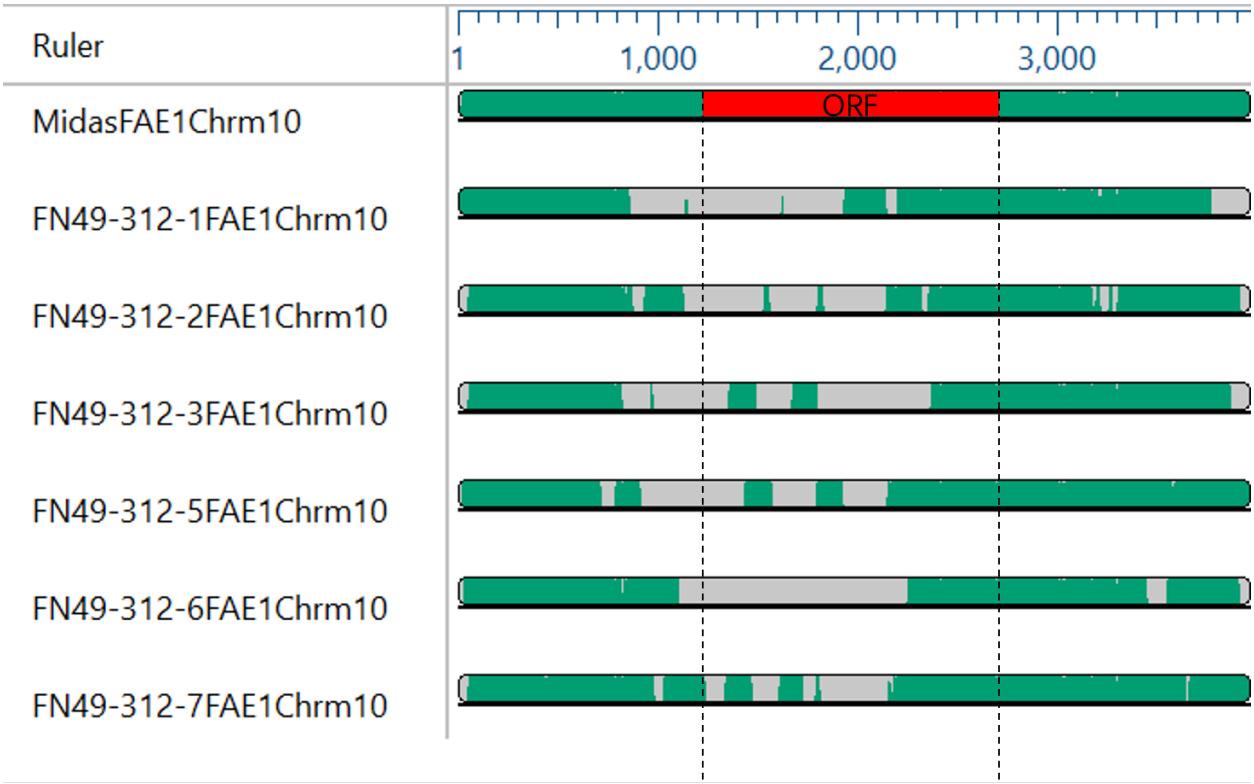

Figure 4.23 Alignment of *FAE1* gene on chromosome 10 of Midas (wild type) and FN49-312-1, -2, -3, -5, -6, -7. The green region indicated the DNA sequence, where was aligned to each other, and grey region represented the missing sequences. The red region represented the ORF of *FAE1B*. The algorithm of alignment applied was MAAFT. The figure was generated by software DNASTAR. The green regions in the middle of the grey part means the deletion was not identical and consecutive in those mutant lines.

5. Discussion

5.1 Influence on camelina seed induced by FN irradiation

Fast neutron (FN) irradiation can generally result in significant damage to the genome, such as large indels and rearrangements resulting from double strand DNA breaking and subsequent repair (Tanaka et al., 1999). To investigate the effects of the FN mutagenesis on camelina seeds, a germination test done on petri dishes was applied (Figure 4.1). After 24 hours, the germination percentage of the group that had received the largest dose of FN was the lowest, while the germination percentage for the group that had not received FN was the highest. Along with the increase in treatment dose, the tendency of germination percentage decreased. According to this observation, it appears that, as expected, stronger FN irradiation results in more substantial genetic damage. This finding is consistent with the result demonstrated by Bolon et al. (2011). They found that the higher the FN dose applied to soybean seeds, the fewer M1 plants emerged, and the fewer individuals produced viable seeds. Nevertheless, the germination percentages of each group became similar after 48 hours and reached almost 100% after 72 hours. *C. sativa* is a hexaploid plant whose genetic redundancy confers strong resilience to induced mutations (Neumann et al., 2021). Thus, unsurprisingly, after FN mutagenesis, camelina seeds still maintain a germination rate comparable to that of the parent Midas after 72 hours of germination (Figure 4.1). The initial delay in germination may be attributed to the seeds needing time to repair DNA damage caused by the FN radiation in the early stages of germination. As described by Waterworth et al. (2019), seeds do not conduct DNA repair until they have been imbibed. Accordingly, the seeds preserved in the dry environment will not be able to initiate the DNA damage repair process, regardless of how long they have been in the inventory. Once the seeds absorb moisture and begin to germinate, and the DNA lesion caused by the FN irradiation begin to be repaired. Consequently, the percentage of germination decreased as the irradiation dose

was increased in the first 24 hours following the imbibing of the seeds. Across all groups, germination was similar after 48 hours. This result demonstrates that the DNA repair was completed within 48 hours and if the damage was repaired, the seeds would continue to germinate. However, if the damage was too severe, the seeds would die. Considering all groups showed close to 100% germination after 72 hours of incubation, this indicated that the effects introduced through the FN mutagenesis were acceptable, meaning that the FN treatment was not intended to affect the percentage of germination of camelina seeds.

Based on the field test (Figure 4.2), about 12% of seeds that were irradiated survived and reached the flowering stage in the field environment. All irradiated groups had a comparable survival rate regardless of the dose of irradiation. In contrast, the Midas seeds at the same age as the FN treated seeds had a 5% higher plant establishment rate at around 18% of planted seeds. Furthermore, when compared to fresh Midas seeds, the 4-year-old Midas seeds had a percentage about 10% lower than that of fresh Midas seeds. There was considerable variation between the plots, and as only 2 plots were planted statistical significance was not determined. According to these results, FN irradiation, at the doses applied, may impact plant vigour but does not influence their ability to germinate. Individuals subjected to lower doses of irradiation displayed better growth. In contrast to plants that were irradiated with higher doses, those that were treated with lower doses were more flourishing, and although no field notes were taken the plants were observed to blossom earlier with more flowers (Figure 4.2 A and B) and had thicker stems. It may be that higher doses of irradiation result in weaker plants due to the genetic redundancy of camelina not being able to sufficiently compensate for the defects caused by irradiation, or the incorrect DNA repair of mutations.

5.2 Useful visual phenotypes

During this work 2 visual phenotypes were briefly investigated, potential male sterility in lineage FN49-33 and leaf morphology in lineage FN49-336. The preliminary study confirmed that the male sterility trait could be rescued via backcrossing to wild type and suggested that it has a genetic basis (Section 4.1.3.1). A reduction in pollen grains was observed in two lineages that resulted from backcrossing. The observation that BLK1-8 produced less pollen grains than BLK1-4 suggests that the first generation of offspring were segregating based on this trait. As all first generations after backcrossing should be heterozygous, and there were two levels of pollen production found in two backcrossing lineages, there could be more than one gene controlling this phenotype. Future research on this trait will continue to focus on determining the mechanisms that result in a reduction in pollen production in these plants. There can be two possible causes of male sterility, either mitochondrial genes coupled to nuclear genes or nuclear genes alone, which are called cytoplasmic male sterility (CMS) or genetic male sterility (GMS) respectively (Chen & Liu, 2014). The CMS can be determined by reciprocal backcrossing experiment. Since mitochondrial genes are predominantly responsible for male sterility and mitochondrial genes are inherited maternally, male sterility would merely be observed in the backcrossing line in which the mutant line served as pollen acceptor, while this phenotype would not be detected in the backcrossing line in which the wild-type line was used as pollen acceptor. In this project, the mutant lineage FN49-33 was used as pollen acceptor. After backcrossing, the progenies produce lower levels of pollen than the wild type. Future research could utilize this mutant line as a pollen donor and monitor the pollen production of its offspring. Having similar levels of pollen in their offspring would suggest that the male sterility is caused by CMS. Male sterility is unusual lasting over generations in the population, because in the nuclear genome, there is a system for restoring fertility, which includes the *Rf* (*restorer of fertility*) gene (Chen & Liu, 2014). The role of these functional *Rf* genes is to influence the functioning of mitochondrial

and chloroplast genes to restore the sterile phenotype. Additionally, genetic factors determining GMS can be identified via GBS. The information regarding GMS genes has been widely published (Chen & Liu, 2014). Those genes' information may serve as targets for retrieving the mutant genome that was revealed by sequencing methods. Further, those genes can be utilized as a reference for the design of probes to be used for the exploration of the mutant genome to uncover the potential GMS genes. Recently, *REDUCED POLLEN NUMBER1* (*RDP1*) is proved to have strongest association regarding the reduction of the number of pollen grains in the arabidopsis (Tsuchimatsu et al., 2020). This *RDP1* could be target for further study. Moreover, based on the research conducted by Wang et al. (2014), they used backcrossing experiment to induce male sterile trait from Naibaicai (*Brassica campestris* L. ssp. *chinensis* L., S01) to Wutacai (*Brassica campestris* L. ssp. *chinensis* (L.) Makino var. *rosularis* Tsen et Lee), whose wild type is fertile. After several generations of backcrossing, they obtained the homozygous Wutacai with male sterile trait and proved that the male sterile was dominated by three allele loci: male sterile gene, fertile gene, and fertility restoration gene. They demonstrated that the dominant to recessive relationship of the alleles was fertility restoration gene > male sterile gene > fertile gene. Their research provided a reference pattern to which the male sterility we found in this project could be similar. Overall, although it is not known if male sterility discovered in this project was either CMS or GMS, it should be noted that in plants pollen production can also be affected by environment. Thus, the male sterility line identified from the mutant population needs further confirmation. For this project the backcrossing work occurred in summer 2021. Unfortunately, Saskatoon experienced a heat wave at that time, and the cooling system in the AAFC greenhouse was not able to maintain the set temperature resulting in some overheating. This may have affected the development of male gametogenesis. Therefore, although a few of their offspring’s flower express the less-pollen and no-pollen phenotype, that was possibly caused not just by genotype, but also jointly affected by the non-ideal environment factor. Thus, the experiment on the pollen viability assays should be repeated in the future to

obtain more reliable results. Furthermore, stain tests of pollen viability are often over-estimated compared to pollen germination assays. Applying a pollen germination assay may increase the accuracy of the result. In short, as hybrid crops can produce 15-50% higher yield than inbred varieties due to heterosis (Tester & Langridge, 2010), this phenotype may be potentially useful as male sterile lines have value in hybrid production.

According to our results, one lineage FN49-336 can produce toothed leaf margins and is inheritable (Figure 4.6). Although the numbers and extents of teeth formation were not clearly recorded, each line appeared to display different numbers and extents of teeth. It has been suggested that leaf patterning and development are influenced by hormone signaling and a genetic network involving transcription factors (Hay et al., 2004, Byrne, 2005, Fleming, 2005). For the toothed leaf, the causes of this altered phenotype remain unclear. There are two genes responsible for the regulation of leaf margin serration in arabidopsis: *CUC2-SHAPED COTYLEDON2* (*CUC2*) and *miR164A*, a microRNA (miRNA) targeting the transcripts of *CUC1* and *CUC2*, with *CUC2* controlling the pattern and *miR164A* determining the extents of serration (Nikovics et al., 2006). Therefore, it is reasonable to speculate that lineage FN49-336 may contain mutations related to the region of *CUC2* and *miR164A*.

5.3 Alteration of fatty acid composition in FN camelina seeds

When characterizing seed oil phenotypes for the mutant lines, seed fatty acid composition, seed weight and seed oil content were all measured. Many of the FN mutant lines had a lower oil content compared to the wild type, although this was significantly affected by poor greenhouse growth conditions. FN irradiation appeared to have a very significant impact on the genome, as this observation indicates. The oil content phenotype is controlled by quantitative trait loci (King et al., 2019). As a result, FN mutagenesis could have damaged multiple loci and restoring all

these lesions proved difficult. Therefore, no line has been identified with significantly higher oil content than the Midas so far.

While screening for fatty acid composition, a number of lines were observed with altered seed fatty acid profiles. The discovery of a reduced amount of C20:1 in the mutant line FN49-38 was promising because C20:1 is an uncommon fatty acid in vegetable oils, but its consumption has no known health enhancing or detrimental effects. From the perspective of biosynthesis, the generation of C20:1 is synthesized through elongation of C18:1 catalyzed by *FAE1* (Sarvas et al., 2021) (Figure 5.1). Nevertheless, C18:1 is the substrate shared by both elongation and desaturation, which means the biosynthetic pathway for omega-6 and omega-3 fatty acids compete this substrate with the generation of C20:1. Accordingly, given limited amount of C18:1, the pathway of synthesizing C20:1 can be the target to be blocked so that more substrates would be used for the biosynthesis of C18PUFAs. According to Fig 5.3.1, the reduction of C20:1 would lead to an increase in the total content of C18:2 and C18:3, as the elongation pathway and desaturation pathway competes the same substrate. Statistical analysis among all M_3 seeds determined via GC showed that the Pearson correlation coefficient of the two was $r = -0.7$, indicating that the amount of C20:1 was indeed negatively associated with the content of C18:2 and C18:3. Thus, the block or lower efficiency of the elongation from C18:1 to C20:1 can result to more C18:1 being utilized in the formation of C18:2 or C18:3. Due to the fact that an oil with a lower content of VLCFAs and a higher amount of n-3 fatty acids can be of nutritional value, this lineage of FN49-38 has the potential to improve camelina oil as an edible oil.

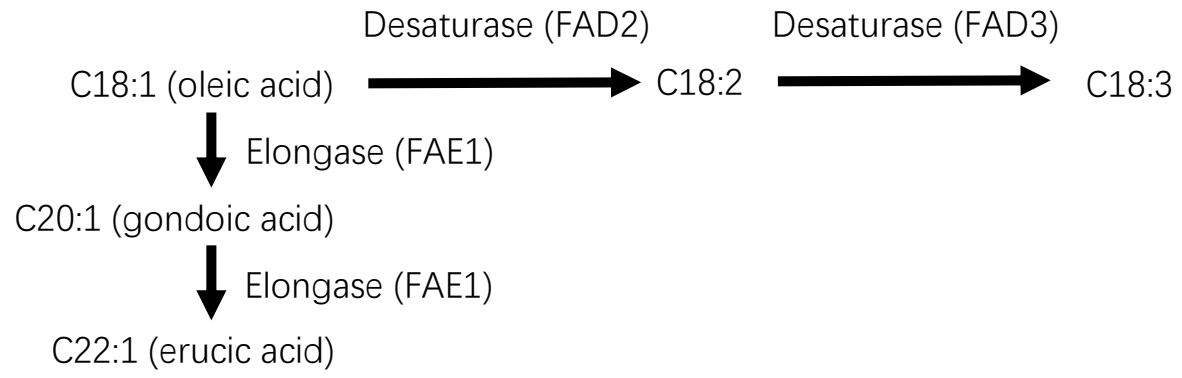

Figure 5.1 Pathway of biosynthesis of C20:1 and other important fatty acids in *C. sativa*.

Higher content of C18:3 was not solely found in FN49-438 lineage. FN49-312-7 is another line showing around 41% of C18:3. Furthermore, FN42-97-1 contains 45.2% of C18:3, the highest among all the lines detected by GC. A possible explanation is that FN caused point mutations in the region of the fatty acid desaturase 3 (*FAD3*) gene, increasing enzyme activity or upregulating the gene expression. Some point mutations change amino acids in the exons, which leads to enhanced enzyme activity. Another scenario is when the point mutations take place in the introns. According to Mazoochian et al. (2018), SNPs located in intron are capable to impact the activity of the corresponding enzyme. Introns can strengthen gene expression by influencing the rate of transcription, nuclear export, and transcript stability as well as enhancing the efficiency of mRNA translation, which is known as 'intron-mediated enhancement (Shaul, 2017). In arabidopsis, a transcriptional increase induced by low temperature was found to be mediated by the nucleotide region from -80 to -6 of the *FAD3* 5'UTR (Wang & Xu, 2010). This regulation results in the FAD3 protein level being higher at low temperature than high temperature (Wang & Xu, 2010). Therefore, the mutations found in the FN mutant lines might be related to this 5'UTR region, which causes translation to increase not only at low temperatures. Furthermore, the alteration of the TAG assembly pathway could also lead to the change of fatty acid composition. The enzymes related to the acyl-editing of the PC pool like PDCT and PDAT

can alter the final fatty acid composition of TAG, the major lipid species in seeds. In other words, if the genes encoding these two enzymes are mutated, it can result in changing C18:3 content in seeds. Therefore, to understand why those lines possess C18:3 content higher than their ancestors, further genotyping analysis will be needed to disclose what mutation has truly occurred in the genome.

The accumulation of fatty acid composition in developing camelina seeds of FN 49-38 lineage (Section 4.4.2) revealed that there was a trend of rising and then falling of the proportion of C18:3. An assumption to explain this observation is that the proportion of C18:3 was increased at first because plants do not generate too many VLCFA at the early period of seed oil development. Furthermore, C18:1 as a substrate for both desaturation and elongation (Figure 5.1), C18:1 may be more likely through the desaturation pathway over the elongation pathway in camelina at the early stage of seed development. This assumption is accordance with the finding from another research conducted on developing camelina seeds (Rodríguez-Rodríguez et al., 2013). According to their results, the content of C18:3 in wild type camelina seeds (CAS-CS0 cultivar) increased from 6 days after flowering to 50 days after flowering (DAF), as well as the content of C20:1 was accumulated throughout the entire period and reaching the highest amount in mature seeds. Their study indicates that C18:1 and C18:2 fatty acids were produced by plastidial *de novo* fatty acid synthesis prior to 18 DAF. In the period 18 to 24 DAF, linoleic and linolenic acids accumulated significantly, while the concentration of C18:1 declined dramatically. From 12 DAF, the proportion of gondoic and erucic acids increased significantly. Accordingly, they concluded that the decrease in C18:1 between 12 and 27 DAF is due to the activation of desaturases and elongase complexes for the generation of PUFAs and VLCFAs, and fatty acid elongase complexes were activated later than desaturases. From the gene expression perspective, gene related to desaturation may be in a higher expression level than the gene controlling the elongation pathway at the early stage of seed development in camelina. Therefore, the proportion of C18:3 was enhanced as more and more substrates were converted to

C18:3. After 30 days, the VLCFAs kept generating and it seems the amount of C18:3 achieve saturation level, which a little or no more C18:3 was generated, so the proportion of C18:3 started to decline. Compared to Rodríguez-Rodríguez et al.'s study, a deletion of *FAE1B* resulted in the maximum level of C18:3 being reached earlier than their study indicates. There may be something limiting the generation of C18:3 once its level reaches a certain level, even though the majority of C18:1 was turned into desaturation pathways rather than elongation pathways. Gene, enzyme, and other factors could play a role in this regulation.

The mutant line FN49-70 contained a higher concentration of erucic acid compared with a lower concentration of oleic acid. Since the proportion of C18:2, C18:3, and C20:1 was very similar to Midas, this showed that genes associated with the elongation pathway were likely mutated to cause more substrate to be converted into C22:1. It is noteworthy that the C20:1 content is only slightly lower than the wild type, suggesting that the mutation in this line caused the elongation pathway to use more substrates than usual and terminate at C22:1. As a result, total VLCFA content of this lineage is higher than Midas. Essentially, more C18:1 was used for elongation than desaturation (Figure 5.1), suggesting that mutations occurred in this lineage might be related to desaturase genes, although. Further research is needed to confirm this hypothesis.

5.4 Genotyping by GBS approach and PCR application

Mutant lineage FN49-38 was identified with a reduced amount of C20:1. According to a comparison of the genome sequences generated by Illumina sequencing of the mutant and reference, deletions in the *FAE1B* gene were identified. The authenticity of the deletion was confirmed by comparison of their genome sequences generated by ONT. The benefit of using two types of sequencing techniques is the complementation of each other. Illumina sequencing, a

short-read sequencing technique, generates short reads of about 200 bp with high accuracy. The ONT, a long-read sequencing technique, can generate long reads over two thousand bp with a slightly lower accuracy. The disadvantage of the short reads is sequence assembly where the algorithm attempts to connect the short reads shared in identical sequences to form contigs. It could discard some reads with low sequencing quality or skip short identity sequences resulting in gaps between contigs. These gaps are not due to deletion but are generated by the assembly algorithm. There is also the possibility that the sequence may correspond to more than one location if there is a high degree of sequence homology between genes, which may result in an incorrect assembly. Moreover, the depth of short-read sequencing data is a crucial parameter to check if the sequencing data is dependable. The errors of assembling contigs will be enhanced dramatically once the depth of short-read sequencing data is below 4x (Jiang et al. 2019). Similarly, Sims et al. (2014) stated that low depths may introduce sequence errors that may contribute to misdirecting conclusions of later analyses. As a matter of fact, in this project, the lower-depth data is frequently combined with many absent regions where no sequence data was present. These regions were confusing because it was not clear whether they represented deletions or just areas that had not been sequenced. Due to different factors causing interference with the assembly of short reads, the gaps noted in the genome sequence following their assembly may not be the true deletion. To confirm whether the gaps in *FAE1* gene in the mutant lines identified from genome sequences generated by Illumina sequencing are authentic deletion induced by FN, ONT was introduced to sequence the genomes for validation. The long reads sequencing technique can generate reads ranging from a few hundred bp to over ten thousand bp, which can improve the sequence assembly in reducing gaps between contigs. Thus, it can be used to validate gaps or deletions identified by Illumina sequencing. In order words, if a gap exists only in contigs from short reads, it is likely to be a true deletion. But if a gap is found in both contigs from short reads and long reads, it is a proved deletion. Therefore, the authenticity of deletions in a *FAE1* gene in the mutant lines identified from genome sequences generated by

Illumina sequencing is confirmed by genome sequences generated by ONT. Furthermore, application of cloning was applied to clone the target region containing the location of the deletion. Those target region was sequenced via the first-generation sequencing technique, Sanger sequencing technology, which yields almost 100% accuracy. This sequencing result was used to confirm on one hand whether the deletion was true, on the other hand, to prove that using ONT's data to validate the deletion observed from the short-read sequencing data was practicable. Therefore, genotyping by sequencing can be achieved by using short-read sequencing technique combined with long-read sequencing technique. In this way, the mutant's genotype can be explored broadly, and sequencing the specific lines that potentially carried remarkable deletion with ONT can effectively validate the facticity of that deletion.

A potential explanation for the inaccurate annotation found on chromosomes 11 and 12 for the *FAE* genes in the published *C. sativa* genome sequences may be related to errors in the assembly process. Short-read sequencing data often experience erroneous assembly, particularly when the genome size is large and polyploid, and the genome structure is complex. The value of long-read sequencing data as a validation of short-read assembly is therefore emphasized. One single read with thousands of base pairs can elevate the accuracy of short-read assembly because it reduces the likelihood of false assemblies caused by repeat regions, regions with extremely high resemblance, or tandem sequence regions.

Genotyping the mutant lines by PCR with specific primers on target genes is more straightforward. When the amplified PCR product of a target gene in mutant lines display a size different from the one from wild type gene, it means a mutation might have occurred. Since Midas's phenotype of gondoic acid content will not change between generations, there is no more segregation in this phenotype seen on Midas. In this respect, Midas can be defined as homozygous and used as a reference. The PCR product generated using gene specific primers with Midas DNA as template should only yield one band on an agarose gel. When a distinct band is observed, it must represent a new genotype resulting from a mutation. Accordingly, PCR

application may be used to ascertain whether an indel has been induced in the target region. Nevertheless, several preconditions must be met with this approach so as to achieve expected results. Firstly, it is important to define the kind of mutations. If it is a large deletion, for example over 500 bp, PCR screening is more straightforward as the length of the mutant gene will be shorter than that of wild type gene. But the difference of band's size might be hardly distinguished when the indel's size is not large enough. Secondly, the primers for PCR should be appropriately designed to cover the whole deletion region. Therefore, if we know where the mutation occurs and have the reference genome information, it will simplify the primer design. These two preconditions also could be the limitations of using PCR to genotype. Firstly, when the mutation is SNPs or point mutations, if the primer is placed flanking the locus, the bands will be similarly sized, resulting in the failure to distinguish the mutant line from the wild-type line. In the case that the primer is designed at the mutation site, the primer may still function because the difference of a single nucleotide could only marginally affect the primer's ability to bind to DNA. Secondly, the reference genome information should be accessible if PCR application is conducted, which means if the working genome is entirely unexplored or the genome information is not reliable, the primer design will be difficult. Thirdly, while the reference genome information is available, it is unclear where the mutation has occurred, which will also hinder the primer design. For this project, FN49-38-1-11 carried a deletion in the *FAE1B* gene from some part of the 5' UTR to the first half of the *FAE1B*-ORF revealed by short-read and long-read sequence result. Thus, PCR amplification with two specific primers flanking the region was successful. Based on this successful experience, applying PCR screening to discover similar mutation among other mutant lines is feasible, which has been demonstrated in this project. A DNA pooling strategy accompanied with PCR can enhance the efficiency of screening. The complexity of the pool tested in this work was up to 20 in 1, which means one mutant line carried deletion mixed with 20 mutant lines without the deletion. The result was robust, so the future pool can be built with increasing complexity. Therefore, the PCR application appears to be

functioning in an appropriate manner for the analysis of large deletions, such as the one found in this project, or also large insertions. There should be a substantial fragment length's difference between the mutant and the wild type, so that the band size of the PCR products differed sufficiently from the wild type, allowing it to be clearly separated from the WT band on an agarose gel. But it is not recommended to apply PCR for detecting small size mutation including point mutations, or small size of indels which are just 1 or 2 bp different. Overall, the success of the PCR in determining mutation depends on if the primers are capable to be designed and whether they function correctly.

Owing to its rapid and high-throughput nature, the GBS approach can be used for the genotyping of mutant lines. A new approach conducted in this project combining Illumina sequencing and ONT sequencing can result in the fact that laboratory work can be performed with a lot less time and effort since bioinformatics can be performed with the help of a powerful computer to deal with the needed analysis. Additionally, the PCR method for testing the deletion was quite involved as the primers needed to be designed ahead for the PCR were specific and the pilot experiment had to be conducted several times to achieve the optimal PCR conditions. Considering the complexity of the hexaploid genome structure of camelina, which has genes related to lipid metabolism in triplicate (Li-Beisson et al., 2013), it becomes challenging to design primers for validation. Therefore, the long-read sequencing data used for validation in our project can be used to bypass the PCR experiment to achieve the same outcome. Nonetheless, detailed PCR genotyping can be applied to confirm the mutant line's zygote type, due to two bands for heterozygote and one band for homozygote displayed on the agarose gel after electrophoresis. Thus, a combination of these two approaches would be effective to genotype a mutant population. PCR is also a convenient way to track a mutation in a breeding program.

5.5 Linkage between genotype and phenotype

The study of the linkage between the deletion genotype and the phenotype of reduced gondoic acid content via backcrossing is essential since the novel phenotype has not been proved to be a direct result of the detected deletion. The mutant line carrying the deletion and showing reduced gondoic acid could be explained by two characteristics being associated with each other rather than confirming they have linkage. The purpose of backcrossing between the mutant lines and Midas was, therefore, to determine if this new genotype could be transferred into Midas's genome for heterozygous progeny. F_1 seeds after backcrossing would have a higher C20:1 content than the mutant line, but lower than Midas as plants were all heterozygous. The cultivation of the heterozygous progeny will enable the segregation of this trait. As a result, three distinct phenotypes could be observed started from the second generation if the trait showed co-dominance. One with C20:1 level as high as the wild-type level would be the dominant homozygote, one with C20:1 level as low as the mutant line's level would be the recessive homozygote, and the third with C20:1 amounts somewhere in between would be a heterozygote. Based on the PCR result (Figure 4.20), the heterozygote was most abundant and around one fourth of the plants were recessive homozygote. This observation was consistent with the expectation, which the recessive homozygote frequency should be around 0.25 and heterozygote should be around 0.5. Although the frequency of the heterozygote is exceeded than 0.5, which was around 0.68, this might be because the sample size was small, which could be not big enough to estimate the allele frequency of the whole population. The investigation of the linkage between the novel genotype and the phenotypic characteristics was conducted by combining the determination of the fatty acid composition of F_3 seeds and the application of PCR to explore their genotype. The fatty acid phenotype agreed with three genotypes, so the linkage was confirmed.

To reduce the effects of FN irradiation to a minimum level and produce a new reliable commercial cultivar, backcrossing the backcrossed offspring with Midas for a few generations will enable unwanted mutations to be eliminated. Additionally, backcrossing is a good strategy to study the relationship between new genotypes and novel fatty acid composition phenotypes. After two generations of backcrossing, the genotype has undergone combination and segregation. According to theory, the segregated mutant homozygote can produce a stable unique mutant phenotype. In consequence, the linkage can be confirmed when the mutant homozygous genotype produces the anticipating phenotype.

5.6 PCR troubleshooting

Since PCR was one of the major experiments in this project, numerous runs of PCR were conducted. A common phenomenon observed in heterozygous genotypes was a heavy bottom band and a light top band in the PCR results. It is possible that the PCR is biased towards amplifying the shorter bands over the longer bands. In addition, this deduction raised another concern, which was the possible failure of the amplification of the top band while the bottom band was amplified successfully. A false-positive, apparently homozygous mutant line carrying the deletion could result from this result. To avoid this problem, a pilot PCR run should be performed in order to determine the optimal PCR conditions. As I observed during the pilot PCR run, although the lower band can be produced over a wide range of conditions, the top band has a more stringent requirement for both annealing temperature and extension time. Additionally, if the PCR conditions were not optimal, the lower band production was not constant. This phenomenon illustrates the importance of conducting a pilot PCR prior to the actual test. Overall, the inherent complexities and limitations of PCR for deletion detection further highlight the advantages of using bioinformatics for detection.

5.7 Mutation frequency

Genetic variations play an imperative role in genetic diversity since they provide allele resources that allow the development of novel traits for plant breeding (Du et al. 2022). Mutations induced by FN mutagenesis are numerous and unpredictable. It is important to note, however, that the mutagenic effect may vary in different plant tissues and on different chromosomes (Du et al. 2022) Additionally, by applying various dosages of irradiation, the chances of causing SNPs, deletions, insertions, or rearrangements would be varied. Therefore, if the mutation frequency of different chromosomes and which dosage of irradiation induces most deletions can be revealed, it could facilitate the development of more controlled FN mutagenesis for plant breeding programs, thus simplifying the design of mutation screening approaches in downstream experiments.

5.8 Similar deletions found in both the FN49-38 and FN49-312 lineages

Based on the GC analysis (Section 4.1.4) and GBS approach (Section 4.4.5), lineages FN49-38 and FN49-312 appeared to contain similar deletion around *FAE1B*. There may be two explanations for this coincidence. One possible explanation is that the region of *FAE1* on chromosome 10 is more susceptible to FN irradiation. Deleting this region might protect the rest of the chromosome from being damaged by FN irradiation. Reduction in the fatty acid composition of these lineages was from 13.3% to around 10% on C20:1. Nevertheless, there were no obvious reductions in other VLCFAs. The reduction of C20:1 content found in our lineages FN49-38 and FN49-312 was not as significant as it was in the study that used EMS to

induce a point mutation on the *FAE1B* gene, which reduced the C20:1+C22:1 content by over 60% compared to the wild type (Ozseyhan et al., 2018). The other possibility is that the DNA was contaminated prior to PCR, or that there were PCR errors. It should be noted, however, that these are merely the assumed causes, and that a further investigation will be necessary to determine the exact causes.

6. Conclusion

Due to camelina's unique hexaploid advantage, most of its genes related to oil synthesis have three homologs. This characteristic of gene redundancy enables camelina to withstand DNA damage caused by FN radiation to the maximum extent, making it possible to use FN mutagenesis to improve the genetic diversity of camelina. In this project, a FN mutant library was established, and the identification and characterization of phenotypes and genotypes of mutant lines related to seed oil composition was the core of this project. The effect of FN irradiation on camelina seed germination was evaluated, and it was determined that no adverse effects were introduced at the dosages used that would reduce the ability to germinate. An effect of irradiation on plant vigour and a delay the maturity of the plants was however apparent from the field study. We carefully observed and recorded the noticeable visual phenotypes of camelina during its growth cycle. A male sterility characteristic has been identified among them, as well as the toothed appearance of the leaf shape. Both altered phenotypes were found to be inheritable. For fatty acid composition phenotypes, according to the GC results, it appeared that varied mutant lines contained different fatty acid composition compared to wild type, including high-content C18:3 lineage (FN49-438), lower C18:2 lineage (FN49-287), lower C20:1 lineage (FN49-38, FN42-353 etc.), higher C22:1 (FN49-70), and so on. Among them, the phenotype of lower C20:1 content was fully studied in this project. Genotyping by sequencing was applied to study FN49-38 offspring's genotypes. Genome information was explored by Illumina sequencing technology, which is a short-read sequencing technique. A deletion spanned from the 5' untranslated region (UTR) to the first half of the *FAE1B* open reading frame (ORF) in chromosome 10 was found from Illumina sequencing data. The ONT sequencing was applied to generate long reads of the mutated region to validate the facticity of the deletion, and its result was in accordance with the Illumina sequencing data. Cloning this mutated region for Sanger sequencing was utilized to prove using ONT for validation was reliable and to confirm the

sequence information generated from that two technique was true. Consequently, the sequencing results were consistent with the data generated from ONT and Illumina, which proved the design of using Illumina sequencing to reveal genome information extensively and using ONT sequencing to validate the result from short-read sequencing technique was practicable. Furthermore, a traditional method PCR was applied to confirm the deletion physically, because sequencing result was possible to be wrong raised by secondary structure, repeat region, or tandem sequence. The PCR products robustly gave two different bands on agarose gel after electrophoresis, which was consistent with the expectation. As a result, the deletion discovered from Illumina data was confirmed in FN49-38 progenies genome, and the genome information was validated. Moreover, the ONT data generated in our project has been used to review and correct the present annotation of *FAE* on chromosomes 11 and 12. PCR application was extended for screening in this project. PCR screening with pooling strategy could enhance the efficiency of finding already known large indels. In this project, the complexity of the pool was proved that it can be built up to 20 in 1, in which 20 lines without mutations in the target gene were mixed with 1 mutated line. Finally, the backcrossing experiment demonstrated that the mutant homozygous plants produced seeds with a lower C20:1 content than the wild-type homozygous plants. This observation illustrated that the deletion was linked to the reduced C20:1 phenotype.

In summary, the FN camelina library has been established. Lineage FN49-336 producing toothed leaves was observed and male sterile line FN49-33-6 was identified. FN49-38, a mutant line exhibiting a decreased content of gondoic acid, was studied and the deletion located around *FAE1B* was identified and characterized. Because of this study, we have solidly demonstrated that FN mutagenesis can change the genotype of *C. sativa* seeds, thus altering their fatty acid composition. The GBS employed an innovative methodology that combined short-read and long-read sequencing to detect deletions, which proved to be effective and generated robust results. The ONT data provided valuable information for the validation of the short-read assembly result.

Backcrossing experiments provided further insight into the linkage between genotype and phenotype. In short, the characterized mutant line can be utilized as genetic resources for future studies and will likely benefit the breeding program for camelina after additional genetic purification work to remove other unwanted mutations is completed.

References

Abramovič, H., & Abram, V. (2005). Physico-Chemical properties, composition and oxidative stability of *Camelina sativa* oil. *Food Technology and Biotechnology, 43*(1), 63.

Al-Hajaj, N., Peterson, G., Horbach, W., Al-Shamaa, C., Tinker, K., & Fu, N. (2018). Genotyping-by-sequencing empowered genetic diversity analysis of Jordanian oat wild relative *Avena sterilis*. *Genetic Resources and Crop Evolution, 65*(8), 2069-2082.

Amarasinghe, S., Su, S., Dong, X., Zappia, L., Ritchie, M., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology, 21*(1), 30.

Athenstaedt, K., & Daum, G. (2006). The life cycle of neutral lipids: Synthesis, storage and degradation. *Cellular and Molecular Life Sciences, 63*(12), 1355-1369.

Baral, K., Coulman, B., Biligetu, B., & Fu, Y. (2018). Genotyping-by-sequencing enhances genetic diversity analysis of crested wheatgrass [*Agropyron cristatum* (L.) gaertn.]. *International Journal of Molecular Sciences, 19*(9), 2587.

Bartel, D. (2009). MicroRNAs: Target recognition and regulatory functions. *Cell, 136*(2), 215-233.

Bates, P.D., Fatihi, A., Snapp, A.R., Carlsson, A.S., Browse, J., Lu, C. (2012). Acyl editing and headgroup exchange are the major mechanisms that direct polyunsaturated fatty acid flux into triacylglycerols. *Plant Physiology (Bethesda), 160*(3),1530-1539.

Bates, P.D. (2016). Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. *Biochimica et Biophysica Acta, 1861*(9), 1214-1225.

Bates, P., Stymne, S., & Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Current Opinion in Plant Biology, 16*(3), 358-364.

Baud, S. (2018). Seeds as oil factories. *Plant Reproduction, 31*(3), 213-235.

Beaudoin, F., Wu, X., Li, F., Haslam, R., Markham, J., Zheng, H., . . . Kunst, L. (2009). Functional characterization of the Arabidopsis β-ketoacyl-coenzyme A reductase candidates of the fatty acid elongase. *Plant Physiology (Bethesda), 150*(3), 1174-1191.

Bekele, W., Wight, C., Chao, S., Howarth, C., & Tinker, N. (2018). Haplotype-based genotyping-by-sequencing in oat genome research. *Plant Biotechnology Journal, 16*(8), 1452-1463.

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.

Bernerth, R., & Frentzen, M. (1990). Utilization of erucoyl-CoA by acyltransferases from developing seeds of *Brassica napus* (L.) involved in triacylglycerol biosynthesis. *Plant Science (Limerick), 67*(1), 21-28.

Bolon, Y., Haun, W., Xu, W., Grant, D., Stacey, M., Nelson, R., . . . Vance, C. (2011). Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiology (Bethesda), 156*(1), 240-253.

Bonaventure, G., Salas, J.J., Pollard, M.R., Ohlrogge, J.B. (2003). Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell* 15: 1020–1033.

Botella, C., Sautron, E., Boudiere, L., Michaud, M., Dubots, E., Yamaryo-Botté, Y., ... Jouhet, J. (2016). ALA10, a phospholipid flippase, controls FAD2/FAD3 desaturation of phosphatidylcholine in the ER and affects chloroplast lipid composition in *Arabidopsis thaliana*. *Plant Physiology (Bethesda), 170*(3), 1300–1314. doi: 10.1104/pp.15.01557

Boyle, C., Hansen, L., Hinnenkamp, C., & Ismail, B. (2018). Emerging camelina protein: extraction, modification, and structural/functional characterization. *Journal of the American Oil Chemists' Society, 95*(8), 1049-1062.

Brash, A. (2001). Arachidonic acid as a bioactive molecule. *The Journal of Clinical Investigation, 107*(11), 1339-1345.

Bremer, J., & Norum, K. (1982). Metabolism of very long-chain monounsaturated fatty acids (22:1) and the adaptation to their presence in the diet. *Journal of Lipid Research, 23*(2), 243-256.

Budin, J., Breene, W., & Putnam, D. (1995). Some compositional properties of camelina (*Camelina sativa* L. Crantz) seeds and oils. *Journal of the American Oil Chemists' Society, 72*(3), 309-315.

Byrne, M. (2005). Networks in leaf development. *Current Opinion in Plant Biology, 8*(1), 59-66.

Cai, G., Wang, G., Kim, S.C., Li, J., Zhou, Y., Wang, X. (2021). Increased expression of fatty acid and ABC transporters enhances seed oil production in camelina. *Biotechnology for Biofuels, 14*(1), 49.

Calder, P.C. (2006). Polyunsaturated fatty acids and inflammation. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, *75*(3), 197–202.

Calder, P. (2020). N-3 PUFA and inflammation: From membrane to nucleus and from bench to bedside. *Proceedings of the Nutrition Society, 79*(4), 404-416.

Campbell, M., Rossi, A., & Erskine, W. (2013). Camelina ((L.) Crantz): Agronomic potential in Mediterranean environments and diversity for biofuel and food uses. *Crop and Pasture Science, 64*(4), 388-398.

Canada, H. (2012, October 11). Government of Canada. Retrieved from https://www.canada.ca/en/health-canada/services/food-nutrition/genetically-modified-foods-other-novel-foods/approved-products/camelina-oil-novel-food-information.html

Canola oil: Heart-healthy oil for all kinds of cooking. (2022, August 22). Retrieved November 21, 2022, from https://www.canolacouncil.org/about-canola/oil/#fact-and-fiction

Chapman, K., & Ohlrogge, J. (2012). Compartmentation of triacylglycerol accumulation in plants. *The Journal of Biological Chemistry, 287*(4), 2288-2294.

Chaudhary, R., Koh, C., Kagale, S., Tang, L., Wu, S., Lv, Z., . . . Parkin, I. (2020). Assessing diversity in the camelina genus provides insights into the genome structure of *Camelina sativa*. *G3: Genes - Genomes - Genetics, 10*(4), 1297-1308.

Chaudhary, J., Deshmukh, R., & Sonah, H. (2019). Mutagenesis approaches and their role in crop improvement. *Plants (Basel), 8*(11), 467.

Chen, G., Harwood, J., Lemieux, M., Stone, S., & Weselake, R. (2022). Acyl-CoA:diacylglycerol acyltransferase: Properties, physiological roles, metabolic engineering and intentional control. *Progress in Lipid Research, 88,* 101181.

Chen, L., & Liu, Y. (2014). Male sterility and fertility restoration in crops. *Annual Review of Plant Biology*, 65(1), 579-606.

Chen, X., Liu, X., Wu, D., & Shu, Q. (2006, May 15). Recent progress of rice mutation breeding and germplasm enhancement in China. Retrieved October 10, 2022, from https://www.osti.gov/etdeweb/biblio/20880300

Chen, Y., Liang, H., Ma, X., Lou, S., Xie, Y., Liu, Z., … Liu, Y. (2013). An efficient rice mutagenesis system based on suspension-cultured cells. *Journal of Integrative Plant Biology, 55*(2), 122-130.

Chen, Y., Kelly, E., Masluk, R., Nelson, C., Cantu, D., & Reilly, P. (2011). Structural classification and properties of ketoacyl synthases. *Protein Science, 20*(10), 1659-1667.

Christie, W., Nikolova-Damyanova, B., Laakso, P., & Herslof, B. (1991). Stereospecific analysis of triacyl-sn-glycerols via resolution of diastereomeric diacylglycerol derivatives by high-performance liquid chromatography on silica. *Journal of the American Oil Chemists' Society, 68*(10), 695-701.

Clemente, T., & Cahoon, E. (2009). Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiology (Bethesda), 151*(3), 1030-1040.

Colbert, T., Till, B., Tompa, R., Reynolds, S., Steine, M., Yeung, A., . . . Henikoff, S. (2001). High-Throughput screening for induced point mutations. *Plant Physiology, 126*(2), 480-484.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., … 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 - PNAS, 97*(12), 6487-6492.

Dawczynski, C., Martin, L., Wagner, A., & Jahreis, G. (2010). N − 3 LC-PUFA-enriched dairy products are able to reduce cardiovascular risk factors: A double-blind, cross-over study. *Clinical Nutrition (Edinburgh, Scotland), 29*(5), 592-599.

D'Andréa, S., Canonge, M., Beopoulos, A., Jolivet, P., Hartmann, M., Miquel, M., . . . Chardot, T. (2007). At5g50600 encodes a member of the short-chain dehydrogenase reductase superfamily with 11beta- and 17beta-hydroxysteroid dehydrogenase activities associated with *Arabidopsis thaliana* seed oil bodies. *Biochimie, 89*(2), 222-229.

D'Amour, C., Reitsma, F., Baiocchi, G., Barthel, S., Güneralp, B., Erb, K., . . . Seto, K. (2017). Future urban land expansion and implications for global croplands. *Proceedings of the National Academy of Sciences - PNAS, 114*(34), 8939-8944.

Deschamps, S., Llaca, V., & May, G. (2012). Genotyping-by-sequencing in plants. *Biology, 1*(3), 460-483.

Dönmez, E., & Belli, O. (2007). Urartian plant cultivation at Yoncatepe (Van), Eastern Turkey. *Economic Botany, 61*(3), 290-298.

Du, Y., Feng, Z., Wang, J., Jin, W., Wang, Z., Guo, T., . . . Zhou, L. (2022). Frequency and spectrum of mutations induced by Gamma rays revealed by phenotype screening and whole-genome re-sequencing in *Arabidopsis thaliana*. *International Journal of Molecular Sciences, 23*(2), 654.

Dumschott, K., Schmidt, M., Chawla, H., Snowdon, R., & Usadel, B. (2020). Oxford nanopore sequencing: New opportunities for plant genomics? *Journal of Experimental Botany, 71*(18), 5313-5322.

Ehrensing, D. & Guy, S.O. (2008). Camelina. Oregon State University Extension Service, EM 8953-E, January

Elshire, R., Glaubitz, J., Sun, Q., Poland, J., Kawamoto, K., Buckler, E., & Mitchell, S. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One, 6*(5), E19379.

Espina, M., Ahmed, C., Bernardini, A., Adeleke, E., Yadegari, Z., Arelli, P., … Taheri, A. (2018). Development and phenotypic screening of an ethyl methane sulfonate mutant population in Soybean. *Frontiers in Plant Science, 9*, 394.

Ewing, T., Julsing, M., & Boeriu, C. (2019, July 11). Isolation of gondoic acid from camelina oil by hydratase-catalysed reactive separation. Retrieved October 10, 2022, from https://research.wur.nl/en/publications/isolation-of-gondoic-acid-from-camelina-oil-by-hydratase-catalyse

Fleming, A. (2005). The control of leaf development. *The New Phytologist, 166*(1), 9-20.

Fox, E., Reid-Bayliss, K., Emond, M., & Loeb, L. (2014). Accuracy of next generation sequencing platforms. *Next Generation, Sequencing & Applications, 1*.

Frankel, E. N. (2012). *Lipid oxidation*. Cambridge: Woodhead Publishing.

Gibellini, F., & Smith, T. (2010). The Kennedy pathway—*de novo* synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life, 62*(6), 414-428. https://doi.org/10.1002/iub.337

Gilchrist, E. & Haughn, G. (2010) Reverse genetics techniques: engineering loss and gain of gene function in plants. *Briefings in functional genomics 9*, 103-110.

Gomes, A., & Korf, B. (2018). Genetic testing techniques. In *Pediatric Cancer Genetics* (pp. 47-64).

Gorkovskiy, A., & Verstrepen, K. (2021). The role of structural variation in adaptation and evolution of yeast and other fungi. *Genes, 12*(5), 699.

Gugel, R., & Falk, K. (2006). Agronomic and seed quality evaluation of *Camelina sativa* in western Canada. *Canadian Journal of Plant Science, 86*(4), 1047-1058.

Güneralp, B., Reba, M., Hales, B., Wentz, E., & Seto, K. (2020). Trends in urban land expansion, density, and land transitions from 1970 to 2010: A global synthesis. *Environmental Research Letters, 15*(4), 44015.

Hagemann, I. (2015). Overview of technical aspects and chemistries of Next-Generation Sequencing. In *Clinical Genomics* (pp.3-19).

Hamid, M. A., Azad, M. A. K., & Howelider, M. A. R. (2006). Development of three groundnut varieties with improved quantitative and qualitative traits through induced mutation. *Plant Mutation Reports*, *1*(2), 14-16.

Hanano, A., Burcklen, M., Flenet, M., Ivancich, A., Louwagie, M., Garin, J., & Blée, E. (2006). Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif. *The Journal of Biological Chemistry, 281*(44), 33140-33151.

Haslam, T., & Kunst, L. (2013). Extending the story of very-long-chain fatty acid elongation. *Plant Science (Limerick), 210*, 93-107.

Hay, A., Craft, J., & Tsiantis, M. (2004). Plant hormones and homeoboxes: Bridging the gap? *BioEssays, 26*(4), 395-404.

He, M., Qin, C., Wang, X., & Ding, N. (2020). Plant unsaturated fatty acids: Biosynthesis and regulation. *Frontiers in Plant Science, 11*, 390.

Henikoff, S., Till, B., & Comai, L. (2004). TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiology (Bethesda), 135*(2), 630-636.

Hess, J., Kohl, T., Kotrová, M., Rönsch, K., Paprotka, T., Mohr, V., . . . Paust, N. (2020). Library preparation for next generation sequencing: A review of automation strategies. *Biotechnology Advances, 41*, 107537.

Hixson, S., Parrish, C., & Anderson, D. (2014). Full substitution of fish oil with camelina (*Camelina sativa*) oil, with partial substitution of fish meal with camelina meal, in diets for farmed Atlantic salmon (*Salmo salar*) and its effect on tissue lipids and sensory quality. *Food Chemistry, 157*, 51-61.

Hixson, S. M., Parrish, C. C., Wells, J. S., Winkowski, E. M., Anderson, D. M., & Bullerwell, C. N. (2016). Inclusion of camelina meal as a protein source in diets for farmed salmonids. *Aquaculture Nutrition*, *22*(3), 615-630.

Hixson, S. M., Parrish, C. C., Wells, J. S., Marie Winkowski, E., & Anderson, D. M. (2016). Inclusion of camelina meal as a protein source in diets for farmed Atlantic cod *Gadus morhua*. *Aquaculture Research*, *47*(8), 2607-2622.

Huang, A. (2018). Plant lipid droplets and their associated proteins: Potential for rapid advances. *Plant Physiology (Bethesda), 176*(3), 1894-1918.

Huang, S., Li, R., & Vossen, van der, E.A.G. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics, 41*(12), 1275-1281.

Hunter, J. & Roth, G. (2010). Camelina production and potential in Pennsylvania, Agronomy Facts 72. College of Agricultural Sciences, Crop and Soil Sciences, Pennsylvania State University.

Hutcheon, C., Ditt, R.F., Belistein, M., Comai, L., Schroeder, J., Goldstein, E., Shewmaker, C.K., Nguyen, T., Rocher, J.D., Kiser, J. (2010). Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. *BMC Plant Biology, 10*(1), 233.

Ibrahim, F.M., & El Habbasha, S.F. (2015). Chemical composition, medicinal impacts and cultivation of camelina (*Camelina sativa*): Review. *International Journal of Pharmtech Research, 8*(10), 114-122.

Illumina sequencing technology. (n.d.). Retrieved March 18, 2022, from http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Innes, J., & Calder, P. (2018) Omega-6 fatty acids and inflammation. *Prostaglandins Leukotrienes and Essential Fatty Acids, 132,* 41-48.

Jiang, W., Henry, I., Lynagh, P., Comai, L., Cahoon, E., & Weeks, D. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnology Journal, 15*(5), 648-657.

Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., & Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. BMC Bioinformatics, 20(1), 556.

Kafetzopoulou, L., Pullan, S., Lemey, P., Suchard, M., Ehichioya, D., Pahlmann, M., . . . Duraffour, S. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science (American Association for the Advancement of Science), 363*(6422), 74-77.

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R. . . . Parkin, I. A. P. (2014). The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature Communications, 5*(1), 3706.

Kagale, S., Nixon, J., Khedikar, Y., Pasha, A., Provart, N., Clarke, W., . . . Parkin, I. (2016). The developmental transcriptome atlas of the biofuel crop *Camelina sativa*. *The Plant Journal: For Cell and Molecular Biology, 88*(5), 879-894.

Kang, J., Snapp, A., & Lu, C. (2011). Identification of three genes encoding microsomal oleate desaturases (FAD2) from the oilseed crop *Camelina sativa*. *Plant Physiology and Biochemistry, 49*(2), 223-229.

Karg, S. (2012). Oil-rich seeds from prehistoric contexts in southern Scandinavia - Reflections on archaeobotanical records of flax, hemp, gold of pleasure, and corn spurrey. *Acta Palaeobotanica, 52*(1), 17-24.

Kaur, N., Chugh, V., & Gupta, A. (2012). Essential fatty acids as functional components of foods- a review. *Journal of Food Science and Technology, 51*(10), 2289-2303.

Keng, J., & Hall, J. (1987). The effects of guard rows on variety test results of maize. *Agricultural Systems, 23*(3), 187-195.

Kibbe, W.A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res. 35*.

Kim, H., Hsieh, K., Ratnayake, C., & Huang, A. (2002). A novel group of oleosins is present inside the pollen of Arabidopsis. *The Journal of Biological Chemistry, 277*(25), 22677-22684.

Kim, S., Yamaoka, Y., Ono, H., Kim, H., Shim, D., Maeshima, M., . . . Lee, Y. (2013). AtABCA9 transporter supplies fatty acids for lipid synthesis to the endoplasmic reticulum. *Proc Natl Acad Sci USA, 110*(2), 773-778.

King, K., Li, H., Kang, J., & Lu, C. (2019). Mapping quantitative trait loci for seed traits in *Camelina sativa*. *Theoretical and Applied Genetics, 132*(9), 2567-2577.

Knorzer, K.H. (1978) Evolution and spread of Gold of Pleasure (*Camelina sativa* S.L.). *Berichte der Deutschen Botanischen Gesellschaft 91*, 187-195.

Knutsen, H. K., Alexander, J., Barregård, L., Bignami, M., Brüschweiler, B., Ceccatelli, S., … Vleminckx, C. (2016). Erucic acid in feed and food. *EFSA Journal*, *14*(11). https://doi.org/10.2903/j.efsa.2016.4593

Kumawat, S. T., Rana, N. K., Bansal, R., Sonah, H., Raj Sharma, T., Deshmukh, R., . . . Kumar Yadav, S. (2019). Expanding avenue of fast neutron mediated mutagenesis for crop improvement. *Plants, 8*(6).

Kwolek-Mirek, M. & Zadrag-Tecza, R. (2014). Comparison of methods used for assessing the viability of yeast cells. *FEMS Yeast Research, 14*(7), 1068-1079.

Larsson, M. (2013). Cultivation and processing of *Linum usitatissimum* and *Camelina sativa* in southern Scandinavia during the Roman Iron Age. *Vegetation History and Archaeobotany, 22*(6), 509-520.

Laurent, P., & Huang, A. (1992). Organ- and development-specific acyl coenzyme a lysophosphatidate acyltransferases in palm and meadowfoam. *Plant Physiology (Bethesda), 99*(4), 1711-1715.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M., Arondel, V., Bates, P., . . . Ohlrogge, J. (2013). Acyl-Lipid Metabolism. *The Arabidopsis Book, 2013*(12), E0161.

Li, G., Chern, M., Jain, R., Martin, J.A., Schackwitz, W.S., Jiang, L., Vega-Sanchez, M.E., Lipzen, A.M., Barry, K.W. & Schmutz, J. (2016) Genome-wide sequencing of 41 rice (*Oryza sativa L.*) mutated lines reveals diverse mutations induced by fast-neutron irradiation. *Molecular plant 9*, 1078-1081.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics, 34*(18), 3094-3100.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics, 26*(5), 589-595.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2

Li, N., Gügel, I.L., Giavalisco, P., Zeisler, V., Schreiber, L., Soll, J. & Philippar, K. (2015). FAX1, a novel membrane protein mediating plastid fatty acid export. *PLoS Biology, 13*(2).

Li, N., Qi, G., Sun, X., Wang, D., Bean, S., & Blackwell, D. (2014). Isolation and characterization of protein fractions isolated from camelina meal. *Transactions of the ASABE, 57*(1), 169-178.

Li, P., Bhattarai, S., Peterson, G., Coulman, B., Schellenberg, M., Biligetu, B., & Fu, Y. (2018). Genetic diversity of northern wheatgrass (Elymus lanceolatus ssp. lanceolatus) as revealed by genotyping-by-sequencing. *Diversity (Basel), 10*(2), 23.

Li, X. (2011). Pollen fertility/viability assay using FDA staining. *Bio-protocol, 1*(10). https://doi.org/10.21769/BioProtoc.75

Li, X., & Mupondwa, E. (2016). Production and value-chain integration of *Camelina Sativa* as a dedicated bioenergy feedstock in the Canadian prairies. *European Biomass Conference and Exhibition Proceedings, 2016*(24), 151-157.

Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., . . . Zhang, Y. (2001). A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant Journal, 27*(3), 235-242.

Lin, L., & Tzen, J. (2004). Two distinct steroleosins are present in seed oil bodies. *Plant Physiology and Biochemistry, 42*(7), 601-608.

Lisa, M., & Holcapek, M. (2008). Triacylglycerols profiling in plant oils important in food industry, dietetics and cosmetics using high-performance liquid chromatography–atmospheric pressure chemical ionization mass spectrometry. *Journal of Chromatography A, 1198*(1-2), 115-130.

Liu, X., Brost, J., Hutcheon, C., Guilfoil, R., Wilson, A., Leung, S., . . . De Rocher, J. (2012). Transformation of the oilseed crop *Camelina sativa* by Agrobacterium-mediated floral dip and simple large-scale screening of transformants. *In Vitro Cellular & Developmental Biology. Plant, 48*(5), 462-468.

Los, D., & Murata, N. (1998). Structure and expression of fatty acid desaturases. *Biochimica et Biophysica Acta (BBA)/Lipids and Lipid Metabolism, 1394*(1), 3-15.

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.

Lu, C., Xin, Z., Ren, Z., Miquel, M., & Browse, J. (2009). Enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. *Proceedings of the National Academy of Sciences - PNAS, 106*(44), 18837-18842.

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

Mandáková, T., Pouch, M., Brock, J., Al-Shehbaz, I., & Lysak, M. (2019). Origin and evolution of diploid and allopolyploid camelina genomes were accompanied by chromosome shattering. *The Plant Cell, 31*(11), 2596-2612.

Mazoochian, L., Mohammad Sadeghi, H., & Pourfarzam, M. (2018). The effect of FADS2 gene rs174583 polymorphism on desaturase activities, fatty acid profile, insulin resistance, biochemical indices, and incidence of type 2 diabetes. *Journal of Research in Medical Sciences, 23*(1), 47–47. https://doi.org/10.4103/jrms.JRMS_961_17

Metzker, M. (2010). Sequencing technologies - the next generation. *Nature Reviews. Genetics, 11*(1), 31-46.

Michael, T., & Jackson, S. (2013). The first 50 plant genomes. *The Plant Genome, 6*(2), 1-7.

Midha, M., Wu, M., & Chiu, K. (2019). Long-read sequencing in deciphering human genetics to a greater depth. *Human Genetics, 138*(11-12), 1201-1215.

Miklaszewska, M., & Banaś, A. (2016). Biochemical characterization and substrate specificity of jojoba fatty acyl-CoA reductase and jojoba wax synthase. *Plant Science (Limerick), 249*, 84-92.

Miklaszewska, M., Zienkiewicz, K., Inchana, P., & Zienkiewicz, A. (2021). Lipid metabolism and accumulation in oilseed crops. *Oléagineux Corps Gras Lipides, 28*, 50.

Miquel, M., & Browse, J. (2017). Lipid biosynthesis in developing seeds. In *Seed Development and Germination* (pp. 169-193).

Mišurcová, L., Ambrožová, J., & Samek, D. (2011). Seaweed Lipids as Nutraceuticals. *Advances in Food and Nutrition Research*, *64*, 339–355. https://doi.org/10.1016/B978-0-12-387669-0.00027-2

Moser, B. (2010). Camelina (*Camelina sativa* L.) oil as a biofuels feedstock: Golden opportunity or false hope? *Lipid Technology, 22*(12), 270-273.

Moser, B., & Vaughn, F. (2010). Evaluation of alkyl esters from *Camelina sativa* oil as biodiesel and as blend components in ultra-low-sulfur diesel fuel. *Bioresource Technology, 101*, 646-653. https://doi:10.1016/j.biotech.2009.08.054

Myllykangas, S., & Ji, H. (2010). Targeted deep resequencing of the human cancer genome using next-generation technologies. *Biotechnology & Genetic Engineering Reviews, 27*(1), 135-158.

Nagarajan, N., & Pop, M. (2013). Sequence assembly demystified. *Nature Reviews Genetics*, *14*(3), 157-167.

Næsted, H., Frandsen, G., Jauh, G., Hernandez-Pinzon, I., Nielsen, H., Murphy, D., . . . Mundy, J. (2000). Caleosins: Ca2 -binding proteins associated with lipid bodies. *Plant Molecular Biology, 44*(4), 463-476.

Neumann, N., Nazarenus, T., Aznar-Moreno, J., Rodriguez-Aponte, S., Mejias Veintidos, V., Comai, L., . . . Cahoon, E. (2021). Generation of camelina mid-oleic acid seed oil by identification and stacking of fatty acid biosynthetic mutants. *Industrial Crops and Products, 159*(C), 113074.

Nowak, M., Boerlijst, M., Cooke, J., & Smith, J. (1997). Evolution of genetic redundancy. *Nature (London), 388*(6638), 167-171.

Nguyen, H., Silva, J., Podicheti, R., Macrander, J., Yang, W., Nazarenus, T., . . . Cahoon, E. (2013). Camelina seed transcriptome: A tool for meal and oil improvement and translational research. *Plant Biotechnol J, 11*(6), 759-769.

Ohlrogge, J., & Jaworski, J. (1997). Regulation of fatty acid synthesis. *Annual Review of Plant Biology, 48*(1), 109-136.

O'Neill, C., Baker, D., Bennett, G., Clarke, J., & Bancroft, I. (2011). Two high linolenic mutants of *Arabidopsis thaliana* contain megabase-scale genome duplications encompassing the FAD3 locus. *The Plant Journal: For Cell and Molecular Biology, 68*(5), 912-918.

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 Biochemistry 123*(C), 1–7.

Pathirana, R. (2011). Plant mutation breeding in agriculture. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 6*(32), 1-20.

Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics (Oxford, England), 35*(13), 2193-2198.

Pecchia, P., Russo, R., Brambilla, I., Reggiani, R., & Mapelli, S. (2014). Biochemical seed traits of *Camelina sativa*–an emerging oilseed crop for biofuel: environmental and genetic influences. *Journal of Crop Improvement*, *28*(4), 465-483.

Perera, S. P., McIntosh, T., Coutu, C., Tyler, R. T., Hegedus, D. D., & Wanasundara, J. P. (2022). Profiling and characterization of *Camelina sativa* (L.) Crantz meal proteins. *Journal of the American Oil Chemists' Society.*

Perry, H., & Harwood, J. (1993). Changes in the lipid content of developing seeds of *Brassica napus*. *Phytochemistry (Oxford), 32*(6), 1411-1415.

Peterson, G., Dong, Y., Horbach, C., & Fu, Y. (2014). Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. *Diversity, 6*(4), 665-680.

Pietras, M., & Orczewska-Dudek, S. (2013). The effect of dietary *Camelina sativa* oil on quality of broiler chicken meat. *Annals of Animal Science, 13*(4), 869-882.

Plant Fatty Acid Synthesis. (n.d.). Retrieved September 27, 2022, from https://lipidlibrary.aocs.org/chemistry/physics/plant-lipid/plant-fatty-acid-synthesis

Pritchard, C., Cheng, H., & Tewari, M. (2012). MicroRNA profiling: Approaches and considerations. *Nature Reviews. Genetics, 13*(5), 358-369.

Poland, J., & Rife, T. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome, 5*(3), 92-102.

Pollard, M., Gurdasani, D., Mentzer, A., Porter, T., & Sandhu, M. (2018). Long reads: Their purpose and place. *Human Molecular Genetics, 27*(R2), R234-R241.

Pompéia, C., Lopes, L., Miyasaka, C., Procópio, J., Sannomiya, P., & Curi, R. (2000). Effect of fatty acids on leukocyte function. *Brazilian Journal of Medical and Biological Research, 33*(11), 1255-1268.

Purkrtova, Z., D'Andrea, S., Jolivet, P., Lipovova, P., Kralova, B., Kodicek, M., & Chardot, T. (2007). Structural properties of caleosin: A MS and CD study. *Archives of Biochemistry and Biophysics, 464*(2), 335-343.

Putnam D.H., Budin, J.T., Field, L.A., Breene W.M. (1993) Camelina: a promising low-input oilseed. In: New Crops (eds: Jacick, J. and Simon, J.E.), Wiley, New York, 1993; 314-322.

Qu, J., Mao, H., Chen, W., Gao, S., Bai, Y., Sun, Y., . . . Ye, J. (2012). Development of marker-free transgenic Jatropha plants with increased levels of seed oleic acid. *Biotechnology For Biofuels, 5*(1), 10.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radatz, W., & Hondelmann, W. (1981). Samenolhaltige pflanzen der Wildflora als potentielle Nutzpflanzen fur die Gewinnung von Industriegrundstoffen Literaturubersicht und Zielsetzung. *Landbauforschung Völkenrode, 31*(4), 227-240.

Rodríguez-Rodríguez, M., Sánchez-García, A., Salas, J., Garcés, R., & Martínez-Force, E. (2013). Characterization of the morphological changes and fatty acid profile of developing *Camelina sativa* seeds. *Industrial Crops and Products, 50*, 673-679.

Sahlin, K., Sipos, B., James, P., & Medvedev, P. (2021). Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nature Communications, 12*(1), 2-13.

Salas, J., Martínez-Force, E., & Garcés, R. (2005). Very long chain fatty acid synthesis in sunflower kernels. *Journal of Agricultural and Food Chemistry, 53*(7), 2710-2716.

Salas, J.J., & Ohlrogge, J.B. (2002). Characterization of substrate specificity of plant FATA and FATB acyl-ACP thioesterases. *Archives of Biochemistry and Biophysics, 403*(1), 25–34.

Sakai, K., Suzuki, S., Nakamura, N., & Okada, S. (1987). Induction and subsequent repair of DNA damage by fast neutrons in cultured mammalian cells. *Radiation Research, 110*(3), 311-320.

Sanger, F., Nicklen, S., & Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences – PNAS, 74*(12), 5463-5467.

Sarvas, C., Puttick, D., Forseille, L., Cram, D., & Smith, M. (2021). Ectopic expression of cDNAs from larkspur and its positional redistribution in seed triacylglycerol of *Camelina sativa*. *Planta, 254*(2).

Sequencing platforms. (n.d.). Retrieved March 18, 2022, from https://www.illumina.com/systems/sequencing-platforms.html

Shaul, O. (2017). How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology, 91*(Pt B), 145–155. https://doi.org/10.1016/j.biocel.2017.06.016

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology, 26*(10), 1135-1145.

Shimada, T., Hayashi, M., & Hara-Nishimura, I. (2018). Membrane dynamics and multiple functions of oil bodies in seeds and leaves. *Plant Physiology (Bethesda), 176*(1), 199-207.

Shukla, R., Sharma, D., Pathak, N., & Bajpai, P. (2018). Genomic DNA isolation from high polyphenolic content *Grewia asiatica* L. leaf without using liquid nitrogen. *Iranian Journal of Science and Technology. Transaction A, Science, 42*(2), 347-351.

Sikora, P., Chawade, A., Larsson, M., Olsson, J. & Olsson, O. (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *International journal of plant genomics 2011*.

Siloto, R., Findlay, K., Lopez-Villalobos, A., Yeung, E., Nykiforuk, C., & Moloney, M. (2006). The accumulation of oleosins determines the size of seed oil bodies in Arabidopsis. *The Plant Cell, 18*(8), 1961-1974.

Simon, J.A., Fong, J., Bernert Jr., J.T. & Browner, W.S. (1995). Serum fatty acids and the risk of stroke. *Stroke 26*(5): 778-782

Simopoulos, A. (2006). Evolutionary aspects of diet, the omega-6/omega-3 ratio and genetic variation: Nutritional implications for chronic diseases. *Biomedicine & Pharmacotherapy, 60*(9), 502-507.

Simopoulos, A. (2008). The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Experimental Biology and Medicine, 233*(6), 674-688.

Sims, D., Sudbery, I., Ilott, N., Heger, A., & Ponting, C. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews. Genetics, 15*(2), 121-132.

Slatko, B., Gardner, A., & Ausubel, F. (2018). Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology, 122*(1), E59.

Smith, M., Cross, A., Jones, O., Griffiths, W., Stymne, S., & Stobart, K. (1990). Electron-transport components of the 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine Δ12-desaturase (Δ12-desaturase) in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons. *Biochemical Journal, 272*(1), 23-29.

Snyder, C., Yurchenko, O., Siloto, R., Chen, X., Liu, Q., Mietkiewska, E., & Weselake, R. (2009). Acyltransferase action in the modification of seed oil biosynthesis. *New Biotechnology, 26*(1), 11-16.

Sørensen, B., Furukawa-Stoffer, T., Marshall, K., Page, E., Mir, Z., Forster, R., & Weselake, R. (2005). Storage lipid accumulation and acyltransferase action in developing flaxseed. *Lipids, 40*(10), 1043-1049.

Sun, J., Guo, X., & Smith, M. A. (2017). Identification of crepenynic Acid in the Seed Oil of *Atractylodes lancea* and *A. macrocephala*. *Journal of the American Oil Chemists' Society*, *94*(5), 655–660. https://doi.org/10.1007/s11746-017-2974-2

Talebi, Ali Benjavad, Talebi, Amin Benjavad, & Shahrokhifar, Behzad. (2012). Ethyl methane sulphonate (EMS) induced mutagenesis in Malaysian rice (cv. MR219) for lethal dose determination. *American Journal of Plant Sciences, 03*(12), 1661-1665.

Tanka, K., Gajendiran, N., Endo, S., Komastu, K., Hoshi, M. & Kamada, N. (1999) Neutron energy-dependent initial DNA damage and chromosomal exchange. *Journal of radiation research 40*, S36-S44.

Taq DNA polymerase, recombinant (5 U/ML). (n.d.). Retrieved October 5, 2022, from https://www.thermofisher.com/order/catalog/product/EP0402

Taylor, D., Weber, N., Barton, D., Underhill, E., Hogge, L., Weselake, R., & Pomeroy, M. (1991). Triacylglycerol bioassembly in microspore-derived embryos of Brassica napus L. cv Reston. *Plant Physiology (Bethesda), 97*(1), 65-79.

Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science (American Association for the Advancement of Science), 327*(5967), 818-822.

Toncea, I. (2014). The seed yield potential of camelia - first Romanian cultivar of camelina (*Camelina sativa* L. Crantz). *Romanian Agricultural research, 31*(31), 17-23.

Troncoso-Ponce, M., Barthole, G., Tremblais, G., To, A., Miquel, M., Lepiniec, L., & Baud, S. (2016). Transcriptional activation of two Delta-9 palmitoyl-ACP desaturase genes by MYB115 and MYB118 is critical for biosynthesis of omega-7 monounsaturated fatty acids in the endosperm of Arabidopsis seeds. *The Plant Cell, 28*(10), 2666-2682.

Truseq DNA Nano. (n.d.). Retrieved September 28, 2022, from https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-nano-dna.html

Tsuchimatsu, T., Kakui, H., Yamazaki, M., Marona, C., Tsutsui, H., Hedhly, A., . . . Shimizu, K. (2020). Adaptive reduction of male gamete number in the selfing plant *Arabidopsis thaliana*. *Nature Communications, 11*(1), 2885.

Van Dijk, E., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The third revolution in sequencing technology. *Trends in Genetics, 34*(9), 666-681.

Vazquez, L., Corzo-Martinez, M., Arranz-Martinez, P., Barroso, E., Reglero, G., & Torres, C. (2019). Bioactive lipid. In *Bioactive molecules in foods*. 467-527.

Vetter, W., Darwisch, V., & Lehnert, K. (2020). Erucic acid in Brassicaceae and salmon – An evaluation of the new proposed limits of erucic acid in food. *NFS Journal*, *19*, 9–15. https://doi.org/10.1016/j.nfs.2020.03.002

Voelker, T., & Kinney, A. (2001). Variations in the biosynthesis of seed-storage lipids. *Annual Review of Plant Biology, 52*, 335-361.

Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T. (2005). Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. *Plant Breeding,* 124(5), 446-453.

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*, 270-277.

Walia, M., Zanetti, F., Gesch, R., Krzyżaniak, M., Eynck, C., Puttick, D., . . . Monti, A. (2021). Winter camelina seed quality in different growing environments across Northern America and Europe. *Industrial Crops and Products, 169*, 113639.

Wang, C., & Xu, Y. (2010). The 5′ untranslated region of the FAD3 mRNA is required for its translational enhancement at low temperature in Arabidopsis roots. *Plant Science (Limerick), 179*(3), 234-240.

Wang, P., Li, X. & Cahoon, E.B. (2012) Chapter 24: biotechnological improvement of soybean oil for lubricant applications. In *Synthetics, Mineral Oils, and Bio-Based Fluids*, 2nd edn (L. Rudnick, ed.), pp. 413– 418. Boca Raton: CRC Press.

Wang, Q., Zhang, X., Li, C., Liu, Z., & Feng, H. (2014). Directional transfer of a multiple-allele male sterile line in *Brassica campestris* L. ssp. *chinensis* (L.) Makino var. *rosularis* Tsen et Lee. *Breeding Science, 64*(2), 149-155.

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology, 39*(11), 1348-1365.

Waterworth, W.M., Bray, C. M., & West, C. E. (2019). Seeds and the art of genome maintenance. *Frontiers in Plant Science, 10*, 706–706. https://doi.org/10.3389/fpls.2019.00706

Wyant, S. R., Rodriguez, M. F., Carter, C. K., Parrott, W. A., Jackson, S. A., Stupar, R. M. & Morrell, P. L. (2020). Fast neutron mutagenesis in soybean creates frameshift mutations. In *Life Science Weekly* (p. 1770).

Yan, H., Bekele, W., Wight, A., Peng, C., Langdon, P., Latta, Y., . . . Tinker, J. (2016). High-density marker profiling confirms ancestral genomes of Avena species and identifies D-genome chromosomes of hexaploid oat. *Theoretical and Applied Genetics, 129*(11), 2133-2149.

Yuan, L., & Li, R. (2020). Metabolic engineering a model oilseed *Camelina sativa* for the sustainable production of high-value designed oils. *Frontiers in Plant Science, 11*, 11.

Zeng, F., Roslinsky, V., & Cheng, B. (2017). Mutations in the promoter, intron and CDS of two FAD2 generate multiple alleles modulating linoleic acid level in yellow mustard. *Scientific Reports, 7*(1), 8284-13.

Zeng, P., Tian, Z., Han, Y., Zhang, W., Zhou, T., Peng, Y., . . . Cai, J. (2022). Comparison of ONT and CCS sequencing technologies on the polyploid genome of a medicinal plant showed that high error rate of ONT reads are not suitable for self-correction. *Chinese Medicine, 17*(1), 1-94.

Zheng, H., Rowland, O., & Kunst, L. (2005). Disruptions of the Arabidopsis enoyl-CoA reductase gene reveal an essential role for very-long-chain fatty acid synthesis in cell expansion during plant morphogenesis. *The Plant Cell, 17*(5), 1467-1481.

Zhuang, P., Shou, Q., Wang, W., He, L., Wang, J., Chen, J., Zhang, Y., & Jiao, J. (2018). Essential fatty acids linoleic acid and α-linolenic acid sex-dependently regulate glucose homeostasis in obesity. *Molecular Nutrition & Food Research*, *62*(17), e1800448–n/a. https://doi.org/10.1002/mnfr.201800448

Zienkiewicz, K., & Zienkiewicz, A. (2020). Degradation of lipid droplets in plants and algae—Right time, many paths, one goal. *Frontiers in Plant Science, 11*, 579019.

Zubr, J. (2003). Dietary fatty acids and amino acids of *Camelina sativa*. *Journal of Food Quality, 26*(6), 451-462.

Appendices

Appendix A. Riptide rapid DNA library prep in 96 well format

Genomic DNA was extracted using the CTAB method, quantified using a fluorometer, and diluted to 50ng/µl for storage in -20 ˚C. The DNA sample was diluted to 12.5 ng/µl immediately before use. The whole library preparation was separated into 3 three days.

The first day was primer extension and termination: the "A" reaction. To prepare A reaction master mix for 100 reactions, Eppendorf tube was used to hold 4 µl of DNA sample at 17 ng/µl, and 200 µl dNTP Mix 1, 100 µl 10x Enzyme 1 Buffer, and 100 µl Enzyme 1 (kept at -20 ˚C until needed) were added. The mixture of the liquids was A reaction master mix, and 50 µl of this solution was put in each well of 8 strip tube. Eight-channel multichannel was used to pipette 4 µl of the A reaction master mixed into each well of the 96-well plate (non-skirted or skirted plate ok with MSB1001 seal). The plate was kept on ice during this and subsequent steps. The next step was pipetting 2 µl of primer of choice (low or high GC-use low and high GC for Camelina in 1:1 ratio because Cs is 49.4% GC, considered intermediate if 40-60% GC) into each well of the plate, and made sure the orientation was correct, i.e.: from A1 to A1. The DNA sample plated was incubated at 98 ˚C for 1 minute in a skirted plate in a PCR machine and set on ice for 2-3 minutes (can skip this step if only working with low GC samples). Briefly spined the DNA samples in a plate centrifuge to collect the contents at the bottom of each tube or well. Then, DNA samples (4 µl at 17 ng/µl, total 68 ng) were pipetted carefully into their corresponding wells. The plate was then put on a PCR machine with MSB 1001. The PCR program was set at 92 ˚C for 3 minutes, 16 ˚C for 5 minutes, slow ramp (0.1 ˚C /sec) to 68 ˚C and stayed for 15 minutes, and finally hold at 4 ˚C. After running the PCR program, slowly ripped off PCR tape slowly and pooled it in a 2 ml tube on ice. Ensured to spin down the plate before pooling. This reaction can be stored temporarily at 4 ˚C overnight or frozen at -20 ˚C for the long term. Kept 1 µl of a library to put on Qubit-this was the starting material.

On the second day, the SPRI Beads I tube should be removed from the fridge at first and placed on the bench for 20-30 minutes to warm up to room temperature. Shaked the tube thoroughly to create a homogeneous solution before use. After removing the SPRI Beads I tube from the fridge, the experiment can start. A 1.5 ml Eppendorf tube was put on ice and added 96 µl of 150mM EDTA to the 1.5 ml Eppendorf tube for 96 samples (the same volume of EDTA to the same number of samples e.g., if there are 48 samples, add 48 µl). Mixed by inverting the tube or pipetting up and down a few times followed by briefly spinning. The volume of the sample pool was measured with a pipette and top up with water to 1000 µl (by weight). Added 1.8 volume (~total 2.8 ml, use 15 ml tube) of well resuspended, room temperature SPRI Beads I. Pipetted up and down well to mix then split equally into 2 tubes, incubated at room temperature for 10 minutes. Placed the tube on a magnetic stand. Allowed the solutions to clear (8 minutes), discarded supernatants, did not disturb the beads. Then made 6 ml of fresh 80% ethanol, by 5 ml of 95% ethanol and 1ml of water and added 850 µl x2 tubes of freshly prepared 80% ethanol to each

tube. Pipetted up and down to resuspend, put on the magnetic rack, discarded the supernatant. Another 850 µl x2 tube of freshly prepared 80% ethanol. Pipetted up and down to resuspend and pooled back into a 2 ml tube, put on the magnetic rack, discarded the supernatant. Ensured that all residual ethanol was removed after this wash. Open the caps of the tubes while they were on the magnetic stand and allowed the beads to air dry for 10 minutes and added 100 µl room-temperature 10 mM Tris-HCL (PH 8.0)-supplied to the beads in 2 ml tubes. Removed the tubes from the magnetic stand and suspended the beads in the liquid by pipetting up and down until the solution was homogeneous. Allowed the tubes to sit at room temperature for 10 minutes. The tube was placed on a magnetic stand, allowing the solutions to clear and transferred the supernatants containing the eluted DNA to a new 0.2 ml tube (to fit a PCR machine). The volume in the tube should be approximately 100 µl and the spent beads should be discarded. Halfway through the second day was accomplished, at this time, the A reaction can be stored temporarily at 4 ˚C or on ice, or frozen at -20 ˚C for long term. Then what should do was the library conversion and B reaction. Before continuing the second-day experiment, the sample should be heated from A reaction at 95 ˚C for 3 minutes and transferred to ice (or can program PCR machine to 95 ˚C for 3 minutes then hold at 4 ˚C). While heating the DNA, prepared the Capture beads: shaked the Capture Beads tube (stored at 4 ˚C) thoroughly to resuspend and transfer 40ul of the beads to a new 1.5 ml Eppendorf DNA tube. Placed the tube on a magnetic stand and waited for the solution to clear (3 min). Carefully removed the supernatant and discarded it. Removed the tube from the magnetic stand and added 200 µl of room temperature HS Buffer to beads. Pipetted the sample up and down to mix the component and returned the tube to the magnetic stand. Waited for the solution to clear. Carefully removed and discarded the supernatant. Removed the tube from the stand and resuspended the beads in 40 µl of HS Buffer by pipetting up and down a few times. Added the heat-denatured sample from the A reaction to the prepared Capture Beads mix prepared in the previous step and incubated the sample at room temperature (or heating block set to 24 ˚C) for 10 minutes. Mixed the sample again by pipetting up and down a few times and incubating for another 10 minutes at room temperature. Placed the tube on the magnetic stand and waited for the solution to clear. Carefully removed and discarded the supernatant. Avoided disturbing the beads. Removed the tube from the stand and resuspend the beads in 100 µl of 0.1N sodium hydroxide, pipetted up and down, left on the bench for 3 minutes, returned tube to magnetic stand allow the solution to clear and discarded the supernatant. Removed the tube from the magnetic stand and resuspended the beads in 200 µL of room-temperature Bead Wash Buffer by pipetting up and down a few times. Returned the tube to the magnetic stand, allowed the solution to clear and discarded the supernatant. Repeated the wash step (previous step) two additional times. Carefully removed any remaining liquid after the final wash. For B reaction, 8 µl 5x Enzyme II buffer, 3 µl dNTP Mix II, 4 µl Primer B, and 24 µl Nuclease-Free water were added to the tube containing the Capture Beads. Pipetted the liquid in the tube up and down several times to resuspend the Capture Beads. Placed the tube on ice. Added 1 µl of Enzyme II to the reaction Mix Gently. Incubated the B reaction for 20 minutes at 24 ˚C, then transferred it back to the ice for 2-3 minutes. Placed the tube on the magnetic stand. Allowed the solution to clear and discarded the supernatant without disturbing the beads. If the solution did not clear, used a pipette to resuspend the beads in the liquid while the tube was on the magnetic stand. The solution should clear more effectively

now. Removed the tube from the magnetic stand and resuspended the beads in 200 µl of room-temperature Bead Wash Buffer by pipetting up and down a few times. Returned the tube to the magnetic stand, allowed the solution to clear and discard the supernatant. Repeated the wash step (previous step) two additional times. Carefully removed any remaining liquid after the final wash. Used a pipette to resuspend the beads in 42 µl of Nuclease-Free Water. Transferred the beads to a thin-walled PCR tube. Then 4 µl Universal PCR Primer, 4 µl Index PCR Primer (barcodes 1 - 12) (chose one barcoded primer), and 50 µl 2X PCR Amplification Mix were added into the PCR tube. The PCR program was set 1 cycle of 98 ˚C for 2 minutes, 8 cycles: 98 ˚C for 20 seconds, 60 ˚C for 30 seconds, and 72 ˚C for 30 seconds, 1 cycle of 72 ˚C for 5 minutes, and lastly, held at 4 ˚C. The sample can be left in the PCR machine at 4 ˚C overnight.

On the third day, the target was size selection. In the beginning, warmed SPRI beads II to room temperature for 20 minutes. In the meanwhile, briefly spined the PCR tube in a microcentrifuge to pellet the beads. Placed the tube on a magnetic stand, waited for the solution to clear and transferred the supernatant to a new 1.5 ml Eppendorf DNA tube. Discarded beads. Added 70 µl of room temperature (warmed at room temp for 20-30 min) well suspended SPRI beads II for 2x150 Illumina paired-end sequencing read. Pipetted 10 times to mix well, incubated at room temp for 10min (mixed the sample by pipetting up and down after 5 min). Placed the tube on a magnetic stand and allowed the tube to clear (8 min). The supernatant was transferred to a new 1.5 ml Eppendorf tube, discarded beads. Added 30 µl of room temperature (warmed at room temp for 20-30 min) well suspended SPRI beads II for 2x150 Illumina paired-end sequencing read. Pipetted 10 times to mix, incubated at room temp for 10min (mix the sample by pipetting up and down after 5 min). Placed the tube on a magnetic stand and allowed the tube to clear (5 min). Discarded supernatant without disturbing the beads. Can keep this supernatant and put more SPRI beads II on to get more out. Made 500ul of fresh 80% ethanol (421 µl of 95% ethanol and 79 µl water). Added 200 µl of freshly prepared 80% ethanol to tube, removed from the rack, pipetted up and down, waited 1 min, then removed and discarded the ethanol (left the tube on stand). Added another 200ul of freshly prepared 80% ethanol to the tube while on the stand, waited 1 min to have clear supernatant, then discarded the ethanol (left the tube on stand). Ensured all ethanol had been removed by using a smaller pipette tip to draw out the residual liquid. Opened the cap of the tube while on the stand, allowed the beads to air dry for 6 minutes. Added 25 µl of 37 ˚C 10 mM Tris-HCL pH=8 (warm to 37 ˚C for higher elution yield) to beads. Removed from stand and pipette to mix, fully resuspended the beads in the liquid. Allowed the tube to sit at 37 ˚C for 20min or longer. Placed tube on the magnetic stand, allowed the solution to clear and transferred the supernatant containing eluted DNA to new Eppendorf tube, avoiding transferring any beads. Kept 1 µl to get concentration using Qubit. Evaluated average peak size using Tape station and qubit for quantity. There should be at least 30 µl at 2 nM for the sequencing, which the average size was supposed to be 500-900 bp.

Appendix B. Genomic DNA Extraction from Nuclei V2

This protocol could be divided into two parts, nuclei isolation, genomic DNA isolation. For the first part nuclei isolation, if two samples were going to be applied, listed materials: cold 1 x HBS (100 ml) and wash buffer (250-300 ml), Cheesecloth (2 layers) and Miracloth (6 layers), 50 ml Falcon tubes, round-bottom centrifuge tubes, 20 ml CTAB extraction buffer, TE, Proteinase K (at 20 mg/ml), RNase A (at 100 mg/ml), Chloroform: IAA (24:1), Isopropanol (or 95% EtOH), and 70% Ethanol were required. The first step of this part was to set aside 1 ml of 1X HBS into a tube for every extracted sample and homogenized tissue in ice-cold 1X HBS in a kitchen blender at medium-low speed for ~40 seconds (for fresh tissue blend 2-3 g fresh leaf into 50 ml 1XHBS; for frozen tissue blend tube contents (2-4 g) into 60 ml 1XHBS). Then transferred the homogenate to an ice-cold 500 ml beaker. Added 20% Triton X-100 in 1XHBX and 0.15% β-ME to the beaker as well (Table b1). Gently swirled the contents with a magnetic stir bar for ~30 minutes on ice in the fume hood.

Table b1. Information shows how much volume of reagent needed for different volume of sample.

Sample Volume (ml)	**20% Triton X-100 in 1XHBX Needed (ml)**	**0.15% β-ME (µl)**
200	5	300
100	2.5	150
60	1.5	90
50	1.25	75
40	1	60

The following step was to filter the solution from the last step into 50 ml Falcon tubes through two layers of cheesecloth and one layer of Miracloth seated in a funnel. Got as much of the extract as possible by gently squeezing it with gloved hands. Pelleted the homogenate by centrifugation in a fixed-angle rotor at 2500 g at 4 ˚C for 20 minutes (based on the size of the genome 1800-3500 g can be used). Carefully discarded the supernatant fluid and added 1 ml of ice-cold wash buffer to each bottle. Gently resuspended the pellet with the assistance of a small soft paintbrush soaked in ice-cold wash buffer, added ~10 ml and then transferred to an ice-cold 50 ml Falcon tube. Filled the tube to ~40 ml with ice-cold wash buffer. Filtered to a new 50 ml tube through 2 layers Miracloth to remove intact cells and tissue residues. Pelleted the nuclei by centrifugation at 2500 g (again dependent on the genome size) 4 ˚C for 15 minutes in a swinging bucket rotor. Discarded the supernatant and resuspended the pellet in 1 ml wash buffer. Added additional cold wash buffer to 40 ml and repeated centrifugation. Washed the pellet 2-3 times by resuspension in wash buffer using a paintbrush followed by centrifugation at 2500 g 4 ˚C for 15 minutes. After the final wash, resuspended the pellet in 940 µl HBS (no Triton), 50 µl Proteinase K, 10 µl RNAse A with a paintbrush in a 50 ml falcon tube. At this stage, 10 µl of this can be examined using a hemocytometer at 100X magnification to ensure the isolation of intact nuclei. Proceeded with DNA extraction immediately or left the sample overnight at 4 ˚C, left in a falcon tube.

The second part of the protocol was genomic DNA isolation. Materials: 50 ml Falcon tubes, 20 ml CTAB: PVP extraction buffer, TE, Chloroform: IAA (24:1), Isopropanol (or 95% EtOH), 70% Ethanol were needed for preparing two samples. To start the second part, added 10 ml 3% CTAB:1% PVP buffer (add BME first) in a 50 ml centrifuge tube, mixed gently by inversion. Incubated at 65 ˚C for 1-2 hrs with occasional gentle inversion to mix tube contents. Added an equal volume of Chloroform: IAA (24:1) and mixed gently by inversion to form an Emulsion. Centrifuged at 3000 g in a benchtop centrifuge for 20 minutes to separate the phases using a slow ramp rate option (ramp rate set to 3). Transferred the aqueous phase using a wide bore transfer pipet to a new tube and repeated the Chloroform: IAA extraction. Ensured in a round-bottom centrifuge tube at the end of this step. Added 2 volumes of 95% EtOH and mixed by inversion. Incubated in -20 ˚C freezer (30 minutes at least, overnight optional). Centrifuged the sample at 5000g for 20 minutes slow ramp rate. Discarded the supernatant and washed the pellet with 1 ml 70% Ethanol and centrifuged again with a slow ramp rate. Removed the supernatant (tap on a paper towel for the last drop) and briefly dried the pellet at room temperature for a half-hour, right side up in the hood, circled the pellet. Dissolved the pellet in 200 µl TE. Transferred to an Eppendorf tube using a wide-bore pipette tip and quantified using Qubit and checked quality using spectrophotometer and sized by femtopulse. The library preparation was accomplished.

Table b2. Component and volume of 10x HB (homogenization Buffer).

Component	**Volume**
00.1 M Trizma Base	50 ml of 1M Tris pH 9.5
0.8 M KCl	133 ml of 3 M KCl
0.1 M EDTA	100 ml of 0.5 M EDTA
10 mM spermidine	0.726 g (solution is aqueous at 1g/ml) or from 3 M
10 mM spermine	1.741 g

Table b3. Component and volume of 20% Triton Solution.

Component	**Volume**
Triton X-100 concentrate	50 ml
Water	200 ml

Table b4. Component and volume of 1x HBS without β-mercaptoethanol.

Component	**Volume 100**	**Volume 200**	**Volume 400**
1/10th HB 10x	10 ml	20 ml	40 ml
0.5M Sucrose	17.1 g	34.2 g	68.4 g

Note. Make up to final with H_2O.

Table b5. Component and volume of Wash Buffer.

Component	**Volume 1000 ml**	**Volume 500 ml**	**Volume 200 ml**

1/10th HB 10x	100 ml	50 ml	20 ml
0.5 M Sucrose	171.05 g	85.5 g	34.2 g
0.5% (v/v) Triton X-100	25 ml	12.5 ml	5 ml
0.15% β-mercaptoethanol	1.5 ml	750 µl	300 µl

Note. Make up to final with H_2O.

Table b6. Reagent and volume of 3% CTAB 1% PVP extraction buffer.

Reagent	**Amount to add (for10 ml)**	**Final concentration**
Cetyltrimethyl ammonium bromide (CTAB) (10% in H_2O)	3 ml	3%
5 M NaCl	2.8 ml	28%
0.5 M EDTA (pH 8.0)	0.4 ml	4%
1 M Tris-Cl (pH 8.0)	1 ml	10%
Polyvinylpyrrolidone (PVP 4% stock)	2.5 ml	1%
β-Mercaptoethanol	0.02 ml	0.2%

Note. 4% PVP stock solution is 0.1 g PVP powder per 2.5 ml water; CTAB is 5 g CTAB powder per 50 ml.

Appendix C. Data of germination tests

Table c1. Germination test 1 (July 2019)

Dosage	24 hours				48 hours			
	G1	G2	G3	G4	G1	G2	G3	G4
0Gy	35	28	38	33	46	42	48	45
7.1Gy	31	42	29	28	46	47	47	47
13.8Gy	21	21	22	27	45	49	44	45
20.9Gy	27	27	26	24	45	48	50	43
28.6Gy	21	26	22	30	44	45	48	46
35.7Gy	24	25	21	21	45	49	49	45
42.4Gy	19	33	20	17	45	50	47	47
49.5Gy	24	30	12	17	50	44	48	46

Note. Each group contained 50 seeds in total.

Table c2. Germination test 2 (October 2020)

Dosage	24 hours			48 hours			72 hours		
	G1	G2	G3	G1	G2	G3	G1	G2	G3
0Gy	32	34	33	49	47	46	49	49	48
7.1Gy	24	21	30	47	47	48	49	49	49
13.8Gy	20	21	26	47	46	47	48	49	49
20.9Gy	19	21	23	45	47	47	49	49	49
28.6Gy	16	16	18	44	45	46	48	48	48
35.7Gy	17	15	24	44	45	49	49	49	49
42.4Gy	17	18	18	44	47	48	50	49	49
49.5Gy	11	14	15	47	43	47	47	50	49

Note. Each group contained 50 seeds in total.

Table c3 Germination test 3 (September 2021)

Dosage	24 hours			48 hours			72 hours		
	G1	G2	G3	G1	G2	G3	G1	G2	G3
0Gy	36	37	34	47	47	46	47	48	48
7.1Gy	35	38	38	46	47	48	48	49	49
13.8Gy	35	34	35	45	47	48	47	47	48
20.9Gy	37	35	36	48	48	48	50	48	48
28.6Gy	37	35	36	49	46	46	49	46	48
35.7Gy	38	27	36	48	45	46	49	45	48
42.4Gy	28	33	34	46	47	48	46	48	48

49.5Gy	35	31	32	49	46	48	50	48	49

Note. Each group contained 50 seeds in total.

Table c4 Germination test 4 (July 2022)

Dosage	24 hours			48 hours			72 hours		
	G1	G2	G3	G1	G2	G3	G1	G2	G3
0Gy	38	37	32	48	48	47	49	49	48
7.1Gy	28	32	30	49	46	46	49	46	47
13.8Gy	28	22	21	49	47	44	50	47	44
20.9Gy	23	24	20	44	46	43	44	47	43
28.6Gy	24	25	23	47	44	48	48	44	48
35.7Gy	25	23	22	43	47	46	44	47	47
42.4Gy	22	23	22	39	44	39	44	44	45
49.5Gy	20	18	20	40	34	40	45	46	44

Note. Each group contained 50 seeds in total.

Table c5 Field test of plant establishment (July 2022)

Dosage	Plot1	Plot2
0Gy	55	55
7.1Gy	36	37
13.8Gy	32	38
20.9Gy	44	43
28.6Gy	34	41
35.7Gy	30	37
42.4Gy	22	32
49.5Gy	34	40
Midas	82	85

Note. Each group of dosage in one plot contain 300 seeds in total.

Appendix D. Fatty acid composition of mutant lines mentioned in the result sections

Table d1. Fatty acid composition (%) of M_3 seeds of lines contained reduced C20:1 content.

Line name	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	20:2	20:3	22:0	22:1	22:2	22:3	24:0	24:1
FN49-70-9	7.3	0.1	3.9	7.5	18.1	34.1	3.9	11.2	2.0	1.5	0.8	5.0	0.4	1.1	0.4	0.7
FN49-70-7	7.6	4.2	7.0	1.1	14.3	38.4	3.9	10.5	1.7	1.8	0.8	4.6	0.3	1.2	0.4	0.8
FN49-70-3	6.2	0.1	2.4	10.2	15.2	38.0	1.8	13.6	2.1	1.9	0.4	4.1	0.2	0.9	0.3	0.7
FN49-70-20	6.6	0.1	3.4	10.8	18.1	33.3	3.0	12.8	2.0	1.4	0.6	3.9	0.3	0.8	0.3	0.6
FN49-70-2	7.4	0.0	3.0	6.8	17.6	35.9	3.0	11.4	2.2	1.6	0.7	5.4	0.4	1.2	0.4	0.8
FN49-70-19	6.7	0.1	2.8	8.5	16.4	37.8	2.2	12.9	2.3	1.9	0.5	3.9	0.3	0.9	0.3	0.7
FN49-70-18	6.2	0.1	2.3	10.6	15.9	36.2	1.9	14.3	2.2	1.9	0.4	4.2	0.3	0.8	0.3	0.7
FN49-70-16	5.8	0.1	2.0	9.0	15.3	37.4	2.2	14.0	2.3	2.0	0.5	5.3	0.4	1.1	0.3	0.7
FN49-70-15	6.3	0.1	2.5	9.4	15.3	38.5	2.0	13.6	2.0	1.8	0.4	3.8	0.3	0.8	0.3	0.7
FN49-70-10	6.5	0.1	2.5	8.5	14.9	39.0	2.5	12.2	1.9	1.8	0.5	4.7	0.3	1.0	0.4	0.8
FN49-70-1	6.9	0.1	2.5	7.2	15.6	39.6	2.3	11.9	2.2	2.0	0.5	4.7	0.3	1.1	0.4	0.8
FN49-438-4	7.0	0.1	2.8	11.1	20.0	38.4	1.5	9.7	1.7	1.4	0.4	2.0	0.2	0.4	0.3	0.9
FN49-438-2	7.1	0.1	2.7	11.9	19.1	39.1	1.3	9.7	1.7	1.4	0.3	1.8	0.1	0.4	0.2	0.8
FN49-438-1	6.3	0.1	2.4	12.3	17.5	41.2	1.2	9.9	1.7	1.5	0.3	1.9	0.1	0.4	0.3	0.8
FN49-38-9	6.5	0.1	2.8	11.8	20.0	36.9	1.5	10.1	2.0	1.5	0.4	2.4	0.2	0.5	0.3	1.0
FN49-38-8	6.7	0.2	2.5	10.5	19.8	39.5	1.3	9.6	1.9	1.6	0.3	2.2	0.2	0.5	0.2	0.9
FN49-38-7	6.9	0.2	3.5	11.5	22.2	34.8	1.9	9.7	1.7	1.2	0.4	1.9	0.2	0.4	0.3	0.8
FN49-38-6	6.7	0.1	3.1	11.3	20.6	36.3	1.7	9.9	1.8	1.4	0.4	2.5	0.2	0.5	0.3	1.0
FN49-38-5	6.8	0.2	2.9	10.3	20.1	37.9	1.6	10.0	1.8	1.4	0.4	2.4	0.2	0.5	0.3	1.0
FN49-38-4	6.8	0.2	2.8	10.6	20.7	38.0	1.5	9.6	1.9	1.5	0.4	2.2	0.2	0.5	0.3	1.0
FN49-38-3	7.3	0.1	3.0	10.6	21.3	35.5	1.6	10.2	2.0	1.4	0.4	2.5	0.2	0.5	0.3	1.0

FN49-38-2	6.8	0.1	3.0	10.7	20.5	36.8	1.6	10.0	2.0	1.5	0.4	2.4	0.2	0.5	0.3	1.1
FN49-38-1	6.8	0.2	2.7	11.2	20.0	38.3	1.4	9.7	1.8	1.5	0.3	2.2	0.2	0.5	0.3	0.9
FN49-287-4	10.2	0.0	3.1	12.0	33.5	22.8	2.1	8.2	1.6	0.5	0.6	2.8	0.0	0.5	0.0	0.7
FN49-287-3	9.9	0.0	3.3	12.4	34.4	21.9	2.2	8.0	1.5	0.5	0.6	3.0	0.0	0.5	0.0	0.7
FN49-287-2	7.9	0.2	2.2	9.2	18.9	34.0	1.8	13.6	2.2	1.6	0.4	4.0	0.3	0.8	0.2	0.8
FN49-287-1	6.5	0.1	1.9	10.4	17.4	35.3	1.6	14.3	2.4	1.7	0.4	4.0	0.2	0.7	0.2	0.8
FN49-72-4	10.3	0.0	2.5	6.8	23.9	33.6	1.9	9.4	2.0	1.2	0.5	3.4	0.3	0.8	0.4	1.0
FN49-72-2	10.2	0.0	3.5	9.5	28.3	29.4	2.2	8.4	1.7	0.7	0.6	2.5	0.0	0.6	0.0	0.8
FN49-213-4	10.7	0.0	3.3	7.6	28.6	29.3	2.0	7.5	1.9	0.9	0.6	2.9	0.4	0.7	0.5	1.1
FN49-213-3	10.0	0.0	3.3	10.7	29.2	28.3	1.8	7.6	1.7	0.8	0.5	2.2	0.0	0.6	0.4	0.9
FN49-213-2	6.7	0.2	2.5	9.4	18.0	39.6	1.6	10.0	2.0	1.9	0.4	2.8	0.2	0.7	0.4	1.1
FN49-312-1	7.4	0.2	2.2	8.5	18.7	38.7	1.6	9.7	2.2	1.8	0.4	3.2	0.3	0.9	0.3	1.0
FN49-312-7	7.2	0.2	2.2	8.4	16.9	41.0	1.6	9.4	2.1	2.0	0.4	3.0	0.3	0.9	0.4	1.0
FN49-312-6	7.6	0.2	2.6	10.0	20.2	37.5	1.3	9.3	2.1	1.6	0.3	2.4	0.2	0.6	0.3	1.0
FN49-312-5	8.9	0.0	3.6	9.6	29.0	27.7	2.3	8.9	1.9	0.8	0.6	2.5	0.2	0.4	0.4	1.0
FN49-312-3	8.2	0.2	2.7	7.5	22.5	35.6	1.9	8.5	2.2	1.4	0.5	3.1	0.3	0.8	0.5	1.1
FN49-312-2	7.1	0.2	3.0	9.1	19.0	38.7	2.1	9.5	1.8	1.6	0.5	2.6	0.2	0.7	0.4	0.8
FN49-312-8	9.3	0.0	2.7	6.9	20.7	35.9	1.8	9.6	2.3	1.7	0.5	3.2	0.3	0.9	0.4	1.1
FN49-620-9	6.2	0.2	3.2	11.5	20.4	36.5	1.9	9.3	1.7	1.4	0.5	2.4	0.2	0.5	0.5	0.9
FN49-620-8	6.9	0.2	2.4	9.3	20.4	37.2	1.7	9.7	2.0	1.6	0.4	2.9	0.3	0.7	0.4	0.9
FN42-50-7	8.5	0.0	3.4	9.4	22.2	34.0	1.9	10.2	1.9	1.3	0.5	2.6	0.2	0.6	0.4	0.9
FN42-50-4	10.9	0.0	3.5	10.2	32.6	26.6	2.0	6.6	1.7	0.7	0.6	1.9	0.0	0.0	0.0	0.8
FN42-50-2	10.4	0.0	3.0	7.6	27.0	30.9	2.0	8.0	2.1	1.1	0.6	2.7	0.0	0.8	0.6	1.1

FN42-50-1	7.8	0.2	3.1	10.4	20.9	36.0	1.8	9.5	1.8	1.5	0.5	2.2	0.2	0.6	0.4	0.9
FN42-5-4	6.2	0.2	2.6	9.3	19.7	38.3	2.0	9.6	2.0	1.6	0.5	2.7	0.2	0.7	0.6	1.0
FN42-5-3	9.7	0.0	2.7	7.0	24.4	33.7	2.0	8.4	2.1	1.2	0.6	2.9	0.3	0.7	0.7	1.1
FN42-81-4	10.6	0.0	2.8	12.7	30.4	35.3	0.0	8.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FN42-81-4	9.4	0.2	2.5	11.4	27.0	31.4	1.2	7.3	1.7	1.1	0.4	1.7	0.2	0.5	0.4	0.8
FN42-81-3	10.3	0.0	3.0	6.9	26.7	32.3	1.9	7.3	2.1	1.1	0.6	2.7	0.4	0.8	0.5	1.0
FN42-81-1	7.1	0.2	2.5	10.1	19.7	38.1	1.4	9.6	2.0	1.7	0.4	2.5	0.2	0.7	0.4	1.0
FN42-97-4	10.7	0.0	3.9	10.6	33.1	25.3	2.0	6.7	1.6	0.6	0.6	1.8	0.0	0.0	0.7	0.8
FN42-97-3	6.7	0.1	2.8	10.9	16.0	40.9	1.5	10.5	1.7	1.9	0.4	2.3	0.1	0.6	0.5	0.9
FN42-97-1	6.9	0.1	2.9	9.0	16.2	45.2	1.5	8.2	1.4	1.7	0.4	2.0	0.1	0.6	0.4	1.1
FN42-353-4	8.8	0.2	2.7	9.1	20.8	35.0	1.8	9.0	2.1	1.6	0.5	2.8	0.3	0.8	0.4	1.2
FN42-353-2	7.1	0.2	2.7	11.5	17.1	41.3	1.3	8.6	1.5	1.6	0.3	2.1	0.2	0.6	0.4	1.0
FN42-353-1	8.0	0.2	2.6	10.1	19.3	37.5	1.5	9.3	2.0	1.7	0.4	2.6	0.2	0.8	0.4	1.1
FN42-119-5	12.8	0.0	3.0	5.3	31.1	28.6	2.6	6.5	2.2	0.9	0.9	3.3	0.0	1.0	0.0	1.0
FN42-119-5	12.4	0.3	3.0	5.2	30.3	27.7	2.5	6.4	2.2	0.8	0.8	3.2	0.6	1.0	0.5	1.0
FN42-119-2	10.6	0.0	3.7	8.4	32.0	24.9	2.7	8.1	2.0	0.7	0.7	2.9	0.5	0.7	0.0	0.7
FN42-119-1	6.9	0.2	2.7	9.0	18.1	39.8	1.6	10.2	2.1	1.8	0.4	2.6	0.2	0.7	0.5	1.0

Table d2. Fatty acid composition (%) of M_3, M_4, and M_5 seeds of lineage FN49-38.

Line name	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	20:2	20:3	22:0	22:1	22:2	22:3	24:0	24:1
FN49-38-1	6.8	0.2	2.7	11.2	20.0	38.3	1.4	9.7	1.8	1.5	0.3	2.2	0.2	0.5	0.3	0.9
FN49-38-2	6.8	0.1	3.0	10.7	20.5	36.8	1.6	10.0	2.0	1.5	0.4	2.4	0.2	0.5	0.3	1.1
FN49-38-3	7.3	0.1	3.0	10.6	21.3	35.5	1.6	10.2	2.0	1.4	0.4	2.5	0.2	0.5	0.3	1.0
FN49-38-4	6.8	0.2	2.8	10.6	20.7	38.0	1.5	9.6	1.9	1.5	0.4	2.2	0.2	0.5	0.3	1.0
FN49-38-5	6.8	0.2	2.9	10.3	20.1	37.9	1.6	10.0	1.8	1.4	0.4	2.4	0.2	0.5	0.3	1.0
FN49-38-6	6.7	0.1	3.1	11.3	20.6	36.3	1.7	9.9	1.8	1.4	0.4	2.5	0.2	0.5	0.3	1.0
FN49-38-7	6.9	0.2	3.5	11.5	22.2	34.8	1.9	9.7	1.7	1.2	0.4	1.9	0.2	0.4	0.3	0.8
FN49-38-8	6.7	0.2	2.5	10.5	19.8	39.5	1.3	9.6	1.9	1.6	0.3	2.2	0.2	0.5	0.2	0.9
FN49-38-9	6.5	0.1	2.8	11.8	20.0	36.9	1.5	10.1	2.0	1.5	0.4	2.4	0.2	0.5	0.3	1.0
FN49-38-8-8	10.9	0.0	3.0	7.2	31.4	28.9	2.1	6.5	2.0	0.8	0.6	2.1	0.0	0.5	0.5	1.2
FN49-38-8-7	7.8	0.2	2.7	12.2	25.9	31.3	1.4	8.7	2.1	1.1	0.4	1.9	0.2	0.4	0.4	1.0
FN49-38-8-5	10.2	0.0	2.9	7.6	29.9	29.7	1.8	7.5	2.2	1.0	0.6	2.1	0.0	0.6	0.5	1.3
FN49-38-8-4	9.9	0.0	3.2	8.8	30.8	27.8	1.9	7.7	2.0	0.8	0.5	2.1	0.0	0.5	0.5	1.2
FN49-38-8-12	6.8	0.2	2.5	9.0	18.4	39.7	1.6	10.5	2.0	1.8	0.4	2.6	0.2	0.7	0.3	1.1
FN49-38-8-11	9.1	0.2	3.1	7.7	25.6	32.0	1.9	9.1	2.2	1.2	0.5	2.7	0.3	0.6	0.3	1.2
FN49-38-8-10	9.7	0.0	3.3	8.9	30.0	28.1	2.1	7.8	2.0	0.8	0.6	2.2	0.0	0.5	0.5	1.2
FN49-38-6-5	7.8	0.0	2.8	10.8	25.4	32.0	1.6	9.1	2.2	1.2	0.4	2.2	0.3	0.5	0.4	1.1
FN49-38-6-2	10.9	0.0	3.2	6.8	31.8	28.1	2.2	6.6	2.1	0.8	0.7	2.4	0.0	0.6	0.5	1.3
FN49-38-4-9	6.8	0.0	2.6	13.7	23.6	32.2	1.4	10.8	1.9	1.1	0.4	1.9	0.1	0.4	0.3	0.8
FN49-38-4-8	6.0	0.1	2.7	14.8	20.7	34.4	1.4	10.7	1.7	1.2	0.3	1.8	0.1	0.3	0.2	0.7
FN49-38-4-7	7.2	0.0	3.3	12.4	25.5	30.0	1.9	10.3	1.8	1.0	0.4	2.1	0.0	0.4	0.3	0.7
FN49-38-4-5	6.5	0.1	2.5	14.1	22.4	33.4	1.4	10.6	1.8	1.2	0.3	2.0	0.1	0.4	0.3	0.8
FN49-38-4-5	7.7	0.0	3.3	12.9	26.5	28.8	1.9	10.2	1.8	0.9	0.4	2.0	0.2	0.4	0.3	0.7
FN49-38-4-4	7.5	0.0	2.8	13.2	25.5	31.2	1.5	9.4	1.8	1.0	0.4	1.8	0.0	0.4	0.2	0.8
FN49-38-4-3	6.7	0.2	2.8	11.4	19.6	36.1	1.6	11.1	1.9	1.5	0.4	2.4	0.2	0.5	0.3	0.8

FN49-38-4-12	7.9	0.0	3.1	15.0	26.7	27.7	1.4	9.5	1.9	0.9	0.4	1.9	0.0	0.3	0.3	0.9
FN49-38-4-10	6.1	0.1	2.7	15.6	20.8	34.8	1.3	10.1	1.7	1.2	0.3	1.6	0.1	0.3	0.3	0.7
FN49-38-4-1	7.2	0.0	3.2	14.3	24.9	29.9	1.7	10.3	1.8	1.0	0.4	1.8	0.0	0.3	0.3	0.8
FN49-38-3-8	7.4	0.1	2.6	12.0	22.6	34.4	1.4	9.5	2.0	1.3	0.4	2.1	0.2	0.5	0.3	1.0
FN49-38-3-7	8.9	0.2	3.1	9.2	27.5	29.9	1.9	8.8	2.0	1.0	0.5	2.4	0.3	0.5	0.4	1.1
FN49-38-3-6	9.7	0.0	3.7	9.7	29.2	27.7	2.2	8.3	1.8	0.8	0.6	2.4	0.0	0.4	0.5	1.1
FN49-38-3-5	8.3	0.0	2.8	11.5	27.6	29.4	1.5	8.4	2.1	1.0	0.4	2.1	0.3	0.5	0.4	1.2
FN49-38-3-2	10.3	0.0	3.8	10.5	32.0	25.7	2.1	7.5	1.8	0.7	0.6	2.2	0.0	0.0	0.0	1.0
FN49-38-3-10	6.6	0.1	2.6	13.6	21.3	35.1	1.2	10.4	2.0	1.4	0.3	1.9	0.2	0.4	0.3	0.9
FN49-38-3-1	9.0	0.0	2.9	10.0	27.7	29.6	1.8	8.8	2.0	1.0	0.5	2.2	0.2	0.5	0.4	1.0
FN49-38-1-5	7.8	0.2	3.2	9.7	23.6	33.1	1.9	9.9	2.0	1.3	0.5	2.5	0.2	0.6	0.4	1.1
FN49-38-1-3	7.7	0.2	3.1	11.1	23.5	32.3	1.8	10.0	2.1	1.3	0.4	2.4	0.2	0.5	0.3	1.1
FN49-38-1-11	9.4	0.2	3.7	9.4	30.2	26.7	2.2	8.3	2.0	0.8	0.6	2.4	0.3	0.5	0.4	1.1
FN49-38-1-11-8	6.1	0.1	2.6	12.6	18.0	38.2	1.3	11.1	1.9	1.7	0.3	2.2	0.1	0.4	0.3	0.9
FN49-38-1-3-1	6.5	0.1	2.6	14.0	22.3	33.3	1.4	10.5	1.9	1.2	0.3	1.9	0.1	0.4	0.3	0.8
FN49-38-1-3-9	6.3	0.1	2.8	12.7	21.3	35.0	1.6	10.6	1.8	1.3	0.4	2.2	0.2	0.4	0.3	0.9
FN49-38-1-3-8	6.1	0.1	2.9	15.6	20.8	34.2	1.4	10.4	1.8	1.3	0.3	1.7	0.1	0.3	0.3	0.7
FN49-38-1-3-7	8.2	0.0	3.2	12.2	27.6	28.3	1.9	9.6	1.9	1.0	0.5	2.1	0.0	0.4	0.3	0.9
FN49-38-1-3-6	8.6	0.0	3.5	11.3	27.6	28.2	2.2	9.4	1.8	0.9	0.5	2.3	0.0	0.5	0.4	1.0
FN49-38-1-3-4	6.3	0.1	2.6	13.3	21.2	35.0	1.4	10.6	1.8	1.3	0.4	2.0	0.1	0.4	0.3	0.9
FN49-38-1-3-3	6.4	0.2	2.6	13.6	23.2	32.2	1.5	10.8	1.9	1.1	0.3	2.0	0.1	0.4	0.3	0.7
FN49-38-1-3-12	7.3	0.0	3.0	12.7	25.4	30.5	1.7	10.3	1.9	1.1	0.4	2.1	0.0	0.4	0.3	0.9
FN49-38-1-3-10	6.1	0.1	2.9	16.6	22.6	31.9	1.3	10.5	1.7	1.1	0.3	1.6	0.1	0.3	0.2	0.7

Table d3. Fatty acid composition (%) of F_2 seeds of backcrossing lines.

Line name	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	20:2	20:3	22:0	22:1	22:2	22:3	24:0	24:1
FNCX 13-1	10.38	0.00	3.13	9.07	31.10	26.08	2.10	8.95	2.04	0.73	0.53	2.54	0.00	0.49	0.00	0.83
FNCX 13-5	8.96	0.00	2.91	10.25	28.30	26.52	1.96	10.56	2.25	0.87	0.48	2.92	0.29	0.49	0.32	0.82
FNCX 13-7	7.17	0.15	2.69	13.39	23.40	29.20	1.63	11.75	2.22	1.16	0.39	2.90	0.26	0.52	0.23	0.84
FNCX 13-8	7.86	0.17	2.40	11.23	25.42	27.16	1.87	12.99	2.33	0.99	0.45	3.28	0.26	0.50	0.25	0.67
FNCX 13-9	8.02	0.18	2.78	12.61	25.55	28.27	1.68	10.62	2.10	0.99	0.41	2.70	0.25	0.49	0.26	0.86
FNCX 16-1	7.34	0.15	2.83	11.72	22.22	31.36	1.80	11.53	2.18	1.26	0.42	3.10	0.27	0.58	0.26	0.91
FNCX 16-10	9.14	0.00	3.22	10.66	29.28	25.03	2.22	10.33	2.18	0.79	0.54	2.84	0.30	0.47	0.33	0.79
FNCX 16-11	7.02	0.14	2.61	12.79	21.64	32.11	1.58	11.78	2.11	1.26	0.37	2.76	0.22	0.50	0.23	0.79
FNCX 16-12	10.07	0.00	2.93	7.99	28.71	27.78	2.23	9.94	2.28	0.92	0.59	3.00	0.00	0.64	0.00	0.88
FNCX 16-3	9.35	0.00	2.84	9.03	27.97	28.22	1.98	9.22	2.30	0.95	0.53	3.02	0.36	0.60	0.34	1.05
FNCX 16-4	8.65	0.00	2.94	10.58	26.51	28.32	1.92	10.20	2.17	0.98	0.48	2.94	0.29	0.56	0.32	0.97
FNCX 16-9	7.39	0.15	2.58	11.85	23.31	30.39	1.74	11.96	2.16	1.17	0.41	2.90	0.22	0.52	0.26	0.82
FNCX 5-1	9.95	0.00	3.31	9.41	28.78	27.37	2.05	9.37	2.00	0.82	0.54	2.67	0.00	0.51	0.34	0.87
FNCX 5-10	9.05	0.00	2.85	9.35	27.91	27.31	2.12	10.12	2.24	0.94	0.55	3.37	0.33	0.57	0.33	0.92
FNCX 5-11	9.01	0.00	3.03	9.80	26.82	27.87	2.03	10.01	2.31	1.01	0.52	3.19	0.34	0.58	0.33	1.01
FNCX 5-2	7.68	0.16	2.75	12.48	24.73	28.81	1.70	11.29	2.13	1.05	0.41	2.85	0.24	0.50	0.26	0.83
FNCX 5-8	8.61	0.00	3.30	11.34	27.35	26.74	2.07	10.33	2.05	0.87	0.50	2.90	0.26	0.47	0.30	0.89
FNCX 9-1	7.11	0.14	2.72	12.70	22.35	30.20	1.69	11.90	2.28	1.25	0.40	3.12	0.28	0.57	0.24	0.87
FNCX 9-12	7.38	0.15	2.90	12.14	23.76	29.21	1.87	11.80	2.20	1.11	0.43	3.03	0.25	0.52	0.27	0.86
FNCX 9-2	8.87	0.23	2.82	8.85	26.53	29.58	1.95	9.83	2.18	0.98	0.52	3.08	0.31	0.58	0.32	0.97
FNCX 9-6	8.18	0.21	3.01	10.95	25.72	28.95	1.87	10.12	2.07	0.98	0.46	2.88	0.26	0.49	0.30	0.95
FNCX 9-7	8.51	0.21	2.66	9.90	26.27	29.03	1.85	10.13	2.34	1.01	0.49	3.09	0.31	0.56	0.29	0.97
FNCX 9-8	6.58	0.14	2.34	12.76	21.37	33.07	1.47	12.02	2.09	1.23	0.34	2.47	0.18	0.44	0.22	0.75

Table d4. Fatty acid composition (%) of F_3 seeds of backcrossing lines.

Line name	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	20:2	20:3	22:0	22:1	22:2	22:3	24:0	24:1
FNCX16-10-2	6.7	0.2	2.4	10.6	19.0	36.2	1.6	11.5	2.2	1.7	0.4	3.2	0.2	0.7	0.3	1.0
FNCX16-10-3	6.2	0.1	2.5	12.5	17.8	36.0	1.7	12.3	2.0	1.7	0.4	3.2	0.2	0.6	0.2	0.8
FNCX16-10-4	6.8	0.1	2.7	10.8	21.2	31.1	2.1	12.8	2.3	1.4	0.5	4.1	0.3	0.7	0.3	0.9
FNCX16-10-6	6.7	0.1	2.6	12.1	18.9	37.6	1.4	10.5	1.9	1.6	0.3	2.4	0.2	0.5	0.3	0.9
FNCX16-10-7	6.1	0.1	2.6	12.5	18.8	32.8	1.8	13.4	2.2	1.6	0.4	3.8	0.2	0.6	0.2	0.9
FNCX16-10-8	6.1	0.2	2.2	12.7	20.6	31.4	1.7	13.0	2.1	1.4	0.4	4.0	0.3	0.6	0.2	0.8
FNCX16-11-1	6.4	0.1	2.8	17.5	22.4	30.9	1.2	11.0	1.5	0.9	0.3	1.7	0.1	0.2	0.2	0.7
FNCX16-11-2	6.5	0.2	3.3	14.1	21.2	30.9	2.2	11.8	1.7	1.1	0.4	2.6	0.1	0.3	0.2	0.8
FNCX16-11-3	6.6	0.2	2.7	15.1	21.5	32.5	1.5	11.0	1.5	1.1	0.3	2.2	0.1	0.3	0.2	0.7
FNCX16-11-6	7.0	0.2	2.8	15.5	22.2	32.0	1.4	10.4	1.5	1.0	0.3	1.9	0.1	0.3	0.2	0.8
FNCX16-11-8	7.0	0.2	3.1	13.5	21.5	30.0	2.0	12.5	1.9	1.1	0.4	2.9	0.2	0.4	0.2	0.8

Appendix E. Information of 96 lines selected for Illumina sequencing

Table e1. Selected M_2 lines for Illumina sequencing and their corresponding M_3 seeds' fatty acid composition (%).

GBS plate 1		16:0	18:0	18:1	18:1	18:2	18:3	20:0	20:1	20:1	20:2	20:3	22:0	22:1	22:2	22:3	24:0	24:1
A01	Midas 3-1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
A02	FN49-35-9	8.87	4.06	7.98	1.22	21.62	34.53	2.12	11.12	0.74	2.07	1.28	0.48	2.01	0.00	0.31	0.35	0.92
A03	FN49-38-8	6.65	2.51	10.49	1.25	19.82	39.46	1.31	9.56	0.69	1.86	1.63	0.31	2.15	0.18	0.55	0.24	0.85
A04	FN49-43-6	7.20	2.59	10.96	1.37	19.12	32.95	2.38	12.59	1.04	2.05	1.42	0.50	3.61	0.25	0.71	0.17	0.58
A05	FN49-52-5	7.84	2.72	9.78	0.98	19.31	35.30	2.11	11.53	0.74	1.91	1.41	0.47	3.56	0.26	0.74	0.29	0.71
A06	FN49-70-9	7.33	3.89	7.50	0.88	18.14	34.11	3.89	11.24	0.69	2.02	1.46	0.81	4.95	0.36	1.07	0.42	0.70
A07	FN49-198-1	9.15	4.87	12.59	0.91	21.83	25.73	4.09	10.84	0.73	1.34	0.73	0.85	3.84	0.26	0.54	0.45	0.90
A08	FN49-312-7	7.24	2.24	8.38	1.35	16.94	41.03	1.62	9.41	1.14	2.08	2.05	0.42	3.00	0.27	0.90	0.41	1.04
A09	FN49-375-11	7.48	3.51	10.02	1.29	14.86	38.07	2.51	11.09	0.98	1.53	1.57	0.55	3.40	0.21	0.78	0.36	1.16
A10	FN49-507	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
A11	FN42-365	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
A12	1149-1-1-1-1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
B01	Midas 3-2	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
B02	FN49-35-10	8.49	3.82	9.91	1.15	19.47	33.20	2.49	11.88	0.75	1.84	1.21	0.52	2.94	0.20	0.51	0.30	0.93
B03	FN49-38-9	6.54	2.82	11.76	1.11	19.98	36.95	1.49	10.06	0.67	1.95	1.52	0.37	2.37	0.19	0.52	0.29	1.01
B04	FN49-43-7	8.90	3.39	8.17	0.72	26.83	25.88	2.74	11.92	0.70	2.35	0.83	0.69	4.92	0.41	0.64	0.00	0.92
B05	FN49-52-7	8.05	3.07	8.23	0.87	19.72	33.74	2.44	12.62	0.55	2.08	1.39	0.53	4.34	0.27	0.78	0.29	0.69
B06	FN49-70-10	6.46	2.52	8.46	1.06	14.91	38.99	2.47	12.24	0.98	1.91	1.82	0.55	4.65	0.30	1.04	0.36	0.81
B07	FN49-198-4	7.08	3.08	9.26	0.86	23.84	29.75	2.63	11.94	0.70	2.40	1.17	0.56	4.08	0.36	0.72	0.30	0.96
B08	FN49-312-9	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
B09	FN49-373	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
B10	FN49-344	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
B11	FN42-730	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/

B12	1149-1-1-1-3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
C01	Midas 3-3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
C02	FN49-38-1	6.78	2.75	11.17	1.17	19.99	38.28	1.41	9.68	0.68	1.79	1.49	0.34	2.17	0.16	0.48	0.29	0.93
C03	FN49-41-5	9.59	3.78	7.82	1.19	24.60	29.90	2.80	9.51	1.04	2.06	1.08	0.67	2.96	0.35	0.74	0.38	1.05
C04	FN49-43-8	7.19	2.80	10.69	1.12	18.79	34.23	2.21	12.36	0.89	2.04	1.44	0.46	3.42	0.23	0.66	0.23	0.76
C05	FN49-52-8	6.88	2.71	13.73	0.81	17.23	33.80	1.59	13.53	0.63	1.97	1.49	0.36	3.21	0.20	0.56	0.25	0.72
C06	FN49-263	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
C07	FN49-198-6	6.22	2.40	9.81	0.85	17.88	35.50	1.90	13.97	0.69	2.24	1.70	0.41	4.06	0.24	0.74	0.24	0.78
C08	FN49-375-1	6.36	3.50	12.04	0.91	19.42	35.21	2.09	12.44	0.58	1.77	1.26	0.41	2.65	0.41	0.69	0.00	0.00
C09	FN49-336	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
C10	FN49-432	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
C11	FN42-706	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
C12	Cx 14-2	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
D01	Midas 3-4	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
D02	FN49-38-2	6.84	2.99	10.71	1.05	20.52	36.79	1.63	9.96	0.66	1.97	1.48	0.40	2.43	0.20	0.54	0.35	1.08
D03	FN49-41-9	6.93	2.87	10.59	1.12	19.84	31.83	2.80	12.32	1.00	2.01	1.24	0.62	4.00	0.30	0.72	0.40	0.97
D04	FN49-43-9	6.79	3.03	11.62	1.03	18.13	33.73	2.58	12.65	0.76	1.93	1.41	0.52	3.48	0.22	0.61	0.28	0.80
D05	FN49-52-9	7.38	2.57	9.28	0.97	17.84	36.27	2.13	12.18	0.82	1.97	1.55	0.48	4.01	0.27	0.82	0.31	0.80
D06	FN49-79-4	8.81	3.22	10.49	1.22	25.91	29.15	2.24	9.83	0.86	1.94	0.91	0.53	2.58	0.25	0.49	0.34	0.79
D07	FN49-312-1	7.42	2.22	8.50	1.34	18.75	38.72	1.57	9.65	1.19	2.24	1.84	0.40	3.17	0.30	0.87	0.33	1.02
D08	FN49-375-4	7.44	2.87	8.87	1.22	15.05	37.68	2.28	13.03	1.06	1.81	1.72	0.47	3.81	0.21	0.74	0.24	0.92
D09	FN49-73	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
D10	FN49-137	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
D11	FN42-621	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
D12	Cx 1-4	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
E01	FN49-35-3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
E02	FN49-38-3	7.25	3.03	10.61	1.12	21.30	35.49	1.63	10.22	0.68	1.96	1.38	0.40	2.46	0.18	0.49	0.33	1.02

E03	FN49-43-1	6.38	3.34	11.69	1.05	18.66	34.03	2.43	12.81	0.83	1.84	1.33	0.52	2.89	0.18	0.51	0.30	0.82
E04	FN49-43-10	8.77	3.60	9.07	0.95	26.45	26.45	2.43	11.72	0.67	2.23	0.90	0.57	3.87	0.33	0.56	0.26	0.84
E05	FN49-70-1	6.87	2.53	7.17	0.85	15.62	39.62	2.28	11.85	0.71	2.24	2.01	0.54	4.69	0.32	1.10	0.37	0.79
E06	FN49-79-5	7.48	3.02	11.32	1.21	21.41	33.44	1.99	10.90	0.89	1.84	1.16	0.43	2.68	0.20	0.50	0.27	0.78
E07	FN49-312-2	7.13	2.95	9.07	1.34	18.99	38.68	2.07	9.53	1.12	1.81	1.59	0.46	2.61	0.22	0.66	0.41	0.84
E08	FN49-375-7	7.69	2.71	8.89	1.03	14.82	39.47	1.83	12.31	0.75	1.75	1.83	0.46	3.81	0.00	0.90	0.30	1.13
E09	FN49-498	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
E10	FN42-187	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
E11	FN42-315	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
E12	Cx 1-2	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
F01	FN49-35-4	6.76	3.19	12.52	1.01	15.10	35.70	2.19	13.55	0.68	1.43	1.33	0.46	3.71	0.17	0.55	0.22	0.94
F02	FN49-38-5	6.77	2.91	10.26	1.32	20.09	37.91	1.59	9.99	0.81	1.85	1.41	0.37	2.35	0.16	0.49	0.30	0.98
F03	FN49-43-3	6.46	3.24	12.43	1.78	17.49	35.41	2.25	10.71	1.45	1.51	1.31	0.52	2.78	0.21	0.65	0.23	0.99
F04	FN49-52-2	7.68	2.57	9.03	1.04	17.41	36.95	2.06	11.85	0.76	2.01	1.61	0.46	3.89	0.26	0.81	0.32	0.83
F05	FN49-70-2	7.45	3.00	6.82	0.91	17.59	35.90	3.04	11.45	0.76	2.15	1.63	0.70	5.39	0.40	1.20	0.41	0.79
F06	FN49-79-6	6.91	3.08	9.01	1.04	19.53	35.19	2.57	11.21	0.94	1.97	1.45	0.56	3.70	0.28	0.77	0.30	0.98
F07	FN49-312-3	8.18	2.67	7.46	1.44	22.49	35.57	1.92	8.50	1.23	2.22	1.44	0.51	3.07	0.33	0.81	0.46	1.13
F08	FN49-375-8	7.51	3.38	11.26	1.12	18.53	35.08	2.05	12.48	0.65	0.00	1.64	1.25	0.42	2.85	0.47	0.21	0.77
F09	FN49-201	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
F10	FN42-51	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
F11	FN42-733	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
F12	Cx 13-3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
G01	FN49-35-7	7.16	3.19	8.71	1.24	19.73	35.42	2.47	10.74	0.91	2.04	1.50	0.54	3.58	0.27	0.78	0.31	0.92
G02	FN49-38-6	6.68	3.09	11.31	1.09	20.60	36.34	1.73	9.86	0.68	1.80	1.39	0.42	2.45	0.18	0.51	0.35	1.04
G03	FN49-43-4	7.05	2.66	11.23	1.13	20.78	31.30	2.13	12.70	0.94	2.17	1.27	0.42	3.79	0.29	0.76	0.20	0.76
G04	FN49-52-3	7.17	2.71	10.15	0.93	18.33	35.94	2.06	12.08	0.62	1.94	1.49	0.46	3.74	0.23	0.71	0.30	0.80
G05	FN49-70-4	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/

G06	FN49-79-8	7.27	2.68	11.42	1.18	20.58	36.80	1.81	10.25	0.73	1.69	1.19	0.41	1.91	0.15	0.42	0.28	0.71
G07	FN49-312-5	8.92	3.59	9.56	1.16	29.01	27.75	2.31	8.87	0.66	1.91	0.80	0.58	2.51	0.24	0.43	0.41	1.04
G08	FN49-375-9	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
G09	FN49-411	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
G10	FN42-377	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
G11	FN42-10	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
G12	Cx 13-1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
H01	FN49-35-8	7.70	3.46	9.20	1.14	17.79	36.01	2.22	12.43	0.83	2.02	1.48	0.48	2.85	0.17	0.48	0.31	0.94
H02	FN49-38-7	6.93	3.52	11.49	1.26	22.20	34.80	1.94	9.73	0.81	1.72	1.21	0.41	1.87	0.16	0.42	0.31	0.78
H03	FN49-43-5	7.58	3.02	8.51	0.89	21.08	30.82	2.49	13.49	0.87	2.35	1.26	0.54	4.68	0.30	0.70	0.22	0.80
H04	FN49-52-4	7.17	2.52	9.95	0.91	17.76	36.58	1.90	12.17	0.69	1.98	1.57	0.42	3.87	0.26	0.76	0.29	0.80
H05	FN49-70-7	7.65	6.97	1.11	0.00	14.33	38.38	3.93	10.46	0.88	1.68	1.81	0.82	4.60	0.28	1.23	0.43	0.80
H06	FN49-79-9	6.70	2.67	11.42	0.93	20.60	32.47	2.09	12.62	0.75	1.95	1.25	0.47	3.61	0.25	0.63	0.26	0.90
H07	FN49-312-6	7.62	2.60	10.00	1.35	20.22	37.49	1.32	9.35	1.08	2.08	1.61	0.35	2.35	0.22	0.58	0.32	0.96
H08	FN49-375-10	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
H09	FN49-665	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
H10	FN42-599	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
H11	FN42-731	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
H12	Cx 19-3	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/

Note. The forward slashes indicate the fatty acid compositions of corresponding lines have not been determined.