Analysis of Genetic Diversity and Expression of Genes Involved in Fatty Acid Composition in Flax (*Linum usitatissimum* L.) and Comparative Genomic Analysis of their Loci

BY

DINUSHIKA PRABODHANI THAMBUGALA

A Thesis

Submitted to the Faculty of Graduate Studies of the

University of Manitoba

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Plant Science
University of Manitoba
Winnipeg, Manitoba, Canada

ACKNOWLEDGEMENTS

First and foremost, I sincerely thank my supervisor Dr. Sylvie Cloutier for her time, patience and her valuable guidance and assistance in all aspects of the research process and my PhD program. I am also most grateful for the fact that her door was always open and she was willing to discuss any issues that I encountered.

I would like to express my sincere gratitude to Dr. Claudio Stasolla, Dr. Michele Piercey-Normore and Dr. Genyi Li who served as members of my advisory committee, for their valuable suggestions, guidance and comments to improve my research.

My special thanks and recognition to Dr. Cloutier's lab members Natasa Radovanovic, Elsa Reimer, Mitali Banik, Kerry Ward, Evelyn Miranda and Andrzej Walichnowski for their excellent technical and moral support. I give my special thanks to Andrzej Walichnowski for manuscript review and suggestions. I am also sincerely thankful to Dr. Santosh Kumar and Dr. Raja Ragupathy, for their valuable suggestions, especially in the later stages of this thesis.

I wish to extend my sincere thanks to Drs Scott Duguid, Gordon Rowland and Helen Booker and their breeding teams for providing the phenotypic data. I also thank Dr. Raju Datla of the National Research Council, Saskatoon, Saskatchewan, Canada for generating the RNA-Seq dataset and Dr. Frank M. You of the Cereal Research Centre, Morden, Manitoba, Canada for bioinformatics help during my research.

This research was conducted as part of the Total Utilization Flax GENomics (TUFGEN) project funded by Genome Canada and other stakeholders including the Flax

Council of Canada, Manitoba Flax Growers Association and the Province of Manitoba. Project management and support by Genome Prairie are also gratefully acknowledged.

Many thanks are due to my mother and sisters for their encouragement of my academic pursuits and support.

Finally I would like to express my utmost gratitude and appreciation to my husband, Kapila, for being patient and supportive during the completion of my PhD program, and Sadew and Sandul my beloved sons that I couldn't spend more time with them as I was busy with my study. This thesis is also dedicated in memory of my father who was the greatest inspiration in my life and is sadly missed by me.

Dedicated to

My husband Kapila

and our sons Sadew and Sandul

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

AA	Arachidonic Acid
ABRE	ABA-Responsive Cis-Elements
ACCase	Acetyl-CoA Carboxylase
ACP	Acyl-Carrier Protein
ALA	Alpha-Linolenic Acid
ANOVA	Analysis of Variance
BAC	Bacterial Artificial Chromosome
DAA	Days After Anthesis
DHA	Docosahexaeonoic Acid
EFA	Essential Fatty Acid
EMS	Ethyl Methanesulfonate
EPA	Eicosapentaenoic Acid
ER	Endoplasmic Reticulum
EST	Expressed Sequence Tag
FA	Fatty Acid
FAD	Fatty Acid Desaturase
FAD2	Fatty Acid Desaturase 2
FAD3	Fatty Acid Desaturase 3
FAE	Fatty Acid Elongase
FAME	Fatty Acids Methyl Ester
FAS	Fatty Acid Synthase

FATA	Acyl-ACP Thioesterase
GC	Gas Chromatography
IOD	Iodine Value
KAS	3-Ketoacyl-ACP Synthase
LACS	Long-Chain-Acyl-CoA Synthetase
LDL	Low-Density Lipoprotein
LIN	Linolenic acid
LIO	Linoleic acid
MAD	Modified Augmented Design
MAT	Malonyl-CoA:Acyl Carrier Protein S-Malonytransferase
MB	Manitoba
MITE	Miniature Inverted Repeat Transposable Element
NCBI	National Center for Biotechnology Information
NJ	Neighbor Joining
OIL	Oil content
OLE	Oleic acid
ORF	Open Reading Frame
PAL	Palmitic acid
PC	Phosphatidylcholine
PGRC	Plant Gene Resources of Canada
PUFA	Polyunsaturated Fatty Acid
QTL	Quantitative Trait Loci
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction

SAD	Stearoyl-ACP Desaturase
SK	Saskatchewan
SNP	Single Nucleotide Polymorphism
STE	Stearic acid
TAG	Triacylglycerol
TE	Transposable Element
TSS	Transcription Start Site
TUFGEN	Total Utilization Flax GENomics
VLC-PUFA	Very Long-Chain PUFA
WGS	Whole Genome Shotgun

LIST OF APPENDICES

Appendix Page

ABSTRACT

Flax (*Linum usitatissimum* L.) is the richest crop source of omega-3 fatty acids praised for their health benefits in human and animals. Here, the extent of the genetic variability for genes encoding stearoyl-ACP desaturase (SAD), fatty acid desaturase 2 (FAD2) and 3 (FAD3) was determined by sequencing the six paralogous genes from 120 flax accessions revealing 71 alleles and 26 isoforms with predicted functional mutations. *Fad3a* and *fad3b* genes showed the highest levels of genetic variations. While most of the single nucleotide polymorphisms (SNPs) and all the indels were silent mutations, both genes carried non-sense SNP mutations resulting in premature stop codons, a feature not observed in *sad* and *fad2* genes.

A subset of six lines including two moderately low, two intermediate and two high LIN lines were selected to establish the relationship between the desaturase expression and FA composition. While the expression of each desaturase differed during seed development, no differential expression of any of the six desaturases was observed between accessions with low, intermediate or high LIN content. Desaturase expression did not correlate with FA composition variations of the six flax genotypes studied, hence other genetic factors were hypothesized to play a role in determining the FA composition of flax.

A comparative structural genomics analysis of the six loci harboring the desaturase genes was performed using bacterial artificial chromosome (BAC) libraries of flax cultivar CDC Bethune and breeding line M5791. With one gene at every 3.2-4.6 Kb, the desaturase loci have a higher gene density than the genome's average of one gene per

7.8-8.2 Kb. High sequence conservation in both genic and intergenic regions of the *sad* and *fad2b* loci contrasted with the significant level of variation of the *fad2a* and *fad3* loci. The organization of the *fad2b* locus was particularly complex with seven copies of the *fad2b* gene in both genotypes. The presence of *Gypsy*, *Copia*, MITE, *Mutator, hAT* and other novel repeat elements at the desaturase loci was similar to that of the whole genome. This structural genomic analysis provided some insight into the genomic organization and composition of the main desaturase loci in flax.

FOREWORD

This thesis follows the paper style format recommended by the Plant Science Department and the Faculty of Graduate Studies of the University of Manitoba. The thesis has seven chapters: a general introduction, a literature review, three manuscripts, a general discussion and conclusion, and references. Manuscripts follow the guidelines of Theoretical and Applied Genetics. Each manuscript contains abstract, introduction, material and methods, results and discussion.

1.0 GENERAL INTRODUCTION

Flax (*Linum usitatissimum* L.) is an annual, self-pollinating, diploid plant (2n=2x=30) with a relatively small genome size of ~370 Mb (Wang et al. 2012). This crop can be commercially grown for its stem fibers (fiber flax) or its seed oil (linseed or flaxseed) (Dillman 1953; Zohary 1999). The fiber is used in the making of textiles such as linen, dyes, specialty papers such as bank notes, eco-friendly insulations, etc. Flax seed produces valuable and unique oil with multiple industrial, food and nutraceutical end-uses (Green and Marshall 1984; Vaisey-Genser and Morris 2001; Fofana et al. 2004). Recently, flax has gained attention in human health mostly because of its high content of health promoting compounds such as α-linolenic acid (ALA) and lignan (Hasler et al. 2000; Simopoulos 2000; Watkins et al. 2001). Current linseed varieties have oil content up to 50% composed of five main fatty acids (FAs): palmitic (PAL, C16:0; ~6%), stearic (STE, C18:0; ~4.4%), oleic (OLE, C18:1cis$^{\Delta 9}$; ~24.2%), linoleic (LIO, C18:2cis$^{\Delta 9,12}$; ~15.3%) and linolenic (LIN, C18:3 cis$^{\Delta 9,12,15}$; ~50.1%) (Muir and Westcott 2003). ALA (also referred to as LIN) constitutes up to 73% of the total fatty acids in high-linolenic acid varieties making flax the richest source of plant-based omega (ω)-3 FAs (Fofana et al. 2010). Solin^{TM} type flax varieties are at the other end of the spectrum in terms of FA composition considering that they are rich in ω-6 FAs but contain only 2-3% ω-3 FAs.

LIO and LIN are essential fatty acids (EFAs) for humans and serve as precursors of the ω-6 and ω-3 FA families, respectively. Mammalian tissues cannot synthesize LIO and LIN and so these FAs must be incorporated into the diet (Simopoulos et al. 2000). LIO and LIN, collectively called polyunsaturated FAs (PUFAs), are important components of

cell membranes that are involved in plant metabolism as a source of stored energy in the form of triacylglycerols (TAGs) and as precursors of signaling molecules such as jasmonic acid (Ohlrogge and Browse 1995). They are also precursors to other very long-chain PUFAs (VLC-PUFAs) such as eicosapentaenoic acid (EPA, C20:5) and docosahexaeonoic acid (DHA, C22:6) whose health benefits in reducing serum cholesterol levels and preventing cardiovascular diseases have been scientifically proven by a number of recent studies (Ander et al. 2004; Wiesenfeld et al. 2003). Flax has recently been granted a health claim related to its role in reducing total and low-density lipoprotein (LDL) cholesterol (Health Canada 2014).The growing interest and increasing demand for oilseed flax is mainly due to the intrinsic quality of its oil. Its FA composition, with potential for high and low content of ALA, is what ultimately determines the applications and end-uses of linseed oil (Green and Marshall 1984; Fofana et al. 2004; Fofana et al. 2006).

During oil biosynthesis, the desaturation of FAs by the sequential action of substrate-specific desaturase is important in determining the relative proportions of saturated and unsaturated FAs and thus determines the end-uses of the oil. Desaturation of STE into OLE, LIO and LIN is consecutively performed by stearoyl-ACP desaturase (SAD), fatty acid desaturase 2 (FAD2) and fatty acid desaturase 3 (FAD3) enzymes. To date, little information is known about the allelic diversity of these genes even though some gene members have been cloned and identified in flax (Singh et al. 1994; Fofana et al. 2004; Vrinten et al. 2005; Banik et al. 2011).

In most plant species, the first FA desaturation step occurs in the plastid where the reaction is catalyzed by SAD (Voelker and Kinney 2001). The *sad* gene codes for the

enzyme responsible for converting stearoyl-ACP to oleoyl-ACP by introducing a double bond at the Δ^9 position, thereby increasing the unsaturated FA content of the plant (Ohlrogge and Jaworski 1997). Two paralogous *sad* loci namely *sad1* and *sad2*, which are differentially expressed in plants, have been identified in flax (Jain et al. 1999). The FAs exported to the endoplasmic reticulum (ER) are then further subjected to elongation and desaturation steps. These steps are catalyzed by two sets of enzymes, namely fatty acid elongases (FAEs) and desaturases (FADs) resulting in ω-6 and ω-3 PUFAs (Voelker and Kinney 2001). The *fad2* genes encode proteins responsible for desaturation of OLE into LIN by addition of a double bond at the Δ^{12} position. Two orthologous copies of the *fad2* gene, namely *fad2a* and *fad2b,* are present in the flax genome as is the case for other plant species (Heppard et al. 1996; Fofana et al. 2004). The *fad3* genes encode proteins responsible for desaturation of LIO into LIN by addition of a double bond at the Δ^{15} position. Multiple *fad3* genes have been identified in the flax genome, namely *fad3a, fad3b,* and more recently *fad3c* (Vrinten et al. 2005; Banik et al. 2011). FAD3A and FAD3B have been shown to be the major enzymes controlling LIN content in flax (Vrinten et al. 2005). A major role for FAD3C has not been established (Banik et al. 2011).

Canada is the world's leading producer and exporter of linseed (Muir and Westcott 2003). Despite extensive breeding efforts in linseed cultivar development (Kenaschuk and Rowland 1995) and the historical significance of flax cultivation, few studies have been conducted to assess the genetic diversity, expression and organization of genes involved in the FA composition of flax. The development of such data would aid

breeding efforts by broadening the genetic basis of the germplasm used in flax improvement (Carter 1993).

While the knowledge of the desaturase genes and enzymes involved in the fatty acid composition of flax seeds is good, little is known about the extent of the genetic variability of these genes, their corresponding isoforms and their role(s) in determining FA composition. The major aim of the first part of this study was to determine the genetic variation for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* genes in flax by sequencing these genes from 120 flax accessions, a subset of a larger core collection comprising ~400 accessions. The sequence data were analyzed to identify DNA variations such as single nucleotide polymorphisms (SNPs) and indels. Deduced amino acid sequences were analyzed. These genotypic data were correlated to the FA composition of the lines grown in the field during three years (2010-2012) and at two locations (Saskatoon and Morden) to hypothesize the functionality of these alleles and isoforms in FA biosynthesis in flax.

Further, little was known about the relationship between expression levels of fatty acid desaturase genes during seed development and FA composition in flax. Only two studies on the regulation and expression of *sad* and *fad* genes during seed development have been reported (Fofana et al. 2006; Banik et al. 2011). Fofana et al. (2006) reported that the expression of *sad* and *fad2* genes in flax was modulated during seed development whereas Banik et al. (2011) found that the expression patterns of *fad3a* and *fad3b* were highly correlated with LIN accumulation during seed development. I hypothesized that FA composition differences among flax accessions could result from differential expression of the desaturase genes during seed development. Based on this hypothesis, three goals were established. First, I wanted to quantify the expression levels of the

desaturase genes at different stages of seed development by semi-quantitative reverse transcriptase (RT)-PCR in relatively low, intermediate and high LIN genotypes expressing identical isoforms for all six desaturases. Second, I planned to study the structural differences in the promoter region of the six desaturase genes. Third, I intended to correlate these structural and expression data with FA composition as determined from the field grown genotypes during four years at two locations with the overall objective of gaining a greater understanding of the genetic factors controlling FA composition in flax.

To better understand the general structural organization of the flax genome and more specifically that of the main loci controlling FA desaturation, a comparative structural genomics analysis of the six loci was performed using bacterial artificial chromosome (BAC) libraries of flax cultivar CDC Bethune and breeding line M5791. CDC Bethune is a high yielding conventional flax variety (55-57% LIN) widely grown in Western Canada (Rowland et al. 2002). M5791 is a high LIN breeding line which contains ~65% LIN under field conditions. Comparative structural genomics of these two lines at the six desaturase loci aimed to reveal insights into putative structural contexts responsible for their different FA composition. From an evolutionary point of view, structural comparative analyses of these independent loci could unearth novel information about the events that have shaped the flax genome.

2.0 LITERATURE REVIEW

2.1 Flax

Flax (*Linum usitatissimum* L., 2n=2x=30) is a diploid, self-pollinated plant belonging to the Linaceae family and is one of about 200 species in the genus *Linum* (Zohary 1999; Diederichsen and Richards 2001). The species is believed to have originated in either the Middle East or Indian regions and was one of the eight 'founder crops' that initiated agriculture in the 'old world' (Vavilov 1951; Zohary 1999). Morphological, cytological and molecular evidence suggest that flax was domesticated for oil and/or fiber use more than 8,000 years ago in the Near East and the wild progenitor of cultivated flax is pale flax (*L. usitatissimum* L. subsp. *angustifolium* (Huds.) Thell.; Diederichsen and Hammer 1995). Flax has been grown for its stem fibers (fiber flax) or its seed oil (linseed or oilseed flax) since ancient times (Zohary and Hopf 2000; Muir and Westcott 2003). Linseed plants (oil morphotype) are shorter, more branched, have larger seeds and are grown in continental climate regions such as Canada, China, India, United States and Argentina. Fiber flax (fiber morphotype) plants are taller, more sparsely branched, have smaller and fewer seeds and are grown in the cool-temperate regions of China, Russia and Western Europe (Zohary and Hopf 2000; Green et al. 2008).

Canada is the world's largest linseed producer and exporter (Muir and Westcott 2003). In 2013, the total world production of linseed reached ~2.3 million tonnes, with Canada, China and Russia being the main producers with 712, 398 and 325 thousand metric tonnes, respectively (FAOSTAT 2014). In 2013, Western Canada's linseed production (712,000 metric tonnes) increased by 223,000 metric tonnes from 489,000

metric tonnes in 2012. Saskatchewan accounted for 82% of the flaxseed production while Manitoba and Alberta produced 8% and 10%, respectively (http://www.grainscanada.gc.ca).

2.2 Utilization of flax as a dual purpose crop

Recently, there has been an interest in total utilization or dual purpose flax, mostly referring to oilseed flax but with a value-added stem fiber and/or chive markets. The increasing demand for the seed component of flax in modern societies during the last two decades is mainly due to an increased understanding of the role of ω-3 in nutrition and its associated health benefits; flax being the richest agricultural plant source of ω-3 fatty acids (FAs). Today, flax seed is considered a functional food because it contains nutrients that provide health benefits (Hasler et al. 2000; Singh et al. 2011) and may prevent certain chronic illnesses such as heart disease, strokes, diabetes, cancers and Alzheimer as well as help combat obesity (Flax Council of Canada). Flax seed oil with its unique drying properties is widely used for industrial purposes, in the manufacture of linoleum and paints and in preserving wood and concrete (Green 1986). Flax's usefulness is not restricted to its seed but value can also be obtained from its straw components, hence the qualifiers 'dual purpose' or 'total utilization'. The straw produces a strong and long lasting fiber that is praised for its quality. Linen is the most well-known product of flax fibres but there are many others such as fiber glass substitutes in composites or in the manufacture of fire logs, dyes, papers and other similar products (Green and Marshall 1984; Vaisey-Genser and Morris 2001).

Despite extensive breeding efforts (Kenaschuk and Rowland 1995) and the historical significance of flax cultivation, few molecular resources are available to enhance breeding programs. In 2009, the Total Utilization Flax GENomics (TUFGEN; http://www.tufgen.ca) project was initiated in Canada to address this gap by generating genomic and molecular resources for flax and applying them for its improvement as a total utilization crop.

2.3 Linseed oil and fatty acid composition

During the last two decades, the popularity of flax as a food source has increased tremendously, mostly because of its nutritional attributes such as omega (ω)-3 FAs, lignans and dietary fibers (Muir and Westcott 2003). Current linseed varieties have oil content up to 50% of which the quality is largely determined by its FA composition (Green 1986). Linseed oil is composed of five major FAs: palmitic (PAL, C16:0; ~6%), stearic (STE, C18:0; ~4.4%), oleic (OLE, C18:1cis$^{\Delta 9}$; ~24.2%), linoleic (LIO, C18:2cis$^{\Delta 9,12}$; ~15.3%) and linolenic (LIN, C18:3 cis$^{\Delta 9,12,15}$; ~50.1%) (Muir and Westcott 2003). Flax is one of the richest sources of alpha-linolenic acid (ALA), the parent fatty acid of the ω-3 family (Fofana et al. 2004). The ALA content in high-LIN varieties developed by flax breeders can be as much as 65–73% (Friedt et al. 1995; Kenaschuk 2005) whereas traditional linseed varieties contain 50-59% ALA. While the high level of ALA is desirable in terms of nutrition, it has hindered food usage of flax oil because it is the reason behind rapid oxidation and instability during frying. On the other hand, this high degree of unsaturation is what makes linseed oil valuable in dry oil applications such as paints, linoleum flooring, inks, soaps and varnishes (Cullis 2007). Mutants with

reduced LIN content (29%) were first obtained using ethyl methanesulfonate (EMS) mediated mutagenesis of the flax cultivar Glenelg (Green and Marshall 1984). Flax cultivars with only 2-4% LIN (Solin^{TM} type) have also been developed through mutation breeding (Green 1986, Rowland 1991) specifically for the fabrication of margarine (Dribnenki and Green 1995; Dribnenki et al. 2007).

It is the proportions of the three major unsaturated FAs, OLE, LIO and LIN that largely determine the end-use of linseed oil for both food and industrial applications. LIO and LIN, important plant polyunsaturated fatty acids (PUFAs), are involved in plant metabolism as essential components of cell membranes, as a source of stored energy in the form of triacylglycerols (TAGs) and as precursors of signaling molecules such as jasmonic acid (Ohlrogge and Browse 1995). LIO and LIN cannot be synthesized by mammals; hence they are by definition considered essential FAs (EFAs). Upon ingestion, LIO and LIN can be further metabolized to produce other very long chain PUFAs (VLCPUFAs) like ecosapentaenoic acid (EPA, C22:5), docosahexaenoic acid (DHA, C22:6) and arachidonic acid (AA, C20:4) (Warude et al. 2006; Dyer et al. 2008). These VLCPUFAs are essential structural components of biological membranes that confer flexibility, fluidity and selective permeability to cellular membranes, and their roles in preventing cardiovascular diseases and reducing bad cholesterol levels have been shown (Ander et al. 2004; Wiesenfeld et al. 2003). VLCPUFAs can be further metabolized to produce a variety of lipid signaling molecules such as eicosanoids that affect growth, development and physiological well-being (Dyer et al. 2008). DHA has a vital role in brain development in infants and in normal brain function in adults (Martinetz 1992). The recommended ratio of ω6/ω3 FAs in the human diet is approximately 2:1 to 6:1 and the

insufficient amount of ω3 FAs in the typical Western diet is a major concern for cardiovascular health (Simopoulos 2000; Lands 2001).

2.4 Oils and fatty acid biosynthesis in plants

2.4.1 Oils

Oils represent an important form of carbon storage in many plants and constitute up to 80% of the total dry matter of a seed. With few exceptions, such as the waxes of jojoba oil, oils consist mainly of TAGs which are esters containing three FAs with chain lengths of C8–C24, with C16 and C18 being predominant. TAGs accumulate during the maturation phase of seed in the embryo, endosperm or both where this oil is an energy source during seed germination (Baud and Lepiniec 2010; Dyer et al. 2008). The usefulness of seed oil depends on its acyl composition such as the number and position of double bonds and the type of functional group.

While plant oils are used primarily for food and feed purposes, they are increasingly being utilized as renewable sources of industrial feedstock and fuel. The four most important oil crops, namely palm, soybean, rapeseed and sunflower, account for approximately 75 percent of the world's production (Dyer et al. 2008). The remaining oil crops account for less than 25% of the market. The demand for vegetable oils has increased rapidly in the past decade. The world production of oilseeds reached 121 million tonnes in 2011, an increase of approximately 50 million tonnes compared to the previous decade (FAOSTAT 2014).

2.4.2 Fatty acid biosynthesis

The main FA biosynthesis pathway is a primary metabolic pathway because it is essential for plant growth. The five major FAs: PAL, STE, OLE, LIO and LIN make up more than 90% of the acyl chains of the glycerolipids found in plant membranes (Ohlrogge and Browse 1995). In membrane glycerolipids, two FAs are esterified at the sn-1 and sn-2 positions of the glycerol backbone and, a polar head group is attached to the sn-3 position. TAGs, the major form of lipid storage in seeds, are formed when all three positions are esterified with FAs (Ohlrogge and Browse 1995).

FA biosynthesis in plants occurs primarily in two subcellular compartments: plastids and the ER; and the biosynthesis is catalyzed by the sequential action of the FA synthase complex and desaturation enzymes located in both compartments (Weselake et al. 2009). The *de novo* synthesis of PAL, STE and OLE and the desaturation of STE into OLE occur in plastids where FAs are attached to acyl-carrier proteins (ACPs) (Dyer et al. 2008). In higher plants, FA synthesis starts in the plastids with the formation of malonyl-CoA from acetyl-CoA and bicarbonate and this ATP-dependent reaction is catalyzed by plastidial acetyl-CoA carboxylase (ACCase) (Turnham and Northcote 1983). The malonyl group of CoA is then transferred to a protein cofactor called ACP by an enzyme called malonyl-CoA:acyl carrier protein S-malonytransferase (MAT). The multi-subunit enzyme complex termed fatty acid synthase (FAS) is responsible for the consecutive attachment of two carbon units to a growing fatty acid chain using acetyl-CoA as a starting unit and malonyl-ACP as the elongator, yielding palmitoyl-ACP and stearoyl-ACP. The initial condensation reaction is catalyzed by 3-ketoacyl-ACP synthase type III (KAS III) and recurring condensation with acetyl-CoA up to C16:0-ACP is catalyzed by KAS I/KAS B isoforms. The final elongation step from 16:0-ACP to C18:0-ACP is

catalyzed by KAS II/KAS A isoforms. In the plastid stroma, most of the stearoyl-ACP is efficiently desaturated by plastidial stearoyl-ACP desaturase (SAD) which inserts a double bond at the Δ9 position yielding oleoyl-ACP as the main product of the plastidial fatty acid biosynthesis (Shanklin and Cahoon 1998; Voelker and Kinney 2001). Following the synthesis of acyl-ACP pools (C16:0, C18:0 and C18:1cis$^{\Delta 9}$), a portion of the resulting FAs is retained in the plastid for the production of plastid glycerolipids (prokaryotic pathway) or acyl groups that are hydrolyzed by acyl-ACP thioesterases (FATA or FATB) and exported to the cytoplasm for further elongation and desaturation on the ER (Browse and Somerville 1991; Ohlrogge and Browse 1995). According to Schnurr et al. (2004), 38% of the *de novo* synthesized acyl-ACP pool (C16:0, C18:0 and C18:1cis$^{\Delta 9}$) remains in the chloroplasts and 62% is cleaved from ACP by acyl-ACP thioesterases in *Arabidopsis* mesophyll cells. The free FAs are converted to acyl-CoA on the outer membrane of the chloroplast by long-chain-acyl-CoA synthetase (LACS) and channelled into the cytoplasm prior to their export towards the ER. In plants, oleoyl-CoA exported to the ER, is further desaturated into PUFAs by membrane-bound fatty acid desaturases (FAD). OLE (C18:1), is first incorporated into phosphatidylcholine (PC) in the ER. It is then desaturated into LIO (C18:2; ω-6 FA) by fatty acid desaturase 2 (FAD2) and can be further desaturated into LIN (C18:3; ω-3 FA) by fatty acid desaturase 3 (FAD3) through the introduction of double bonds at the Δ12 and Δ15 positions in the fatty acid acyl chain, respectively.

2.4.3 Fatty acid desaturation

Fatty acid desaturases (FADs) are ubiquitous and found in all groups of organisms. Desaturases are classified into two groups: the soluble desaturases found in plants and the particulate or integral membrane desaturases found in yeast and mammals (Shanklin and Cahoon 1998; Sperling et al. 2003). Free fatty acids are esterified to different substrates including acyl carrier protein (ACP) for the soluble plastid desaturases, or to coenzyme A (CoA) or lipids for the integral membrane desaturases (Murata and Wada 1995; Shanklin and Cahoon 1998).

Desaturation is an important biochemical process in the FA biosynthesis pathway because it determines saturated and unsaturated fatty acid content and ultimately the end-use of the oil (Knutzon et al. 1992; Mikkilineni and Rocheford 2003). In the FA biosynthetic pathway, FADs are considered the key enzymes that drive the entire pathway leading to the synthesis of PUFAs (Rajwade et al. 2014). FADs can introduce double bonds into the hydrocarbon chains of FAs to produce unsaturated FAs. They are ion containing enzymes that convert single bonds (C-C) into double bonds (C=C) via an oxygen dependent dehydrogenation reaction that occurs at a specific position in the FA acyl chain (Shanklin and Cahoon 1998; Los and Murata 1998). The FA desaturation reaction requires two electrons in addition to one molecule of oxygen. Ferredoxin is used as the electron donor in the desaturation reactions catalyzed by acyl-ACP desaturases and acyl-lipid desaturases in cyanobacteria and in the plastids of plants (Wada et al. 1993). The desaturation reactions catalized by the acyl-lipid desaturases, localized in the cytoplasm of plants, and the acyl-CoA desaturases of animals and fungi, use cytochrome b_5 as the electron donor (Jaworski 1987; Shanklin and Cahoon 1998; Los and Murata 1998).

FAD2 and FAD3 are membrane-bound proteins with three highly conserved HIS-box motifs essential for enzyme activity (Shanklin at al. 1994; Los and Murata 1998) while SAD, the only known soluble desaturase, has two characteristic HIS-box motifs (Singh et al. 1994; Luo et al. 2009; Shilman et al. 2011). Genes encoding desaturase enzymes involved in fatty acid biosynthesis pathway have been cloned and characterized from several plant species including flax (Browse and Somerville 1991; Voelker and Kinney 2001; Thelen and Ohlrogge 2002). Commonly, in plants cells, at least two isoforms of each desaturase enzyme that differ in their cellular location (plastid and ER), lipid substrate and electron donor system, are present (Harwood 1996; Shanklin and Cahoon 1998).

2.5 Genetic control of fatty acid biosynthesis in flax

In flax, many of the genes encoding the enzymes that perform FA synthesis have been identified and characterized (Green 1986; Fofana et al. 2004; Sorensen et al. 2005; Vrinten et al. 2005; Fofana et al. 2006; Krasowska et al. 2007; Khadake et al. 2009; Banik et al. 2011). SAD, through its desaturation activity, has the potential to increase the unsaturated FA content of plants (Ohlrogge and Jaworski 1997). Two paralogous *sad* loci, *sad1* and *sad2*, differentially expressed in plants, have been identified in flax (Jain et al. 1999). Singh et al. (1994) reported the isolation and characterization of a cDNA sequence encoding the SAD protein from flax cultivar Glenelg and Fofana et al. (2004), from AC McDuff. Two closely related *fad2* genes, namely *fad2a* and *fad2b*, were cloned and characterized from flax genotypes Nike and NL97 (Krasowska et al. 2007; Khadake et al. 2009). Three *fad3* genes have been identified in the flax genome: *fad3a* and *fad3b*

from cultivar Normandy (Vrinten et al. 2005) and more recently *fad3c* from flax cultivar AC McDuff and breeding lines UGG5-5 and SP2047 (Banik et al. 2011). FAD3A and FAD3B have been shown to be the major enzymes controlling the LIN content of the storage lipids in flax seeds (Vrinten et al. 2005) while a major role for FAD3C has not been established. Recently, an *in silico* gene mining approach was used to identify genome wide putative gene families involved in FA biosynthesis from flax cv. CDC Bethune (You et al. 2014). Two new genes from the SAD family, 13 new genes from the FAD2 family and three new genes from the FAD3 gene family were identified but their roles in fatty acid composition were not determined. Structurally, they were all present as duplicated copies indicative of recent whole genome duplication events (You et al. 2014).

GENETIC VARIATION OF SIX DESATURASE GENES IN FLAX AND THEIR IMPACT ON FATTY ACID COMPOSITION

Dinushika Thambugala[1,4], Scott Duguid[2], Evelyn Loewen[2], Gordon Rowland[3], Helen Booker[3], Frank M You[4], Sylvie Cloutier[1,4]

[1]*Department of Plant Science, University of Manitoba, 66 Dafoe Rd, Winnipeg, MB, Canada, R3T 2N2*

[2]*Morden Research Station, Agriculture and Agri-Food Canada, 101 Route 100, Unit 100, Morden, MB, Canada, R6M 1Y5*

[3]*Crop Development Centre, University of Saskatchewan, 51 Campus Drive, Saskatoon, SK, Canada, S7N 5A8*

[4]*Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Rd, Winnipeg, MB, Canada, R3T 2M9*

Author Dinushika Thambugala conducted this work as part of her PhD thesis. Dinushika Thambugala carried out the experiment, analysed and interpreted the data and co-wrote the manuscript. The major supervisor Dr. Sylvie Cloutier designed the experiments, participated in the interpretation of data and co-wrote the manuscript. Dr. Scott Duguid, Dr. Gordon Rowland, Dr. Helen Booker and Evelyn Loewen designed and conducted the field experiments and generated the phenotypic data. Dr. Frank M You participated in data analysis and its description in the manuscript. All the authors read and approved the final manuscript. **The manuscript was published in Theoretical and Applied Genetics 2013, 126:2627-2641**

3.0 GENETIC VARIATION OF SIX DESATURASE GENES IN FLAX AND THEIR IMPACT ON FATTY ACID COMPOSITION

3.1 Abstract

Flax (*Linum usitatissimum* L.) is one of the richest plant sources of omega-3 fatty acids praised for their health benefits. In this study, the extent of the genetic variability for genes encoding stearoyl-ACP desaturase (SAD), fatty acid desaturase 2 (FAD2) and 3 (FAD3) was determined by sequencing the six paralogous genes from 120 flax accessions representing a broad range of germplasm including some EMS mutant lines. A total of 6 alleles for *sad1* and *sad2*, 21 for *fad2a*, 5 for *fad2b*, 15 for *fad3a* and 18 for *fad3b* were identified. Deduced amino acid sequences of the alleles predicted 4, 2, 3, 4, 6, and 7 isoforms, respectively. Allele frequencies varied greatly across genes. *Fad3a*, with 110 SNPs and 19 indels, and *fad3b*, with 50 SNPs and 5 indels, showed the highest levels of genetic variation. While most of the SNPs and all the indels were silent mutations, both genes carried non-sense SNP mutations resulting in premature stop codons, a feature not observed in *sad* and *fad2* genes. Some alleles and isoforms discovered in induced mutant lines were absent in the natural germplasm. Correlation of these genotypic data with fatty acid composition data of 120 flax accessions phenotyped in six field experiments revealed statistically significant correlations of some of the SAD and FAD isoforms on fatty acid composition, oil content and iodine value. The novel allelic variants and isoforms identified for the six desaturases will be a resource for the development of oilseed flax with unique and useful fatty acid profiles.

3.2 Introduction

Flax (*Linum usitatissimum* L.) is an annual, self-pollinating, diploid (2n=2x=30) crop belonging to the *Linaceae* family. Flax has been grown for its stem fibres (fibre flax) or its seed oil (linseed or oilseed flax) for several thousand years (Zohary 1999). During the last two decades, flax has attracted great attention in human health mostly because of its desirable fatty acid composition. Current linseed varieties have oil content up to 50% (Cloutier et al. 2010) and the major fatty acids are palmitic (PAL, C16:0; ~6%), stearic (STE, C18:0; ~4.4%), oleic (OLE, C18:1cis$^{\Delta 9}$; ~24.2%), linoleic (LIO, C18:2cis$^{\Delta 9,12}$; ~15.3%) and linolenic (LIN, C18:3 cis$^{\Delta 9,12,15}$; ~50.1%) (Muir and Westcott 2003). Flax is the leading source of plant based omega-3 fatty acids. Alpha-linolenic acid (ALA), the parent fatty acid of the omega-3 family, constitutes up to 73% of the total fatty acids in high-lin varieties, whereas traditional linseed varieties have 50-59% ALA. Solin type flax varieties are rich in LIO, the parent fatty acid of the omega-6 family, and contain generally 2-4% ALA (Fofana et al. 2010).

LIO and LIN, important plant polyunsaturated fatty acids (PUFAs), are involved in plant metabolism as structural components, as a source of energy storage in the form of triacylglycerols (TAGs), as essential components of cell membranes and as precursors of signaling molecules such as jasmonic acid (Ohlrogge and Browse 1995). Mammalian tissues cannot synthesize LIO and LIN and hence these fatty acids are considered essential. Upon ingestion, LIO and LIN can be further elongated and desaturated to form other long chain PUFAs (LCPUFAs) like ecosapentaenoic acid (EPA, C22:5), docosahexaenoic acid (DHA, C22:6) and arachidonic acid (AA, C20:4) (Warude et al. 2006). These LCPUFAs are essential structural components of biological membranes,

especially in the brain and the retina, and, are associated with developmental and physiological processes that affect human health (Dyer et al. 2008). DHA plays a vital role in brain development in infants and in normal brain function in adults (Martinetz 1992). LCPUFAs are important in maintaining the flexibility, fluidity and selective permeability of cellular membranes and their roles in preventing cardiovascular diseases and reducing bad cholesterol levels have been shown (Ander et al. 2004; Wiesenfeld et al. 2003). The insufficient amount of ALA in the typical Western diets is a major concern in cardiovascular diseases (Simopoulos 2000; Lands 2001).

Fatty acid desaturases and elongases are key enzymes involved in the fatty acid biosynthesis pathway (Warude et al., 2006). During oil biosynthesis in plants, the stepwise desaturation of fatty acids is an important process that determines the saturated to unsaturated fatty acid ratio and, ultimately, the end-use of the oil as a food source or for industrial applications (Knutzon et al. 1992; Mikkilineni and Rocheford 2003). Fatty acid desaturases are responsible for the insertion of double bonds into the hydrocarbon chain of fatty acids (Shanklin and Cahoon 1998; Los and Murata 1998). Fatty acid desaturases FAD2 and FAD3 are membrane-bound proteins with three highly conserved histidine box motifs essential for the enzyme activity (Shanklin at al. 1994; Los and Murata 1998) while Stearoyl-ACP desaturase (SAD) is the only known soluble desaturase with two characteristic HIS-box motifs (Singh et al. 1994; Luo et al. 2009; Shilman et al. 2011). Genes encoding desaturases involved in the fatty acid biosynthesis pathway have been cloned and characterized from many species (Chi et al. 2008; Lu et al. 2010; Chen et al. 2010).

Many of the genes encoding the enzymes that perform *de novo* fatty acid biosynthesis in flax have also been identified and characterized (Green 1986a; Fofana et al. 2004; Sørensen et al. 2005; Vrinten et al. 2005; Fofana et al. 2006; Krasowska et al. 2007; Khadake et al. 2009; Banik et al. 2011). SAD is responsible for converting stearoyl-ACP to oleoyl-ACP by introducing a double bond at the Δ9 position and thereby has the potential to increase the unsaturated FA content of the plant (Ohlrogge and Jaworski 1997). Two paralogous *sad* loci, *sad1* and *sad2*, differentially expressed in plants, have also been identified in flax (Jain et al. 1999). Singh et al. (1994) reported the isolation and characterization of a cDNA sequence encoding the SAD protein from flax cultivar Glenelg and Fofana et al. (2004), from AC McDuff. The *fad2* genes encode proteins responsible for desaturation of OLE into LIO by addition of a double bond at the Δ12 position. Two closely related *fad2* genes namely, *fad2a* and *fad2b*, were cloned and characterized from flax genotypes Nike and NL97 (Krasowska et al. 2007; Khadake et al. 2009). The *fad3* genes encode proteins responsible for the desaturation of LIO into LIN by performing the addition of a double bond at the Δ15 position. Three *fad3* genes have been identified in the flax genome: *fad3a* and *fad3b* from cultivar Normandy (Vrinten et al. 2005) and more recently *fad3c* (Banik et al. 2011). FAD3A and FAD3B have been shown to be the major enzymes controlling the LIN content of the storage lipids in flax seeds (Vrinten et al. 2005) while a major role for FAD3C has not been established.

While our knowledge of the major desaturase genes and enzymes involved in the fatty acid composition of flax seeds is good, little is known about the extent of the genetic variability of these genes, their corresponding isoforms and their relationship to fatty acid composition. The major aim of this study was to determine the genetic variation for *sad*,

fad2 and *fad3* genes in flax by sequencing these genes from 120 flax accessions. These genotypic data were correlated to the fatty acid composition phenotyped in multiple field experiments during three years at two locations to hypothesize the functionality of these alleles and isoforms.

3.3 Materials and Methods

3.3.1 Plant material and DNA extraction

A total of 120 *Linum usitatissimum* (L.) accessions representing both oil and fibre types of flax were selected for this study (Appendix I). Seeds obtained from the Plant Gene Resources of Canada (PGRC) were grown in a greenhouse. DNA was extracted from lyophilized young leaf tissues using the DNeasy 96 Plant kit (Qiagen, Missisauga, ON, Canada) according to the manufacturer's instructions and quantified by fluorometry.

3.3.2 Primer design and PCR amplification

Gene-specific primers used for PCR amplification of *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* were designed based on the 5' and 3'-UTR regions of their Genbank genomic sequences using the Primer3 software (Rozen and Skaletsky, 2000) (Appendix II). PCR reactions were carried out in a final volume of 10 µl containing 40 ng of total genomic DNA, 0.4 µM each primer, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.8 mM dNTPs, 0.1 µl of 10X BSA (1mg/ml) and 1 Unit Taq DNA polymerase. PCR reactions were performed using the following conditions: an initial denaturation of 4 min at 94°C followed by 35 cycles at 94°C for 30 s, 60°C for 30 s, 72°C for 1-3 min depending on the target and a final extension of 10 min at 72°C. A total of 6-12 independent PCR reactions were

performed for each target gene of each genotype. The PCR products from each gene/genotype were pooled and aliquots were visualized by agarose gel electrophoresis to verify amplicon specificity.

3.3.3 DNA sequencing

The pooled PCR amplicons were purified with Multiscreen$_{384}$-PCR filter plates according to manufacturer's instructions (Millipore Corp., Billerica, MA, USA). Aliquots of each purified pooled PCR amplicon were resolved on 1% agarose gels to estimate DNA concentration. Aliquots of the purified PCR amplicons were sequenced with Big-Dye V3.1 Terminator chemistry (Applied BioSystems, Foster City, CA, USA) using amplicon specific primers designed to span the entire amplicons with overlap in both orientations (Appendix II). Sequencing reactions were performed in a volume of 6 µL containing 40 ng of purified amplicons, 1 µl of 5X sequencing buffer, 8.7 µM primer and 0.4 µl BigDye reaction mix. Reactions were carried out under the following conditions: an initial denaturation of 5 min at 92°C followed by 60 cycles at 92°C for 10 s, 55°C for 5 s, 60°C for 4 min and a final extension step of 10 min at 60°C. Unincorporated dideoxynucleotides were removed by ethanol precipitation prior to resolution of the sequences on an ABI 3130xl Genetic Analyzer (Huang and Cloutier, 2008).

3.3.4 Genetic diversity analysis

DNA trace files from the ABI 3130xl Genetic Analyzer were processed and assembled using an internal data pipeline called SOOMOS v0.6 (T. Banks, personal communication) which implement the base calling software PHRED (Ewing *et al.* 1998) and the assembly

software CAP3 (Huang and Madan 1999). Multiple alignment, translation and identification of open reading frames (ORFs) were conducted using clustalW v1.82 (Thompson et al. 1994) and DNAMAN v3.2 (Lynnon Corp., Vaudreuil-Dorion, Quebec, Canada). Assemblies were manually curated to correct sequencing errors.

3.3.5 Phylogenetic analysis

Phylogenetic analyses were performed using MEGA 4.0 (Tamura *et al*., 2007). Phylogenetic trees based on the alignment of full length DNA sequences of each gene were constructed using the Neighbor-joining (NJ) algorithm (Saitou and Nei, 1987) as implemented in MEGA 4.0. Bootstrap values were estimated using 1,000 replications.

3.3.6 Field trials and phenotyping of fatty acid compositions

The 120 flax accessions were grown in a type 2 modified augmented design (MAD) (Lin and Pouschinsky, 1985) at the Kernen farm near Saskatoon (SK, Canada) and at the Morden Research Station (MB, Canada) in 2009, 2010 and 2011. In the MAD, plots were arranged in 10 × 10 grids and each main plot was split into five subplots where the central subplot was occupied by the main plot control cultivar ‘CDC Bethune’. Two additional subplot controls, ‘Macbeth’ and ‘Hanley’, were assigned to two random subplots of five randomly selected whole plots. The 120 flax accessions were randomly allocated to the remaining subplots. The design and assignment of flax accessions were conducted using the Agrobase software (Agronomix Software Inc, Winnipeg, Canada). Oil content (OIL) was determined by Nuclear Magnetic Resonance calibrated against the FOSFA (Federation of Oils, Seeds and Fats Associations Limited) extraction method.

Fatty acid profiles were obtained using fatty acids methyl esters (FAMEs) extracted from seeds (AOAC method 996.06) (Daun and Mazur 1983). FA composition of each line was measured on a Varian 3800 gas chromatograph (GC) (Varian Analytical Instruments, Mississauga, ON, Canada). FA compositions of PAL, STE, OLE, LIO and LIN were expressed as a percentage of the total fatty acid composition. Iodine value (IOD) was calculated from the GC determined fatty acid composition (AOCS Method Cd 1c-85).

3.3.7 Statistical analysis

All observed values for FA composition, OIL and IOD obtained from six individual experiments (three years and two locations; Appendix III) were analyzed individually and adjusted for soil heterogeneity based on the MAD statistical analysis method and pipeline programs described by You et al. (2013). To assess the differences among isoforms, one-way analysis of variance (ANOVA) with unequal sample sizes was used, followed by the Duncan's multiple range comparison tests at 0.05 probability level. All statistical analyses were carried out using SAS v9.2 (SAS Institute, Cary, USA). The ANOVAs were repeated excluding the seven EMS mutant lines from the data set.

3.4 Results

DNA sequences spanning the entire coding region of *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* from 120 flax genotypes were obtained. BLASTN and BLASTX searches against the NCBI non-redundant (nr/nt) database were used to identify the coding regions and the open reading frames of each gene. Exon and intron structure were determined and amino acid sequences were deduced. Alleles were numbered and isoforms were identified with

letters. The number of alleles ranged from five for *fad2b* to 21 for *fad2a* and the number of corresponding isoforms ranged from 2 to 7 (Table 3.1).

Table 3.1 Allelic diversity for six fatty acid desaturases and their deduced isoforms

Gene	Length	Exons	Introns	Exon Length	Amino Acids	Alleles	Isoforms
sad1	2515	3	2	1191	396	6	4
sad2	2519	3	2	1191	396	6	2
fad2a	1137	1	-	1137	378	21	3
fad2b	1149	1	-	1149	382	5	4
fad3a	3280	6	5	1179	392	15	6
fad3b	3002	6	5	1176	391	18	7

3.4.1 *Sad1* and *sad2*

The two *sad* genes shared a similar overall structure with three exons and two introns (Fig. 3.1a, 3.1b). The length of the coding region was 2,515 bp for *sad1* and 2,519 bp for *sad2*. Both genes encode proteins of 396 amino acid residues and share 91% identity at the DNA level and 99% at the amino acid level (Appendix IV). A total of 10 SNPs in the *sad1* coding region defined six alleles (Table 3.2). Three mutations were missense and seven were silent (Fig. 3.1a). Allele 1 was found in 109/120 accessions while the other five alleles were present in only one to five accessions. Alleles 1, 2 and 3 encoded isoform A present in 116 accessions. Alleles 4, 5 and 6 each caused amino acid substitutions resulting in isoform B, C and D, respectively (Fig. 3.1a). Seven SNPs were identified in *sad2* thus defining six alleles (Fig. 3.1b, Table 3.2). Allele 1, 2 and 3 were found in 51, 33 and 31 accessions, respectively, while the other three were rare. Of the two SNPs in exon 3, only one caused an amino acid substitution of a glycine to a serine, thus defining isoform A and B, present in 86 and 34 accessions, respectively (Fig. 3.1b).

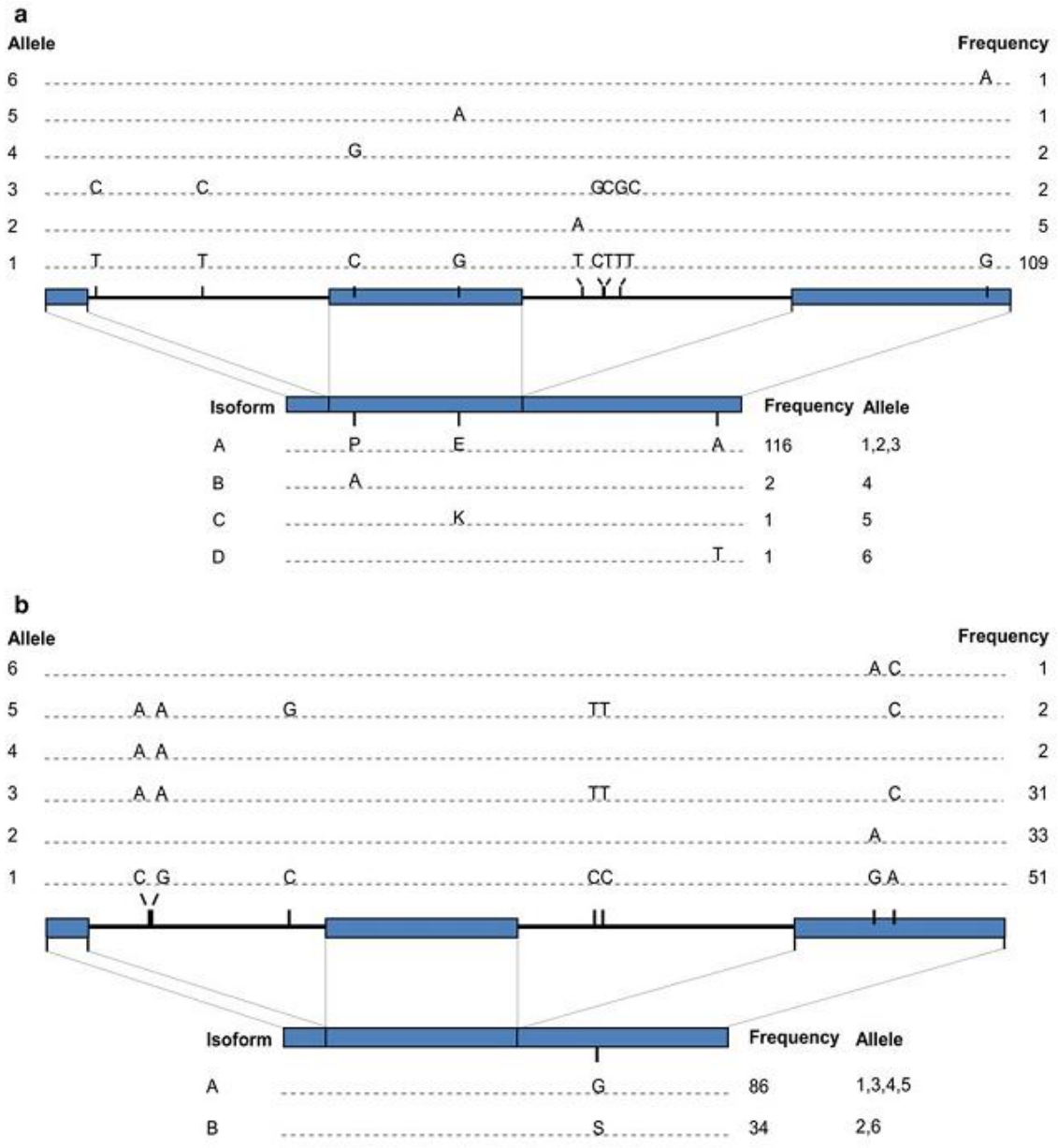


Fig 3.1 Schematic diagram of alleles and predicted isoforms of six desaturase genes and proteins obtained from sequencing these genes in 120 flax genotypes. SNPs and indels defining the alleles and their frequency in the germplasm are illustrated in the upper panels. Amino acid substitutions defining the isoforms, their frequency and corresponding alleles are illustrated in the lower panels. Exons are drawn as boxes and introns as lines. (▼) Deletions are represented by inverted triangles with the number of bases deleted. a) *sad1*, b) *sad2*, c) *fad2a*, d) *fad2b*, e) *fad3a*, f) *fad3b*. *Fad3a* alleles 13, 14, and 15 are not illustrated because they were hypervariable.

Fig. 3.1 Continued

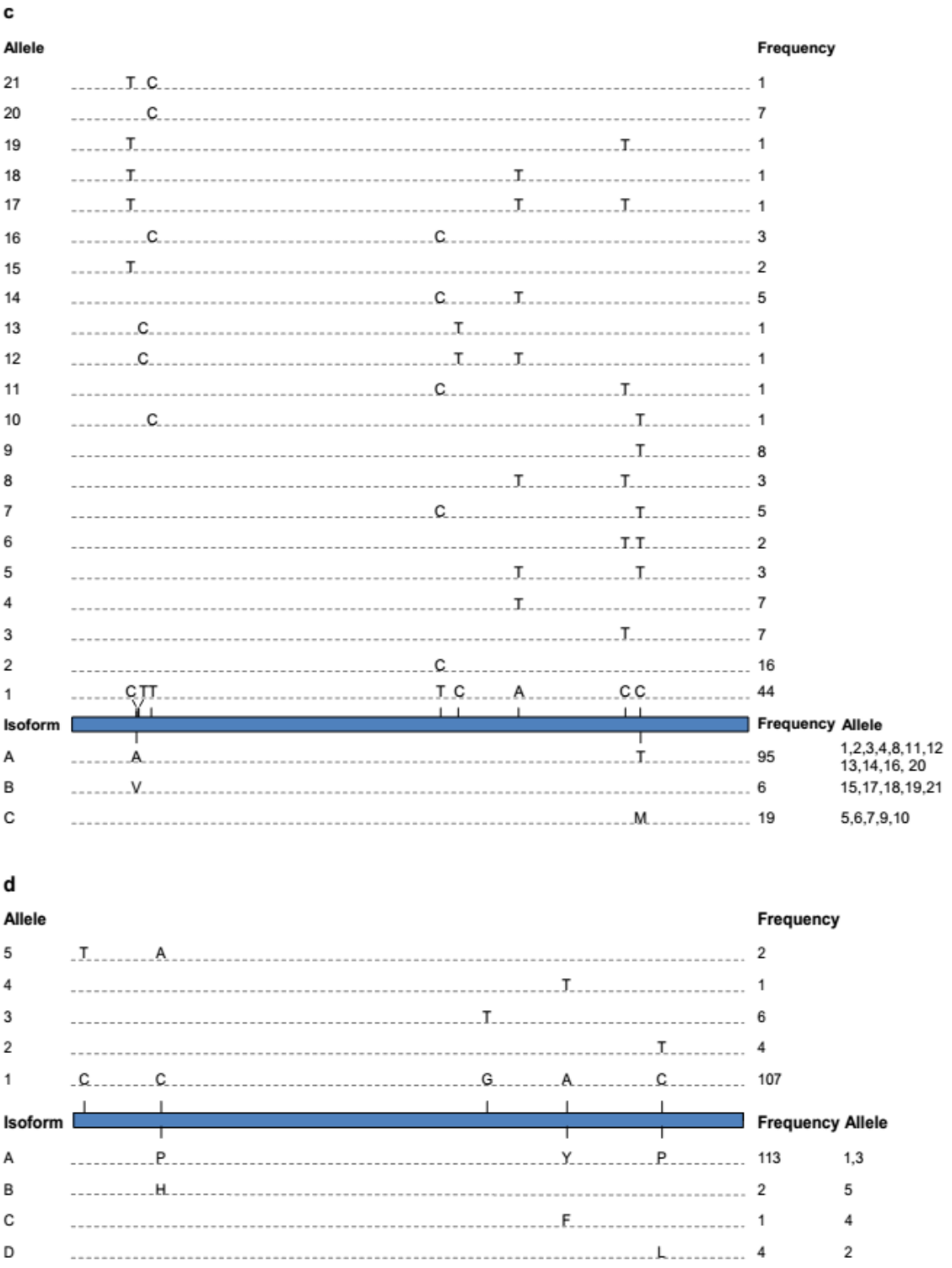
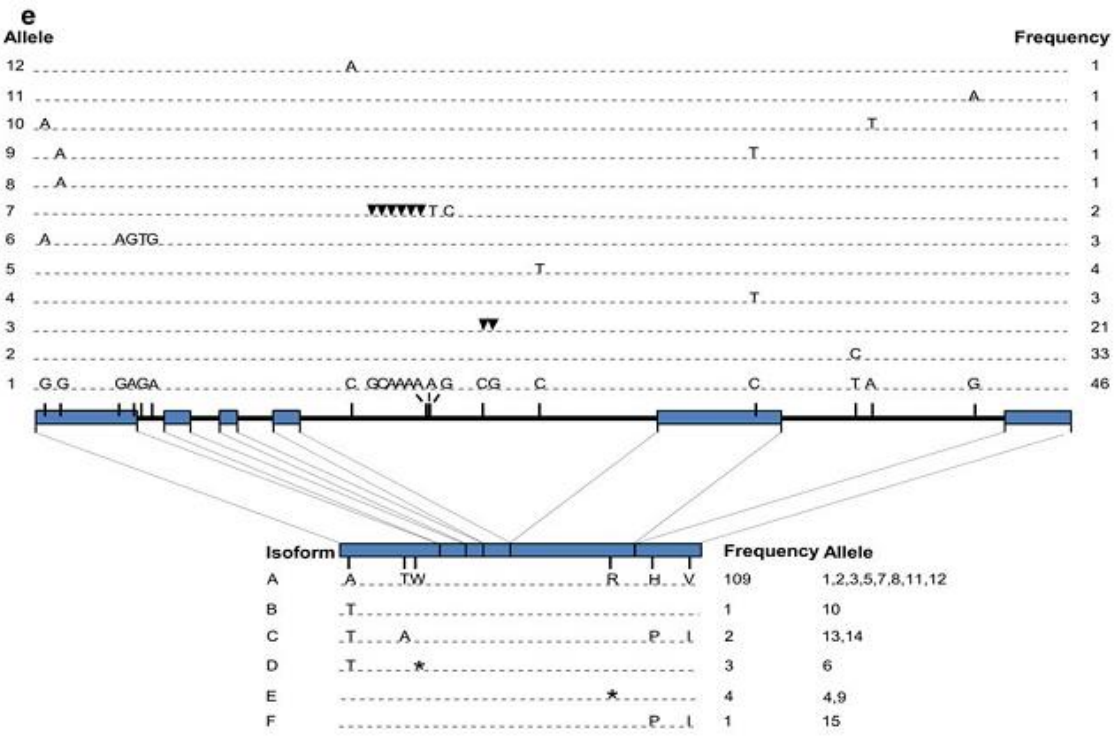


Fig 3.1 Continued


e
Allele
Frequency
12 A 1
11 A 1
10 A T 1
9 A T 1
8 A 1
7 T C 2
6 A AGTG 3
5 T 4
4 T 3
3 21
2 C 33
1 G G GAGA C GCAAAAAG CG C C TA G 46
Isoform
Frequency
Allele
A A TW R H V 109 1,2,3,5,7,8,11,12
B T 1 10
C T A P I 2 13,14
D T * 3 6
E * 4 4,9
F P I 1 15

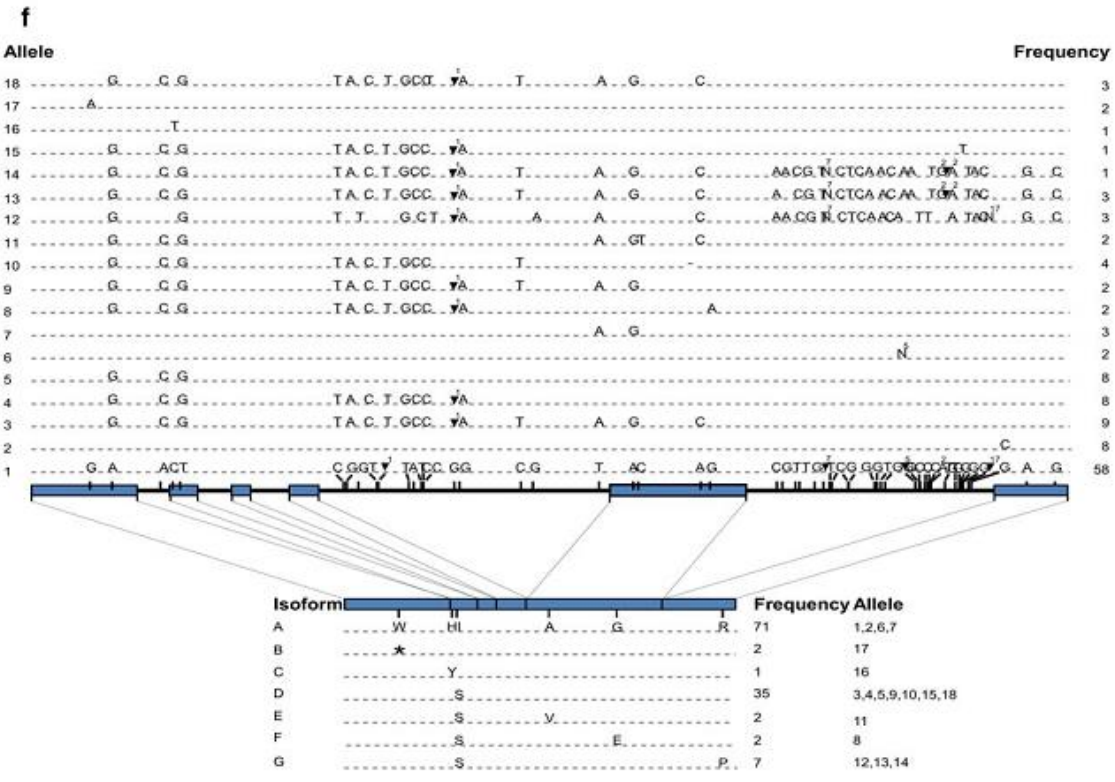

f
Allele
Frequency
18 G C G TA C T GCC A T A G C 3
17 A 2
16 T 1
15 G C G TA C T GCC A T 1
14 G C G TA C T GCC A T A G C AACG N CTCAACAA TGA TAC G C 1
13 G C G TA C T GCC A T A G C A CG N CTCA ACAA TGA TAC G C 3
12 G G T T GCT A A A C AACG N CTCAACA TT A TAN G C 3
11 G C G A GT C 2
10 G C G TA C T GCC T 4
9 G C G TA C T GCC A T A G 2
8 G C G TA C T GCC A A 2
7 A G 3
6 N 2
5 G C G 8
4 G C G TA C T GCC A 8
3 G C G TA C T GCC A T A G C 9
2 C 8
1 G A ACT C GGT TA CC GG C G T AC AG CGTTGT CG GGTG G 58
Isoform
Frequency
Allele
A W HI A G R 71 1,2,6,7
B * 2 17
C Y 1 16
D S 35 3,4,5,9,10,15,18
E S V 2 11
F S E 2 8
G S P 7 12,13,14

Table 3.2 SNPs and indels identified in *sad* and *fad* genes sequenced from 120 flax accessions

Gene	SNP location		SNP frequency	Indel location		Indel frequency
	Exon	Intron	(SNP/100bp)	Exon	Intron	(Indel/100bp)
sad1	3	7	0.40	-	-	-
sad2	2	5	0.28	-	-	-
fad2a	8	-	0.70	-	-	-
fad2b	5	-	0.44	-	-	-
fad3a	14	96	3.35	-	19	0.58
fad3b	10	40	1.66	-	5	0.17

3.4.2 *Fad2a* and *fad2b*

The *fad2a* and *fad2b* intron-less genes spanned 1,137 and 1,149 bp encoding proteins of 378 and 382 amino acid residues, respectively (Fig. 3.1c, 3.1d). The two sequences shared 82% identity at the DNA level and 87% at the amino acid level (Appendix V). Of the six desaturase genes sequenced herein, the *fad2a* gene had the most alleles with 21 as defined by eight SNPs (Table 3.2). The most frequent allele was present in 44 accessions while the remaining 20 alleles were found in one to 16 accessions (Fig. 3.1c). Despite the high allelic variation, only three isoforms of FAD2A were deduced because only two of the eight SNPs were non-synonymous. A total of five SNPs formed the five *fad2b* alleles, of which, three were non-synonymous, hence the four FAD2B isoforms (Fig. 3.1d, Table 3.2).

3.4.3 *Fad3a* and *fad3b*

The coding regions of *fad3a* and *fad3b* were 3,280 and 3,002 bp, respectively, both encompassing six exons and five introns (Fig. 3.1e, 3.1f). The deduced protein sequences had 392 and 391 amino acid residues. The two sequences displayed only 85% identity at the DNA level but 94% at the amino acid level (Appendix VI). Indels located in introns and ranging from one to 29 bp were responsible for the 278 bp length difference and

divergence between the two *fad3* paralogs (Table 3.2). Because three of the *fad3a* alleles were hypervariable, the assembly of their coding sequences yielded two distinct contigs. The first contig included the *fad3a* sequences from 117 accessions (Fig. 3.1e) while the remaining three formed the second contig (Appendix VII). Taken together, 110 SNPs and 19 indels were identified in this gene. However, the majority were the results of the hypervariable alleles 13, 14 and 15 as only 14 SNPs and two indels were detected in the remaining 12 alleles. Considering all 120 *fad3a* sequences, only 14 SNPs and no indels were detected in exons thereby encoding six isoforms, of which, isoform A was present in 109 accessions. The hypervariable alleles 13, 14 and 15 were predicted to encode two different isoforms (C and F). Isoform D and isoform E had premature stop codons, a feature not observed in *sad* and *fad2* genes.

Fifty SNPs and five indels defined 18 different alleles of *fad3b* (Table 3.2). The most common allele was present in nearly half of the accessions. Among the 50 SNPs, only six were non-synonymous (Fig. 3.1f). A transition from G to A in the first exon of allele 17 resulted in a nonsense mutation leading to a stop codon near the N-terminus of the FAD3B protein which is predicted to encode a truncated desaturase of 53 amino acids. Another single point mutation, identified in the second exon, was responsible for a histidine to tyrosine substitution in the first HIS-box (isoform C). Isoforms D, E, F and G all shared an isoleucine to serine substitution, but isoform E, F and G each had one additional but different amino acid substitution (Fig. 3.1f).

3.4.4 Phylogeny of desaturase genes

The Neighbor-joining trees of *sad1* and *sad2* had very similar topology with the clustering of each of the six alleles into two major clades (Appendix VIII). NJ trees derived from *fad2a* and *fad2b* sequences differed significantly in their topology (Appendix IX). The *fad2a* sequences formed four distinct clades including the number of sub-clades showing a higher sequence divergence among *fad2a* alleles. In contrast, the *fad2b* NJ tree showed a substantially reduced nucleotide diversity among alleles in which, four of the five alleles grouped together to form clade II (Appendix IX). NJ trees of *fad3a* and *fad3b* had two distinct clades with higher bootstrap support. However, the *fad3a* NJ tree showed less divergence between alleles with the exception of the hypervariable alleles which grouped together to form clade I. In contrast, the *fad3b* NJ tree showed significant sequence divergence between alleles implying higher accumulation of mutations through evolution (Appendix X). The third position of codon often creates noise in a dataset and interferes with phylogenetic signal. To test this, we repeated the phylogenetic analysis using deduced amino acid sequences of all six genes excluding the third position of codon. As illustrated by the NJ trees of all six desaturase genes, the noise associated with the third position did not interfere with phylogenetic signal (Appendix XI, XII, XIII).

3.4.5 Association between fatty acid composition and desaturase isoforms

A one-way ANOVA was conducted to compare the effect of the predicted isoforms of SAD, FAD2 and FAD3, individually and in combinations on the fatty acid composition, OIL and IOD. Although lines carrying SAD2 isoform B accumulated significantly more OLE, all SAD combinations and the two SAD1 isoforms individually had no significant

effect on OLE (Appendix XIV). However, the effect of SAD1/2 combinations was significant on STE (Table 3.3). Significant differences were observed between FAD2A/B combinations for OLE and OIL content and FAD2A isoforms for LIO and OIL content. Among the three different FAD2A isoforms, lines carrying isoform C accumulated significantly more LIO (Appendix XV). FAD3A/B isoform combinations were significant for LIN, LIO, OLE, PAL, OIL and IOD traits. FAD3A and FAD3B isoforms were also individually significant for LIN, LIO, PAL, OIL and IOD with FAD3A not being significant for OLE acid (Table 3.3). Lines carrying FAD3A isoforms D and E and FAD3B isoforms B, C and F individually as well as in combinations (EF, DC and EB) accumulated significantly less LIN (Fig. 3.2a, b, c). LIO content was significantly elevated in lines having the EF, DC and EB combinations revealing a strong inverse association between LIN and LIO (Fig. 3.2d).

We suspected that some of the significance was attributed to non-functional isoforms exclusively found in EMS mutant lines. To test this hypothesis, we repeated the one-way ANOVA using only the 113 non-mutant lines. This had the effect of eliminating isoforms D and E for FAD3A and B, C and F for FAD3B. Overall, fewer desaturases were significant for most of the traits but several significant associations found with the whole dataset remained significant with the reduced dataset (Appendix XVI). Of particular interest are the significant associations between FAD3B isoforms and PAL, SAD2 and FAD2A isoforms and OLE and FAD3B isoforms and LIO. Also, OIL was significantly correlated by FAD2A isoforms and FAD2A/B isoform combinations.

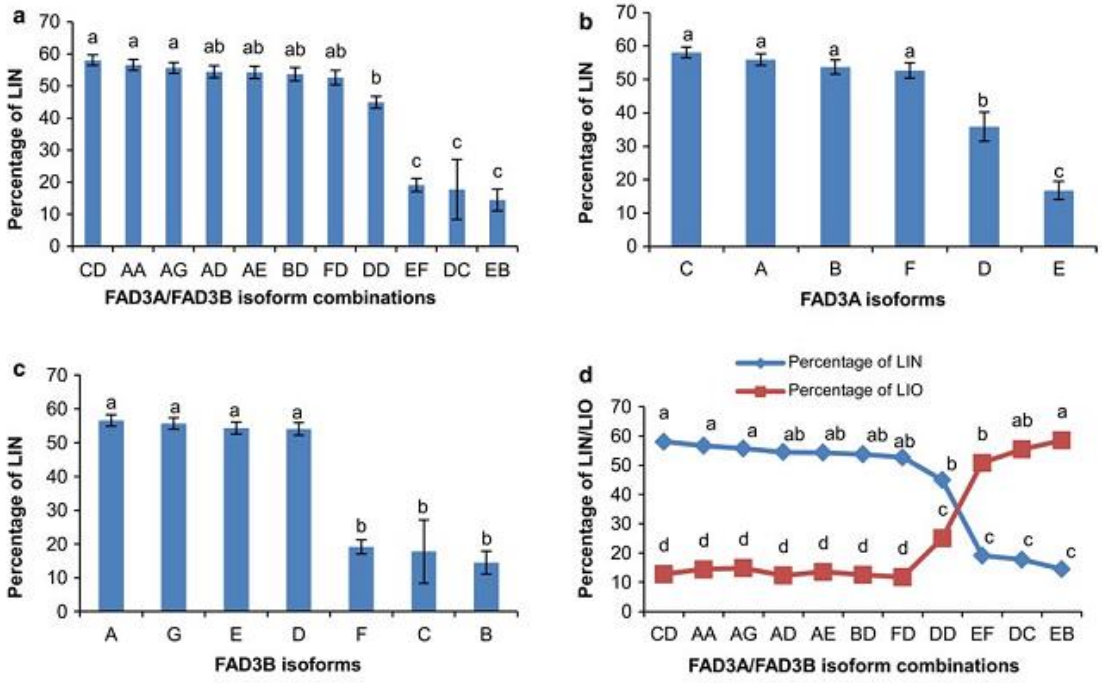


Fig 3.2 Association between the predicted isoforms of (a) FAD3A/B and LIN content, (b) FAD3A and LIN content, (c) FAD3B and LIN content and (d) FAD3A/B and LIN and LIO contents. Vertical bars represent standard error of the mean. Letters on top of the bar indicate statistical significance of Duncan's multiple range tests.

Table 3.3 Effect of *sad* and *fad* predicted isoforms on palmitic, stearic, oleic, linoleic and linolenic acid composition, oil content and iodine value

Trait	Gene	*P*-value	Predicted isoforms or combinations[†]
Palmitic acid (PAL)	*sad1*	0.8225	
	sad2	0.6520	
	fad2a	0.1886	
	fad2b	0.8978	
	fad3a	<.0001*	[E(7.84)][a], [A,B,C,D,F(5.96-5.59)][b]
	fad3b	<.0001*	[F(9.42)][a], [A,B,C,D,E,G(6.38-5.48)][b]
	sad1/sad2	0.8547	
	fad2a/fad2b	0.6547	
	fad3a/fad3b	<.0001*	[EF(9.42)][a],[AA,AD,AE,AG,BD,CD,DC,DD,EB,FD(6.37-5.48)][b]
Stearic acid (STE)	*sad1*	0.1731	
	sad2	0.0832	
	fad2a	0.2950	
	fad2b	0.1558	
	fad3a	0.8507	
	fad3b	0.2350	
	sad1/sad2	0.0425*	[CB(6.62)][a], [DB(4.60)][ab], [AA, AB,BB,(4.29-3.78)][b]
	fad2a/fad2b	0.3162	
	fad3a/fad3b	0.4552	

Trait	Gene	*P*-value	Predicted isoforms or combinations[†]
Oleic acid (OLE)	*sad1*	0.9703	
	sad2	0.0083*	[B(21.64)][a], [A(19.71)][b]
	fad2a	0.0067*	[A(20.79)][a], [C(18.35)][ab], [B(17.90)][b]
	fad2b	0.3225	
	fad3a	0.2275	
	fad3b	0.0009*	(16.48-22.46)[a]
	sad1/sad2	0.0944	
	fad2a/fad2b	0.0054*	(15.59-20.99)[a]
	fad3a/fad3b	0.0031*	[FD(24.18)][a], [BD(23.77)][ab], [AA,AD,AE,AG,CD,DD(22.79-19.23)][abc], [DC(17.08)][bc], [EB,EF(16.78-16.46)][c]
Linoleic acid (LIO)	*sad1*	0.9220	
	sad2	0.7126	
	fad2a	0.0076*	[C(21.45)][a], [A,B(14.76-12.91)][b]
	fad2b	0.9755	
	fad3a	<.0001*	[E(54.67)][a], [D(35.23)][b], [A,B,C,F(13.88-11.76)][c]
	fad3b	<.0001*	[B(58.52)][a], [C(55.49)][ab], [F(50.81)][b],[A,D,E,G(14.87-13.06)][c]
	sad1/sad2	0.9722	
	fad2a/fad2b	0.1268	
	fad3a/fad3b	<.0001*	[EB(58.52)][a], [DC(55.49)][ab], [EF(50.81)][b], [DD(25.11)][c], [AA,AD,AE,AG,BD,CD,FD(14.87-11.76)][d]
Linolenic acid (LIN)	*sad1*	0.9643	
	sad2	0.6721	
	fad2a	0.0707	
	fad2b	0.7270	
	fad3a	<.0001*	[A,B,C,F(58.08-52.70)][a], [D(35.86)][b], [E(16.80)][c]
	fad3b	<.0001*	[A,D,E,G(56.64-54.07)][a], [B,C,F(19.12-14.48)][b]
	sad1/sad2	0.9596	
	fad2a/fad2b	0.3172	
	fad3a/fad3b	<.0001*	[AA,CD,AG(58.08-55.70)][a], [AD, AE,BD,FD(54.484-52.70)][ab], [DD(44.91)][b], [DC,EB,EF(19.12-14.48)][c]
Oil content (OIL)	*sad1*	0.9779	
	sad2	0.9024	
	fad2a	0.0004*	[A,C(44.14-42.80)][a], [B(39.90)][b]
	fad2b	0.2108	
	fad3a	0.0044*	[C,D(47.49-45.48)][a], [A,B,E(43.18-42.71)][ab], [F(39.68)][b]
	fad3b	0.0136*	[C(48.46)][a],[A,B,D,E,F,G(43.81-42.31)][b]
	sad1/sad2	0.9955	
	fad2a/fad2b	0.0068*	[CA(44.20)][a], [AA,AB,AD,BA,CD(43.10-40.17)][ab],[BC(38.52)][b]
	fad3a/fad3b	0.0056*	[DC(48.46)][a], [CD,DD(47.00-45.48)][ab], [AD,AE,EF(43.81-43.54)][abc], [AA,AG,BD,EB(42.85-42.31)][bc], [FD(39.68)][c]
Iodine value (IOD)	*sad1*	0.9949	
	sad2	0.4642	
	fad2a	0.3869	
	fad2b	0.3647	
	fad3a	<.0001*	[C(190.68)][a], [A,B,F(187.97-179.11)][ab], [D(170.86)][b], [E(152.96)][c]
	fad3b	<.0001*	[A,E,D,G(189.88-183.40)][a], [C,B,F(157.33-152.15)][b]
	sad1/sad2	0.9516	
	fad2a/fad2b	0.3689	[AA,AD,AE,AG,BD,CD,DD,FD(190.68-177.63)][a],
	fad3a/fad3b	<.0001*	[DC,EB,EF(157.33-152.15)][b]

*Statistical significance ($p < 0.05$)

[†] Means for the isoform(s) or isoform combinations are in bracket. They represent data collected from two locations during three years

[a,b,c,d] Statistical significance of Duncan's multiple range tests

3.5 Discussion

FA desaturases introduce double bonds at specific locations of fatty acid acyl chains and, hence are considered targets for manipulation of fatty acid composition of oil seed crops. FA desaturases exhibit significant diversity in their sequences and expression (Loss and Murata 1998; Warude et al. 2006). Minor changes in the primary structure of proteins may result in modification of the enzyme function including altered substrate specificity, region selectivity or loss of function (Avelange-Macherel et al. 1995; Broadwater et al. 2002; Khadake et al. 2011. In the present study, the extent of the genetic variability for six desaturase genes (*sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b*) was determined by sequencing them from 120 genotypes of flax. DNA sequences were obtained to quantify the scope of the genetic variations and to predict structural changes of the encoded desaturases. These genotypic data were correlated to the fatty acid composition obtained from multi-year, multi-location field grown material to determine the functional role of these alleles and isoforms on fatty acid composition and oil content.

Sequence analysis of the six genes revealed a significant level of variation at the nucleotide level with SNPs being the most frequently observed mutation type. Most of these point mutations were synonymous substitutions that did not alter the underlying amino acid sequences. SNPs are the most abundant type of DNA variation of plant genomes (Brookes 1999; Wei et al. 2011). However, their frequency varies among the different plant species (Wei et al. 2011). Inbred rice and *Arabidopsis* display one SNP in every 300 bp (Schmid et al. 2003), while outbreeding maize has one SNP/60 bp (Ching et al. 2002). This higher SNP rate was observed in *fad3b* while the rate of *fad3a* was twice as high with one SNP every 30 bp. Nucleotide variations in exons were lower than in

introns and most were missense or silent mutations. Both SNPs and indels were present in introns, whereas exons contained only SNPs (Table 3.2). Exons are under stronger selection pressure resulting in a slower mutation rate caused by the elimination of deleterious mutations from the gene pool (Gaut 1998; Wei et al. 2011).

Plant membrane-bound desaturases are characterized by the presence of three highly conserved HIS-box motifs essential for the enzyme activity (Shanklin et al. 1994; Los and Murata 1998). These motifs are involved in the formation of di-ion active sites (Fox et al. 1993; Shanklin et al. 1994). Consistent with other plant desaturases, the four membrane-bound *fad* genes sequenced herein are predicted to encode FADs with the highly conserved histidine-rich motifs. Only accession SP2047 had a point mutation in one of the HIS-box of FAD3B which was previously shown to be non-functional (Banik et al. 2011). The level of sequence conservation observed between the paralogous desaturases was also reported for other plant desaturases (Scheffler et al. 1997). *Sad1* and *sad2* have the most highly conserved exon structure. Several SADs have been cloned and characterized from various crops such as castor bean, soybean, safflower, *Arabidopsis* and flax (Knutzon et al. 1991; Shanklin and Somerville 1991; Singh et al. 1994; Jain et al. 1999). The high sequence identity at both DNA and amino acid level between *sad* sequences was reported in other plants (Shanklin and Sommerville 1991; Browse and Somerville 1991; Singh et al. 1994; Luo et al. 2009;) and can be interpreted as an indication of the essential role of Δ9-desaturase in lipid biosynthetic pathway in plants.

C18 unsaturated fatty acids of the plastid and the microsomal membranes originate from the desaturation of stearoyl-ACP in the plastid by SAD, thus serving as an attractive

target for altering the unsaturated fatty acid content in oil crops (Ohlrogge and Jaworski 1997). Modification in the synthesis of C18 unsaturated fatty acids may impair membrane fluidity because they are part of structural membranes as well as major components of the seed storage oil (Lightner et al. 1994). The complexity of the multi-gene *sad* family is another indication of its essential role in plants (Ohlrogge and Jaworski 1997; Jain et al. 1999; Fofana et al. 2006). The more conserved nature of *sad2* with a few nucleotide changes is supported by previous studies showing stronger expression of *sad2* in flax (Allaby et al. 2005). The *sad2* locus appeared to be the more physiologically important of the two. The significant effect of SAD2 isoforms on the OLE content corroborates this assumption.

Unsaturated fatty acids in plants play essential roles in membrane integrity and function, cellular signalling, thermal adaptation and energy storage (Browse and Somerville 1991; Mikami and Murata 2003; Zhang et al. 2012). Desaturation of OLE into LIO is considered an important step affecting the quality of seed storage oils, as it initiates the synthesis of PUFAs from monounsaturated OLE. Although a single *fad2* gene was identified in *Arabidopsis*, this gene appears to exist as a gene family in most other plants including flax, soybean, cotton and safflower (Heppard et al. 1996; Fofana et al. 2004; Krasowska et al. 2007; Li et al. 2007; Khadake et al. 2009; Zhang et al. 2009; Cao et al. 2013). Recent studies have demonstrated that genetic variation in *fad2* was associated with consequent changes in fatty acid profiles (Pham et al. 2011; Wang et al. 2011). *Fad2* genes are thought to be rate limiting in fatty acid biosynthesis pathway in flax and are highly influenced by the environment (Fofana et al. 2006). *Fad2a*'s highest allelic diversity of 21 alleles contrasted with the conservation of *fad2b* with only five.

The *fad2b* gene plays a major role in producing LIO and remains a house-keeping microsomal Δ12 oleate desaturase with constitutive expression throughout the plant (Cao et al. 2013; Schlueter et al, 2007). In soybean, *fad2-2*, the orthologue of *fad2b*, was shown to be the most important gene for increasing LIO content (Schlueter et al, 2007). Site-directed mutagenesis of *fad2* altering a few amino acid residues modified the enzymatic activity of the encoded FAD2 in *Lesquerella fendleri* and *Synechocystis* sp. (Avelange-Macherel et al. 1995; Broun et al. 1998). Similarly, through our characterization of the *fad2* genetic variability and its correlation with fatty acid composition, we identified the FAD2A-C isoform to be correlated with a significantly higher level of LIO, suggesting a positive effect of the threonine to methionine substitution located in the vicinity of the third HIS-box.

Two distinct pathways operating in the plastids and the endoplasmic reticulum are responsible for the synthesis of C18:3 fatty acids in plants. Several endoplasmic *fad3* genes have been cloned and characterized from various plants such as *Brassica*, safflower, flax and *Arabidopsis* (Arondel et al. 1992; Yadav et al. 1993; Vrinten et al. 2005). Flax displays wide genetic variability for LIN content with traditional linseed varieties having 50-59% linolenic acid, high-LIN varieties with 60-70% (Friedt et al. 1995; Kenaschuk 2005) and Solin varieties with 2-4%. The first Solin lines were developed using mutation breeding of flax varieties Glenelg and McGregor (Green 1986a; Rowland 1991). As revealed by sequence analysis, the *fad3* genes were hypervariable with numerous SNPs and indels. *Fad3* genes also carried non-sense mutation resulting in premature stop codons, a feature only observed in induced mutants or lines derived from them. Naturally occurring allelic diversity in plants has been

considered an important genetic factor for phenotypic variation (Buckler and Thornsberry 2002). Induced mutations eliminate or cause a large reduction of a functional gene product, whereas naturally occurring allelic variation alters the gene products and may be the fundamental mechanism for quantitative trait variation (Yano and Sasaki 1997; Mackay 2001). Similarly, the portion of the allelic diversity and novel isoforms discovered in induced mutant lines were not present in the natural germplasm. EMS mutant lines (Double Low, UGG146-1, SP2047, E1747, YSED18, M96006 and S95407) carrying stop codons in the FAD3A-D, FAD3A-E and FAD3B-B isoforms and HIS-box mutation in the FAD3B-C isoform accumulated significantly reduced levels of LIN. The inability of the mutated FAD3 to perform the desaturation of LIO to LIN in induced mutant lines is supported by the previous studies showing additive gene effects across the two loci on desaturation of LIO into LIN (Green 1986b) and the impaired biochemical activity of the mutant alleles (Stymne et al. 1992). These studies validate the conclusion that the mutant *fad3* alleles either producing truncated proteins or carrying HIS-box mutations are inactive.

The amino acid substitutions in FAD3B-D and -F isoforms may also have a negative influence on FAD3B activity. FAD3A and FAD3B have been shown to be the major enzymes controlling the LIN content in flax seeds (Vrinten et al. 2005). Vrinten et al. (2005) showed that both *fad3a* and *fad3b* carried point mutations leading to premature stop codons in line 593–708 resulting 2-3% LIN content. Similarly, HIS-box mutation in *fad3b* gene in Solin line SP2047 caused the enzyme inactivity (Banik et al. 2011).

The significant inverse relationship of LIN content with LIO found in this study was also reported in a number of crops including flax, soybean and almonds (Wakjira et

al. 2004; Thomas et al. 2003; Abdallah et al. 1998). This inverse association is in agreement with the fact that the biosynthesis of LIN occurs through the step wise desaturation of OLE via LIO (Ayerza 2009). Thus, LIO accumulates in FAD3 mutant lines (Bocianowski et al. 2012). Since the IOD measures the degree of unsaturation, lines with elevated LIN content also show higher IOD (Cloutier et al. 2011). Several studies have demonstrated the correlation between oil content and levels of saturated fatty acids (Velasco et al. 2007). An increase in PAL by 1% led to a decrease in oil content of 1.4% in rapeseed (Mollers and Schierholt 2002). In soybean, both mutants with, reduced and elevated PAL led to a decrease in seed oil content in comparison with lines with standard fatty acid composition (Ndzana et al. 1994; Hartmann et al. 1996; Stoltzfus et al. 2000). However, the correlation between oil content and the levels of unsaturated fatty acids has not been fully elucidated. While a few studies found no adverse effect of high oleic acid soybeans with 80% OLE content on yield and oil content (Kinney 1996; Graef et al. 2009), Brace et al. (2011) showed a statistically significant reduction in both oil content and yield in high oleic acid soybeans. The significant differences observed in OIL content with respect to FAD2A/B and FAD3A/B isoform combinations are not consistent with elevated or reduced LIO or LIN content, suggesting further studies to elucidate correlations. However, seed oil content is a complex quantitative trait governed by a number of genes and also influenced by the environment (Burton 1987; Cloutier et al. 2011; Lee et al. 2007, Eskandari et al. 2013). Therefore, it is likely that these correlations may not only be determined by genetic factors, but also influenced by the environment.

Genetic redundancy drives evolution by allowing functional diversification while simultaneously retaining the original function(s) of essential genes (Cao et al. 2013). In

flax, the SAD, FAD2 and FAD3 enzymes are encoded by duplicated genes (Fofana et al. 2010). Following duplication, paralogs can retain their original gene function, gain new function(s), or be silenced (Force et al. 1999). In flax, functional redundancy of six paralog desaturases provides additional buffering capacity for mutation tolerance even in exons as exemplified by the predicted non-functional FAD3A-D, FAD3A-E and FAD3B-B. A duplicated pair of genes can have an altered selective pressure, leading to the loss of one copy or to an increased rate of divergence in sequence when both copies are preserved (Fischer et al. 2001). Here, *fad3a* and *fad3b* seem to be functionally preserved and the redundancy may have allowed for the higher divergence between sequences, consistent with the previous studies showing the additive role of *fad3a* and *fad3b* and their equal contribution to LIN content in flax (Vrinten et al. 2005; Banik et al. 2011).

Selection pressure plays a prominent role in decreasing nucleotide diversity in domesticated crops (Wei et al. 2011). In flax, the overall reduction of nucleotide diversity during domestication is at a moderate level (27% with respect to pale flax) when compared to other inbred species such as wheat and barley (Fu 2011). Selection pressure over the process of domestication might have a significant impact on the observed variation in *fad2, sad1* and *sad2* as illustrated by the NJ trees (Appendix VIII, IX, X). The impact of selection on *fad2* and *fad3* diversity during domestication was also reported in cultivated sunflower (Chapman and Burke 2012).

FA composition can be altered by manipulating one or more steps of their biosynthesis pathway (Ohlrogge and Jaworski 1997, Thelen and Ohlrogge 2002, Cahoon et al. 2010). Most domesticated oilseed crops have been modified to obtain optimized FA profiles providing specific end-uses through approaches such as classical breeding or

genetic engineering (Drexler et al. 2003). Suppression of the Δ12-desaturase gene in soybean, sunflower, cotton and canola has successfully increased OLE content in their seed oils (Metzger and Bornscheuer 2006). Novel allelic variants and isoforms identified for the six desaturases provide useful genetic and molecular resources and information for the development of oilseed flax with unique and useful oil profiles that would not require a transgenic or mutagenesis approach.

CONNECTING TEXT BETWEEN CHAPTER 3 AND 4

In the previous study, the genetic variability for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* genes in flax was characterized by sequencing the six genes from 120 flax accessions. Between five and 21 alleles corresponding to two to seven isoforms were identified for the six desaturases. Thirty-four accessions had an identical isoform composition for all six desaturase genes but their FA composition varied significantly. These results lead to the formulation of the hypothesis "FA composition differences in these lines could result from differential expression of the desaturase genes during seed development" for the next study. Based on this hypothesis, three goals were established. The first goal was to quantify the expression levels of the desaturase genes at different stages of seed development by semi-quantitative reverse transcriptase (RT)-PCR in relatively low, intermediate and high LIN genotypes expressing identical isoforms for all six desaturases. The second goal aimed to study the structural differences in the promoter region of the six desaturase genes. The third goal was to correlate these structural and expression data with FA composition as determined by phenotyping the field grown genotypes during four years at two locations with the overall objective of gaining a greater understanding of the genetic factors controlling the FA composition in flax.

FATTY ACID COMPOSITION AND DESATURASE GENE EXPRESSION IN FLAX (*Linum usitatissimum* L.)

Dinushika Thambugala[1,2], Sylvie Cloutier[1,2,3*]

[1]*Department of Plant Science, University of Manitoba, 66 Dafoe Rd, Winnipeg, MB, Canada, R3T 2N2*

[2]*Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Rd, Winnipeg, MB, Canada, R3T 2M9*

[3]*Current address: Eastern Cereal and Oilseed Research Centre, K.W. Neatby Building, 960 Carling Ave, Ottawa, ON, Canada, K1A 0C6*

Author Dinushika Thambugala conducted this work as part of her PhD thesis. Dinushika Thambugala carried out the experiment, analysed and interpreted the data and co-wrote the manuscript. The supervisor Dr. Sylvie Cloutier designed the experiments, participated in the interpretation of data and co-wrote the manuscript. The authors read and approved the final manuscript.

The manuscript was published in Journal of Applied Genetics 2014, 55:423-432

4.0 FATTY ACID COMPOSITION AND DESATURASE GENE EXPRESSION IN FLAX (*Linum usitatissimum* L.)

4.1 Abstract

Little is known about the relationship between expression levels of fatty acid desaturase genes during seed development and fatty acid (FA) composition in flax. In the present study, we looked at promoter structural variations of six FA desaturase genes and their relative expression throughout seed development. Computational analysis of the nucleotide sequences of the *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* promoters showed several basic transcriptional elements including CAAT and TATA boxes, and several putative target-binding sites for transcription factors, which have been reported to be involved in the regulation of lipid metabolism. Using semi-quantitative reverse transcriptase PCR, the expression patterns throughout seed development of the six FA desaturase genes were measured in six flax genotypes that differed for FA composition but that carried the same desaturase isoforms. FA composition data were determined by phenotyping the field grown genotypes over four years in two environments. All six genes displayed a bell-shaped pattern of expression peaking at 20 or 24 days after anthesis. *Sad2* was the most highly expressed of all six genes. The expression of all six desaturase genes did not differ significantly between genotypes (P=0.1400), hence there were no correlations between FA desaturase gene expression and variations in FA composition in relatively low, intermediate and high linolenic acid genotypes expressing identical isoforms for all six desaturases. These results provide further clues towards understanding the genetic factors responsible for FA composition in flax.

4.2 Introduction

Flax (*Linum usitatissimum* L.) is the leading source of plant-based omega-3 fatty acids (FAs) praised for their health benefits in humans and animals. Oilseed flax, also known as linseed, generally contains 40-50% oil and its quality is largely determined by its FA composition (Green 1986; Cloutier et al. 2011). Linseed oil is primarily composed of palmitic (PAL, C16:0; ~6%), stearic (STE, C18:0 ~4.4%), oleic (OLE, C18:1 ~24.2%), linoleic (LIO, C18:2 ~15.3%) and linolenic (LIN, C18:3 ~50.1%) acids (Muir and Westcott 2003). The high levels of alpha-linolenic acid (ALA or LIN) and moderate levels of LIO in linseed oil not only contribute to a healthy diet but are considered essential FAs because humans lack the Δ^{12} and Δ^{15} desaturase enzymes that convert OLE to LIO and LIO to LIN, respectively (Damude and Kinney 2008). Humans can use these FAs as substrates for further elongation and desaturation leading to the formation of very long chain polyunsaturated FAs (VLCPUFAs) like ecosapentaenoic acid (EPA, C22:5), docosahexaenoic acid (DHA, C22:6) and arachidonic acid (AA, C20:4) (Warude et al. 2006). These VLCPUFAs also have health benefits and studies have established their important role in reducing total and low-density lipoprotein (LDL) cholesterol levels in humans and preventing chronic diseases including cardiovascular diseases, hormonal cancers and arthritis (Oomah 2001; Wiesenfeld et al. 2003; Ander et al. 2004; Dyer et al. 2008). However, the high LIN content of flaxseed oil makes it more susceptible to oxidation and rancidity (Zuk et al. 2012), thus limiting its use as an edible oil, but simultaneously providing it with unique drying properties that makes it valuable in various industrial applications (Green 1986).

Genetic control of FA biosynthesis in flax has been studied and many of the genes encoding the enzymes that perform FA synthesis have been identified and characterized (Green 1986; Fofana et al. 2004; Sorensen et al. 2005; Vrinten et al. 2005; Fofana et al. 2006; Krasowska et al. 2007; Khadake et al. 2009; Banik et al. 2011; Thambugala et al. 2013). FA desaturation and elongation are important biochemical processes that drive the multi-step FA biosynthetic pathway in a sequential manner, leading to synthesis of polyunsaturated FAs (Warude et al. 2006; Khadake et al. 2011). Fatty acid desaturases (FADs) are the key enzymes that introduce double bonds into FA acyl chains in a stepwise manner starting from STE (Los and Murata 1998; Shanklin and Cahoon 1998; Smooker et al. 2011). The desaturation of STE is sequentially catalyzed by desaturases namely, stearoyl-ACP desaturase (SAD) (Singh et al. 1994; Jain et al. 1999), fatty acid desaturase 2 (FAD2) (Krasowska et al. 2007; Khadake et al. 2009) and fatty acid desaturase 3 (FAD3) (Vrinten et al. 2005; Banik et al. 2011). In flax, these three enzymes are encoded by duplicated genes (Fofana et al. 2010). The two FAD3 enzymes, FAD3A and FAD3B, have been shown to be the major enzymes controlling LIN content in linseed (Vrinten et al. 2005).

Although the genetic variability of desaturase genes and their impact on FA composition in flax have been studied (Thambugala et al. 2013), only two studies on the regulation and expression of *sad* and *fad* genes during seed development have been reported (Fofana et al. 2006, Banik et al. 2011). Fofana et al. (2006) reported that the expression of *sad* and *fad2* genes in flax was modulated during seed development whereas Banik et al. (2011) found that the expression patterns of *fad3a* and *fad3b* were highly correlated with LIN accumulation during seed development.

In our previous study, we characterized the genetic variability for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* genes in flax by sequencing the six genes from 120 flax accessions (Thambugala et al. 2013). Between five and 21 alleles corresponding to two to seven isoforms were identified for the six desaturases. Thirty-four accessions had an identical isoform composition for all six desaturase genes but their FA composition varied significantly. We hypothesized that FA composition differences in these lines could result from differential expression of the desaturase genes during seed development. Based on this hypothesis, this study had three goals. First, to quantify the expression levels of the desaturase genes at different stages of seed development by semi-quantitative reverse transcriptase (RT)-PCR in relatively low, intermediate and high LIN genotypes expressing identical isoforms for all six desaturases. Second, to study the structural differences in the promoter region of the six desaturase genes. Third, to correlate these structural and expression data with FA composition as determined by phenotyping the field grown genotypes during four years at two locations with the overall objective of gaining a greater understanding of the genetic factors controlling the FA composition in flax.

4.3 Materials and Methods

4.3.1 Plant material

FA composition of 34 flax accessions carrying identical isoforms for the desaturase genes *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* were analysed as previously described (Thambugala et al. 2013) and six linseed genotypes showing significantly different ($P<0.05$) FA profiles were selected for this study (Appendix XVII). These six genotypes,

including two high (UGG5-5, M5791), two intermediate (FP2270, CN30861) and two relatively low LIN (CN97334, CN97407), were grown in a growth chamber under the following conditions: 22°C with a 16-h photoperiod at a photon density of approximately 145 $\mu E \cdot m^{-2} \cdot s^{-2}$ (Fofana *et al*. 2004). Flowers were tagged at anthesis and developing bolls harvested at 8, 12, 16, 20, 24, 28 and 32 days after anthesis (DAA) were immediately frozen in liquid nitrogen where they were stored until RNA extraction.

4.3.2 RNA extraction

Total RNA was extracted from 8-32 DAA developing bolls of each genotype using the RNA extraction procedure described in Banik et al. (2011). For each extraction, 0.2 g of bolls was used. Final total RNA pellets were resuspended in 50 µl RNase-free water and stored at –80°C. The RNA was quantified by nano-spectrophotometer (Implen GmbH, Munich, Germany).

4.3.3 First strand cDNA synthesis

To remove any potential remnant DNA, total RNA from each developmental stage and genotype was treated with TURBO DNase according to the manufacturer's instructions (Ambion, Austin, Texas, USA). The DNase treated RNA was used as template to synthesize first strand cDNA using oligo(dT) primer and Superscript^{TM} II reverse transcriptase followed by RNaseH treatment as per the manufacturer's recommendations (Invitrogen, Carlsbad, CA, USA). An amount of 800 ng total RNA was used to synthesize the cDNA in three independent 20 µl reactions for each developmental stage and each genotype. Pooled cDNA samples of each developmental stage and genotype

were stored at –20°C.

4.3.4 cDNA quantification

A fluorometric method was used to precisely quantify the first strand cDNA. In this method, the RNA was digested and the single-stranded cDNA was quantified by fluorescence using RiboGreen (Invitrogen) as described by Libus and Storchova (2006). cDNA quantification was performed in duplicate for each development stage of each genotype.

4.3.5 RT-PCR of *sad* and *fad* genes

Semi-quantitative RT-PCR was performed using the quantified cDNA samples from 8-32 DAA using 28 cycles as previously recommended (Kumar et al. 2013). Gene-specific PCR primers (Table 4.1) for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* were designed using Primer Express (Applied Biosystems) and Primer 3 software (Rozen and Skaletsky 2000). Optimized amplification reactions (25 µL) contained 4 ng cDNA, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.8 mM dNTPs, 0.4 µM each primer and 1.5 U Taq DNA polymerase. The PCR reactions were first denatured at 94°C for 5 min followed by 28 cycles consisting of 94°C for 30 s; 62°C for 30 s and 72°C for 60 s, prior to a final extension at 72°C for 10 min. RT-PCR products were resolved on 2.5% agarose gels stained with ethidium bromide. The flax adenine phosphoribosyl-transferase 1 (*apt1*) gene was used as a reference control in all RT-PCR experiments (Banik et al. 2011). Three independent RT-PCR replications were performed for each gene, each developmental stage and genotype, including *apt1*. The expression of the target genes relative to the reference gene

*apt1*was evaluated by densitometric analysis of the signal strength of the semi-quantitative RT-PCRs with the AlphaImagerHP software (version 3.4, proteinsimple, Santa Clara, CA, USA). Desaturase gene to *apt1* ratios were calculated by dividing the background corrected signal of the PCR amplicon of the desaturase gene to that of *apt1*.

4.3.6 Promoter analysis of *sad* and *fad* genes

Promoter sequences corresponding to bases -800 to +200 relative to the transcription start site (TSS) of *sad* and *fad* genes were obtained from the six flax genotypes by sequencing the amplicons with Big-Dye V3.1 Terminator chemistry and resolving them on an ABI 3130xl Genetic Analyser (Applied BioSystems, Foster City, CA, USA). All sequences were processed using PHRED (Ewing et al. 1998) and assembled with CAP3 (Huang and Madan 1999) as implemented in the internal data pipeline called SOOMOS v0.6 (T. Banks, personal communication). Sequence alignments were performed with Clustal W (Thompson et al. 1994). TSSs were identified using the bioinformatics pipeline for TSS signals (http://fruitfly.org/seq_tools/promoter.html) (Reese 2001). Promoter analysis was performed with PLAnt Cis-acting regulatory DNA Elements (PLACE) (Higo et al. 1999) and PLANT Promoter Analysis Navigator (PlantPAN) (http://plantpan.mbc.nctu.edu.tw/seq_analysis.php) (Chang et al. 2008).

4.3.7 Fatty acid composition

Field plots of the 34 flax accessions were grown in a type 2 modified augmented design (MAD) (Lin and Pouschinsky, 1985) at the Kernen farm near Saskatoon (SK, Canada) and at the Morden Research Station (MB, Canada) in 2009, 2010, 2011 and 2012

(Appendix XVII). FA composition and oil content (OIL) were obtained from all eight field experiments as previously described (Thambugala et al. 2013).

Table 4.1 Sequences and melting temperature of primers used for semi-quantitative RT-PCR and promoter analysis (amplification and sequencing)

Gene	Primer name	Sequence (5' to 3')	Tm (°C)
apt1	APT1-319F	TAGAGCTGACCAGGACAAACA	62
	APT1-409R	GTTTATGAATGCGCTTGTCTCA	62
sad1	SAD1-F770	TCGCAGCAGACGAGAAACG	60
	SAD1-R840	AGGGTCGATCTCGAAGAGCTT	58
sad2	SAD2-F202	AAGCTGGAGATCTTTAAGTCCCTTGA	59
	SAD2-R307	GTTCGGGCAGGAAATCTTGT	58
fad2a	FAD2A-F873	CGTGGATCGAGACTACGGGTTA	60
	FAD2A-R942	ATGGTGCGCGACATGTGT	58
fad2b	FAD2B-F496	TGGCACTCAAAGTACCTCAACAA	58
	FAD2B-R571	AAGGCCAGCCGAGAGTGA	58
fad3a	FAD3A-F40	GACTTCAAAACTGTGGCTCT	55
	FAD3A-R132	GATAGCCACACCATTGGTGC	62
fad3b	FAD3B-F662	GCAGCGGTCTTGATTTCAACA	48
	FAD3B-R759	ATTTTGAGGACCGGAGCGAA	50
Promoter analysis			
sad1	SAD1-F53350	AATGCCTCCAAAGTGCTCTC	59
	SAD1-R54516	GCTTACTTGGTGGAGGTGGA	60
	SAD1-F53628	TTTGGTGACTCGAAAGTTCT	55
	SAD1-R53974	CATATGACATTGCAAGACGA	56
sad2	SAD2-F10585	CGTCCCAATTGATGACAATG	60
	SAD2-R11671	TGGAATTGAAAGTGGAAGCA	59
	SAD2-F10975	CCAAAGTGCTCTCTACTTGC	55
	SAD2-R11296	GCGTTTCATCAGTTCTATCG	56
fad2a	FAD2A-F22	CGGCGATTTTGAAGTGCAT	62
	FAD2A-R868	CTCACCGAGCGTGAATGGT	62
	FAD2A-F252	GCCCTCCTTCATATTCTTCT	55
	FAD2A-R566	TCCTTTCCAGTTTTCAGTTG	55
fad2b	FAD2B-F1036	AAGGGTGATGGTCTTGATGC	60
	FAD2B-R2094	AGGGACGGCTTTCTTGATCT	60
	FAD2B-F1239	TACCCTAAAGTGATCAATGG	53
	FAD2B-R1569	AGCAAGTAGTGCTATCCTGA	53
fad3a	FAD3A-F949	TCGATTGCAAAGCAAGAGAG	59
	FAD3A-R2018	AACGGCGAAGCTGAGGAT	61
	FAD3A-F1228	TTAGTCGATTTCACCCTAGC	55
	FAD3A-R1503	AACGGTTGTTGTTACTTGCT	55
fad3b	FAD3B-F15266	CCCAACCCATTACATGACG	60
	FAD3B-R16252	GCTGAGGATGACAAGGAGGT	59
	FAD3B-F15435	AACATTGCAATTCAGAGTCC	55
	FAD3B-R15777	TCTGCTCTTTATTGGGTTTC	55

4.3.8 Statistical analysis

The phenotypic data for FA composition and OIL (Appendix XVII) were adjusted for soil heterogeneity based on the MAD statistical analysis method using the recently described pipeline (You et al. 2013). Variance components were calculated using adjusted phenotypic data to determine the effect of year, location, genotype and their interactions on FA composition and OIL using the PROC GLM procedure (SAS Institute, Cary, NC, USA). To assess the differences in FA composition among the six genotypes, one-way analysis of variance (ANOVA) was used followed by the Duncan's multiple range comparison test at 0.05 probability level. A similar analysis was performed to determine statistical significance between the ratios of FA desaturase gene to *apt1* of the six desaturase genes, across all six genotypes and for all seven stages of seed development. All statistical analyses were carried out using SAS v9.2 (SAS Institute, Cary, NC, USA).

4.4 Results

4.4.1 Fatty acid composition

The FA composition of the 34 flax accessions with identical isoform composition for all the six desaturase genes displayed significant variations ($P < 0.0001$) in LIN content ranging from 46 to 72% (Fig. 4.1). Six of those accessions were selected on the basis of their significantly different ($P<0.05$) FA compositions. All five FAs showed significant difference across the six genotypes (Fig. 4.2). UGG5-5 and M5791 had high LIN but low LIO content. In contrast, CN97334 and CN97407 had lower LIN but higher LIO than UGG5-5 and M5791. Significant FA compositional variations were also found for PAL, STE and OLE (Fig. 4.2).

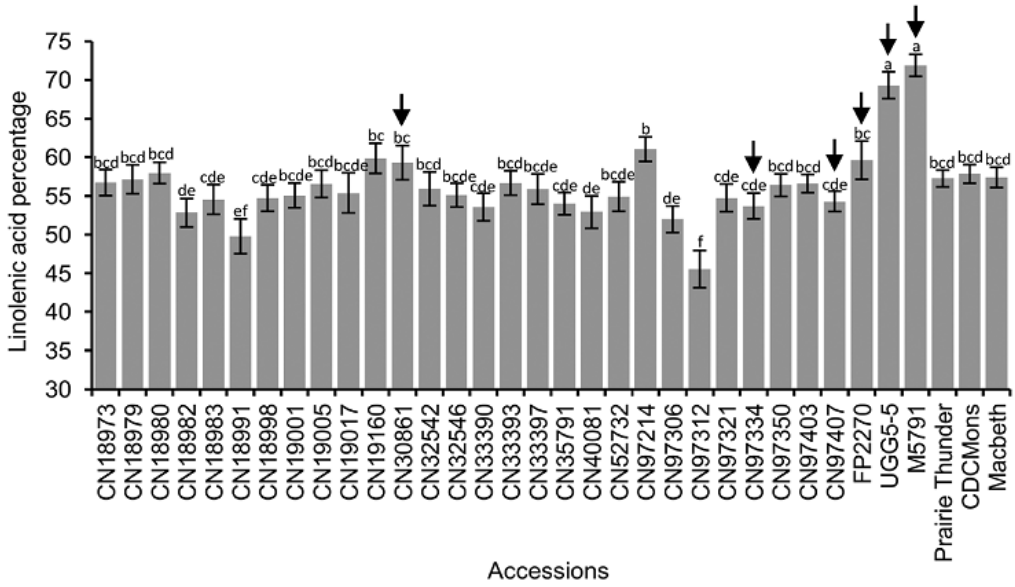


Fig 4.1 Linolenic acid content of 34 flax accessions carrying identical isoforms for the fatty acid desaturases SAD1, SAD2, FAD2A, FAD2B, FAD3A and FAD3B. The sample means were averaged from two locations (MB and SK) over four years (2009, 2010, 2011 and 2012). Arrows indicate the accessions selected for the fatty acid desaturase gene expression study. Error bars represent the standard error of the mean. Letters above the bars indicate statistical significance of the Duncan's multiple range tests.

4.4.2 *Sad* and *fad* gene expression during seed development

A semi-quantitative RT-PCR method was used for quantification of RNA transcripts and the detection of any variation in the expression of the six FA desaturases during the seed developmental stages of flax from 8-32 DAA in six flax genotypes (Appendix XVIII). With the exception of *sad1,* expression was significantly modulated ($P < 0.0001$) for the other desaturases during seed development (Appendix XIX). All six genes followed a similar pattern where gene expression tended to increase from eight to 20 DAA, peaked at 20 or 24 DAA and decreased during the later stages of seed maturation (Fig. 4.3). *Sad2* was the most highly expressed gene (Appendix XVIII) throughout all stages of seed development, peaking at 24DAA (Fig. 4.3). *Sad1* expression was lower than *sad2* and generally remained more constant throughout all developmental stages (Fig. 4.3).

Fad2 and *fad3* displayed a similar pattern where gene expression increased from eight to 24DAA and 20DAA respectively and decreased towards maturity. Over all seed developmental stages, *sad* and *fad* expression were not significant between genotypes (P=0.1400) (Appendix XVIII, Appendix XIX). However, at 32 DAA, significant differential expression between genotypes was observed for *fad2a*, *fad3a* and *fad3b* (Appendix XX). *Fad2a* and *fad3a* were more highly expressed in the high-LIN line M5791 whereas *fad3b* was expressed at a lower level in FP2270 than in the other five genotypes (Table 4.2).

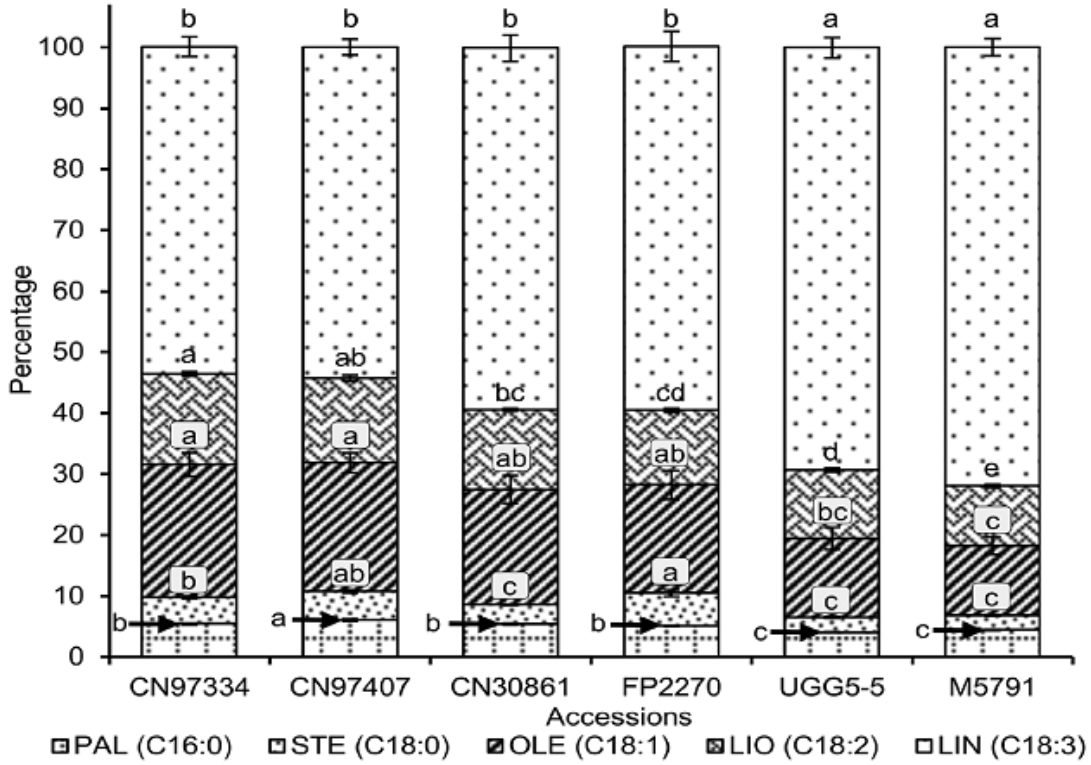


Fig 4.2 Fatty acid composition of CN97334, CN97407, CN30861, FP2270, UGG5-5 and M5791. Percentages of the five main fatty acids, namely palmitic (PAL, C16:0), stearic (STE, C18:0), oleic (OLE, C18:1), linoleic (LIO, C18:2) and linolenic (LIN, C18:3) acids for each genotype are illustrated. Error bars represent standard error of the mean. Letters beside or above the bars indicate statistical significance of the Duncan's multiple range tests.

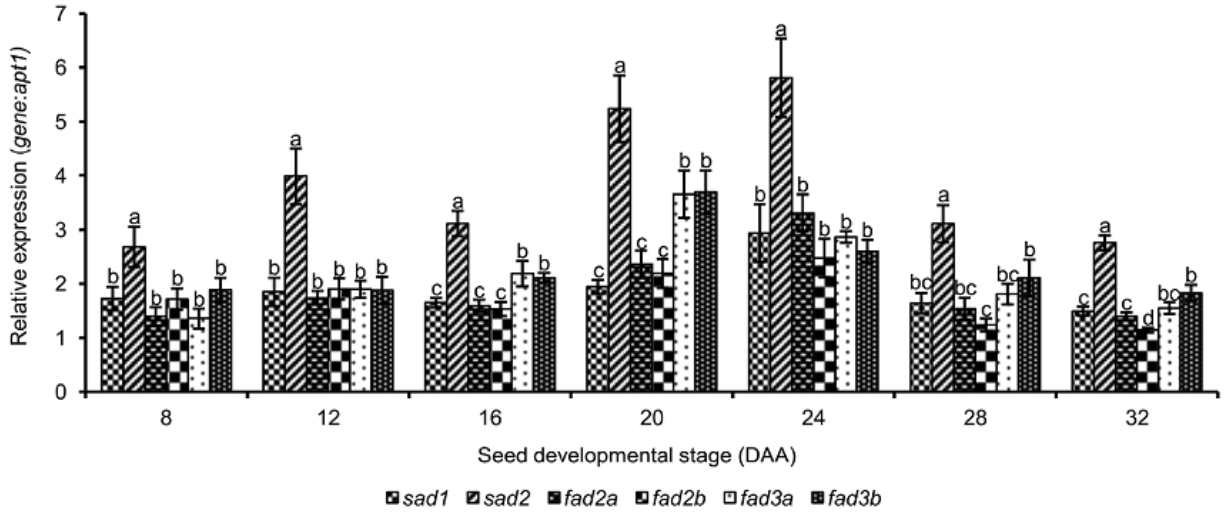


Fig 4.3 Relative expression of the fatty acid desaturase genes *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* in flax during seed development. Error bars represent standard error of the mean and are based on three independent RT-PCR replicates. Letters above the bars indicate statistical significance of the Duncan's multiple range tests.

4.4.3 Promoter analysis

Promoter sequence analysis of *sad* and *fad* genes corresponding to bases from -800 to +200 relative to the TSS revealed single point mutations in the promoter region of *sad1*, *sad2* and *fad3b* of CN30861 (Appendix XXI). Computational analysis of promoter regions of the six desaturase genes indicated the presence of several basic transcriptional elements such as CAAT and TATA boxes. In addition to these basic elements, many seed or endosperm specific and ABA-responsive cis-elements (ABRE) including several Dof core (AAAG motif), DPBP core (ACACNNG motif), E-box (CANNTG motif), Myb-core (CNGTTR motif) and ACGT-box (AACGTT/ABRE motif) were identified. The sequence analysis also revealed the presence of motifs similar to pollen-specific cis-acting elements; POLLEN1LELAT52 (AGAAA) is one of two co-dependent regulatory elements responsible for the pollen-specific activation of genes (Appendix XXI).

Table 4.2 Expression of desaturase genes *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* during seed development of six flax genotypes

Seed developmental stage (DAA)	Genotype	Relative expression (*gene*:*apt1*)					
		sad1	*sad2*	*fad2a*	*fad2b*	*fad3a*	*fad3b*
8	CN97334	1.0	1.9	0.8	1.4	0.9	1.3
	CN97407	2.1	3.8	1.5	2.0	1.4	2.0
	CN30861	1.6	3.3	1.4	2.1	1.8	2.3
	FP2270	2.0	2.3	1.3	1.9	0.8	1.9
	UGG5-5	1.7	2.2	1.2	1.3	1.2	1.4
	M5791	1.9	2.6	2.6	1.7	2.1	2.4
12	CN97334	1.4	3.1	1.4	1.6	1.4	1.6
	CN97407	2.0	4.3	2.0	1.9	2.2	2.3
	CN30861	1.7	4.2	1.8	1.9	2.2	2.1
	FP2270	1.7	3.7	1.5	1.7	1.7	1.9
	UGG5-5	2.2	4.6	2.1	2.4	2.1	1.9
	M5791	2.1	4.1	1.7	1.9	1.8	1.6
16	CN97334	1.6	2.3	1.7	1.5	1.5	2.0
	CN97407	1.4	2.6	1.5	1.4	2.2	2.0
	CN30861	2.1	4.1	1.9	1.8	3.2	2.5
	FP2270	1.6	3.3	1.7	1.6	2.4	2.0
	UGG5-5	1.5	3.2	1.4	1.3	2.1	2.1
	M5791	1.8	3.2	1.3	1.6	1.7	2.1
20	CN97334	1.9	5.1	2.2	2.4	3.0	3.2
	CN97407	1.9	6.2	2.0	2.2	3.7	4.1
	CN30861	2.2	6.0	3.0	2.7	4.5	4.4
	FP2270	1.8	5.2	2.1	2.1	3.7	3.7
	UGG5-5	1.7	4.6	2.4	1.8	3.5	3.5
	M5791	2.2	4.3	2.4	2.0	3.6	3.3
24	CN97334	3.1	5.9	3.8	2.8	2.8	2.4
	CN97407	2.7	5.5	3.0	2.0	2.6	2.3
	CN30861	3.0	6.5	3.0	2.6	3.0	2.3
	FP2270	2.8	6.1	3.2	2.4	2.8	2.7
	UGG5-5	2.7	5.0	3.4	2.3	3.1	3.1
	M5791	3.3	6.0	3.4	2.8	2.9	2.8
28	CN97334	1.6	3.8	1.4	1.3	1.6	3.0
	CN97407	1.4	3.3	1.5	1.1	1.5	2.1
	CN30861	1.7	4.0	1.6	1.5	2.2	2.4
	FP2270	1.9	2.9	1.4	1.4	2.0	2.0
	UGG5-5	1.7	3.2	1.7	1.3	1.9	1.8
	M5791	1.5	1.6	1.7	0.9	1.7	1.3
32	CN97334	1.4	2.9	1.3^{b}	1.2	1.5^{b}	2.0^{a}
	CN97407	1.3	2.8	1.3^{b}	1.1	1.6^{ab}	2.2^{a}
	CN30861	1.6	2.5	1.3^{b}	1.2	1.6^{ab}	2.3^{a}
	FP2270	1.6	3.1	1.3^{b}	1.2	0.8^{c}	0.7^{b}
	UGG5-5	1.4	2.6	1.3^{b}	0.9	1.8^{ab}	1.6^{a}
	M5791	1.7	2.7	1.9^{a}	1.3	2.1^{a}	2.0^{a}

[a,b,c] Statistical significance ($p < 0.05$) of Duncan's multiple range tests among genotypes within each gene and developmental stage.

4.4.4 Phenotypic data

All FA traits showed significant genotype (G), location (L) and year (Y) effects ($P < 0.001$; Appendix XXII). Genotype by environment interactions (GEs: G*L, G*Y, L*Y and G*L*Y) were also significant for all FA composition traits and oil content (Appendix XXII).

4.5 Discussion

FADs display significant diversity in their sequences and expression (Los and Murata 1998; Warude et al. 2006) and hence are considered biotechnological targets for manipulation of FA composition of oilseed crops (Khadake et al. 2009). Although the genetic variability for FADs and their impact on FA composition in flax has been recently reported (Thambugala et al. 2013), little is known about how the *fad* expression levels during seed development affect FA composition. In the present study, expression patterns of *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* of six flax genotypes at various seed developmental stages were studied using semi-quantitative RT-PCR analysis.

Semi-quantitative RT-PCR is a highly sensitive and specific method to analyse the expression of genes (Choquer et al. 2003). The reliability of this method depends on a number factors including RNA quality, primer specificity, technique precision and use of a stable house-keeping gene (Wong and Medrano 2005; Banik et al. 2011). Although semi-quantitative RT-PCR is mostly used as a qualitative method of analysing gene expression, the determination of both *fad* and *apt1* reference gene products by densitometric analysis enabled quantification and permitted comparisons across developmental stages, genes and genotypes (Choquer et al. 2003; Libus and Storchova

2006). The *Arabidopsis apt1* gene has been identified as one of the most stable internal controls (Gutierrez et al. 2008) and the flax *apt1* orthologue used in this study confirmed its consistent expression across all seed developmental stages and genotypes (Livak and Schmittgen 2001; Pfaffl 2005; Banik et al. 2011). Furthermore, the approach of quantifying cDNA precisely with the RiboGreen method reduces the variability associated with variations in starting material, hence adding precision to the evaluation method (Libus and Storchova 2006).

To establish correlations between gene expression and variation in FA composition, we examined expression patterns of *sad* and *fad* genes using six flax genotypes varying in LIN content. To our knowledge, this is the first report in flax comparing expression levels of all six desaturase genes and FA composition during seed development from genotypes that differed significantly in LIN content. Oil accumulation is a highly controlled developmental process. Genetic studies indicated that genes of the FA biosynthetic pathways, including triacylglycerol (TAG) synthesis, are regulated at the level of transcription (Baud and Graham 2006; O'Hara et al. 2002; Ohlrogge and Jaworski 1997). Gene expression programs related to FA synthesis are activated during the maturation phase and most genes encoding FA synthesis enzymes display a bell-shaped pattern of expression during seed development (Baud and Lepiniec 2009). Similarly, the six FA desaturases studied herein all displayed the bell-shaped pattern of expression with a peak at or after 20 DAA. Although the flax genome contains two paralogous *sad* loci, *sad1* and *sad2*, they are differentially expressed (Jain et al. 1999) with *sad1* having lower and more constant expression throughout seed development. The highly conserved nature of *sad2*

and its higher expression are in agreement with its essential Δ9-desaturase role in the lipid biosynthetic pathway in flax (Allaby et al. 2005; Thambugala et al. 2013).

Transcriptional control of *fad* gene expression in flax has been demonstrated (Fofana et al. 2006; Banik et al. 2011). The two paralogous *fad2a* and *fad2b* genes have been cloned and characterized from flax (Krasowska et al. 2007; Khadake et al. 2009). Both genes displayed relatively low but steady expression patterns throughout seed development. Fofana et al. (2006) demonstrated the seed-specific expression of *fad2a* while constitutive expression of *fad2b* has also been reported (Cao et al. 2013; Schlueter et al. 2007). *Fad3a* and *fad3b* had similar expression patterns as the other four desaturase genes except that their expression peaked at 20 DAA instead of 24 as previously shown for other flax accessions by Banik et al. (2011) who also demonstrated that the *fad3* expression correlated with LIN accumulation during seed development.

Fad2 genes are thought to be the rate-limiting genes of the FA biosynthesis pathway in flax and are highly influenced by the environment (Fofana et al. 2006). The significant GE interaction observed for FA traits also suggests the complex interactions of gene and environmental cues on FA composition. Temperature during the growing season affects FA composition of flax and other oil crops (Casa et al. 1999; Fofana et al. 2006; Baud and Lepiniec 2010). The higher thermal stability of safflower's FAD2 compared to that of sunflowers was proposed to explain the more stable FA composition of safflower irrespective of the temperature during seed development (Esteban et al. 2004). Effects of varying temperature on *sad* gene expression alter the FA composition of soybean seeds (Byfield and Upchurch 2007).

Identification of molecular components that regulate expression patterns of FA biosynthesis genes is important for understanding the variation in FA composition in flax. Activators and repressors fine-tune the expression level of FA biosynthetic pathway genes (Bene et al. 2001; Baud and Lepiniec 2010; Saed Taha et al. 2012). An AW-box sequence [CnTnG](n)$_7$[CG] identified in the promoter regions of several FA biosynthetic genes was proposed to be recognized by WRI1, a transcription factor responsible for activating these genes (Maeo et al. 2009; Baud et al. 2007). Although promoters of the six desaturase genes share a functionally similar promoter core, their expression can be modulated by differences in upstream regulatory elements. Kim et al. (2006) reported that the strong seed specific expression of the sesame *fad2* gene is controlled by negative cis-regulatory elements of the promoter and enhancers located in the 5'-UTR (untranslated region). Temporal control of *fad2* via ABA-response elements was also reported (Kim et al. 2006). However, the functions of these putative regulatory elements for the expression of FAD genes in flax are yet to be determined.

Millar and Kunst (1999) reported that the natural genetic variation of the FA composition of seed oils in different ecotypes of *Arabidopsis thaliana* are probably due to altered expression levels or activities of FA biosynthetic enzymes. Here, we looked at the expression levels of six desaturase genes in six genotypes that differed for FA composition but that encoded the same desaturase isoforms and we were unable to establish correlations between the expression of any of the desaturase genes with the FA composition of the six flax genotypes. Two main hypotheses provide potential explanations for our results. First, genetic factors other than the six desaturases studied here may play an important role in determining the FA composition of flax. Although

desaturases and thioesterases are the major enzymes responsible for FA composition in oilseed crops (Ohlrogge and Jaworski 1997; Baud and Lepiniec 2010), minor genes may play an important role in determining the FA composition variation. Complete genome sequences of many plant species have indeed allowed the identification of a number of genes involved in plant oil biosynthesis (Ying et al. 2012). However, the factors leading to variations in FA composition remain to be fully understood (Hobbs et al. 2004; Keurentjes et al. 2006). Genes of minor effect have generally been considered responsible for variation observed in LIN content in several oil crops including flax (Das and Rai 1974; Doucet and Filipescu 1981; Green 1986; Cloutier et al. 2011), rapeseed (Kondra and Thomas 1975; Pleines and Friedt 1989) and soybean (White et al. 1961). Green (1986) reported that the small differences in LIN content between 'Glenelg' and the majority of current flax varieties are most probably due to the cumulative effects of several minor genes that modify the expression of *fad3a* and *fad3b* (Vrinten et al. 2005). QTL and association mapping may help in identifying minor genes controlling FA composition (Soto-Cerda et al. 2013). Of these putative minor genes, Lei et al. (2012) suggested that acyl carrier protein (ACP), 3-ketoacyl-ACP-synthase (KAS) and acyl-ACP thioesterase (FATA) play a role in FA composition and may also be rate limiting. ACP, KAS and FATA gene expression correlated significantly with monounsaturated FA and PUFA synthesis (Lei et al. 2012). Similarly, natural variation in long chain FA content in *Arabidopsis thaliana* was found to be controlled by a new isoform of β-ketoacyl-CoA synthase 18 (KCS18; Jasinski et al. 2012).

The second hypothesis is that differential expression of desaturases can indeed explain FA compositional differences in lines with identical isoforms grown in the field

but that we could not demonstrate it because we measured the expression under controlled conditions. Environment and genotype by environment significantly affect FA composition (Deng and Scarth 1998; Hernández et al. 2011). The effect of the environment on *fad2* expression in flax has been documented (Fofana et al. 2006). It is therefore conceivable that correlations may exist between desaturase expression levels and FA composition if the former was measured in the same field environment as the latter. This second hypothesis is not mutually exclusive to the minor genes hypothesis because minor genes could be responsible for translating the environmental cues that affect expression of desaturases. Further studies of desaturase gene expression in field environments combined with phenotyping of the genotypes in the same environments are needed to elucidate and partition the role of the genetic and environmental factors in FA composition in flax.

Seed FA composition has become a major target for modification by plant breeding and genetic engineering in many oil crops for food and non-food purposes (Murphy, 1996; Damude and Kinney 2008; Cahoon et al. 2010). A greater understanding of the genetic control of the FA composition of linseed oil is essential to develop flax cultivars producing oils for specific end-uses. The current study provides some thought-provoking hints to understand the genetic components such as transcription factors and genes other than the FA desaturases, controlling FA composition in flax, but further investigations are required to fill the knowledge gap.

CONNECTING TEXT BETWEEN CHAPTER 4 AND 5

In chapter 3, the genetic variability for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* was determined by sequencing 71 alleles corresponding to 26 isoforms with predicted functional mutations. We also demonstrated that different fatty acid compositions were not likely the result of differential desaturase gene expression in chapter 4. To gain additional knowledge of the genetic determinants of fatty acid composition, to better understand the general structural organization of the flax genome and, more specifically, of the loci controlling FA desaturation, a comparative structural genomics analysis of the six loci harboring the desaturase genes was performed using BAC clones of flax cultivars that differed in FA composition, i.e., CDC Bethune (~55-57% LIN) and breeding line M5791 (~65% LIN). The results of this comparative structural analysis are described in chapter 5.

STRUCTURAL ORGANIZATION OF FATTY ACID DESATURASE LOCI IN LINSEED LINES WITH CONTRASTING LINOLENIC ACID CONTENTS

Dinushika Thambugala[1,2], Raja Ragupathy[1,2], Sylvie Cloutier[1,2,3*]

[1]*Department of Plant Science, University of Manitoba, 66 Dafoe Rd, Winnipeg, MB, Canada, R3T 2N2*

[2]*Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Rd, Winnipeg, MB, Canada, R3T 2M9*

[3]*Current address: Eastern Cereal and Oilseed Research Centre, K.W. Neatby Building, 960 Carling Ave, Ottawa, ON, Canada, K1A 0C6*

* Corresponding author: Sylvie.J.Cloutier@agr.gc.ca

Author Dinushika Thambugala conducted this work as part of her PhD thesis. Dinushika Thambugala carried out the experiment, analysed and interpreted the data and co-wrote the manuscript. Dr. Raja Ragupathy participated in the analysis of the repeat elements. The supervisor Dr. Sylvie Cloutier designed the experiments, generated the sequence data, participated in the interpretation of the data and co-wrote the manuscript. The authors read and approved the final manuscript.

5.0 STRUCTURAL ORGANIZATION OF FATTY ACID DESATURASE LOCI IN LINSEED LINES WITH CONTRASTING LINOLENIC ACID CONTENTS

5.1 Abstract

Flax (*Linum usitatissimum* L.), the richest crop source of omega-3 fatty acids is a diploid plant with an estimated genome size of ~370 Mb that is well suited for studying genomic organization of agronomically important traits. In this study, 12 bacterial artificial chromosome (BAC) clones, harbouring the six FA desaturase loci *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* from the conventional variety CDC Bethune and the high linolenic acid line M5791 were sequenced, analysed and compared to understand the structural organization of these loci and to gain insight into the genetic mechanism(s) underlying fatty acid composition in flax. With one gene at every 3.2-4.6 Kb, the desaturase loci have a higher gene density than the genome's average of gene 7.8-8.2 Kb. The gene order and orientation across the two genotypes were generally conserved with the exception of *sad1* loci. The gene-rich loci harbored many other genes that may play a role in the lipid or carbohydrate metabolic/catabolic pathways. Sequence analysis of the six desaturase loci revealed a significant level of variation at the nucleotide level at the *fad2a* and *fad3b* loci with SNPs being the most frequently observed mutation type. The *fad2a* locus had 297 SNPs and 36 indels over ~95 kb region contrasting with the *fad2b* locus which had a mere seven SNPs and four indels in the ~110kb overlapping region. The organization of the *fad2b* locus was particularly complex with seven copies of the *fad2b* gene in both genotypes. The presence of *gypsy*, *copia*, MITE, *Mutator*, *hAT* and other novel repeat elements at the desaturase loci was comparable to the whole genome

composition. This structural genomic analysis provided some insights into the genomic organization and composition of the main desaturase loci of linseed and, of their complex evolution through both tandem and whole genome duplications.

5.2 Introduction

Flax (*Linum usitatissimum* L., 2n=2x=30), also referred to as linseed or flaxseed, is the richest crop source of omega (ω)-3 fatty acids (FAs) praised for their health benefits in humans and animals. Conventional linseed varieties contain 55 to 57% α-linolenic acid (ALA or LIN, C18:3 cis$^{\Delta 9,12,15}$) but genetic variation for this trait exists with Solin or Linola^{TM} lines having less than 5% ALA and high-LIN lines exceeding 65% (Green 1986; Rowland 1991; Friedt et al. 1995; Kenaschuk 2005).

In the FA biosynthesis pathway, desaturases are the key enzymes that drive the multi-step sequential pathway leading to the synthesis of PUFAs (Warude et al. 2006; Khadake et al. 2011). Two paralogous *sad* loci, *sad1* and *sad2*, have been identified and characterized from flax cultivar Glenelg and AC McDuff (Singh et al. 1994; Fofana et al. 2004). Two *fad2* genes namely, *fad2a* and *fad2b*, were cloned and characterized from flax genotypes Nike and NL97 (Krasowska et al. 2007; Khadake et al. 2009). Two *fad3* genes (*fad3a* and *fad3b*) from flax cultivar Normandy and more recently *fad3c* from flax cultivar AC McDuff and breeding lines UGG5-5 and SP2047 were identified (Vrinten et al. 2005; Banik et al. 2011). *Fad3a* and *fad3b* are the major enzymes controlling the LIN content in flax seeds (Vrinten et al. 2005) while the exact function of *fad3c* has not been established (Banik et al. 2011). Recently, an *in silico* gene mining approach was used to identify genome wide putative gene families related FA biosynthesis from flax cv. CDC

Bethune (You et al. 2014). Two new genes from the *sad* family, 13 new genes from the *fad2* family and three new genes from the *fad3* gene family were identified but their roles in fatty acid composition were not established. Structurally, they were all present as duplicated copies suggestive of recent whole genome duplication events (You et al. 2014).

Understanding the genomic organization of genes involved in the FA composition may be useful for the improvement of oil and FA traits through marker assisted breeding, mutation breeding or genetic manipulation. Flax is a self-pollinated diploid crop with a small estimated genome size of ~370 Mb making it well suited for genomic studies. In the last few years, important genetic resources have been developed for this crop including a whole genome shotgun (WGS) assembly of the widely cultivated flax cultivar CDC Bethune (Rowland et al. 2002), which contains 302 Mb of the estimated 373 Mb nuclear genome (Wang et al. 2012). The flax genome differs somewhat from other plant genomes with a large portion (~13.8%) of the genome consisting of ribosomal DNA and a small portion (6.1%) represented by known transposable elements (TEs), although it contained several novel TEs (Ragupathy et al. 2011; Gonzalez and Deyholos 2012). The high predicted gene content (43484) and the relatively high number of duplicated genes (9920) is consistent with the recent whole genome duplication (Wang et al. 2012; Sveinsson et al. 2014; You et al. 2014).

While linseed lines are believed to possess a stable genome, some fiber flax lines have the ability to respond to nutrient stress with heritable genomic alterations (Durrant 1962; Cullis 1973; Johnson et al. 2011; Bickel et al. 2012). For this reason, flax has been studied as a model of genome plasticity (Cullis 1981; 2005; Cullis and Cleary 1986).

These heritable genomic modifications include changes in total DNA content, copy number variation of repetitive regions including rDNA genes and satellite regions and, the appearance of unusual insertion elements such as the Linum Insertion Sequence 1 (LIS-1) (Bickel et al. 2012). In the fiber flax lines Stormont Cirrus (called "plastic" or Pl), Rembrandt, Hollandia and Liral Monarch, individuals exposed to certain stresses are capable of giving rise to lines termed genotrophs that show stable phenotypic and genotypic changes (Bickel et al. 2012; González and Deyholos 2012). There is so far no evidence of such plasticity in linseed morphotypes.

Despite its inherent unstable genome (Cullis, 1973; Johnson et al. 2011), current evidence suggests flax's adaptive evolution (Cullis 2005; Chen et al. 2009). Studying the genomic organization of loci harboring FA desaturase genes and their variations provide an opportunity to study the less dramatic changes occurring in the linseed genome. Moreover, comparison of the loci harboring the desaturase genes of high and intermediate LIN lines may provide some insights into their phenotypic differences. In our previous studies, we quantified the genetic variability for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* by sequencing 71 alleles corresponding to 26 isoforms with predicted functional mutations (Thambugala et al. 2013). Functional characterization of all seven FAD2 and 13 FAD3 isoforms in yeast confirmed the extensive variations in activities with some of the isoforms having no detectable activity and others having significantly greater activity than the most common isoforms (Radovanovic et al. 2014). We also demonstrated that different fatty acid compositions were not likely the result of differential desaturase gene expression (Thambugala and Cloutier 2014). To gain additional knowledge of the genetic determinants of fatty acid composition and of the structural genomic organization of

linseed, we conducted a comparative structural genomic analysis of the six desaturase loci of two linseed genotypes contrasting in their fatty acid composition.

5.3 Materials and Methods

5.3.1 BAC selection, DNA extraction and fingerprinting

Comparative genomic analyses of *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* loci were performed using BAC clones isolated from the linseed BAC libraries of CDC Bethune, a high yielding conventional linseed variety containing ~55-57% LIN (Rowland et al. 2002), and M5791, a high-LIN breeding line containing ~65% LIN (Fofana et al. 2010). Two-dimensional plate, row and column pools of the two BAC libraries were created as previously described for wheat but without super plate pools (Nilmalgoda et al. 2003). These library pools were screened with PCR primers designed from each of the six target desaturase genes and individual clones were identified (Table 5.1). The identified BAC clones were streaked and BAC DNA was prepared using the Eppendorf Perfectprep BAC 96 purification kit (5 Prime, Hamburg, Germany) according to manufacturer's instructions. The positive BAC clones were fingerprinted according to Luo et al. (2003).

Table 5.1 Primers used to identify BAC clones harboring the desaturase genes and names of the sequenced clones

Target loci	Primer sequences		BAC clone sequenced	
	Forward	Reverse	CDC Bethune	M5791
sad1	AGCCAGCCTTACGCCGTG	TCTGCTGCGATGATCCCG	317I17	212N17
sad2	ATGGGTTACGCTATTACACTCGA	CAGATGATGCAGTCTGAAGAAAG	375N24	6M22
fad2a	AGTGCCTCCACCATCCAGA	CAGAACACGCCTTGGCTG	346C18	139G15
fad2b	CGAATAAGGCGGACTCCGA	GTACCAGAACACGCCTTTGTTCT	364K11	28C5
fad3a	GCACCAATGGTGTGGCTATC	GCAGCATACATCAGATCAGAGCC	395P20	44E4
fad3b	TCAAAACTGTGGCTCTGCAG	GAAAAGAAAGCTGGGGGC	356B4	27L18

5.3.2 BAC sequencing, contig assembly, gap closing

A total of 12 BAC clones corresponding to one BAC clone for each of the six loci and two genotypes were sequenced by the Sequencing Unit of the National Research Council (Saskatoon, SK) using the 454 next generation sequencing technology (Roche, Brandford, CT). The 454 sequences were assembled using CAP3 with a minimum of 40 bp overlap between reads and greater than 90% identity (Huang and Madan 1999). The WGS sequence assembly of CDC Bethune (Wang et al. 2012; Linum usitatissimum v1.1, http://phytozome.jgi.doe.gov/pz/portal.html) was used as reference to verify the orientation and order of the contigs in the assembly using BLAST (Altschul et al. 1990) as implemented in the SOOMOS v0.6 in-house software tool (Banks, personal communication). WGS resequencing of CDC Bethune and M5791 was also performed by the Michael Smith Genome Sciences Centre of the BC Cancer Agency (Vancouver, BC) on the Illumina HiSeq 2000 (Illumina Inc., San Diego, USA). Reads were aligned using BWA (version 0.6.1) using default settings. The software package SAMtools was used to convert the sequence alignment/map (SAM) files to sorted binary alignment/map (BAM) files. A combined BAM file was created for visualization using Tablet (Milne et al. 2013) using default settings. This additional sequencing information was used to assist in the manual correction of the 454 assembly of the BAC clones.

5.3.3 Annotation of transposable elements

TE annotation was performed using an in-house repeat element database for flax. In brief, the draft genome assembly of flax was mined with a novel pipeline called LTR annotator to identify LTR retroelements. Sequences were mined using two *de novo* LTR identification programs, LTR_FINDER (Xu and Wang 2007) and LTRHarvest

(Ellinghaus et al. 2008). Homology search against the reference plant genome repeat database MIPS-REdat 9.0 (Nussbaumer et al. 2013) was used to identify known LTR elements. The draft genome assembly of flax was mined for miniature inverted repeat transposable elements (MITEs) using MITE Digger (Yang 2013). Seven superfamilies of DNA transposons (*Harbinger*, *hAT*, *L1*, *Mutator*, *TC1/Mariner*, *Helitron*, *En-Spm*) were extracted manually and curated from a set of repeat elements previously reported for flax (González and Deyholos 2012). From the remaining uncharacterized repeat elements, a total of 5573 non-redundant sequences were identified by excluding all MITEs, LTR elements and DNA transposons. Repeat element annotation for the BAC clones was extracted from this whole genome annotation process.

5.3.4 Annotation of genes

Gene prediction algorithm of FGENESH (http://www.softberry.com/) and the predicted gene sequences were further verified using FLAX GFF3 (Wang et al. 2012). BLASTx homology searches against the NCBI non-redundant protein database and BLASTn against dbEST (http://www.ncbi.nlm.nih.gov/BLAST/) were performed to confirm the predicted gene sequences and to annotate genes encoding known proteins. Candidates with an expected e-value <e−10 were reported. Candidate genes predicted only by the gene prediction software without any protein or EST matches were identified as hypothetical genes.

5.3.5 Comparative analysis

Dot plot analyses were performed using JDOTTER (Brodie et al. 2004), BLAST 2 (Tatusova and Madden 1999) and DNAMAN version 3.2 (Lynnon Biosoft, Vaudreuil, Canada) to show the pairwise alignments of the BAC sequence loci from CDC Bethune and M5791 over the common BAC regions.

5.4 Results

5.4.1 Sequence analysis and gene organization

More than 226K reads totalling 58M nucleotides representing between 17- and 77-fold coverage of the 12 BAC clones were obtained by 454 sequencing (Appendix XXIII). Complete sequence assembly was obtained for all 12 clones as detailed in Table 5.2. The BAC clones were annotated (Fig. 5.1, Appendix XXIV). Annotation revealed that ~75% of the BAC sequences consisted of protein coding genes. Each clone contained 18-43 genes for a gene density of 1 per 3.2-4.6 Kb (Table 5.2). Overall, the majority of the predicted proteins had significant similarity with proteins from flax, *Populus*, *Jatropha* and *Ricinus*. Homology searches against the NCBI-EST database indicated that the high proportion (~92%) of predicted proteins aligned with flax-ESTs and the remaining 8% with ESTs from other species.

5.4.2 Annotation of transposable elements

A total of 6.2% and 3.8% of the BAC sequences were made of TEs in CDC Bethune and M5791, respectively (Table 5.3). Retrotransposons were the most common mobile elements found in the BAC sequences and these were represented primarily by LTR type retroelements: *copia*, *gypsy* and unknown LTRs. DNA transposons represented a much

smaller proportion of the BAC sequences whereas unclassified repeats represented higher proportion of TEs than DNA transposons (Table 5.3, Appendix XXV). Most of the LTRs were solo LTRs, truncated or degenerated. One complete *copia* retroelements (spanning from 129728 to 134667bp) was identified in CDC Bethune BAC clone 317I7 (Fig. 5.1a) and one complete unclassified retroelement spanning positions 59029 to 65700 bp of BAC346C18 was identified (Fig. 5.1c). In addition, one MITE was identified in both *fad3b* BAC clones (Fig. 5.1f).

5.4.3 Annotation of genes

5.4.3.1 Structural organization of *sad1* and *sad2* loci

Despite the high conservation of the *sad1* locus with only 14 SNPs and four indels, the CDC Bethune 317I7 clone was predicted to have four additional genes in the overlapping region. However, only G33 (Lus10027481), predicted to encode a putative cytochrome P450 protein, had significant BLAST matches and evidence of expression as illustrated by the RNA-seq data (Fig. 5.1a, Appendix XXIV, Appendix XXVI). Two large deletions in BAC clone 212N17 of 1777 and 4945 bp were responsible for this structural difference. Both *sad1* and *sad2* loci contained several genes involved in carbohydrate and lipid metabolic pathways (Appendix XXIV). Of the 12 clones sequenced herein, the *sad2* BAC clones 375N24 and 6M22 displayed the shortest overlap with only seven predicted genes (Fig. 5.1b, Appendix XXIV). A total of 21 SNPs and 5 indels were found in the ~32 Kb overlap, corresponding to a higher mutation rate than the *sad1* locus (Table 5.2).

Table 5.2 Assembly and annotation summary of the BAC sequences harbouring the fatty acid desaturase loci containing *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* of CDC Bethune and M5791

FA gene	Genotype	BAC clone	Assembly size (bp)	Scaffold name	Scaffold start	Scaffold end	Orientation in scaffold	Predicted genes	Gene density (kp)	Overlapping region		
										Identity (%)	SNPs	Indels
sad1	CDC Bethune	317I7	172117	scaffold96	439292	614681	+	42	1 per 4.0	98.4	14	4
	M5791	212N17	115454	scaffold96	554750	684999	+	25	1 per 4.6			
sad2	CDC Bethune	375N24	155839	scaffold33	1502	158992	+	36	1 per 4.3	99.9	21	5
	M5791	6M22	138726	scaffold33	127220	277245	+	36	1 per 3.8			
fad2a	CDC Bethune	346C18	140743	scaffold931	52118	185768	+	43	1 per 3.2	99.1	297	36
	M5791	139G15	107370	scaffold931	92016	205725	+	33	1 per 3.2			
fad2b	CDC Bethune	346K11	127580	scaffold155	235413	361689	+	33	1 per 3.8	100	7	4
	M5791	28C5	110505	scaffold155	238161	348067	+	32	1 per 3.4			
fad3a	CDC Bethune	395P20	112909	scaffold28	617569	729850	+	26	1 per 4.3	96.9	143	8
	M5791	44E4	108314	scaffold28	622153	729849	+	26	1 per 4.1			
fad3b	CDC Bethune	356B4	68175	scaffold27	1	67337	+	18	1 per 3.7	89.5	268	40
	M5791	27L18	78447	scaffold27	1	77534	+	21	1 per 3.7			

Table 5.3 Summary of transposable elements (TEs) identified in the 12 BAC sequences

Genotype	Class	Order	Super family	Number of elements	Sequence length (bp)
CDC Bethune	Retrotransposons	LTR	*Copia*	1	4940
			Gypsy	2	23905
			Unknown	2	11778
	DNA transposons	TIR	*hAT*	1	1663
			Mutator	1	1622
		MITE	*MITE*	1	362
	Unclassified	Unclassified	*Unclassified*	4	3604
Total				**12**	**47874 (6.2%)**
M5791	Retrotransposons	LTR	*Copia*	2	12275
			Gypsy	0	0
			Unknown	1	4815
	DNA transposons	TIR	*hAT*	1	1663
			Mutator	0	0
		MITE	*MITE*	1	362
	Unclassified	Unclassified	*Unclassified*	7	6196
Total				**12**	**25311 (3.8%)**

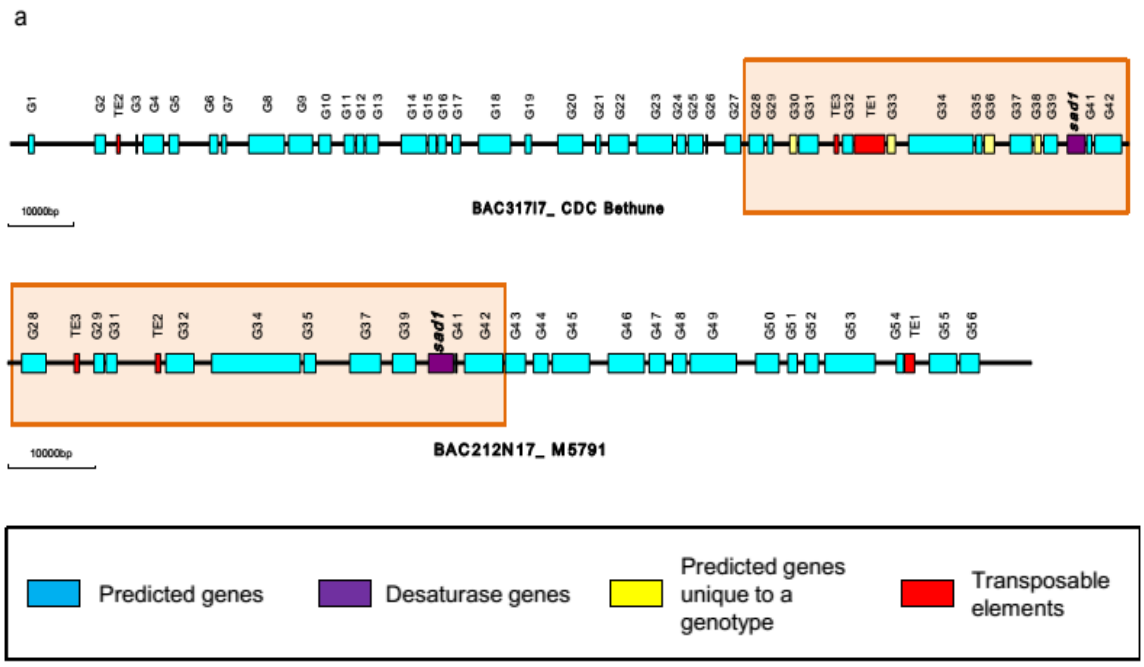


Fig 5.1 Schematic representation of the annotation of 12 BAC clones harbouring **a** *sad1*, **b** *sad2*, **c** *fad2a*, **d** *fad2b*, **e** *fad3a* and **f** *fad3b* loci of CDC Bethune and M5791.the overlapping regions are indicated by shaded boxes. Transposable elements are listed in appendix XXV and predicted genes (G) are numbered as described in appendix XXIV.

Fig 5.1 Continued

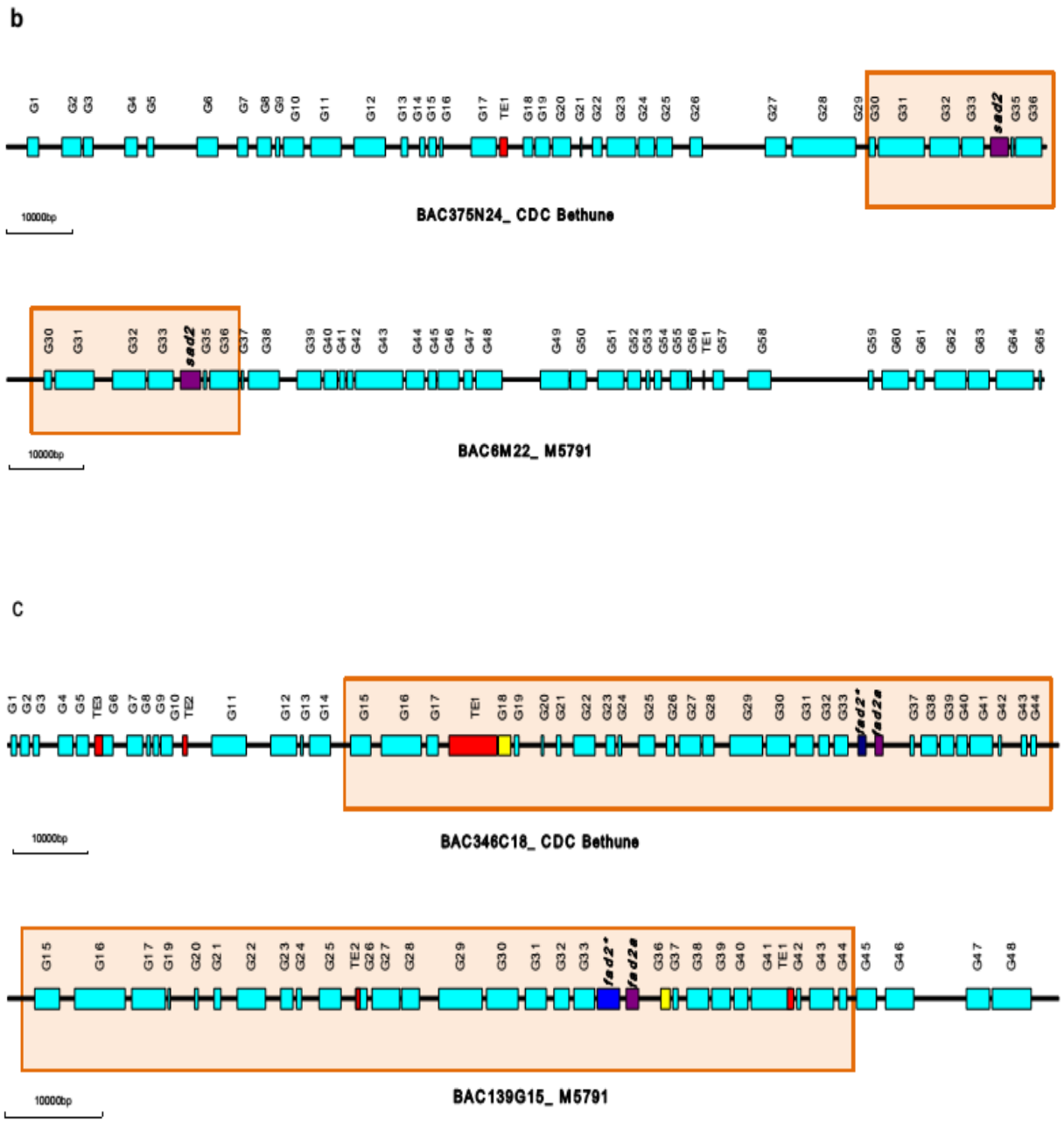


Fig 5.1 Continued

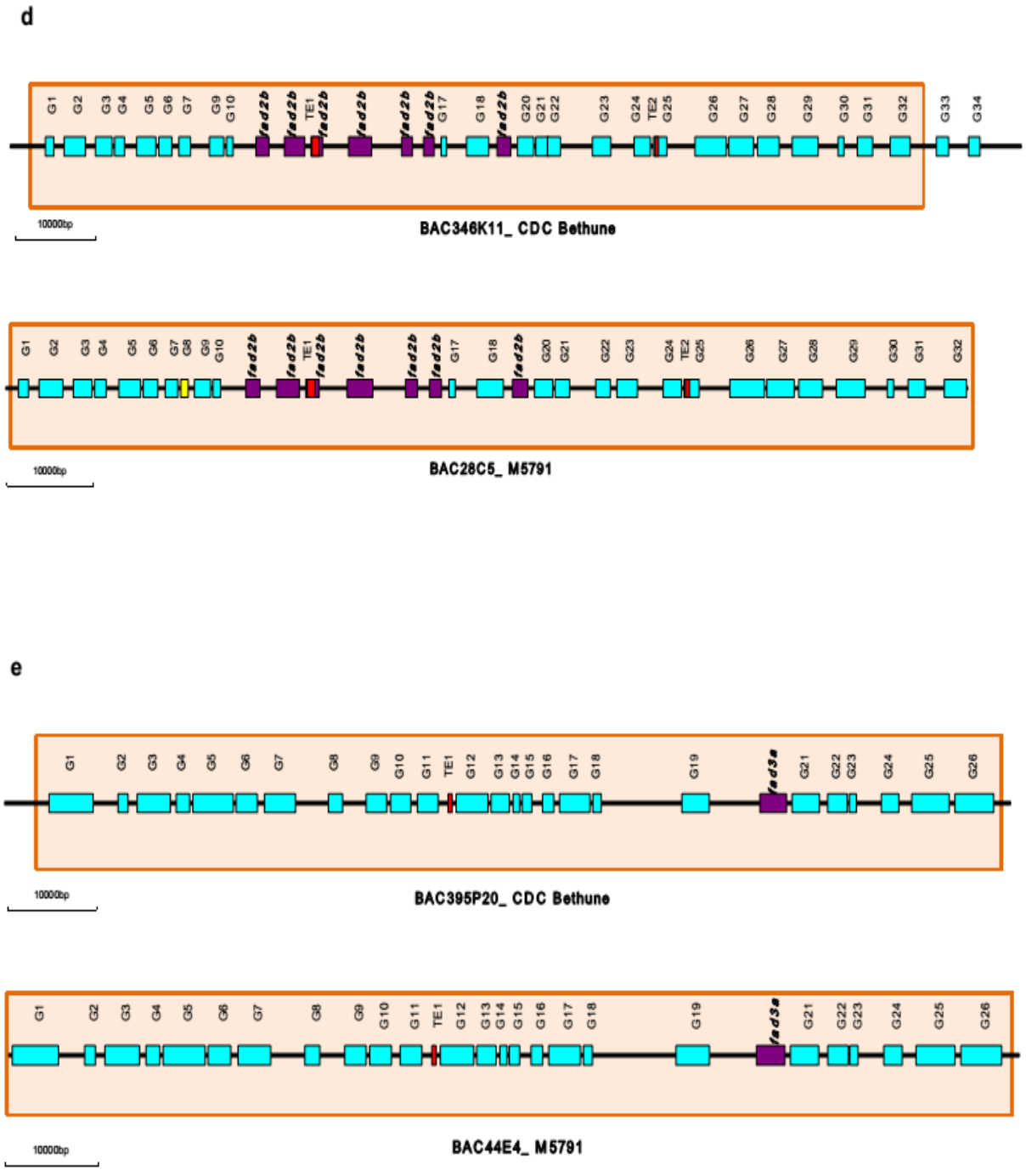
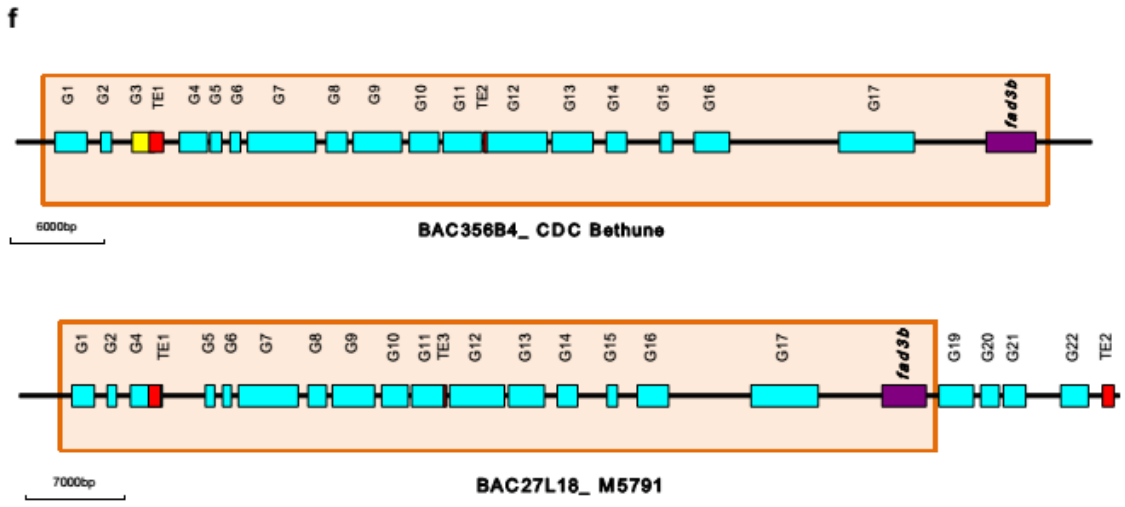


Fig 5.1 Continued



5.4.3.2 Structural organization of *fad2a* and *fad2b* loci

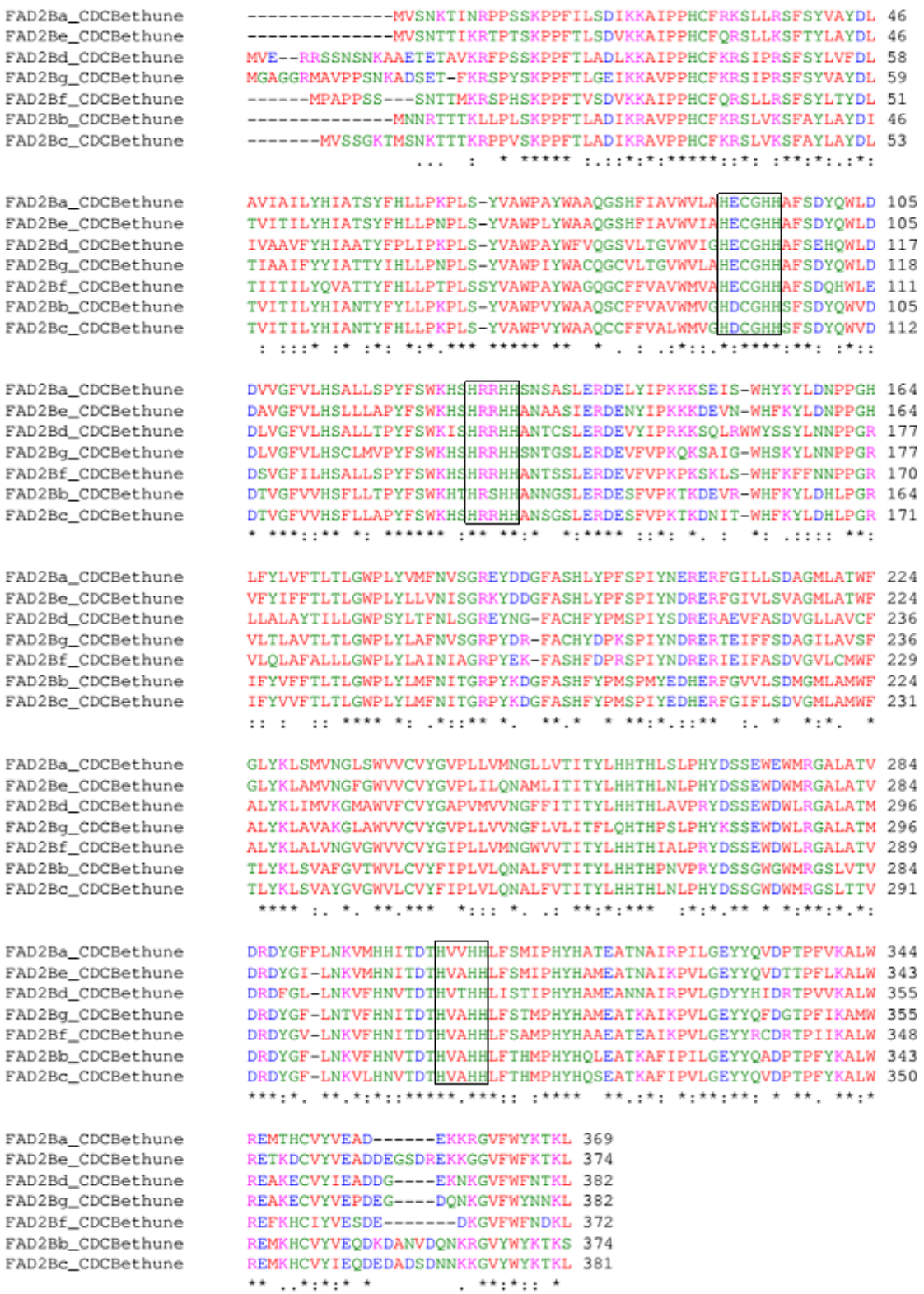
The *fad2a* locus displayed the second highest level of structural variation with a total of 297 SNPs and 36 indels in ~95 kb overlap (Table 5.2). Despite the highest gene density of one gene per 3.2 kb, only 12 genes had evidence of expression according to the RNA-seq data (Fig. 5.1c, Appendices XXIV and XXVI). The 5' end of the fad2a gene was defined by a predicted gene with significant similarity to an omega-6 desaturase (Lus10012006) which seemed expressed in all 13 RNA-seq tissue libraries (Appendix XXVI). Of the four copies of the gene encoding GDSL-like lipase acylhydrolases (Lus10011997, Lus10011998, Lus10011999 and Lus10012000+Lus10012001) identified at the *fad2a* locus, two were supported by flax EST matches but none by RNA-Seq (Appendix XXIV, Appendix XXVI). The *fad2b* locus was highly conserved with only seven SNPs and four indels over the overlapping region of ~110kb with one additional gene (G8) in M5791 (Fig. 5.1d, Appendix XXIV). Unlike the other loci, which carried a single copy of the desaturase gene, both clones carried seven copies of *fad2b*. ClustalW alignment of the deduced amino acid sequences revealed the presence of three highly

conserved histidine-rich motifs in all seven copies (Fig. 5.2a, Fig. 5.2b). The seven *fad2b* genes were predicted to encode proteins that differ at the amino acid level including some amino acid substitutions in the His-boxes but all were predicted to encode full length proteins without any premature stop codons (Fig. 5.2a, Fig. 5.2b). Although all seven copies of the *fad2b* produced significant BLAST matches to flax ESTs, RNA-Seq indicated expression for a single copy (Appendix XXVI).

5.4.3.3 Structural organization of *fad3a* and *fad3b* loci

The overall organization of ~112 and 108 kb of the fad3a BAC clones of CDC Bethune and M5791 was highly similar with a total of 26 predicted genes (Table 5.2, Fig. 5.1e). The 16 genes that were supported by flax EST included six genes with putative role in FA and carbohydrate metabolic pathways **(**Appendix XXIV**).** Of the 12 clones sequenced herein, the *fad3b* clones 356B4 and 27L18 had the highest SNP and indel rate with a total of 268 SNPs and 40 indels (Table 5.2), but their genomic organization was similar (Fig. 5.1f, Appendix XXIV).

a


FAD2Ba_CDCBethune ----------MVSNKTINRPPSSKPPFILSDIKKAIPPHCFRKSLLRSFSYVAYDL 46
FAD2Be_CDCBethune ----------MVSNTTIKRTPTSKPPFTLSDVKKAIPPHCFQRSLLKSFTYLAYDL 46
FAD2Bd_CDCBethune MVE--RRSSNSNKAAETETAVKRFPSSKPPFTLADLKKAIPPHCFKRSIPRSFSYLVFDL 58
FAD2Bg_CDCBethune MGAGGRMAVPPSNKADSET-FKRSPYSKPPFTLGEIKKAVPPHCFKRSIPRSFSYVAYDL 59
FAD2Bf_CDCBethune ------MPAPPSS---SNTTMKRSPHSKPPFTVSDVKKAIPPHCFQRSLLRSFSYLTYDL 51
FAD2Bb_CDCBethune --------------MNNRTTTKLLPLSKPPFTLADIKRAVPPHCFKRSLVKSFAYLAYDI 46
FAD2Bc_CDCBethune -------MVSSGKTMSNKTTTKRPPVSKPPFTLADIKRAVPPHCFKRSLVKSFAYLAYDL 53
FAD2Ba_CDCBethune AVIAILYHIATSYFHLLPKPLS-YVAWPAYWAAQGSHFIAVWVLAHECGHHAFSDYQWLD 105
FAD2Be_CDCBethune TVITILYHIATSYFHLLPNPLS-YVAWPLYWAAQGSHFIAVWVIAHECGHHAFSDYQWLD 105
FAD2Bd_CDCBethune IVAAVFYHIAATYFPLIPKPLS-YVAWPAYWFVQGSVLTGVWVIGHECGHHAFSEHQWLD 117
FAD2Bg_CDCBethune TIAAIFYYIATTYIHLLPNPLS-YVAWPIYWACQGCVLTGVWVLAHECGHHAFSDYQWLD 118
FAD2Bf_CDCBethune TIITILYQVATTYFHLLPTPLSSYVAWPAYWAGQGCFFVAVWMVAHECGHHAFSDQHWLE 111
FAD2Bb_CDCBethune TVITILYHIANTYFYLLPKPLS-YVAWPVYWAAQSCFFVAVWMVGHDCGHHSFSDYQWVD 105
FAD2Bc_CDCBethune TVITILYHIANTYFHLLPKPLS-YVAWPVYWAAQCCFFVALWMVGHDCGHHSFSDYQWVD 112
FAD2Ba_CDCBethune DVVGFVLHSALLSPYFSWKHSHRRHHSNSASLERDELYIPKKKSEIS-WHYKYLDNPPGH 164
FAD2Be_CDCBethune DAVGFVLHSLLLAPYFSWKHSHRRHHANAASIERDENYIPKKKDEVN-WHFKYLDNPPGH 164
FAD2Bd_CDCBethune DLVGFVLHSALLTPYFSWKISHRRHHANTCSLERDEVYIPRKKSQLRWWYSSYLNNPPGR 177
FAD2Bg_CDCBethune DLVGFVLHSCLMVPYFSWKHSHRRHHSNTGSLERDEVFVPKQKSAIG-WHSKYLNNPPGR 177
FAD2Bf_CDCBethune DSVGFILHSALLSPYFSWKHSHRRHHANTSSLERDEVFVPKPKSKLS-WHFKFFNNPPGR 170
FAD2Bb_CDCBethune DTVGFVVHSFLLTPYFSWKHTHRSHHANNGSLERDESFVPKTKDEVR-WHFKYLDHLPGR 164
FAD2Bc_CDCBethune DTVGFVVHSFLLAPYFSWKHSHRRHHANSGSLERDESFVPKTKDNIT-WHFKYLDHLPGR 171
FAD2Ba_CDCBethune LFYLVFTLTLGWPLYVMFNVSGREYDDGFASHLYPFSPIYNERERFGILLSDAGMLATWF 224
FAD2Be_CDCBethune VFYIFFTLTLGWPLYLLVNISGRKYDDGFASHLYPFSPIYNDRERFGIVLSVAGMLATWF 224
FAD2Bd_CDCBethune LLALAYTILLGWPSYLTFNLSGREYNG-FACHFYPMSPIYSDRERAEVFASDVGLLAVCF 236
FAD2Bg_CDCBethune VLTLAVTLTLGWPLYLAFNVSGRPYDR-FACHYDPKSPIYNDRERTEIFFSDAGILAVSF 236
FAD2Bf_CDCBethune VLQLAFALLLGWPLYLAINIAGRPYEK-FASHFDPRSPIYNDRERIEIFASDVGVLCMWF 229
FAD2Bb_CDCBethune IFYVFFTLTLGWPLYLMFNITGRPYKDGFASHFYPMSPMYEDHERFGVVLSDMGMLAMWF 224
FAD2Bc_CDCBethune IFYVVFTLTLGWPLYLMFNITGRPYKDGFASHFYPMSPIYEDHERFGIFLSDVGMLAMWF 231
FAD2Ba_CDCBethune GLYKLSMVNGLSWVVCVYGVPLLVMNGLLVTITYLHHTHLSLPHYDSSEWEWMRGALATV 284
FAD2Be_CDCBethune GLYKLAMVNGFGWVVCVYGVPLILQNAMLITITYLHHTHLNLPHYDSSEWDWMRGALATV 284
FAD2Bd_CDCBethune ALYKLIMVKGMAWVFCVYGAPVMVVNGFFITITYLHHTHLAVPRYDSSEWDWLRGALATM 296
FAD2Bg_CDCBethune ALYKLAVAKGLAWVVCVYGVPLLVVNGFLVLITFLQHTHPSLPHYKSSEWDWLRGALATM 296
FAD2Bf_CDCBethune ALYKLALVNGVGWVVCVYGIPLLVMNGWVVTITYLHHTHIALPRYDSSEWDWLRGALATV 289
FAD2Bb_CDCBethune TLYKLSVAFGVTWVLCVYFIPLVLQNALFVTITYLHHTHPNVPRYDSSGWGWMRGSLVTV 284
FAD2Bc_CDCBethune TLYKLSVAYGVGWVLCVYFIPLVLQNALFVTITYLHHTHLNLPHYDSSGWDWMRGSLTTV 291
FAD2Ba_CDCBethune DRDYGFPLNKVMHHITDTHVVHHLFSMIPHYHATEATNAIRPILGEYYQVDPTPFVKALW 344
FAD2Be_CDCBethune DRDYGI-LNKVMHNITDTHVAHHLFSMIPHYHAMEATNAIKPVLGEYYQVDTTPFLKALW 343
FAD2Bd_CDCBethune DRDFGL-LNKVFHNVTDTHVTHHLISTIPHYHAMEANNAIRPVLGDYYHIDRTPVVKALW 355
FAD2Bg_CDCBethune DRDYGF-LNTVFHNITDTHVAHHLFSTMPHYHAMEATKAIKPVLGEYYQFDGTPFIKAMW 355
FAD2Bf_CDCBethune DRDYGV-LNKVFHNITDTHVAHHLFSAMPHYHAAEATEAIKPVLGEYYRCDRTPIIKALW 348
FAD2Bb_CDCBethune DRDYGF-LNKVFHNVTDTHVAHHLFTHMPHYHQLEATKAFIPILGEYYQADPTPFYKALW 343
FAD2Bc_CDCBethune DRDYGF-LNKVLHNVTDTHVAHHLFTHMPHYHQSEATKAFIPVLGEYYQVDPTPFYKALW 350
FAD2Ba_CDCBethune REMTHCVYVEAD------EKKRGVFWYKTKL 369
FAD2Be_CDCBethune RETKDCVYVEADDEGSDREKKGGVFWFKTKL 374
FAD2Bd_CDCBethune REAKECVYIEADDG----EKNKGVFWFNTKL 382
FAD2Bg_CDCBethune REAKECVYVEPDEG----DQNKGVFWYNNKL 382
FAD2Bf_CDCBethune REFKHCIYVESDE--------DKGVFWFNDKL 372
FAD2Bb_CDCBethune REMKHCVYVEQDKDANVDQNKRGVYWYKTKS 374
FAD2Bc_CDCBethune REMKHCVYIEQDEDADSDNNKKGVYWYKTKL 381

b

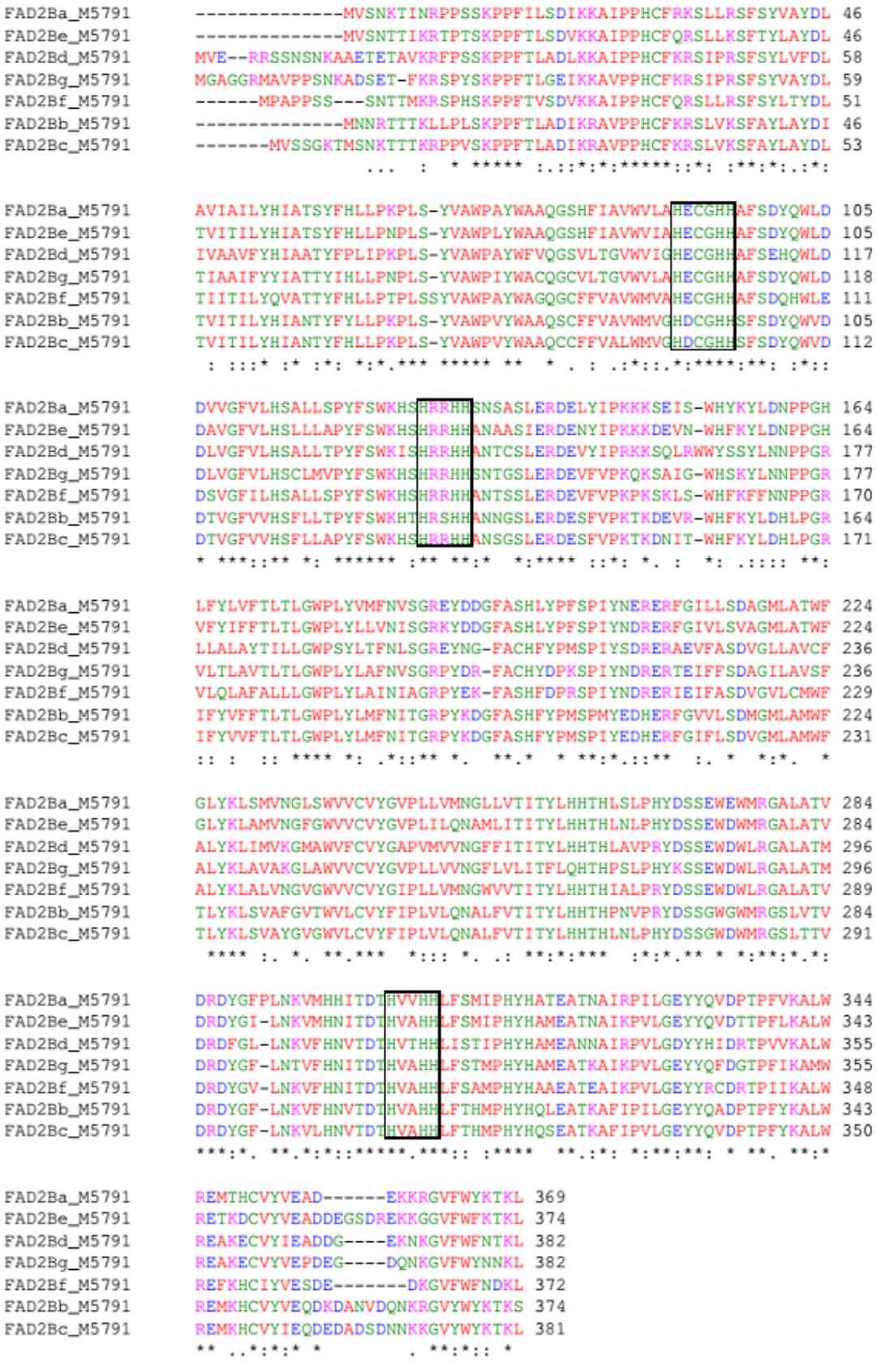

FAD2Ba_M5791 ----------MVSNKTINRPPSSKPPFILSDIKKAIPPHCFRKSLLRSFSYVAYDL 46
FAD2Be_M5791 ----------MVSNTTIKRTPTSKPPFTLSDVKKAIPPHCFQRSLLKSFTYLAYDL 46
FAD2Bd_M5791 MVE--RRSSNSNKAAETETAVKRFPSSKPPFTLADLKKAIPPHCFKRSIPRSFSYLVFDL 58
FAD2Bg_M5791 MGAGGRMAVPPSNKADSET-FKRSPYSKPPFTLGEIKKAVPPHCFKRSIPRSFSYVAYDL 59
FAD2Bf_M5791 ------MPAPPSS---SNTTMKRSPHSKPPFTVSDVKKAIPPHCFQRSLLRSFSYLTYDL 51
FAD2Bb_M5791 ----------MNNRTTTKLLPLSKPPFTLADIKRAVPPHCFKRSLVKSFAYLAYDI 46
FAD2Bc_M5791 -------MVSSGKTMSNKTTTKRPPVSKPPFTLADIKRAVPPHCFKRSLVKSFAYLAYDL 53
AVIAILYHIATSYFHLLPKPLS-YVAWPAYWAAQGSHFIAVWVLAHECGHHAFSDYQWLD 105
TVITILYHIATSYFHLLPNPLS-YVAWPLYWAAQGSHFIAVWVIAHECGHHAFSDYQWLD 105
IVAAVFYHIAATYFPLIPKPLS-YVAWPAYWFVQGSVLTGVWVIGHECGHHAFSEHQWLD 117
TIAAIFYYIATTYIHLLPNPLS-YVAWPIYWACQGCVLTGVWVLAHECGHHAFSDYQWLD 118
TIITILYQVATTYFHLLPTPLSSYVAWPAYWAGQGCFFVAVWMVAHECGHHAFSDQHWLE 111
TVITILYHIANTYFYLLPKPLS-YVAWPVYWAAQSCFFVAVWMVGHDCGHHSFSDYQWVD 105
TVITILYHIANTYFHLLPKPLS-YVAWPVYWAAQCCFFVALWMVGHDCGHHSFSDYQWVD 112
DVVGFVLHSALLSPYFSWKHSHRRHHSNSASLERDELYIPKKKSEIS-WHYKYLDNPPGH 164
DAVGFVLHSLLLAPYFSWKHSHRRHHANAASIERDENYIPKKKDEVN-WHFKYLDNPPGH 164
DLVGFVLHSALLTPYFSWKISHRRHHANTCSLERDEVYIPRKKSQLRWWYSSYLNNPPGR 177
DLVGFVLHSCLMVPYFSWKHSHRRHHSNTGSLERDEVFVPKQKSAIG-WHSKYLNNPPGR 177
DSVGFILHSALLSPYFSWKHSHRRHHANTSSLERDEVFVPKPKSKLS-WHFKFFNNPPGR 170
DTVGFVVHSFLLTPYFSWKHTHRSHHANNGSLERDESFVPKTKDEVR-WHFKYLDHLPGR 164
DTVGFVVHSFLLAPYFSWKHSHRRHHANSGSLERDESFVPKTKDNIT-WHFKYLDHLPGR 171
LFYLVFTLTLGWPLYVMFNVSGREYDDGFASHLYPFSPIYNERERFGILLSDAGMLATWF 224
VFYIFFTLTLGWPLYLLVNISGRKYDDGFASHLYPFSPIYNDRERFGIVLSVAGMLATWF 224
LLALAYTILLGWPSYLTFNLSGREYNG-FACHFYPMSPIYSDRERAEVFASDVGLLAVCF 236
VLTLAVTLTLGWPLYLAFNVSGRPYDR-FACHYDPKSPIYNDRERTEIFFSDAGILAVSF 236
VLQLAFALLLGWPLYLAINIAGRPYEK-FASHFDPRSPIYNDRERIEIFASDVGVLCMWF 229
IFYVFFTLTLGWPLYLMFNITGRPYKDGFASHFYPMSPMYEDHERFGVVLSDMGMLAMWF 224
IFYVVFTLTLGWPLYLMFNITGRPYKDGFASHFYPMSPIYEDHERFGIFLSDVGMLAMWF 231
GLYKLSMVNGLSWVVCVYGVPLLVMNGLLVTITYLHHTHLSLPHYDSSEWEWMRGALATV 284
GLYKLAMVNGFGWVVCVYGVPLILQNAMLITITYLHHTHLNLPHYDSSEWDWMRGALATV 284
ALYKLIMVKGMAWVFCVYGAPVMVVNGFFITITYLHHTHLAVPRYDSSEWDWLRGALATM 296
ALYKLAVAKGLAWVVCVYGVPLLVVNGFLVLITFLQHTHPSLPHYKSSEWDWLRGALATM 296
ALYKLALVNGVGWVVCVYGIPLLVMNGWVVTITYLHHTHIALPRYDSSEWDWLRGALATV 289
TLYKLSVAFGVTWVLCVYFIPLVLQNALFVTITYLHHTHPNVPRYDSSGWGWMRGSLVTV 284
TLYKLSVAYGVGWVLCVYFIPLVLQNALFVTITYLHHTHLNLPHYDSSGWDWMRGSLTTV 291
DRDYGFPLNKVMHHITDTHVVHHLFSMIPHYHATEATNAIRPILGEYYQVDPTPFVKALW 344
DRDYGI-LNKVMHNITDTHVAHHLFSMIPHYHAMEATNAIKPVLGEYYQVDTTPFLKALW 343
DRDFGL-LNKVFHNVTDTHVTHHLISTIPHYHAMEANNAIRPVLGDYYHIDRTPVVKALW 355
DRDYGF-LNTVFHNITDTHVAHHLFSTMPHYHAMEATKAIKPVLGEYYQFDGTPFIKAMW 355
DRDYGV-LNKVFHNITDTHVAHHLFSAMPHYHAAEATEAIKPVLGEYYRCDRTPIIKALW 348
DRDYGF-LNKVFHNVTDTHVAHHLFTHMPHYHQLEATKAFIPILGEYYQADPTPFYKALW 343
DRDYGF-LNKVLHNVTDTHVAHHLFTHMPHYHQSEATKAFIPVLGEYYQVDPTPFYKALW 350
REMTHCVYVEAD------EKKRGVFWYKTKL 369
RETKDCVYVEADDEGSDREKKGGVFWFKTKL 374
REAKECVYIEADDG----EKNKGVFWFNTKL 382
REAKECVYVEPDEG----DQNKGVFWYNNKL 382
REFKHCIYVESDE-------DKGVFWFNDKL 372
REMKHCVYVEQDKDANVDQNKRGVYWYKTKS 374
REMKHCVYIEQDEDADSDNNKKGVYWYKTKL 381

Fig 5.2 Multiple sequence alignment of the deduced amino acid sequences of the seven tandem duplicates of *fad2b* of **a** CDC Bethune and **b** M5791. Large boxes identify the three His-boxes. Identical residues are indicated by asterisks (*) and gaps are identified with dashes (-). Conserved amino acid substitutions are denoted with colons (:) and semi-conserved substitutions are indicated by dots (.).

5.4.4 Comparative analysis

The *sad1* genomic region shared 98.4% identity over the ~60 kb overlap with two large deletions of 1777 and 4945 bp in the 212N17 BAC clone as well as several other small indels and SNPs which were responsible for the size difference and the divergence between the two clones (Table5.2, Appendix XXVII). As visualized by the *fad2a* dot plot, the two sequences shared a repeat region, one large indel and a number of SNPs and small indels (Table 5.2, Appendix XXVII). Several small and one large indels as well as a large number of SNPs were responsible for the ~10% divergence in sequence identity between the two *fad3b* clones (Table 5.2, Appendix XXVII). The *fad2b* locus region had 100% sequence identity over their 111 kb overlap. The same was true for the comparison of *fad3a* sequences (Table 5.2).

5.5 Discussion

The WGS reference sequence of flax variety CDC Bethune contains ~302 Mb of the estimated nuclear genome of 373 Mb and its *in silico* gene prediction and annotation are publicly available (Wang et al. 2012; *Linum usitatissimum* v1.1, http://phytozome.jgi.doe.gov/pz/portal.html). The availability of the genome sequence and its annotated genes facilitates comparative genomic studies and provides an effective resource to unravel the mechanisms and genetic factors responsible for important traits including FA composition (Bowman et al. 2007).

Flax is the leading source of plant-based omega-3 FAs with great genetic variation for LIN content (Green 1986; Rowland 1991; Friedt et al. 1995; Kenaschuk 2005). Because desaturases are the major enzymes responsible for FA composition in oilseed crops (Ohlrogge and Jaworski 1997; Baud and Lepiniec 2010), efforts to date to understand the molecular basis of high LIN content in flax have mainly focused on the genetic variability and expression of FA desaturase genes (Pan et al. 2013). However, much less emphasis was placed on the structural analysis of the loci that harboured the fatty acid desaturase genes. Here, we performed a comparative analysis of six fatty acid desaturase loci in two lines that differed in FA composition, i.e., CDC Bethune (~55-57% LIN) and breeding line M5791 (~65% LIN) to gain insight into the potential structural genetic alterations that may affect FA composition in linseed and, to more broadly understand the structural organization and evolution of these loci**.**

TEs can make up a large fraction of plant genomes ranging from approximately 14 % in Arabidopsis to more than 80% in maize (Schnable et al. 2009). The flax genome contains ~24% TEs of which, only ~6.1% are known elements (Ragupathy et al. 2011; Wang et al. 2012; González and Deyholos 2012). The presence of *Gypsy*, *Copia*, MITE and novel (unclassified) repeat elements at the desaturase loci resembled the whole genome composition although the regions characterized herein represented only a small fraction (~0.2%) of the genome, precluding a generalized conclusion. TEs have been shown to influence neighboring gene expression (Feschotte et al. 2002; Defraia and Slotkin 2014). Genes within TE-rich regions were reported to have lower transcript expression (González and Deyholos 2012) although TE can also contain enhancer elements that increase expression of neighboring genes (Makarevitch et al. 2015). Our

data set which comprised six loci and the related publically available RNA-Seq data for CDC Bethune is not comprehensive enough to lend support either way to the specific role(s) of TEs in enhancing or repressing expression of the neighboring genes. However, the knowledge of the TE composition of two genotypes provides the necessary information to test the hypotheses through Real-Time PCR for example. Such experiment must; however, seriously consider the paralogs because the TE silencing effect was shown to be reduced by gene redundancy (González and Deyholos 2012).

The predicted gene content of CDC Bethune is relatively high (1 per 7.8-8.2 Kb) compared to other plant species (Ragupathy et al. 2011; Wang et al. 2012). This is in line with the relatively low content of repetitive sequences of the flax genome (Cullis et al. 2005; Ragupathy et al. 2011; Wang et al. 2012) and was hypothesized to be the result of a recent whole genome duplication (Wang et al. 2012; Sveinsson et al. 2014; You et al. 2014). With one gene every 3.2-4.6 Kb, the desaturase loci have a gene density that is approximately twice that of the genome's average.

Over the past several decades, the flax genome has been well documented for its plasticity (Cullis 1981, 1986, 2005) and was, in fact, reported to be inherently unstable (Cullis, 1973; Johnson et al. 2011). These conclusions were; however, obtained from a small number of fibre varieties which seemed able to respond to stressful growth conditions by reshuffling their genomes in an organized fashion producing stable genotrophs, an ability that has not been shown in linseed lines which seem to possess stable genomes (Durrant 1962, Johnson et al. 2011, Bickel et al. 2012). This is in line with our observations at the *fad* and *sad* loci where the gene order and orientation were

generally conserved between the two genotypes as previously reported for soybean (Schlueter et al. 2006).

Although a single *fad2* gene was identified in *Arabidopsis*, in most other plant species, FAD2 is encoded by small gene families such as in soybean and cotton where three and four members were reported, respectively (Heppard et al. 1996; Li et al. 2007; Zhang et al. 2009; Cao et al. 2013). The existence of multiple *fad2* gene copies in *Brassica napus* and soybean arose from repeated rounds of genome duplication (Scheffler et al. 1997; Mikkilineni and Rocheford 2003; Schlueter et al. 2007; Schmutz et al. 2010). Current evidence surrounding the evolution of the flax genome indicate an ancient whole genome duplication event (20-40 MYA, Sveinsson et al. 2014) and a more recent one (5-9 MYA, Wang et al. 2012) in line with the results obtained by You et al. (2014) who found in the CDC Bethune genome, two unlinked copies of *fad2a*, *fad3a* and *fad3b*. The *fad2b* locus was however comprised of a cluster of *fad2* genes and the whole cluster, with the exception of one gene, was duplicated. Indeed, the cluster on linkage group 8 contained seven copies of *fad2b-like* genes and the cluster on linkage group 6 contained six (You et al. 2014). It was thus hypothesized that the *fad2b* locus underwent tandem duplication events sometimes between the two whole genome duplication events while the gene loss event on linkage group 6 would be more recent.

The *fad2b* gene plays a major role in producing LIO with constitutive expression in both vegetative tissues and developing seeds (Cao et al. 2013; Schlueter et al. 2007). The amino acid substitutions identified in the three histidine boxes were previously reported in functionally divergent FAD2 enzymes in other plant species (Cao et al. 2013). For example, the first histidine motif was HECGHH in FAD2B-a, e, d, g and f, except

FAD2B-b and c that had HDCGHH (Fig.2a, Fig.2b). The actual number of genes with functionality in seeds may be smaller than the number of duplicated copies because some of these loci may represent genes not expressed in seeds or pseudogenes that are not functional (Scheffler et al. 1997). The RNA-seq data indicated that only one of the *fad2b* genes was expressed and that expression had been detected in all 13 tissues tested (Appendix XXVI) with significantly higher expression in developing seed tissues. We caution to the interpretation of this result because the assembly criteria may not have been stringent enough to distinguish among family members. Real-time PCR with primers specific to each family member would be required to confirm the expression levels and the tissue-specificity of each one. From a structural point of view, the seven *fad2b* genes were predicted to encode proteins that differ at the amino acid level including some amino acid substitutions in the His-boxes but all were predicted to encode full length proteins (no premature stop codons), albeit of slightly different length, once again reiterating the need to be cautious with the RNA-Seq data interpretation.

Following duplication, paralogs may either retain their original gene function or evolve sub-functionalization or neo-functionalization or, alternatively, be silenced through the accumulation of non-sense mutations, frame shifts, insertions or deletions (Force et al. 1999). Two such paralogous loci, *fad2a* and *fad2b*, seem to have evolved at a different pace with a much higher mutation tolerance in *fad2a* contrasting with the conservation of the *fad2b* locus with few nucleotide changes. While the redundancy may have allowed for higher divergence between sequences, the more conserved nature of *fad2b* may indicate its essential role.

In conclusion, understanding the genomic organization of genes involved in the FA composition may be useful for the improvement of oil and FA traits through molecular assisted breeding, mutation breeding or genetic manipulation. Detailed comparative analysis of 12 BAC clones harbouring six FA desaturase loci in high and intermediate LIN lines has provided an initial insight in to the genomic organization and uniqueness of these *fad* and *sad* genomic regions and stability of linseed genome compared to fibre flax genome.

6.0 GENERAL DISCUSSION AND CONCLUSION

Flax (*Linum usitatissimum* L) is the richest crop source of omega (ω)-3 fatty acids (FAs) and has attracted great attention recently in human health mostly because of its desirable fatty acid composition. With an estimated genome size of ~370 Mb and the availability of the genome sequence and its annotated genes, flax offers good resources for studying mechanisms and genetic factors responsible for important traits including FA composition. Despite importance and high market prospects of this crop, few studies have been conducted to assess the genetic diversity, expression and organization of genes involved in its fatty acid (FA) composition.

Here we reported on the extent of the genetic variability for genes encoding stearoyl-ACP desaturase (SAD), fatty acid desaturase 2 (FAD2) and 3 (FAD3) by sequencing the six paralogous genes from 120 flax accessions representing a broad range of germplasm including some EMS mutant lines. These genotypic data were correlated to the FA composition of the lines grown in the field during three years (2010-2012) and at two locations (Saskatoon and Morden) to hypothesize the functionality of these alleles and isoforms in FA biosynthesis in flax. Sequence analysis of the six genes revealed a significant level of variation at the nucleotide level with SNPs being the most frequently observed mutation type. Most of these point mutations were synonymous substitutions that did not alter the underlying amino acid sequences. Consistent with other plant desaturases, the four membrane-bound *fad* genes sequenced herein were predicted to encode FADs possessing the highly conserved histidine-rich motifs. Only accession SP2047 had a point mutation in one of the HIS-box of FAD3B which was previously

shown to be non-functional. The level of sequence conservation observed between the paralogous desaturases was also reported for other plant desaturases. The high sequence identity at both DNA and amino acid levels between *sad* sequences was reported in other plants and can be interpreted as an indication of the essential role of Δ9-desaturase in the lipid biosynthetic pathway in plants. *Fad2a*'s highest allelic diversity of 21 alleles contrasted with the conservation of *fad2b* with only five. The *fad2b* gene plays a major role in LIO production and remains a house-keeping microsomal Δ12 oleate desaturase with constitutive expression throughout the plant. In soybean, *fad2-2,* the orthologue of *fad2b,* was shown to be the most important gene for increasing LIO content.

Correlation of these genotypic data with fatty acid composition data of 120 flax accessions phenotyped in six field experiments revealed statistically significant effects of some of the SAD and FAD isoforms on fatty acid composition, oil content and iodine value. In flax, functional redundancy of six paralog desaturases provides additional buffering capacity for mutation tolerance even in exons as exemplified by the predicted non-functional FAD3A-D, FAD3A-E and FAD3B-B isoforms. For example, redundancy may have allowed for higher divergence between *fad3a* and *fad3b* sequences while preserving its role, consistent with the previous studies showing the additive role of *fad3a* and *fad3b* and their equal contribution to LIN content in flax. Selection pressure over the process of domestication might have a significant impact on the observed variation in *fad2, sad1* and *sad2* as illustrated by the NJ trees. The impact of selection on *fad2* and *fad3* diversity during domestication was also reported in cultivated sunflower.

Further, little was known about the relationship between expression levels of fatty acid desaturase genes during seed development and FA composition in flax. At the onset

of this research, only two studies on the regulation and expression of *sad* and *fad* genes during seed development had been reported. Our first study revealed that thirty-four lines had the same six isoforms but differed in FA composition. A subset of six lines including two moderately low, two intermediate and two high LIN lines were selected to establish the relationship between the desaturase expression and FA composition. Semi-quantitative reverse transcriptase (RT-) PCR conducted on developing seeds from eight to 32 days after anthesis for all six lines revealed that all six FA desaturase genes displayed a bell-shaped pattern of expression during seed development peaking around 20 days after anthesis which is in line with other gene expression pattrens related to FA synthesis. The highly conserved nature of *sad2* and its higher expression are in agreement with its essential Δ9-desaturase role in lipid biosynthesis in flax. Although promoters of the six desaturase genes share a functionally similar promoter core, their expression can be modulated by differences in upstream regulatory elements. The strong seed specific expression of the sesame *fad2* gene controlled by negative cis-regulatory elements of the promoter and enhancers located in the 5'-UTR (untranslated region) has been demonstrated.

Desaturase expression did not correlate with FA composition variations of the six flax genotypes studied, hence other genetic factors were hypothesized to play a role in determining the FA composition of flax. Genes of minor effect have generally been considered responsible for variation observed in LIN content in several oil crops including flax. Differential expression between field and growth cabinet conditions was also hypothesized to explain the lack of correlation. Environment and genotype by

environment significantly affect FA composition and the effect of the environment on *fad2* expression in flax has been documented.

To better understand the general structural organization of the flax genome and, more specifically, of the loci controlling FA desaturation, a comparative structural genomics analysis of the six loci harboring the desaturase genes (*sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a*, *fad3b*) was performed using bacterial artificial chromosome (BAC) libraries of flax cultivars that differed in FA composition, i.e., CDC Bethune (~55-57% LIN) and breeding line M5791 (~65% LIN). The presence of *Gypsy*, *Copia*, MITE and novel (unclassified) repeat elements at the desaturase loci resembled the whole genome composition although the regions characterized herein represented only a small fraction (~0.2%) of the genome, precluding a generalized conclusion. The flax genome differs somewhat from other plant genomes with a large portion (~13.8%) of the genome consisting of ribosomal DNA and a small portion (6.1%) represented by known transposable elements TEs. With one gene at every 3.2-4.6 Kb, the desaturase loci have a higher gene density than the genome's average of one gene per 7.8-8.2 Kb. The gene order and orientation across the two genotypes were generally conserved with the exception of the *sad1* locus as reported for linseed lines which seem to possess a stable genome.

Two paralogous loci, *fad2a* and *fad2b*, seem to have evolved at a different pace with a much higher mutation tolerance in *fad2a* contrasting with the conservation of the *fad2b* locus. While the redundancy may have allowed for higher divergence between sequences, the more conserved nature of *fad2b* may indicate its essential role. The organization of the *fad2b* locus was particularly complex with seven copies of the *fad2b*

gene in both genotypes. It was hypothesized that the *fad2b* locus underwent tandem duplication events sometimes between the two whole genome duplication events. From a structural point of view, the seven *fad2b* genes were predicted to encode proteins that differ at the amino acid level including some amino acid substitutions in the His-boxes but all were predicted to encode full length proteins albeit of slightly different length. However, the actual number of genes expressed in seeds and encoding functional isoforms is unknown. Some of these paralogs may represent genes not expressed in seeds or encode non-functional proteins. The RNASeq data seem to corroborate the former considering that expression of a single *fad2b* gene was identified.

FA composition can be altered by manipulating one or more steps of their biosynthesis pathway. Novel allelic variants and isoforms identified for the six desaturases provide useful genetic and molecular resources and information for the development of oilseed flax with unique and useful oil profiles that would not require a transgenic or a mutagenesis approach. The structural genomic analysis provided some insights into the genomic organization and composition of the main desaturase loci and, of their complex evolution through both tandem and whole genome duplications. Expression analysis of fatty acid desaturase genes during seed development provided some hints to understand the genetic components such as transcription factors and genes other than the FA desaturases, controlling FA composition in flax, but further investigations are required to fill the knowledge gap.

7.0 LITERATURE CITED

Abdallah A, Ahumada MH, Gradziel TM (1998) Oil content and fatty acid composition of almond kernels from different genotypes and California production regions. J Am Soc Hort 123:1029-1033

Allaby RG, Peterson GW, Merriwether DA, Fu YB (2005) Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the *sad2* locus. Theor Appl Genet 112:58-65

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Ander BP, Weber AR, Rampersad PP, Gilchrist JSC, Pierce GN, Lukas A (2004) Dietary flaxseed protects against ventricular fibrillation induced by ischemia-reperfusion in normal and hypercholesterolemic rabbits. J Nutr 134:3250-3256

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258:1353-1355

Avelange-Macherel MH, Macherel D, Wada H, Murata N (1995) Site-directed mutagenesis of histidine in the Δ12 acyl lipid desaturase of *Synechocystis*. FEBS Lett 361:111-114

Ayerza R(h) (2009) The seed's protein and oil content, fatty acid composition and growing cycle length of a single genotype of chia (*Salvia hispanica* L.) as affected by environmental factors. J Oleo Sci 58(7):347-354

Banik M, Duguid S, Cloutier S (2011) Transcript profiling and gene characterization of three fatty acid desaturase genes in high, moderate and low linolenic acid genotypes of flax (*Linum usitatissimum* L.) and their role in linolenic acid accumulation. Genome 54:471-483

Baud S, Graham IA (2006) A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of *Arabidopsis* using *in situ* histochemistry. Plant J 46:155-169

Baud S, Lepiniec L (2009) Regulation of *de novo* fatty acid synthesis in maturing oilseeds of *Arabidopsis*. Plant Physiol Biochem 47:448-455

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

Baud S, Mendoza MS, To A, Harscoët E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. Plant J 50:825-838

Bene H, Lasky D, Ntambi JM (2001) Cloning and characterization of the human stearoyl-coa desaturase gene promoter: transcriptional activation by sterol regulatory element binding protein and repression by polyunsaturated fatty acids and cholesterol. Biochem Biophys Res Commun 284:1194-1198

Bickel CL, Lukacs M, Cullis CA (2012) The loci controlling plasticity in flax. Research and Reports in Biology 3:1-11

Bocianowski J, Mikołajczyk K, Bartkowiak-Broda I (2012) Determination of fatty acid composition in seed oil of rapeseed (*Brassica napus* L.) by mutated alleles of the FAD3 desaturase genes. J Appl Genet 53:27-30

Bowman JL, Floyd SK, Sakakibara K (2007) Green genes-comparative genomics of the green branch of life. Cell 129:229-234

Brace RC, Fehr WR, Schnebly SR (2011) Agronomic and seed traits of soybean lines with high oleate concentration. Crop Sci 51:534-541

Broadwater JA, Whittle E, Shanklin J (2002) Desaturation and hydroxylation-residues 148 and 324 of *Arabidopsis* FAD2, in addition to substrate chain length, exert a major influence in partitioning of catalytic specificity. J Biol Chem 277:15613-15620

Brodie R, Roper RL, Upton C (2004) JDOTTER: a java interface to multiple dotplots generated by dotter. Bioinformatics 20:279-281

Brookes AJ (1999) The essence of SNPs. Gene 234:177-186

Broun P, Boddupalli S, Somerville C (1998) A bifunctional oleate 12-hydroxylase: desaturase from *Lesquerella fendleri*. Plant J 13:201-210

Browse J, Somerville C (1991) Glycerolipid biosynthesis: biochemistry and regulation. Ann Rev Plant Physiol Plant Mol Biol 42:467-506

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107-111

Burton JW (1987) Quantitative genetics: results relevant to soybean breeding. In: Wilcox JR (ed) Soybeans: improvement, production, and uses, 2nd edn. ASA, CSSA, and SSSA, Madison, pp 211-242

Byfield GE, Upchurch RG (2007) Effect of temperature on delta-9 stearoyl-ACP and microsomal omega-6 desaturase gene expression and fatty acid content in developing soybean seeds. Crop Sci 47:1698-1704

Cahoon E, Clemente T, Damude H, Kinney A (2010) Modifying vegetable oils for food and non-food purposes. In: Vollmann J, Rajcan I (eds) Oil Crops. Springer, New York, pp 31-56

Cao S, Zhou XR, Wood CC, Green AG, Singh SP, Liu L, Liu Q (2013) A large and functionally diverse family of *Fad2* genes in safflower (*Carthamus tinctorius* L.). BMC Plant Biol 13:5

Carter JF (1993) Potential of flaxseed and flaxseed oil in baked goods and other products in human nutrition. Cereal Foods World 38:753-759

Casa R, Russell G, Lo Cascio B, Rossini F (1999) Environmental effects on linseed (*Linum usitatissimum* L.) yield and growth of flax at different stand densities. Eur J Agron 11:267-278

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: Plant Promoter Analysis Navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene group. BMC Genomics 9:561

Chapman MA, Burke JM (2012) Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. Theor Appl Genet 125:897-907

Chen Y, Lowenfeld R and Cullis CA (2009) An environmentally induced adaptive (?) insertion event in flax. Int J Genet Mol Biol 1:38-47

Chen Z, Wang M, Barkley NA, Pittman RN (2010) A simple allele-specific PCR assay for detecting FAD2 alleles in both A and B genomes of the cultivated peanut for high-oleate trait selection. Plant Mol Biol Rep 28:542-548

Chi X, Yang Q, Zhao F, Qin S, Yang Y, Shen J, Lin H (2008) Comparative analysis of fatty acid desaturases in cyanobacterial genomes. Comp Funct Genomics 2008:284508

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19-33

Choquer M, Boccara M, Vidal-Cros A (2003) A semi-quantitative RT-PCR method to readily compare expression levels within *Botrytis cinerea* multigenic families in vitro and in planta. Curr Genet 43:303-309

Cloutier S, Ragupathy R, Niu Z, Duguid S (2011) SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. Mol Breed 28:437-451

Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E, Walichnowski A, Ward K, Rowland G, Duguid S, Banik M (2012) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:1783-1795

Cullis CA (1973) DNA differences between flax genotrophs. Nature 243:515-516

Cullis CA (1981) DNA-Sequence Organization in the Flax Genome. Biochim Biophys Acta 652:1-15

Cullis CA (2005) Mechanisms and control of rapid genomic changes in flax. Annals Bot 95:201-206

Cullis CA (2007) Flax. In: Kole C (ed) Genome mapping and molecular breeding in plants, vol 2. Springer, Berlin, pp 275-295

Cullis CA, Cleary W (1986) Rapidly Varying DNA-Sequences in Flax. Can J Genet Cytol 28:252-259

Damude HG, Kinney AJ (2008) Engineering oilseeds to produce nutritional fatty acids. Physiol Plant 132:1-10

Das K, Rai M (1974) A diallel analysis of iodine value in linseed. Indian J Genet Plant Breed 34:718-725

Daun JK and Mazur PB (1983) Use of gas liquid chromatography for monitoring the fatty acid composition of Canadian rapeseed. J Am Oil Chem Soc 60:1751-1754

Defraia C, Slotkin RK (2014) Analysis of retrotransposon activity in plants. Methods Mol Biol 1112:195-210

Deng X, Scarth R (1998) Temperature effects on fatty acid composition during development of low-linolenic oilseed rape (*Brassica napus*). J Am Oil Chem Soc 75:759-766

Diederichsen A, Hammer K (1995) Variation of cultivated flax (*Linum usitatissimum* L. subp. *usitatissimum*) and its wild progenitor pale flax (subsp. *angustifolium* (Huds.) Thell.). Genet Resour Crop Evol 42:263-272

Diederichsen A, Richards K (2001) Cultivated flax and the genus *Linum* L.:Taxonomy and germplasm conservation. In: Westscott AM, Amsterdam N (eds) Flax: The genus *Linum*, Hardwood Academic Publishers, pp 22-54

Dillman AC (1953) Classification of flax varieties, USDA. Technical Bulletin No. 1064. U.S. Government Print Office, Washington, DC

Doucet I, Filipescu H (1981) Inheritance and content of unsaturated fatty acids in linseed. An Inst Cercet Cereale Plante Teh - Fundulea 46:35-48

Drexler H, Spiekermann P, Meyer A, Domergue F, Zank T, Sperling P, Abbadi A, Heinz E (2003) Metabolic engineering of fatty acids for breeding of new oilseed crops: strategies, problems and first results. J Plant Physiol 160:779-802

Dribnenki JCP, Green AG (1995) Linola '947' low linolenic acid flax. Can J Plant Sci 75:201-202

Dribnenki JCP, McEachern SF, Chen Y, Green AG, Rashid KY (2007) 2149 Solin (low linolenic flax). Can J Plant Sci 87:297-299

Durrant A (1962) The environmental induction of heritable change in *Linum*. Heredity 17:27-61

Dyer JM, Stymne S, Green AG, Carlsson AS (2008) High-value oils from plants. Plant J 54:640-655

Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons, BMC Bioinformatics 2008, 9:18

Eskandari M, Cober ER, Rajcan I (2013) Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. Theor Appl Genet 126:483-495

Esteban AB, Sicardo MD, Mancha M, Martinez-Rivas JM (2004) Growth temperature control of the linoleic acid content in safflower (*Carthamus tinctorius*) seed oil. J Agric Food Chem 52:332-6

Ewing B, Hillier L, Wendl MC, Green G (1998) Base-calling of automated sequencer traces using PHRED: I. accuracy assessment. Genome Res 8:175-185

FAOSTAT (2014) Production of crops: linseed: area harvested and production (tonnes). Available at http://faostat3.fao.org/home/index.html

FAOSTAT (2014) Trade: Crops and livestock products. Available at http://faostat3.fao.org/browse/T/TP/E

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 3:329-341

Fischer G, Neuve´glise C, Durrens P, Gaillardin C, Dujon B (2001) Evolution of gene order in the genomes of two related yeasts species. Genome Res 11:2009-2019

Fofana B, Duguid S, Cloutier S (2004) Cloning of fatty acid biosynthetic genes β-Ketoacyl CoA synthase, fatty acid elongase, stearoyl-ACP desaturase, and fatty acid desaturase and analysis of expression in the early developmental stages of flax (*Linum usitatissimum* L.) seeds. Plant Sci 166:1487-1496

Fofana B, Cloutier S, Duguid S, Ching J, Rampitch C (2006) Gene expression of stearoyl-ACP desaturase and Δ12 fatty acid desatuare 2 is modulated during seed development of flax (*Linum usitatissimum*). Lipids 41:705-712

Fofana B, Ragupathy R, Cloutier S (2010) Flax Lipids: Classes, biosynthesis, genetics and the promise of applied genomics for understanding and altering of fatty acids. In: Gilmore PL (ed) Lipids: Categories, biological functions and metabolism, nutrition, and health. Nova Science Publishers Inc, New York, pp 71-98

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postle-thwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545

Fox BG, Shanklin J, Somerville CR, Munck E (1993) Stearoyl-acyl carrier protein Δ9 desaturase from *Ricinus communis* is a di-iron oxoprotein. Proc Nat Acad Sci USA 90:2486-2490

Friedt W, Bickert C, Schaub H (1995) In vitro breeding of high linolenic, doubled haploid lines of linseed (*Linum usitatissimum* L.) via androgenesis. Plant Breed 114:322-326

Fu YB (2011) Population-based resequencing revealed an ancestral winter group of cultivated flax: implication for flax domestication processes. Ecol Evol 2:622-635

Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. Evol Biol 30:93-120

González LG, Deyholos MK (2012) Identification, characterization and distribution of transposable elements in the flax (*Linum usitatissimum* L.) genome. BMC Genomics13:644

Graef G, LaVallee B, Tenopir P, Tat M, Schweiger B, Kinney A, Gerpen J, Clemente T (2009) A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. Plant Biotechnol J 7:41-421

Green AG, Marshall AR (1984) Isolation of induced mutants in linseed (*Linum usitatissimum* L.) having reduced linolenic acid content. Euphytica 33:321-328

Green AG (1986) Genetic control of polyunsaturated fatty acid biosynthesis in flax (*Linum usitatissimum*) seed oil. Theor Appl Genet 72:654-661

Green AG (1986a) A mutant genotype of flax (*Linum usitatissimum* L) containing very low levels of linolenic acid in its seed oil. Can J Plant Sci 66:499-503

Green AG (1986b) Genetic control of polyunsaturated fatty acid biosynthesis in flax (*Linum usitatissimum*) seed oil. Theor Appl Genet 72:654-661

Green AG, Chen Y, Singh SP, Dribnenki JCP (2008) Flax. In: Kole C, Hall TC (eds) Compendium of transgenic crop plants: Transgenic oilseed crops. Blackwell Publishing, Chicester, pp 199-226

Gutierrez L, Mauriat M, Guénin S, Pelloux J, Lefebvre JF, Louvet R et al. (2008) The lack of a systematic validation of references genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol J 6:609-618

Hartmann RB, Fehr WR, Welke GA, Hammond EG, Duvick DN, Cianzio SR (1996) Association of elevated palmitate content with agronomic and seed traits of soybean. Crop Sci 36:1466-1470

Harwood JL (1996) Recent advances in the biosynthesis of plant fatty acids. Biochim Biophys Acta 1301:7-56

Hasler CM, Kundrat S, Wool D (2000) Functional foods and cardiovascular disease. Curr Arteroscler Rep 2:467-475

Heppard EP, Kinney AJ, Stecca KL, Miao GH (1996) Developmental and growth temperature regulation of two different microsomal omega-6 desaturase genes in soybeans. Plant Physiol 110:311-319

Hernández ML, Padilla MN, Sicardo MD, Mancha M, Martínez-Rivas JM (2011) Effect of different environmental stresses on the expression of oleate desaturase genes and fatty acid composition in olive fruit. Phytochemistry 72:178-87

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297-300

Hobbs DH, Flintham JE, Hills MJ (2004) Genetic control of storage oil synthesis in seeds of *Arabidopsis*. Plant Physiol 136:3341-3349

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868-877

Huang XQ, Cloutier, S (2008) Molecular characterization and genomic organization of low molecular weight glutenin subunit genes at the *GLU-3* loci in hexaploid wheat (*Triticum aestivum* L). Theor Appl Genet 116:953-966

Jain RK, Thomson RG, Taylor DC, MacKenzie SL, McHughen A, Rowland GG, Tenaschuk D, Coffey M (1999) Isolation and characterization of two promoters from linseed for genetic engineering. Crop Sci 39:1696-1701

Jasinski S, Lécureuil A, Miquel M, Loudet O, Raffaele S, Froissard M, Guerche P (2012) Natural variation in seed very long chain fatty acid content is controlled by a new isoform of KCS18 in *Arabidopsis thaliana.* PLoS One 7:e49261

Jaworski JG (1987) Biosynthesis of monoenoic and polyenoic fatty acids. In: Stumpf PK (ed.) The Biochemistry of Plants, Vol. 9, Academic Press, Orlando, FL, pp 159-174

Johnson C, Moss T, Cullis C (2011) Environmentally induced heritable changes in flax. J Vis Exp 47:2332

Kenaschuk EO (2005) High linolenic acid flax. US patent 6870077 issued on March 22, 2005

Kenaschuk EO, Rowland GG (1995) Flax. In: Slinkard AE, Knott DR (eds), Harvest of gold: the history of field crop breeding in Canada. University of Saskatchewan, SK, Canada, pp. 173-176

Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, vander Plas LH, Jansen RC,Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet 38:842-8

Khadake R, Khonde V, Mhaske V, Ranjekar P, Harsulkar A (2011) Functional and bioinformatic characterisation of sequence variants of *Fad3* gene from flax. J Sci Food Agric 91:2689-2696

Khadake RM, Ranjekar PK, Harsulkar AM (2009) Cloning of a novel omega-6 desaturase from flax (*Linum usitatissimum*) and its functional analysis in *Saccharomyces cerevisiae*. Mol Biotechnol 42:168-174

Kim MJ, Kim H, Shin JS, Chung CH, Ohlrogge JB, Suh MC (2006) Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative cis-regulatory elements in the SeFAD2 promoter and enhancers in the 5'-UTR intron. Mol Genet Genomics 276:351-68

Kinney AJ (1996) Development of genetically engineered soybean oils for food applications. J Food Lipids 3:273-292

Knutzon DS, Scherer DE, Schreckengost WE (1991) Nucleotide sequence of a complementary DNA clone encoding stearoyl-acyl carrier protein desaturase from castor bean, *Ricinus communis*. Plant Physiol 96:344-345

Knutzon DS, Thompson GA, Radke SE, Johnson WB, Knauf VC, Kridl JC (1992) Modification of Brassica seed oil by antisense expression of a stearoyl-acyl carrier protein desaturase gene. Proc Nat Acad of Sci USA 89:2624-2628

Kondra ZP, Thomas PM (1975) Inheritance of oleic, linoleic and linolenic acids in seed oil of rapeseed (*Brassica napus*). Can J Plant Sci 55:205-210

Krasowska A, Dziadkowiec D, Polinceusz A, Plonka A, Łukaszewicz M (2007) Cloning of flax oleic fatty acid desaturase and its expression in yeast. J Am Oil Chem Soc 84:809-816

Kumar S, Jordan MC, Datla R, Cloutier S (2013) The *LuWD40-1* gene encoding WD repeat protein regulates growth and pollen viability in flax (*Linum usitatissimum* L.). PLoS One 8:369124

Lands WEM (2001) Impact of daily food choices on health promotion and disease prevention. In: Hamazaki T, Okuyama H (eds) Fatty acids and lipids-new findings, S Karger AG, Basel, Switzerland, pp 1-5

Lee GJ, Wu X, Shannon JG, Sleper DA, Nguyen HT (2007) Soybean. In: Kole C (ed) Genome mapping and molecular breeding in plants, volume 2, oilseeds. Springer, Berlin, pp 1–3

Lei A, Chen H, Shen G, Hu Z, Chen L, Wang J (2012) Expression of fatty acid synthesis genes and fatty acid accumulation in *haematococcus pluvialis* under different stressors. Biotechnol Biofuels 5:18

Li LY, Wang XL, Gai JY, Yu DY (2007) Molecular cloning and characterization of a novel microsomal oleate desaturase gene from soybean. J Plant Physiol 164:1516-1526

Libus J, Storchová H (2006) Quantification of cDNA generated by reverse transcription of total RNA provides a simple alternative tool for quantitative RT-PCR normalization. BioTechniques 41:156-158

Lightner J, Wu J, Browse J (1994) A mutant of *Arabidopsis* with increased levels of stearic acid. Plant Physiol 106:1443-1451

Lin CS, Poushinsky G (1985) A modified augmented design (type 2) for rectangular plots. Can J Plant Sci 65:743-749

Livak, KJ, Schmittgen, TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta}$CT method. Methods 25:402-408

Los DA, Murata N (1998) Structure and expression of fatty acid desaturases. Biochim Biophys Acta 1394:3-15

Lu Y, Chi X, Li Z, Yang Q, Li F, Liu S, Gan Q, Qin S (2010) Isolation and characterization of a stress-dependent plastidial Δ12 fatty acid desaturase from the Antarctic microalga *Chlorella vulgaris* NJ-7. Lipids 45:179-187

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378-389

Luo T, Deng WY, Zeng J, Zhang FL (2009) Cloning and characterization of a stearoyl-acyl carrier protein desaturase gene from *Cinnamomum longepaniculatum*. Plant Mol Biol Rep 27:13-19

Mackay TFC (2001) Quantitative trait loci in *Drosophila*. Nat Rev Genet 2:11-20

Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H, Ishiguro S, Nakamura K (2009) An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J 60:476-487

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet 11:e1004915

Martinetz M (1992) Tissue levels of polyunsaturated fatty acids during early human development. J Pediatrics 120:S129-S138

Metzger JO, Bornscheuer U (2006) Lipids as renewable resources: current state of chemical and biotechnological conversion and diversification. Appl Microbiol Biotechnol 71:13-22

Mikami K, Murata N (2003) Membrane fluidity and the perception of environmental signals in cyanobacteria and plants. Prog Lipid Res 42:527-543

Mikkilineni V, Rocheford TR (2003) Sequence variation and genomic organization of fatty acid desaturase-2 (*Fad2*) and fatty acid desaturase-6 (*Fad6*) cDNAs in maize. Theor Appl Genet 106:1326-1332

Millar AA, Kunst L (1999) The natural genetic variation of the fatty-acyl composition of seed oils in different ecotypes of *Arabidopsis thaliana*. Phytochemistry 52:1029-33

Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D (2013) Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics 14:193-202

Mollers C, Schierholt A (2002) Genetic variation of palmitate and oil content in a winter oilseed rape doubled haploid population segregating for oleate content. Crop Sci 42:379-384

Muir A, Westcott N (2003) Flax: the genus Linum. Taylor & Francis, London, UK, pp 307

Murata N, Wada H (1995) Acyl-lipid desaturases and their importance in the tolerance and acclimatization to cold of cyanobacteria. Biochem J 308:1-8

Murphy DJ (1996) Engineering oil production in rapeseed and other oil crops. Trends Biotechnol 14:206-213

Ndzana X, Fehr WR, Welke GA, Hammond EG, Duvick DN, Cianzio SR (1994) Influence of reduced palmitate content on agronomic and seed traits of soybean. Crop Sci 34:646-649

Nilmalgoda SD, Cloutier S, Walichnowski AZ (2003) Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome 46:870-8

Nussbaumer T, Martis MM, Roessner SK, Pfeifer M, Bader KC, Sharma S, Gundlach H, Spannagl M (2013) MIPS PlantsDB: a database framework for comparative plant genome research. Nucleic Acids Res. 4: D1144-D1151

O'Hara P, Slabas AR, Fawcett T (2002) Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. Plant Physiol 129:310–20

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7:957-970

Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acids synthesis. Annu Rev Plant Physiol Plant Mol Biol 48:109-136

Oomah BD (2001) Flaxseed as a functional food source. J Sci Food Agric 81:889-894

Pan X, Siloto RM, Wickramarathna AD, Mietkiewska E, Weselake RJ (2013) Identification of a pair of phospholipid:diacylglycerol acyltransferases from developing flax (*Linum usitatissimum* L.) seed catalyzing the selective production of trilinolenin. J Biol Chem 1288:24173-24188

Pfaffl, MW (2005) Quantification strategies in real time PCR. In: Bustin SA (ed) A-Z of quantitative PCR. International University Line, La Jolla, CA, pp 87-112

Pham AT, Lee JD, Shannon JG, Bilyeu KD (2011) A novel FAD2-1A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. Theor Appl Genet 123:793-802

Pleines S, Friedt W (1989) Genetic control of linolenic acid concentration in seed oil of rapeseed (*Brassica napus* L.). Theor Appl Genet 78:793-797

Radovanovic N, Thambugala D, Duguid S, Loewen E, Cloutier S (2014) Functional characterization of flax fatty acid desaturase FAD2 and FAD3 isoforms expressed in yeast reveals a broad diversity in activity. Mol Biotechnol 56:609-620

Ragupathy R, Rathinavelu R, Cloutier S (2011) Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. BMC Genomics 12:217

Rajwade AV, Kadoo NY, Borikar SP, Harsulkar AM, Ghorpade PB, Gupta VS (2014) Differential transcriptional activity of SAD, FAD2 and FAD3 desaturase genes in

developing seeds of linseed contributes to varietal variation in α-linolenic acid content. Phytochemistry 98: 41-53

Reese MG (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comp Chem 26:51-56

Rowland GG (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). Can J Plant Sci 71:393-396

Rowland GG, Hormis YA, Rashid KY (2002) CDC Bethune flax. Can J Plant Sci 82:101-102

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology, Humana Press, Totowa, NJ pp 365-386

Saed Taha R, Ismail I, Zainal Z, Abdullah SN (2012) The stearoyl-acyl-carrier-protein desaturase promoter (Des) from oil palm confers fruit specific GUS expression in transgenic tomato. J Plant Physiol 169:1290-1300

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Scheffler JA, Schimdt H, Sperling P, Parkin IAP, Luhs W, Lydiate DJ, Heinz E (1997) Desaturase multigene families of *Brassica napus* arose through genome duplication. Theor Appl Genet 94:583-591

Schlueter JA, Scheffler BE, Schlueter SD, Shoemaker RC (2006) Sequence conservation of homeologous bacterial artificial chromosomes and transcription of homeologous genes in soybean (*Glycine max* L. Merr.). 174:1017-28

Schlueter JA, Vasylenko-Sanders IF, Deshpande S, Yi J, Siegfried M, Roe BA, Schlueter SD, Scheffler BE, Shoemaker RC (2007) The *FAD2* gene family of soybean: insights into the structural and functional divergence of a paleopolyploid genome. Crop Sci 47:S14-S26

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar, B (2003) Large scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res 13:1250-1257

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183

Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, Liang CZ, Zhang JW, Fulton L, Graves TA, et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326:1112-1115

Schnurr J, Shockey J, Browse J (2004) The acyl-CoA synthetase encoded by LACS2 is essential for normal cuticle development in *Arabidopsis*. Plant Cell 16:629-642

Shanklin J, Cahoon EB (1998) Desaturation and related modifications of fatty acids. Annu Rev Plant Physiol Plant Mol Biol 49:611-641

Shanklin J, Somerville C (1991) Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc Nat Acad Sci USA 88:2510-2514

Shanklin J, Whittle E, Fox BG (1994) Eight histidine residues are catalytically essential in a membrane associate diron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. Biochemistry 33:12787-12794

Shilman F, Brand Y, Brand A, Hedvat I, Hovav R (2011) Identification and molecular characterization of homologous *Δ9-stearoyl acyl carrier protein desaturase 3* genes from the allotetraploid peanut (*Arachis hypogaea*). Plant Mol Biol Rep 29:232-241

Simopoulos AP (2000) Human requirement for N-3 polyunsaturated fatty acids. Poultry Sci 79:961-970

Singh KK, Mridula D, Rehal J, Barnwal P (2011) Flaxseed: a potential source of food, feed and fiber. CRC Crit Rev Food Sci Nutr 51:210-222

Singh S, McKinney S, Green A (1994) Sequence of a cDNA from *Linum usitatissimum* encoding the Stearoyl-ACP carrier protein desaturase. Plant Physiol 140:1075

Smooker AM, Wells R, Morgan C, Beaudoin F, Cho K, Fraser F, Bancroft I (2011) The identification and mapping of candidate genes and QTL involved in the fatty acid desaturation pathway in *Brassica napus*. Theor Appl Genet 122:1075-1090

Sorensen BM, Furukawa-Stoffer TL, Marshall KS, Page EK, Mir Z, Forster RJ, Weselake RJ (2005) Storage lipid accumulation and acyltransferase action in developing flaxseed. Lipids 40:1043-1049

Soto-Cerda BJ, Duguid S, Booker H, Diederichsen A, Cloutier S (2013) Association mapping of seed quality traits using the flax (*Linum usitatissimum* L.) core collection. Theor Appl Genet (in press)

Sperling P, Ternes P, Zank TK, Heinz E (2003) The evolution of desaturases. Prostaglandins Leukot Essent Fatty Acids 68:73-95

Stoltzfus DL, Fehr WR, Welke GA (2000) Relationship of elevated palmitate to soybean seed traits. Crop Sci 40:52-54

Stymne S, Tonnet ML, Green AG (1992) Biosynthesis of linolenate in developing embryos and cell-free preparations of high-linolenate linseed (*Linum usitatissimum*) and low-linolenate mutants. Arch Biochem Biophys 294:557-563

Sveinsson S, McDill J, Wong GK, Li J, Li X, Deyholos MK, Cronk QC (2014) Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (*Linum*) using transcriptomics. Ann Bot 113:753-61

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599

Tatusova TA, Madden TL (1999) BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247-250

Thambugala D, Duguid S, Loewen E, Rowland G, Booker H, You FM, Cloutier S (2013) Genetic variation of six desaturase genes in flax and their impact on fatty acid composition. Theor Appl Genet 126:2627-2641

Thambugala D, Cloutier S (2014) Fatty acid composition and desaturase gene expression in flax (*Linum usitatissimum* L.). J Appl Genet 55:423-432

Thelen JJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Eng 4:12-21

Thomas JMG, Boote KJ, Allen LH Jr, Gallo-Meagher M, Davis JM (2003) Elevated temperature and carbon dioxide effects on soybean seed composition and transcript abundance. Crop Sci 43:1548-1557

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680

Turnham E, Northcote DH (1983) Changes in the activity of acetyl-CoA carboxylase during rape-seed formation. Biochem J 212:223-9

Vaisey-Genser M, Morris DH (2001) History of cultivation and uses of flaxseed. In: Westscott AM, Amsterdam N (eds) Flax: The genus *Linum*, Hardwood Academic Publishers, pp 1-21

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Chronica Botanica 13:1-366

Velasco L, Perez-Vich B, Fernandez-Martınez JM (2007) Relationships between seed oil content and fatty acid composition in high stearic acid sunflower. Plant Breed 126:503-508

Voelker T, Kinney AJ (2001) Variation in the biosynthesis of seed-storage lipids. Ann Rev Plant Physiol Plant Mol Biol 52:335-361

Vrinten P, Hu Z, Munchinsky MA, Rowland G and Qiu X (2005) Two FAD3 desaturase genes control the level of linolenic acid in flax seed. Plant Physiol 139:79-87

Wada H, Schmidt H, Heinz E, Murata N (1993) In vitro ferredoxin-dependent desaturation of fatty acids in cyanobacterial thylakoid membranes. J Bacteriol 175:544-547

Wakjira A, Labuschagne MT, Hugo A (2004) Variability in oil content and fatty acid composition of Ethiopian cultivars of linseed. J Sci Food Agric 84:601-607

Wang ML, Barkley NA, Chen Z, Pittman RN (2011) FAD2 gene mutations significantly alter fatty acid profiles in cultivated peanuts (*Arachis hypogaea*). Biochem Genet 49:748-759

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK, Wang J, Deyholos MK (2012) The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. Plant J 72:461-473

Warude D, Joshi K, Harsulkar A (2006) Polyunsaturated fatty acids: biotechnology. Crit Rev Biotechnol 26:83-93

Watkins BA, Devitt AA, Feng S (2001) Designed eggs containing conjugated linoleic acids and omega-3 polyunsaturated fatty acids. World Rev Nutr Diet 90:162-182

Wei S, Peng Z, Zhou Y, Yang Z, Wu K, Ouyang Z (2011) Nucleotide diversity and molecular evolution of the *WAG-2*gene in common wheat (*Triticum aestivum* L) and its relatives . Genet Mol Biol 34:606-615

Weselake RJ, Taylor DC, Rahman MH, Shah S, Laroche A, McVetty PB, Harwood JL (2009) Increasing the flow of carbon into seed oil. Biotechnol Adv 27:866-78

White Jr HB, Quackenbush FW, Probst AH (1961) Occurrence and inheritance of linolenic and linoleic acids in soybean seeds. J Am Oil Chem Soc 38:113-117

Wiesenfeld PW, Babu US, Collins TFX, Sprando R, O'Donnell MW, Flynn TJ, Black T, Olejnick N (2003) Flaxseed increased α-linolenic and eicosapentaenoic acid and decreased arachidonic acid in serum and tissues of rat dams and offspring. Food Chemical Toxicol 41:841-855

Wong ML, Medrano JF (2005) Real-time PCR for mRNA quantitation. BioTechniques 39:75-85

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35: W265-W268

Yadav NS, Wierzbicki A, Aegerter M, Caster CS, Perez-Girau L, Kinney AJ, Hitz WD, Booth JR Jr, Schweiger B, Stecca KL, Allen SM, Blackwell M, Reiter RS, Carlson TJ, Russell SH, Feldmann KA, Pierce J, Browse J (1993) Cloning of higher plant ω-3 fatty acid desaturases. Plant Physiol 103:467-476

Yang G (2013) MITE Digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements. BMC Bioinformatics14:186

Yano M, Sasaki T (1997) Genetic and molecular dissection of quantitative traits in rice. Plant Mol Biol 35:145-153

Ying JZ, Shan JX, Gao JP, Zhu MZ, Shi M, Lin HX (2012) Identification of quantitative trait loci for lipid metabolism in rice seeds. Mol Plant 5:865-875

You FM, Duguid S, Thambugala D, Cloutier S (2013) Statistical analysis and field evaluation of the type 2 modified augmented design (MAD) in phenotyping of flax (*Linum usitatissimum*) germplasm in multiple environments. Austr J Crop Sci 7:1789-1800

You FM, Li P, Kumar S, Ragupathy R, Li Z, Fu YB and Cloutier S (2014) Genome wide identification and characterization of the gene families controlling fatty acid biosynthesis in flax (*Linum usitatissimum* L). J Proteomics Bioinform 7:310-326

Zhang D, Pirtle IL, Park SJ, Nampaisansuk M, Neogi P, Wanjie SW, Pirtle RM, Chapman KD (2009). Identification and expression of a new delta-12 fatty acid

desaturase (*FAD2-4*) gene in upland cotton and its functional expression in yeast and *Arabidopsis thaliana* plants. Plant Physiol Biochem 47:462-471

Zhang JT, Liu H, Sun J, Li B, Zhu Q, Chen SL, Zhang HX (2012) *Arabidopsis* fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. PLoS One 7:e30355

Zohary D (1999) Monophyletic and polyphyletic origin of the crops on which agriculture was formed in the Near East. Genet Resour Crop Evol 46:133-142

Zohary D, Hopf M (2000) Domestication of plants in the Old World. 3rd ed. Oxford Univ Press, Oxford, UK, pp 125-13

Zuk M, Prescha A, Stryczewska M, Szopa J (2012) Engineering flax plants to increase their antioxidant capacity and improve oil composition and stability. J Agric Food Chem 60:5003-5012

APPENDICES

Appendix I List of accessions from which the six desaturase genes were sequenced.

Accession/Description	Accession number	Species/type[2]	Origin[3]	*sad* allele and isoform				*fad2* allele and isoform				*fad3* allele and isoform			
				sad1		*sad2*		*fad2a*		*fad2b*		*fad3a*		*fad3b*	
AC Watson	CN18973	*L. usitatissimum*/O	CAN	1	A	1	A	11	A	1	A	1	A	2	A
Flanders	CN18979	*L. usitatissimum*/O	CAN	1	A	3	A	3	A	1	A	3	A	1	A
Somme	CN18980	*L. usitatissimum*/O	CAN	1	A	1	A	1	A	1	A	3	A	2	A
CDC Valour	CN18981	*L. usitatissimum*/O	CAN	1	A	2	B	2	A	1	A	1	A	2	A
Evelin	CN18982	*L. usitatissimum*/F	FRA	1	A	1	A	2	A	1	A	1	A	1	A
Laura	CN18983	*L. elongatum*/F	NLD	1	A	1	A	1	A	1	A	1	A	1	A
Hermes	CN18986	*L. usitatissimum*/F	FRA	1	A	2	B	1	A	1	A	2	A	1	A
Viking	CN18987	*L. elongatum*/F	NLD	1	A	2	B	1	A	1	A	2	A	1	A
Ariane	CN18988	*L. elongatum*/F	FRA	1	A	2	B	2	A	1	A	1	A	1	A
Atalante	CN18989	*L. usitatissimum*/O	FRA	1	A	1	A	7	C	1	A	3	A	1	A
Nike	CN18991	*L. elongatum*/F	POL	1	A	1	A	1	A	1	A	1	A	1	A
Linda	CN18993	*L. usitatissimum*/O	NLD	1	A	1	A	1	A	1	A	2	A	4	D
Verne	CN18994	*L. usitatissimum*/O	USA	1	A	2	B	1	A	1	A	3	A	1	A
Raisa	CN18997	*L. elongatum*/F	NLD	1	A	1	A	2	A	1	A	1	A	4	D
Escalina	CN18998	*L. elongatum*/F	NLD	1	A	1	A	1	A	1	A	2	A	1	A
Marina	CN19001	*L. elongatum*/F	NLD	1	A	1	A	1	A	1	A	2	A	1	A
AC McDuff	CN19003	*L. usitatissimum*/O	CAN	1	A	2	B	5	C	1	A	3	A	1	A
AC Emerson	CN19004	*L. usitatissimum*/O	CAN	1	A	2	B	1	A	1	A	2	A	1	A
AC Linora	CN19005	*L. usitatissimum*/O	CAN	1	A	1	A	1	A	1	A	1	A	1	A
no name	CN19007	*L. usitatissimum*/O	ETH	1	A	1	A	12	A	1	A	1	A	15	D
CDC Normandy	CN19017	*L. usitatissimum*/O	CAN	1	A	1	A	1	A	1	A	3	A	2	A
Ottawa 829-C	CN19157	*L. elongatum*/O	CAN	1	A	1	A	8	A	2	D	1	A	1	A
Ottawa 770B	CN19158	*L. usitatissimum*/O	CAN	1	A	1	A	1	A	2	D	1	A	13	G
Diadem	CN19159	*L. usitatissimum*/O	CAN	1	A	2	B	1	A	1	A	12	A	1	A
Bolley Golden	CN19160	*L. elongatum*/O	USA	1	A	1	A	1	A	1	A	2	A	1	A
Kirovogradskij 71	CN30860	*L. usitatissimum*/O	UKR	1	A	4	A	14	A	1	A	1	A	5	D
Kubanskij	CN30861	*L. usitatissimum*/O	UNK	1	A	3	A	13	A	1	A	1	A	1	A
Vniil-17	CN32542	*L. elongatum*/F	RUS	1	A	1	A	1	A	1	A	3	A	2	A
Korostenskij 3	CN32546	*L. elongatum*/F	UKR	1	A	1	A	3	A	1	A	1	A	1	A
Linott	CN33385	*L. usitatissimum*/O	CAN	1	A	2	B	4	A	1	A	1	A	2	A
Noralta	CN33386	*L. usitatissimum*/O	CAN	1	A	1	A	5	C	1	A	1	A	1	A

Accession/Description	Accession number	Species/type[2]	Origin[3]	*sad* allele and isoform				*fad2* allele and isoform				*fad3* allele and isoform			
				sad1		*sad2*		*fad2a*		*fad2b*		*fad3a*		*fad3b*	
Redwood 65	CN33388	*L. usitatissimum*/O	CAN	1	A	1	A	5	C	1	A	3	A	1	A
Rocket	CN33389	*L. usitatissimum*/O	CAN	1	A	1	A	6	C	1	A	3	A	1	A
Natasja	CN33390	*L. elongatum*/F	NLD	1	A	1	A	3	A	1	A	2	A	1	A
Domtar Selection	CN33393	*L. elongatum*/F	UNK	1	A	1	A	4	A	1	A	1	A	1	A
Dufferin	CN33397	*L. usitatissimum*/O	CAN	1	A	1	A	1	A	1	A	3	A	1	A
Bison	CN33399	*L. usitatissimum*/O	USA	1	A	2	B	1	A	1	A	3	A	1	A
Norstar	CN33400	*L. elongatum*/O	USA	1	A	2	B	9	C	2	D	3	A	13	G
Culbert	CN33992	*L. usitatissimum*/O	USA	1	A	3	A	9	C	1	A	3	A	1	A
Tverca	CN35791	*L. elongatum*/F	RUS	1	A	1	A	1	A	1	A	1	A	1	A
McGregor	CN37286	*L. elongatum*/O	CAN	1	A	3	A	9	C	1	A	1	A	1	A
Natasja	CN40081	*L. usitatissimum*/F	NLD	1	A	1	A	1	A	1	A	2	A	1	A
Norlin	CN52732	*L. usitatissimum*/O	CAN	1	A	1	A	2	A	1	A	3	A	2	A
Clli-642	CN96845	*L. usitatissimum*/O	RUS	1	A	3	A	21	B	1	A	2	A	1	A
Clli-643	CN96846	*L. usitatissimum*/O	RUS	1	A	2	B	20	A	1	A	13	C	10	D
Clli-1407	CN96911	*L. usitatissimum*/O	TUR	1	A	1	A	4	A	1	A	1	A	11	E
Clli-1455	CN96958	/O	TUR	1	A	3	A	2	A	1	A	2	A	18	D
Clli-1458	CN96962	*L. usitatissimum*/O	TUR	1	A	2	B	16	A	1	A	2	A	10	D
Clli-1470	CN96974	/O	IND	1	A	2	B	2	A	1	A	5	A	3	D
Clli-1499	CN96988	*L. usitatissimum*/O	ETH	1	A	3	A	19	B	1	A	2	A	1	A
Clli-1502	CN96991	*L. usitatissimum*/O	ETH	1	A	3	A	18	B	4	C	2	A	6	A
Clli-1503	CN96992	*L. usitatissimum*/O	ETH	1	A	3	A	17	B	1	A	2	A	6	A
Clli-1519	CN97004	*L. usitatissimum*/O	ETH	1	A	3	A	15	B	1	A	2	A	1	A
Clli-1924	CN97050	*L. usitatissimum*/O	IRN	1	A	2	B	1	A	3	A	1	A	3	D
Clli-1930	CN97056	*L. usitatissimum*/O	PAK	1	A	2	B	2	A	3	A	2	A	3	D
Clli-1938	CN97064	*L. usitatissimum*/O	PAK	1	A	2	B	2	A	3	A	1	A	3	D
Clli-1946	CN97072	/O	PAK	1	A	3	A	1	A	1	A	2	A	3	D
Clli-1957	CN97083	/O	PAK	1	A	2	B	2	A	3	A	1	A	3	D
Clli-1991	CN97092	*L. usitatissimum*/O	PAK	1	A	5	A	1	A	3	A	1	A	3	D
Clli-1995	CN97096	*L. usitatissimum*/O	PAK	1	A	2	B	8	A	3	A	1	A	9	D
Clli-2002	CN97103	*L. usitatissimum*/O	PAK	1	A	2	B	3	A	1	A	1	A	9	D
Clli-2028	CN97129	*L. usitatissimum*/O	IRN	4	B	2	B	3	A	1	A	2	A	10	D
Clli-2028B	CN97129B	/O	IRN	4	B	2	B	4	A	1	A	2	A	10	D
Clli-2038	CN97139	*L. usitatissimum*/O	IRN	1	A	3	A	14	A	1	A	5	A	3	D
Clli-2046	CN97147	*L. usitatissimum*/O	TUR	1	A	1	A	1	A	1	A	11	A	14	G
Clli-2052	CN97153	*L. usitatissimum*/O	TUR	6	D	6	B	1	A	1	A	2	A	11	E
Horal	CN97176	*L. usitatissimum*/O	CZE	1	A	3	A	2	A	1	A	5	A	18	D

Accession/Description	Accession number	Species/type[2]	Origin[3]	*sad* allele and isoform				*fad2* allele and isoform				*fad3* allele and isoform			
				sad1		*sad2*		*fad2a*		*fad2b*		*fad3a*		*fad3b*	
Sorth Behbehan	CN97180	*L. elongatum*/F	IRN	1	A	1	A	2	A	1	A	1	A	4	D
noname	CN97214	*L. usitatissimum*/O	ARG	1	A	1	A	1	A	1	A	1	A	1	A
No. 1048	CN97238	*L. usitatissimum*/O	HUN	1	A	1	A	20	A	1	A	1	A	12	G
Lina Deta	CN97287	*L. usitatissimum*/O	HUN	1	A	1	A	16	A	1	A	10	B	4	D
Raja	CN97300	*L. elongatum*/O	HUN	1	A	2	B	2	A	5	B	3	A	12	G
N.P. (R.R.) 9	CN97306	*L. mediterraneum*/O	IND	1	A	3	A	1	A	1	A	2	A	1	A
N.P. (R.R.) 37	CN97307	*L. mediterraneum*/O	IND	1	A	3	A	1	A	1	A	2	A	5	D
N.P. (R.R.) 38	CN97308	*L. usitatissimum*/O	IND	1	A	3	A	1	A	1	A	2	A	18	D
T.126	CN97312	*L. mediterraneum*/O	IND	1	A	3	A	20	A	1	A	2	A	1	A
Clli-2528	CN97321	*L. usitatissimum*/O	ROM	1	A	1	A	20	A	1	A	1	A	1	A
Mocoreta	CN97334	*L. usitatissimum/O*	ARG	2	A	3	A	1	A	1	A	2	A	1	A
H723 F3-6-3-3-4-2-2	CN97341	/UN	ARG	5	C	2	B	1	A	5	B	2	A	1	A
de metcha 1-3-3 Vilm	CN97350	*L. usitatissimum*/O	FRA	2	A	1	A	1	A	1	A	2	A	1	A
de metcha 1-3-6 Vilm	CN97351	*L. elongatum*/F	FRA	1	A	1	A	4	A	1	A	15	F	5	D
Texas S. 4-6 Walsh x New Golden	CN97366	*L. usitatissimum*/O	USA	1	A	1	A	3	A	1	A	2	A	12	G
Reserve (N. Dak. Res. 155)	CN97377	*L. usitatissimum*/O	USA	1	A	3	A	8	A	1	A	2	A	5	D
Novelty	CN97392	*L. usitatissimum*/O	CAN	1	A	1	A	1	A	1	A	1	A	5	D
Sel. C.I. 21-2 Jalaun	CN97393	*L. usitatissimum*/O	USA	1	A	3	A	15	B	1	A	2	A	1	A
Res. x Hoshangabad (C.I. 19 x C.I.	CN97396	*L. usitatissimum*/O	USA	1	A	5	A	4	A	1	A	5	A	4	D
Sel. C.I. 19-47 Pale Blue	CN97397	*L. usitatissimum*/UN	USA	1	A	2	B	4	A	1	A	1	A	5	D
No. Dak. No. 40,013	CN97402	/UN	USA	2	A	2	B	14	A	1	A	1	A	7	A
Linota	CN97403	*L. usitatissimum*/O	USA	1	A	3	A	3	A	1	A	1	A	1	A
Buda Sel.	CN97404	*L. usitatissimum*/O	USA	1	A	2	B	2	A	1	A	1	A	7	A
Buda Sel.B	CN97404B	/O	USA	1	A	2	B	14	A	1	A	1	A	7	A
No.Dak.Res. No.52	CN97406	/UN	USA	1	A	1	A	9	C	1	A	1	A	1	A
Rio (Long 79)	CN97407	*L. usitatissimum*/O	USA	2	A	1	A	1	A	1	A	1	A	1	A
Tammes #3 White Involute	CN97424	*L. elongatum*/F	NLD	1	A	1	A	1	A	2	D	1	A	13	G
N.D. Nur. No. 1740 (G.36 a/21)	CN97430	*L. usitatissimum*/O	DEU	1	A	1	A	9	C	1	A	7	A	1	A
TMP 2998-9	CN97430B	*L. usitatissimum* /O	DEU	1	A	1	A	9	C	1	A	7	A	1	A
CDC Bethune	CDC Bethune	*L. usitatissimum* /O	CAN	1	A	2	B	1	A	1	A	8	A	1	A
FP2214	FP2214	*L. usitatissimum* /O	CAN	1	A	1	A	9	C	1	A	1	A	1	A
SP2047*	SP2047*	*L. usitatissimum* /O	CAN	1	A	3	A	10	C	1	A	6	D	16	C
FP2270	FP2270	*L. usitatissimum* /O	CAN	1	A	3	A	1	A	1	A	3	A	1	A
UGG5-5	UGG5-5	*L. usitatissimum* /O	CAN	1	A	3	A	20	A	1	A	3	A	1	A
Hanley	Hanley	*L. usitatissimum* /O	CAN	1	A	2	B	7	C	1	A	1	A	1	A
E1747*	E1747*	*L. usitatissimum* /O	CAN	1	A	3	A	7	C	1	A	4	E	17	B

Accession/Description	Accession number	Species/type[2]	Origin[3]	*sad* allele and isoform				*fad2* allele and isoform				*fad3* allele and isoform			
				sad1		*sad2*		*fad2a*		*fad2b*		*fad3a*		*fad3b*	
Lirina	Lirina	*L. usitatissimum* /O	CAN	2	A	4	A	7	C	1	A	3	A	1	A
Atlas	Atlas	*L. usitatissimum* /F	SWE	1	A	1	A	2	A	1	A	1	A	3	D
M5791	M5791	*L. usitatissimum* /O	CAN	1	A	3	A	2	A	1	A	3	A	2	A
Crepitam Tabor	Crepitam Tabor	*L. usitatissimum* /F	HUN	3	A	1	A	1	A	1	A	2	A	5	D
Prairie Blue	Prairie Blue	*L. usitatissimum* /O	CAN	1	A	3	A	6	C	1	A	3	A	1	A
Viking(European)	Viking (European)	*L. usitatissimum* /F	EU	1	A	2	B	1	A	1	A	2	A	1	A
Prairie Grande	Prairie Grande	*L. usitatissimum* /O	CAN	1	A	2	B	7	C	1	A	1	A	1	A
Double Low*	Double Low*	*L. usitatissimum* /O	CAN	1	A	1	A	1	A	1	A	6	D	4	D
Prairie Thunder	Prairie Thunder	*L. usitatissimum* /O	CAN	1	A	1	A	1	A	1	A	1	A	1	A
UGG102-2	UGG102-2	*L. usitatissimum* /O	CAN	1	A	3	A	20	A	1	A	14	C	4	D
S95407*	S95407*	*L. usitatissimum* /O	CAN	3	A	3	A	1	A	1	A	4	E	8	F
UGG146-1*	UGG146-1*	*L. usitatissimum* /O	CAN	1	A	1	A	20	A	1	A	6	D	4	D
YSED18*	YSED18*	*L. usitatissimum* /O	CAN	1	A	2	B	9	C	1	A	4	E	17	B
G-1186-94	G-1186-94	*L. usitatissimum* /F	GER	1	A	1	A	1	A	1	A	3	A	5	D
CDCMons	CDCMons	*L. usitatissimum* /O	CAN	1	A	3	A	14	A	1	A	1	A	1	A
M96006*	M96006*	*L. usitatissimum* /O	CAN	1	A	3	A	1	A	1	A	9	E	8	F
Macbeth	Macbeth	*L. usitatissimum* /O	CAN	1	A	1	A	16	A	1	A	1	A	1	A

[1] Canadian National accession number at the Plant Gene Resources of Canada (PGRC), Saskatoon, SK, Canada

[2] UN = unknown, O = oil seed/linseed type, F = fiber type

[3] Country code

*Induced mutants or derived from induced mutant line

Appendix II Description of primers used for PCR amplification of *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* and for sequencing.

Gene	GeneBank Accession Number	Primer name	Sequence (5' to 3')	Tm (°C)
sad1	AJ006957.1	SAD1-F212N17	AAGCTGCTGCCAAGATTCAT	60.0
		SAD1-R212N17	TACAAGTGGCGAGTGCTGAC	60.1
		SAD1-F49	CTCAACAACTTCTCCTCCAG	55
		SAD1-R357	CTGTGCAAACACTGACAATC	55.1
		SAD1-F626	AGGTCCTTGGAAGAATGACT	55.3
		SAD1-R958	CTGAGCTCCTTCACTTGCTC	57.9
		SAD1-F1213	TTCAGTATCTCATCGGCTCT	55
		SAD1-R1525	TCACGTGTATCCTTTCACAA	55
		SAD1-F1657	TAGGTTTGCCCGTTAGTAAG	54.8
		SAD1-R1938	GAACAAAGTCCAAAAAGCAC	55
		SAD1-F2188	GGACATGATGAGGAAGAAGA	55.1
		SAD1-R2505	AATTCTCTGCTGAAGATCCA	55
sad2	AJ006958.1	SAD2-F6M22	TACAAGTGGCGAGTGCTGAC	60.1
		SAD2-R6M22	GGTTGAATCGCTGAATTGGT	59.9
		SAD2-F49	CTCAACAACTTCTCCTCCAG	55
		SAD2-R339	AATCCAGACTCGTGAAACAC	55.1
		SAD2-F653	TGACATGATTAAGGCGTAGT	53.3
		SAD2-R1000	CCAGCACAACAAAATAGTCA	54.8
		SAD2-F1211	TTCAGTATCTCATCGGCTCT	55
		SAD2-R1559	TTAATTGCACACGACATCAC	55
		SAD2-F1805	TTGTTTGTATCGCTGTATCG	54.9
		SAD2-R2143	ACGATCTTGGTGTATGCTGT	55.6
		SAD2-F2192	GGACATGATGAGGAAGAAGA	55.1
		SAD2-R2509	AATTCTCTGCTGAAGATCCA	55
fad2a	EU660502.1	FAD2A-F139G15	TTGATGTAGGGGAAGAATCCA	59.4
		FAD2A-R139G15	TTCGAAGACCCTCACAGCTT	60.0
		F1143-139G15	ATTCGTCCCTCCTTGTTCCT	59.9
		R1382-139G15	CGAACCGGTCATATGGTCTC	60.3
		FAD2A-F1453	AGGGATATTCACCGTGTGCT	59.4
		FAD2A-R2229	GTTTTGCGATTGCATCATTG	60.1
fad2b	EU660501.1	FAD2B-F25C5	GAACGAAAGCCAAATCCAAA	60.1
		FAD2B-R25C5	GGGAGGGCATTATCCTTGTT	60.2
		F26443-25C5	GCCAGCATCGGAGAAGAATA	60.3
		R26667-25C5	TTTGTCCCCAAGCAGAAATC	60.1
fad3a	HM991829.1	F1499-395P20	CCGTTGCCTAAACTGAAACC	59.6
		R5039-395P20	AGCCTGCAGCATACATCAGA	59.6
		F2063-395P20	GATTGCTCAAGGAACCATGT	57.6
		R2194-395P20	GCAACAGCCCAGATAAAAAG	57.5
		F2604-395P20	CTCTCCCAATGTTTGCGTAT	57.7
		R2730-395P20	TTACCCACCCGAAACATATC	57.3

Gene	GeneBank Accession Number	Primer name	Sequence (5' to 3')	Tm (°C)
		F3177-395P20	TATGGTTTTACCCCAATGGA	57.7
		R3336-395P20	AAGTACATCCATCCACGACA	56.3
		F3781-395P20	CGCATTTCAACCCATACAG	57.6
		R3839-395P20	AATCATGACCGATGTCCTCT	56.9
		F4257-395P20	AGCTGGTCCCTTGAATTTCT	57.9
		R4377-395P20	GGAAAACTGCTCAGGACATT	56.8
fad3b	HM991834.1	FAD3B-F356B4	CCATCCACTTGGCATCCTAC	60.3
		FAD3B-R356B4	AACGCAACCAGAGAGCAGTT	60.1
		F16473-27L18	TGAACAATGTGATGGGACAT	57
		F16946-27L18	GACTCGGGTCGATTTATTTC	56.2
		F17485-27L18	CATTCTCGACCGAAAAGATT	56.9
		F18095-27L18	TCAAATGCCACACTATCACC	56.9
		F18661-27L18	CCCTTATATTCTCAGCCGTTA	56.2
		R16437-27L18	CATGCAACAACCCAGATAAA	57.1
		R16980-27L18	GGTAAGTACCCACCCGAAAT	57.8
		R17583-27L18	GCGGAAAACGACACCTATAC	57.2
		R18114-27L18	GGTGATAGTGTGGCATTTGA	56.9
		R18741-27L18	ATGGATCCGATCTACAAAGC	56.7

Appendix III Fatty acid composition, oil content, iodine value and stearate, oleic, and linoleic desaturation proportion of 120 flax accessions averaged from two locations (MB and SK) over three years (2009, 2010 and 2011).

Accession description	Accession no.	Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)		Iodine value (IOD)[3]		Stearate desaturation proportion (SDP)[4]		Oleic desaturation proportion (ODP)[5]		Linoleic desaturation proportion (LDP)[6]	
		Percent[1]	SE[2]	Percent	SE	Percent	SE	Percent	SE	Percent	SE	Percent	SE	IOD	SE	SDP	SE	ODP	SE	LDP	SE
AC Watson	CN18973	4.85	0.14	4.00	0.23	19.40	1.75	15.05	0.26	56.72	1.69	44.22	0.78	191.12	3.34	0.96	0.00	0.79	0.02	0.79	0.00
Flanders	CN18979	5.01	0.07	4.33	0.34	19.68	1.97	13.79	0.49	57.15	1.87	44.48	1.01	190.32	3.90	0.95	0.00	0.78	0.02	0.81	0.00
Somme	CN18980	5.60	0.21	3.23	0.18	19.06	1.67	14.22	0.24	57.96	1.38	43.32	0.91	192.62	2.58	0.97	0.00	0.79	0.02	0.80	0.00
CDC Valour	CN18981	5.19	0.10	3.18	0.17	23.05	1.97	13.06	0.08	55.62	2.10	43.05	0.91	187.93	3.83	0.97	0.00	0.75	0.02	0.81	0.01
Evelin	CN18982	4.76	0.07	4.87	0.37	19.22	1.96	18.28	0.57	52.82	1.84	39.57	0.57	186.36	3.87	0.95	0.00	0.79	0.02	0.74	0.01
Laura	CN18983	5.17	0.08	5.75	0.40	19.74	1.86	14.81	0.41	54.53	1.93	38.32	0.61	185.29	4.16	0.94	0.00	0.78	0.02	0.79	0.00
Hermes	CN18986	5.03	0.03	4.06	0.27	19.92	1.86	17.45	0.31	53.54	2.00	40.55	0.34	187.40	3.86	0.96	0.00	0.78	0.02	0.75	0.01
Viking	CN18987	4.94	0.02	3.84	0.20	18.96	1.89	16.82	0.43	55.36	1.93	38.75	0.32	190.26	3.69	0.96	0.00	0.79	0.02	0.77	0.01
Ariane	CN18988	5.41	0.08	3.35	0.24	17.78	1.89	16.75	0.27	56.70	1.86	38.98	0.33	192.63	3.55	0.96	0.00	0.80	0.02	0.77	0.00
Atalante	CN18989	6.27	0.17	4.50	0.29	16.67	1.25	12.97	0.19	59.49	1.29	42.07	0.94	192.43	2.64	0.95	0.00	0.81	0.01	0.82	0.00
Nike	CN18991	5.24	0.10	5.42	0.38	22.53	1.77	16.93	0.70	49.76	2.24	40.23	0.81	178.86	4.18	0.94	0.00	0.75	0.02	0.75	0.01
Linda	CN18993	6.14	0.08	5.40	0.27	22.09	1.98	12.68	0.33	53.63	1.94	43.03	0.49	181.24	3.94	0.94	0.00	0.75	0.02	0.81	0.00
Verne	CN18994	5.35	0.08	3.76	0.18	20.12	1.90	15.42	0.15	55.21	1.99	43.23	0.60	188.45	3.74	0.96	0.00	0.78	0.02	0.78	0.01
Raisa	CN18997	5.31	0.06	3.54	0.12	22.95	1.81	16.02	0.36	52.30	1.66	41.11	0.29	184.29	3.20	0.96	0.00	0.75	0.02	0.77	0.00
Escalina	CN18998	5.19	0.06	5.34	0.39	17.73	1.68	16.98	0.52	54.71	1.73	39.27	0.50	187.77	3.71	0.94	0.00	0.80	0.02	0.76	0.00
Marina	CN19001	5.13	0.10	5.23	0.32	19.55	1.70	14.91	0.46	55.05	1.61	39.55	0.69	186.65	3.51	0.94	0.00	0.78	0.02	0.79	0.00
AC McDuff	CN19003	6.01	0.12	4.11	0.26	18.41	1.51	16.67	0.41	54.75	1.49	46.11	0.93	187.93	3.03	0.96	0.00	0.79	0.02	0.77	0.00
AC Emerson	CN19004	5.63	0.11	2.65	0.13	17.43	1.62	14.77	0.14	59.55	1.60	43.68	0.72	196.35	2.98	0.97	0.00	0.81	0.02	0.80	0.00
AC Linora	CN19005	5.73	0.13	2.85	0.15	18.25	1.76	16.60	0.35	56.55	1.77	44.63	0.91	192.39	3.31	0.97	0.00	0.80	0.02	0.77	0.01
no name	CN19007	5.68	0.22	4.54	0.34	22.18	2.05	13.84	0.45	53.75	2.07	41.76	1.40	183.67	4.18	0.95	0.00	0.75	0.02	0.79	0.00
CDC Normandy	CN19017	5.42	0.16	2.99	0.15	23.66	2.71	12.47	0.19	55.36	2.61	42.85	0.86	186.76	4.83	0.97	0.00	0.74	0.03	0.82	0.00
Ottawa 829-C	CN19157	4.92	0.05	2.80	0.08	15.12	1.00	14.46	0.33	62.81	0.67	40.53	0.40	202.35	1.44	0.97	0.00	0.84	0.01	0.81	0.00
Ottawa 770B	CN19158	5.65	0.06	4.23	0.26	15.65	1.28	14.68	0.24	59.99	1.34	42.58	0.45	195.81	2.61	0.96	0.00	0.83	0.01	0.80	0.00

Accession description	Accession no.	Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)		Iodine value (IOD)[3]		Stearate desaturation proportion (SDP)[4]		Oleic desaturation proportion (ODP)[5]		Linoleic desaturation proportion (LDP)[6]	
		Percent[1]	SE[2]	Percent	SE	Percent	SE	Percent	SE	Percent	SE	Percent	SE	IOD	SE	SDP	SE	ODP	SE	LDP	SE
Diadem	CN19159	5.52	0.06	4.19	0.21	21.36	1.67	14.91	0.35	54.03	1.51	43.69	0.59	185.53	3.03	0.96	0.00	0.76	0.02	0.78	0.00
Bolley Golden	CN19160	6.12	0.15	3.78	0.25	17.10	1.71	13.06	0.30	59.85	1.94	34.10	8.65	193.88	3.76	0.96	0.00	0.81	0.02	0.82	0.01
Kirovogradskij 71	CN30860	5.81	0.14	3.92	0.27	21.01	1.87	14.56	0.27	54.58	2.05	43.68	0.66	186.06	4.11	0.96	0.00	0.77	0.02	0.79	0.00
Kubanskij	CN30861	5.35	0.11	3.23	0.12	18.90	2.28	13.10	0.26	59.28	2.21	45.49	1.16	194.01	4.23	0.97	0.00	0.79	0.03	0.82	0.00
Vniil-17	CN32542	5.49	0.21	2.85	0.19	23.44	2.35	12.23	0.26	55.92	2.18	42.56	0.77	187.63	3.99	0.97	0.00	0.74	0.03	0.82	0.00
Korostenskij 3	CN32546	4.62	0.06	4.33	0.25	20.16	1.84	15.84	0.50	55.10	1.54	40.08	0.48	188.93	3.19	0.95	0.00	0.78	0.02	0.78	0.00
Linott	CN33385	4.99	0.04	2.63	0.11	22.57	1.94	14.21	0.31	55.72	1.79	43.06	0.42	189.79	3.28	0.97	0.00	0.76	0.02	0.80	0.01
Noralta	CN33386	6.35	0.15	3.99	0.28	18.14	1.59	15.26	0.28	56.28	1.70	42.02	0.73	189.25	3.49	0.96	0.00	0.80	0.02	0.79	0.00
Redwood 65	CN33388	6.27	0.19	3.16	0.18	19.19	1.50	16.14	0.28	55.31	1.38	43.62	0.97	189.14	2.59	0.97	0.00	0.79	0.02	0.77	0.00
Rocket	CN33389	5.95	0.13	4.16	0.30	18.80	1.63	12.38	0.20	58.76	1.91	42.91	0.65	191.32	3.52	0.96	0.00	0.79	0.02	0.83	0.01
Natasja	CN33390	4.87	0.07	5.33	0.49	18.73	1.63	17.47	0.43	53.55	1.79	38.71	0.68	186.45	3.88	0.94	0.01	0.79	0.02	0.75	0.00
Domtar Selection	CN33393	4.73	0.09	4.85	0.33	18.15	1.58	15.68	0.54	56.67	1.57	39.46	0.50	191.01	3.28	0.95	0.00	0.80	0.02	0.78	0.01
Dufferin	CN33397	5.31	0.10	4.34	0.32	19.66	1.90	14.74	0.40	55.88	1.97	43.47	1.14	188.62	4.06	0.95	0.00	0.78	0.02	0.79	0.00
Bison	CN33399	5.62	0.06	3.17	0.12	24.74	2.45	14.44	0.36	52.12	2.26	42.93	0.50	182.64	4.22	0.97	0.00	0.73	0.03	0.78	0.01
Norstar	CN33400	6.24	0.07	2.94	0.17	20.94	1.90	16.56	0.30	53.25	2.13	43.10	0.56	185.99	3.90	0.97	0.00	0.77	0.02	0.76	0.01
Culbert	CN33992	4.48	0.06	3.49	0.16	16.72	1.51	15.40	0.19	59.88	1.64	43.91	0.34	197.69	3.10	0.96	0.00	0.82	0.02	0.80	0.00
Tverca	CN35791	5.03	0.04	3.66	0.20	22.22	1.69	15.24	0.55	53.98	1.46	42.53	0.17	186.70	3.00	0.96	0.00	0.76	0.02	0.78	0.01
McGregor	CN37286	6.08	0.19	4.27	0.27	17.13	1.36	17.39	0.51	55.38	1.07	42.26	0.98	189.73	2.31	0.95	0.00	0.81	0.02	0.76	0.00
Natasja	CN40081	5.03	0.09	5.38	0.49	19.60	1.99	17.14	0.41	52.91	2.08	38.45	0.76	184.96	4.33	0.94	0.01	0.78	0.02	0.75	0.00
Norlin	CN52732	5.32	0.13	3.05	0.15	24.27	2.11	12.38	0.19	54.89	1.92	43.11	0.74	185.90	3.53	0.97	0.00	0.73	0.02	0.82	0.00
Clli-642	CN96845	5.73	0.11	3.92	0.17	14.68	1.40	12.96	0.46	62.85	1.62	41.60	0.77	199.50	2.72	0.96	0.00	0.84	0.02	0.83	0.01
Clli-643	CN96846	5.74	0.09	3.43	0.17	19.98	1.61	11.71	0.37	59.23	1.39	43.38	0.63	192.40	2.84	0.96	0.00	0.78	0.02	0.84	0.00
Clli-1407	CN96911	6.44	0.21	4.44	0.23	22.02	1.70	13.76	0.32	53.39	1.64	43.37	0.97	182.44	3.36	0.95	0.00	0.75	0.02	0.79	0.00
Clli-1455	CN96958	6.00	0.21	4.63	0.40	25.22	1.98	9.22	1.07	54.95	1.93	46.28	1.70	181.40	4.37	0.95	0.00	0.72	0.02	0.86	0.01
Clli-1458	CN96962	6.14	0.14	4.31	0.33	23.36	1.81	12.03	0.54	54.24	1.60	42.87	0.85	182.81	3.52	0.95	0.00	0.74	0.02	0.82	0.00

Accession description	Accession no.	Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)		Iodine value (IOD)[3]		Stearate desaturation proportion (SDP)[4]		Oleic desaturation proportion (ODP)[5]		Linoleic desaturation proportion (LDP)[6]	
		Percent[1]	SE[2]	Percent	SE	Percent	SE	Percent	SE	Percent	SE	Percent	SE	IOD	SE	SDP	SE	ODP	SE	LDP	SE
Clli-1470	CN96974	5.94	0.11	5.23	0.23	40.06	2.74	5.73	0.61	43.02	2.40	43.56	0.89	156.93	4.98	0.94	0.00	0.55	0.03	0.88	0.00
Clli-1499	CN96988	5.80	0.17	4.11	0.24	15.32	1.72	12.72	0.22	62.06	2.02	42.64	0.58	197.56	3.88	0.96	0.00	0.83	0.02	0.83	0.00
Clli-1502	CN96991	6.01	0.22	4.73	0.16	20.79	2.32	12.29	0.51	56.13	2.36	38.52	0.99	185.99	3.97	0.95	0.00	0.77	0.03	0.82	0.01
Clli-1503	CN96992	6.31	0.18	5.61	0.42	19.64	1.73	12.87	0.85	55.57	1.50	36.71	2.01	184.55	3.87	0.94	0.00	0.78	0.02	0.81	0.01
Clli-1519	CN97004	5.80	0.16	4.88	0.25	22.10	0.83	13.81	0.28	53.53	0.53	39.41	1.45	182.96	0.88	0.95	0.00	0.75	0.01	0.79	0.00
Clli-1924	CN97050	5.64	0.12	5.21	0.25	24.60	1.97	9.97	0.55	54.54	1.73	43.11	0.57	181.09	3.76	0.94	0.00	0.72	0.02	0.85	0.00
Clli-1930	CN97056	5.83	0.14	5.26	0.28	26.59	2.17	9.57	0.51	52.71	2.00	44.13	0.97	177.33	4.23	0.94	0.00	0.70	0.03	0.85	0.00
Clli-1938	CN97064	5.72	0.15	5.49	0.33	27.00	1.89	9.60	0.43	52.22	1.93	44.42	1.01	176.45	4.12	0.94	0.00	0.70	0.02	0.84	0.00
Clli-1946	CN97072	5.97	0.17	10.47	0.84	21.39	1.53	11.00	0.44	51.20	1.99	43.54	1.23	171.38	4.64	0.89	0.01	0.74	0.02	0.82	0.00
Clli-1957	CN97083	5.45	0.11	5.17	0.32	26.34	2.27	8.64	0.48	54.59	2.12	43.93	0.71	180.42	4.39	0.95	0.00	0.71	0.03	0.86	0.00
Clli-1991	CN97092	6.51	0.24	5.10	0.33	21.91	1.81	14.12	0.24	52.38	1.82	45.76	1.27	180.33	3.56	0.95	0.00	0.75	0.02	0.79	0.00
Clli-1995	CN97096	6.08	0.13	4.44	0.25	19.90	1.97	13.66	0.61	55.92	1.95	46.16	1.33	187.05	3.88	0.95	0.00	0.78	0.02	0.80	0.01
Clli-2002	CN97103	5.73	0.16	4.18	0.32	22.81	1.81	15.26	0.44	52.01	1.80	43.03	0.62	182.10	3.64	0.96	0.00	0.75	0.02	0.77	0.01
Clli-2028	CN97129	5.92	0.16	4.09	0.29	21.84	2.42	12.37	0.46	55.62	2.82	43.28	0.74	185.70	5.22	0.96	0.00	0.76	0.03	0.82	0.01
Clli-2028B	CN97129B	6.16	0.16	4.00	0.23	19.60	1.91	11.97	0.43	58.40	1.84	42.81	0.92	190.36	3.63	0.96	0.00	0.78	0.02	0.83	0.00
Clli-2038	CN97139	5.72	0.11	5.78	0.38	22.22	1.66	12.09	0.35	54.16	1.78	43.38	0.81	181.72	3.75	0.94	0.00	0.75	0.02	0.82	0.00
Clli-2046	CN97147	6.03	0.07	4.16	0.20	20.86	1.70	13.37	0.31	55.60	1.60	43.74	0.47	186.55	3.25	0.96	0.00	0.77	0.02	0.81	0.00
Clli-2052	CN97153	6.31	0.17	4.60	0.33	20.56	1.69	13.27	0.17	55.22	1.94	43.70	0.96	185.11	3.83	0.95	0.00	0.77	0.02	0.81	0.00
Horal	CN97176	6.00	0.07	5.57	0.26	30.08	2.39	9.44	0.41	48.93	2.38	43.58	0.75	170.22	4.69	0.94	0.00	0.66	0.03	0.84	0.01
Sorth	CN97180	5.19	0.15	4.10	0.14	19.08	1.35	13.94	0.31	57.71	1.15	43.70	1.07	191.52	2.27	0.96	0.00	0.79	0.02	0.81	0.00
Behbehan noname	CN97214	6.07	0.15	3.29	0.23	17.52	1.46	12.07	0.29	61.02	1.58	43.01	1.23	195.58	3.29	0.96	0.00	0.81	0.02	0.83	0.00
No. 1048	CN97238	6.03	0.15	4.95	0.27	25.38	2.05	13.87	0.44	49.81	1.97	42.34	1.11	176.14	4.12	0.95	0.00	0.71	0.02	0.78	0.00
Lina Deta	CN97287	5.89	0.09	4.06	0.28	23.77	2.24	12.53	0.38	53.73	2.13	42.85	1.06	182.69	4.30	0.96	0.00	0.74	0.03	0.81	0.00
Raja	CN97300	5.41	0.14	4.32	0.21	21.12	1.47	15.45	0.67	53.76	1.56	41.15	0.57	185.54	3.09	0.95	0.00	0.77	0.02	0.78	0.01
N.P. (R.R.) 9	CN97306	5.95	0.15	6.00	0.28	23.50	1.53	12.56	0.28	51.97	1.69	46.29	0.95	177.93	3.52	0.94	0.00	0.73	0.02	0.81	0.00
N.P. (R.R.)	CN97307	5.98	0.30	5.34	0.38	23.56	2.30	11.11	0.69	54.11	2.23	44.33	1.08	181.05	4.93	0.94	0.00	0.73	0.03	0.83	0.00

Accession description	Accession no.	Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)		Iodine value (IOD)[3]		Stearate desaturation proportion (SDP)[4]		Oleic desaturation proportion (ODP)[5]		Linoleic desaturation proportion (LDP)[6]	
		Percent[1]	SE[2]	Percent	SE	Percent	SE	Percent	SE	Percent	SE	Percent	SE	IOD	SE	SDP	SE	ODP	SE	LDP	SE
37																					
N.P. (R.R.) 38	CN97308	5.67	0.26	5.08	0.86	24.72	1.41	10.39	0.67	54.17	1.58	44.70	1.97	180.96	3.45	0.95	0.01	0.72	0.02	0.84	0.01
T.126	CN97312	6.24	0.22	7.77	0.51	28.98	2.14	11.45	0.30	45.55	2.42	42.23	0.83	163.91	5.04	0.92	0.01	0.66	0.03	0.80	0.00
Clli-2528	CN97321	5.86	0.10	4.44	0.24	21.59	1.90	13.40	0.34	54.73	1.80	43.15	0.92	184.96	3.63	0.95	0.00	0.76	0.02	0.80	0.00
Mocoreta	CN97334	5.40	0.17	4.39	0.21	21.76	1.95	14.86	0.40	53.67	1.68	44.18	0.91	184.83	3.35	0.95	0.00	0.76	0.02	0.78	0.00
H723 F3-6-3-3-4-2-2	CN97341	5.40	0.09	6.62	0.35	18.59	1.03	13.12	0.45	56.28	1.01	43.44	0.31	185.93	2.44	0.93	0.00	0.79	0.01	0.81	0.00
de metcha 1-3-3 Vilm	CN97350	5.81	0.09	3.73	0.17	19.11	1.49	14.81	0.41	56.40	1.46	41.51	0.81	189.63	2.93	0.96	0.00	0.79	0.02	0.79	0.00
de metcha 1-3-6 Vilm	CN97351	5.71	0.16	5.61	0.45	24.18	2.33	11.76	0.60	52.70	2.25	39.68	0.73	179.03	4.69	0.94	0.00	0.73	0.03	0.82	0.01
Texas S. 4-6 Walsh x New Golden	CN97366	5.79	0.10	3.02	0.18	19.22	1.76	14.69	0.25	57.29	1.75	44.84	0.79	191.83	3.24	0.97	0.00	0.79	0.02	0.80	0.00
Reserve (N. Dak. Res. 155)	CN97377	5.56	0.07	3.47	0.17	18.15	1.65	13.24	0.32	59.59	1.53	42.71	0.60	194.42	3.10	0.96	0.00	0.80	0.02	0.82	0.00
Novelty	CN97392	5.13	0.04	3.52	0.11	18.65	1.83	15.95	0.47	56.62	1.51	41.94	0.50	191.77	3.16	0.96	0.00	0.80	0.02	0.78	0.00
Sel. C.I. 21-2 Jalaun	CN97393	5.84	0.08	3.93	0.18	14.88	1.23	12.80	0.33	62.65	1.41	40.49	0.60	198.87	2.49	0.96	0.00	0.84	0.01	0.83	0.01
Res. x Hoshangabad (C.I. 19 x C.I.	CN97396	5.84	0.06	3.94	0.11	24.70	1.93	13.38	0.24	52.14	1.82	46.28	0.77	180.81	3.48	0.96	0.00	0.73	0.02	0.80	0.00
Sel. C.I. 19-47 Pale Blue	CN97397	5.85	0.05	3.10	0.15	17.47	1.41	16.68	0.44	56.93	1.14	42.45	0.40	192.85	2.47	0.97	0.00	0.81	0.02	0.77	0.00
No. Dak. No. 40,013	CN97402	5.65	0.05	2.32	0.07	21.09	1.88	13.38	0.49	57.58	1.54	41.09	0.37	191.93	3.09	0.98	0.00	0.77	0.02	0.81	0.00
Linota	CN97403	4.69	0.05	2.84	0.06	18.47	1.37	17.34	0.50	56.59	1.17	39.83	0.22	193.95	2.32	0.97	0.00	0.80	0.01	0.77	0.01
Buda Sel.	CN97404	5.67	0.13	2.65	0.24	22.84	3.11	13.42	0.33	55.42	2.91	42.07	1.18	187.87	5.44	0.97	0.00	0.75	0.03	0.80	0.01
Buda Sel.B	CN97404B	5.65	0.06	2.41	0.14	23.56	2.58	13.57	0.42	54.83	2.33	42.03	0.83	187.21	4.49	0.97	0.00	0.74	0.03	0.80	0.00
No.Dak.Res. No.52	CN97406	4.74	0.09	3.02	0.12	17.75	1.52	14.99	0.14	59.38	1.48	42.06	0.64	196.55	2.75	0.97	0.00	0.81	0.02	0.80	0.00
Rio (Long 79)	CN97407	6.03	0.22	4.76	0.29	20.99	1.61	13.96	0.52	54.28	1.32	42.69	1.01	184.22	2.92	0.95	0.00	0.76	0.02	0.80	0.00

Accession description	Accession no.	Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)		Iodine value (IOD)[3]		Stearate desaturation proportion (SDP)[4]		Oleic desaturation proportion (ODP)[5]		Linoleic desaturation proportion (LDP)[6]	
		Percent[1]	SE[2]	Percent	SE	Percent	SE	Percent	SE	Percent	SE	Percent	SE	IOD	SE	SDP	SE	ODP	SE	LDP	SE
Tammes #3 White Involute	CN97424	5.00	0.06	3.36	0.12	16.01	1.23	15.50	0.26	60.23	1.14	40.84	0.19	198.17	2.20	0.96	0.00	0.83	0.01	0.80	0.00
N.D. Nur. No. 1740 (G.36 a/21)	CN97430	8.62	0.88	4.61	0.24	20.07	1.28	12.25	0.74	54.39	1.34	44.71	0.88	180.75	3.52	0.95	0.00	0.77	0.02	0.82	0.01
TMP 2998-9	CN97430B	8.51	0.66	4.74	0.25	20.43	1.78	11.40	0.50	54.85	1.93	44.03	0.93	180.80	3.31	0.95	0.00	0.76	0.02	0.83	0.01
CDC Bethune	CDCBethune	5.14	0.13	3.80	0.25	21.29	1.96	14.98	0.36	54.79	1.86	44.40	0.75	187.59	3.55	0.96	0.00	0.77	0.02	0.78	0.00
FP2214	FP2214	4.96	0.13	3.90	0.25	18.36	2.69	14.74	0.69	57.82	2.34	46.45	0.98	192.57	4.81	0.96	0.00	0.80	0.03	0.80	0.00
SP2047*	SP2047	5.94	0.32	3.71	0.20	17.08	1.76	55.49	12.77	17.77	11.49	48.46	1.19	157.28	9.27	0.96	0.00	0.81	0.02	0.25	0.17
FP2270	FP2270	5.06	0.10	5.47	0.55	17.69	2.35	12.27	0.33	59.60	2.47	42.71	1.08	192.39	5.00	0.94	0.01	0.80	0.03	0.83	0.00
UGG5-5	UGG5-5	4.00	0.10	2.48	0.14	12.96	1.80	11.17	0.32	69.31	1.71	44.54	0.91	211.79	3.42	0.97	0.00	0.86	0.02	0.86	0.00
Hanley	Hanley	5.74	0.10	2.83	0.12	16.52	1.47	17.16	0.30	57.82	1.42	43.04	0.91	195.18	2.60	0.97	0.00	0.82	0.02	0.77	0.00
E1747*	E1747	6.10	0.21	4.08	0.38	16.21	1.68	52.23	3.96	21.38	5.66	41.53	1.24	160.33	6.87	0.96	0.00	0.82	0.02	0.29	0.07
Lirina	Lirina	5.39	0.08	3.78	0.28	19.63	2.75	13.77	0.30	57.36	3.26	48.61	1.32	190.79	5.78	0.96	0.00	0.78	0.03	0.81	0.01
Atlas	Atlas	4.92	0.10	3.85	0.22	20.00	2.88	14.18	0.43	56.97	2.75	42.25	0.45	190.79	5.42	0.96	0.00	0.78	0.03	0.80	0.00
M5791	M5791	4.40	0.06	2.44	0.12	11.41	1.44	9.89	0.28	71.85	1.43	43.73	0.86	214.90	2.69	0.97	0.00	0.88	0.02	0.88	0.00
Crepitam Tabor	Crepitam Tabor	5.40	0.10	2.77	0.11	19.93	2.72	14.45	0.33	57.49	2.44	44.47	0.61	192.56	4.58	0.97	0.00	0.78	0.03	0.80	0.00
Prairie Blue	Prairie Blue	5.02	0.10	4.24	0.31	19.29	1.91	13.30	0.40	58.20	1.91	45.01	0.98	191.87	3.74	0.96	0.00	0.79	0.02	0.81	0.00
Viking(European)	Viking	4.96	0.06	3.80	0.26	18.34	2.49	16.38	0.51	56.65	2.18	38.83	0.46	192.35	4.44	0.96	0.00	0.80	0.03	0.78	0.00
Prairie Grande	Prairie Grande	4.69	0.07	3.92	0.21	19.89	1.61	14.55	0.28	56.90	1.57	45.23	0.52	191.15	3.16	0.96	0.00	0.78	0.02	0.80	0.00
Double Low*	Double Low	5.79	0.19	4.43	0.35	19.50	2.46	25.13	0.83	45.23	2.79	45.55	1.51	178.61	4.87	0.95	0.00	0.78	0.03	0.64	0.02
Prairie Thunder	Prairie Thunder	5.25	0.21	4.09	0.27	18.04	1.61	15.37	0.99	57.25	1.08	44.13	0.87	191.89	2.35	0.96	0.00	0.80	0.02	0.79	0.01
UGG102-2*	UGG102-2	6.18	0.11	4.38	0.35	18.48	2.56	13.80	0.66	56.93	2.32	47.57	0.91	188.72	4.85	0.95	0.00	0.79	0.03	0.80	0.00
S95407*	S95407	6.33	0.15	4.29	0.38	17.62	1.80	67.50	1.68	4.22	0.62	45.57	0.86	143.10	2.95	0.95	0.00	0.80	0.02	0.06	0.01
UGG146-1*	UGG146-1	6.13	0.34	4.72	0.27	19.47	1.89	25.08	0.21	44.59	1.77	48.46	0.42	176.83	3.11	0.95	0.00	0.78	0.02	0.64	0.01

Accession description	Accession no.	Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)		Iodine value (IOD)[3]		Stearate desaturation proportion (SDP)[4]		Oleic desaturation proportion (ODP)[5]		Linoleic desaturation proportion (LDP)[6]	
		Percent[1]	SE[2]	Percent	SE	Percent	SE	Percent	SE	Percent	SE	Percent	SE	IOD	SE	SDP	SE	ODP	SE	LDP	SE
YSED18*	YSED18	6.44	0.24	3.65	0.33	17.34	1.69	64.82	4.02	7.58	2.59	43.56	0.69	147.00	2.11	0.96	0.00	0.81	0.02	0.11	0.04
G-1186-94	G-1186-94	5.95	0.14	3.38	0.17	13.48	1.48	12.12	0.66	65.17	1.58	42.97	0.92	203.07	2.64	0.96	0.00	0.85	0.02	0.84	0.01
CDCMons	CDCMons	5.66	0.25	3.64	0.18	17.58	1.42	15.39	0.35	57.84	1.20	43.68	1.41	193.07	2.23	0.96	0.00	0.81	0.02	0.79	0.00
M96006*	M96006	12.50	1.92	3.63	0.27	15.33	1.62	34.12	3.12	34.02	4.41	42.04	1.21	161.27	7.27	0.96	0.00	0.82	0.02	0.50	0.06
Macbeth	Macbeth	4.93	0.16	3.87	0.20	17.78	1.53	16.11	0.55	57.40	1.31	46.32	0.82	193.34	2.49	0.96	0.00	0.81	0.02	0.78	0.01

[1]Expressed as a percentage of the total fatty acid composition
[2]Standard error
[3]IOD=(0.86×OLE)+(1.732×LIO)+(2.616×LIN); Reference: AOCS Method Cd 1c-85
[4]SDP=(%OLE+%LIO+%LIN)/(%STE+%OLE+%LIO+%LIN)
[5]ODP=(%LIO+%LIN)/(%OLE+%LIO+%LIN)
[6]LDP=%LIN/(%LIO+%LIN)

a

```
sad1-a          ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTC 60
sad2-a          ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTC 60
                ************************************ ** ******************

sad1-a          TCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCCACC 120
sad2-a          TCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCCACC 120
                ************************************************** ******

sad1-a          AAGTAAG----------CATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGA 170
sad2-a          AAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCG-TTTCATTTAAGCGA 179
                *******          ********* ****************** **** *********

sad1-a          TTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTC 230
sad2-a          TTGATCGTAGA-AAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCGTTTC 238
                *********** **** ************* *******  ******** ***********

sad1-a          TTCTTCTGC---------GCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACT 281
sad2-a          TTTTTCTTTCTTCAGACTGCCTCGTCTGCATTATGTTATTGTTCGTTT-CCGATTTGACT 297
                ** ****           **** *  ********** *********** ***********

sad1-a          GACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATT 341
sad2-a          AACCTACATAA-TCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAATTGATT 356
                 *** ****** ********************** ***** ******************

sad1-a          GTCAGTGTTTGCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACC 401
sad2-a          GTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACC 416
                ***** ****************  ****** ****************************

sad1-a          ATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATC 461
sad2-a          -TACCAATTTCAGCGTCTTTCTCTGGAA-----ATTTCTG----------TCTCTGTATC 460
                 *  *********  ** ******* **      * ****          **********

sad1-a          TACTATCCTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCT 521
sad2-a          TACTATCCTATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCT 520
                *********** *********  ********* ***** **** *** ****** *****

sad1-a          TACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTT 581
sad2-a          TACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTT 580
                **** ***** ********************** ** * **********************

sad1-a          GTCTATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAAT 641
sad2-a          TTCTATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTT-AAGGTCCTGGGAAGAAT 639
                 ****************************************** ******** ********

sad1-a          GACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATA 701
sad2-a          GATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATA 699
                ** ********************* * **********************************

sad1-a          TAGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACAT 761
sad2-a          TAGGGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACAT 759
                ***************** ********* ********************************

sad1-a          GGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG 821
sad2-a          GGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG 819
                ************************************************************

sad1-a          ATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTT 881
sad2-a          ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTT 879
                ** ******** ** *********** ***** ***** ** ************** ***

sad1-a          GAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCGAGGAG 941
sad2-a          GAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAGGAG 939
                ******************* ***************** ********************

sad1-a          CAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTG 1001
sad2-a          CAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTG 999
                ************************** ******** ************************
```

```
sad1-a          GTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCAACACCCTT 1061
sad2-a          GTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAACACCCTT 1059
                ***************************** ****************************

sad1-a          GACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCG 1121
sad2-a          GACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCG 1119
                ************************************************************

sad1-a          TGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTATACCTCTCTGGA 1181
sad2-a          TGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTACCTCTCTGGA 1179
                ********************************************* ************

sad1-a          AGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTA 1241
sad2-a          AGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTA 1239
                ************************************************************

sad1-a          TGTAATCACATACT----------TCATCCTTTTCTATTAATCTTTGCGTGAACAAAATT 1291
sad2-a          TGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGATTGAACAAAATT 1299
                **** ****** **          * ********* ***********  ***********

sad1-a          CACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTA-CTGCCTAGCTTCTATGA 1350
sad2-a          CAATAAACTGGTAGC---TGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGA 1356
                ** ** *********   ****************** ** ** *****************

sad1-a          AACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTTTTCTTA 1410
sad2-a          GA--AAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAAAGTTTC---A 1411
                 *  *******  ************ ********* *** ** *** **** ***    *

sad1-a          TATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTC 1470
sad2-a          TATACCATCTTCCATCTATTTTACATGACATACCAACTTCTACTTCGGAGAAAATTCGCC 1471
                ******* **** * * * **** ************** *** *****  *******  *

sad1-a          --------------CGTCGAAGATATAATACTTGG---ATTGGTTAAATGAATTGTGAAA 1513
sad2-a          GTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATGAACTGTTAAA 1531
                              * ********* ***** *   ******** ***** *** ***

sad1-a          GGATACACGTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGA 1573
sad2-a          CAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGA 1591
                  ***** ******* ** ** *********** * ****** ** ********** ***

sad1-a          CAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTC 1633
sad2-a          CAAATGTGAACTGGAACC---CTA-GTAAA-------------CTATGAATT-GAGGTTG 1633
                *** **************   *** *****             ***  **** ******

sad1-a          TCCTTGATCACCTTAGTCCATCTTAGGTTTG----CCCGTTAGTAAGATCTGCATTTAGC 1689
sad2-a          TCCTTCATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGC 1693
                ***** ***** *** **  ** * ******    ** *** * *********** ****

sad1-a          AGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAA 1749
sad2-a          AGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAA 1753
                ****************** ** ** * ********************************

sad1-a          CCATCGGAC-AAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGT 1808
sad2-a          ACATCGGACCAAGTATCTGGTT-AGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGT 1812
                 ******** *****  ***** ************************************

sad1-a          ATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG 1868
sad2-a          ATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG 1872
                ***** ***** ************************************************

sad1-a          CAACGGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTG 1928
sad2-a          CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTG 1932
                ******************************* ***************************

sad1-a          GACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCA 1988
sad2-a          GACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCA 1992
                *************** ******************************************

sad1-a          TCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCG 2048
sad2-a          TCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCG 2052
                ******************************************* ** **********
```

```
sad1-a          CCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGAGA 2108
sad2-a          CCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGAGA 2112
                ********** ************************************************

sad1-a          AACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACG 2168
sad2-a          AGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACG 2172
                * *********** ************************************************

sad1-a          GTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCACTTGA 2228
sad2-a          GTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCACACTTGA 2232
                ******** **** ************************************ *******

sad1-a          TGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCG 2288
sad2-a          TGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCG 2292
                **************************************** ****************

sad1-a          GGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAGGTGGAAAG 2348
sad2-a          GGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGGGGAGGTGGAAAG 2352
                ************************ **********************************

sad1-a          TGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGC 2408
sad2-a          TGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGC 2412
                *************** ***************************** ******** ****

sad1-a          TTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAAGCAAACGTCGA 2468
sad2-a          TTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGGCAAAGCAAACGTCGA 2472
                **** ************** ****************************************

sad1-a          AATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA 2515
sad2-a          AATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA 2519
                ********** ***********************************
```

b

```
SAD1-A          MALKLNPVTTFPSTRSLNNFSSRSPRTFLMAASTFNSTSTKEAEKLKKSHGPPKEVHMQV 60
SAD2-A          MALKLNPVTTFPSTRSLNNFSSRSPRTFLMAASTFNSTSTKEAEKLKKSHGPPKEVHMQV 60
                ************************************************************

SAD1-A          THSMPPQKLEIFKSLEGWAEDVLLPHLKPVEKCWQPQDFLPEPESDGFEEQVKELRARAK 120
SAD2-A          THSMPPQKLEIFKSLEGWAEDVLLPHLKPVEKCWQPQDFLPEPESDGFEEQVKELRARAK 120
                ************************************************************

SAD1-A          ELPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTPWAIWTRAWTAEENRHGD 180
SAD2-A          ELPDDYFVVLVGDMITEEALPTYQTMLNTLDGVRDETGASLTPWAIWTRAWTAEENRHGD 180
                ************************************************************

SAD1-A          LLNKYLYLSGRVDMRQIEKTIQYLIGSGMDPKTENNPYLGFIYTSFQERATFISHGNTAR 240
SAD2-A          LLNKYLYLSGRVDMRQIEKTIQYLIGSGMDPKTENNPYLGFIYTSFQERATFISHGNTAR 240
                ************************************************************

SAD1-A          LAKDHGDMKLAQICGIIAADEKRHETAYTKIVEKLFEIDPDGTVLALADMMRKKISMPAH 300
SAD2-A          LAKDHGDMKLAQICGIIAADEKRHETAYTKIVEKLFEIDPDGTVLALADMMRKKISMPAH 300
                ************************************************************

SAD1-A          LMYDGEDDNLFDNYSSVAQRIGVYTAKDYADILEFLVGRWKVDAFTGLSGEGNKAQDFVC 360
SAD2-A          LMYDGEDDNLFDNYSSVAQRIGVYTAKDYADILEFLVGRWKVDAFTGLSGEGNKAQEFVC 360
                ************************************************************:***

SAD1-A          GLPARIRKLEERAAGRAKQTSKSVPFSWIFSRELVL 396
SAD2-A          GLPARIRKLEERAAGRAKQTSKSVPFSWIFSRELVL 396
                ************************************
```

Appendix IV CLUSTAL alignment of (a) DNA sequences and (b) deduced amino acid sequences of *sad1*-a and *sad2*-a. Identical residues indicated by asterisks (*) and gaps are identified by dashes. Conserved amino acid substitutions are denoted with colon (:) and semi-conserved substitutions are indicated by a dot (.). Numbers on the right indicate the position number.

a

```
fad2a-a         ATGGGTGCCGGTGGCAGAATGTCAGTGCCTCCATCA------------TCCAAACCTATG 48
fad2b-a         ATGGGTGCTGGCGGAAGAATGGCCGTGCCTCCATCGAACAAGGCGGACTCCGAAACCTTT 60
                ******** ** ** ****** * ***********             *** ** *  *

fad2a-a         AAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATTCCT 108
fad2b-a         AAGCGGTCTCCTTACTCAAAACCTCCCTTCACTCTTGGTGAGATCAAGAAAGCCGTCCCT 120
                *** ***************** ** ** ***** ** ****** ******* *** * ***

fad2a-a         CCACACTGTTTCAAACGTTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACC 168
fad2b-a         CCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACC 180
                ******** ****** * ** ******** ****** ********** ** *********

fad2a-a         ATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTC 228
fad2b-a         ATAGCCGCCATCTTCTACTACATCGCCACCACTTACATCCACCTCCTCCCCAATCCTCTC 240
                ** ** ** ************************** ************* *  ******

fad2a-a         AACTACCTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGG 288
fad2b-a         TCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGG 300
                  **** * ** ****** ***************** ****** ********** ** ***

fad2a-a         GTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATG 348
fad2b-a         GTCCTAGCCCACGAATGCGGTCACCATGCCTTCAGCGACTACCAATGGCTCGACGACTTG 360
                **  * ** ********************************** ************ **

fad2a-a         GTTGGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCAC 408
fad2b-a         GTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCAC 420
                ** ***** ******** ** * **** * ** ** ******** ***************

fad2a-a         CGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG 468
fad2b-a         CGTCGCCACCACTCCAATACTGGGTCCCTCGAACGAGACGAGGTTTTTGTCCCCAAGCAG 480
                ** ******** ***** ** ** ** ** ** ** ** ***** ***************

fad2a-a         AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACA 528
fad2b-a         AAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACA 540
                **  * *  ** ** *** **** ******** ********************* *****

fad2a-a         TTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGGAGA 588
fad2b-a         CTTGCAGTCACTCTCACTCTCGGCTGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGG 600
                 * ** *****  * ** ***** ****** ********************** ** **

fad2a-a         CCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAG 648
fad2b-a         CCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAG 660
                ** ** *********** *******  ** *** *    ** ** ***** ** ******

fad2a-a         CGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTATACAGA 708
fad2b-a         CGAACGGAGATATTCTTCTCCGATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAG 720
                ** * ******** * * ***** ** ** **  *  * *** ***     ** ****

fad2a-a         CTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATA 768
fad2b-a         CTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTA 780
                ** *   ** * ** ***** * ****** ** *   * ******** *****  *  **

fad2a-a         GTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCTCTTCCGCAC 828
fad2b-a         GTGAATGGATTCCTTGTCTTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCAC 840
                ************** *** * ******************** ** ** **  * ******

fad2a-a         TACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTAC 888
fad2b-a         TACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTAC 900
                ***** ****************** ** **** ** **  ***** ****  ********

fad2a-a         GGGTTACTCAACACCGTGTTCCACAACATCACCGACACACATGTCGCGCACCATCTCTTC 948
fad2b-a         GGGTTTCTGAACACGGTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTC 960
                ***** ** ***** ******** ******** ** ** ** ** ******** ** ***

fad2a-a         TCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGG 1008
fad2b-a         TCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGA 1020
                ** ***************** ** ***** ***** ****************  * **
```

```
fad2a-a         GAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGGAGGGAGGCAAAGGAG 1068
fad2b-a         GAGTACTACCAATTCGACGGGACTCCATTCATCAAGGCGATGTGGAGGGAGGCTAAGGAG 1080
                ***** ***** ***** ******** **  * ***** ************** ******

fad2a-a         TGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAAC 1128
fad2b-a         TGTGTTTATGTCGAGCCCGACGAAGGTGACCAGAACAAAGGCGTGTTCTGGTACAACAAC 1140
                **  * *********** ** ***** ****  * * **********************

fad2a-a         AAGCTGTGA 1137
fad2b-a         AAGCTGTGA 1149
                *********
```

b

```
FAD2A-A         MGAGGRMSVPPS----SKPMKRSPYSKPPFTLGELKKAIPPHCFKRSIPRSFAYVAYDLT 56
FAD2B-A         MGAGGRMAVPPSNKADSETFKRSPYSKPPFTLGEIKKAVPPHCFKRSIPRSFSYVAYDLT 60
                *******:****    *:.:**************:***:*************:*******

FAD2A-A         IAAIFYYIATTYFHLLPSPLNYLAWPVYWACQGCILTGVWVLAHECGHHAFSDYQWLDDM 116
FAD2B-A         IAAIFYYIATTYIHLLPNPLSYVAWPIYWACQGCVLTGVWVLAHECGHHAFSDYQWLDDL 120
                ************:****.**.*:***:*******:************************:

FAD2A-A         VGFVLHSSLLVPYFSWKHSHRRHHSNTGSLDRDEVFVPKQKAEIGWYSKYLNNPPGRVIT 176
FAD2B-A         VGFVLHSCLMVPYFSWKHSHRRHHSNTGSLERDEVFVPKQKSAIGWHSKYLNNPPGRVLT 180
                *******.*:*********************:**********: ***:***********:*

FAD2A-A         LAVTLTLGWPLYLAFNVSGRPYDRFACHFDPHGPIYNDRERMEIYLSDAGIFTVCYILYR 236
FAD2B-A         LAVTLTLGWPLYLAFNVSGRPYDRFACHYDPKSPIYNDRERTEIFFSDAGILAVSFALYK 240
                ****************************:**:.******** **::*****::*.: **:

FAD2A-A         LVLTKGLVWVVSIYGVPLLIVNGFLVLITFLQHTHPSLPHYKSSEWDWMRGALSTVDRDY 296
FAD2B-A         LAVAKGLAWVVCVYGVPLLVVNGFLVLITFLQHTHPSLPHYKSSEWDWLRGALATMDRDY 300
                *.::***.***.:******:*****************************:****:*:****

FAD2A-A         GLLNTVFHNITDTHVAHHLFSTMPHYHAMEATKAIKPVLGEYYQFDGTPFVKAMWREAKE 356
FAD2B-A         GFLNTVFHNITDTHVAHHLFSTMPHYHAMEATKAIKPVLGEYYQFDGTPFIKAMWREAKE 360
                *:***********************************************:*********

FAD2A-A         CIYVEPDEGDPSQGVFWYNNKL 378
FAD2B-A         CVYVEPDEGDQNKGVFWYNNKL 382
                *:******** .:*********
```

Appendix V CLUSTAL alignment of (a) DNA sequences and (b) deduced amino acid sequences of *fad2a-a* and *fad2b-a.* Identical residues indicated by asterisks (*) and gaps are identified by dashes. Conserved amino acid substitutions are denoted with colon (:) and semi-conserved substitutions are indicated by a dot (.). Numbers on the right indicate the position number.

a

```
fad3a-a         ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATC 60
fad3b-a         ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCA------ATGGTGTGGCTATG 54
                ************************** ************      *************

fad3a-a         AATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATTGCA 120
fad3b-a         AATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCA 114
                ***************  ********************** ******************

fad3a-a         GACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGC 180
fad3b-a         GACATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGC 174
                ***************** *********** ********************** ******

fad3a-a         TACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAGCTGGAC 240
fad3b-a         TACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCAAAGCTGGAC 234
                ******************* ************************** ***********

fad3a-a         AGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTT 300
fad3b-a         AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTT 294
                ********* ********** ****** ******************************

fad3a-a         GTTCTTGGACATGATTGGTAA-----TTTCACATGATCTTTCTGGTAATGTGGGTTTTCT 355
fad3b-a         GTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTTT-- 352
                ********************     ********  **********************

fad3a-a         TTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGA 415
fad3b-a         ------ATTGAAAAAGATTAAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGA 406
                      ***************************** **********************

fad3a-a         GCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCCTCAATCC 475
fad3b-a         GCTTCTCAGACATCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCC 466
                ************  ************** ******************************

fad3a-a         TCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTT 535
fad3b-a         TCGTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATT---ATTGCTAGT 523
                ************************************ ***** ****   *** **  *

fad3a-a         CTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTT--------ACCAGGAGAATTA 587
fad3b-a         TCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTA 583
                  ***** *********   ********* * ****  *        **************

fad3a-a         GCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAG 647
fad3b-a         GCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTG 643
                ************************************************************ *

fad3a-a         TAAGTTGACATGCAGTTTGCTCTAAAA-TGCAGAGTCCTCTGTTTTTTGTGTGTTCTTGT 706
fad3b-a         TAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTT-TGTGTTCTTGT 702
                ******************** ***** ******** *********** ***********

fad3a-a         GCTTTAATGACGATGATAATGAAATTG---AAATTTGTAATAGCTGCCGGAGAAGGTGTA 763
fad3b-a         GCTTTAATGGTGATAATAATGAAATTGTTGAAA--TGTAACAGCTACCGGAGAAAGTGTA 760
                *********  *** ************   ***  ***** **** ******** *****

fad3a-a         CAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGTTTGC 823
fad3b-a         CAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTGC 820
                *************** **** ****************** ****************

fad3a-a         GTATCCTATCTACTTGGTAAGTAAACAGACTGA-----CTCCAAAGTAGGAACTAATGAC 878
fad3b-a         TTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGAC 880
                 ********************** *******     ************** *******

fad3a-a         AATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATTAC 938
fad3b-a         AATTTTGGACCCGA----GCTT--CCGACTCGGGTCGATTTATTTCGGGTGGGTACTTAC 934
                **************    * **    ************* * ************* ****

fad3a-a         CCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAA 998
fad3b-a         CCGAT--------------GCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAA 980
                *****              ************** ***************** *********
```

```
fad3a-a         CCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTAGTTTT 1058
fad3b-a         CCCGCCCCATTCTTGACGAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTT 1040
                ************************** ******************** **********

fad3a-a         TATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCC-TCAATTCTAGCCGGAATTTGAAT 1117
fad3b-a         TATGTTACTAGTTTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGTCAGAAATTTGAAT 1100
                ************************ ******* ** *******   * * *********

fad3a-a         TCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGTAATTAGT 1177
fad3b-a         TCAAACTGACATTGGGTAATTCTGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGT 1160
                ********* ** ********* ******** * ************ *************

fad3a-a         TGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACG 1237
fad3b-a         TGGAATTAGACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCG------- 1213
                ** ** ******** **** ************************  ***  *

fad3a-a         GAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGAATTATTGGAC 1297
fad3b-a         -------------------------------GAAAAGCGTTTCCAGTGGAGTTAGACG-C 1241
                                             ***** *** ** ****** ***   * *

fad3a-a         CCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACAC 1357
fad3b-a         CCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCT-----------ATTTT 1290
                ********************************** * *  *           *

fad3a-a         TACTGT-CATGCATTTTA-ATCTTGTATGGTTTTACCCCAATGGATGCAGCGATGGATCC 1415
fad3b-a         TAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCC---------------------- 1329
                ** *** *********** ***** ** *** *******

fad3a-a         GGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAA 1475
fad3b-a         ------------------------------------------------------------


fad3a-a         AATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGTGTCGTGGATGGATG 1535
fad3b-a         ---------------------------------------------------GATGGATA 1337
                                                                   *******

fad3a-a         TACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGA 1595
fad3b-a         TTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGG 1397
                * ****** ************  **** ***********  *  ********** ****

fad3a-a         TGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGACACGATCTGTTTTGGGT 1655
fad3b-a         TGC-----TG---------------------------------CACGCTCT--------- 1410
                ***     **                                  **** ***

fad3a-a         GGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCT 1715
fad3b-a         -------------------------TCGTGTTCCAACCCGCACCCAAACCGCCCCATTCT 1445
                                        ***** ***** ******* ********  *****

fad3a-a         CGACCAAAAAGATTTTATGAATACATATCAACAGAAAAATCTAGTTTTCATGTTACTAGT 1775
fad3b-a         CGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGT 1505
                ***** ********************** ********** ****** ******

fad3a-a         TTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAA 1835
fad3b-a         TTAATGTACAACAGTAT-AGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAA 1564
                ** ********** *** **************  * ******************** ****

fad3a-a         TTCGCATTCAAACCGGAATTGGATGGAATCGGTATACCTCGTCACGGATGCATTGTCAAT 1895
fad3b-a         TTCGCATTCAAACAGGAATTGGATGGAATTGGT---------CCCGGATGCATAGTCATT 1615
                ************* *************** ***         * ********* **** *

fad3a-a         TCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACA 1955
fad3b-a         TCCCAGGCAGTTTCATGGTTTTATAACCAATCAATCTAATCT-TATGCTTTTGATAAACA 1674
                *** **  **************  ************** *** **** ******* ****

fad3a-a         GTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGTGACCTGTTCGC 2015
fad3b-a         GTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGC 1734
                ***** ****************************************** ***** *****

fad3a-a         ACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACT 2075
fad3b-a         ACCAAACGAGAGGGCAGCGGTCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACT 1794
                *** ********* ** ***** ******* ****************************
```

```
fad3a-a         CCTCTGCTACTCATCGTTCATCTACGGCTTCCTTCCGGTCTTCAAAATCTACGGCGTCCC 2135
fad3b-a         CCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACC 1854
                ********************* *********  ******* *************** **

fad3a-a         TTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGA 2195
fad3b-a         TTATCTGATATTCGTGGCATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGA 1914
                ****** *********** ************************* **************

fad3a-a         GCAGAAGCTGCCGTGGTACAGAGGCAAAGAGTGGAGCTACCTACGTGGAGGGCTGACGAC 2255
fad3b-a         GCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGAC 1974
                ***************************** ****************************

fad3a-a         CGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTAT 2315
fad3b-a         CGTTGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCAT 2034
                *** ************************************************** **

fad3a-a         TCACCATCTCTTCCCTCAAATGCCACACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTG 2375
fad3b-a         TCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAA--TTTG 2092
                ********************************** ** **********  *    ***

fad3a-a         ATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTT 2435
fad3b-a         ATTATTAATTTACTGTTTTTGTTGTTATAATTTGAGTC-------------GGGAGATTT 2139
                ******** *** ************************             ** *****

fad3a-a         CCTTCTTGGATCCGATCCAGGATCAAGCTGGTCCCTTGAATTTCTATATGATCT------ 2489
fad3b-a         CCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAAT 2199
                ***** *  *********  * **** ** * ********* *  **** ****

fad3a-a         -TATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTT----------- 2537
fad3b-a         CTAGATTAATCAGGAACAATATGATCATGTTGTTTAAACTAATTTTGTTGGACCATAACC 2259
                 ** ****** * * * *** ** **** *  *** ** ** *******

fad3a-a         TACC-----------------------------------ATCATTTTCGATCACCGGA 2560
fad3b-a         TACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAG 2319
                ****                                   ******* ** * ***

fad3a-a         AAATGTCCTGAGCAGTTTTCCGGTCACTTTAACCTCCATTGAC--AAATTTTTTCACCCA 2618
fad3b-a         AAGTTTCTCGATCAGTTTTCCGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCA 2379
                ** * **  ** ************ ***** **** * ****   ***** *********

fad3a-a         CATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTT 2678
fad3b-a         CGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTT-----------------C 2421
                * * * ***  * * *   ********* * **** *****

fad3a-a         CGATGACCAACTGTACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATG-- 2736
fad3b-a         CGATGATTGGCCGTACAATTTTGTACGAAGAGTTGTACGGATCATATAAATGTGTATAAG 2481
                ******    * ****** ******   * ******* ******** **********

fad3a-a         --TACAAAAGTATATTCT-AAGTACTATACTAAGCATTACTTA-GTATTACGTTTC-TAC 2791
fad3b-a         TTTCTAGAAATCCGTACTGAAATA-TATAC-ATATTTGACTTTTGTATAAAGTGTAATAC 2539
                  *  * ** *   * ** ** ** ***** *    * ****  **** * ** *  ***

fad3a-a         -AAAC-CTATAGAGAAATGCATACAATTTTGTATAGAACTTAGTATACACGT------AG 2843
fad3b-a         TAAATACTATACT-AAGTGC-TGTA-CTCAGTATGATACTTAGTACACACATTTGTATGA 2596
                 ***  *****   ** *** *  *  *  ****   ******** **** *

fad3a-a         CTGTGAAATGTCAATTTC-CCTCCGTATTTTCAG-----AGAC--AAGACATGATTTTTA 2895
fad3b-a         CTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTA 2656
                ** **************  **    **** ****     ***   ******   ******

fad3a-a         GACTGGC--AGATTTTT------TTT----ATCGGAT-----AG--ATTTCTCCAACTTC 2936
fad3b-a         TAC-GGCTGAAATTTGTGGGGGCTTTGTAGATCGGATCCATAAGTCATTTCTT-GGCTCA 2714
                 ** ***  * **** *      ***    *******     **  ******    **

fad3a-a         AGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGT 2996
fad3b-a         AGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGT 2774
                *********** ************************************************

fad3a-a         GACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGAACCGAAGAAATCAG 3056
fad3b-a         GACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGAGCCGAAGAAATCAG 2834
                ******************************************* *************
```

```
fad3a-a         GGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACG 3116
fad3b-a         GGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACG 2894
                ********************************* ************************

fad3a-a         TTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACATATTCCCAAGTTCCCTA 3176
fad3b-a         TTAGCGACACAGGCGACGTCGTTTTCTATCAGTCTGACCCACATATTCCCAAGTTCCGTA 2954
                ******* ********************** ************************ **

fad3a-a         CCAG---TGCCACCACCAAGTCCAAATCTAGCTGA 3208
fad3b-a         CCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGA 2989
                ****   ******************** ******
```

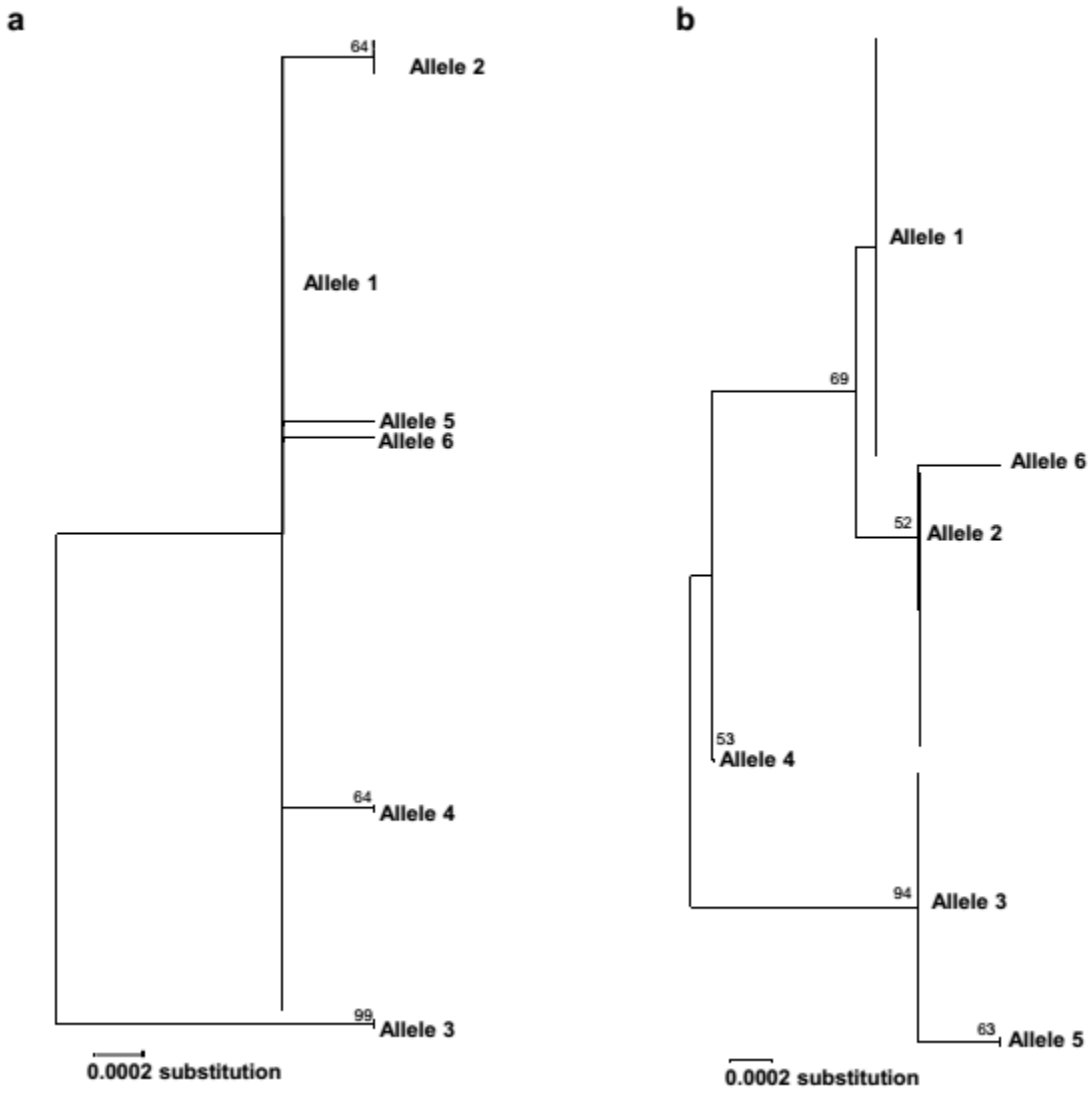
b

```
FAD3A-A         MSPPNSMSPATNGSTNGVAINGAKKLLDFDPSAAPPFKIADIRAAIPPHCWVKNPWRSLS 60
FAD3B-A         MSPPNSMSPTTNG--NGVAMNGAKKQLDFDPSAAPPFKIADIRAAIPPHCWVKNPWRSLS 58
                *********:***  ****:***** ********************************

FAD3A-A         YVLRDLLVILSFAVAATKLDSWTVWPLYWIAQGTMFWAVFVLGHDCGHGSFSDSWLLNNV 120
FAD3B-A         YVLRDLLVILSFAVAAAKLDSWTFWPLYWVAQGTMFWAVFVLGHDCGHGSFSDIWLLNNV 118
                ****************:******.*****:********************** ******

FAD3A-A         MGHILHSSILVPYHGWRISHKTHHQNHGNVEKDESWVPLPEKVYKSLDTGTKFMRFTIPL 180
FAD3B-A         MGHILHSSILVPYHGWRISHKTHHQNHGNVEKDESWVPLPEKVYKSLDTSTKFMRFTIPL 178
                ************************************************.**********

FAD3A-A         PMFAYPIYLWRRSPGKKGSHFNPYSDLFAPNERTSVMISTLCWTAMALLLCYSSFIYGFL 240
FAD3B-A         PMFAYPIYLWTRSPGKKGSHFNPYSDLFAPNERAAVLISTLCWTAMALLLCYSSFIYGFA 238
                ********** **********************::*:********************** 

FAD3A-A         PVFKIYGVPYLIFVAWLDMVTYLHHHGYEQKLPWYRGKEWSYLRGGLTTVDRDYGVINNI 300
FAD3B-A         PVLKIYGVPYLIFVAWLDMVTYLHHHGYEQKLPWYRGKEWSYLRGGLTTVDRDYGVINNI 298
                **:************************************************************

FAD3A-A         HHDIGTHVIHHLFPQMPHYHLVEATQAAKHVLGKYYREPKKSGPFPFHLFGYLVRSLGED 360
FAD3B-A         HHDIGTHVIHHLFPQMPHYHLVEATQAAKHVLGKYYREPKKSGPFPFHLFGYLVRSLGED 358
                ************************************************************

FAD3A-A         HYVSDTGDVVFYQSDPHIPKFPTS-ATTKSKSS 392
FAD3B-A         HYVSDTGDVVFYQSDPHIPKFRTSSATTKSKSS 391
                ******************** ** ********
```

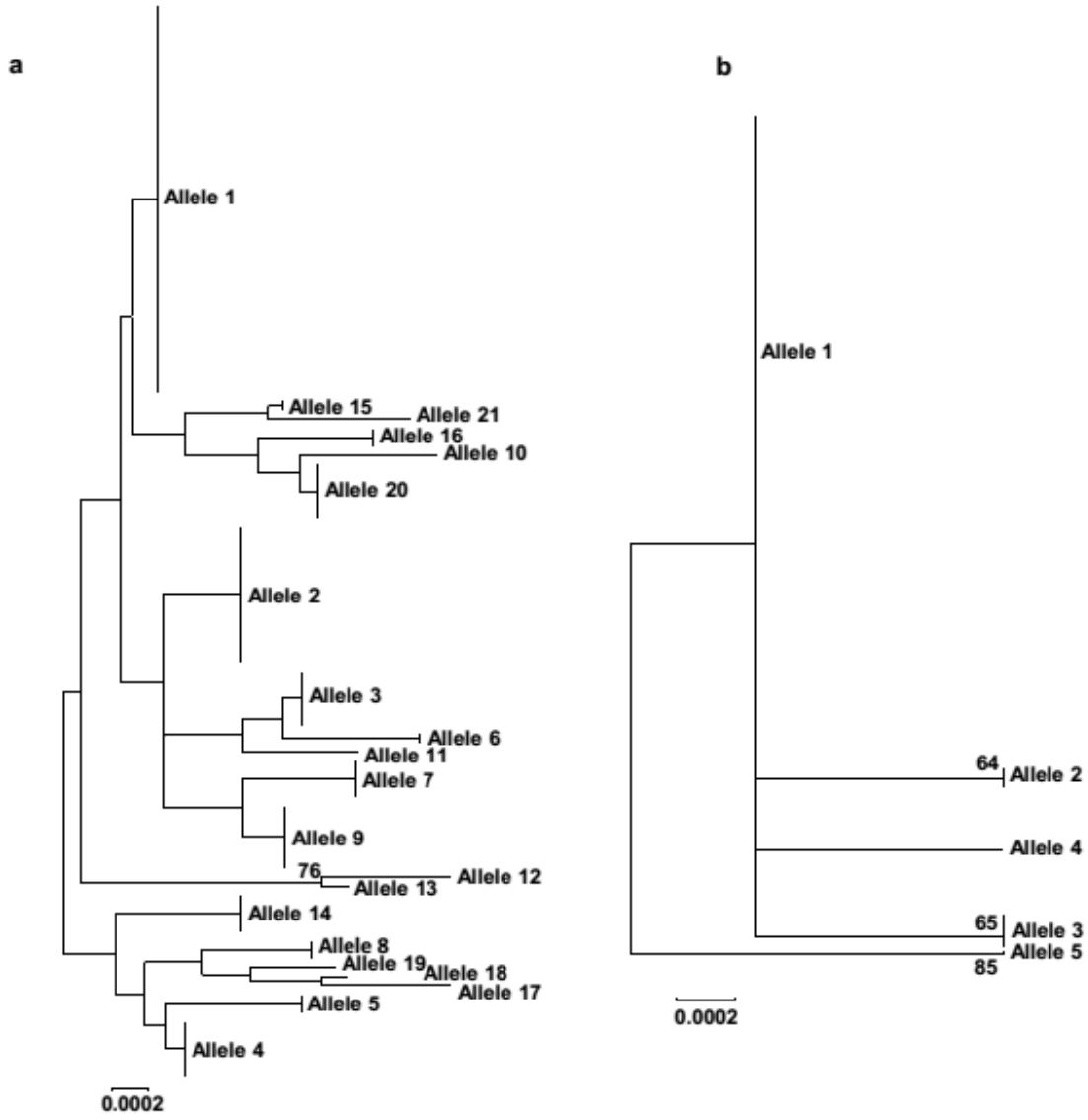
Appendix VI CLUSTAL Alignment of (a) DNA sequences and (b) deduced amino acid sequences of *fad3a-a* and *fad3b-a.* Identical residues indicated by asterisks (*) and gaps are identified by dashes. Conserved amino acid substitutions are denoted with colon (:) and semi-conserved substitutions are indicated by a dot (.). Numbers on the right indicate the position number.

Appendix VII Summary of SNPs and indels identified from *fad3a* allele 13, 14 and 15 (assembled into contig 2).

Accession	Exon I		Intron I		Exon II		Intron II		Exon III		Intron III		Exon IV		Intron IV		Exon V		Intron V		Exon VI		Allele	Isoform
	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel		
CN96846	3	-	3	1	2	-	2	1	-	-	4	1	1	-	14	4	1	-	66	11	3	-	13	C
UGG102-2	3	-	3	1	2	-	2	1	-	-	4	1	1	-	15	4	1	-	65	11	3	-	14	C
CN97351	-	-	-	-	-	-	-	-	-	-	-	1	1	-	7	2	1	-	65	11	3	-	15	F

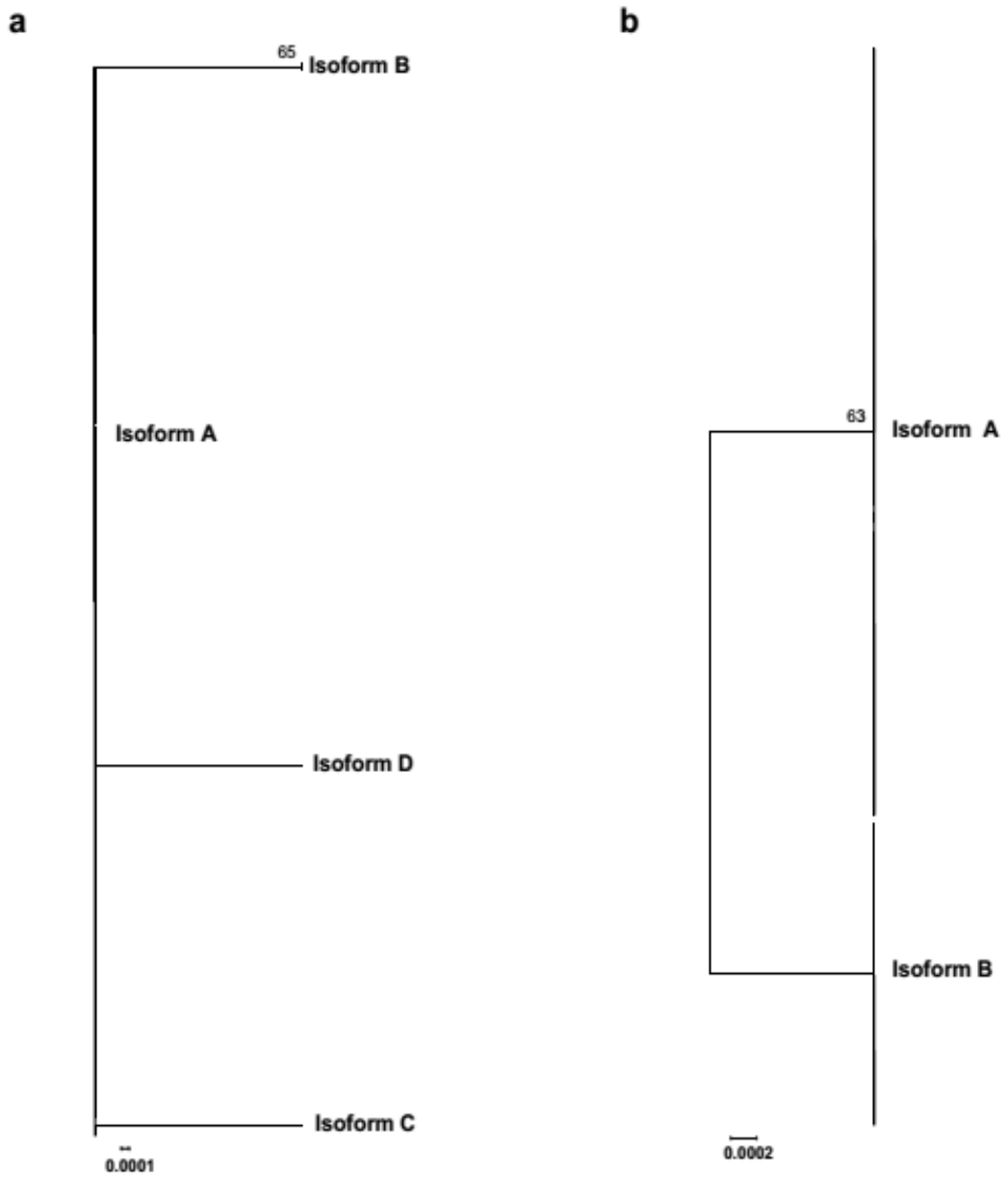


Appendix VIII Neighbour-joining tree of (a) *sad1* and (b) *sad2* full length gene sequences from 120 accessions of flax. Accessions are identified by their allele number and vertical branch length represents number of accessions. Bootstrap values greater than 50 were shown.

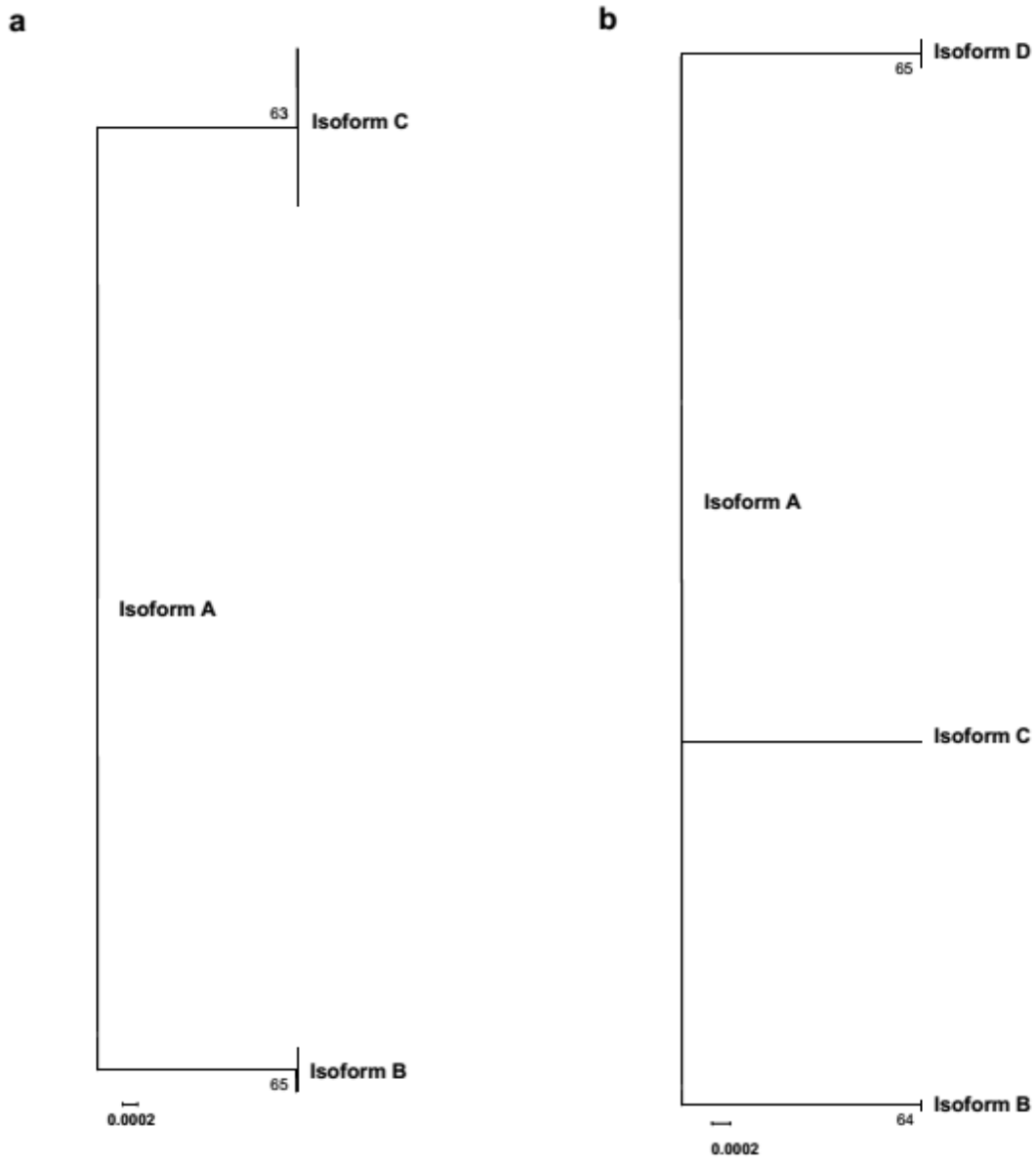


Appendix IX Neighbour-joining tree of (a) *fad2a* and (b) *fad2b* full length gene sequences from 120 accessions of flax. Accessions are identified by their allele number and vertical branch length represents number of accessions. Bootstrap values greater than 50 were shown.

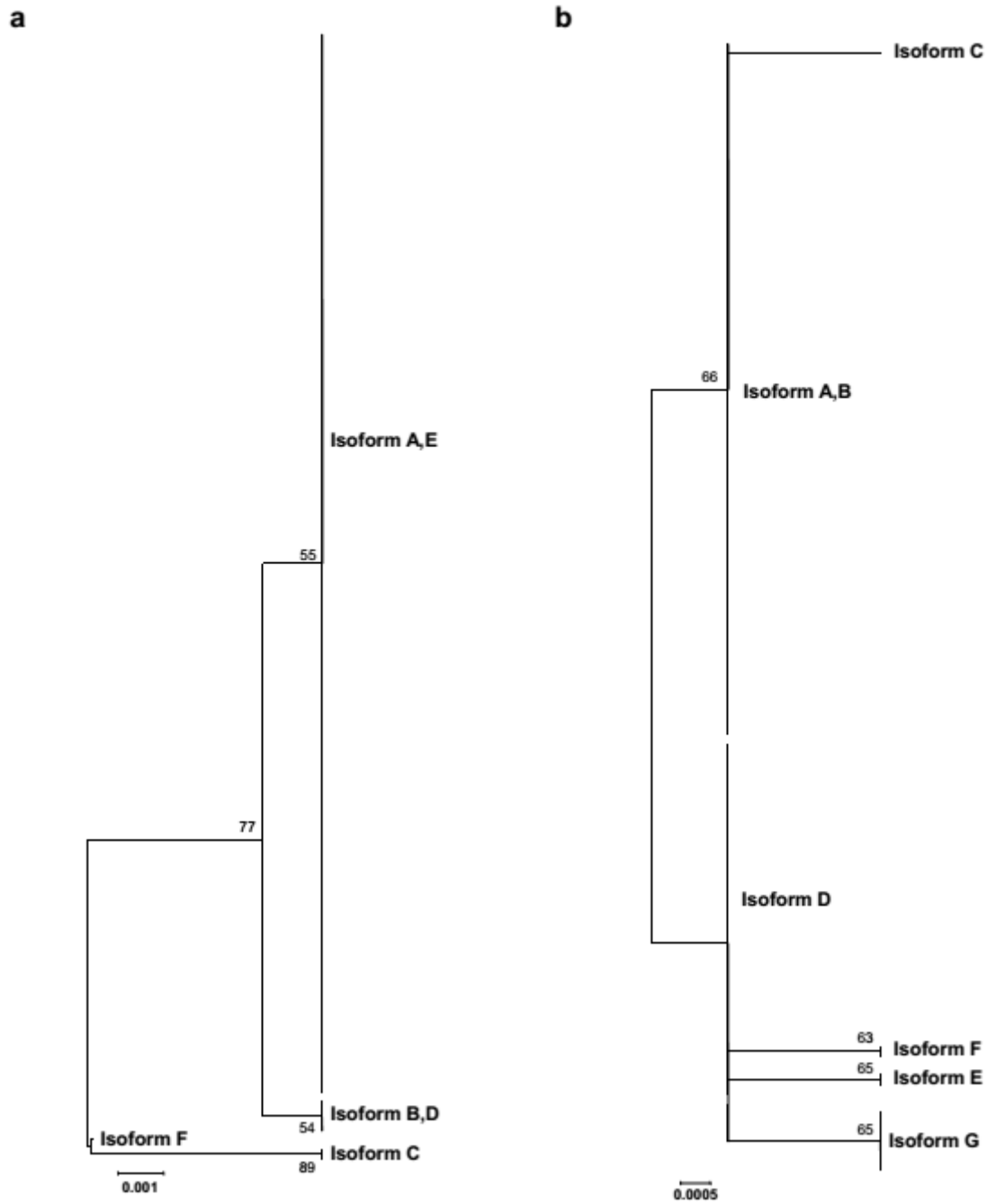
Appendix X Neighbour-joining tree of (a) *fad3a* and (b) *fad3b* full length gene sequences from 120 accessions of flax. Accessions are identified by their allele number and vertical branch length represents number of accessions. Bootstrap values greater than 50 were shown.



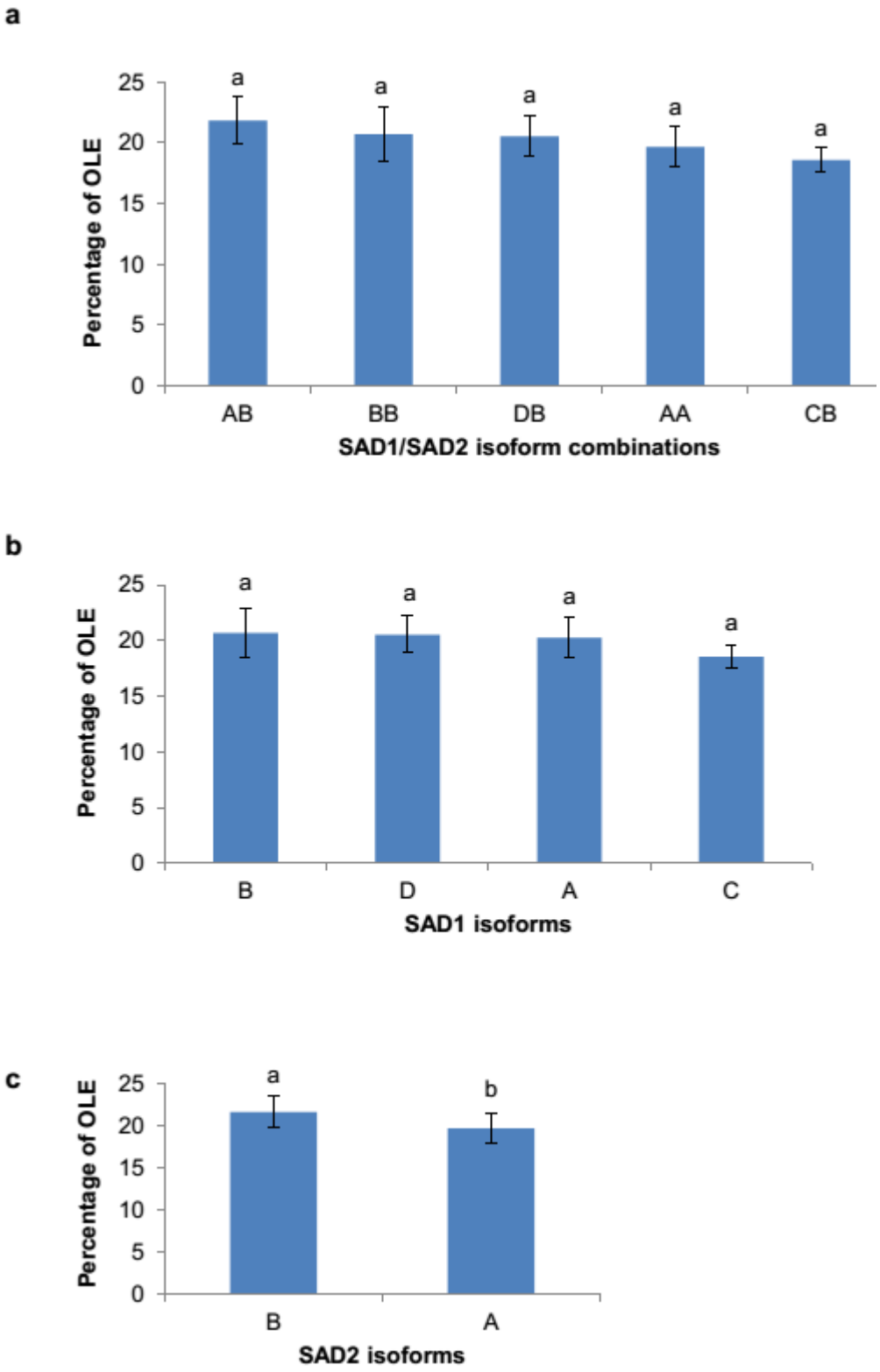
Appendix XI Neighbour-joining tree of (a) *sad1* and (b) *sad2* deduced amino acid sequences from 120 accessions of flax. Accessions are identified by their isoform and vertical branch length represents number of accessions. Bootstrap values greater than 50 were shown.



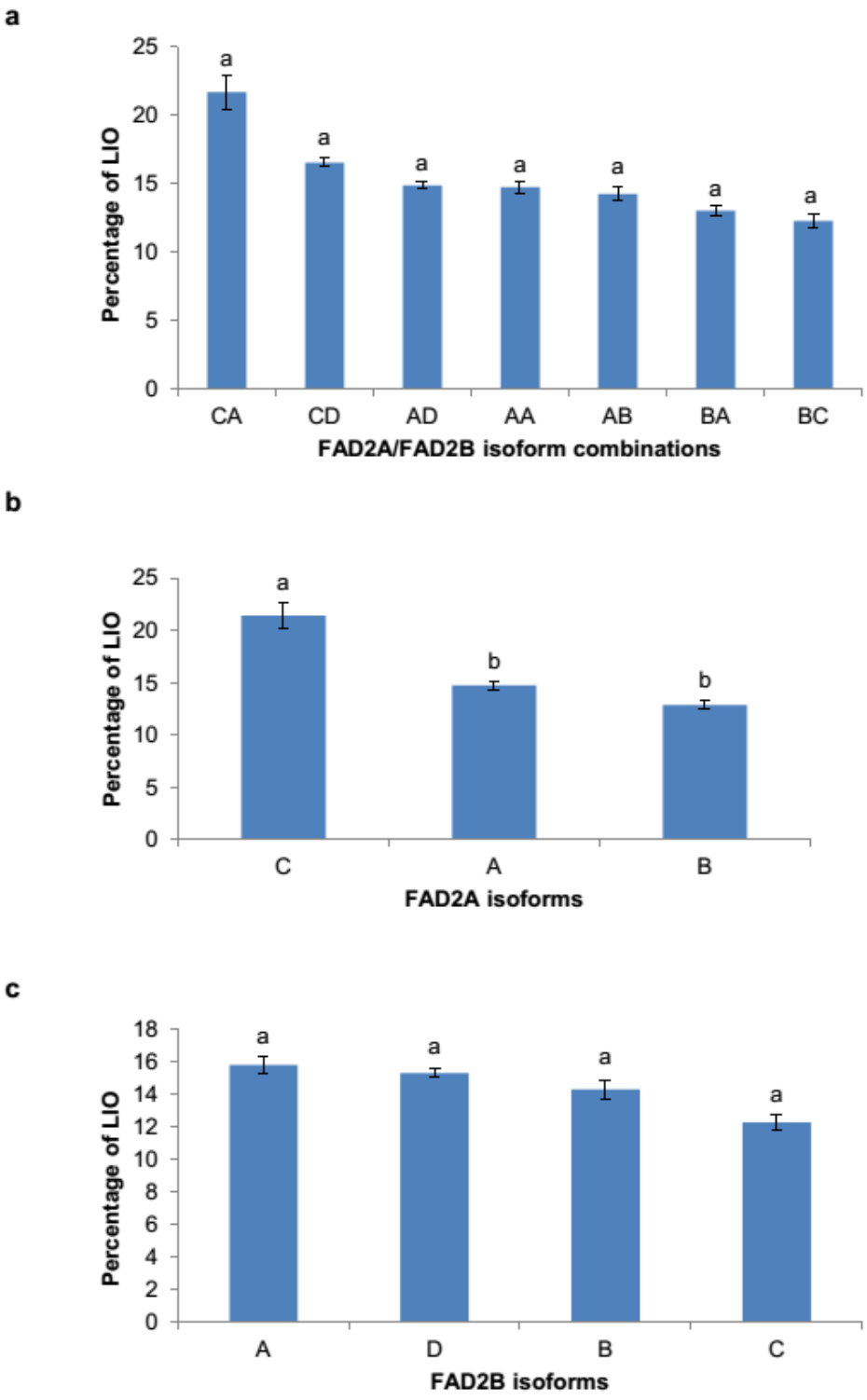
Appendix XII Neighbour-joining tree of (a) *fad2a* and (b) *fad2b* deduced amino acid sequences from 120 accessions of flax. Accessions are identified by their isoform and vertical branch length represents number of accessions. Bootstrap values greater than 50 were shown.



Appendix XIII Neighbour-joining tree of (a) *fad3a* and (b) *fad3b* deduced amino acid sequences from 120 accessions of flax. Accessions are identified by their isoform and vertical branch length represents number of accessions. Bootstrap values greater than 50 were shown.



Appendix XIV Association between OLE and the predicted isoforms of (a) SAD1/2 and OLE content, (b) SAD1 and OLE content and (c) SAD2. Vertical bars represent standard error of the mean. Letters on top of the bar indicate statistical significance of Duncan's multiple range tests.



Appendix XV Association between LIO and the predicted isoforms of (a) FAD2A/B and LIO content, (b) FAD2A and LIO content and (c) FAD2B. Vertical bars represent standard error of the mean. Letters on top of the bar indicate statistical significance of Duncan's multiple range tests.

Appendix XVI Effect of SAD and FAD isoforms identified from the accessions representing the non-mutant flax germplasm on palmitic, stearic, oleic, linoleic and linolenic acid composition, oil content and iodine value

Trait	Genes/gene	*P*-value	Predicted isoforms or combinations
Palmitic acid (PAL)	*sad1*	0.5153	
	sad2	0.9991	
	fad2a	0.0169*	(5.95-5.51)[a]
	fad2b	0.8519	
	fad3a	0.8295	
	fad3b	0.0364*	[E(6.37)][a], [D,G(5.73-5.78)][ab], [A(5.47)][b]
	sad1/sad2	0.6638	
	fad2a/fad2b	0.1652	
	fad3a/fad3b	0.1961	
Stearic acid (STE)	*sad1*	0.1955	
	sad2	0.0955	
	fad2a	0.3710	
	fad2b	0.1749	
	fad3a	0.6480	
	fad3b	0.0701	
	sad1/sad2	0.0551	
	fad2a/fad2b	0.3845	
	fad3a/fad3b	0.1845	
Oleic acid (OLE)	*sad1*	0.9674	
	sad2	0.0118*	[B (21.77)][a], [A(19.87)][b]
	fad2a	0.0144*	[A(20.91)][a], [C(18.62)][ab], [B(17.90)][b]
	fad2b	0.2806	
	fad3a	0.5581	
	fad3b	0.0003*	(19.43-22.64)[a]
	sad1/sad2	0.1146	
	fad2a/fad2b	0.0097*	(15.59-21.12)[a]
	fad3a/fad3b	0.0021*	(19.23-24.18)[a]
Linoleic acid (LIO)	*sad1*	0.7191	
	sad2	0.7904	
	fad2a	0.1604	
	fad2b	0.4892	
	fad3a	0.6222	
	fad3b	<.0001*	[G(14.87)][a], [A,E(13.51-14.43)][ab], [D(12.33)][b]
	sad1/sad2	0.8547	
	fad2a/fad2b	0.4840	
	fad3a/fad3b	0.0004*	(11.76-14.87)[a]
Linolenic acid (LIN)	*sad1*	0.9777	
	sad2	0.0799	
	fad2a	0.0770	
	fad2b	0.4077	

Trait	Genes/gene	*P*-value	Predicted isoforms or combinations
	fad3a	0.6365	
	fad3b	0.0743	
	sad1/sad2	0.4339	
	fad2a/fad2b	0.0374*	(53.25-61.01)[a]
	fad3a/fad3b	0.1772	
Oil content (OIL)	*sad1*	0.9563	
	sad2	0.6414	
	fad2a	0.0004*	[C,A(42.68-44.07)][a], [B(39.89)][b]
	fad2b	0.2268	
	fad3a	0.1946	
	fad3b	0.0569	
	sad1/sad2	0.9804	
	fad2a/fad2b	0.0078*	[CA,CD (43.10-44.13)][a], [AA, AB, AD, BA(40.17-42.73)][b], [BC(38.52)][b]
	fad3a/fad3b	0.0599	
Iodine value (IOD)	*sad1*	0.9785	
	sad2	0.1757	
	fad2a	0.1898	
	fad2b	0.2593	
	fad3a	0.5835	
	fad3b	0.0016*	(183.70-189.88)[a]
	sad1/sad2	0.7473	
	fad2a/fad2b	0.0932	
	fad3a/fad3b	0.0081*	(179.11-190.68)[a]

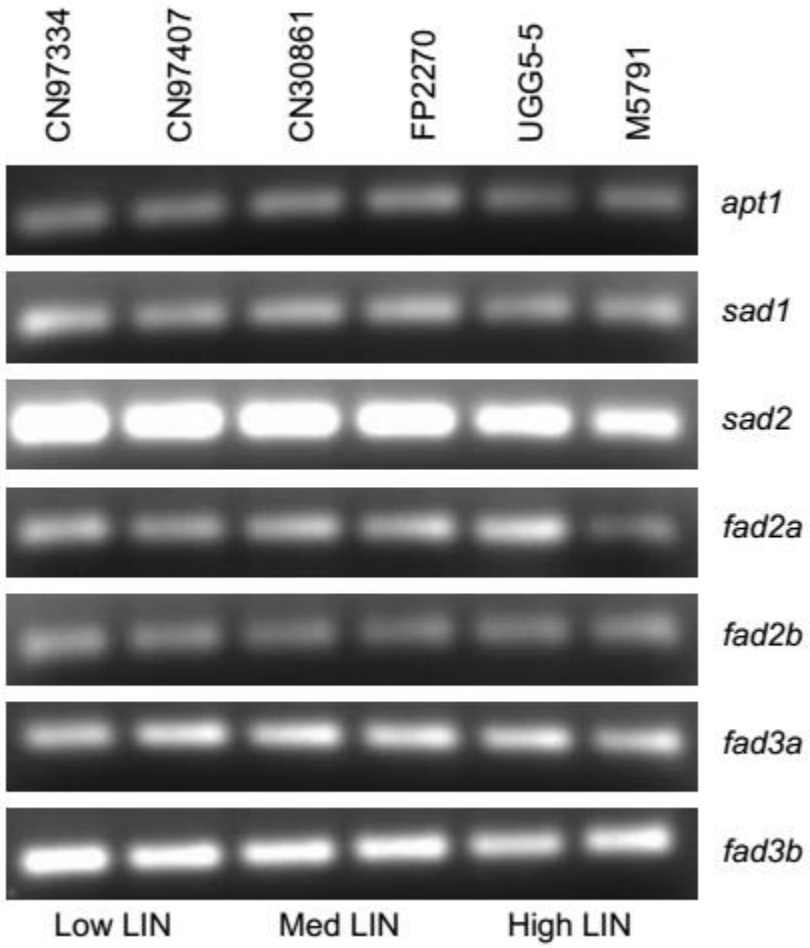
[1]Means for the isoform(s) or isoform combinations are in bracket. They represent data collected from two locations during three years. Superscript letters indicate statistical significance of Duncan's multiple range tests. *Statistical significance ($p < 0.05$)

Appendix XVII Description of flax accessions used for *sad* and *fad* gene expression study with the predicted *sad* and *fad* alleles and isoforms according to the nomenclature previously described (Thambugala et al. 2013). Phenotypic data for fatty acid composition and oil content were averaged from two locations (MB and SK) over four years (2009, 2010, 2011 and 2012).

Accession description	Accession number	Allele and Isoform												Palmitic acid (PAL)		Stearic acid (STE)		Oleic acid (OLE)		Linoleic acid (LIO)		Linolenic acid (LIN)		Oil (OIL)	
		SAD1		SAD2		FAD2A		FAD2B		FAD3A		FAD3B		%[1]	SE[2]	%	SE	%	SE	%	SE	%	SE	%	SE
Mocoreta	CN97334	2	A	3	A	1	A	1	A	2	A	1	A	5.40	0.17	4.39	0.21	21.76	1.95	14.86	0.40	53.67	1.68	44.18	0.91
Rio (Long 79)	CN97407	2	A	1	A	1	A	1	A	1	A	1	A	6.03	0.22	4.76	0.29	20.99	1.61	13.96	0.52	54.28	1.32	42.69	1.01
Kubanskij	CN30861	1	A	3	A	13	A	1	A	1	A	1	A	5.35	0.11	3.23	0.12	18.90	2.28	13.10	0.26	59.28	2.21	45.49	1.16
FP2270	FP2270	1	A	3	A	1	A	1	A	3	A	1	A	5.06	0.10	5.47	0.55	17.69	2.35	12.27	0.33	59.60	2.47	42.71	1.08
UGG5-5	UGG5-5	1	A	3	A	20	A	1	A	3	A	1	A	4.00	0.10	2.48	0.14	12.96	1.80	11.17	0.32	69.31	1.71	44.54	0.91
M5791	M5791	1	A	3	A	2	A	1	A	3	A	2	A	4.40	0.06	2.44	0.12	11.41	1.44	9.89	0.28	71.85	1.43	43.73	0.86

[1]Expressed as a percentage of the total fatty acid composition

[2]Standard error



Appendix XVIII Reverse transcriptase PCR of six fatty acid desaturase genes and the reference *apt1* gene from six flax genotypes comparing the level of expression at 20 days after anthesis showing the consistent amplification of the control *apt1* gene and the relative differential expression of the six desaturase genes.

Appendix XIX Analysis of variance for expression (*gene:apt1*) of six desaturase genes. Mean square values and statistical significance for *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* during seed development of six flax genotypes are shown.

Gene(s)	**Source of variation**	**Mean Square**	**Pr > F**
sad1, sad2, fad2a, fad2b, fad3a, fad3b	Gene	74.23	<.0001*
	Genotype	3.12	0.1400
	Stage	49.29	<.0001*
	Gene*Genotype	0.53	0.9998
	Gene*Stage	3.24	0.0101*
	Genotype*Stage	1.20	0.9306
	Gene*Genotype*Stage	0.27	1.0000
sad1	Genotype	0.35	0.9572
	Stage	4.27	0.0233
	Genotype*Stage	0.20	1.0000
sad2	Genotype	2.41	0.7873
	Stage	28.28	<.0001*
	Genotype*Stage	1.01	1.0000
fad2a	Genotype	0.31	0.8895
	Stage	8.88	<.0001*
	Genotype*Stage	0.27	0.9999
fad2b	Genotype	0.31	0.9147
	Stage	4.25	0.0013*
	Genotype*Stage	0.19	1.0000
fad3a	Genotype	1.50	0.2420
	Stage	11.83	<.0001*
	Genotype*Stage	0.36	0.9994
fad3b	Genotype	0.90	0.6805
	Stage	7.96	<.0001*
	Genotype*Stage	0.49	0.9992

* Statistical significance ($P < 0.01$)

Appendix XX *P*-value of *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* for genotypes during seed development.

Seed developmental stage (DAA)	***P*-value**					
	sad1	***sad2***	***fad2a***	***fad2b***	***fad3a***	***fad3b***
8	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
12	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
16	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
20	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
24	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
28	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
32	n.s.	n.s.	0.01*	n.s.	0.001*	0.002*

* Significant at $P < 0.01$; *n.s.* = non-significant

sad1

-849 TTGCAGATCTCGTTGCTGATAACACATACATGGCTACAAGACTCTAAAACGGTTCAAAGTGAAATTGTTTTGGTGGTAGAGTTGTGTGTTTGGTGACTCG

ACGTBOX

-749 AAAGTTCTGGATTCGAATCCAGCATTCCCCACAAAATAGACACCAACGTAGTGTTTATTTACCGTCTTCTATCTTGTATTGACCGAGAGTTTCGATATAC

T(CN30861)

-649 TCCGACAAAAAAAGACATCTTCACATCATCAAATGGATCCGTAGTTAGTGCAGTGGCTCGAATAACATAAATGAAAAAAGGAAAAAATTTGCCTGAAATC

-549 GATGCTCAAAACAAGTAGAAATTCATTCAAACATATTTAGACAAACACGATCATTTAGCATCATCAAATTAATAACAAGAGCAAACAATAAAGCACATAG

-449 CAAAACATACAATAGTCGTCTTGCAATGTCATATGATAATAAGCCAGTGAAACCATGAAGCCCAAGTGAAGTGGTCAAGTGGAGCTGAAAGCTTCCGAAC

DPBFCORE MYBCORE

-349 CCAAGCCCCCGCTACCGGGTTAGGACATACGACACGCGACATGCTACGAAACTTAAAAATCGGTCACGCAGTTAATGGAACAAATGAAACGCAACGACTA

POLLEN1LELAT52 WRKY

-249 TTAAGTGACCATTTTGCAGAAATGATATGAAAAAGTGACCATTTAGACAAATGAGCAAAGAAAATACAAGTGGCGAGTGCTGACATAATAAACCGAATGC

DOFCORE HSE EBOX

-149 AGGCGTTACCATCCAATTTTACAACCATTCAATTCAAAAGTTTTTCCAATTTCCATTTCCTCATCTGCCTTACCCATAAATCTCGACGGACACCAAAAAA

TSS

-49 CTCAGCCAGCTTGCCCCCAAACAACAGCGCAGAAAAACCTTCAACAACAATG

+1

sad2

-954 AACATCAATGTCAATCTCTGCAGATTTTTGTTAGCAGCAGGTCATGATTCTTTTTTGGTTGATTCTTGTGAATGTAAGCTATTTGTTGTTGTAATATATG

-854 CATTGATTGTGATTTTGTTTTAGCTTTGATCAATGAAATAAATCTCGTTCAACCCAACCATCAGGCTCTTTCATATTCATTTTGACGACTATATATACAT

MYBCORE POLLEN1LELAT52

-754 AATCGTACAAACTATTCGGTTAACTAATCTACAGAAAGTCGGAGTTAGCTAGAGATTGTCAAGGAGGAGGAGATCATACACCTAATTTTGAAGCTGATTC

WRKY

-654 TTCATCTATGATTTCGAGTTTTGACTTGATTTGGCTCTTCGATATTCGAAATTAAATGCCTCAATGCCTCCAAAGTGCTCTCTACTTGCGGGTGGACCTA

EBOX ACGTBOX

-554 CAAAACTAGACAAACAGGTGCAAAAAACATGTGTTTACACGTCCATGTTATCTTGCATTGGCCCATGTTTTCTGCATTGTAAATCTTTCCCCAAACACAT

-454 AGTTAGACGAAGTCGATAATCTAGCACCATCAAATCAATAACACGAGCAAATAATAAAGTAAATAGTGAAACCATGAAGCCTAATTGGTCGAGTGGAGCT

DOFCORE TSS

-354 GAAAGCTTTCATCGGTATCGAACCCAACCCCCCCTGCTACGAAACTTAAAAATGGGTTACGCAATTACACTCGATAGAACTGATGAAACGCAACGATTGT

DPBFCORE

-254 TAAGTAACCATTTTGCAGAAACGATAATTAACAAGTGACCATTTGGATAAATGACCAGAGAAAATACAAGTGGCGAGTGCTGACATAATAAACCGAATGC

-154 GGGCGTTACCATCCAATTTTACAACCATTCAATTCAATATCTCACATTCAAGTTTTTCCAACTTCCATTTCCTCATCTGCCTTACCCATAAATCTCGACA

A(CN30861)

-54 CCAAAACACTCAGCCAGCTTCGTCCCAAACAACGCAGAAAAACCTTCAACAACAATG

+1

fad2a

-702 ACACCCCTCCTGCATGCGCGAATCTGTGGGATTTTTCCTGCAATTGAAATTGATTTTCCGCTAATTAGGGGGTGTTTGGCTGAGTTCTTCGTTCACCAGT

POLLEN1LELAT52

-602 AGAAATTGCAAAACTGGAAATGAGAACACCAAAAATTGAATGAAAAAACTACATTAGCTGGTTTATTAAACGCCTGCCCTCCTTCATATTCTTCTTTTGG

EBOX MYBCORE

-502 GTTCCGGGTCCTAATCATATGCTGATTCAGTCAATTCTCTGTTGCTTTATTGGTGAAATTGGAAGGAATTCAAACTTTTGTTTGCTTCAGTGCGTGAATG

DOFCORE

-402 AATGAACCCACTCTATCACCTTCAGGGAAATCTTGGTGGGTTGGTAGTTAGGTGGTCATAGTTGGGTTTCTTTTCAAAGGTTGAAAAGAGGGTTCTTTTC

WRKY TSS

-302 ACAAGTTATTTGACTGGTGGAATCATTAACTAATTTGTTCCTAAAAATGGACAAGTTGTGCTAAATCACGCAACTGAAAACTGGAAAGGAAAGACCAATT

-202 CAAAATCGGGTAATTAGGTTTGGTGGTGATAATGTTAAGTAGGACTTTTGATGTAGGGGAAGAATCCAAGTAATCATTCCTTGTCAAGAATCCTTCTCTG

-102 TTTCTTCCTGCAAGTTATTGTGATAATTCTGATGAATCTTTTCTTCTGATCATATATGATATATATGCAGGTGCTGTAACAAAATACACAGAAAGAAGAA

-2 AA**ATG**
+1

fad2b

EBOX
-928 TCTTGATGCAATCTCTCTGAGTTTGAAAAAATATCTTTTTCTTTTGGAATGATTGATCGTCTCTTACTATTGTTTGAACATTACACGTGACATTGGTGTG

HSE
-828 ATGAACTGTAACAGGGCCAATAAGAGATTTGTTTCTGTGCTCTGAATTTTCTTTCTTTCTTTGAATCTGGAAAGTCAATAATACTGATATGTTACCCTAA

POLLEN1LELAT52 **DOFCORE**
-728 AGTGATCAATGGATTTTCTATTCTTAGAAAGCCTGATTTTTTTGGGGGAAAGGGTTCAGAATTGAGATTGCAACACTGTACCAAACATTGTATGTTATTC

-628 AATAGGGGTTAGATTCTGAGCTTCTTGGTCTGTGACTTTCATAAGAGAGAATTTAATTCTCGTAAATGGTCATTTTCAATTATACTTGCTGTGGCCTGCA

MYBCORE
-528 TATATCTACCTTGCCGCTATGCTTACAGCATGCTGTTATTTTGAATGTAACCTGATTTGTGGAATTTAGCTTACTTTTAAACTTACTAATAATTAGATTC

WRKY **DPBFCORE** **TSS**
-428 GTATCAGGATAGCACTACTTGCTTAAACTGACTATAAAAACACAAGTCTCCCATGGTTGTACGCTTGTCCTGGTTCAATACGCAACTTCAGAGAGACATT

-328 TTTTCTGACGACGATTGATAAAGACGGTAAACTGTCTCGTGGAACCGAGAGTAGTTGTTTCCTCTCAAATTTGAACCTTAACGTTTCTGCATCGGATATT

-228 CTCGATTTGGGGAGGGCATTATCCTTGTTGTTACAGTATATCTATTGGTATCCTAGGCATTGTAAAATAAGATTGTCCTTTGTAACCTAAATTGTAACGG

-128 TTGTTTCCCAATACTCCAAACACCTCAACTCTGTGATTCTGCGCGCACACAGTTTCTTTCCATCACATCACAAACATGATTTTTTTCTCTGGTGCCAAAC

-28 TTCTGCAGGTGCTGTTGATAAAGCAAGA**ATG**
+1

fad3a

-853 AAGAGAGTTGTAAGAGTTTCGACTAAGTTCAAATGGAGCCCAAAGTTTGATCATCAGTTTGTGAAAACAAAGTCAAGCTCGTCCATATCTCTGCCTTGTT

HSE **DPBFCORE**
-753 CCCAACCCACTACATAGCATCTGGAAGACCTCGTACTTCACATTCTCGGACCGAAGGACAACCAATACCCCCCTTGTGATCCTAAACACATGCACAAATC

-653 CCTCTGCCCGAAACTTGCCCGAACTTACTCCCTAAGACCGATGCCCACTTGAGTCACATGAGTTGATTAGTCGATTTCACCCTAGCTCCCGCGAACTCAG

-553 CAGTGCCCGTTGCGACTCCGCCAAATCACTAATCCTTAATTAAAAGAACTAATAAGTTGATATCATCACATTTGTGGTAACTCATGCATGCACATAGGTT

EBOX
-453 TCCTAGATACCATTGAAGGAAGTTGCCATGTGTTTGAATCAAAGATTTGCCCACCACCATTGATACTGAAATTGAAGAACCTAGCAGCCAGCAACGGCTC

MYBCORE
-353 CTTTTCATTTGTCTTTCAACAGAGCAAGTAACAACAACCGTTGCCTAAACTGAAACCCAATAAAGAGCAAAAAAAAAGGGGTTGGGTGGTGTAGGCTAGT

-253 TTGTCTGAAATCAGTGTACATTTTGCATTTCCATTTACTCTTCTCCATCCACTTGGCATCCTGCATTACTTCTTCTTCGTTAGTTCTCACCAACCTACAT

DOFCORE
-153 ACTCTTCGGTTATAAATACTGTGAGGCTGAAACCAAAGGCCACTCAGTCTATTCATTATTATTCAAAAATATATATTGGGTTTGTTTGGTGCAGATTACA

WRKY **TSS**
-53 GTGACTTCAAAACTGTGGCTCTGCA**C**GACCAAACTATGAGCCCTCCAAACTCA**ATG**
+1

fad3b

```
-780   CCCAACCCATTACATGACGTCAGAAAGAGCTCGTACTTCACATTCTCGGACCGAAGGACAACCATGCTTTGTTGACATCTCCGAAAGGCAACCAATTATA
                  ACGTBOX                                                                 WRKY                HSE
                     EBOX
-680   CCCCCTAGTGATCCTAAACACATGCACATGTCCCTATGCCCGAACTTGCTCCCTAAATATTGCAGCATCAACATTGCAATTCAGAGTCCCCCAAACAAGA

-580   CCGATGCCCACTTGAGCCACATTAGTTGATTAGTCGATTTCGCCCTAGCTCCCGCTAACCGATCGTAGCCACTCCACCAAGTCACTAATCCTTGATTAAA

                        DPBFCORE
-480   GAGTTAAACAAGTTGATATCACACCTGTGGTAACTCATGCACATACCATTGAAGAAAGTTGCCATGTGTTAGAATCAAGGATTTGCCCACCACCATTGAT

                                                                                    MYBCORE
-380   ACTGAAATTGAAGAATGCTAGCTAGCAGGCAGCAACGGCTCCTTTTCATTTGTCTTTCAACAGAGCAAGTAACAACAACCGTTGCCTAAACTGAAACCCA

            POLLEN1LELAT52
-280   ATAAAGAGCAGAAAAAAGGGTTGGGTGGTGTAGGCTAGTTTGTCTGAAATCAATGTACATTTTGCACTTCCATTTACTCTTCTCCATCCACTTGGCATCC

                                                  A(CN30861)                                DOFCORE      TSS
-180   TACATTATTACTTCTTCTTCGTTAGCTCTCACCAAACTTTACATACACTTTTTCGGTTATAAATACTGTGAGCCTCAAAGCAAAGGCCACTCACTCTATT

-80    CATTATTATTAAAAAAAAAATATTGGTTGTTTGGTGCAGATTATAGTGACTTCAAAACTGTGGCTCTGCAGGACCAAACTATG
                                                                                          +1
```

Appendix XXI Nucleotide sequences of the promoter region of *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* genes. Sequence numbering is relative to the ATG codon (+1). 5'-UTR is highlighted and putative transcription initiation site (TSS) is indicated. The TATA and CAAT boxes are boxed. Putative *cis*-acting regulatory elements are underlined and designated with the names of each motif (Higo et al. 1999; Chang et al. 2008). Red colour letters indicate the single point mutations identified in the promoter regions of the corresponding genotype.

Appendix XXII Analysis of variance for fatty acid composition for the data collected from six environments. Mean square values and statistical significance for palmitic acid (PAL), stearic acid (STE), oleic acid (OLE), linoleic acid (LIO) and linolenic acid (LIN) are shown.

Source of variation	**PAL**	**STE**	**OLE**	**LIO**	**LIN**
Genotype (G)	3.96*	6.48*	63.43*	280.06*	299.00*
Location (L)	6.00*	59.75*	4399.06*	89.49*	4492.78*
Year (Y)	6.88*	15.56*	715.02*	46.41*	444.19*
G * L	0.21*	0.24*	2.95*	5.46*	6.32*
G * Y	0.25*	0.18*	2.74*	4.99*	7.96*
L * Y	0.24*	15.75*	878.42*	25.58*	822.69*
G * L * Y	0.15*	0.11*	2.08*	3.60*	5.43*

* Statistical significance ($P < 0.0001$

Appendix XXIII Summary of 454 sequencing statistics of 12 BAC clones

BAC clone	Number of reads	Number of nucleotides	Average read length (bp)	Assembly size (bp)	Coverage (X)
6M22	20558	5316732	259	138726	38
27L18	17431	4595038	264	78447	59
28C5	34481	8506145	247	110505	77
44E4	17528	4572177	261	108314	42
139G15	9939	2562165	258	107370	24
212N17	20153	5325129	264	115454	46
317I7	16536	4358130	264	172117	25
346C18	9146	2414566	264	140743	17
356B4	17096	4408089	258	68175	65
346K11	26544	6727863	253	127580	53
375N24	24341	6266192	257	155839	40
395P20	12266	3174401	258	112909	28
Total	226019	58226627	259	1436179	41

Appendix XXIV Annotation of genes located on the BAC sequences harbouring the fatty acid desaturase loci *sad1*, *sad2*, *fad2a*, *fad2b*, *fad3a* and *fad3b* of CDC Bethune and M5791

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
Sad1									
1	BAC317I7_1	2825	3911	+	273	2	90	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
2	BAC317I7_2	12980	14847	+	882	3	293	51% identity with hypothetical protein POPTR_0015s13190g (*Populus trichocarpa*) XP_002321912.2(E=1e-84), EST-78% identity over 239 bp of GW864295.1(E=1e-32)	MYB-LIKE DNA-BINDING PROTEIN MYB- Transcription factor, Myb superfamily (Lus10027458)
3	BAC317I7_3	19306	19798	+	363	1	120	48% identity with gag-pol polyprotein (*Phaseolus vulgaris*) AAR13298.1 (E= 2e-28), EST-No significant similarity found	Not found in flax GFF3
4	BAC317I7_4	20422	23775	+	867	3	288	100% identity with hypothetical protein (*Linum usitatissimum*) AFN53628.1(E=0), EST-99% identity over 357 bp of JG233726.1(E=0)	MYB-LIKE DNA-BINDING PROTEIN MYB-Transcription factor, Myb superfamily (Lus10027459)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
5	BAC317I7_5	24360	26161	–	603	5	200	87% identity with hypothetical protein (*Linum usitatissimum*) AFN53683.1(E= 9e-86), EST-99% identity over 476 bp of JG101327.1(E= 0)	Remorin, C-terminal region (Lus10027460)
6	BAC317I7_6	30629	32194	–	513	2	170	95% identity with hypothetical protein (*Linum usitatissimum*) AFN53683.1(E= 1e-83), EST-99% identity over 505 bp of JG264061.1(E= 0)	CYTOCHROME B561-RELATED- Cytochrome b-integral to membrane (Lus10027461)
7	BAC317I7_7	32489	33468	+	360	3	119	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
8	BAC317I7_8	36607	42358	+	3000	11	999	92% identity with putative ATP-binding protein (*Linum usitatissimum*) AFN53629.1 (E= 0), EST-99% identity over 647 bp of CA483297.1(E= 0)	No functional annotations for this locus (Lus10027462)
9	BAC317I7_9	42822	46729	+	1596	8	531	84% identity with hypothetical protein JCGZ_10179 (*Jatropha curcas*) KDP46339.1 (E= 0), EST-95% identity over 420 bp of JG049216.1(E=	NUCLEOTIDE-BINDING PROTEIN NBP35(YEAST)-RELATED- Predicted ATPase, nucleotide-binding (Lus10027463)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								0)	
10	BAC317I7_10	47364	49541	+	930	5	309	97% identity with putative ATP-binding protein (*Linum usitatissimum*) AFN53629.1(E= 2e-164), EST-99% identity over 753 bp of JG216873.1(E= 0)	GCIP-INTERACTING PROTEIN P29- Cyclin D-interacting protein GCIP (Lus10027464)
11	BAC317I7_11	51312	53143	+	1419	1	472	65% identity with pentatricopeptide repeat-containing protein, putative (*Ricinus communis*) XP_002510791.1 (E= 0), EST-100% identity over 267 bp of JG074990.1 (E= 1e-135)	FAMILY NOT NAMED (Lus10027465)
12	BAC317I7_12	53150	54609	+	561	4	186	No significant similarity found, EST-99% identity over 185 bp of JG074990.1 (E= 6e-89)	No functional annotations for this locus (Lus10027466)
13	BAC317I7_13	54639	56845	–	849	4	282	99% identity with putative aquaporin PIP2-8 (*Linum usitatissimum*) AFN53685.1 (E= 0), EST-97% identity over 849 bp of EH791987.1 (E= 0)	AQUAPORIN TRANSPORTER-membrane, transporter activity, transport (Lus10027467)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
14	BAC317I7_14	60030	64212	–	1170	11	389	91% identity with putative aquaporin PIP2-8 (*Linum usitatissimum*) AFN53630.1 (E= 0), EST-No significant similarity found	No functional annotations for this locus (Lus10027468)
15	BAC317I7_15	64255	65710	+	444	3	147	68% identity with hypothetical protein CICLE_v10010015mg (*Citrus clementina*) XP_006450258.1(E= 8e-15), EST-83% identity over 227 bp of JG216873.1 (E=9e-47)	Not found in flax GFF3
16	BAC317I7_16	65717	67193	–	288	3	95	99% identity with putative metal ion-binding protein (*Linum usitatissimum*) AFN53686.1(E= 4e-28), EST-97% identity over 288 bp of JG096228.1(E=8e-135)	No functional annotations for this locus (Lus10027468)
17	BAC317I7_17	67864	69438	–	372	2	123	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
18	BAC317I7_18	71958	77161	–	4305	1	1434	57% identity with unknown (*Oryza sativa* Japonica Group) ACY72569.1(E=0), EST-84% identity over 288	Not found in flax GFF3

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								bp of DR007262.1(E= 4e-29) NCBI-EST	
19	BAC317I7_19	79131	80345	–	639	1	212	56% identity with PREDICTED: ethylene-responsive transcription factor 1B-like (*Nicotiana sylvestris*) XP_009764919.1 (E= 2e-42), EST-92% identity over 95 bp of FS224028.1 (E= 4e-29) NCBI-EST	AP2 domain- sequence-specific DNA binding transcription factor activity, regulation of transcription, DNA-dependent (Lus10027469)
20	BAC317I7_20	84116	88245	+	2421	2	806	45% identity with hypothetical protein VITISV_042091 (*Vitis vinifera*) CAN65591.1 (E= 7e-171), EST-No significant similarity found	Not found in flax GFF3
21	BAC317I7_21	89953	90966	–	429	3	142	100% identity with putative copper ion-binding protein (*Linum usitatissimum*) AFN53631.1(E= 5e-77), EST-99% identity over 429bp of JG152685.1 (E= 0)	COPPER TRANSPORT PROTEIN ATOX1-RELATED-metal ion transport, metal ion binding (Lus10027470)
22	BAC317I7_22	91957	95262	+	1188	5	395	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
									(Lus10027471)
23	BAC317I7_23	96321	102053	+	3294	11	1097	64% identity with KDEL motif-containing protein 1 precursor, putative (*Ricinus communis*) XP_002510788.1 (E=0), EST-100% identity over 554bp of JG084614.1 (E= 0)	KDEL (LYS-ASP-GLU-LEU) CONTAINING - RELATED- Endoplasmic reticulum protein EP58, contains filamin rod domain and KDEL motif (Lus10027472+ Lus10027473)
24	BAC317I7_24	102413	104020	–	414	3	137	100% identity with actin-depolymerizing factor 12 (*Linum usitatissimum*) AFN53632.1 (E=2e-92), EST-78% identity over 554bp of FL825102.1 (E=1e-60), NCBI-EST	COFILIN-RELATED-actin binding, intracellular (Lus10027474)
25	BAC317I7_25	104128	106718	–	951	7	316	88% identity with actin-depolymerizing factor 12 (*Linum usitatissimum*) AFN53688.1 (E=2e-139), EST-99% identity over 451bp of JG131778.1 (E=0)	ZINC FINGER CCHC DOMAIN CONTAINING PROTEIN-nucleic acid binding, zinc ion binding (Lus10027475)
26	BAC317I7_26	106851	107437	+	330	1	109	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
27	BAC317I7_27	109825	112514	+	1350	3	449	100% identity with a putative tubulin beta-1 chain protein (*Linum usitatissimum*) AFN53689.1 (E=0), EST-97% identity over 866bp of JG218340.1 (E=0)	TUBULIN- Tubulin/FtsZ family, GTPase domain and Tubulin C-terminal domain- GTPase activity, protein complex, protein polymerization, GTP binding, GTP catabolic process, microtubule-based process (Lus10027476)
28	**BAC317I7_28**	113480	116096	–	1539	4	512	98% identity with Acyl CoA ligase (*Linum usitatissimum*) AFN53690.1(*E*=0), EST-92% identity over 1049 bp of GW864295.1(*E* = 0)	AMP-binding enzyme, catalytic activity, metabolic process (Lus10027477+ Lus10027478)
	BAC212N17_28	1590	4554	–	1539	4	512		
29	**BAC317I7_29**	116314	117477	+	345	2	114	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC212N17_29	9697	11048	+	303	2	100		
30	**BAC317I7_30**	119810	121161	+	267	2	88	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC212N17_30							No predicted gene corresponds to BAC317I7_30	
31	**BAC317I7_31**	121174	124437	–	1101	2	366	56% identity with putative	Glycosyl hydrolases

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC212N17_31	11061	12486	–	813	1	270	glucan endo-1,3-beta-glucosidase GVI (*Nicotiana sylvestris*) XP_009767610.1, (*E*= $2e^{-91}$)	family 17, hydrolase activity, hydrolyzing O-glycosyl compounds, carbohydrate metabolic process (Lus10027479)
32	**BAC317I7_32**	127937	129829	+	1161	3	386	99% identity with putative cytochrome P450 protein (*Linum usitatissimum*) AFN53633.1 (*E*=0), no EST evidence	FAMILY NOT NAMED-Cytochrome P450-oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, heme binding, iron ion binding, electron carrier activity (Lus10027480)
	BAC212N17_32	17830	21215	+	1470	4	489	92% identity with putative cytochrome P450 protein (*Linum usitatissimum*) AFN53633.1 (*E*=0), EST-94% identity over 428 bp of JG078817.1 (*E* = 0)	
33	**BAC317I7_33**	134813	136280	+	432	1	143	100% identity with putative cytochrome P450 protein (*Linum usitatissimum*) AFN53633.1 (E=$6e^{-86}$), EST-94% identity over 428 bp of JG078817.1 (*E*=$3e^{-180}$)	FAMILY NOT NAMED-Cytochrome P450-oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, heme binding, iron ion binding, electron carrier activity (Lus10027481)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC212N17_33							No predicted gene corresponds to BAC317I7_34	
34	**BAC317I7_G34**	138036	148271	–	4170	34	1389	75% identity with hypothetical protein POPTR_0012s13100g (*Populus trichocarpa*) XP_002318216.1 (*E*=0), EST-95% identity over 858 bp of JG218033.1(*E*=0)	PROPROTEIN CONVERTASE SUBTILISIN/KEXIN-Subtilase family and Tripeptidyl peptidase II-serine-type endopeptidase activity, proteolysis, tripeptidyl-peptidase II (Lus10027482)
	BAC212N17_34	22971	33206	–	4137	34	1378		
35	**BAC317I7_35**	148412	149618	–	579	2	192	83% identity with UDP-glycosyltransferase 1 (*Linum usitatissimum*) AFJ52976.1 (*E*=1e^{-52}), EST-88% identity over 567 bp of JG140008.1 (*E*= 1e^{-124})	No functional annotations for this locus (Lus10027483)
	BAC212N17_35	33347	34927	–	390	1	129		
36	**BAC317I7_36**	149678	151489	–	243	2	80	100% identity with tripeptidyl peptidase II (*Linum usitatissimum*) AFN53634.1 (E=1e-35), EST-93% identity over 243bp of JG239206.1 (E=3e-93)	Not found in flax GFF3

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC212N17_36							No predicted gene corresponds to BAC317I7_37	
37	**BAC317I7_37**	153593	157277	+	1116	3	371	88% identity with D-alanyl-D-alanine endopeptidase, putative (*Ricinus communis*) XP_002513691.1 (*E*=0), EST-96% identity over 764 bp of JG106645.1 (*E*=0)	SIGNAL PEPTIDE PEPTIDASE- integral to membrane, aspartic-type endopeptidase activity (Lus10027484)
	BAC212N17_37	38530	42214	+	1116	3	371		
38	**BAC317I7_38**	157389	158677	–	366	3	121	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC212N17_38							No predicted gene corresponds to BAC317I7_39	
39	**BAC317I7_39**	158809	161219	+	1338	5	445	78% calmodulin-binding heat-shock protein, putative (*Ricinus communis*) XP_002513688.1(*E*=0),	CALMODULIN-BINDING HEAT-SHOCK PROTEIN-Lipase (class 3) and

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC212N17_39	43333	46156	+	1338	5	445	EST-100% identity over 937 bp of JG217705.1 (*E*=0)	Lipase 3 N-terminal region- lipid catabolic process, carboxylesterase activity, triglyceride lipase activity (Lus10027485)
40 *	**BAC317I7_40**	162459	165444	+	1191	3	396	100% identity with stearoyl-acyl carrier protein desaturase (*Linum usitatissimum*) CAA07349.1(*E*=0), EST-97% identity over 947 bp of GW865443.1 (*E*=0)	FATTY ACID DESATURASE-Acyl-[acyl-carrier-protein] desaturase- desaturase activity, oxidation reduction process, fatty acid metabolic process (Lus10027486)
	BAC212N17_40	47396	50382	+	1191	3	396		
41	**BAC317I7_41**	165474	166499	–	336	2	111	89% hypothetical protein MTR_1g100890 (*Medicago truncatula*) XP_003592261.1 (*E*=1e^{-08}), EST-93% identity over 207 bp of JG036658.1 (*E*=0)	No functional annotations for this locus (Lus10027487)
	BAC212N17_41	50412	50804	–	207	1	68		
42	**BAC317I7_42**	166683	171044	+	2448	18	815	92% identity with putative chloroplastic acyl-acyl carrier desaturase protein	FAMILY NOT NAMED-Kinesin motor domain-microtubule motor

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC212N17_42	51458	55982	+	2247	17	748	(*Linum usitatissimum*) AFN53635.1 (*E*=0), EST-96% identity over 604 bp of JG256682.1 (*E*=0)	activity, ATP binding, microtubule-based movement (Lus10027488)
43	BAC212N17_43	56009	58508	—	1566	6	521	52% identity with hypothetical protein JCGZ_26090 (*Jatropha curcas*) KDP22259.1 (E= 2e-138), EST-92% identity over 352 bp of GW864249.1(E= 2e-142)	No functional annotations for this locus (Lus10027489)
44	BAC212N17_44	59212	61084	—	750	3	249	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
45	BAC212N17_45	61334	65745	+	2292	10	763	56% identity with hypothetical protein JCGZ_26089 (*Jatropha curcas*) KDP22258.1 (E=0), EST-99% identity over 542 bp of JG024335.1 (E=0)	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE- ubiquitin ligase complex, protein ubiquitination, protein kinase activity, ubiquitin-protein ligase activity, protein phosphorylation (Lus10027490)
46	BAC212N17_46	67641	71897	+	2070	9	689	52% identity with hypothetical protein JCGZ_26088 (*Jatropha*	No functional annotations for this locus (Lus10027491)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								curcas) KDP22257.1 (E= 0), EST-97% identity over 206 bp of EB713711.1 (E= 1e-91)	
47	BAC212N17_47	72227	74237	–	915	2	304	67% identity with hypothetical protein POPTR_0006s24590g (*Populus trichocarpa*) XP_002308564.1 (E=5e-106), EST-94% identity over 597 bp of JG212960.1 (E=0)	Protein of unknown function (DUF1230) (Lus10027492)
48	BAC212N17_48	74842	76660	+	618	1	205	54% identity with Uncharacterized protein TCM_015201 (*Theobroma cacao*) XP_007038735.1 (E=8e-41), EST-99% identity over 485 bp of JG086866.1 (E=0)	No functional annotations for this locus (Lus10027495)
49	BAC212N17_49	76855	82269	–	2859	12	952	75% identity with hypothetical protein JCGZ_26084 (*Jatropha curcas*) KDP22253.1 (E=0), EST-97% identity over 574 bp of JG105720.1 (E=0)	MITOGEN-ACTIVATED KINASE KINASE KINASE-protein kinase activity, ATP binding, protein phosphorylation (Lus10027496)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
50	BAC212N17_50	84232	87056	–	1509	3	502	74% identity with PREDICTED: putative transporter arsB (*Nicotiana sylvestris*) XP_009763772.1 (E=0), EST-77% identity over 600 bp of EY810625.1 (E=4e-85) NCBI-EST	SOLUTE CARRIER FAMILY 13 MEMBER-citrate transmembrane transporter activity, citrate transport, integral to membrane, transmembrane transport (Lus10027497)
51	BAC212N17_51	87849	89152	+	432	2	143	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
52	BAC212N17_52	89731	91496	–	819	1	272	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
53	BAC212N17_53	92005	97866	–	1659	2	552	64% identity with hypothetical protein JCGZ_26079 (*Jatropha curcas*) KDP22248.1 (E=0), EST-99% identity over 485 bp of JG086866.1 (E=0)	No functional annotations for this locus (Lus10027498)
54	BAC212N17_54	100053	101095	+	336	1	111	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
55	BAC212N17_55	103793	107053	+	873	6	290	63% identity with PREDICTED: sec-independent protein translocase protein TATB, chloroplastic-like (*Citrus*	GAG-POL-RELATED RETROTRANSPOSON-protein transport, actin filament polymerization, cytoskeleton, protein

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								sinensis) XP_006490440.1 (E=6e-52), EST-94% identity over 621 bp of JG181065.1 (E=0)	transporter activity (Lus10027499)
56	BAC212N17_56	107223	109659	–	1365	3	454	99% identity with defective in cuticular ridge 1 (*Linum* usitatissimum) AHA57444.1 (E=0), EST-88% identity over 610 bp of GW866355.1 (E=0)	Transferase family-transferase activity, transferring acyl groups other than amino-acyl groups (Lus10027500)
Sad2									
1	BAC375N24_1	3154	5101	+	294	1	97	74% identity with F-box family protein (*Populus trichocarpa*) XP_002323765.1 (E= 4e-41), EST-No significant similarity found	Not found in flax GFF3
2	BAC375N24_2	8307	11423	+	873	3	290	100% identity with Myb-like DNA-binding domain protein (*Linum usitatissimum*) AFN53682.1 (E= 2e-151), EST-100% identity over 457 bp of JG237984.1 (E=0)	MYB-LIKE DNA-BINDING PROTEIN MYB-Transcription factor, Myb superfamily (Lus10039214)
3	BAC375N24_3	11532	13153	–	612	5	203	100% identity with hypothetical protein (*Linum*	Remorin, C-terminal region and Remorin, N-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								usitatissimum) AFN53683.1 (E= 2e-92), EST-100% identity over 612bp of GW867454.1 (E=0)	terminal region (Lus10039215)
4	BAC375N24_4	17706	19840	–	696	4	231	98% identity with hypothetical protein (*Linum usitatissimum*) AFN53683.1 (E=4e-123), EST-95% identity over 558bp of JG264061.1 (E=0)	CYTOCHROME B561-RELATED- integral to membrane (Lus10039216)
5	BAC375N24_5	21015	22288	–	534	3	177	Predicted by FGENESH, No significant similarity found	ASPARTYL PROTEASES- aspartic-type endopeptidase activity, proteolysis (Lus10039217)
6	BAC375N24_6	28524	31900	+	2673	7	890	98% identity with putative ATP-binding protein (*Linum usitatissimum*) AFN53684.1 (E=0), EST-97% identity over 647bp of CA483297.1 (E=0)	No functional annotations for this locus (Lus10039218)
7	BAC375N24_7	34579	36388	+	1083	3	360	87% identity with hypothetical protein JCGZ_10179 (*Jatropha* curcas) KDP46339.1 (E=0), EST-100% identity over	NUCLEOTIDE-BINDING PROTEIN NBP35(YEAST)-RELATED- Predicted ATPase, nucleotide-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								424bp of JG049216.1 (E=0)	binding (Lus10039219)
8	BAC375N24_8	37496	39853	+	1008	5	335	97% identity with putative ATP-binding protein (*Linum usitatissimum*) AFN53684.1 (E=0), EST-83% identity over 601bp of XM_002305492.2 (E=4e-151) NCBI-EST	GCIP-INTERACTING PROTEIN P29- Cyclin D-interacting protein GCIP, pre-mRNA-splicing factor SYF2 (Lus10039220)
9	BAC375N24_9	40301	41197	–	246	1	81	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
10	BAC375N24_10	41414	44739	+	1470	2	489	65% identity with pentatricopeptide repeat-containing protein, putative (*Ricinus communis*) XP_002510791.1 (E=0), EST-98% identity over 190bp of JG074990.1 (E=6e-88)	FAMILY NOT NAMED, PPR repeat (Lus10039221)
11	BAC375N24_11	45522	50347	–	1161	6	386	99% identity with putative aquaporin PIP2-8 (*Linum usitatissimum*) AFN53630.1 (E=0), EST-100% identity over 841bp of EH791987.1 (E=0)	AQUAPORIN TRANSPORTER-membrane, transporter activity, Transport (Lus10039222)
12	BAC375N24_12	52014	56966	–	1305	12	434	94% identity with putative	Protein of unknown

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								aquaporin PIP2-8 (*Linum usitatissimum*) AFN53685.1 (E=0), EST-79% identity over 136bp of DC587880.1 (E=1e-15) (NCBI-EST)	function (DUF1624) (Lus10039223)
13	BAC375N24_13	59086	60324	–	288	3	95	100% identity with putative metal ion-binding protein (*Linum usitatissimum*) AFN53686.1 (E=1e-28), EST-100% identity over 288bp of JG096228.1 (E=4e-148)	COPPER TRANSPORT PROTEIN ATOX1-RELATED (Lus10039224)
14	BAC375N24_14	61886	62836	+	399	1	132	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
15	BAC375N24_15	63197	64542	–	786	1	261	55% identity with PREDICTED: ethylene-responsive transcription factor 1B (*Solanum lycopersicum*) XP_004247350.1 (E=4e-42), EST-87% identity over 95bp of FS224028.1 (E=2e-20) (NCBI-EST)	Not found in flax GFF3
16	BAC375N24_16	64771	65516	+	306	2	101	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
17	BAC375N24_17	69540	73506	—	1356	6	451	99% identity with hypothetical protein (*Linum usitatissimum*) AFN53687.1 (E=1e-66), EST-98% identity over 433bp of JG105784.1 (E=0)	COPPER TRANSPORT PROTEIN ATOX1-RELATED- metal ion transport, metal ion binding (Lus10039225)
18	BAC375N24_18	77311	79000	—	471	2	156	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
19	BAC375N24_19	79105	81531	+	1623	6	540	68% identity with KDEL motif-containing protein 1 precursor, putative (*Ricinus communis*) XP_002510788.1 (E=0), EST-98% identity over 802bp of GW868434.1 (E=0)	KDEL (LYS-ASP-GLU-LEU) CONTAINING - RELATED- Endoplasmic reticulum protein EP58, contains filamin rod domain and KDEL motif (Lus10039226+ Lus10039227)
20	BAC375N24_20	81664	84689	+	1545	5	514	72% identity with hypothetical protein JCGZ_26108 (*Jatropha curcas*) KDP22277.1 (E=0), EST-77% identity over 642bp of CK117680.1 (E=5e-89) (NCBI-EST)	KDEL (LYS-ASP-GLU-LEU) CONTAINING - RELATED- Endoplasmic reticulum protein EP58, contains filamin rod domain and KDEL motif (Lus10039228)
21	BAC375N24_21	85890	86378	—	393	2	130	88% identity with putative actin-depolymerizing factor 12 (*Linum usitatissimum*)	COFILIN-RELATED- actin binding, intracellular

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								AFN53688.1 (E=0), EST-94% identity over 757bp of JG131778.1 (E=0)	(Lus10039229)
22	BAC375N24_22	87695	89346	–	804	5	267	99% identity with putative actin-depolymerizing factor 12 (*Linum usitatissimum*) AFN53688.1 (E=2e-161), EST-94% identity over 757bp of JG131778.1 (E=0)	ZINC FINGER CCHC DOMAIN CONTAINING PROTEIN- nucleic acid binding, zinc ion binding (Lus10039230)
23	BAC375N24_23	89871	94410	+	1995	5	664	90% identity with putative tubulin beta-1 chain protein (*Linum usitatissimum*) AFN53689.1 (E=0), EST-100% identity over 638bp of JG218340.1 (E=0)	TUBULIN- GTPase activity, protein complex, protein polymerization, GTP binding, microtubule, GTP catabolic process (Lus10039231)
24	BAC375N24_24	94609	97206	+	249	3	82	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
25	BAC375N24_25	97253	99917	–	1539	4	512	100% identity with acyl CoA ligase (*Linum usitatissimum*) AFN53690.1 (E=0), EST-95% identity over 967bp of GW864295.1 (E=0)	FAMILY NOT NAMED-catalytic activity, metabolic process ((Lus10039232)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
26	BAC375N24_26	102252	104385	–	1038	2	345	52% identity with PREDICTED: putative glucan endo-1,3-beta-glucosidase GVI (*Nicotiana sylvestris*) XP_009767610.1 (E=1e-108), EST-No significant similarity found	FAMILY NOT NAMED-hydrolase activity, hydrolyzing O-glycosyl compounds, carbohydrate metabolic process (Lus10039233)
27	BAC375N24_27	113566	116866	+	1716	5	571	100% identity with cytochrome P450 (*Linum usitatissimum*) AFN53691.1 (E=0), EST-99% identity over 417bp of JG078817.1 (E=0)	FAMILY NOT NAMED-Cytochrome P450, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, heme binding, iron ion binding (Lus10039234)
28	BAC375N24_28	117577	127342	–	4002	34	1333	77% identity with hypothetical protein POPTR_0012s13100g (*Populus trichocarpa*) XP_002318216.1 (E=0), EST-100% identity over 840bp of JG218033.1 (E=0)	PROPROTEIN CONVERTASE SUBTILISIN/KEXIN-Subtilase family, Tripeptidyl peptidase II-serine-type endopeptidase activity, proteolysis (Lus10039235)
29	BAC375N24_29	128249	128430	–	183	1	60	100% identity with pectinacetylesterase (*Linum*	No functional annotations for this locus

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								usitatissimum) AFN53693.1 (E=6e-34), EST-No significant similarity found	(Lus10039236)
30	**BAC375N24_30**	**129037**	**130226**	**–**	**870**	**2**	**289**	pectinacetylesterase (*Linum usitatissimum*), E=0, 100%, AFN53693.1, EST-No significant similarity found	GLUCOSYL/GLUCURONOSYL TRANSFERASES-transferase activity, transferring hexosyl groups, metabolic process (Lus10039237)
	BAC6M22_30	**4971**	**6160**	**–**	**870**	**2**	**289**		
31	**BAC375N24_31**	130529	137617	–	1299	12	432	80% identity with pectinacetylesterase (*Linum usitatissimum*) AFN53693.1 (E=0), EST-99% identity over 539 bp of JG225748.1 (*E*=0)	NOTUM-RELATED-Pectin acetylesterase and similar proteins (Lus10039238)
	BAC6M22_31	6463	11825	–	1143	11	380	69% identity with pectinacetylesterase (*Linum usitatissimum*) AFN53693.1 (E=$2e^{-140}$), EST-99% identity over 539 bp of JG225748.1 (*E*=0)	
32	**BAC375N24_32**	138176	142853	+	1116	3	371	87% identity with D-alanyl-D-alanine endopeptidase, putative (*Ricinus communis*) XP_002513691.1 (*E*=0),	SIGNAL PEPTIDE PEPTIDASE-integral to membrane, aspartic-type endopeptidase activity
	BAC6M22_32	14105	18783	+	1116	3	371		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								EST-99% identity over 764 bp of JG106645.1 (*E*=0)	(Lus10039239)
33	**BAC375N24_33**	142938	146495	+	1740	7	579	75% identity with calmodulin-binding heat-shock protein, putative (*Ricinus communis*) XP_002513688.1 (*E*=0), EST-95% identity over 634 bp of JG217705.1 (*E*=0)	CGI-141-RELATED/LIPASE CONTAINING PROTEIN-Lipase (class 3), Lipase 3 N-terminal region- lipid catabolic process, carboxylesterase activity, triglyceride lipase activity (Lus10039240)
	BAC6M22_33	18868	22424	+	1251	5	416		
34 *	**BAC375N24_34**	147244	150152	+	1191	3	396	100% identity with stearoyl ACP desaturase 2 (*Linum usitatissimum*) AFJ53117.1 (*E*=0), EST-95% identity over 947 bp of GW865443.1 (*E*=0)	FATTY ACID DESATURASE-acyl-[acyl-carrier-protein] desaturase activity, oxidation-reduction process, fatty acid metabolic process (Lus10039241)
	BAC6M22_34	23173	26081	+	1191	3	396		
35	**BAC375N24_35**	150295	150942	—	207	1	68	77% identity with hypothetical protein TRIUR3_23694 (*Triticum*	No functional annotations for this locus (Lus10039242)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC 6M22_35	26224	26871	–	207	1	68	*urartu*) EMS54319.1(E=3e^{-08}), EST-100% identity over 947 bp of JG036658.1 (*E*=3e^{-108})	
36	**BAC375N24_36**	151002	155212	+	2376	18	791	91% identity with class 3 lipase (*Linum usitatissimum*) AFN53694.1 (*E*=0), EST-99% identity over 604 bp of JG256682.1 (*E*=0)	FAMILY NOT NAMED-Kinesin motor domain-microtubule motor activity, ATP binding, microtubule-based movement (Lus10039243)
	BAC6M22_36	27088	31107	+	2373	18	790		
37	BAC6M22_37	31316	31829	–	429	2	142	65% identity with ATP binding protein, putative (*Theobroma cacao*) XP_007038731.1 (E=2e-29), EST-92% identity over 417bp of GH171786.1 (E=7e-177)	No functional annotations for this locus (Lus10039244)
38	BAC6M22_38	32265	36577	+	2310	11	769	49% identity with hypothetical protein JCGZ_26088 (*Jatropha curcas*) KDP22257.1 (E=2e-171), EST-83% identity over 243bp of GH171786.1 (E=2e-53) (NCBI-EST)	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE- ubiquitin ligase complex, protein ubiquitination, protein kinase activity, ubiquitin-protein ligase activity,

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
									protein phosphorylation (Lus10039245)
39	BAC6M22_39	38753	42223	+	1854	8	617	56% identity with hypothetical protein JCGZ_26089 (*Jatropha curcas*) KDP22258.1 (E=0), EST-94% identity over 542bp of JG024335.1 (E=0)	No functional annotations for this locus (Lus10039246)
40	BAC6M22_40	42381	44343	–	822	3	273	76% identity with hypothetical protein POPTR_0006s24590g (*Populus trichocarpa*) XP_002308564.1 (E=9e-115), EST-99% identity over 542bp of JG229613.1 (E=0)	Protein of unknown function (DUF1230) (Lus10039247)
41	BAC6M22_41	44397	45408	+	765	1	254	53% identity with PREDICTED: uncharacterized protein LOC103333189 (*Prunus mume*) XP_008234208.1 (E=6e-66), EST-No significant similarity found	No functional annotations for this locus (Lus10039248)
42	BAC6M22_42	45428	46442	+	621	1	206	47% identity with unknown (*Lotus japonicas*) AFK37681.1 (E=9e-48),	No functional annotations for this locus (Lus10039249)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								EST-100% identity over 496bp of JG141047.1 (E=0)	
43	BAC6M22_43	46584	53102	–	2835	12	944	77% identity with hypothetical protein JCGZ_26084 (*Jatropha curcas*) KDP22253.1 (E=0), EST-99% identity over 576bp of JG105720.1 (E=0)	MITOGEN-ACTIVATED KINASE KINASE KINASE-protein kinase activity, ATP binding, protein phosphorylation (Lus10039250)
44	BAC6M22_44	53339	55972	–	1512	3	503	75% identity with PREDICTED: putative transporter arsB-like (*Solanum tuberosum*) XP_006363328.1 (E=0), EST-No significant similarity found	SOLUTE CARRIER FAMILY 13 MEMBER-citrate transmembrane transporter activity, citrate transport, integral to membrane, transmembrane transport (Lus10039251)
45	BAC6M22_45	56278	57573	–	624	2	207	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus (Lus10039252)
46	BAC6M22_46	57609	60590	+	1572	4	523	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
47	BAC6M22_47	61063	62405	+	516	1	171	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
48	BAC6M22_48	62611	66355	–	1725	3	574	62% identity with hypothetical protein	No functional annotations for this locus

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								JCGZ_26079 (*Jatropha curcas*) KDP22248.1 (E=0), EST-99% identity over 576bp of JG239469.1 (E=0)	(Lus10039253)
49	BAC6M22_49	71256	75292	+	1032	7	343	43% identity with unknown (*Populus trichocarpa*) ABK94883.1 (E=7e-35), EST-100% identity over 346bp of JG181065.1 (E=8e-180)	mttA/Hcf106 family-protein transporter activity (Lus10039254)
50	BAC6M22_50	75305	77608	–	1401	3	466	92% identity with defective in cuticular ridge 1 (*Linum usitatissimum*) AHA57444.1 (E=0), EST-96% identity over 638bp of GW866355.1 (E=0)	Transferase family-transferase activity, transferring acyl groups other than amino-acyl groups (Lus10039255+ Lus10039256)
51	BAC6M22_51	78924	82627	+	366	4	121	No significant similarity found, EST-100% identity over 638bp of JG076221.1 (E=9e-86)	No functional annotations for this locus (Lus10039257)
52	BAC6M22_52	82925	84871	–	1155	2	384	60% identity with conserved hypothetical protein (*Ricinus communis*) XP_002513670.1 (E=4e-131), EST-No significant	Protein of unknown function (DUF677) (Lus10039258)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								similarity found	
53	BAC6M22_53	85416	86093	+	390	1	129	46% identity with conserved hypothetical protein (*Ricinus communis*) PREDICTED: uncharacterized protein LOC101297786 (*Fragaria vesca* subsp. vesca) XP_004308652.1 (E=3e-20), EST-No significant similarity found	Not found in flax GFF3
54	BAC6M22_54	86450	87649	–	858	1	285	69% identity with hypothetical protein EUGRSUZ_A02756 (*Eucalyptus grandis*) KCW90655.1 (E=7e-55), EST-94% identity over 201bp of JG129365.1 (E=5e-77)	C2 domain-protein binding (Lus10039259)
55	BAC6M22_55	88618	91048	+	942	7	313	97% identity with fibrillarin, putative (*Ricinus communis*) XP_002513667.1 (E=3e-168), EST-97% identity over 613bp of JG199575.1 (E=0)	RRNA 2-O-METHYLTRANSFERASE FIBRILLARIN- RNA binding, rRNA processing, tRNA processing, methyltransferase activity

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
									(Lus10039260)
56	BAC6M22_56	91063	91658	–	324	2	107	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus (Lus10039261)
57	BAC6M22_57	94309	95965	–	306	2	101	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus (Lus10039262)
58	BAC6M22_58	99049	102232	–	1476	5	491	76% identity with multidrug resistance pump, putative (*Ricinus communis*) XP_002513666.1 (E=0), EST-100% identity over 581bp of JG060686.1 (E=0)	MULTIDRUG RESISTANCE PROTEIN- drug transmembrane transporter activity, antiporter activity, membrane, transmembrane transport, drug transmembrane transport (Lus10039263)
59	BAC6M22_59	115078	115959	+	393	2	130	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus (Lus10039264)
60	BAC6M22_60	116900	120647	+	2046	1	681	57% identity with hypothetical protein JCGZ_26073 (*Jatropha curcas*) KDP22242.1 (E=0), EST-93% identity over 151bp of JG269786.1	GRAS domain family ((Lus10039265)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								(E=2e-54)	
61	BAC6M22_61	121358	122726	–	351	3	116	65% identity with PREDICTED: non-functional NADPH-dependent codeinone reductase 2 (*Vitis vinifera*) XP_002285202.1 (E=2e-10), EST-83% identity over 70bp of CB655626.1 (E=7e-07) (NCBI-EST)	ALDO/KETO REDUCTASE-oxidoreductase activity, oxidation-reduction process ((Lus10039266)
62	BAC6M22_62	123942	128365	+	1656	11	551	87% identity with hypothetical protein JCGZ_26071 [Jatropha curcas] KDP22240.1 (E=0), EST-83% identity over 801bp of JG639862.1 (E=0)	TRNA SYNTHETASE-RELATED- nucleotide binding, aminoacyl-tRNA ligase activity, proline-tRNA ligase activity, tRNA aminoacylation for protein translation, prolyl-tRNA aminoacylation (Lus10039267)
63	BAC6M22_63	128437	131442	–	2160	8	719	55% identity with PREDICTED: ferric reduction oxidase 2 (*Vitis vinifera*) XP_002272804.1 (E=0), EST-No significant	NADPH OXIDASE-integral to membrane, oxidoreductase activity, flavin adenine dinucleotide binding, iron

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								similarity found	ion binding, electron carrier activity (Lus10039268)
64	BAC6M22_64	132159	137416	+	1359	8	452	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
65	BAC6M22_65	137894	138409	+	360	1	119	44% identity with PREDICTED: non-specific lipid-transfer protein C, cotyledon-specific isoform-like (*Cleome hassleriana*) XP_010537417.1 (E=1e-17), EST-No significant similarity found	Protease inhibitor/seed storage/LTP family (Lus10039270)
Fad2a									
1	BAC346C18_1	469	1457	–	824	3	277	98% identity with putative Fe(II) oxygenase superfamily protein (*Linum usitatissimum*) AFN53719.1 (E= 3e-180), EST-80% identity over 131 bp of EH042969.1 (E=2e-16) (NCBI-EST)	FAMILY NOT NAMED-integral to membrane, ATPase activity, ATPase activity, coupled to transmembrane movement of substances, transmembrane transport, ATP binding, transport (Lus10011976)
2	BAC346C18_2	1732	3141	–	915	2	304	59% identity with multidrug resistance protein	FAMILY NOT NAMED-ATPase activity, ATP binding

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								1, 2, putative (*Ricinus communis*) XP_002519757.1 (E= 5e-75), EST-No significant similarity found	(Lus10011977)
3	BAC346C18_3	3412	4482	–	582	3	193	50% identity with hypothetical protein VITISV_009891 (*Vitis vinifera*) CAN77320.1 (E= 1e-52), EST-No significant similarity found	FAMILY NOT NAMED-integral to membrane, ATPase activity, ATPase activity, coupled to transmembrane movement of substances, transmembrane transport, ATP binding, transport (Lus10011978)
4	BAC346C18_4	6790	8974	–	1005	4	334	58% identity with hypothetical protein JCGZ_21657 (*Jatropha curcas*) KDP21186.1 (E= 3e-139), EST-96% identity over 347 bp of JG036975.1 (E=1e-158)	OXIDOREDUCTASE, 2OG-FE(II) OXYGENASE FAMILY PROTEIN-oxidoreductase activity (Lus10011979)
5	BAC346C18_5	9223	11067	–	1065	4	354	86% identity with putative Fe(II) oxygenase superfamily protein (*Linum usitatissimum*) AFN53719.1	OXIDOREDUCTASE, 2OG-FE(II) OXYGENASE FAMILY PROTEIN-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								(E= 1e-162), EST-99% identity over 584 bp of JG204147.1 (E=0)	oxidoreductase activity(Lus10011980)
6	BAC346C18_6	11847	14328	–	1158	5	385	94% identity with putative Fe(II) oxygenase superfamily protein (*Linum usitatissimum*) AFN53719.1 (E= 0), EST-99% identity over 584 bp of JG204147.1 (E=0)	OXIDOREDUCTASE, 2OG-FE(II) OXYGENASE FAMILY PROTEIN- oxidoreductase activity (Lus10011981)
7	BAC346C18_7	15967	18307	–	864	3	287	59% identity with xyloglucan endo-1 family protein (*Populus trichocarpa*) XP_002317309.1 (E=4e-101), EST-No significant similarity found	GLYCOSYL HYDROLASE-RELATED-xyloglucan:xyloglucosyl transferase activity, hydrolase activity, hydrolyzing O-glycosyl compounds, apoplast, cell wall, carbohydrate metabolic process, cellular glucan metabolic process (Lus10011982)
8	BAC346C18_8	18566	19343	–	303	1	100	64% identity with Xyloglucan endotransglucosylase/hydrol	Xyloglucan endo-transglycosylase (XET) C-terminus- apoplast, cell

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								ase 16 (*Theobroma cacao*) XP_007025220.1 (E=4e-38), EST-84% identity over 98bp of EX269128.1 (E=8e-16) -NCBI- EST	wall, cellular glucan metabolic process (Lus10011983)
9	BAC346C18_9	19491	20456	–	435	3	144	73% identity with xyloglucan endo-1 family protein (*Populus trichocarpa*) XP_002317310.1 (E=2e-58), EST-76% identity over 98bp of JG168586.1 (E=4e-25)	GLYCOSYL HYDROLASE-RELATED- hydrolase activity, hydrolyzing O-glycosyl compounds, carbohydrate metabolic process (Lus10011984)
10	BAC346C18_10	20501	22314	–	1125	4	374	72% identity with hypothetical protein JCGZ_21651 (*Jatropha curcas*) KDP21180.1 (E=0), EST-100% identity over 600bp of JG218249.1 (E=0)	OXIDOREDUCTASE, 2OG-FE(II) OXYGENASE FAMILY PROTEIN-oxidoreductase activity (Lus10011985)
11	BAC346C18_11	27292	32156	+	1503	6	500	44% identity with putative Fe(II) oxygenase superfamily protein (*Linum usitatissimum*) AFN53719.1 (E=1e-93), EST-No significant similarity found	OXIDOREDUCTASE, 2OG-FE(II) OXYGENASE FAMILY PROTEIN-oxidoreductase activity (Lus10011986)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
12	BAC346C18_12	35190	38893	+	1842	9	613	57% identity with Flavonol synthase/flavanone 3-hydroxylase, putative (*Ricinus communis*) XP_002519770.1 (E=2e-133), EST-No significant similarity found	OXIDOREDUCTASE, 2OG-FE(II) OXYGENASE FAMILY PROTEIN- oxidoreductase activity (Lus10011987)
13	BAC346C18_13	39196	39722	–	354	2	117	42% identity with DNA-binding bromodomain-containing family protein (*Populus trichocarpa*) XP_002300324.2 (E=2e-05), EST-100% identity over 236bp of JG208453.1 (E=4e-119)	EUKARYOTIC TRANSLATION INITIATION FACTOR 4E RELATED- RNA binding, translation initiation factor activity, cytoplasm, translational initiation (Lus10011988)
14	BAC346C18_14	40383	43384	–	1221	9	406	86% identity with putative bromodomain-containing protein (*Linum usitatissimum*) AFN53717.1 (E=6e-98), EST-100% identity over 521bp of GW864352.1 (E=0)	FALZ-RELATED BROMODOMAIN-CONTAINING PROTEINS- protein binding (Lus10011989)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
15	**BAC346C18_15**	45854	48815	–	753	5	250	74% identity with PREDICTED: bidirectional sugar transporter SWEET2a (*Vitis vinifera*), XP_002269484.1 (E=4e^{-89}), EST-96% identity over 753 bp of JG217941.1 (E=0)	RAG1-ACTIVATING PROTEIN 1-Sugar efflux transporter for intercellular exchange (Lus10011990)
	BAC139G15_15	2789	5468	–	753	5	250		
16	**BAC346C18_16**	49959	55533	+	1500	10	499	72% identity with hypothetical protein EUTSA_v10006968mg (*Eutrema salsugineum*), XP_006415560.1 (E=6e^{-149}), EST-85% identity over 713 bp of JG038739.1 (E=8e^{-82})	XP-G/RAD2 DNA REPAIR ENDONUCLEASE FAMILY-nuclease activity, DNA repair (Lus10011991)
	BAC139G15_16	6844	12195	+	1671	12	556		
17	**BAC346C18_17**	56035	57787	+	810	7	269	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus (Lus10011992)
	BAC139G15_17	12697	16272	+	1146	9	381		
18	**BAC346C18_18**	65637	67542	+	1485	1	494	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC139G15_18							No predicted gene corresponds to BAC346C18_18	
19	**BAC346C18_19**	67779	68626	+	201	1	66	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC139G15_19	16374	16784	+	201	1	66		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
20	**BAC346C18_20**	71349	71921	+	345	1	114	100% identity with putative DUF584 protein (*Linum usitatissimum*), AFN53716.1 (E=1e^{-56}), EST-98% identity over 644 bp of GW866122.1 (E=4e^{-163})	Senescence regulator (Lus10011993)
	BAC139G15_20	19085	19657	–	345	1	114		
21	**BAC346C18_21**	73360	74211	+	207	1	68	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC139G15_21	21096	21943	+	207	1	68		
22	**BAC346C18_22**	75675	78767	+	1107	4	368	Predicted by FGENESH, No significant similarity found	No functional annotations for this locus (Lus10011994)
	BAC139G15_22	23430	26522	+	1107	4	368		
23	**BAC346C18_23**	80016	81468	+	1017	3	338	65% identity with hypothetical protein JCGZ_21640 (*Jatropha curcas*) KDP21169.1 (E=9e^{-167})	ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN- hydrolase activity, acting on ester bonds, lipid metabolic process (Lus10011996)
	BAC139G15_23	27840	29291	+	1149	3	382		
24	**BAC346C18_24**	81664	82342	+	300	1	99	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC139G15_24	29487	30164	+	300	1	99		
25	**BAC346C18_25**	84399	86791	+	1119	3	372	80% identity with PREDICTED: GDSL esterase/lipase At5g45670-	ZINC FINGER FYVE DOMAIN CONTAINING

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC139G15_25	31792	34184	+	1119	3	372	like (*Cucumis melo*), XP_008443704.1 (*E*= 0), EST-99% identity over 731 bp of JG123050.1 (*E*=0)	PROTEIN-hydrolase activity, acting on ester bonds, lipid metabolic process (Lus10011997)
26	**BAC346C18_26**	88099	89438	+	435	1	144	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC139G15_26	35492	36831	+	435	1	144		
27	**BAC346C18_27**	89787	92881	+	1098	4	365	100% identity with putative GDSL-like lipase acylhydrolase protein (*Linum usitatissimum*), AFN53715.1 (*E*= 0)	ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN-hydrolase activity, acting on ester bonds, lipid metabolic process (Lus10011998)
	BAC139G15_27	37180	40193	+	1098	4	365		
28	**BAC346C18_28**	92898	94714	+	1113	5	370	90% identity with putative GDSL-like lipase acylhydrolase protein (*Linum usitatissimum*) AFN53715.1 (*E*=5e^{-164})	ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN-hydrolase activity, acting on ester bonds, lipid metabolic process (Lus10011999)
	BAC139G15_28	40438	42186	+	1113	5	370		
29	**BAC346C18_29**	96531	101163	+	1176	7	391	43% identity with GDSL-like Lipase/Acylhydrolase superfamily protein, putative isoform 2 (*Theobroma cacao*), XP_007014321.1 (*E*=2e^{-50}),	ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN-hydrolase activity, acting on ester bonds, lipid metabolic

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								EST-89% identity over 695 bp of JG083121.1 (E=5e^{-78})	process (Lus10012000 +Lus10012001)
	BAC139G15_29	44003	48600	+	1188	7	395	44% identity with GDSL-like Lipase/Acylhydrolase superfamily protein, putative isoform 2 (*Theobroma cacao*), XP_007014321.1 (E=8e^{-49}), EST-89% identity over 695 bp of JG083121.1 (E=5e^{-78})	
30	**BAC346C18_30**	101446	104823	+	1053	5	350	55% identity with hypothetical protein JCGZ_21635 (*Jatropha curcas*), KDP21164.1 (E=7e^{-116})	ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN-hydrolase activity, acting on ester bonds, lipid metabolic process (Lus10012002)
	BAC139G15_30	48883	52253	+	1053	5	350		
31	**BAC346C18_31**	105379	107964	+	1023	7	340	85% identity with PAP fibrillin protein (*Linum usitatissimum*), AFN53714.1 (E=1e^{-158}), EST-90% identity over 950 bp of GW865201.1 (E=0)	PAP_fibrillin-structural molecule activity, chloroplast (Lus10012003)
	BAC139G15_31	52809	55157	+	1023	7	340		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
32	**BAC346C18_32**	108506	110040		900	7	299	76% identity with PREDICTED: beta-carotene isomerase D27, chloroplastic (*Eucalyptus grandis*), XP_010069831.1 (*E*=6e-118), EST-99% identity over 577bp of JG025709.1 (*E*=0)	No functional annotations for this locus (Lus10012004)
	BAC139G15_32	55740	57459	–	900	7	299		
33	**BAC346C18_33**	110493	112592	–	822	6	273	60% identity with Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein isoform 1 (*Theobroma cacao*), XP_007051133.1 (*E*=8e^{-85}), EST-99% identity over 549 bp of JG264515.1 (*E*=0)	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE-peptidyl-prolyl cis-trans isomerase activity (Lus10012005)
	BAC139G15_33	57797	60062	–	792	7	263	63% identity with Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein isoform 1 (*Theobroma cacao*), XP_007051133.1 (*E*=3e^{-88}), EST-99% identity over 758 bp of JG264515.1 (*E*=0)	

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
34	**BAC346C18_34**	113687	115007	–	936	2	311	72% identity with omega-6 desaturase (*Linum usitatissimum*), AFN53713.1 (E=8e^{-64})	No functional annotations for this locus (Lus10012006)
	BAC139G15_34	60192	62599	–	1068	2	355	93% identity with omega-6 desaturase (*Linum usitatissimum*), AFN53713.1 (E=1e^{-63})	
35 *	**BAC346C18_35**	116028	117239	–	1137	1	378	99% identity with fatty acid desaturase 2 (*Linum usitatissimum*), AFJ53078.1 (E= 0), EST-98% identity over 876 bp of CD760583.1 (E=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation-reduction process - lipid metabolic process (Lus10012007+ Lus10012008)
	BAC139G15_35	63098	64569	–	1137	1	378		
36	**BAC346C18_36**							No predicted gene corresponds to BAC139G15_36	
	BAC139G15_36	66633	67762	–	402	1	133	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
37	**BAC346C18_37**	120683	121428	+	285	2	94	100% identity with Late embryogenesis abundant group 1 (*Linum usitatissimum*), AFN53712.1 (*E*=3e^{-38}), EST-100% identity over 609 bp of JG200127.1 (*E*=2e^{-146})	Late embryogenesis abundant (LEA) group 1-embryo development (Lus10012009)
	BAC139G15_37	67832	68576	+	285	2	94		
38	**BAC346C18_38**	122146	124542	+	1860	2	619	86% identity with late embryogenesis abundant group 1 (*Linum usitatissimum*), AFN53712.1 (*E*=0), EST-100% identity over 610 bp of JG266537.1 (*E*=3e^{-112})	Late embryogenesis abundant (LEA) group 1-embryo development (Lus10012010)
	BAC139G15_38	69297	71692	+	2001	1	666		
39	**BAC346C18_39**	124665	126773	–	1482	4	493	90% identity with putative UDP-glucosyltransferase (*Linum usitatissimum*), AFN53711.1 (*E*=0), EST-91% identity over 230 bp of JG077394.1 (*E*=8e^{-82})	GLUCOSYL/GLUCURONOSYL TRANSFERASES-UDP-transferase activity, transferring hexosyl groups, metabolic process, UDP-glucosyl transferase, UDP-glucosyl transferase (Lus10012011)
	BAC139G15_39	71815	73923	–	1482	4	493		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
40	**BAC346C18_40**	126939	128549	+	1065	2	354	88% identity with hypothetical protein (*Linum usitatissimum*), AFN53710.1 (*E*=0), EST-94% identity over 865 bp of JG177336.1 (*E*=0)	No functional annotations for this locus (Lus10012013)
	BAC139G15_40	74089	75699	+	1098	1	365		
41	**BAC346C18_41**	128701	131956	–	858	6	285	84% identity with oxidoreductase family protein (*Populus trichocarpa*), XP_002312720.1 (*E*=$2e^{-176}$), EST-99% identity over 896 bp of GW867070.1 (*E*=0)	PROLYL 4-HYDROXYLASE ALPHA SUBUNIT-oxidoreductase activity, prolyl 4-hydroxylase (Lus10012014)
	BAC139G15_41	75849	80123	–	897	6	298		
42	**BAC346C18_42**	132505	133121	+	207	1	68	62% identity with KH domain-containing protein / zinc finger family protein isoform 2 (*Theobroma cacao*), XP_007028982.1 (*E*=$2e^{-18}$), EST-87% identity over 110 bp of GW867070.1 (*E*=$5e^{-26}$)	KH DOMAIN CONTAINING RNA BINDING PROTEIN-nucleic acid binding, RNA binding, zinc ion binding (Lus10012016)
	BAC139G15_42	80485	81101	+	207	1	68		
43	**BAC346C18_43**	135541	136490	+	489	1	162	67% identity with PREDICTED: zinc finger CCCH domain-containing	KH DOMAIN CONTAINING RNA BINDING PROTEIN-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC139G15_43	81796	84458	+	945	3	314	protein 36-like (*Cleome hassleriana*) XP_010520038.1 (*E*=1e-58), EST-94% identity over 480bp of EH792333.1 (*E*=0)	nucleic acid binding, RNA binding, zinc ion binding (Lus10012017)
44	**BAC346C18_44**	136841	137803	–	492	2	163	100% identity with hypothetical protein (*Linum usitatissimum*), AFN53709.1 (*E*=2e^{-100}), EST-96% identity over 491 bp of JG192645.1 (*E*=0)	Late embryogenesis abundant (LEA) group 1-embryo development (Lus10012018)
	BAC139G15_44	84809	85771	–	492	2	163		
45	BAC139G15_45	86627	88812	–	1173	7	390	77% identity with PREDICTED: enoyl-[acyl-carrier-protein] reductase [NADH] 1, *chloroplastic-like* (*Populus* euphratica) XP_011033222.1 (*E*=0), EST-87% identity over 820bp of JG185479.1 (*E*=0)	No annotated domains for this protein-Reductases with broad range of substrate specificities, enoyl-[acyl-carrier protein] reductase (Lus10012019)
46	BAC139G15_46	89558	92609	–	915	3	304	51% identity with unnamed protein product (*Vitis vinifera*) CBI30036.3 (*E*=3e-52), EST-89% identity over 476bp of JG158484.1 (*E*=4e-158)	SBP domain-DNA binding, nucleus (Lus10012020)
47	BAC139G15_47	97790	100348	–	1737	2	578	49% identity with	CGI-12 PROTEIN-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								hypothetical protein JCGZ_25523 (*Jatropha curcas*) KDP24607.1 (*E*=2e-169), EST-No significant similarity found	RELATED-Mitochondrial transcription termination factor, mTERF (Lus10012021)
48	BAC139G15_48	100436	104600	+	1953	10	650	53% identity with PREDICTED: protein STRUBBELIG-RECEPTOR FAMILY 2 (*Populus euphratica*) XP_011004818.1 (*E*=0), EST-95% identity over 498bp of JG194149.1 (*E*=0)	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE-protein kinase activity; protein binding, protein phosphorylation (Lus10012022)
Fad2b									
1	**BAC364K11_1**	4414	5726	—	267	1	88	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC28C5_1	1449	2761	—	267	1	88		
2	**BAC364K11_2**	6800	9697	—	927	7	308	73% identity with hypothetical protein POPTR_0001s10500g (*Populus trichocarpa*), XP_002297982.1 (*E*=9e^{-113}), EST-99% identity over 551 bp of JG154733.1 (*E*=0)	TESTIS DEVELOPMENT PROTEIN PRTD-putative glutamine amidotransferase (Lus10021037)
	BAC28C5_2	3834	6731	—	927	7	308		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
3	**BAC364K11_3**	10729	13036	+	1053	8	350	63% identity with Emsy N Terminus/ plant Tudor-like domains-containing protein isoform 3 (*Theobroma cacao*), (E=5e^{-126}), EST-99% identity over 419 bp of JG214971.1 (E=0)	ENT domain-Uncharacterized conserved protein, contains ENT domain (Lus10021038)
	BAC28C5_3	7764	10071	+	1053	8	350		
4	**BAC364K11_4**	13158	14657	–	726	2	241	63% identity with catalytic, putative (*Ricinus communis*), XP_002533454.1 (E=4e^{-86}), EST-99% identity over 627 bp of JG077441.1 (E=0)	X-Pro dipeptidyl-peptidase (S15 family)-aminopeptidase activity, proteolysis(Lus10021039)
	BAC28C5_4	10193	11692	–	726	2	241		
5	**BAC364K11_5**	15927	18601	+	1170	10	389	84% identity with hypothetical protein POPTR_0016s04600g (*Populus trichocarpa*), XP_002322669.1 (E=0), EST-98% identity over 844 bp of JG185479.1 (E=0)	Reductases with broad range of substrate specificities, enoyl-(acyl-carrier protein) reductase (Lus10021040)
	BAC28C5_5	12962	15636	+	1170	10	389		
6	**BAC364K11_6**	18728	20569	–	567	6	188	45% identity with PREDICTED: non-structural maintenance of	UNCHARACTERIZED (Lus10021041)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC28C5_6	15763	17604	–	567	6	188	chromosomes element 4 homolog A-like (*Phoenix dactylifera*), XP_008782862.1 (*E*=2e^{-24})	
7	**BAC364K11_7**	21259	22911	+	1098	1	365	100% identity with hypothetical protein (*Linum usitatissimum*), JG071200.1 (*E*= 0), EST-100% identity over 758 bp of JG071200.1 (*E*=0)	No functional annotations for this locus (Lus10021042)
	BAC28C5_7	18294	19946	+	1098	1	365		
8	**BAC364K11_8**							No predicted gene corresponds to BAC28C5_8	
	BAC28C5_8	20058	21099	+	306	1	101	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
9	**BAC364K11_9**	25099	27142	+	924	3	307	66% identity with conserved hypothetical protein (*Ricinus communis*), XP_002533457.1, (*E*=7e^{-128}) , EST-99% identity over 722 bp of JG180867.1 (*E*=0)	KH DOMAIN CONTAINING RNA BINDING PROTEIN-nucleic acid binding, RNA binding, zinc ion binding (Lus10021043)
	BAC28C5_9	21632	23675	+	924	3	307		
10	**BAC364K11_10**	27248	28289		759	2	252	100% identity with putative seed maturation protein (*Linum usitatissimum*),	Late embryogenesis abundant (LEA) group 1-embryo development

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC28C5_10	23790	24831	–	759	2	252	AFN53649.1 (*E*=3e^{-53}), EST-100% identity over 131 bp of JG192701.1 (*E*=2e^{-60})	(Lus10021044)
11	**BAC364K11_11**	30990	32821	+	1110	1	369	100% identity with omega-6 desaturase (*Linum usitatissimum*), (*E*= 0), EST-84% identity over 736 bp of GW865322.1 (*E*=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process (Lus10021045)
	BAC28C5_11	27532	29362	+	1110	1	369		
12	**BAC364K11_12**	34541	37312	+	1125	1	374	100% identity with omega-6 desaturase (*Linum usitatissimum*), (*E*= 0), EST-86% identity over 834 bp of JG228031.1 (*E*=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process (Lus10021046)
	BAC28C5_12	31082	33853	+	1125	1	374		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
13	**BAC364K11_13**	37894	39609	+	1146	1	381	100% identity with omega-6 desaturase (*Linum usitatissimum*), AFN53646.1 (*E*= 0), EST-92% identity over 836 bp of JG228031.1 (*E*=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process (Lus10021047)
	BAC28C5_G13	34435	36150	+	1146	1	381		
14 *	**BAC364K11_14**	42618	45745	+	1149	1	382	100% identity with omega-6 desaturase (*Linum usitatissimum*), AFN53645.1 (*E*= 0), EST-77% identity over 642 bp of JG220343.1 (*E*=$6e^{-92}$)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process (Lus10021048)
	BAC28C5_14	39159	42282	+	1149	1	382		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
15	**BAC364K11_15**	49320	50873	–	1125	1	374	97% identity with omega-6 desaturase (*Linum usitatissimum*), AFN53644.1 (*E*= 0), EST-99% identity over 588 bp of JG235393.1 (*E*=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process (Lus10021049)
	BAC28C5_15	45857	47410	–	1125	1	374		
16	**BAC364K11_16**	52089	53634	+	1119	1	372	100% identity with omega-6 desaturase (*Linum usitatissimum*), AFN53643.1 (*E*= 0), EST-78% identity over 502 bp of JG184171.1 (*E*=$2e^{-77}$)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process (Lus10021050)
	BAC28C5_16	48626	50171	+	1119	1	372		
17	**BAC364K11_17**	54317	55215	–	339	3	112	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC28C5_17	50854	51752	–	339	3	112		
18	**BAC364K11_18**	57523	60495	+	1149	4	382	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC28C5_18	54060	57255	+	1353	5	450		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
19	**BAC364K11_19**	61371	63295	+	1149	1	382	100% identity with omega-6 desaturase (*Linum usitatissimum*), ACF49507.1 (*E*= 0), EST-94% identity over 1140 bp of EH792177.1 (*E*=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, oxidation reduction process, lipid metabolic process, omega-6 fatty acid desaturase (delta-12 desaturase), (Lus10021051)
	BAC28C5_19	58131	60055	+	1149	1	382		
20	**BAC364K11_20**	63908	66173	+	705	7	234	69% identity with Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein isoform 1 (*Theobroma cacao*), XP_007051133.1 (*E*=3e^{-94}), EST-100% identity over 385 bp of JG244576.1 (*E*=0)	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE- peptidyl-prolyl cis-trans isomerase activity, protein folding (Lus10021052)
	BAC28C5_20	60668	62933	+	705	7	234		
21	**BAC364K11_21**	66225	68093	–	1551	2	516	100% identity with pentatricopeptide repeat-containing protein (*Linum*	FAMILY NOT NAMED-PPR repeat (Lus10021053)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC28C5_21	63014	64882	–	1551	2	516	*usitatissimum*), AFN53641.1 (*E*= 0), EST-100% identity over 469 bp of JG202499.1 (*E*=0)	
22	**BAC364K11_22**	70928	72762	+	771	1	256	36% identity with hypothetical protein PRUPE_ppa020526mg (*Prunus persica*), XP_007204031.1 (*E*=2e^{-17})	VQ motif (Lus10021054)
	BAC28C5_22	67688	69522	+	771	1	256		
23	**BAC364K11_23**	73365	75905	–	969	7	322	71% identity with PREDICTED: probable plastid-lipid-associated protein 13, chloroplastic-like isoform X1 (*Citrus sinensis*), XP_006492860.1 (*E*=5e^{-141}), EST-99% identity over 676 bp of JG265370.1 (*E*=0)	No functional annotations for this locus (Lus10021055)
	BAC28C5_23	70125	72665	–	969	7	322		
24	**BAC364K11_24**	78650	80897	+	1188	4	395	74% identity with PREDICTED: probable protein phosphatase 2C 43 (*Vitis vinifera*), XP_002279599.1 (*E*= 0)	PROTEIN PHOSPHATASE 2C-Protein phosphatase 2C/pyruvate dehydrogenase (lipoamide) phosphatase-catalytic activity (Lus10021056)
	BAC28C5_24	75410	77657	+	1188	4	395		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
25	**BAC364K11_25**	81072	82958	–	1134	4	377	100% identity with putative actin-97 protein (*Linum usitatissimum*), AFN53639.1 (*E*= 0), EST-99% identity over 922 bp of JG174412.1 (*E*=0)	ACTIN-Actin and related proteins (Lus10021057)
	BAC28C5_25	77832	79718	–	1134	4	377		
26	**BAC364K11_26**	86292	90452	+	1341	10	446	63% identity with PREDICTED: metal tolerance protein C2 isoform X1 (*Pyrus x bretschneideri*), XP_009369486.1 (*E*=6e^{-82})	CATION EFFLUX PROTEIN/ ZINC TRANSPORTER-integral to membrane, cation transmembrane transporter activity (Lus10021058)
	BAC28C5_26	83052	87213	+	1341	10	446		
27	**BAC364K11_27**	90559	93863	+	1731	5	576	73% identity with transducin family protein (*Populus trichocarpa*), XP_002308919.2 (*E*= 0), EST-99% identity over 497 bp of JG244946.1 (*E*=0)	WD40 REPEAT PROTEIN-U4/U6 small nuclear ribonucleoprotein Prp4 (contains WD40 repeats) (Lus10021059)
	BAC28C5_27	87320	90624	+	1731	5	576		
28	**BAC364K11_28**	94211	97132	–	1827	3	608	100% identity with putative ATP-binding protein (*Linum usitatissimum*),	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC28C5_28	90972	93893	–	1827	3	608	AFN53637.1 (*E*= 0)	KINASE-protein kinase activity, binding, ATP binding, protein phosphorylation (Lus10021060)
29	**BAC364K11_29**	98534	102002	+	1050	1	349	42% identity with hypothetical protein POPTR_0016s05830g, partial (*Populus trichocarpa*), XP_002322724.2 (*E*=2e^{-64}), EST-89% identity over 407 bp of JG171781.1 (*E*=2e^{-137})	No functional annotations for this locus (Lus10021061)
	BAC28C5_29	95295	98764	+	1050	1	349		
30	**BAC364K11_30**	104353	105257	–	222	1	73	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC28C5_30	101117	102020	–	222	1	73		
31	**BAC364K11_31**	106748	108899	–	882	4	293	42% identity with unnamed protein product (*Coffea canephora*), CDP10155.1 (*E*=1e^{-18}),	WRC-protein binding (Lus10021062)
	BAC28C5_31	103511	105662	–	882	4	293		
32	**BAC364K11_32**	110896	113614	+	1257	1	418	63% identity with rnf5, putative (*Ricinus communis*), XP_002530725.1 (*E*=3e^{-138}), EST-99% identity over 646 bp of JG258669.1 (*E*=0)	RNF5-Zinc finger, C3HC4 type (RING finger)- Predicted E3 ubiquitin ligase (Lus10021063)
	BAC28C5_32	107660	110378	+	1257	1	418		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
33	BAC364K11_33	116732	118462	–	342	2	113	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
34	BAC364K11_34	120768	122412	+	1056	4	351	99% identity with putative homeobox-leucine zipper protein (*Linum usitatissimum*) AFN53636.1 (E=3e-144), EST-99% identity over 623bp of EH791697.1 (E=0)	FAMILY NOT NAMED-Homeobox domain and Homeobox associated leucine zipper-DNA binding, sequence-specific DNA binding transcription factor activity, sequence-specific DNA binding, nucleus, regulation of transcription, DNA-dependent (Lus10021064)
Fad3a									
1	**BAC395P20_1**	5015	10125	–	1107	12	368	55% identity with hypothetical protein POPTR_0004s03700g (*Populus trichocarpa*), XP_006384013.1 (E=1e^{-143})	Protein of unknown function (DUF707) (Lus10038303)
	BAC44E4_1	403	5539	–	1107	12	368		
2	**BAC395P20_2**	12720	14057	–	555	1	184	64% identity with PREDICTED: kelch-like	Kelch motif-protein binding (Lus10038304)

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC44E4_2	8134	9471	–	555	1	184	protein 40b-like XP_003545867.1 (*Glycine max*), (E=2e^{-78})	
3	**BAC395P20_3**	14884	18794	–	1329	14	442	82% identity with alpha-galactosidase family protein (*Populus trichocarpa*), XP_002305678.2 (E=0), EST-98% identity over 562 bp of JG024700.1 (E=0)	Alpha-D-galactosidase (melibiase)- hydrolase activity, hydrolyzing O-glycosyl compounds and/or carbohydrate metabolic process (Lus10038305)
	BAC44E4_3	10298	14208	–	1329	14	442		
4	**BAC395P20_4**	19270	20916	+	750	2	249	63% identity with hypothetical protein POPTR_0019s11870g (*Populus trichocarpa*), XP_002325576.1 (E=1e^{-89}), EST-99% identity over 575 bp of JG235563.1 (E=0)	Helix-loop-helix DNA-binding domain-transcription regulator activity (Lus10038306)
	BAC44E4_4	14684	16330	+	750	2	249		
5	**BAC395P20_5**	21115	25806	–	2682	10	893	69% identity with hydroxyproline-rich glycoprotein (*Populus trichocarpa*), XP_002317226.1 (E=7e^{-128}), EST-98% identity over 349 bp of JG044297.1 (E=1e^{-167})	No fuctional annotation for this locus (Lus10038307)
	BAC44E4_5	16529	21222	–	2682	10	893		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
6	**BAC395P20_6**	25945	28528	+	1740	6	579	43% identity with cyclic nucleotide gated channel 1 isoform 1 (*Theobroma cacao*), XP_007136556.1 ($E=1e^{-126}$)	Cyclic nucleotide-binding domain and Ion transport protein-ion channel activity, transmembrane transport, ion transport (Lus10038308)
	BAC44E4_6	21361	23942	+	1740	6	579		
7	**BAC395P20_7**	29134	32793	–	1821	2	606	100% identity with putative phenol oxidase (*Linum usitatissimum*), AFN53703.1 ($E = 0$)	TYROSINASE-oxidoreductase activity, catechol oxidase activity, oxidation-reduction process and metabolic process (Lus10038309)
	BAC44E4_7	24548	28207	–	1821	2	606		
8	**BAC395P20_8**	36290	38097	–	771	1	256	100% identity with hypothetical protein (*Linum usitatissimum*), AFN53702.1 ($E=4e^{-69}$), EST-100% identity over 513 bp of JG068566.1 ($E=0$)	Dof domain, zinc finger-DNA binding, regulation of transcription, DNA-dependent and zinc ion binding (Lus10038310)
	BAC44E4_8	31705	33510	–	771	1	256		
9	**BAC395P20_9**	40549	43036	–	1305	3	434	30% identity with uncharacterized protein isoform 1 (*Theobroma cacao*), XP_007019231.1 ($E=7e^{-15}$)	No functional annotation for this locus (Lus10038311)
	BAC44E4_9	35962	38449	–	1305	3	434		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
10	**BAC395P20_10**	43239	45738	+	954	5	317	100% identity with putative epoxide hydrolase 3 (*Linum usitatissimum*), AFN53701.1 (*E*=0)	Alpha/beta hydrolase fold-Soluble epoxide hydrolase (Lus10038312)
	BAC44E4_10	38652	41151	+	954	5	317		
11	**BAC395P20_11**	46295	48802	—	873	3	290	100% identity with hypothetical protein (*Linum usitatissimum*), AFN53700.1 (*E*=0), EST-95% identity over 809 bp of JG226016.1 (*E*=0)	DEHYDRODOLICHYL DIPHOSPHATE SYNTHASE-transferase activity, transferring alkyl or aryl (other than methyl) groups, dehydrodolichyl diphosphate synthase (Lus10038313)
	BAC44E4_11	41888	44394	—	873	3	290		
12	**BAC395P20_12**	50623	54381	+	1788	11	595	65% identity with PREDICTED: protein DAMAGED DNA-BINDING 2 (*Vitis vinifera*), XP_002279214.1 (*E*=0), EST-99% identity over 373 bp of JG135499.1 (*E*=0)	DAMAGE-SPECIFIC DNA BINDING PROTEIN 2 -WD40 protein, DNA damage-binding protein 2 (Lus10038314)
	BAC44E4_12	46212	49968	+	1770	11	589		
13	**BAC395P20_13**	54520	56756	—	690	6	229	76% identity with peptidyl-prolyl cis-trans isomerase (*Linum usitatissimum*),	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE-peptidyl-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC44E4_13	50107	52343	–	690	6	229	AFN53672.1 (*E*=5e^{-106}), EST-84% identity over 357 bp of JG129214.1(*E*=1e^{-91})	prolyl cis-trans isomerase activity, protein folding (Lus10038315)
14	**BAC395P20_14**	57004	57937	+	396	3	131	43% identity with exopolygalacturonase clone GBGE184 precursor, putative (*Ricinus communis*), XP_010269646.1 (*E*=5e^{-23}),	Glycosyl hydrolases family 28-polygalacturonase activity, carbohydrate metabolic process (Lus10038316)
	BAC44E4_14	52591	53524	+	396	3	131		
15	**BAC395P20_15**	58000	59268	+	465	1	154	61% identity with aspartate aminotransferase, partial (*Prunus persica*), AGF95103.1 (*E*=4e^{-45}), EST-91% identity over 339 bp of CA483040.1 (*E*=7e^{-123})	Putative GTPase activating protein for Arf-regulation of ARF GTPase activity, ARF GTPase activator activity (Lus10038317)
	BAC44E4_15	53587	54855	+	465	1	154		
16	**BAC395P20_16**	60316	61743	+	936	1	311	100% identity with hypothetical protein (*Linum usitatissimum*), AFN53699.1 (*E*= 0), EST-99% identity over 741 bp of JG230412.1(*E*=0)	TRYPTOPHAN BIOSYNTHESIS PROTEIN-Stationary phase-induced protein, SOR/SNZ family, pyridoxal phosphate biosynthetic process (Lus10038318)
	BAC44E4_16	55903	57330	+	936	1	311		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
17	**BAC395P20_17**	62225	65782	+	1743	8	580	84% identity with hypothetical protein (*Linum usitatissimum*), AFN53699.1 (*E*= 0), EST-96% identity over 502 bp of JG212978.1 (*E*=0)	PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C FAMILY PROTEIN-Phosphoinositide-specific phospholipase C, intracellular signal transduction, phosphatidylinositol phospholipase C activity, phospholipase C activityphospholipase C activity, lipid metabolic process (Lus10038319)
	BAC44E4_17	57812	61369	+	1743	8	580		
18	**BAC395P20_18**	65945	67069	+	819	1	272	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC44E4_18	61532	62656	+	819	1	272		
19	**BAC395P20_19**	75892	79153	+	3078	3	1025	47% identity with ATP-dependent Clp protease ClpB family protein (*Populus trichocarpa*), XP_002314097.2 (*E* =0), EST-94% identity over 782 bp of GW864743.1 (*E*=0)	ATP-DEPENDENT CLP PROTEASE-Chaperone HSP104 and related ATP-dependent Clp proteases, ATP binding (Lus10038320)
	BAC44E4_19	71400	75199	+	3078	3	1025		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
20 *	**BAC395P20_20**	84641	87847	+	1179	6	392	100% identity with fatty acid desaturase 3A (*Linum usitatissimum*), ABA02172.1 (*E*=0), EST-99% identity over 857 bp of JG191662.1 (*E*=0)	FATTY ACID DESATURASE 2-With oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water, oxidation-reduction process, lipid metabolic process, omega-3 fatty acid desaturase (delta-15 desaturase) (Lus10038321)
	BAC44E4_20	80030	83235	+	1179	6	392		
21	**BAC395P20_21**	88260	91491	—	1515	8	504	78% identity with unnamed protein product (*Coffea canephora*), CDP09204.1 (*E*=0), EST-95% identity over 616 bp of JG214299.1 (*E*=0)	NADH DEHYDROGENASE-RELATED-NADH dehydrogenase, oxidoreductase activity, oxidation-reduction process, NADH dehydrogenase (Lus10038322)
	BAC44E4_21	83663	86894	—	1515	8	504		
22	**BAC395P20_22**	92234	94645	+	1035	7	344	100% identity with mitochondrial malate dehydrogenase (*Linum usitatissimum*),	MALATE AND LACTATE DEHYDROGENASE-oxidoreductase activity,

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC44E4_22	87637	90049	+	1035	7	344	AFN53696.1 (*E*=0), EST-99% identity over 623 bp of JG174936.1 (*E*=0)	L-malate dehydrogenase activity, malate metabolic process, malate dehydrogenase (Lus10038323)
23	**BAC395P20_23**	94652	95670	—	288	2	95	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC44E4_23	90056	91074	—	288	2	95		
24	**BAC395P20_24**	98252	100408	+	867	4	288	76% identity with tonoplast intrinsic protein, putative (*Ricinus communis*), XP_002525663.1 (E=3e^{-140}), EST-100% identity over 394 bp of JG203377.1 (*E*=0)	AQUAPORIN TRANSPORTER-transporter activity (Lus10038324)
	BAC44E4_24	93657	95813	+	867	4	288		
25	**BAC395P20_25**	101680	106056	+	1209	12	402	76% identity with alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative (*Ricinus communis*), XP_002525666.1 (*E*=0)	ALPHA-GALACTOSIDASE/ALPHA-N-ACETYLGALACTOSAMINIDASE-hydrolase activity, hydrolyzing O-glycosyl compounds, carbohydrate metabolic process, alpha-galactosidase (Lus10038325)
	BAC44E4_25	97085	101463	+	1209	12	402		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
26	**BAC395P20_26**	106491	111041	–	2454	9	817	55% identity with conserved hypothetical protein (*Ricinus communis*), XP_002525667.1 (*E*=0), EST-98% identity over 189 bp of JG267990.1 (*E*=0)	TEX2 PROTEIN-RELATED-Uncharacterized conserved protein TEX2, contains PH domain (Lus10038326)
	BAC44E4_26	101898	106447	–	2490	8	829		
Fad3b									
1	**BAC356B4_1**	2453	4600	+	759	1	252	90% identity with hypothetical protein (*Linum usitatissimum*), AFN53702.1 (*E*=2e^{-57}), EST-94% identity over 513 bp of JG068566.1 (*E*=0)	Dof domain, zinc finger-DNA binding, regulation of transcription, DNA-dependent, zinc ion binding (Lus10036172)
	BAC27L18_1	3754	5532	+	768	1	255		
2	**BAC356B4_2**	5335	6169	–	168	1	55	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC27L18_2	6266	7100	–	168	1	55		
3	**BAC356B4_3**	7332	8888	+	261	1	86	59% identity with PREDICTED: zinc finger BED domain-containing protein DAYSLEEPER-like (*Nicotiana sylvestris*), XP_009791292.1 (*E*=8e^{-28})	Not found in flax GFF3

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC27L18_3							No predicted gene corresponds to BAC356B4_3	
4	**BAC356B4_4**	10312	12242	+	1608	4	535	66% identity with hydroxyproline-rich glycoprotein (*Populus trichocarpa*), XP_002317226.1 (E=1e^{-92})	No functional annotations for this locus (Lus10036173)
	BAC27L18_4	10337	12267	+	1608	4	535		
5	**BAC356B4_5**	12284	13149	+	582	4	193	67% identity with hydroxyproline-rich glycoprotein (*Populus trichocarpa*), XP_002317226.1 (E=2e^{-80}), EST-99% identity over 349 bp of JG044297.1 (E=5e^{-180})	No functional annotations for this locus (Lus10036174)
	BAC27L18_5	13214	14079		582	4	193		
6	**BAC356B4_6**	13537	14328	–	510	2	169	58% identity with hypothetical protein POPTR_0019s11870g (*Populus trichocarpa*), XP_002325576.1 (E=3e^{-54}), EST-95% identity over 497 bp of JG235563.1 (E=0)	No functional annotations for this locus (Lus10036175)
	BAC27L18_6	14469	15260	–	510	2	169		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
7	**BAC356B4_7**	14657	19090	+	1821	15	606	74% identity with alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative (*Ricinus communis*), XP_002519009.1 (*E* = 0), EST-99% identity over 395 bp of JG024700.1 (*E*=0)	ALPHA-GALACTOSIDASE/ALPHA-N-ACETYLGALACTOSAMINIDASE- hydrolase activity, hydrolyzing O-glycosyl compounds, carbohydrate metabolic process (Lus10036176)
	BAC27L18_7	15589	20022	+	1821	15	606		
8	**BAC356B4_8**	19613	21097	+	399	1	132	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC27L18_8	20545	22029	+	399	1	132		
9	**BAC356B4_9**	21328	24519	—	1257	5	418	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC27L18_9	22260	25451	—	1257	5	418		
10	**BAC356B4_10**	24866	26905	+	1089	5	362	95% identity with putative epoxide hydrolase 3 (*Linum usitatissimum*), AFN53701.1 (*E*=0), EST-75% identity over 976 bp of JG216808.1 (*E*=$3e^{-105}$)	ALPHA/BETA HYDROLASE FOLD-CONTAINING PROTEIN-Soluble epoxide hydrolase (Lus10036177)
	BAC27L18_10	25798	27837	+	1089	5	362		
11	**BAC356B4_11**	26991	29601	—	1146	4	381	93% identity with hypothetical protein (*Linum usitatissimum*), AFN53700.1 (*E*=$1e^{-168}$),	DEHYDRODOLICHYL DIPHOSPHATE SYNTHASE-transferase activity, transferring alkyl

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC27L18_11	27923	30532	–	1146	4	381	EST-100% identity over 811 bp of JG226016.1 (*E*=0)	or aryl (other than methyl) groups, cis-prenyltransferase, dehydrodolichyl diphosphate synthase (Lus10036178)
12	**BAC356B4_12**	29675	33761	+	1779	11	592	75% identity with hypothetical protein JCGZ_08474 (*Jatropha curcas*), KDP46502.1 (*E*= 0), EST-100% identity over 368 bp of JG135499.1 (*E*=2e^{-153})	DAMAGE-SPECIFIC DNA BINDING PROTEIN 2- WD40 protein, DNA damage-binding protein 2 (Lus10036179)
	BAC27L18_12	30606	34693	+	1779	11	592		
13	**BAC356B4_13**	33896	36639	–	684	6	227	74% identity with peptidyl-prolyl cis-trans isomerase (*Linum usitatissimum*), AFN53672.1 (E=9e^{-103}), EST-85% identity over 438 bp of JG135153.1 (*E*=6e^{-120})	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE- peptidyl-prolyl cis-trans isomerase activity, protein folding (Lus10036180)
	BAC27L18_13	34828	37571	–	684	6	227		
14	**BAC356B4_14**	37340	38782	+	543	4	180	Predicted by FGENESH, No significant similarity found	Not found in flax GFF3
	BAC27L18_14	38290	39889	+	351	3	166		
15	**BAC356B4_15**	40743	41678	+	936	1	311	98% identity with hypothetical protein (*Linum usitatissimum*),	TRYPTOPHAN BIOSYNTHESIS PROTEIN-pyridoxal

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
	BAC27L18_15	41828	42763	+	936	1	311	AFN53699.1 (*E*=0), EST-96% identity over 741 bp of JG230412.1 (*E*=0)	phosphate biosynthetic process-pyridoxine biosynthesis protein (Lus10036181)
16	**BAC356B4_16**	42911	45291	+	1689	8	562	87% identity with hypothetical protein (*Linum usitatissimum*), AFN53699.1 (*E*=0), EST-100% identity over 789 bp of JG212978.1 (*E*=0)	PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C FAMILY PROTEIN-intracellular signal transduction, phosphatidylinositol phospholipase C activity, phospholipase C activity, protein binding, lipid metabolic process (Lus10036182)
	BAC27L18_16	43986	46370		1689	8	562		
17	**BAC356B4_17**	52058	56998	+	3387	7	1128	49% identity with hypothetical protein JCGZ_11198 (*Jatropha curcas*), KDP34836.1 (*E* = 0), EST-99% identity over 698 bp of GW864743.1 (*E*=0)	ATP-DEPENDENT CLP PROTEASE-Chaperone HSP104 and related ATP-dependent Clp proteases, ATP binding (Lus10036183)
	BAC27L18_17	52067	57030	+	3588	8	1195		

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
18 *	**BAC356B4_18**	61394	64687	+	1176	6	391	100% identity with fatty acid desaturase 3B (*Linum usitatissimum*), ABA02173.1 (*E* =0), EST-99% identity over 857 bp of JG191662.1 (*E*=0)	FATTY ACID DESATURASE 2-oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen, lipid metabolic process, omega-3 fatty acid desaturase (Lus10036184)
	BAC27L18_18	61434	64727	+	1176	6	391		
19	BAC27L18_19	65442	68104	—	1515	8	504	78% identity with hypothetical protein PRUPE_ppa004595mg (*Prunus persica*) XP_007205047.1 (E=0), EST-96% identity over 616bp of JG214299.1 (E=0)	NADH DEHYDROGENASE-RELATED-Pyridine nucleotide-disulphide oxidoreductase, Pyridine nucleotide-disulphide oxidoreductase-oxidoreductase activity, flavin adenine dinucleotide binding, NADH dehydrogenase (Lus10036185)
20	BAC27L18_20	68443	69907	+	582	5	193	92% identity with mitochondrial malate dehydrogenase (*Linum*	MALATE AND LACTATE DEHYDROGENASE- L-

No	Predicted gene[a]	Start position	End position	Strand	CDS length (bp)	Number of exons	Predicted protein (aa)[b]	BLAST result[c]	GFF3 verification and functional annotation
								usitatissimum) AFN53696.1 (E=2e-123), EST-93% identity over 597bp of JG174936.1 (E=0)	malate dehydrogenase activity, oxidation-reduction process, malate metabolic process (Lus10036186)
21	BAC27L18_21	70050	71822	–	288	2	95	No significant similarity found, EST-99% identity over 288bp of JG058777.1 (E=2e-146)	Not found in flax GFF3
22	BAC27L18_22	74116	76262	+	774	3	257	92% identity with mitochondrial malate dehydrogenase (*Linum usitatissimum*) AFN53696.1 (E=2e-123), EST-93% identity over 597bp of JG174936.1 (E=0)	AQUAPORIN TRANSPORTER-membrane, transporter activity, transport (Lus10036187)

[a] Predicted genes were named according to the name of the BAC clone and the order in which they appear in the assembly. Predicted genes located in the overlapping region are in bold.

[b] Number of amino acid residues

[c] Best BLASTx match against the NCBI non-redundant (nr) database and/or best BLASTn match against dbEST

*FA gene used to design the primers for the BAC identification

Appendix XXV Transposable elements identified in the 12 BACs harbouring *sad1, sad2, fad2a, fad2b, fad3a* and *fad3b* loci of CDC Bethune and M5791

Locus	Genotype	Order	Super family	Transposable element	Length (bp)	E value
sad1	CDC Bethune	LTR	*Copia*	LTRAnnotator_Lu-RLC-1651#full	4940	0
		LTR	*Gypsy*	LTRAnnotator_Lu-RLG-860#full	12108	0
		LTR	*Gypsy*	LTRAnnotator_Lu-RLG-1434#full	11797	0
	M5791	LTR	*Unknown*	LTRAnnotator_Lu-RLX-1596#full	4815	0
		Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=1391475	1475	0
		Unclassified	*Unclassified*	Flax_Undetermined_repeat_piler_fam85.64908	908	0
sad2	CDC Bethune	TIR	*Mutator*	Flax_Mutatorrepeatscout_R=631622	1622	0
	M5791	LTR	*Copia*	LTRAnnotator_Lu-RLC-597#full	5312	1.0E-142
fad2a	CDC Bethune	LTR	*Unknown*	LTRAnnotator_Lu-RLC-1805#full	6675	0
		LTR	*Unknown*	LTRAnnotator_Lu-RLG-1389#full	5103	0
		Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=203681200	1200	0
	M5791	Unclassified	*Unclassified*	Flax_Undetermined_repeat_piler_fam2.10883	883	0
		Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=1346526	526	0
fad2b	CDC Bethune	Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=33121077	1077	0
		Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=1352618	618	0
	M5791	Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=33121077	1077	0
		Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=1352618	618	0
fad3a	CDC Bethune	Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=563709	709	0
	M5791	Unclassified	*Unclassified*	Flax_Undetermined_repeat_repeatscout_R=563709	709	0
fad3b	CDC Bethune	TIR	*hAT*	Flax_hATrepeatscout_R=1961663	1663	0
		MITE	*MITE*	MITE_Digger-candidate-3-assembled-By-SOOMOS-flax-Contig3	362	1.0E-100
	M5791	TIR	*hAT*	Flax_hATrepeatscout_R=1961663	1663	0
		LTR	*Copia*	LTRAnnotator_Lu-RLC-1235#full	6963	0
		MITE	*MITE*	MITE_Digger-candidate-3-assembled-By-SOOMOS-flax-Contig3	362	1.0E-100

Appendix XXVI List of expressed genes identified in the 12 BAC sequences and their level of expression in different tissues of flax based on normalized RNA-seq data and expressed as fragments per kb of transcripts per million mapped reads (FPKM) as previously described (Kumar et al. 2013; http://linum.ca/downloads/RNAseq)

Fatty acid desaturase locus	Gene ID	Globular embryo	Heart embryo	Torpedo embryo	Cotyledon embryo	Mature embryo	Seed	Anther	Ovary	Flower	Root	Stem	Etiolated seedling	Leaf
sad1	Lus10027459	0.555	0.862	1.186	0.389	0.000	4.276	1.165	2.762	0.173	4.728	0.058	0.589	0.626
	Lus10027462	1.097	1.999	2.091	1.238	1.002	9.119	2.066	5.669	15.976	8.517	17.198	17.317	6.815
	Lus10027463	34.131	37.085	18.913	25.674	55.026	14.416	7.118	16.212	20.829	23.407	15.695	11.269	1.225
	Lus10027464	27.821	23.853	12.895	16.008	20.986	14.758	25.685	13.222	11.284	19.586	14.015	15.451	6.112
	Lus10027471	1.191	1.674	1.537	1.890	4.176	1.033	1.713	0.854	0.857	1.864	1.359	1.601	0.554
	Lus10027476	1.191	1.674	1.537	1.890	4.176	1.033	1.713	0.854	0.857	1.864	1.359	1.601	0.554
	Lus10027480	0.000	0.000	0.000	0.000	0.000	0.388	0.028	0.723	0.507	1.045	0.201	0.015	0.000
	Lus10027481	0.033	0.012	0.005	0.000	0.000	0.063	0.590	0.038	0.019	0.063	0.027	0.003	0.006
	Lus10027482	0.227	0.184	0.471	3.451	1.604	1.319	0.698	1.284	2.977	2.941	2.185	3.520	1.018
	Lus10027484	26.516	42.945	77.437	103.047	27.155	20.758	18.855	19.463	15.112	21.158	16.630	13.299	2.118
	Lus10027485	1.106	1.836	10.296	7.535	4.143	3.076	0.648	1.791	4.124	3.785	1.693	12.386	0.250
	Lus10027486*	94.856	161.745	426.637	490.648	161.480	72.299	12.209	28.887	20.129	47.666	28.811	23.971	9.491
	Lus10027488	2.881	2.337	3.180	0.455	0.020	1.149	0.077	3.623	2.908	2.822	1.632	1.823	1.554
	Lus10027490	0.160	0.236	0.344	0.467	0.226	1.215	0.463	0.850	1.891	1.668	2.067	0.991	0.812
	Lus10027491	5.389	5.074	3.602	4.058	9.222	3.208	21.346	3.383	4.978	4.953	3.862	2.730	0.470
	Lus10027495	11.238	19.591	18.597	47.538	42.928	3.518	1.166	2.170	1.243	6.114	3.725	5.765	0.802
	Lus10027496	0.429	0.086	0.952	0.612	0.098	2.527	2.249	3.403	3.418	7.091	6.188	5.528	0.539
	Lus10027498	0.618	0.638	0.839	0.612	0.283	1.532	3.661	1.110	1.964	3.186	1.912	2.023	0.284
	Lus10027499	12.127	14.333	4.658	5.942	6.366	11.818	3.940	17.027	6.956	22.072	11.532	14.427	9.991
sad2	Lus10039215	2.282	0.793	0.061	0.091	0.241	59.534	58.380	19.787	50.143	54.207	45.826	35.200	1.971
	Lus10039216	0.546	0.995	19.695	9.024	0.101	10.051	0.633	0.521	0.000	1.737	3.463	1.806	1.983
	Lus10039222	125.782	193.015	388.377	80.548	12.966	256.317	366.835	236.338	92.151	189.674	155.190	363.890	0.105
	Lus10039223	0.147	0.314	3.653	2.698	3.976	0.760	0.093	2.659	0.254	0.424	0.019	3.082	0.169
	Lus10039224	4.015	2.255	6.346	7.234	21.319	24.041	77.806	52.750	38.940	48.316	34.124	29.775	3.753
	Lus10039225	9.529	10.376	0.204	0.211	0.000	0.000	0.000	0.000	0.078	0.000	0.049	0.318	0.053

Fatty acid desaturase locus	Gene ID	Globular embryo	Heart embryo	Torpedo embryo	Cotyledon embryo	Mature embryo	Seed	Anther	Ovary	Flower	Root	Stem	Etiolated seedling	Leaf
	Lus10039229	3.862	0.863	0.044	0.016	0.025	0.810	500.423	0.891	0.069	3.944	0.936	0.073	0.000
	Lus10039230	11.578	11.733	2.737	1.142	3.685	2.403	0.013	2.718	2.163	0.900	2.382	1.214	0.573
	Lus10039232	17.647	24.034	20.591	6.950	16.518	14.284	102.746	14.149	80.486	11.999	12.095	11.798	1.259
	Lus10039233	0.008	0.013	0.034	0.025	0.000	0.034	0.054	0.093	0.000	0.072	0.013	0.004	0.000
	Lus10039235	2.100	1.832	4.771	12.478	11.347	30.883	18.650	20.735	25.926	8.820	26.386	37.852	9.932
	Lus10039236	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.596	0.000	0.000	0.000
	Lus10039237	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.123	0.000	0.012	0.000
	Lus10039238	0.368	0.102	0.380	0.437	0.192	9.459	1.842	8.820	10.586	5.214	6.671	6.925	0.861
	Lus10039241*													
	Lus10039242	4.701	0.000	0.000	6.196	10.475	0.408	3.946	1.304	0.391	6.267	1.930	0.876	0.000
	Lus10039244	0.404	0.205	0.217	0.195	0.104	0.040	0.152	0.074	0.283	0.028	0.033	0.031	0.013
	Lus10039250	2.433	2.791	2.718	1.051	2.193	4.561	4.840	5.572	5.183	2.585	5.866	5.228	1.813
	Lus10039251	0.699	1.504	1.802	3.797	3.076	0.191	0.101	0.404	1.197	0.124	0.018	0.840	0.202
	Lus10039252	0.012	0.000	0.000	0.029	0.000	0.000	1.156	0.000	0.015	0.069	0.000	0.018	0.000
	Lus10039253	2.608	2.587	3.273	1.302	2.653	13.690	29.100	8.081	18.628	7.205	12.830	16.059	1.581
	Lus10039258	0.000	0.000	0.000	0.000	0.000	0.036	0.037	0.186	0.505	0.178	1.460	0.377	0.050
	Lus10039259	0.197	0.067	0.371	0.112	0.022	2.039	0.678	62.939	4.759	7.659	1.288	6.551	0.000
	Lus10039261	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Lus10039262	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Lus10039263	0.010	0.000	0.000	0.000	0.012	0.158	0.150	0.051	0.216	0.044	1.075	0.472	0.000
	Lus10039264	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Lus10039266	0.000	0.000	0.161	0.233	0.152	0.502	0.000	0.000	0.000	0.252	0.254	0.499	0.154
	Lus10039268	0.014	0.005	0.015	0.000	0.015	0.008	0.075	0.037	0.079	0.022	0.091	0.011	0.000
fad2a	Lus10011976	2.909	3.555	3.097	2.273	1.551	0.429	0.053	0.524	0.574	0.756	0.566	0.435	0.235
	Lus10011977	0.000	0.000	0.000	0.000	0.000	0.024	0.000	0.007	0.036	0.037	0.000	0.000	0.000
	Lus10011978	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.000	0.023	0.000
	Lus10011982	3.479	0.299	0.000	0.000	0.022	0.481	83.103	0.418	0.013	9.088	0.125	0.000	0.000
	Lus10011989	5.418	6.466	9.552	7.401	13.303	16.146	26.883	13.182	26.087	30.738	20.738	16.120	5.331
	Lus10011990	1.870	1.957	5.131	21.187	18.814	12.125	8.479	10.340	10.524	11.320	7.759	28.967	1.836

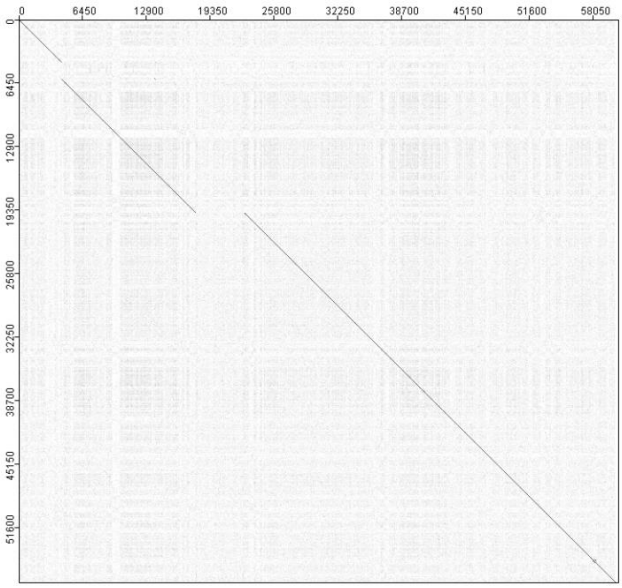
Fatty acid desaturase locus	Gene ID	Globular embryo	Heart embryo	Torpedo embryo	Cotyledon embryo	Mature embryo	Seed	Anther	Ovary	Flower	Root	Stem	Etiolated seedling	Leaf
	Lus10011993	0.000	0.119	0.405	0.000	0.144	7.575	150.340	2.572	0.889	38.871	0.684	14.510	0.000
	Lus10012006	0.549	0.763	0.506	0.493	3.244	0.727	3.146	0.730	0.787	0.912	0.900	0.745	0.153
	Lus10012007+ Lus10012008*													
	Lus10012011	0.161	0.094	0.191	0.000	0.332	0.449	0.049	0.065	15.243	2.183	0.752	6.002	0.492
	Lus10012018	2.200	4.084	13.303	17.501	185.148	3.968	1.141	0.432	1.099	1.835	2.169	4.627	0.000
	Lus10012019	14.020	27.239	21.486	5.866	3.619	5.640	1.078	7.164	1.613	8.912	4.385	3.213	5.844
	Lus10012021	6.143	6.320	4.698	3.712	6.149	3.146	3.542	2.900	2.884	4.926	2.360	2.076	0.827
fad2b	Lus10021037	4.944	6.121	5.957	12.523	4.312	2.153	0.400	2.483	1.234	2.761	2.033	1.073	21.604
	Lus10021039	1.220	1.100	1.109	0.694	0.131	0.098	20.017	0.077	0.029	1.504	0.057	0.028	0.256
	Lus10021041	3.609	7.992	25.673	26.587	0.050	5.144	0.393	5.766	2.598	4.728	1.896	2.379	3.496
	Lus10021044	1.619	0.516	9.195	542.327	1372.630	11.660	0.695	0.008	0.273	0.083	0.090	1.417	0.152
	Lus10021045 (FAD2Ba)													
	Lus10021045 (FAD2Bb)													
	Lus10021047 (FAD2Bc)													
	Lus10021048* (FAD2Bd)													
	Lus10021049 (FAD2Be)	183.496	218.113	235.199	480.693	37.897	11.567	17.091	16.484	8.199	14.017	6.012	11.421	18.469
	Lus10021050 (FAD2Bf)													
	Lus10021051 (FAD2Bg)													
	Lus10021053	13.036	9.637	3.005	3.282	11.708	1.916	0.135	1.762	1.062	1.605	0.803	0.709	7.230
	Lus10021055	4.422	7.781	4.221	4.771	0.319	4.193	0.662	5.141	1.791	9.538	6.873	5.736	206.601
	Lus10021057	7.432	6.694	2.314	2.337	0.583	5.212	144.722	11.756	9.275	70.619	9.174	15.356	170.042
	Lus10021058	1.364	2.586	2.000	4.002	6.993	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Lus10021060	5.390	5.523	7.902	9.711	8.763	1.013	1.341	0.922	0.264	1.716	0.778	1.154	10.932
	Lus10021062	0.000	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.042	0.000	0.000
fad3a	Lus10038303	8.127	5.365	2.537	4.378	4.395	1.847	7.573	1.906	2.195	0.739	1.679	1.713	0.159
	Lus10038304	0.148	0.404	0.000	0.000	0.057	0.000	0.016	0.038	0.084	0.011	0.011	0.056	0.000
	Lus10038305	1.752	4.765	13.211	21.719	1.809	3.141	1.566	7.404	1.369	1.816	5.070	2.502	2.369
	Lus10038307	0.196	0.203	0.127	0.200	0.106	0.078	45.970	0.152	0.102	0.375	0.104	0.071	0.089

Fatty acid desaturase locus	Gene ID	Globular embryo	Heart embryo	Torpedo embryo	Cotyledon embryo	Mature embryo	Seed	Anther	Ovary	Flower	Root	Stem	Etiolated seedling	Leaf
	Lus10038309	0.190	2.185	0.727	0.671	2.407	0.003	0.000	0.005	1.083	0.000	0.003	0.588	0.000
	Lus10038310	0.024	0.000	0.201	1.000	0.271	3.071	1.376	2.660	4.699	1.210	1.660	2.316	0.000
	Lus10038311	0.408	0.819	1.136	0.363	0.016	0.876	0.009	0.858	0.434	0.245	0.648	0.581	0.427
	Lus10038313	0.000	0.015	0.000	0.000	0.000	0.051	0.000	0.840	0.060	0.114	0.137	0.074	0.135
	Lus10038315	0.074	0.000	0.000	0.000	0.000	0.015	107.178	0.066	0.000	2.688	0.152	0.000	0.000
	Lus10038316	0.000	0.000	0.000	0.000	0.000	0.039	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Lus10038319	0.057	0.063	0.020	0.015	0.126	0.010	0.263	0.078	0.234	0.055	0.107	0.078	0.013
	Lus10038321*													
	Lus10038322	0.632	0.465	3.769	1.368	4.570	6.155	3.861	8.490	10.817	4.072	6.452	5.051	0.461
	Lus10038326	0.573	0.699	3.095	1.439	0.515	7.813	5.738	6.120	5.708	2.239	6.528	6.511	2.178
fad3b	Lus10036172	0.236	0.055	0.059	2.257	1.156	10.646	7.275	9.430	13.968	4.420	5.022	3.093	0.000
	Lus10036176	3.931	5.250	12.726	48.251	2.469	2.639	0.110	6.039	2.327	1.065	1.220	1.404	0.670
	Lus10036177	0.027	0.151	0.194	0.130	0.019	0.550	0.011	0.124	0.703	0.028	0.293	0.342	0.000
	Lus10036179	0.806	1.068	0.291	0.336	0.624	0.053	1.076	0.068	0.058	0.027	0.105	0.099	0.025
	Lus10036183	0.629	1.350	2.364	3.408	1.341	22.878	3.954	14.099	16.633	8.265	35.201	14.843	11.673
	Lus10036184*	0.898	2.735	228.380	395.749	0.473	407.903	122.808	214.312	0.511	71.523	1.083	0.427	1.195
	Lus10036186	0.000	0.144	0.000	0.000	0.037	0.026	0.128	0.000	0.023	0.022	0.000	0.074	0.000
	Lus10036187	7.679	10.932	657.456	6575.420	3103.280	114.857	8.288	1.521	3.844	4.822	6.947	8.077	1.890

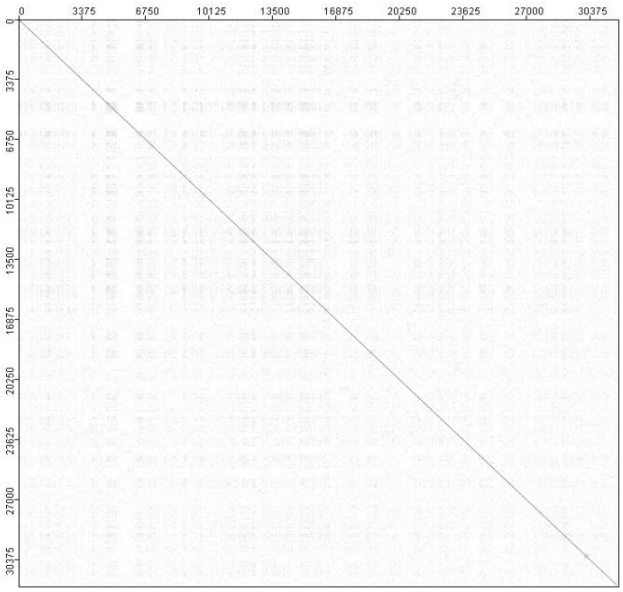
* Desaturase gene

Shaded boxes represent the additional copies of the corresponding desaturase gene

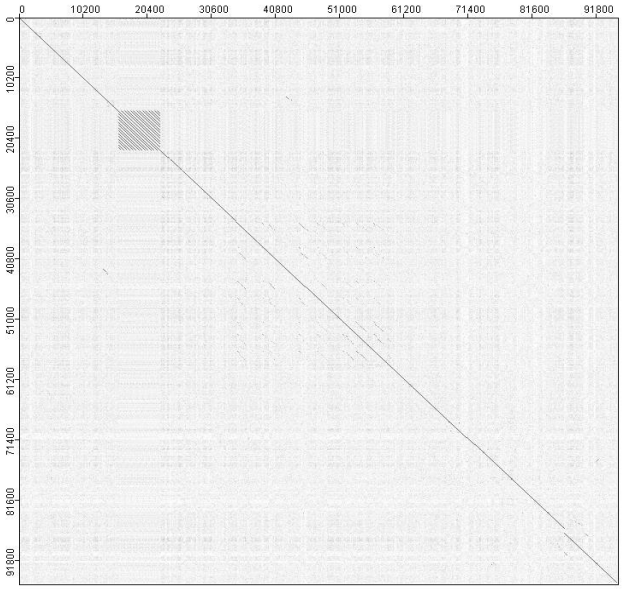
a


0
6450
12900
19350
25800
32250
38700
45150
51600
58050

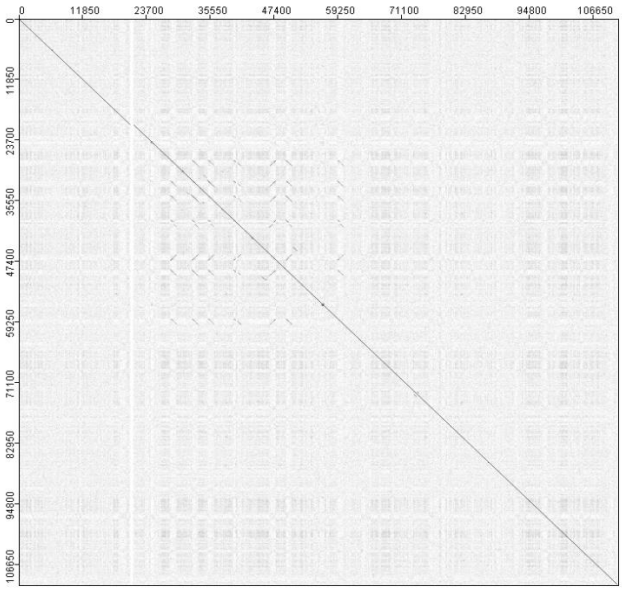
b


0
3375
6750
10125
13500
16875
20250
23625
27000
30375

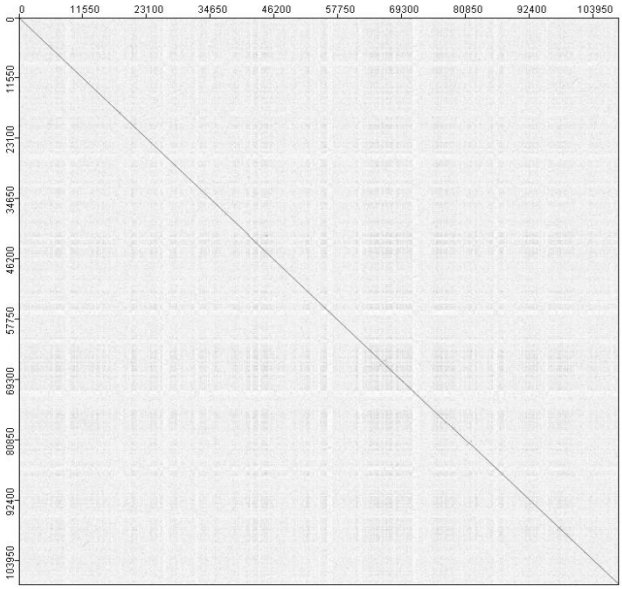
c


0
10200
20400
30600
40800
51000
61200
71400
81600
91800

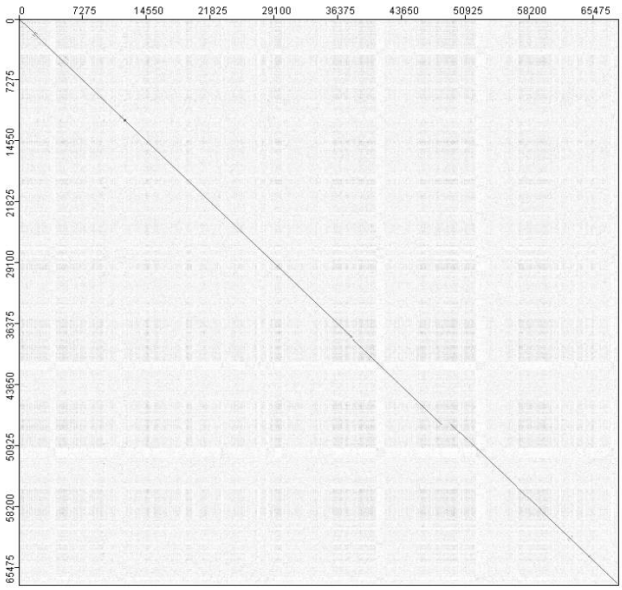
d


0
11850
23700
35550
47400
59250
71100
82950
94800
106650

e



f



Appendix XXVII Dot plots of desaturase loci BAC sequences of CDC Bethune (x-axis) and M5791 (y-axis) performed using JDOTTER (Brodie et al. 2004) with default parameters. **a** *sad1*, **b** *sad2*, **c** *fad2a*, **d** *fad2b*, **e** *fad3a* and **f** *fad3b*