Molecular and biochemical characterization of high and low α-linolenic acid containing Indian flax (*Linum usitatissimum* L.) varieties

A thesis submitted
to the
University of Mumbai
for the
Ph.D. (Science) Degree
in Biotechnology

Submitted By
Ashwini V. Rajwade

Under the guidance of
Dr. Vidya S. Gupta

Division of Biochemical Sciences,
National Chemical Laboratory
Pune 411008 (INDIA)

December 2012

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **'Molecular and biochemical characterization of high and low α-linolenic acid containing Indian flax (*Linum usitatissimum* L.) varieties'** submitted by **Mrs. Ashwini V. Rajwade,** for the degree of **Doctor of Philosophy**, was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.

Date:

Place: Pune

Dr. Vidya S. Gupta

(Research Guide)

STATEMENT BY THE CANDIDATE

As required by the University Ordinance 770, I wish to state that the work embodied in this thesis titled "**Molecular and biochemical characterization of high and low α- linolenic acid containing Indian flax (*Linum usitatissimum* L.) varieties**" forms my own contribution to the research work carried out under the guidance of **Dr. Vidya S. Gupta** at the **National Chemical Laboratory, Pune, India**. This work has not been submitted for any other degree of this or any other University. Whenever references have been made to previous works of others, it has been clearly indicated as such and included in the Bibliography.

Signature of the Candidate

Name: Ashwini V. Rajwade

Certified by

Signature of the Guide

Name: Dr. Vidya S. Gupta

Dedicated to my
beloved parents
Ian Crowson

TABLE OF CONTENTS

Acknowledgements **i**

List of abbreviations **iv**

List of tables **vi**

List of figures **viii**

Thesis Abstract **xii**

CHAPTER 1: Introduction and review of literature

Section 1: Fatty acids: saturated and unsaturated **1**

1.1 Monounsaturated fatty acids **3**

1.2 Polyunsaturated fatty acids **3**

1.2.1 Importance of ω-6 and ω-3 fatty acid balance in the diet **6**

1.3 Fatty acid biosynthesis in plants **7**

1.4 Fatty acid desaturases: essential for PUFA synthesis **10**

1.4.1 Key fatty acid desaturases responsible for accumulation of PUFA in plants **13**

1.4.2 Role of desaturases in plants **18**

Section 2: Flax: the richest agricultural source of ALA **21**

2.1 Economic importance **21**

2.2 Flax as human food **23**

2.3 Taxonomic status **26**

2.4 Origin and domestication **27**

2.5 Plant morphology **28**

2.6 Flax as a crop **29**

2.7 Germplasm diversity **31**

2.8 Omics studies **34**

2.9 Indian flax agriculture and germplasm **37**

Section 3: Genesis of the thesis **41**

3.1 Objectives **41**

3.2 Organization of the thesis **42**

CHAPTER 2: Assessment of genetic and fatty acid diversity among Indian flax varieties

2.1 Introduction **43**

2.2 Materials and methods **44**
2.2.1 Plant material **44**
2.2.2 DNA extraction and PCR amplification **44**
2.2.3 Fatty acid analysis **46**
2.2.4 Total oil content **46**
2.2.5 Data analysis and softwares used for genetic diversity studies **47**
2.3 Results **48**
2.3.1 Genotypic diversity analysis based on molecular studies **48**
2.3.2 Fatty acid analysis **57**
2.4 Discussion **61**
2.4.1 ISSR marker based genetic diversity **61**
2.4.2 Fatty acid diversity and its correlation with the genetic diversity among the genotypes **62**
2.4.3 Implications of genetic diversity **63**

CHAPTER 3: Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content

3.1 Introduction **64**
3.2 Materials and methods **64**
3.2.1 Plant material **64**
3.2.2 Fatty acid analysis **66**
3.2.3 Transcript profiling **66**
3.3 Results **69**
3.3.1 Fatty acid accumulation patterns in developing seeds **69**
3.3.2 Desaturase gene expression in developing seeds **74**
3.4 Discussion **83**
3.4.1 Expression dynamics of three desaturase genes in flax **84**
3.4.2 Transcriptional correlation of desaturase genes with ALA content **87**

CHAPTER 4: Sequence characterization and *in silico* structure prediction of desaturases from the flax varieties varying in α-linolenic acid content

4.1 Introduction **89**
4.2 Materials and methods **91**
4.2.1 Plant material **91**

4.2.2 DNA extraction and PCR amplification **91**

4.2.3 Nucleotide sequencing and analysis of the fatty acid desaturase genes **92**

4.2.4 Haplotype analysis **94**

4.2.5 Three-dimensional protein modeling **94**

4.3 Results **94**

4.3.1 Molecular characterization of *SAD*, *FAD2* and *FAD3* genes of ten flax varieties **94**

4.3.2 Amino acid sequence analysis for SAD, FAD2 and FAD3 among the ten varieties **98**

4.3.3 Haplotype analysis **107**

4.3.4 Structure prediction and analysis of desaturases **111**

4.4 Discussion **121**

4.4.1 Unique and distinct haplotypes in desaturase genes across flax varieties **122**

4.4.2 AA sequence comparison and structural features of desaturase proteins across flax varieties **122**

4.4.3 Correlation of desaturase gene mutation with ALA content **124**

CHAPTER 5: Thesis summary and future prospects

5.1 Genetic and fatty acid diversity amongst Indian flax varieties **126**

5.2 Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content **127**

5.3 Sequence characterization and *in silico* structure prediction of desaturases from the flax varieties varying in α-linolenic acid content **128**

5.4 Future prospects **128**

5.4.1 Promoter region analysis **129**

5.4.2 Studies towards post-transcriptional and post-translational gene regulation **129**

5.4.3 Functional analysis of desaturase sequence variants obtained from this study **129**

5.4.4 Protein structure determination of membrane bound desaturases **129**

5.4.5 Genetic resource development **130**

Bibliography **131**

Annexure I **148**

Synopsis of the thesis **163**

Curriculum Vitae **174**

ACKNOWLEDGEMENTS

Along the path of research towards the goal of achieving PhD, I came across and worked with a great number of people, who influenced me in many different ways. Here, I take the opportunity to acknowledge and thank them for their contribution in helping me reach this stage.

At the outset, I wish to express my sincere gratitude towards my Research advisor, Dr. Vidya Gupta, National Chemical Laboratory (NCL), Pune for accepting me as a student and for her continued academic guidance. Her scientific foresight, enthusiastic attitude and openness to ideas were the constant source of inspiration. Every meeting or discussion with her, used to be scientifically and emotionally an enriching experience. She made me think positively and to have confidence in whatever I do. Especially, in the past six months, during the writing of this thesis, I had the privilege of spending a lot of time with her. Daily sessions of discussion and corrections were a great learning experience about patience, diligence and the proper use of scientific language. I am fortunate to have worked under her supervision and grateful for her constant support, encouragement, and sensitive motherly care during the course of my work

I express my sincere thanks to Dr. Sivaram and Dr. Sourav Pal, the former and present Directors of NCL, for providing the research facilities at this institute. I also acknowledge the Council of Scientific and Industrial Research, Government of India for the research fellowship that enabled me to pursue PhD at NCL.

I would like to thank Dr. Abhay Harsulkar, Interactive Research School for Health Affairs (IRSHA) for the initial help during planning of my research objectives and providing valuable inputs regarding the biochemical processes as well as plant system on which I worked. The discussions with him helped me address various inherent problems present in the system. His carefree attitude towards any problem taught me a lot about how to deal with difficult situations. I would also like to thank Dr. P.K. Ranjekar, Director, IRSHA for allowing me to carry out certain experiments at IRSHA as well as few scientifically stimulating interactions that we had regarding my research progress.

I want to express my heartfelt gratitude towards Dr. Narendra Kadoo, NCL for his help in planning some of the experiments, tissue collection trips as well as valuable guidance regarding the use of various software programs for analyzing the data. Naren used to be always there to discuss the day to day experimental outcomes and his innovative suggestions used to be always helpful in achieving the target in a much efficient way. His thorough knowledge about computers and technicalities about various instruments had always been an inspiring fact. Without his help this thesis would not have been in the present form.

I sincerely thank Dr. P.B. Ghorpade and his team at AICRP on Linseed, College of Agriculture, Nagpur for providing the seed material and field facility for growing flax varieties used in my study, for three consecutive years. Without their help, this work won't have been possible. Further, I would also like to thank Dr. Waghmare, CICR, Nagpur for providing liquid N_2 cans and tissue storage facility at his institute during the period of tissue collection.

Dr. Meena Lagu deserves a special mention, as her presence in NCL during the initial phase of my work inspired me in many different ways. Her meticulous planning of any work and perfect record keeping abilities are worth mentioning. Even after she left NCL, she used to be enthusiastic about knowing my work progress and always encouraged me to move ahead towards my goal. I would also like to thank Dr. Mohini Sainani and Dr. Ashok Giri, NCL for encouragements during my research tenure. Further, I thank Dr. S.P. Borikar, Organic chemistry Division, NCL for helping me with part of GC analysis.

I am grateful to Dr. Bhushan Dholakia for being there during my manuscript and thesis writing phase. I sincerely thank him for healthy scientific discussions, sound advice, good company and lastly for critically proof reading my thesis draft, giving an unbiased opinion about it. I would also like to thank Dr. Vishal Dawkar for carefully proof reading my thesis and giving proper advice about the required changes. I sincerely thank Rakesh for helping me with the bioinformatic work required for protein structure determination. His untiring efforts to fulfill all my queries and requests are highly appreciated.

I would like to acknowledge all my friends who helped me during the tissue collection trips to Nagpur. I would specially like to thank Rupali and Vithal from IRSHA, Shweta and all the Flax group members, Varsha, Sandip and Vitthal, from NCL. Without their help my work won't have started. Rupali had been there throughout my working phase and was a great support during my rough patches. I will always cherish her encouragement and friendship.

Plant molecular biology group not only provided a very healthy scientific environment but gave me a chance to get some very loving and caring friends like, Ramya, Rasika and Gayatri. They were there to share the pleasurable as well as sad moments of life. I will always remember the wonderful days spent in their company.

I am thankful to all my past and present lab mates for creating a pleasant working atmosphere in the laboratory. I shared many happy moments with Trupti, Gauri A, Rashmi, Shashi, Sofia, Sadhana, Suvarna, Rohini, Neeta, Aditi, Richa, Laxmi, Ritu, Tejas, Jidnyasa and Charu. I was also happy to have the company of Ajay, Manisha, Manasi T, Vaiju, Ajit, Sirsha, Sagar, Savita, Hemangi, Ram, Krishna, Prashant, Pawan, Radhika, Manje, Elan, Gauri, Pradeep, Hemlata, Prajakta, Manasi M, Medha, Arun, Atul, Priya, Neha K, Yojana, Neha M, Priyanka, Rahul, Amey, BK, Gouri, Sheon, Reema, Sheetal, Yashashree, Nikhil, Rebecca, Yashwant and Ashish. Few new PMB members like Tejas, Amol, Purshottam… are also nice to interact with and always ready to help in their possible ways. I am also thankful to Jagtap Kaka, Trehan, Ramakant and Indira for their help. Affectionate insistence of having the afternoon tea by Jagtap Kaka will be always remembered.

The encouragement, help and good wishes of number of my friends from Pune as well as elsewhere like Aditi, Archana, Vanita, Vinita, Anita, Rasika, Prajakta, Shubhada, Deepali, Chetna, Mayuri, Manini, Anisha and Binoy will be fondly remembered. My dear friends Ashwini and Sunil Tatkar require a special mention, as I will be always in debt to their unflinching support during my most difficult times. Ashwini's gentle persuasion and encouragement to act on any plan, her positive attitude and simple solutions to overcome any problem were really inspiring. Thanks for always being there to talk.

My relatives, more like an extended family have been a constant source of love and encouragement. I would specially mention and thank Taru Maushi, Saroja Kaku, Rahul and Radhika for always extending their help and blessings. Their unconditional timely support will be forever remembered.

My husband, Vikram had been by my side throughout this long and tough journey. He had been always supportive of all the decisions that I took and felt a lot of pride in what I am doing. Thanks for being there with our daughter in my absence or whenever I was busy with my work. Especially these last few months of thesis writing had been really trying and I see a lot of relief and joy as this phase comes to an end.

My dearest daughter Mahika, requires the most heartfelt mention. Her seamless love for everyone and politeness are the most touching part of her nature. I am really sorry for not being there many a times, when she needed me the most. I salute her for always being happy and contented with whatever time I could spend with her. Mahika, you are the source of happiness in our home and even though a small kid, the most understanding and adjusting person I have ever seen. Many thanks for your endless patience and love during this period of my work.

This thesis is like a gift of love to my dear parents, without whose support and perseverance this difficult task would never have been completed. Aai and Baba, I am really short of words to express my feelings towards you and instead, bow to your unflinching support, love and faith in me. I thank you for always being there besides my husband and daughter in most difficult periods and the silent support that you provided to the whole family, sacrificing your comfort. Knowing that you two are there to steer me out of any problem, I could concentrate on my work and reach this stage. I would also like to thank my dear brother, Nitin and sister-in-law, Priya for always being supportive and encouraging about my research work. Dada helped me focus on my priorities and I could always discuss any ups and downs that I underwent during this phase, with him.

A special mention goes to my parent-in-laws who have been really supportive throughout my career. I have no words to express my gratitude for them, especially my mother in-law who is like a pillar of strength and knowledge in my life. Her never say die attitude towards life is worth mentioning. Aai, you have influenced me greatly in so many ways and without your support this work would have remained incomplete. I thank my brother-in-law Girish and co-sister, Gauri for their encouragement and good wishes.

Lastly, I thank all my well-wishers and friends whose names I failed to mention here. Above all, I thank God almighty for giving me this opportunity in life and helping me complete it. This journey had been tough yet, a wonderful and an unforgettable experience.

Ashwini

LIST OF ABBREVIATIONS

AA	Amino acid
ACCase	Acetyl-CoA carboxylase
Acetyl-CoA	Acetyl-coenzyme A
ACP	Acyl carrier protein
AFLP	Amplified fragment length polymorphism
AHC	Agglomerative hierarchical clustering
ALA	α-linolenic acid
AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Variance
bp, kb, Mb	Base pair, kilo-base pair, megabase pair
cDNA	Complementary deoxyribonucleic acid
Ct	Thresh hold cycles
CTAB	Hexadecyl-trimethyl-ammonium bromide
DAA	Days after anthesis
DHA	Docosahexaenoic acid
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
EDTA	Ethylenediamine tetra acetate
EFA	Essential fatty acid
EPA	Eicosapentaenoic acid
ER	Endoplasmic reticulum
FA	Fatty acid
FAD2	Fatty acid desaturase 2/ Δ12 desaturase
FAD3	Fatty acid desaturase 3/ Δ15 desaturase
FAME	Fatty acid methyl esters
FID	Flame ionization detector
g	Relative centrifugal force
g, mg, µg, ng	Gram, milligram, microgram, nanogram
GC	Gas chromatography
ha	Hectare
hr, min, s	Hour, minute, second
IRAP	Inter-retrotransposon amplified polymorphism
ISSR	Inter-simple sequence repeat
KAS	3-ketoacyl-ACP synthases
l, ml, µl	Liter, milliliter, microliter
LA	Linoleic acid
LCPUFA	Long chain polyunsaturated fatty acids
M, mM, µM, N	Molar, millimolar, micromolar, normal
MA	Myristic acid
MUFA	Monounsaturated fatty acid
NMR	Nuclear magnetic resonance

OA	Oleic acid
PA	Palmitic acid
PCA	Principal Component Analysis
PCR	Ploymerase chain reaction
PIC	Polymorphism information content
PUFA	Polyunsaturated fatty acids
qRT-PCR	Real-time polymerase chain reaction
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rpm	Revolutions per minute
RQ	Relative expression
SA	Stearic acid
SAD	Stearoyl-ACP desaturase/ Δ9 desaturase
SDG	Secoisolariciresinoldiglycoside
SD	Standard deviation
SE	Standard error
SFA	Saturated fatty acids
SPI	ISSR primer index
SSR	Simple sequence repeat
U	Units (enzymatic)
UPGMA	Unweighted Pair Group Method with Arithmetic Mean

LIST OF TABLES

Table 1.1: Examples of saturated fatty acids

Table 1.2: Examples of unsaturated fatty acids

Table 1.3: Fatty acid desaturase classes

Table 1.4: List of plants from which Δ 9, Δ12 and Δ 15 desaturase genes are isolated

Table 1.5: Proximate composition of flax based on common measures

Table 2.1: Fatty acid composition and seed oil content in flax genotypes

Table 2.2: The ISSR primers used, number of amplified & polymorphic loci and percentage polymorphism in evaluation of genetic diversity

Table 2.3: Fatty acid content (%) among the groups of flax genotypes

Table 2.4A: Probability values among the groups of flax genotypes based on ALA data.

Table 2.4B: Probability values among the groups of flax genotypes based on LA data

Table 2.4C: Probability values among the groups of flax genotypes based on OA data

Table 2.4D: Probability values among the groups of flax genotypes based on SA data

Table 3.1: Primers used for real-time PCR

Table 3.2: ALA contents (%), oil content (%) and absolute ALA content of ten flax varieties

Table 3.3: Fatty acid (FA) levels in variable ALA groups. Average content of six FAs (% ± SE) *viz*., Myristic (MA), Palmitic (PA), Stearic (SA), Oleic (OA), Linoleic (LA) and Linolenic (ALA) in high and low ALA groups and TL23

Table 3.4: Relative expression (RQ) maxima of *SAD* genes. Average RQ maxima (in folds) of SAD genes in high and low ALA groups and TL23

Table 3.5: Relative expression (RQ) maxima of *FAD2* genes. Average RQ maxima (in folds) of *FAD2* genes in high and low ALA groups and TL23

Table 3.6: Relative expression (RQ) maxima of *FAD3* genes. Average RQ maxima (in folds) of *FAD3* genes in high and low ALA groups and TL23

Table 4.1: Gene specific PCR primer sequences (5′→3′) used for amplification of *SAD*, *FAD2* and *FAD3* genes

Table 4.2: *SAD2* gene and AA variation in ten flax varieties

Table 4.3: *FAD2* gene and AA variation in ten flax varieties

Table 4.4: *FAD3A* gene and AA variation in ten flax varieties

Table 4.5: *FAD3B* gene and AA variation in ten flax varieties

Table 4.6: Allelic variations and number of protein isoforms identified in six desaturase genes among ten flax varieties

Table 4.7: Representation of number of haplotypes and their allelic frequency, for each desaturase gene, in the ten varieties. Values in parentheses represent allelic frequency

Table 4.8: Haplotype definition of *SAD1* gene

Table 4.9: Haplotype definition of *SAD2* gene

Table 4.10: Haplotype definition of *FAD2* gene

Table 4.11: Haplotype definition of *FAD2-2* gene

Table 4.12: Haplotype definition of *FAD3A* gene for nine varieties except NL260

Table 4.13: Haplotype definition of *FAD3B* gene

LIST OF FIGURES

Figure 1.1: Molecular structure of **(A)** Linoleic acid, LA, 18:2, ω-6 (Contains 18 carbon atoms and 2 double bonds); **(B)** α-Linolenic acid, ALA, 18:3, ω-3 (Contains 18 carbon atoms and 3 double bonds); **(C)** Arachidonic acid, AA, 20:4, ω-6 (Contains 20 carbon atoms and 4 double bonds); **(D)** Eicosapentaenoic acid, EPA, 20:5, ω-3 (Contains 20 carbon atoms and 5 double bonds); **(E)** Docosahexaenoic acid, DHA, 22:6, ω-3 (Contains 22 carbon atoms and 6 double bonds); and **(F)** Triglycerol molecule esterified with three fatty acids

Figure 1.2: Metabolic pathway of ω-6 and ω-3 fatty acids in humans. Arachidonic acid (AA) and eicosapentaenoic acid (EPA) produce pro-inflamatory and anti-inflammatory eicosanoids, respectively

Figure1.3: Fatty acid composition of some nuts, oilseeds and fruit

Figure1.4: Hypothetical scheme of the relative percentages of fat and various fatty acid families in human nutrition as extrapolated from cross sectional analyses of contemporary hunter-gatherer populations and from longitudinal observations and their putative changes during the preceding 100 years (y). *Trans* fatty acids, the result of the hydrogenation process, have increased dramatically in the food supply during this century

Figure 1.5: Fatty acid biosynthesis pathway in plants. After each condensation with 3-ketoacyl-ACP synthase (KAS), the 3-ketoacyl-ACP product is reduced (reaction 4), dehydrated (reaction 5), and reduced again (reaction 6), by 3-ketoacyl-ACP reductase, 3-hydroxyacyl-ACP dehydrase, and enoyl-ACP reductase, respectively

Figure 1.6: Cartoon depicting structure of membrane-bound desaturases. 'Fe' represents a putative di-iron active site and 'H' represents histidine-rich structural motifs involved in coordinating the di-iron active site

Figure 1.7: Comparison of saturated and unsaturated fatty acids in dietary fats and oils

Figure 1.8: Flax plant morphology

Figure 1.9: Variation of colour in **(A)** flowers and **(B)** seeds of flax varieties

Figure 1.10: Stem length (defined by the length of the stem without side branches) variation observed in the flax varieties. It can be measured in cm or as a fraction of the stem with branching compared to the entire stem length (see figures beside branches). Basal branches are not considered. **(A)** and **(B)** typical fiber flax with less than 1/4 of the stem branched; **(C)** intermediate flax; **(D)** large-seeded flax (oilseed flax)

Figure 1.11: Production of oilseed flax or Linseed in top few countries in the world FAOSTAT data 2010)

Figure 1.12: Major flax/Linseed producing states in India

Figure 1.13: Position of India in the world for **(A)** area under flax cultivation, **(B)** yield per hector and **(C)** production in tones

Figure 2.1: Amplification profiles of 70 genotypes using **(A)** UBC 836 primer and **(B)** UBC 841 primer. 'M' represents ΦX174/*Hae*III molecular marker

Figure 2.2: Dendrogram based on Jaccard's coefficient of genetic distances

Figure 2.3: Principal Component Analysis scatter plot of the flax genotypes

Figure 2.4: Principal Component Analysis (PCA) - distribution of genotypes based on the fatty acid composition (all the five FAs *viz*., PA, SA, OA, LA and ALA, were considered for the analysis)

Figure 3.1: **(A)** Flower tagging, **(B)** Eight developing seed stages of flax

Figure 3.2: Trend of fatty acid accumulation in eight developing seed stages of **(A)** high and **(B)** low ALA containing varieties (considering average FA values of high and low ALA flax varieties)

Figure 3.3: Fatty acid profiles of ten flax varieties in eight developing seed stages. **(A-E):** high ALA varieties, **(F-I):** low ALA varieties and **(I):** mutant variety

Figure 3.4: Fatty acid profiles of ten flax varieties in eight seed developmental stages. We analyzed six fatty acids [Myristic (MA), Palmitic (PA) Stearic (SA), Oleic (OA), Linoleic (LA) and Linolenic (ALA)] from eight seed developmental stages, from flower to mature stage (48 DAA) in flax

Figure 3.5: PCA plots. Score plot and loading plot (inset) of Principal Component Analysis (PCA) of four fatty acids *viz*., Stearic (SA), Oleic (OA), Linoleic (LA) and Linolenic (ALA) acids in the mature seeds of the ten flax varieties

Figure 3.6: *SAD* gene expression. Relative expression of *SAD* genes in seed developmental stages of high and low ALA flax variety groups, **(A)** *SAD1* gene expression in high ALA group, **(B)** *SAD1* gene expression in low ALA group , **(C)** *SAD2* gene expression in high ALA group, **(D)** *SAD2* gene expression in low ALA group

Figure 3.7: Expression range of *SAD* genes in high and low ALA groups and TL23. **(A)** Expression range of *SAD1*, **(B)** Expression range of *SAD2*

Figure 3.8: *FAD2* gene expression. Relative expression of *FAD2* genes in seed developmental stages of high and low ALA flax variety groups, **(A)** *FAD2* gene expression in high ALA group, **(B)** *FAD2* gene expression in low ALA group , **(C)** *FAD2-2* gene expression in high ALA group, **(D)** *FAD2-2* gene expression in low ALA group

Figure 3.9: Expression range of *FAD2* genes in high and low ALA groups and TL23. **(A)** Expression range of *FAD2*, **(B)** Expression range of *FAD2-2*

Figure 3.10: *FAD3* gene expression. Relative expression of *FAD3* genes in seed developmental stages of high and low ALA flax variety groups, **(A)** *FAD3A* gene expression in high ALA group, **(B)** *FAD3A* gene expression in low ALA group, **(C)** *FAD3B* gene expression in high ALA group, **(D)** *FAD3B* gene expression in low ALA group

Figure 3.11: Expression range of *FAD3* genes in high and low ALA groups and TL23. **(A)** Expression range of *FAD3A*, **(B)** Expression range of *FAD3B*

Figure 4.1: Alignment of deduced amino acid sequences of *SAD1* gene from ten flax varieties with *SAD1* sequence in NCBI and Phytozome databases. Solid boxes represent conserved di-iron binding motifs [(D/E) X2 H]

Figure 4.2: Alignment of deduced amino acid sequences of *SAD2* gene from ten flax varieties with *SAD2* sequence in NCBI and Phytozome databases. Solid boxes represent conserved di-iron binding motifs [(D/E) X2 H] and **(.)** represents variation in AA sequence

Figure 4.3: Alignment of deduced amino acid sequences of *FAD2* gene from ten flax varieties with two *FAD2* sequences in NCBI database. Solid boxes represent conserved histidine motifs and dotted box represents ER retrieval motif. Variation in the AA sequences is represented by **(.)** and **(:)**

Figure 4.4: Alignment of deduced amino acid sequences of *FAD2-2* gene from ten flax varieties with *FAD2-2* sequence in NCBI database. The boxes represent conserved histidine motifs

Figure 4.5: Alignment of deduced amino acid sequences of *FAD3A* gene from ten flax varieties with *FAD3A* sequence in NCBI, Phytozome and Linum.ca databases. The boxes represent conserved histidine motifs and dotted box represents di-lysine motif. Variation in the AA sequence is represented by **(:)**. Stop codon in TL23 is represented by * in the AA sequence

Figure 4.6: Alignment of deduced amino acid sequences of *FAD3B* gene from ten flax varieties with *FAD3B* sequence in NCBI, Phytozome and Linum.ca databases. The boxes represent conserved histidine motifs and dotted box represents di-lysine motif. Variation in the AA sequence is represented by **(:)**

Figure 4.7: Ramchandran plot of the model generated for **(A)** SAD1 and **(B)** SAD2 protein

Figure 4.8: Z-score plots of the model generated for **(A)** SAD1 and **(B)** SAD2 protein indicated by black filled circle (Z-score for SAD1: -7.06 and Z-score for SAD2: -5.29)

Figure 4.9: ProSA based energy profiles of the model generated for **(A)** SAD1 and **(B)** SAD2 proteins

Figure 4.10: **(A)** SAD1 and **(B)** SAD2 protein models (helix region in blue represents two conserved di-iron binding motifs)

Figure 4.11: NCBI SAD2 protein model (grey) superimposed with **(A)** SAD2 protein model of NL260 (green) (representative of all the varieties), **(B)** SAD2 protein model of TL23 (green). The region in red indicates the position of the AA change while helix region in blue represents two conserved di-iron binding motifs

Figure 4.12: **(A)** Ramchandran plot **(B)** Z-score plot (Z-score: -3.7, indicated by black filled circle) **(C)** ProSA based energy profile of the I-TASSER generated model for FAD2 protein

Figure 4.13: **(A)** FAD2 protein model, **(B)** NCBI FAD2 protein model (grey) superimposed with Padmini FAD2 protein model (green). The regions in red (in the superimposition model, **B**) indicate the positions of the AA change. Region in blue represents three conserved His-box motifs H1, H2 and H3 in both the structures **A** and **B**

Figure 4.14: Ramchandran plots of the I-TASSER generated model for **(A)** FAD3A and **(B)** FAD3B protein

Figure 4.15: Z-score plots of the homology model generated for **(A)** FAD3A **(B)** FAD3B protein indicated by black filled circle (Z-score of FAD3A: -2.56 and FAD3B: -2.79)

Figure 4.16: ProSA based energy profiles of the homology model generated for **(A)** FAD3A and **(B)** FAD3B protein

Figure 4.17: FAD3A protein model (helix region in blue represents three conserved His-box motifs, H1, H2 and H3)

Figure 4.18: NCBI FAD3A protein model (grey) superimposed with **(A)** FAD3A protein model of NL260 (green), **(B)** FAD3A protein model of NL97 (green), **(C)** FAD3A protein model of TL23 (brick red). The regions in red indicate the position of AA changes while helix regions in blue represent three conserved His-box motifs. In TL23, the protein is truncated and the H3 motif is absent. The superimposition model shows incomplete alignment of TL23 model with the NCBI model (Only grey motifs are visible in the figure C from where the TL23 protein is truncated)

Figure 4.19: FAD3B protein model (region in blue represents three conserved His-box motifs, H1, H2 and H3)

Figure 4.20: NCBI FAD3A protein model (grey) superimposed with **(A)** FAD3A protein model of NL97 (green) (representative of 8 varieties other than NL260 and Padmini), **(B)** FAD3A protein model of TL23 (green). The regions in red indicate the positions of AA change while the regions in blue represent three conserved His-box motifs

---

THESIS ABSTRACT

Common flax or linseed (*Linum usitatissimum* L.) is a versatile crop grown since prehistoric times for seeds and fiber. More recently, flax seed oil has come into focus due to its fatty acid (FA) composition. It is the richest agricultural source of α-linolenic acid (ALA), an essential dietary polyunsaturated fatty acid of ω-3 class (Morris, 2007), which serves as a precursor for biologically active longer chain poly-unsaturated fatty acids (PUFA), mainly eicosapentanoic acid (EPA) and decosahexanoic acid (DHA). They are components of cell membranes and play important metabolic roles. The presence of high levels of ALA in flax seed oil has lead to its use in both, industrial as well as food and feed products including a wide variety of nutraceuticals and health foods. Further, flax seeds are also a rich source of soluble and insoluble dietary fibers as well as lignan, a phytoestrogen compound, which add to its positive health attributes. Keeping in view the nutritional importance and future market requirement of flax, I initiated my work towards the molecular and biochemical analysis of Indian flax varieties.

The first part of my work involved analysis of genetic and fatty acid diversity among 70 flax genotypes actively utilized in Indian flax breeding programs. I used the PCR based ISSR markers to reveal the genetic polymorphism. We detected 63.9% average loci polymorphism among these genotypes using 12 consistently amplifying polymorphic ISSR primers and the average gene diversity (H_i) observed among the genotypes was 0.15, which was low. Both the results were indicative of overall low genetic diversity in these 70 flax genotypes. Based on their genetic diversity, the 70 flax genotypes were grouped in five clusters using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA). Fatty acid analysis of the same genotypes showed the presence of five major fatty acids with predominance of 18 carbon species, namely, α-linolenic acid (ALA 18:3), linoleic acid (LA 18:2), oleic acid (OA 18:1) and stearic acid (SA 18:0) and 16 carbon species, namely palmitic acid (PA 16:0). Correlation of FA contents with the genetic clusters exhibited a meaningful segregation of two clusters with high ALA content and one cluster with that of LA. The genotypes in these clusters suggested genetic differences associated with these chemotypes. Further, the Mantel's test revealed significant statistical correlation between the molecular variation and the LA and ALA variation.

Flax varieties differ markedly in ALA levels. Fatty acid desaturases play key roles in accumulating ALA in seed. Considering this, we planned our second aspect of the work where we performed FA profiling of various seed developmental stages of ten Indian flax varieties including one mutant variety, TL23. Depending on their ALA contents, these varieties were grouped under high ALA and low ALA groups. Transcript profiling of six microsomal desaturase genes (*SAD1*, *SAD2*, *FAD2*, *FAD2-2*, *FAD3A* and *FAD3B*), which act sequentially in the fatty acid desaturation pathway, was performed using real-time PCR. We observed gene specific as well as temporal expression pattern of all the desaturases and the expression profiles correlated well with the FA data for the groups. From our analysis we concluded that, though all the desaturase genes are required for the final ALA accumulation, four genes (*SAD1, FAD2, FAD3A* and *FAD3B*) might be specifically responsible for its differential accumulation.

The third and the last part of my study consists of sequence characterization of all the desaturase genes from the same ten varieties to find out if there is presence of any nucleotide sequence variation leading to AA variation and whether these AA sequence variants have any functional implications in the form of differential ALA accumulation. The *SAD* and *FAD2* genes did not show much sequence variations which would translate into nonsynonymous AA substitutions and hence resulted into only one or two haplotypes and protein isoforms among the ten varieties. On the other hand, both *FAD3A* and *FAD3B* genes showed more nucleotide variations and few of them resulted into nonsynonymous AA substitutions, and showed more haplotypes and protein isoforms compared to the other two genes. *FAD3A* gene of TL23 had a premature stop codon, which resulted in a truncated protein and consequently very low levels of ALA accumulation. Besides the mutant variety TL23, some of the high ALA group varieties, like NL260 and Padmini, formed different haplotypes and protein isoforms for both *FAD3A* and *FAD3B*. Further, protein structures were also predicted for all the desaturases and they were compared among the ten flax varieties. There were no structural variations observed due to any of the AA changes in the desaturase proteins of these flax varieties except for TL23.

CHAPTER 1

Introduction and review of literature

Section 1: Fatty acids: saturated and unsaturated

Fatty acids (FAs) are carboxylic acids with a long unbranched aliphatic chain, which is either saturated or unsaturated. Saturated FAs are the ones without any double bond between the carbon atoms, and hence have a straight chain configuration. On the other hand, unsaturated FAs contain double bonds (-CH=CH-) in the FA chain and can be further differentiated as monounsaturated, i.e., having a single double bond, or polyunsaturated, i.e., having more than one double bonds. Fatty acid fluidity increases with increasing number of double bonds. These double bonds may have a *cis* or *trans* configuration. Fatty acids with *cis* bonds are less flexible and a chain with many *cis* bonds becomes curved. As a result of this, the ability of fatty acids with *cis* bonds to be closely packed is limited when they are part of a phospholipid in a lipid bilayer, or triglycerides in storage lipid and which in turn affects the melting temperature of the membrane or of the fat. In contrast, *trans* bonds do not cause the chain to bend much, and FAs with *trans* bonds have shape similar to straight saturated FAs. However, most naturally occurring unsaturated FAs contain *cis* bonds while *trans* bonds are characteristically produced due to human processing, such as during industrial hydrogenation of plant oils. The differences in geometry between the various types of unsaturated FAs, as well as between saturated and unsaturated FAs, play an important role in biological processes, and in the construction of biological structures such as cell membranes. Some examples of both saturated and unsaturated FAs are listed in Tables 1.1 and 1.2.

Table 1.1: Examples of saturated fatty acids

Fatty acid	Chemical structure	Carbon chain length: double bonds (C:D)
Caprylic acid	$CH_3(CH_2)_6COOH$	8:0
Capric acid	$CH_3(CH_2)_8COOH$	10:0
Lauric acid	$CH_3(CH_2)_{10}COOH$	12:0
Myristic acid	$CH_3(CH_2)_{12}COOH$	14:0
Palmitic acid	$CH_3(CH_2)_{14}COOH$	16:0
Stearic acid	$CH_3(CH_2)_{16}COOH$	18:0
Arachidic acid	$CH_3(CH_2)_{18}COOH$	20:0
Behenic acid	$CH_3(CH_2)_{20}COOH$	22:0
Lignoceric acid	$CH_3(CH_2)_{22}COOH$	24:0
Cerotic acid	$CH_3(CH_2)_{24}COOH$	26:0

Fatty acids, being lipids, are generally water-insoluble but highly soluble in organic solvents. They function as fuel molecules or sources of reserve energy and serve as components of many other classes of lipids, including triglycerides (commonly known as “fats”) and phospholipids, which are important building blocks of biological membranes. FA derivatives also function as hormones, examples are eicosanoids (the principal groups of hormones of this class are prostaglandins, prostacyclins, leukotrienes and thromboxanes) (Lands, 1991; Wada et al., 2007) and intracellular messengers, which negatively or positively control the steps involved in the signal transduction, particularly at the level of cell membranes (Sumida et al., 1993; Graber et al., 1994). In plants, FAs play a significant role in pathogen defense as biosynthetic precursors for cuticular components or the phytohormone jasmonic acid as well as participate in modulating basal, effector-triggered, and systemic immunity (Kachroo and Kachroo, 2009).

Table 1.2: Examples of unsaturated fatty acids

Fatty acid	Chemical structure	Δ^x	C:D	ω−x
Myristoleic acid	$CH_3(CH_2)_3$**CH=CH**$(CH_2)_7COOH$	*cis*-Δ^9	14:1	ω −5
Palmitoleic acid	$CH_3(CH_2)_5$**CH=CH**$(CH_2)_7COOH$	*cis*-Δ^9	16:1	ω −7
Sapienic acid	$CH_3(CH_2)_8$**CH=CH**$(CH_2)_4COOH$	*cis*-Δ^6	16:1	ω −10
Oleic acid	$CH_3(CH_2)_7$**CH=CH**$(CH_2)_7COOH$	*cis*-Δ^9	18:1	ω −9
Elaidic acid	$CH_3(CH_2)_7$**CH=CH**$(CH_2)_7COOH$	*trans*-Δ^9	18:1	ω −9
Vaccenic acid	$CH_3(CH_2)_5$**CH=CH**$(CH_2)_9COOH$	*trans*-Δ^{11}	18:1	ω −7
Linoleic acid	$CH_3(CH_2)_4$**CH=CH**CH_2**CH=CH** $(CH_2)_7COOH$	*cis,cis*-Δ^9,Δ^{12}	18:2	ω −6
Linoelaidic acid	$CH_3(CH_2)_4$**CH=CH**CH_2**CH=CH** $(CH_2)_7COOH$	*trans,trans*- Δ^9,Δ^{12}	18:2	ω −6
α-Linolenic acid	CH_3CH_2**CH=CH**CH_2**CH=CH**CH_2 **CH=CH**$(CH_2)_7COOH$	*cis,cis,cis*- Δ^9,Δ^{12},Δ^{15}	18:3	ω −3
Arachidonic acid	$CH_3(CH_2)_4$**CH=CH**CH_2**CH=CH**CH_2 **CH=CH**CH_2**CH=CH**$(CH_2)_3COOH$	*cis,cis,cis,cis*- $\Delta^5\Delta^8$,Δ^{11},Δ^{14}	20:4	ω −6
Eicosapentaenoic acid	CH_3CH_2**CH=CH**CH_2**CH=CH**CH_2 **CH=CH**CH_2**CH=CH**CH_2**CH=CH** $(CH_2)_3COOH$	*cis,cis,cis,cis,cis*- Δ^5,Δ^8,Δ^{11},Δ^{14},Δ^{17}	20:5	ω −3
Erucic acid	$CH_3(CH_2)_7$**CH=CH**$(CH_2)_{11}COOH$	*cis*-Δ^{13}	22:1	ω −9
Docosahexaenoic acid	CH_3CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=** **CH**CH_2**CH=CH**CH_2**CH=CH**CH_2**CH=** **CH**$(CH_2)_2COOH$	*cis,cis,cis,cis,cis,* *cis*- Δ^4, Δ^7, Δ^{10}, Δ^{13}, Δ^{16}, Δ^{19}	22:6	ω −3

Δ^x : Position of the double bond from the carboxylic end.

ω−x: Position of the double bond from the methyl end or ω-end

1.1 Monounsaturated fatty acids

Monounsaturated fatty acids (MUFAs) are distinguished from the other fatty acid classes on the basis of having only one double bond in their fatty acid chain. The carbon chain length can vary from 14 to 18, with a predominance of 16 and 18 carbon MUFAs. Compared with saturated fatty acids (SFAs), MUFAs lower total and LDL cholesterol levels, and relative to carbohydrate, they increase HDL cholesterol levels and decrease plasma triglyceride levels (Kris-Etherton et al., 1999). High fat fruits such as olives and avocados are some of the richest sources of MUFAs. Besides these, they are found in natural foods such as red meat, whole milk products, nuts etc.

1.2 Polyunsaturated fatty acids

Polyunsaturated fatty acids (PUFAs) are essential components of human health. The lengths of carbon chain in PUFAs vary from 18 to 22 carbons. Depending on the position of the first double bond from the ω end (methyl end) in the chain, PUFAs are denoted as omega-6 (ω-6, at 6th carbon) and omega-3 (ω-3, at 3rd carbon) fatty acid (alternatively, n-6 or n-3). These are considered as essential PUFAs for humans, as they cannot be synthesized in the human body and must be obtained from dietary sources. Linoleic acid (LA) and α-linolenic acid (ALA) are the parent essential fatty acids (EFAs) of ω-6 and ω-3 class, respectively. LA and ALA serve as precursors to longer chain polyunsaturated fatty acids (LCPUFAs) like arachidonic acid (AA), eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (Figure 1.1). Humans and animals can in general convert LA to AA and ALA to EPA and DHA (de Gomez and Brenner, 1975). They are further required in the synthesis of biologically active prostaglandins, thromboxanes and leukotrienes. Depending on the parent compound from which they are synthesized, these compounds are either involved in the pro- or anti-inflammatory pathways, active during disease conditions (Figure 1.2). The balance between these mutually antagonistic compounds determines the final outcome of the disease process. EFAs also help to maintain the barrier of the skin and are involved in cholesterol metabolism (Horrobin, 1990). While LA is plentiful in human diet and is found in the seeds of most food plants, sources of ALA are few. In high amount, ALA is only found in the chloroplast of green leafy vegetables like spinach, purslane etc., and in seeds of a few plants like flax, soybean, canola, hemp, walnut etc. (Figure 1.3).

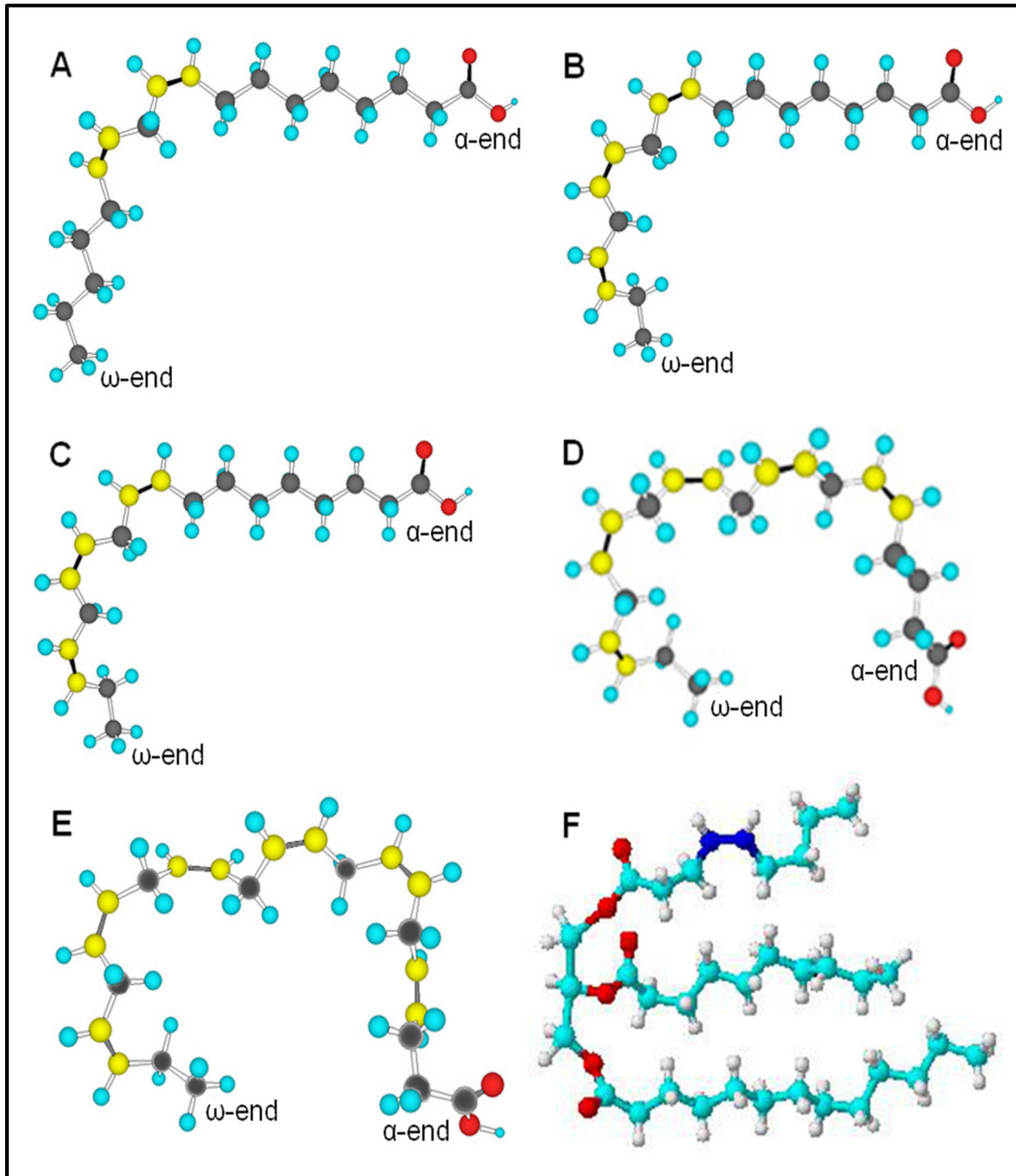


Figure 1.1: Molecular structure of **(A)** Linoleic acid, LA, 18:2, ω-6 (contains 18 carbon atoms and 2 double bonds); **(B)** α-Linolenic acid, ALA, 18:3, ω-3 (contains 18 carbon atoms and 3 double bonds); **(C)** Arachidonic acid, AA, 20:4, ω-6 (contains 20 carbon atoms and 4 double bonds); **(D)** Eicosapentaenoic acid, EPA, 20:5, ω-3 (contains 20 carbon atoms and 5 double bonds); **(E)** Docosahexaenoic acid, DHA, 22:6, ω-3 (contains 22 carbon atoms and 6 double bonds); and **(F)** Triglycerol molecule esterified with three FAs. [Grey balls: carbon atoms; Yellow balls: carbon atoms between which double bonds are present; Turquoise balls: hydrogen atoms; Red balls: oxygen atoms]. (Adapted from http://www.omega3learning.uconn.edu)

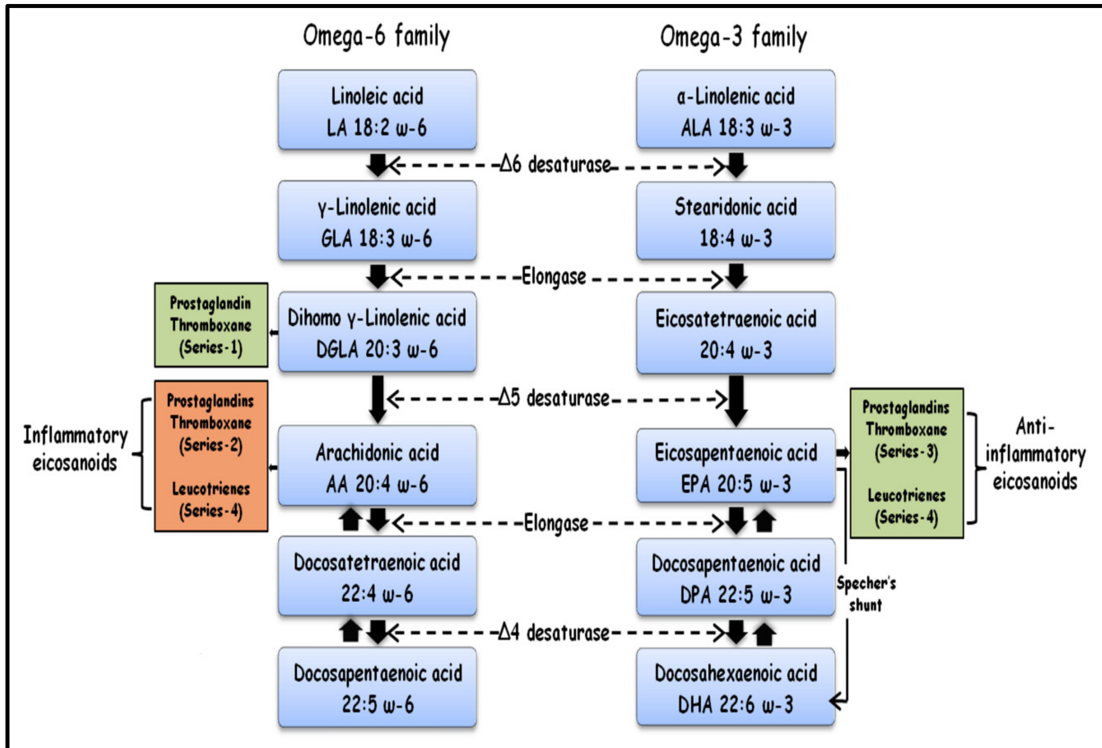


Figure 1.2: Metabolic pathway of ω-6 and ω-3 fatty acids in humans. Arachidonic acid (AA) and eicosapentaenoic acid (EPA) produce pro-inflamatory and anti-inflamatory eicosanoids, respectively

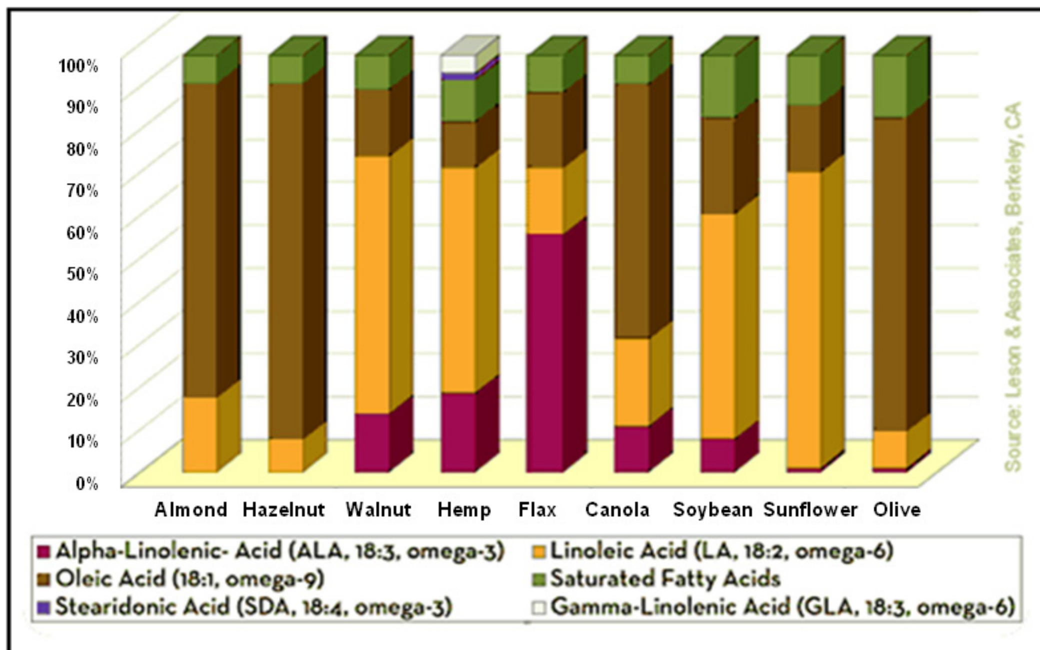


Figure 1.3: Fatty acid composition of some nuts, oilseeds and fruit compared to flax. (Source: Lesson & associate, Berkeley, CA)

1.2.1 Importance of ω-6 and ω-3 fatty acid balance in the diet

Several studies have suggested that man evolved on a diet that was much lower in saturated FAs than is today's diet (Eaton and Konner, 1985). Moreover, the diet contained small and nearly equal proportion of ω-6 and ω-3 long-chain FAs, having a ratio of about 1:1. Whereas, today this ratio has increased and is 20-25:1, indicating that modern diets are severely deficient in ω-3 FAs compared to the diet on which humans evolved and their genetic patterns were established (Eaton and Konner, 1985; Simopoulos, 1995).

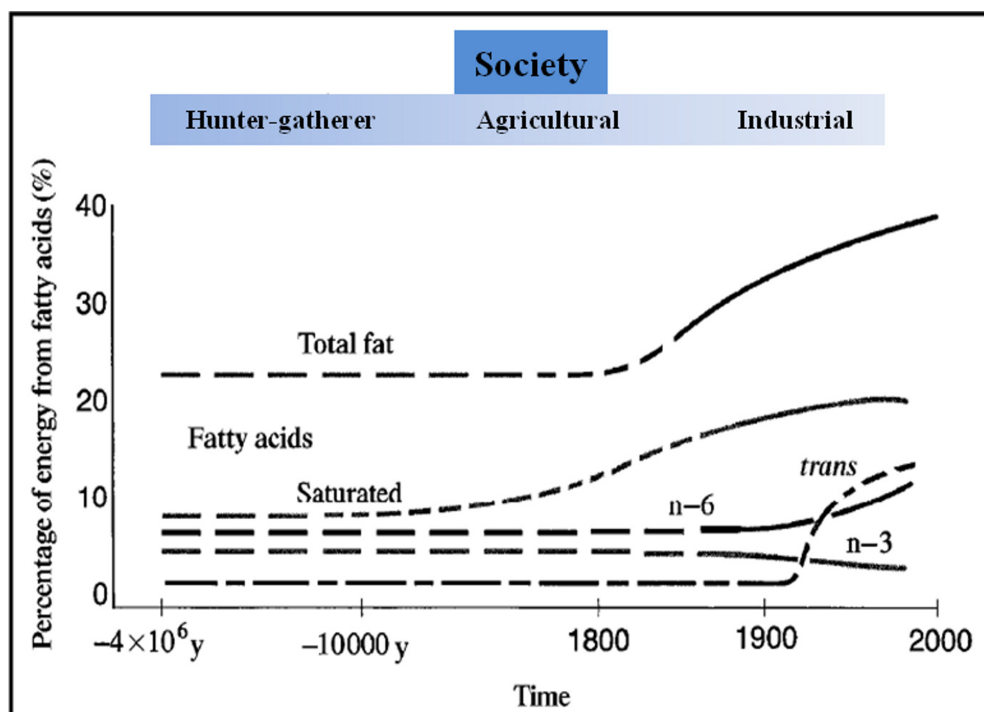


Figure1.4: Hypothetical scheme of the relative percentages of fat and various FA acid families in human nutrition as extrapolated from cross sectional analyses of contemporary hunter-gatherer populations and from longitudinal observations and their putative changes during the preceding 100 years (y). *Trans* fatty acids, the result of the hydrogenation process, have increased dramatically in the food supply during this century (adapted from Leaf and Weber, 1987)

There has been a major shift in the human fat consumption pattern over the past century because of the technological developments (Figure 1.4). Specifically, there has been an increase in the intake of *trans* FAs, found mainly in products made with hydrogenated vegetable oils, and ω-6 FAs, found in vegetable oils and animal products derived from grain-fed livestock. Further, modern agriculture and animal domestication, with their emphasis on production, have decreased ω-3 FA contents in

many food sources, such as, green leafy vegetables, animal meats, eggs and even fish (Simopoulos, 1991). All of these have led to the increase in the ratio of ω-6/ω-3 in the present day diets. High amount of ω-6 FAs in the body leads to the production of pro-inflammatory eicosanoids, which may contribute in the pathogenesis of many diseases, including cardiovascular diseases, cancer, osteoporosis and inflammatory and autoimmune diseases; whereas, increased levels of ω-3 PUFA (a lower ω-6/ω-3 ratio), exert suppressive effects by its anti-inflammatory eicosanoid production (Simopoulos, 2002; Mozaffarian, 2005; Mori, 2006).

According to the joint recommendation by Food and Agriculture Organization and World Health Organization (FAO/WHO) committee, the ideal ω-6/ω-3 ratio should be between 5:1 and 10:1. On the other hand, various studies indicate that the optimal ω-6/ω-3 ratio may vary in disease conditions (Mozaffarian, 2005). The therapeutic dose of ω-3 FAs will depend on the degree of severity of disease resulting from the genetic predisposition of an individual. A lower ratio of ω-6/ω-3 FAs is more desirable in reducing the risk of many chronic diseases of high prevalence in Western societies, as well as in the developing countries. Therefore, to promote normal growth and development as well as to maintain good human health, it is essential to rectify the imbalance in the ω-6/ω-3 ratio by restoring sources of ω-3 FAs in the diet. There are few agricultural sources of high ALA (ω-3 FA), such as green leafy vegetables, chia, perilla, hemp, flax, purslane etc. However, flaxseed with its high level of ALA and ω-6/ω-3 ratio of 0.3 to1.0 can help to restore the ω-6 and ω-3 FA balance in the human body (Anonymous, 2001).

1.3 Fatty acid biosynthesis in plants

The fatty acid biosynthesis pathway is considered as a primary metabolic pathway, since it is found in every cell of a plant and is essential for its growth. In plants, the major site of the FA synthesis is within plastid, which is unlike that in animals and fungi, where the fatty acids are primarily synthesized in the cytosol. The pool of acetyl-coenzyme A (Acetyl-CoA) present in the plastid serves as the carbon source for fatty acid synthesis (Post-Beittenmiller et al., 1991; Post-beittenmiller et al., 1992). In the first committed step for *de novo* FA synthesis, malonyl-CoA is formed from acetyl-CoA by addition of CO_2 using the biotin prosthetic group of the enzyme acetyl-CoA carboxylase (ACCase) and serves as the carbon donor for subsequent elongation reactions of fatty acyl chain (Harwood, 1988). Further, FA synthesis in the

plastids occurs through a repeated series of condensation, reduction, and dehydration reactions that add two carbon units derived from malonyl ACP to the elongating FA chain. Overall, nearly 30 enzymatic reactions are required to produce a 16- or 18-carbon FA from acetyl-CoA and malonyl-CoA (Ohlrogge and Browse, 1995). Before entering the FA synthesis pathway, the malonyl group of malonyl-CoA is transferred from CoA to a protein cofactor, acyl carrier protein (ACP) with the help of enzyme malonyl-CoA:ACP transacylase (Shimakata and Stumpf, 1983a). All the further reactions of the pathway involve ACP until the 16- or 18-carbon product is ready for transfer to glycerolipids or for export from the plastid. Malonyl-ACP undergoes a number of condensation reactions mediated by a set of at least three separate condensing enzymes, known as 3-ketoacyl-ACP synthases (KAS) to produce an 18-carbon fatty acid. The first condensation of acetyl-CoA and malonyl-ACP to form a four-carbon product is catalyzed by KAS III. A second condensing enzyme, KAS I, is responsible for producing chain lengths from 6 to 16 carbons. Finally, KAS II is the enzyme required for the elongation of the 16 carbon palmitoyl-ACP to 18 carbon stearoyl-ACP (Shimakata and Stumpf, 1983b; Jaworski et al., 1989). In addition to these three enzyme catalyzed reactions, after each condensation step, the 3-ketoacyl-ACP product is reduced, dehydrated, and reduced again, by 3-ketoacyl-ACP reductase, 3-hydroxyacyl-ACP dehydrase, and enoyl-ACP reductase, respectively. As a result, saturated FA that is two carbons longer than at the start of the cycle is obtained (Figure 1.5).

The elongation of FAs in the plastids is terminated when the acyl group is removed from ACP by either of the two enzyme systems, an acyl-ACP thioesterase or acyltransferases in the plastid. Acyl-ACP thioesterase hydrolyzes the acyl-ACP and releases free FA, whereas acyltransferases transfer the FA from ACP to glycerol-3-phosphate or to monoacylglycerol-3-phosphate. The transport of FA outside plastid is determined on the basis of their release from ACP by either a thioesterase or an acyltransferase. Only when thioesterase acts on acyl-ACP, free FAs can leave the plastid. On the outer membrane of the chloroplast envelope, an acyl-CoA synthetase assembles an acyl-CoA thioester that is then available for acyltransferase reactions to form glycerolipids in the endoplasmic reticulum (ER) (Harwood, 1988; Ohlrogge and Browse, 1995).

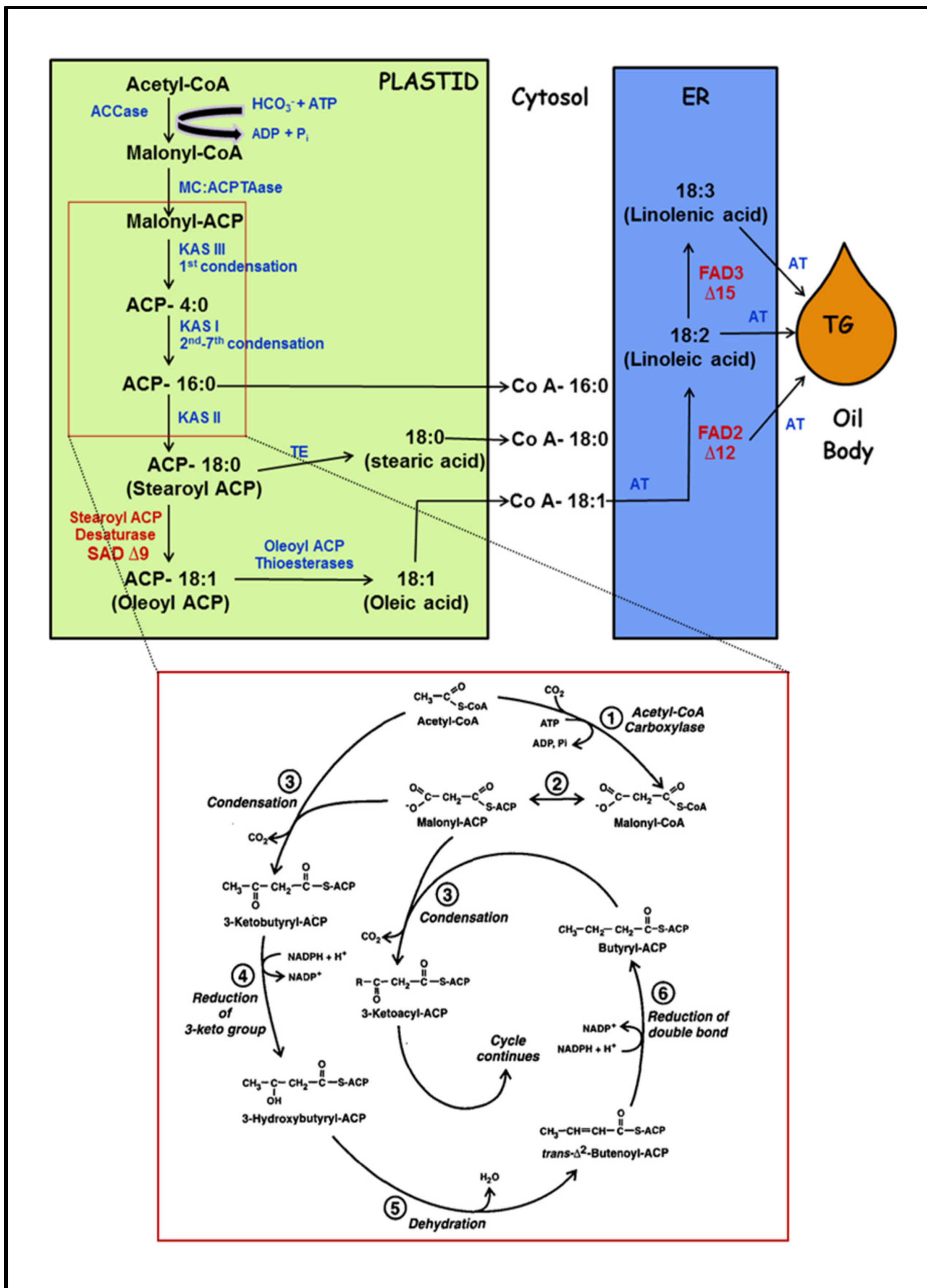


Figure 1.5: Fatty acid biosynthesis pathway in plants. After each condensation with 3-ketoacyl-ACP synthase (KAS), the 3-ketoacyl-ACP product is reduced (reaction 4), dehydrated (reaction 5), and reduced again (reaction 6), by 3-ketoacyl-ACP reductase, 3-hydroxyacyl-ACP dehydrase, and enoyl-ACP reductase, respectively (Adapted from Ohlrogge and Browse, 1995). (AT- acyletransferases, TE- acyl-ACP thioesterase and TG- Triglycerides)

The above mentioned FAs are saturated in nature, but it has been observed that over 75% of the FAs found in most plant tissues, are unsaturated (Ohlrogge and Browse, 1995). So, these saturated FAs are further converted to unsaturated FAs with the help of desaturase enzymes by addition of double bonds at specific locations in their fatty acyl chains. The first double bond is introduced in the plastid by the soluble enzyme stearoyl-ACP desaturase (SAD, Δ9) converting saturated stearic to monounsaturated oleic. Subsequent desaturation of the FAs to the highly unsaturated forms is carried out by membrane bound desaturases of the chloroplast and the endoplasmic reticulum (ER), where FAD2 or Δ12 desaturase is the next in the series of desaturases, which converts oleic acid to linoleic acid by introduction of second double bond. Similarly, the third double bond is added to linoleic acid by FAD3 or Δ15 desaturase to form α-linolenic acid (Somerville and Browse, 1991; Heinz, 1993).

1.4 Fatty acid desaturases: essential for PUFA synthesis

The desaturase class includes all the enzymes, which can activate oxygen and use it for subsequent modification of C–H bonds at saturated or monounsaturated carbons in substrates as diverse as alkyl groups, acyl residues in thio-, amide- or oxygen-ester linkage, carotenoids, sphingolipids, aldehydes and sterols (Shanklin and Cahoon, 1998; Tocher et al., 1998). Fatty acid desaturases are enzymes that convert a single bond between two carbon atoms to a double bond in a fatty acyl chain at specific positions (Los and Murata, 1998). The resultant double bond is often referred to as unsaturated bond, and the reactions catalyzed by these enzymes are known as desaturation reactions. Desaturases catalyze this reaction with strict regioselectivity and stereoselectivity. Based on the regioselectivity, there are two groups of desaturases, methyl-end desaturases, which introduce the double bond on a specific position counted from the methyl or ω-end of a FA; whereas, front-end desaturases position a double bond counted from the carboxyl end or Δ-end of FAs (Sperling et al., 2003). Examples of methyl end desaturases are ω-3/Δ15 and ω-6/Δ12 desaturases, while that of front end are Δ9, Δ6, Δ5 desaturases etc. (Napier et al., 1999).

These enzymes can be further classified into two phylogenetically unrelated groups; the membrane bound FA desaturases and the soluble desaturases (Shanklin and Cahoon, 1998; Sperling et al., 2003), although both are reported to be di-iron-oxo enzymes (Fox et al., 1993; Shanklin et al., 2009). Soluble desaturases are acyl-acyl carrier protein (ACP) desaturases, represented by the stearoyl ACP desaturase (SAD

or Δ9) (EC 1.14.99.6), which desaturates stearoyl-ACP to produce ACP-bound oleic acid (18:1). Membrane desaturases like FAD2 (Δ12) and FAD3 (Δ15) introduce double bonds into FAs that are either esterified as acyl-CoA or bound to the glycerol moiety of glycerolipids (Los and Murata, 1998). Soluble desaturases are known to have two conserved histidine motifs (Shanklin and Cahoon, 1998), while membrane desaturases contain three separate histidine-boxes and four trans-membrane domains (Los and Murata, 1998; Murphy, 1999). The distribution of FAs and FA desaturases is ubiquitous, observed in all aerobic organisms including algae, fungi, mosses, higher plants and mammals. They play a key role in the maintenance of the proper structure and functioning of biological membranes (Ohlrogge and Browse, 1995; McConn and Browse, 1998; Schmid and Ohlrogge, 2002).

Free FAs are never desaturated *in vivo*. FAs are either esterified to acyl carrier protein (ACP) for the action of soluble plastid desaturases, or to coenzyme A (CoA) or lipids for the integral membrane desaturases. Depending on whether the FA is bound to ACP, CoA or lipids, there are three types of desaturases that are responsible for desaturation in the FA biosynthesis pathway. First is acyl-lipid desaturase, which introduces unsaturated bonds into FAs bound in lipid form. Second, acyl-ACP desaturase introduces the first double bond into fatty acids that are bound to ACP and the third, acyl-CoA, introduces unsaturated bonds into fatty acids that are bound to CoA. Δ9, Δ12 and Δ15 desaturases are acyl-ACP or acyl-lipid desaturases and are distributed in cyanobacteria and plants (Murata and Wada, 1995).

Desaturases are well characterized enzymes containing Fe-binding active centers (Figure 1.6). Amino acid (AA) sequences of membrane bound FA desaturases from plants, animals and other organisms have three strongly conserved histidine rich sequences with the general motifs of HXXXH, HXXHH and H/QXXHH (Los and Murata, 1998) (Table 1.3). Desaturation reaction requires molecular oxygen, NAD(P)H, an electron transport system (ferredoxin-NADPH reductase and ferredoxin, or cytochrome b5 reductase and cytochrome b5) and a terminal desaturase.

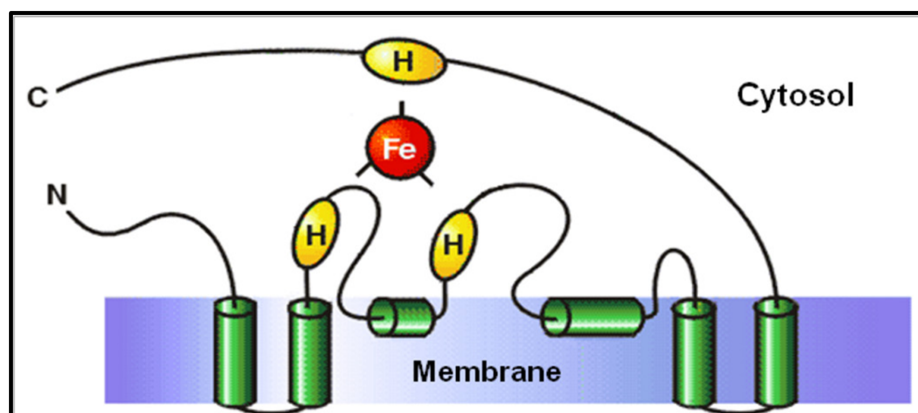


Figure 1.6: Cartoon depicting structure of membrane-bound desaturases. 'Fe' represents a putative di-iron active site and 'H' represents histidine-rich structural motifs involved in coordinating the di-iron active site. Adapted from Covello Lab, Plant Biotechnology Institute, Saskatoon, SK, Canada (http://cbr.pbi.nrc.ca/covellop/r-fattyacid.html)

Table 1.3: Fatty acid desaturase classes

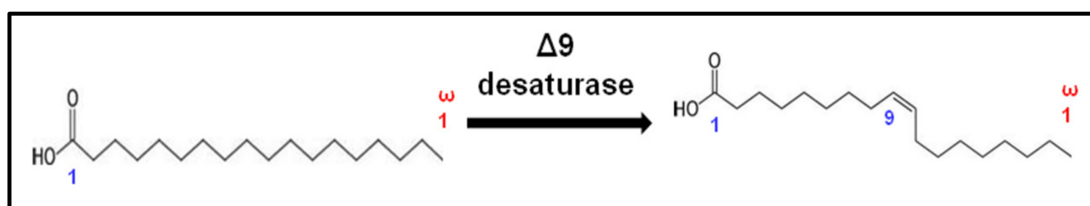
	Soluble desaturases	**Membrane bound desaturases**	
	Acyl ACP desaturases	**Acyl lipid desaturases**	**Acyl CoA desaturases**
Occurrence	Higher plant plastids	Cyanobacteria andhigher plants	Animals, yeasts and fungi
Substrate	Fatty acyl-CoAs	Fatty acids bound to the glycerol moiety of polar glycerolipids,	Fatty acids bound to ACP
Electron donor	Ferredoxin-NADPH reductase and ferredoxin	Ferredoxin (in cyanobacteria and plant plastid) or cytochrome *b*5 (in plant ER)	Cytochrome *b*5
Structure	Amino acid residues involved in binding the di-iron complex form two characteristic D/EXXH motifs	Histidine boxes $H(X)_{3-4}H$, $H(X)_{2-3}HH$ and $H/Q(X)_{2-3}HH$ involved in binding the di-iron complex	Histidine boxes $H(X)_{3-4}H$, $H(X)_{2-3}HH$ and $H/Q(X)_{2-3}HH$ involved in binding the di-iron complex

1.4.1 Key fatty acid desaturases responsible for accumulation of PUFA in plants

Various studies have shown that the three desaturases, SAD (Δ9), FAD2 (Δ12) and FAD3 (Δ15), drive the poly-unsaturated fatty acids (PUFA) synthesis pathway and they exist in multiple forms and locations within cells in different plant systems. In addition to such diversity, they show spatial and temporal specificity in their expression pattern. Some genes express specifically in seed while some in vegetative tissue. Further, few genes express at early developmental stages, few at mid or late stages while some of them have a constitutive expression pattern in the developing seeds.

1.4.1.1 Δ9 desaturase (SAD)

The Δ9 desaturase, also known as stearoyl-acyl-carrier-protein desaturase (SAD), is the first in the series of desaturases, catalyzing the desaturation of stearoyl-ACP (18:0-ACP) by introduction of the first double bond into the FA chain between C9 and C10 to form oleoyl-ACP (18:1-ACP). This marks the beginning of unsaturation of saturated FAs in PUFA biosynthesis. The product of its desaturation reaction, monounsaturated oleic acid, can alone serve as a substrate for further PUFAs, such as linoleic acid and linolenic acid. SAD plays a key role in determining the ratio of saturated to unsaturated FAs in plants.



It is a well characterized, soluble enzyme and is exclusively localized in the plant plastid. The SAD gene was first cloned from safflower (*Carthamus tinctorius*) embryos and it was proved that ferredoxin is required for enzyme activity (Thompson et al., 1991). This was followed by the cloning of SAD gene of castor (*Ricinus communis*) (Knutzon et al., 1991). Later, crystal structure and some characters, such as substrate specificity and di-iron-oxo centre of *Ricinus* SAD were further elucidated (Fox et al., 1993; Lindqvist et al., 1996). The SAD gene has also been isolated and characterized from several other plants like spinach (Nishida et al., 1992),

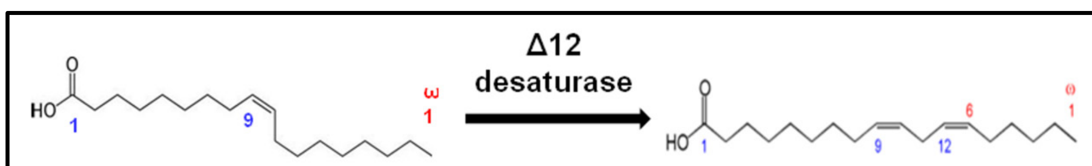
Arabidopsis (Fukuchi-Mizutani et al., 1998), Brassica (Slocombe et al., 1994), flax (Singh et al., 1994), sunflower (Serrano-Vega et al., 2003), soybean (Byfield et al., 2006), and *Jatropha curcas* (Tong et al., 2006).

Till now, nearly seventy *SAD* gene sequences or their segments have been deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/). The presence of multiple *SAD* genes was also reported in plants like Arabidopsis and soybean. In Arabidopsis, the two Δ9 desaturases were characterized as *ADS1* and *ADS2*. They showed variation in expression on change in temperature, where cold temperature up-regulated *ADS2* and down-regulated *ADS1* expression (Fukuchi-Mizutani et al., 1998). *SACPD-A* and *SACPD-B* were isolated from developing soybean seeds and quantifiable variation was observed in the expression levels of these SAD enzymes (Byfield et al., 2006). In flax, two *SAD* genes, *SAD1* and *SAD2,* were reported by Jain et al. (1999) and expression profile of *SAD1* gene isoform during seed development of flax variety AC McDuff was analyzed by Fofana et al. (2006).

Several studies were performed to analyze the relationship between the structure and function of SAD and the regulation of its gene activity (Knutzon et al., 1991; Shanklin and Somerville, 1991; Fox et al., 1993; Kodama et al., 1995; Murata and Wada, 1995; Lindqvist et al., 1996; Behrouzian et al., 2002). The emerging evidences suggest that the composition of fatty acids in plants can be modified by regulating *SAD* gene activity through genetic engineering (Knutzon et al., 1992; Ohlrogge, 1999; Whittle and Shanklin, 2001).

1.4.1.2 Δ12 desaturase (FAD2)

The Δ12 desaturase, also known as ω-6 desaturase and commonly abbreviated as FAD2 (Fatty acid desaturase 2), introduces the second double bond in the biosynthesis of 18:2 FAs, a process in which oleic acid is converted to linoleic acid by addition of a double bond at Δ12 or ω-6 (between C12 and C13) position of oleic acid. There are two distinct sites for Δ12 desaturation in the cell, one is plastidial and the other on the ER, both having distinct genes with specialized signals for their respective localization (Ohlrogge and Browse, 1995). The 'YNNKL' motif at the C terminus of the protein ensures ER localization of Δ12 desaturase (McCartney et al., 2004).



The *FAD2* gene encoding microsomal ω-6 desaturase was originally characterized in *Arabidopsis thaliana* (Okuley et al., 1994). Several *FAD2* genes were later cloned, characterized and their expression patterns were analyzed in various plants such as soybean (Heppard et al., 1996; Tang et al., 2005; Schlueter et al., 2007), Brassica (Scheffler et al., 1997), sunflower (Martinez-Rivas et al., 2001), sesame (Jin et al., 2001), cotton (Liu et al., 2001; Pirtle et al., 2001; Zhang et al., 2009), olive (Banilas et al., 2005; Hernandez et al., 2005; Hernandez et al., 2009) and *Camelina sativa* (Kang et al., 2011). A plastidial form of membrane bound ω-6 desaturase was isolated from spinach as well as from olives (Schmidt et al., 1994; Banilas et al., 2005). Soybean being a paleopolyploid, there exist multiple *FAD2* gene copies, which surprisingly show a lot of diversity and complexity in their genic and genomic organization as well as in their expression specificity. Initially, two copies, *FAD2-1* and *FAD2-2*, were cloned from soybean (Heppard et al., 1996), where it was observed that *FAD2-1* expressed primarily in developing seeds, while *FAD2-2* expressed in both, vegetative tissues and developing seeds. Later, EST based searches by Tang et al. (2005) identified an additional copy of *FAD2-1* in soybean (*FAD2-1A* and *FAD2-1B*) differing in just 24 amino acids residues. Their expression studies in yeast revealed that the *FAD2-1A* isoform is more unstable than *FAD2-1B*, particularly when cultures were maintained at elevated growth temperatures. Further analysis of the EST libraries from which the *FAD2* genes were identified revealed that all *FAD2-1* ESTs were from seed-related libraries while *FAD2-2* ESTs came from a variety of tissues (Tang et al., 2005). *Brassica napus* also has 4-6 gene copies of Δ12 desaturase. It was further suggested that the desaturase multi gene family of Brassica arose as the result of duplication of chromosome segments rather than duplication of individual genes (Scheffler et al., 1997).

Martinez-Rivas et al. (2001) isolated and characterized three microsomal oleate desaturases (*FAD2*), *Ha89FAD2-1*, *Ha89FAD2-1* and *Ha89FAD2-3* from sunflower. When the expression of these genes was analyzed in normal and high oleic sunflower, it was observed that *FAD2-2* and *FAD2-3* genes were weakly expressed in

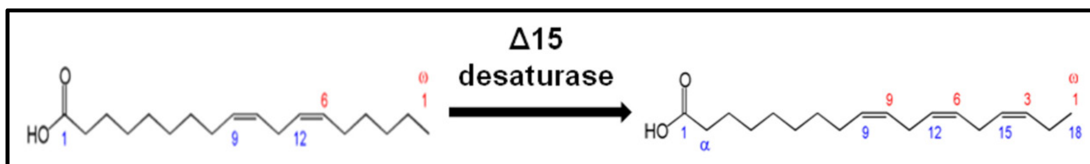
all tissues studied from both the varieties. On the other hand, *FAD2-1* gene expressed strongly and exclusively in developing embryos of normal-type sunflower, while its expression in high-oleic developing embryos was drastically reduced. One or two copies of the seed specific *SeFAD2* gene encoding microsomal Δ12 desaturase were reported in sesame. According to the development of sesame oilseeds, changes in linoleic acid levels occurred coincidently with changes in the *SeFAD2* transcript levels (Jin et al., 2001).

In cotton, four different isoforms of *FAD2* gene were identified (Liu et al., 2001; Pirtle et al., 2001; Zhang et al., 2009). Two microsomal oleate desaturases, *OepFAD2-1* and *OepFAD2-2* were also isolated from olive (*Olea europaea*) (Hernandez et al., 2005). The *OepFAD2-1* transcripts were strongly detected in very young seeds and in leaves, showing low levels in mesocarps, while the *OepFAD2-2* gene was moderately expressed in developing seeds, ripening mesocarp and leaves. These expression data suggested differential functions for the two olive microsomal oleate desaturase genes, with *FAD2-1* possibly responsible for the desaturation of reserve lipids in the young seed, while *FAD2-2* might be mainly involved in storage lipid desaturation in the mature seeds and the mesocarp (Hernandez et al., 2005; Hernandez et al., 2009). Three FAD2 genes were identified in *Camelina sativa* and designated *CsFAD2-1* to *-3*. Although these three CsFAD2 genes share very high sequence similarity, they showed different expression patterns. *CsFAD2-1* gene expressed in the vegetative, as well as developing seeds; whereas, transcripts of *CsFAD2-2* and *CsFAD2-3* were mainly detected in developing seeds, suggesting their major role in storage oil desaturation in seed. In flax, Fofana et al. (2004) identified two copies of *FAD2* gene by southern hybridization, which expressed in developing seeds with closely matching sequences. They later studied the expression profiles of *FAD2* genes during seed development in Canadian flax variety AC McDuff and observed a temporal expression pattern (Fofana et al., 2006). Later, Krasowska et al. (2007) cloned and characterized a full length *FAD2* gene from flax. Khadake et al. (2009) cloned a novel Δ12 desaturase gene (*LuFAD2-2*) from flax. These two genes showed 85% similarity at the amino acid (AA) level.

1.4.1.3 Δ15 desaturase (FAD3)

The Δ15 desaturase is also known as ω-3 desaturase and commonly abbreviated as FAD3 (fatty acid desaturase 3). It converts linoleic acid to α-linolenic acid (18:3) by

introduction of a third double bond at Δ15 or ω-3 position (between C15 and C16) of linoleic acid. Multiple forms of Δ15 desaturase gene have been identified in plants, of which microsomal and plastidial Δ15 desaturase are the primary forms.



In Arabidopsis, one microsomal desaturase, encoded by the *Fad3* gene and two distinct plastidial desaturases, encoded by *Fad7* and *Fad8* genes have been described (Arondel et al., 1992; Yadav et al., 1993; Gibson et al., 1994; Nishiuchi et al., 1994). The deduced amino acid sequence of the *Fad3* gene showed 70% identity to that of the *Fad7* gene (Iba et al., 1993). The high degree of homology between the *Fad3* and the *Fad7* gene products suggested that these genes were derived from a common ancestral gene. Among these desaturases, the microsomal enzymes were shown to be the major contributors to seed ALA content. Further, it was shown that *Fad8* had a predominant role in lipid desaturation at low temperatures (McConn et al., 1994), whereas *Fad7* appeared to be more important in regulating the levels of trienoic FAs during leaf maturation (Horiguchi et al., 1996). In further experiments, it was also observed that the triple mutant plants at *fad3-fad7-fad8* loci contained almost negligible (0.1%) trienoic FA (McConn and Browse, 1996).

Besides Arabidopsis, the *FAD3* gene has been isolated from several plants such as Brassica (Arondel et al., 1992), *Nicotiana tabacum* (Hamada et al., 1994), castor (van de Loo and Somerville, 1994), *Limanthes douglasii* (Bhella and Mackenzie, 1995), *Triticum aestivum* (Horiguchi et al., 1998), maize (Berberich et al., 1998), *Perilla frutescens* (Chung et al., 1999), *Oryza sativa* (Wang et al., 2006) and *Descurainia sophia* (Tang et al., 2007). Soybean contained three microsomal FAD3 homologs, *GmFAD3A*, *GmFAD3B* and *GmFAD3C,* that contributed to seed linolenic acid levels (Bilyeu et al., 2003). *GmFAD3A* was significantly upregulated in developing seeds, while *GmFAD3B* and *GmFAD3C* levels remained relatively low, indicating predominance of *GmFAD3A* for ALA accumulation in developing seeds. In flax, three microsomal Δ15 desaturase *viz*., *FAD3A, FAD3B* (Vrinten et al., 2005) and *FAD3C* genes have been reported (Banik et al., 2011). Transcript profiling of these

three genes during seed development was performed in four Canadian flax varieties varying in their ALA content (Banik et al., 2011). Based on the correlation of expression levels of these genes with linolenic acid accumulation during seed development, they showed that the *FAD3A* and *FAD3B* genes were responsible for linolenic acid accumulation in flax seeds, whereas the *FAD3C* did not play a major role in it. Table 1.4 enlists the desaturase genes isolated from various plant systems.

1.4.2 Role of desaturases in plants

Desaturases play a vital role in deciding the membrane and storage lipid FA composition. Polyunsaturated lipids, which are products of desaturation, constitute approximately 50% of the hydrophobic membrane barriers that outline the cell compartments. The composition of these lipids is critically important for many membrane functions, determined by its structure and fluidity and, thus, for proper growth and development of all living organisms (Wallis and Browse, 2002; Mikami and Murata, 2003). For example, membrane fluidity was shown to affect bilayer permeability (Schuler et al., 1991), ATPase activity (Cooke and Burden, 1990), and carrier-mediated transport (Deuticke and Haest, 1987).

Further, changes in the degree of FA desaturation are observed in plant responses to various abiotic stresses, such as cold, heat, salt and drought. Various studies have shown an increase in production of trienoic acids as a response to cold acclimation. Expression of *FAD8* gene in *Arabidopsis thaliana* was strongly inducible by low temperature (Gibson et al., 1994). In tomato leaves, the expression of *LeFAD7* was induced by chilling stress (4°C) and inhibited by high temperature (45°C) (Liu et al., 2006). FA desaturation during chilling acclimation is one of the factors involved in conferring low temperature tolerance to young leaves in tobacco as well (Kodama et al., 1995). In transplastomic tobacco, the expression of a FA desaturase gene from either wild potato *Solanum commersonll* or the cyanobacterium *Anacystis nidulans* altered the FA profiles, and improved the cold tolerance (Craig et al., 2008). Transgenic tobacco plants over-expressing *FAD7* also showed enhanced cold tolerance (Kodama et al., 1994; Khodakovskaya et al., 2006), whereas those with silenced *FAD7* gene contained a lower level of trienoic FAs than wild-type plants, and were more tolerant to high temperature (Murakami et al., 2000).

Table 1.4: List of plants from which Δ 9, Δ12 and Δ 15 desaturase genes are isolated

Plant	Δ 9 desaturase	Δ 12 desaturase	Δ 15 desaturase	References
Arabidopsis	ADS1, ADS2	FAD2, FAD6	FAD3, FAD7, FAD8	Iba et. al., 1993; Yadav et. al., 1993; Fukuchi-Mizutani, 1998; Watahiki and Yamamoto, 1994; Okuley et. al., 1994; Schmidt et.al., 1994; Falcone et al., 1994
Arachis hypogaea		Δ^{12}, *AhFAD2A, AhFAD2B*		Jung et al., 2000; Lopez et al., 2000
Brassica	SAD	FAD2	FAD3	Arondel et. al., 1992; Slocombe et al., 1994; Scheffler, 1997
Castor	-	FAD2	FAD3	Fox et al., 1993;van de Loo and Somerville, 1994
Cotton	-	FAD2, FAD2-3, FAD2-4		Liu, 2001; Pirtle, 2001; Zhang, 2009
Descurainia sophi	-	-	FAD3, FAD7, FAD8	Tang et. al., 2007
Flax	SAD	linFAD1, linFAD2, FAD2-1;FAD2-2	FAD3A, FAD3B, FAD3C	Singh, 1994; Fofana et al., 2004; Vrinten et. al., 2005; Krasowska et al., 2007; Khadake et al., 2009; Banik et al., 2011
Jatrophacurcas	SAD	-	FAD3	Tong, 2006
Limanthesdouglasii	-	-	FAD3	Bhella and MacKenzie, 1995
Maize	-	-	FAD3	Berberich et al., 1998
Nicotiana tabacum	-	-	FAD3	Hamada et. al., 1996
Oryza sativa	-	-	FAD3	Wang et al., 2006
Perillafrutescens	-	-	FAD3	Chung et.al., 1999
Safflower	SAD	-	-	Thompson, 1991
Sesame		FAD2		Jin, 2001
Soybean	SAD	FAD2-1,FAD2-1A, FAD2-1B,FAD2-2, FAD2-2A,FAD2-2B, FAD2-3, FAD6	FAD3A,FAD3B, FAD3C, FAD3-2b	Heppard, 1996; Bilyeu et. al., 2003; Tang, 2005; Byfield, 2006; Schlueter, 2007; Li et al., 2007
Sunflower	SAD	FAD2-1,FAD2-2, FAD2-3,	-	Hongtrakul et al., 1998; Martínez-Rivas et. al., 2001; Serrano-Vega, 2003
Triticum aestivum	-	-	FAD3	Horiguchi et al. 1998

A temperature dependent chloroplast ω-3 desaturase was cloned from rice, which was highly expressed at 15°C than at normal temperature (Wang et al., 2006). These results give an indication that there is a direct relationship between the membrane FA unsaturation and the surrounding temperature. Further, it was shown that the antisense expression of the Arabidopsis *FAD7* reduced salt/drought tolerance in transgenic tobacco plants (Im et al., 2002), whereas over expression of either *FAD3* or *FAD8* increased tolerance to drought in tobacco plants, and to osmotic stress in cultured cells (Zhang et al., 2005). In Arabidopsis, Zhang et al. (2009; 2012) provided the evidence of FAD2, an endoplasmic reticulum localized ω-6 desaturase, being required for salt tolerance. Their results suggested that FAD2 mediated high-level vacuolar and plasma membrane FA desaturation is essential for the proper function of membrane attached Na^+/H^+ exchangers, and thereby to maintain a low cytosolic Na^+ concentration for salt tolerance during seed germination and early seedling growth in Arabidopsis. FAD2 is the main enzyme responsible for polyunsaturated lipid synthesis in developing seeds of oil crops. The *FAD2* mutants of Arabidopsis were deficient in activity of the ER oleate desaturase and when grown at low temperature, the seed development of FAD2 mutant was observed to be impaired (Miquel and Browse, 1994).

The membrane lipids of higher plants are characterized by a high proportion of trienoic FAs (16:3+18:3) (Somerville and Browse, 1991), especially linolenic acid (18:3), which play an important role as a precursor to a defense-related signal molecule, jasmonate (Nishiuchi and Iba, 1998). In plants like tobacco (Nishiuchi and Iba, 1998), Arabidopsis (Nishiuchi et al., 1994), potato (Todoroki et al., 1998), etc., higher transcript levels of *FAD7* genes on wound induction in different plant parts, accompanied by an increase in trienoic fatty acids as major polar lipids have been reported. This gene serves to supply 18:3 as a precursor to stress-related signaling molecules, such as jasmonic acid (JA) (Farmer, 1994). The high level of JA induces transcriptional activation of numerous defensive genes (Creelman et al., 1992; Pena-Cortes et al., 1993). Similarly, the plasma membrane also provides a surface on which some plant resistance proteins perceive pathogen-derived effectors and thus confer race specific resistance (Shah, 2005). Kacharoo et al. (2001) reported the role of fatty acid desaturase in modulating the activation of defense signaling pathways in plants. In their study, a 10-fold reduction in the activity of mutant *SAD* gene was observed in

Arabidopsis, resulting in elevation of the 18:0 FA content in the these plants. The reduced *SAD* activity led to the induction of certain defense responses from the salicylic acid pathway while inhibition of pathogenesis response related with activation of jasmonic acid pathway. On this basis, it was proposed that a FA-derived signal modulates crosstalk between different defense signaling pathways.

In plants, apart from having role in biotic and abiotic stress management mechanism, desatrurases also affect other cellular and physiological processes such as modulation of ion channels (Hourton-Cabassa et al., 2009), pollen formation (McConn and Browse, 1996) and chloroplast development (Wallis and Browse, 2002).

Section 2: Flax: the richest agricultural source of ALA

2.1 Economic importance

As justified by the species name given by Linnaeus, "*usitatissimum*" meaning "most useful", almost the whole flax plant is used and has a great importance in human life. Flax has two major uses, oil from seed (linseed oil) and fibers from stem. It is a source of many products in high value textile markets, in the composites and paper/pulp industry, as well as in industrial and nutritional oil sectors (Hamilton, 1986; Sharma,1992). Flax seed represents only 1% of the world supply of oilseeds, while flax fiber represents about 3% of the world natural fiber production. However, flax products are recently being considered to have higher potential for increased industrial use, as well as for human food and feed markets (Cullis, 2007).

The outer portion of theflax stem forms the long, strong bast fibers, used for weaving linen fabric, which is valued for its distinctive appearance as well as comfort. Linen is further used in designing apparel and interior textiles (upholstery, carpets etc.). Besides pure flax products, these fibers are blended with other natural fibers like cotton, silk, wool etc., and periodically appear as part of the fashion trends or for niche markets (Borland, 2003). In addition to linen, the shorter flax fibers, called tow, have other industrial applications, including nonwoven fabrics, composites, and specialty papers used for printing banknotes and rolling paper for cigarettes and tea bags (van Dam et al., 1994; Berglund, 2002). Flax fiber composites have a particularly huge market potential in automotive components (Lepsch and Horal,

1998; Domier and Kerr, 2000). Composite stiffness values for flax are comparable to glass fiber composites and therefore are suitable for low cost engineering uses (Heijenrath and Peijs, 1996; Gassan and Bledzki, 1997). Further, shives, which are the woody core of the flax stem after separation of fibers, are used in making strong, lightweight particleboard panels and as reinforcing materials in plastics (Domier and Kerr, 2000; Papadopoulos and Hague, 2003).

Flax seed is the source of linseed oil, which has both industrial as well as edible/nutritional use. The presence of high levels of α-linolenic acid (ALA), an ω-3 FA, makes linseed oil industrially (drying property) and nutritionally (essential FA) very important. For a long time, linseed oil has been extensively used industrially as a drying oil in paints and varnishes and also for making linoleum flooring, oil cloth, certain inks, plastics, and resins etc. Top quality seed and cold-pressed oil goes in human use market as human food or animal feed ingredient. Ground flax blends well with many other food ingredients and hence can be added to almost any baked product such as bread, waffles, pancakes etc., giving it a nutty flavor. Some other food uses of flax seed are: ready-to-eat breakfast cereals, breakfast drinks, salad dressings made with cold-pressed flax seed oil, salad toppings, biscuits, meat extenders, crackers, soups, bagels, fiber bars etc. All these food items are categorized as “functional foods” and consumed as a part of usual diet. Further, both flax seed oil and ground flax are also used as nutraceuticals in the form of soft gel capsules or as microencapsulated powder, respectively. Thus, in the health and nutrition industry, flax forms a subcategory as “specialty supplement” of unsaturated fatty acids (Fitzpatrick, 2006).

Flax with its unique seed FA content as well as the presence of other seed constituents, contributes to good animal health. Linseed meal is a byproduct of flax seed after it is crushed for linseed oil and is increasingly used as an ingredient in animal and poultry feeds for better animal health and for enriched food products obtained from them. Linseed meal has a unique combination of amino acids in the protein, which produces a glossy, healthy coat for animals. Further, flax seed has also been incorporated into rations of laying hen and in broiler diets, so as to increase the ω-3 FAs in the egg yolk and in poultry meat, as the egg yolk and meat lipids can be influenced through dietary fats. Laying hens consuming 10 to 20% flax in their rations produce eggs having a desirable, balanced FA composition. These “Omega eggs” are

being produced in many countries including India and contain increased amounts of ω-3 FAs (300 mg/egg) and decreased amounts of saturated FAs compared to normal eggs. The Omega eggs are even consistently lower in cholesterol content, from 210 mg/egg (Standard USDA egg level) to 180 mg/egg (Scheideler and Lewis, 1997). Adding flax to cattle and goat feed can also increase the content of healthy fats in meat and dairy products. These products provide consumers with foods that can have a healthier fat profile. Such food choices appeal to all health-conscious consumers and now a days, are high in demand.

2.2 Flax as human food

Flax seed and flax seed oil have been used for food and herbal products for centuries in Asia, Europe, and Africa. In recent times, it is gaining renewed attention because of its health and nutritional benefits. Flax seed has three major components making it beneficial in human and animal nutrition: (1) a very high content of ALA (45-60%), an essential FA for humans; (2) the highest content of plant lignans of all plant or seed products used for human food; and (3) a high percentage of dietary fiber, both soluble and insoluble. A typical flax seed composition consists of approximately 40% fat, 28% dietary fiber, 21% protein, 4% ash, and 1% carbohydrates (Anonymous, 2001) (Table1.5). It also has one of the most nutritious plant protein compositions and is very similar to that of soybean protein. The nutrient composition of flax also includes a number of important essential minerals and minor amounts of water and fat-soluble vitamins. One tablespoon of milled flax (also known as ground flax or flax meal) contains the same amount of magnesium as a banana (34 mg) and 66 mg potassium. Vitamin E, primarily present as gamma-tocopherol, is the main vitamin in flax and it functions as an antioxidant (Daun and Przybylski, 2000).

Flax oil, compared to many other sources of dietary fats, has a very healthy FA profile, with low levels (approximately 9%) of saturated fat, moderate levels (18%) of monounsaturated fat and high concentrations (73%) of PUFAs (Figure 1.7). The PUFA content comprises about 16% ω-6 fatty acids, primarily as linoleic acid (LA), and 57% α-linolenic acid (ALA), an ω-3 fatty acid (Bhatty and Cherdkiatgumchai, 1990; Daun and DeClercq, 1994; POS, 1994).

Table 1.5: Proximate composition of flax based on common measures (Adapted from Flax: A health and nutrition primer Morris 2007). (CHO- Carbohydrate)

Form of flax	Weight (g)	Common measure	Energy (kcal)	Total fat(g)	ALA (g)	Protein (g)	Total CHO (g)	Total dietary fiber (g)
Proximate analysis	100	-	450	41.0	23.0	20.0	29.0	28.0
Whole Seed	180	1 cup	810	74.0	41.0	36.0	52.0	50.0
	11	1 tbsp	50	4.5	2.5	2.2	3.0	3.0
	4	1 tsp	18	1.6	0.9	0.8	1.2	1.1
Ground Seed	130	1 cup	585	53.0	30.0	26.0	38.0	36.0
	8	1 tbsp	36	3.3	1.8	1.6	2.3	2.2
	2.7	1 tsp	12	1.1	0.6	0.5	0.8	0.8
Flax oil	100	-	884	100.0	57.0	-	-	-
	14	1 tbsp	124	14.0	8.0	-	-	-
	5	1 tsp	44	5.0	2.8	-	-	-

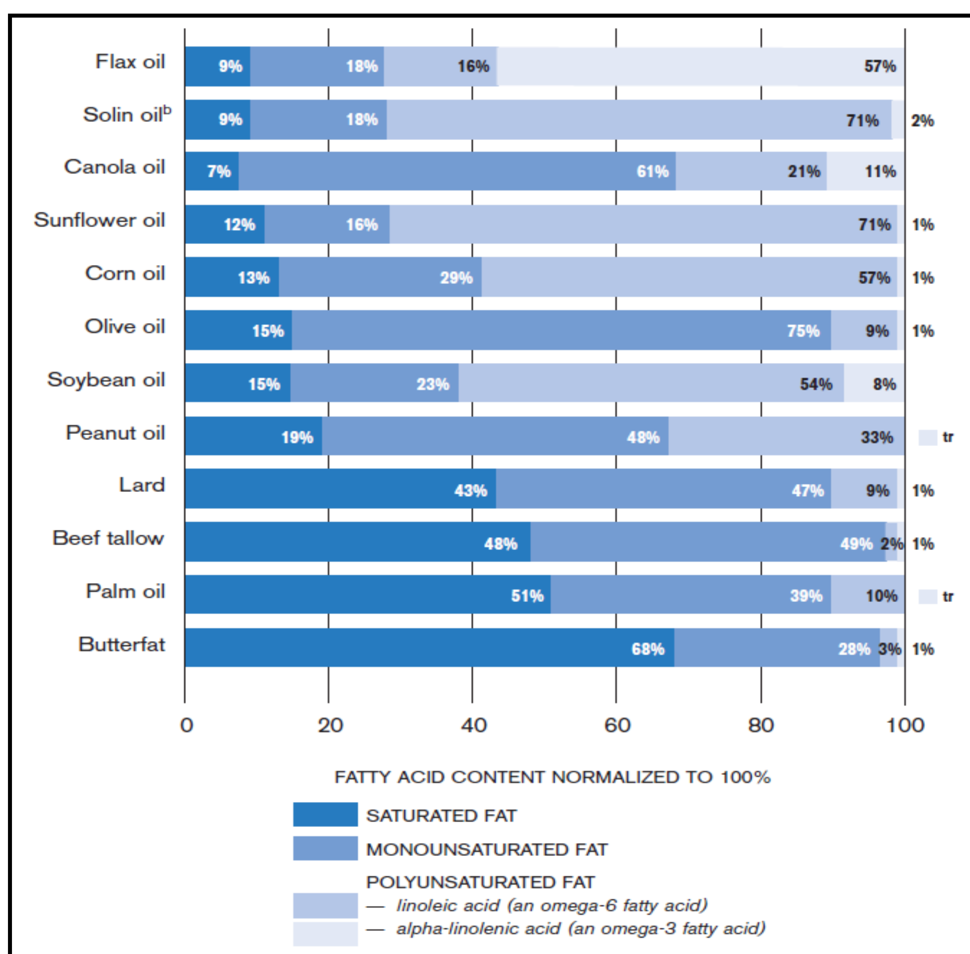


Figure 1.7: Comparison of saturated and unsaturated fatty acids in dietary fats and oils (Adapted from Flax: A health and nutrition primer, Morris 2007)

As described earlier, LA and ALA are essential fatty acids (EFA), since the human body cannot synthesize them and hence must be obtained from external food sources or diet. AA and EPA, which are the products of further conversion of LA and ALA in the human body, are metabolized to produce hormone like substances known as eicosanoids that effect physiological functions such as cell growth and division, inflammatory responses, muscle activity, blood pressure, and immune function. Since LA and ALA compete with one another for the enzymes responsible for their conversion to AA and EPA, respectively, it is important to have a proper balance of ω-6 and ω-3 fatty acids in the diet. Flax has a lower ratio of ω-6/ω-3 and hence helps to improve this FA balance in our body.

Various biologic effects of ALA together contribute to its positive health effects. It is important in the growth and development of infants (especially the brain and retina), as being the precursor of DHA, which has a functional role in the membrane intensive organs like brain. Inflammation is a characteristic of many chronic diseases, including atherosclerosis or "hardening of the arteries", the underlying condition that contributes to heart attacks and strokes, rheumatoid arthritis, Lupus nephritis etc. (Morris, 2007; Zhao et al., 2007). ALA reduces such inflammatory reactions by blocking the formation of compounds that promote inflammation (anti-inflammatory eicosanoid production) and hence has an important role in prevention of these diseases. It also helps in lowering high blood pressure and cholesterol (Pellizzon et al., 2007; Pan et al., 2009) and thus used in treating diseases of the heart and blood vessels. Further, it also has a role in prevention of certain types of cancer (Narisawa et al., 1994; Williams et al., 2007), neurodegenerative diseases (Joshi et al., 2006), and also in lowering blood glucose levels.

Flax is one of the richest sources of plant lignans, which is basically a phytoestrogen with a strong antioxidant property (Thompson et al., 1997). It is particularly rich in the lignan secoisolariciresinol diglycoside (SDG), and also contains small amounts of other lignans, such as, matairesinol, pinoresinol and isolariciresinol. Each gram of flax seeds contain 0.82-10.55 mg SDG, corresponding to nearly 0.7 to 1.9% of whole seed (Thompson, 2003). These lignans are further converted to the mammalian lignans, enterodiol and enterolactone, by bacteria in the gut of humans and other animals. These compounds have a protective role against certain cancers, particularly hormone-sensitive cancers such as those of the breast,

endometrium and prostate, by interfering with sex hormone metabolism. Presence of phytoestrogens in flax makes it useful in reducing menopause symptoms and is as effective as hormone replacement therapy in such group of women (Lemay et al., 2002). SDG is also a strong antioxidant and scavenges free radicals, which have implications in pathologies associated with many diseases like atherosclerosis, cancer and Alzheimer's disease etc. (Kitts et al., 1999).

Lastly, flaxseed with its high dietary fiber content (28%) provides additional health benefits on its consumption. About two-thirds of flax seed fiber is water insoluble and consists of non-starch polysaccharides such as cellulose, hemi cellulose and lignin. The insoluble dietary fiber fraction of flax plays an important role in improving laxation, hence gives relief in constipation, improves colon health, and may have protective effects against colon cancer. The soluble dietary fiber fraction of flax is found primarily as mucilage gums, which have been shown to play a role in lowering serum cholesterol levels, hence reduces coronary heart diseases and incidence of obesity. It also appears to have a reducing effect on blood glucose level and hence controls diabetes (Wolever and Jenkins, 1993).

Increasing interest in the above mentioned health benefits of flax has led to its use as a constituent of designer health foods and nutraceuticals. There has been an increase in demand of whole or milled flax and flax oil from the bakery and commercial food companies for fortification of their products.

2.3 Taxonomic status

Flax or linseed (*Linum usitatissimum* L.) is an annual plant belonging to the family Linaceae under the Order, Malpighiales and Division, Magnoliophyta, with a cosmopolitan distribution. Linaceae comprises 16 genera of which *Linum* is the largest genus with about 200 annual and perennial species (Plant List 2010 http://www.theplantlist.org). Cultivated flax belongs to the species *L. usitatissimum* – the most useful and most used species of the entire genus. The chromosome number of the *Linum* species shows a wide range, varying from 2n = 16 to 2n = 80 (Gill, 1987). However, *L. usitatissimum* and its wild relatives contain 2n = 30 chromosomes (Muravenko et al., 2003). The flax genome was estimated to contain a C-value of 0.7 pg, equivalent to ~675 Mb in size (http://data.kew.org/cvalues/) and to have unique characteristics (Cullis, 1973). However, a recent estimate of the size of the CDC

Bethune flax genome of 0.38 pg/C would translate into only 370 Mb (Ragupathy et al., 2011).

2.4 Origin and domestication

Flax is one of the eight founder crops (first group of grain crops that started agriculture) of the Near Eastern agriculture (Zohary and Hopf, 2000; Muir and Westcott, 2003). It was a principal source of oil and fiber from prehistoric times until the early twentieth century and was extensively cultivated in vast areas of Eurasia (Durrant, 1976). As revealed by various archaeological findings (Helbaek, 1959; van Zeist and Bakker-Heeres, 1975), flax was domesticated for both oil and fiber use more than 8,000 years ago in the Near East, but its domestication history remains uncertain (Allaby et al., 2005; Zohary et al., 2012). The center of origin of cultivated flax is believed to be the Near East, although secondary centers of diversity were identified in the Mediterranean basin, Ethiopia, Central Asia, and India (Vavilov, 1926; Zohary and Hopf, 2000).

Morphological (Diederichsen and Hammer, 1995), cytogenetic (Gill and Yermanos, 1967) and molecular (Fu et al., 2002; Allaby et al., 2005; Fu and Allaby, 2010) evidences suggest that the wild progenitor of cultivated flax is pale flax (*L. bienne* Mill. or previously *L. usitatissimum* L. subsp. *Angustifolium* (Huds.) Thell.; Hammer, 1986), with which, when crossed, it is fertile (Tammes, 1928). Pale flax has a very wide biogeographical range spanning Western Europe and the Mediterranean, North Africa, Western and Southern Asia, and the Caucasus regions. It was earlier thought that these diversity regions of flax may represent independent domestication events within the wide biogeographical range of pale flax (Harlan, 1971). However, recent molecular and phylogenetic evidences (Allaby et al., 2005; Fu, 2005; Fu and Allaby, 2010; Fu, 2011), suggest that domesticated/cultivated flax is of monophyletic origin and has probably evolved from a single domestication event of pale flax. Though, the geographic location of this event is still obscure. In addition, these data also indicate that the flax was first domesticated for its oil, rather than for fiber use (Allaby et al., 2005; Fu and Allaby, 2010; Fu, 2011). The pale flax had strong branches, small seeds and dehiscent (splitting) capsules. On the other hand, the main changes that were observed because of domestication were the shift to non-dehiscent (non-splitting) capsules, increase in seed size, selection for high oil yielding, or longer stem with high amount of long fiber, varieties.

2.5 Plant morphology

The flax plant is erect and grows to about three feet in height. It is a cool-season annual that has one main stem and a short-branched taproot. Two or more branches may develop from the base of the plant and produce a multi-branched arrangement of flowers. Leaves are glaucous green, slender, lanceolate, 20–40 mm long and 3 mm broad. The hermaphrodite flowers are 15–25 mm in diameter, with five petals and generally are pale blue in colour (Figure 1.8). Flax varieties can be distinguished by the colour of the flower that can range from dark to light blue, white or pale pink. The fruit is a round, dry capsule/pod/boll 5–9 mm in diameter, containing several glossy brown seeds 4–7 mm long. The seeds may vary in colour from reddish brown to golden yellow (Figure 1.9) (Cullis, 2007)

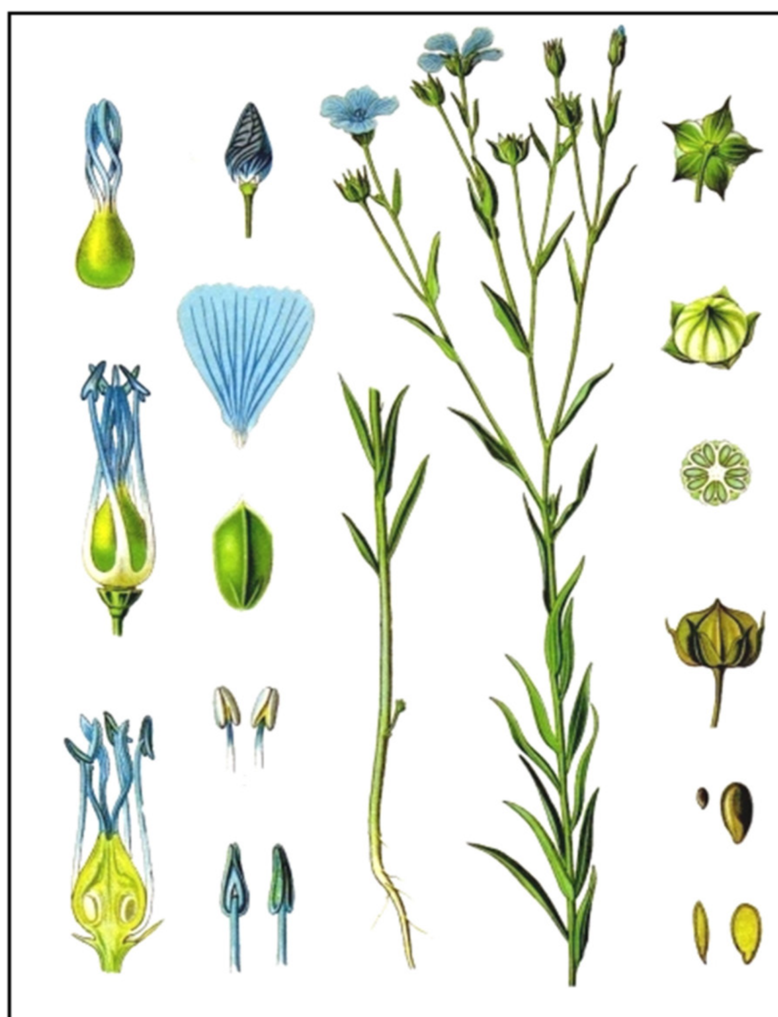


Figure 1.8: Flax plant morphology (Source: from "Koehler's Medicinal-Plants" 1887)

(A)

(B)

Figure 1.9: Variation of colour in **(A)** flowers and **(B)** seeds of flax varieties (Adapted from Flax genetic diversity as the raw material for future success, Diederichsen and Fu) (http://www.saskflax.com/documents/presentations/06A_Diederichsen.pdf)

2.6 Flax as a crop

Flax or linseed is purely a cool season crop. Temperate and cool climatic conditions are best suited for its growth and the temperature required for its cultivation ranges from minimum 10°C to maximum 38°C. Flax can be raised in almost all types of soils, where sufficient moisture is available, but does better on heavier soils having greater water-retention capacity. It is also tolerant to wide range of soil pH (5.0-7.0). However, it grows the best on well drained loam to clay loam soils rich in humus. Annual rainfall of 700 to 750 mm is optimum; while under dry climates, irrigation is essential. The crop takes about 120 to 140 days to mature depending on the cultivar where, the vegetative period is of 45 to 60 days, flowering period is 20 to 30 days and the maturation period consists of 30 to 40 days (Vittal et al., 2005).

Commercially grown flax crops are grouped into two main types: fiber flax (*L. usitatissimum* convar. *Elongatum* Vav. et Ell.) and seed flax (linseed) [convar. *mediterraneum* (Vav. et Ell.) Kulpa et Danert], alternatively referred to as long-stalked flax and crown flax, respectively. A third type, generally called dual-purpose flax is also grown for both seed and fiber (convar. *usitatissimum*). Fiber varieties usually have longer stem, 80–120 cm tall, with fewer branches, fewer seed capsules, and smaller seeds. Whereas, oil type has shorter and heavily branched stems, 60–80 cm tall, with a higher number of seed capsules and larger seeds (Figure 1.10).

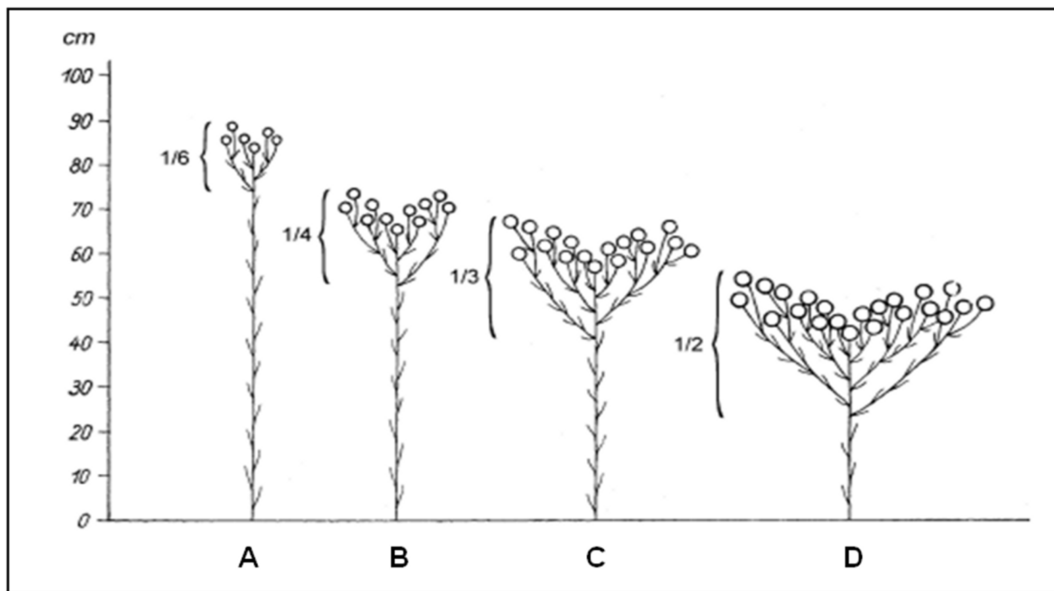


Figure 1.10: Stem length (defined by the length of the stem without side branches) variation observed in the flax varieties. It can be measured in cm or as a fraction of the stem with branching compared to the entire stem length (see figures beside branches). Basal branches are not considered. **(A)** and **(B)** typical fiber flax with less than 1/4 of the stem branched; **(C)** intermediate flax; **(D)** large-seeded flax (oilseed flax). (Adapted from Kulpa and Danert, 1962)

From the historical accounts it was observed that there was wider use of fiber flax landraces, with both fiber and dual uses relatively more important than at present (Zohary et al., 2012). However, fiber flax cultivation has been steadily declining in favor of cotton. The largest production of fiber flax presently occurs mainly in France, China, Russia, Belarus and Belgium, with 0.23 million ha of land all over world under fiber flax cultivation producing 0.6 million tonnes of fiber. Production wise, France ranks first, followed by China, with 0.37 tonnes and 0.12 tonnes of fiber in year 2010. On the other hand, oil seed flax or linseed currently occupies nearly 2.2 million ha of

land worldwide, with a production of 1.9 million tonnes and the main oil seed producing countries are India, Canada, China, Russia, Kazakhstan, Ethiopia, and United States. Production wise, Canada is the foremost country producing nearly 0.4 million tonnes of flax seeds in year 2010 (FAOSTAT, 2010) (Figure 1.11).

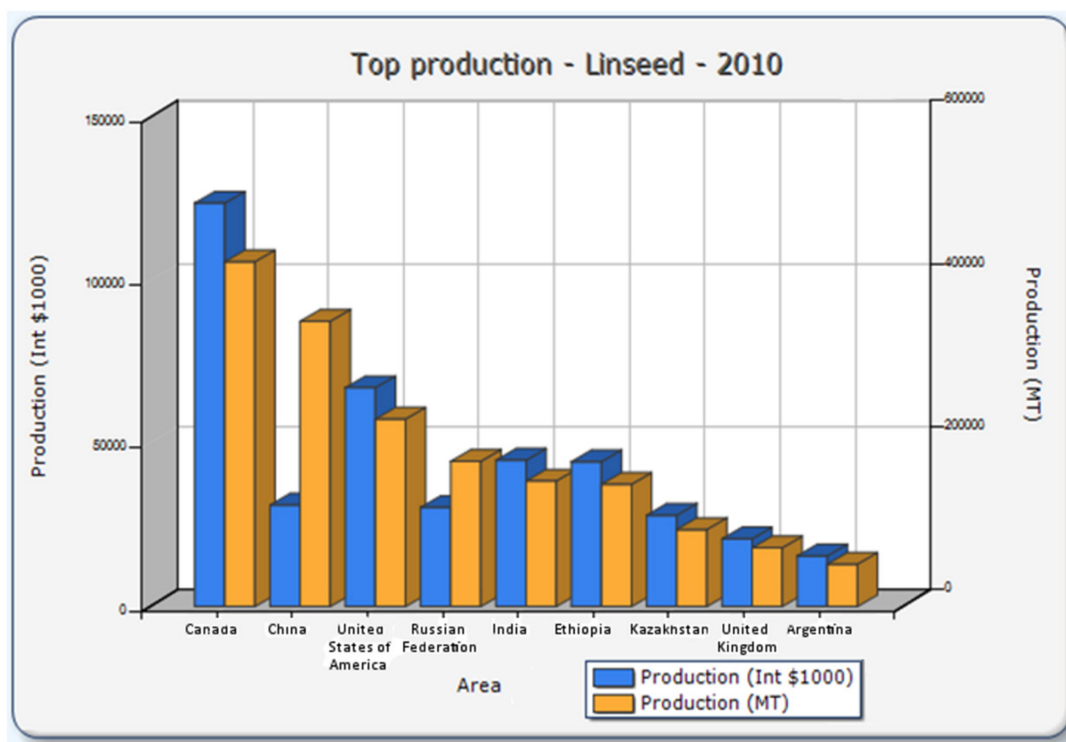


Figure 1.11: Production of oilseed flax or Linseed in top few countries in the world (FAOSTAT, 2010)

2.7 Germplasm diversity

Plant breeders are continuously challenged to develop flax cultivars adapted to changing market needs, environmental conditions and with resistance to new diseases. The availability of well characterized diverse germplasm is of the greatest importance for the success of this work. World genebanks maintain about 48,000 accessions of cultivated flax, of which around 10,000 might be unique ones (Diederichsen, 2007). Some of the well characterized flax germplasm collections are with N. I. Vavilov Institute, St Petersburg, Russia; All-Russian Flax Research Institute, Torzhok, Russia; IPK-Genebank, Gatersleben, Germany; AGRITEC, Šumperk, Czech Republic; NC7, North Central Regional Plant Introduction Station, USDA-ARS, NCRPIS, Ames, USA; CGN, Centre for Genetic Resources, The Netherlands Plant Research International, Wageningen etc. (Diederichsen, 2007).

The early and rapid distribution of flax cultivation in the Old World resulted in a wide range of flax landraces adapted to different uses and environmental conditions (Vavilov, 1926; Elladi, 1940; Helbaek, 1959). Diederichsen (2001) described the genetic diversity in the world collection of cultivated flax maintained by Plant Gene Resources of Canada (PGRC) using morphological and agrobotanical characters, where a total 2331 flax accessions were considered for the study. A wide range of diversity was observed in the qualitative and quantitative characters as well as for FA compositions of this world collection. The coefficient of variation (CV) for some quantitative characters like plant height, days while flowering, 1000 seed weight, stearic acid content etc., were quite high (≥20% CV). While certain qualitative characters like colour of various floral parts, capsule dehiscence, number of branches, seed colour etc., showed higher number of observed classes (≥6). Further, it was also observed that the plant height of flax from Eastern Asia was very high, due to the strong representation of fiber flax from China in the PGRC collection while, flax from India was generally short, but had large seeds and high oil content. Flax from the Mediterranean area also had large seeds.

Molecular markers are highly useful for identifying potentially novel genotypes among the various flax accessions, and to assess genetic diversity for both germplasm management and core collection assembly (Frankel and Brown, 1984). A variety of marker systems, including random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR) and inter-retrotransposon amplified polymorphism (IRAP), have been used to analyze flax germplasm all over world. The use of DNA based markers to study flax diversity was first reported by Oh et al. (2000), who compared RAPD and RFLP techniques and generated a genomic map based on these marker data.

Fu et al. (2002) reported low genetic variability in 61 flax varieties including Canadian cultivars and land races using RAPD markers. They also showed that more variation existed in landraces than cultivars and could clearly differentiate fiber and oil flax, suggesting their distinct genetic makeup. RAPD markers were even used to analyze the genetic variation, genetic erosion and relationship in 54 North American flax cultivars (Fu et al., 2003). It was observed that the North American linseed cultivars had more variation than those from the other countries. Further, Fu (2005)

studied the geographic pattern of flax variability in the world collection using RAPD markers. It revealed that the accessions from the East Asia and European regions were most diverse, but accessions from the Indian subcontinent and Africa were most distinct. Surprisingly, accessions from West Asia were genetically more related to those from the African region and less to those from the Indian subcontinent. These findings were significant for understanding flax domestication and also were useful in classifying intraspecific diversity of cultivated flax, establishing a core subset of the flax collection and exploring new sources of genes for flax improvement.

Our group has also developed a core collection for Indian flax/linseed germplasm (Kale et al. unpublished). It comprises 222 accessions out of nearly 3000 accessions present in the Indian germplasm, selected on the basis of morphological and agronomic traits. Further, the core collection was analyzed for genetic diversity and population structure using morphological and molecular markers (SSRs). On the basis of these analyses, the accessions were divided into two sub-populations and overall a low genetic diversity was observed among them.

Diederichsen and Fu (2006) studied phenotypic and RAPD variation within and among the four infraspecific groups of flax (dehiscent flax, fiber flax, large-seeded flax and intermediate flax) to understand phenotypic and genotypic differentiation within the cultivated genepool. Van Treuren et al. (2001) addressed the question of redundancy in the flax germplasm from CGN, The Netherlands, through AFLP analysis and revealed similar appearing genotypes with different accession numbers in the collection. The ISSR technique for flax fingerprinting was optimized by Wiesner and Wiesnerova (Wiesner and Wiesnerova, 2003, 2004a) using re-amplification method and through statistical correlation of free energy of dissociation of ISSR primers. They also showed statistical correlation between the value of thousand seed mass (TSM) and ISSR based clustering of 53 flax accessions from Czech Republic (Wiesnerova and Wiesner, 2004b). Uysal et al. (2010) used ISSRs to assess genetic diversity of 34 Pale flax (progenitor of the cultivated flax) accessions collected from Turkey and to compare genetic variation with 493 individual plants belonging to three major types of cultivated flax (dehiscent, winter, and landrace). It was observed that Pale flax displayed more ISSR variation than landraces and dehiscent type, but less than winter type, of cultivated flax. Further, it was also shown

that 493 individual plants largely clustered according to their plant types and that pale flax was genetically closer to the dehiscent type.

Simple sequence repeat (SSR) markers were used by Cloutier et al. (2009), Fu and Peterson (2010) and Fu (2011) to study the genetic relationships between cultivars, landraces and wild relatives as well as to study flax domestication process. Smýkal et al. (2011) developed a novel method for fingerprinting flax germplasm based on inter-retrotransposon amplified polymorphism (IRAP) and applied it to over 700 accessions of Czech National Flax Germplasm collection in order to assess genetic diversity and germplasm structure. Most of the studies show a lower genetic diversity in cultivated flax than in wild relatives or landraces. All these findings provide a basis for better flax germplasm management, core collection establishment and exploration of genetic diversity in breeding material.

2.8 Omics studies

To assist breeding efforts towards improvements of both seed and stem traits, breeders need a proper understanding of the complexity of the genetic mechanisms underlying traits such as oil content, FA composition, stem fiber content and fiber composition etc. Quantitative trait loci (QTL) and association mapping (AM) studies have the ability to provide some insights into the genetic mechanisms of complex traits and provide molecular markers to implement marker assisted breeding. In the last few years, there has been a surge in the area of flax genomics research and important genetic resources have been developed for this crop.

The first molecular map for flax was developed using RAPDs and RFLPs where, 19 RFLP and 69 RAPD markers covering 15 linkage groups (LGs) were identified (Cullis et al., 1995). Later, Oh et al. (2000) analyzed 20 RFLP and 520 RAPD markers of which 13 RFLP and 80 RAPD markers were located on the 15 LGs, in addition to one STS. Thus, the initial linkage map for flax was consisted of these 94 markers assigned to 15 LGs covering about 1,000 cM (Oh et al., 2000). During the similar period, an AFLP map was also generated for flax (Spielmeyer et al., 1998). This map was used to identify two QTLs on independent LGs with a major effect on resistance to Fusarium wilt. Two hundred thirteen marker loci covered 1,400 cM of the flax genome, with an average spacing of 10 cM and comprising 18 LGs. The map also incorporated RFLP markers tightly linked to flax rust (*Melamspora lini*)

resistance genes and markers detected by disease resistance gene like sequences. However, there were some limitations of these maps because of both, the type as well as limited number of markers present in them. Thus, there was need for a reliable, high density genetic map of flax that would serve as reference for a wide variety of applications such as QTL mapping, map based gene cloning, marker assisted crop improvement, linkage disequilibrium (LD) mapping, phylogenetic analysis etc.

In the next phase of resource development and to understand the molecular processes associated with the development of fiber and oil in a better manner, transcriptomic studies were initiated. Roach and Deyholos (2007, 2008) constructed the first cDNA custom flax array system and successfully used it to characterize gene expression profiles in fiber-bearing stem tissues. This platform was constructed by spotting 9,600 anonymous cDNA clones obtained from a flax stem-peel cDNA library. Further, a high-density microarray platform was developed by Fenart et al. (2010) using Nimblegen technology. Flax cDNAs generated from various tissues and/or developmental stages from two different flax cultivars (oil-seed and fiber flax respectively) were 454-sequenced and assembled into contigs. These contigs representing genes found in both oil-seed and fiber flax were used to produce a 48K array. On the same lines, Venglat et al. (2011) reported a total of 261,272 expressed sequence tags (EST) generated using 13 cDNA libraries representing genes active in developing embryos, seed coats and endosperm as well as genes expressed in flowers, etiolated seedlings, leaves and stem tissue. These EST libraries even included transcription factor genes that are typically expressed at low levels.

In recent years, a large number of SSR markers were developed from various previously constructed flax EST libraries (Cloutier et al., 2009; Soto-Cerda et al., 2011) and from SSR-enriched genomic libraries or other genomic sequences (Roose-Amsaleg et al., 2006; Bickel et al., 2011; Deng et al., 2011; Rachinskaya et al., 2011; Kale et al., 2012). Ragupathy et al. (2011) identified 4,064 putative SSRs from bacterial artificial chromosome (BAC) end sequences (BES). There are currently 1,326 SSR markers published in flax (Cloutier et al., 2012a). These markers were further used for constructing genetic maps and genetic diversity assessment (Fu and Peterson, 2010; Cloutier et al., 2011; Soto-Cerda et al., 2011). The first integrated consensus genetic and physical map of flax has been published by Cloutier et al. (2012b), where they have constructed three linkage maps using three different

populations containing about 385-469 mapped markers each. The consensus map of flax incorporated 770 markers based on 371 shared markers including 114 that were shared by all the three populations and 257 shared between any two populations. The map with 15 linkage groups corresponds to the haploid number of chromosomes of this species. The total length of the consensus genetic map is 1,551 cM with a mean marker density of 2.0 cM. A total of 670 markers were anchored to 204 of the 416 fingerprinted contigs of the physical map corresponding to ~274 Mb or 74 % of the estimated flax genome size of 370 Mb. Apart from identification of genomic and EST-SSR, various genes involved in the fatty acid and lignin biosynthetic pathways have been identified and characterized (Singh et al., 1994; Fofana et al., 2004; Vrinten et al., 2005; Krasowska et al., 2007; Khadake et al., 2009; Banik et al., 2011).

As a part of the Genome Canada TUFGEN flax genomics project, whole-genome shotgun sequencing (WGS) of nuclear genome of the CDC Bethune variety was conducted. Seven paired-end libraries ranging from 300 bp to 10 kb in size, were sequenced using an Illumina genome analyzer. A *de novo* assembly, comprised exclusively of deep-coverage (approximately 94× raw, approximately 69× filtered) short-sequence reads (44–100 bp), produced a set of scaffolds with N_{50} = 694 kb, including contigs with N_{50} = 20.1 kb. The contig assembly contained 302 Mb of non-redundant sequence representing 81% genome coverage. Up to 96% of published flax ESTs aligned to the whole-genome shotgun scaffolds (Wang et al., 2012).

The availability of whole flax genome in the public domain led to certain *in silico* studies, identifying novel miRNAs and their targets among the flax genes (Barozai, 2012; Barvkar et al., 2012; Moss and Cullis, 2012; Neutelings et al., 2012). The genome database has also been used for identifying UDP glycosyltransferase (UGT) genes from flax, which are involved in the glycosylation process of secondary metabolites in plants, using a UGT specific conserved signature motif. Total 137 genes were identified and were phylogenetically clustered into 14 major groups (Barvkar et al., 2012). Using publicly available expressed sequence tag (EST), microarray data and reverse transcription quantitative real time PCR (RT-qPCR), it was observed that 73% of these genes showed expression evidence in 15 tissues examined and indicated varied expression profiles.

Besides the above mentioned genomic and transcriptomic studies, certain proteomic studies in developing as well as mature seeds of flax have been performed.

Hradilová et al. (2010) reported a comparative analysis of proteomic changes in contrasting flax cultivars upon cadmium exposure. They observed significant changes in the expression of 14 proteins (related to disease/defense, metabolism, protein destination and storage, signal transduction, energy and cell structure) out of total proteins analyzed from Cd-treated cell suspension cultures derived from these contrasting cultivars with varying degree of cadmium tolerance, indicating the presence of natural variation in phytoremediation potential of flax. Flax proteome analysis was also used to investigate plant adaptation to an environment with a permanently increased level of radioactivity in the Chernobyl area (Klubicova et al., 2010). Systematic proteomic characterization of the mature and developing flax seed from the remediated Chernobyl area was done, yielding quantitative information for 318 and 379 proteins respectively. Further, these proteins were identified and categorized into various metabolic functional classes. However, due to the lack of genomic resources for flax at that time, proteins of unknown function comprised the largest group (Klubicova et al., 2011a; Klubicova et al., 2011b).

2.9 Indian flax agriculture and germplasm

In India, flax is a winter season (rabi) crop and is next in importance to rapeseed-mustard grown during this season. Thus, the main season for sowing linseed is October to November depending upon the availability of soil moisture. It is often grown on marginal and sub-marginal rainfed soils. The major linseed growing states are Madhya Pradesh, Maharashtra, Chhattisgarh, Uttar Pradesh, Jharkhand, Bihar, Rajasthan, Karnataka, Andhra Pradesh and Orissa (Vittal et al., 2005; Ahlawat, 2007) (Figure 1.12).

Presently India ranks first in the world, in terms of the area, with more than 0.45 million ha of land under flax cultivation. However, average yield of flax in the country is low with, 317.4 kg/ ha, leading to low production of 0.14 million tonnes, despite of highest land available. Thus, production wise, this positions India at the fifth place in the world (along with Ethiopia) with Canada leading the chart (FAOSTAT, 2010) (Figure 1.13).

In India, flax for seed is usually grown as a sole crop with adequate fertilization. It is normally rotated with hybrid maize, sorghum, pearl millet, groundnut and cowpea. Another traditional practice used in the country for flax

cultivation is *Utera* cropping system. This system has been in practice for efficient use of residual moisture in rice fields, where tillage is a problem. About 25% of the flax growing area is under *Utera* where, linseed is broadcast in the standing rice fields when the crop is between flowering and dough stages. It is allowed to complete its life cycle under the moisture stress, with inadequate nutrients and plant protection measures, resulting in poor yields. Alternatively, on rainfed lands, intercropping of linseed particularly with chickpea, lentil, durum wheat, barley and safflower is being popularized on a priority basis to increase the production. This provides stability to total production from a unit land and higher monetary returns to the farmers.

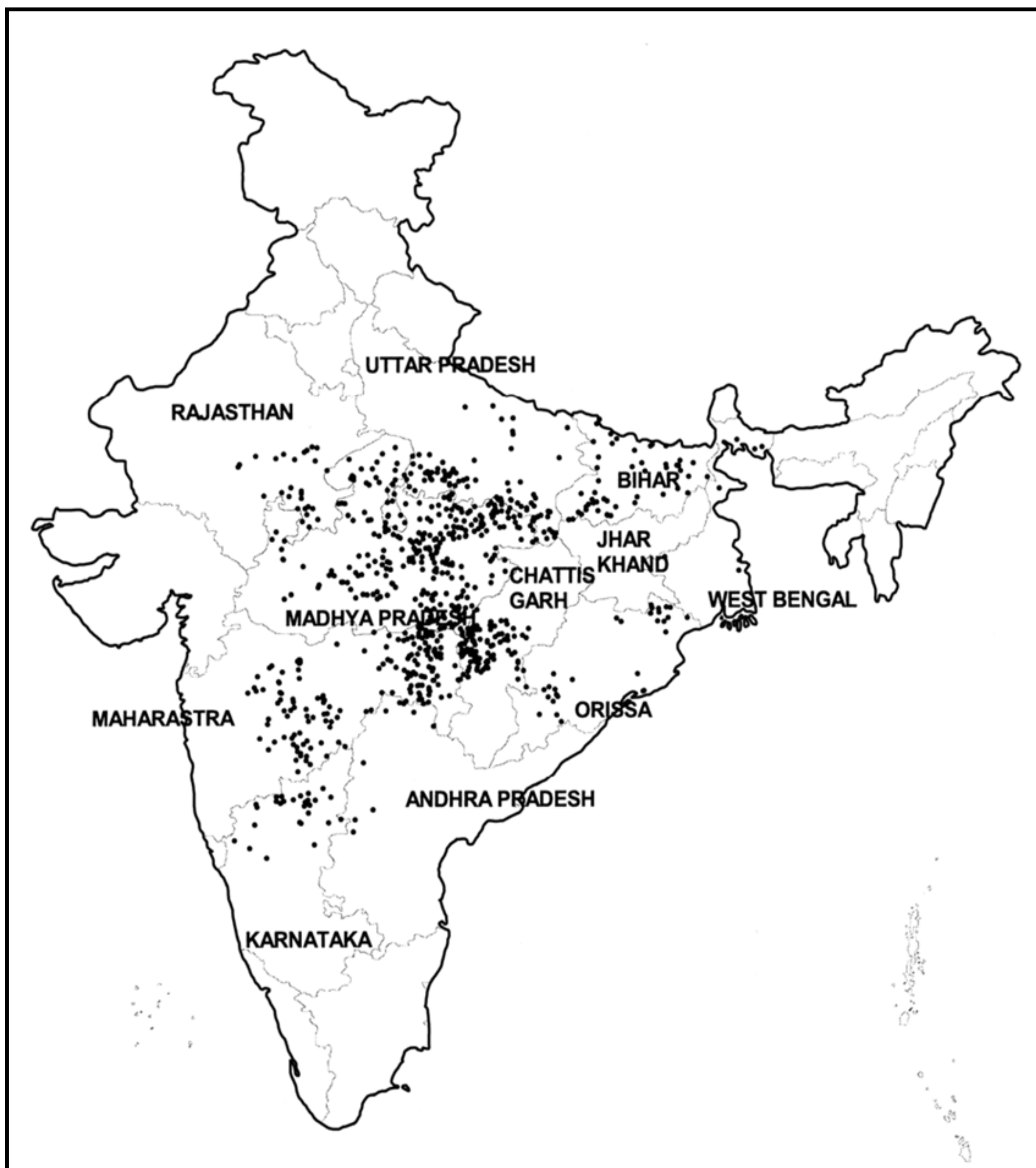


Figure 1.12: Major flax/Linseed producing states in India. Dots indicate various flax growing places in the state. (Adapted from Vittal et al. 2005)

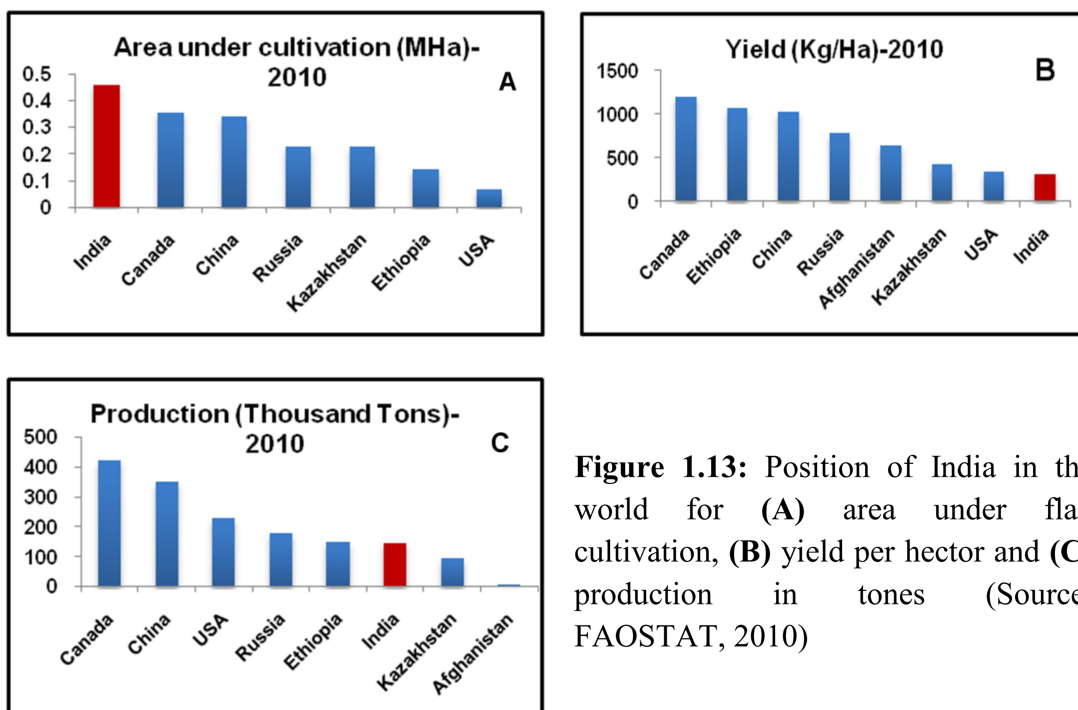


Figure 1.13: Position of India in the world for **(A)** area under flax cultivation, **(B)** yield per hector and **(C)** production in tones (Source: FAOSTAT, 2010)

Indian flax varieties are of two types: peninsular types, which have deep root system; and alluvial types, which are shallow-rooted and are characterized by profuse tillering. There are a number of varieties available for cultivation in different-cropping systems, situations and for varied agro-climatic conditions in linseed. During the last decade, a considerable number of double purpose varieties were developed which have good seed and fiber yielding potentials. Varieties differ in maturity, growth habit, and seed size and colour. Some of the popular flax varieties are Jawahar-17, Jawahar-7 (R-7), Mayurbhanj, LC 185, Hira, Mukta, Neelum, Shekhar, NL97, NL260, Sheetal, Padmini, Gaurav, Rashmi, Shubhra, Shweta, etc.

Disease and insect pests are the major constraints limiting flax productivity. In India, alternaria blight is the most serious disease followed by powdery mildew. Wilt, powdery mildew and linseed bud fly are predominant in Uttar Pradesh, Madhya Pradesh, Orissa, Maharashtra, Rajasthan and Karnataka. Further, in Chhattisgarh and adjoining areas of Maharashtra and Orissa, Cuscuta (a phenerogamic plant parasitic weed) causes heavy losses. Insect pests like termite, cutworm, wireworm, semilooper, leaf minor and gram pod borer also infest the crop and cause considerable damage. The reported losses in yield due to diseases and pests in severe conditions are 40-100% by rust, 60% and above by powdery mildew, 27- 60% by alternaria blight, upto 80% by wilt and 97% by linseed bud fly (Vittal et al., 2005). Most of the farmers still

practice the *Utera* system and grow low yielding disease susceptible varieties leading to low economic returns and hence the area under flax cultivation is slowly decreasing. This is even supported by the statistical data, which showed that the flax area decreased from 1.424 million ha in 1986 to 0.754 million ha in 2001 and further in 2010 it has gone down to 0.45 million ha (FAOSTAT, 2010). All these observations indicate that flax is largely a neglected crop in India. There is an urgent need to replace the old varieties from the farmers' field with new improved high yielding and disease resistant varieties so as to increase the flax productivity in the country.

Presently, the Indian flax germplasm consists of nearly 3000 accessions maintained at the Project Coordinating Unit (Linseed), C.S. Azad University of Agriculture and Technology Campus, Kanpur, UP. These are categorized in three groups as oil type, fiber type and dual purpose. Indian flax varieties are known for bold seeds and good oil content (Diederichsen, 2001; Diederichsen and Fu, 2006) and produce 30 to 40% oil of which 40 to 50% is ALA. Some varieties with 60-65% of ALA are also available; while on the other hand, varieties with less than 3% ALA were developed to produce oil with improved tolerance to rancidity (AICRP on linseed, Annual Report 2009-2010). So far, the Indian flax breeding program had objectives such as, development of varieties producing oil with improved shelf life (containing low ALA), dual purpose varieties that can yield both fiber and oil, and varieties resistant to pest and pathogens. However, with the knowledge about the importance of ω-3 FAs in human diet, the trend is changing and now there is an increased emphasis on development of high ALA varieties. Moreover, flax is the cheapest agricultural source of high ALA, which can be incorporated in our diet to balance the ω-6/ω-3 ratio, which is presently skewed towards ω-6 FAs. Similarly, although the dual purpose varieties were developed with the aim of combining high oil content with good fiber quality traits, it was found that such varieties were in fact inferior in both, the oil content and fiber quality. Thus, the current breeding focus has shifted to develop specific flax varieties for oil and fiber quality traits.

Section 3: Genesis of the thesis

Adequate intake and proper balance of essential fatty acids are necessary to ensure proper physiological functions of human body. However, in our diet, the levels of ω-3 FAs are decreasing in an alarming way. Since flax is the only agricultural source of high ALA (45-65%) it is essential to exploit it for omega-3 nutrition. However, varietal variations exist in the biochemical parameters, oil content and quality and yield. Even though, India ranks the first with 0.45 million ha of land under flax cultivation, production-wise it is placed at 5th in the world, due to its low productivity. The reports detailing systematic efforts towards flax breeding for omega-3 fatty acid enhancement are in scanty and in general, it is largely a neglected crop in India. In view of the nutritional importance and market requirement of flax, it is essential to breed varieties with disease resistance, higher yield potential and oil with high ALA content. It was therefore necessary to initiate molecular and biochemical research in flax to genetically improve its yield as well as nutritional quality. Keeping the Indian flax agricultural scenario in view, certain objectives for my thesis were designed to understand the existing diversity in Indian germplasm and mechanism and regulation ALA synthesis in flax.

3.1 Objectives

1. Assessment of genetic diversity among various Indian flax cultivars using molecular markers
2. Gas chromatographic analysis of diversity in various fatty acid (FA) contents of these flax cultivars
3. FA profiling of various seed developmental stages of ten flax varieties varying in ALA contents
4. Expression analysis of microsomal desaturase genes (*SAD*, *FAD2* and *FAD3*) from the seed developmental stages of the ten varieties
5. Isolation and molecular characterization of complete desaturase genes from the ten varieties

3.2 Organization of the thesis

The thesis is presented in five chapters and the content of each chapter as follows:

Chapter 1: Introduction and review of literature (the current chapter)

Chapter 2: Assessment of genetic and fatty acid diversity among Indian flax varieties

This chapter describes molecular genetic analysis of elite flax genotypes actively involved in Indian flax breeding programs using ISSR markers and its correlation with their ALA contents.

Chapter 3: Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content

This chapter describes the transcriptional correlation of six microsomal desaturase genes (*SAD1*, *SAD2*, *FAD2*, *FAD2-2*, *FAD3A* and *FAD3B*), which act sequentially in the fatty acid desaturation pathway; with the fatty acid accumulation in the various seed developmental stages of ten Indian flax varieties varying in their ALA contents.

Chapter 4: Sequence characterization and *in silico* structure prediction of desaturases from the flax varieties varying in α-linolenic acid content

This chapter describes haplotype analysis based on the sequencing of *SAD1, SAD2, FAD2, FAD2-2, FAD3A* and *FAD3B* genes as well as protein structure predictions based on the deduced AA sequences of these genes from the same ten flax varieties.

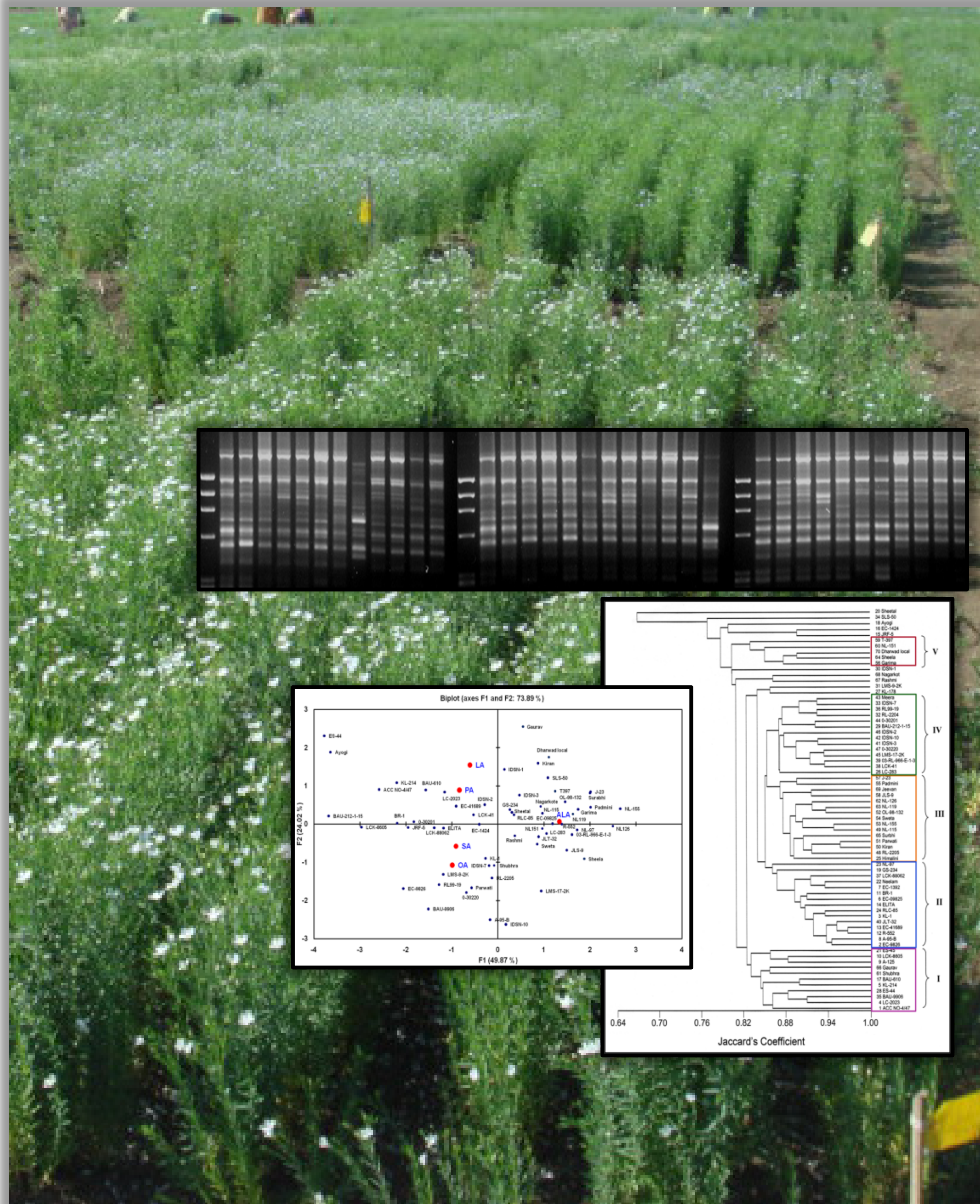
Chapter 5: Summary and future prospects

Bibliography

Annexure I

CHAPTER 2

Assessment of genetic and fatty acid diversity among Indian flax varieties



A part of this work has been published in *Molecular Biotechnology* (2010) 45: 161-170

2.1 Introduction

The knowledge about diversity and genetic relationships among the primary germplasm and breeding material gives insight into crop improvement strategies. Considering the nutritional importance and market requirement of flax, breeding for higher yielding and disease resistant varieties is imperative. Agronomically superior and genetically diverse genotypes are expected to contain unique alleles that can be exploited by crossing them.

Diversity assessment in flax was earlier attempted using morphological parameters (von Kulpa and Danert, 1962; Diederichsen, 2001) and isozyme markers (Mansby et al., 2000). The use of DNA based markers to study the same was first reported by Oh et al. (2000), and later by several groups (Everaert et al., 2001; von Treuren et al., 2001; Fu et al., 2002a; Fu et al., 2002b; Wiesnerova and Wiesner, 2004b; Fu, 2005; Diederichsen and Fu, 2006; Cloutier et al., 2009; Fu and Allaby, 2010; Uysal et al., 2010; Fu, 2011; Smykal et al., 2011). These groups used molecular markers not only to study the genetic diversity in the cultivated flax from different countries all over the world but also, i) to differentiate infraspecific groups, landraces and wild relatives of flax, ii) to address the question of redundancy in the germplasm collections, iii) to find any geographic pattern of variability in the world collection, iv) to study domestication process, etc. Taking into account most of these studies; it was observed that the cultivated flax shows a low genetic diversity compared to its wild relatives or landraces. This is probably because of its domestication process, for the desired characters over so many years.

Apart from the Middle East, which is believed to be the center of origin of cultivated flax, India is proposed to be one of the secondary center of diversity (Vavilov, 1926; Zohary and Hopf, 2000). Indian flaxes are predominantly oil-flax types, as the warmer Indian climate favors oil accumulation in seeds (Diederichsen, 2001; Diederichsen and Fu, 2006). They are bold-seeded with higher oil content (30 to 40%) of which 40 to 50% is ALA. Varieties with 60 to 65% ALA as well as with less than 3% ALA have been developed. The latter are targeted to produce oil with improved tolerance to rancidity (AICRP on linseed, Annual Report 2009-2010). However, the Indian flax germplasm has not been characterized at molecular level for genetic diversity. In the present study, the genetic diversity of Indian flax germplasm was analyzed using ISSR technique, because it achieves higher reproducibility than

RAPD, detects higher percentage of genomic polymorphism than RFLP, is cheaper than AFLP, and gives a complex but polymorphic amplification pattern due to abundance of SSRs in the genome (Zietkiewicz et al., 1994). To the best of our knowledge, this is the first effort to assess molecular genetic relationship and its correlation with ALA content among the elite flax genotypes actively utilized in Indian flax breeding programs.

2.2 Materials and methods

2.2.1 Plant material

Leaf tissue was harvested from flax germplasm collection grown at College of Agriculture, Nagpur, Maharashtra, India. Alternatively, the seeds of the genotypes were grown in pots and individual seedlings were harvested for DNA isolations. Total 70 flax genotypes (Table 2.1) were used for the analysis, which included germplasm accessions, elite cultivars and better performing varieties released for cultivation. To assess the intravarietal genetic variation, five seedlings each were raised for 18 randomly selected genotypes and the DNAs extracted from these individual seedlings were subjected to ISSR analysis.

2.2.2 DNA extraction and PCR amplification

Total genomic DNA was extracted from the young leaf tissue using modified hexadecyl-trimethyl-ammonium bromide (CTAB) method (Richards et al., 1994). In the first step, 1 g of leaf tissue was ground to fine powder in liquid nitrogen and was quickly transferred to a polypropylene tube containing 10 ml of pre-warmed extraction buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris, 20 mM EDTA, 100 mM Sodium metabisulphite, 1% *β*-Mercaptoethanol, pH 8.0).The mixture was swirled gently and incubated at 65°C in water bath for 60 min with intermittent shaking for efficient lysis of cells. After incubation, the tube was cooled, equal volume of chloroform: isoamyl alcohol (24:1) was added, shaken gently, and was then centrifuged at 12,000 *g*, for 10 min at room temperature. The aqueous layer was removed carefully with a Pasteur pipette and collected in fresh polypropylene tube. $1/10^{th}$ volume of 10% CTAB was added followed by addition of equal volume of chloroform: isoamyl alcohol. The mixture was swirled gently and the tube was centrifuged at 12,000 *g*, for 10 min at room temperature. The aqueous layer was removed carefully and collected in a fresh polypropylene tube to which ¼ volume of

precipitation buffer (2% CTAB, 50 mM Tris, 10 mM EDTA, pH 8.0) was added followed by equal volume of sterile deionized water. The tube was incubated overnight at room temperature and next day centrifuged at 10,000 rpm, for 10 min at room temperature and the solution was decanted carefully. As flax contains high amount of polysaccharides and mucilage, the pellet was dissolved in HSTE buffer (High salt Tris EDTA-10 mM Tris, 0.5 mM EDTA, 1M NaCl, pH 8.0) and kept at room temperature till it dissolved. This solution was then precipitated with twice the volume of chilled ethanol and centrifuged at 12,000 *g*, for 10 min at 4°C followed by two 70% ethanol washes. The pellet was dried at room temperature. Sufficient volume of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) was added to dissolve the pellet following which RNase treatment was given. The DNA was quantified spectrophotometrically at 260/280 nm and visually by ethidium bromide staining on 0.8% agarose gel and by comparing it with uncut lambda DNA samples with known concentrations (Sambrook et al., 1989).

A set of 100 ISSR primers (UBC 801-900, UBC set # 9), procured from the University of British Columbia (Canada), was used in the present analysis. These primers were used for initial screening with a subset of flax genotypes and the primers that generated good scorable patterns were further used with the full set of 70 genotypes. Amplification of 20 ng DNA was performed in 25 µl reaction volume containing 10X PCR buffer (10 mM Tris HCl, pH 8.3; 50 mM KCl, 1.5 mM $MgCl_2$), 0.1 mM dNTPs (Amersham Biosciences, USA), 0.4 mM spermidine, 0.3 µM primer and 0.48 U *Taq* DNA polymerase (Bangalore Genei, India) using PTC 225 thermal cycler (MJ Research, USA) programmed for 45 cycles. After initial denaturation at 94°C for 5 min, each cycle comprised 30 sec denaturation at 94°C, 45 sec annealing at 50°C or 60°C depending on the annealing temperature of primers used and 2 min extension at 72°C with 5 min final extension at 72°C at the end of 45 cycles.

The amplification products were resolved along with ΦX174/*Hae*III marker (Bangalore Genei, India) on 1.5% agarose gel and electrophoresed in 0.5X TAE buffer. The gels were stained with ethidium bromide and documented on ImageMaster VDS gel documentation system (Amersham Biosciences, USA). PCR amplifications were repeated at least twice to establish reproducibility of the results.

2.2.3 Fatty acid analysis

2.2.3.1 Preparation of fatty acid methyl esters (FAMEs)

Flax seeds were powdered by using mortar and pestle in liquid nitrogen and this powder (100 mg) was taken in 10 ml screw cap glass tubes. Fatty acids present in the seed oil were esterified adding 5 ml 0.6N methanolic-HCl with 0.2% Butylated hydroxyl toluene (BHT) to the seed powder. Tube with the above mixture was incubated in water bath at 85°C for 2 hr. The tube was removed from the water bath and kept on ice for cooling. The fatty acid methyl esters (FAMEs) thus formed were extracted in 3 ml hexane by following way: In the cooled esterified seed mixture 3 ml hexane was added, the tube was then vortexed properly and centrifuged at 3000 rpm for 15 min at room temperature. The upper hexane layer consisted of FAMEs and with the help of a Pasture pipette; it was separated in a fresh 5 ml screw capped glass vial. Extracts were dried in argon current and reconstituted in 1:200 volumes of chloroform (Manku et al., 1983).

2.2.3.2 Gas chromatography

One µl of chloroform reconstituted FAMEs was injected in AutoSystem XL GC (Perkin Elmer, USA) with SP-2330 Supelco capillary column, 30 m long and 0.32 mm diameter. The temperature program was 150°C for 10 min, followed by 10°C rise/min up to 220°C and steady for 10 min. Helium 1 ml/min was used as the carrier gas. The injector port was maintained at 240°C and FID detector temperature was 275°C. Appropriate fatty acid standards were procured from Sigma Aldrich (USA) and the fatty acid peaks were identified by integrating them with the standards' profiles. The area under the peak was expressed as percentage fatty acid content. Estimation of each sample was repeated minimum three times.

2.2.4 Total oil content

The total oil content data of the genotypes were compiled from All India Coordinated Research Project on Linseed, College of Agriculture, Nagpur, MH, India and All India Coordinated Research Project on Linseed: *Annual Report* Kanpur, UP, India; 2007-2008. Further, absolute ALA content was calculated on the basis of percentage ALA present in the total oil of each genotype.

2.2.5 Data analysis and softwares used for genetic diversity studies

Only consistently reproducible ISSR bands on the gel were scored as present (1) or absent (0) and a binary matrix for each marker was prepared. Similarity matrix within the population was constructed using the software WINDIST of WINBOOT package (Yap and Nelson, 1996). Cluster analysis was performed using MVSP (Multi Variate Statistical Package) version 3.13 (Kovach, 1998) and a dendrogram was generated based on the similarity matrix by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), where the Jaccard's coefficient was used as the measure of similarity or distance. Principal Component Analysis (PCA) was also carried out in MVSP and a scatter plot was developed on the first two components.

The polymorphism information content (PIC) for each locus is calculated as $PIC=1-\sum P_i^2$, where P_i is the band frequency of the *i*th allele (Smith et al., 1997). However, in case of RAPD and ISSR, the PIC was considered to be $1-p^2-q^2$, where p is band frequency and q is no band frequency (Ghislain et al., 1999). The PIC values were then used to calculate an ISSR primer index (SPI), which was generated by summing up the PIC values of all the loci amplified by the same primer. The probability (PI) that two randomly selected genotypes would exhibit identical DNA fragment profiles using the same set of ISSR primers was calculated as, $PI = (X_D)^n$, where X_D represents the average genetic similarity index for all the pair wise comparisons and n is the mean number of fragments obtained per genotype (Wetton et al., 1987).

The heterozygosity per locus (or intralocus gene diversity) (h_i) for a genetic marker is calculated as $h_i=1-\sum p_i^2$ (Nei, 1978). However, for biallelic DNA markers, it is calculated as $h_i =1-p_i^2 - q_i^2$, where p_i is the frequency of the present allele and q_i is the frequency of the second allele. For ISSR markers, which are dominant in nature, the second allele is considered to be null. The frequency of the null allele at locus *i* (q_i) was estimated as $q_i = [x_i]^{1/2}$, where x_i is the no. of individuals with null alleles per locus in the population. Whereas, the frequency of the present allele was estimated as $p_i = 1- q_i$. In case of dominant markers, total and average heterozygosity cannot be estimated, as these do not allow discrimination between heterozygous and homozygous individuals. Therefore, 'average gene diversity' (H_i) given by the formula $H_i = \sum_i^L h_i/L$, was estimated, where L=total number of loci amplified by the marker system.

Based on the ALA content, the genotypes were ranked in ascending order, with the genotypes having the highest ALA content ranked 1 (Table2.1). The ALA contents of the genotypes were correlated with the clusters of the dendrogram. Mean values of each cluster were calculated and subjected to the analysis of variance (ANOVA). The 'F' ratio was calculated and used to compare the variance among the clusters to the variance within the clusters. The probability of 'F' value obtained was compared with the critical alpha level to estimate the significance at 95 and 99% levels. Following this, 'post-hoc' or 'after the fact' test was performed by calculating 't' values between all the clusters. Significant difference among the cluster is a measure of cohesiveness and goodness of fit of the 'chemotype' with the UPGMA cluster.

PCA was performed to study the distribution of 70 genotypes based on their fatty acid (FA) profiles using XLSTAT ver. 2006. Both, the score plot and the loading plot were constructed and compared to analyze the contribution of each FA present in the seed oil for distribution of flax genotypes in various quadrants of the PCA plot. Further, the FA data of all the 70 genotypes were standardized to have mean 0 and standard deviation 1. The standardized data were used for calculating distance matrix based on Euclidean distance method and clustering was done using agglomerative hierarchical clustering (AHC) with Ward's agglomeration method in XLSTAT. The distance matrix and the clusters were formed on the basis of both, all the five FAs profiled for flax varieties as well as considering individual FAs. To find statistically significant correlation between the distance matrices based on FA content and the ISSR molecular data for genotypes, Mantel's test was carried out using XLSTAT.

2.3 Results

2.3.1 Genotypic diversity analysis based on molecular studies

2.3.1.1 ISSR analysis

One hundred UBC ISSR primers were initially screened with a subset of 18 flax genotypes and 12 primers that gave clear and reproducible patterns were used for the analysis of 70 genotypes enlisted in Table 2.1. A total of 136 loci were obtained of which 87 were polymorphic. A maximum of 51 polymorphic loci were amplified in the genotype Rashmi and a minimum in the genotype LC-2023 (33 loci). The percentage of polymorphic loci for all the genotypes ranged from 37.93 to 58.62%,

with an average of 47.41%. The number of amplified loci by the primers varied from 6 (UBC 815) to 20 (UBC 841) (Table 2.2), with a size range of 180 to 2000 bp. The average number of amplified loci and polymorphic loci per primer were 11.3 and 7.25, respectively. The percent loci polymorphism ranged from 11.1 (UBC 809) to 81.8 (UBC 808) with an average of 63.9 across all the genotypes. Figures 2.1A and 2.1B depict the representative amplification profiles of 70 genotypes with primer UBC 835 and UBC 841. It was observed that most of the primers showing polymorphism had AG/GA repeats and the average number of polymorphic loci produced by $(AG)_n$ and $(GA)_n$ repeat containing primers were also similar (7.4). A genetic distance matrix was generated based on these 87 polymorphic loci.

Table 2.1: Fatty acid composition and seed oil content in flax genotypes

S. No.	Genotypes	Fatty acid content (%)					ALA Rank	Total oil content (%)	Absolute ALA content	Absolute ALA Rank
		PA	SA	OA	LA	ALA				
1	ACC NO-4/47	8	9	28	17	38	17	32.00	12.16	11
2	EC-9826	8	7	36	10	39	16	40.00	15.60	8
3	KL-1	6	6	31	14	42	13	38.00	15.96	8
4	LC-2023	8	7	27	15	43	12	41.00	17.63	6
5	KL-214	8	8	28	17	39	16	37.00	14.43	10
6	EC-09825	7	6	24	12	51	4	40.00	20.40	4
7	EC-1392*	-	-	-	-	-	-	36.00	-	-
8	A-95-B	6	6	35	10	43	12	39.40	16.94	7
9	A-125*	-	-	-	-	-	-	39.20	-	-
10	LCK-8605	8	10	30	15	37	18	37.50	13.88	10
11	BR-1	8	10	27	14	40	15	37.23	14.89	9
12	R-552	6	6	24	14	50	5	38.75	19.38	5
13	EC-41689	7	6	29	16	41	14	39.00	15.99	8
14	ELITA	7	7	30	15	41	14	42.00	17.22	7
15	JRF-5	7	7	32	16	37	18	31.00	11.47	12
16	EC-1424	7	7	27	14	44	11	34.11	15.01	9
17	BAU-610	7	6	30	18	38	17	40.00	15.20	9
18	Ayogi	9	4	36	19	31	20	28.00	8.68	14

S. No.	Genotypes	Fatty acid content (%)					ALA Rank	Total oil content (%)	Absolute ALA content	Absolute ALA Rank
		PA	SA	OA	LA	ALA				
19	GS-234	7	7	24	14	47	8	41.00	19.27	5
20	Sheetal	7	6	26	14	47	8	41.00	19.27	5
21	ES-45*	-	-	-	-	-	-	-	-	-
22	Neelam*	-	-	-	-	-	-	41.00	-	-
23	NL-97	6	6	23	13	51	4	42.00	21.42	3
24	RLC-85	6	7	25	16	46	9	43.36	19.95	4
25	Himalini*	-	-	-	-	-	-	-	-	-
26	LC-283	6	6	26	14	49	6	41.00	20.09	4
27	KL-178*	-	-	-	-	-	-	38.82	-	-
28	ES-44	8	11	25	22	34	19	29.00	9.86	13
29	BAU-212-1-15	8	9	31	16	30	21	38.61	11.58	11
30	IDSN-1	7	5	25	17	45	10	38.00	17.10	7
31	LMS-9-2K	7	7	33	12	41	14	40.77	16.72	7
32	RL-2204*	-	-	-	-	-	-	42.44	-	-
33	IDSN-7	7	7	29	11	45	10	42.00	18.90	5
34	SLS-50	7	6	21	15	50	5	37.87	18.94	5
35	BAU-9906	7	9	33	10	41	14	35.06	14.37	10
36	RL99-19	8	7	33	9	42	13	41.00	17.22	7
37	LCK-88062	8	7	30	13	42	13	42.00	17.64	6
38	LCK-41	7	5	30	15	42	13	41.00	17.22	7
39	03-RL-966-E-1-3	6	6	24	13	51	4	41.00	20.91	3
40	JLT-32	6	5	28	14	47	8	38.96	18.31	6
41	IDSN-3	7	5	26	15	47	8	39.00	18.33	6
42	IDSN-10	6	8	31	9	46	9	42.00	19.32	5
43	Meera*	-	-	-	-	-	-	39.00	-	-
44	0-30201	7	9	28	16	39	16	43.00	16.77	7
45	LMS-17-2K	6	6	29	10	47	8	44.00	20.68	3
46	IDSN-2	8	5	28	13	45	10	41.00	18.45	6
47	0-30220	7	8	31	10	44	11	40.00	17.60	6
48	RL-2205	7	8	28	10	46	9	38.60	17.76	6
49	NL-115	7	6	24	13	50	5	42.00	21.00	3

S. No.	Genotypes	Fatty acid content (%)					ALA Rank	Total oil content (%)	Absolute ALA content	Absolute ALA Rank
		PA	SA	OA	LA	ALA				
50	Kiran	8	4	23	14	49	6	43.00	21.07	3
51	Parwati	7	8	30	10	44	11	40.00	17.60	6
52	OL-98-132	7	5	23	13	51	4	40.00	20.40	4
53	NL-155	6	5	20	13	54	1	38.00	20.52	3
54	Sweta	7	6	26	11	49	6	42.00	20.58	3
55	Padmini	7	5	22	12	54	1	43.00	23.22	1
56	Garima	7	6	21	12	53	2	42.00	22.26	2
57	J-23	7	5	21	13	54	1	41.00	22.14	2
58	JLS-9	6	6	25	12	50	5	42.00	21.00	3
59	T-397	7	5	23	14	50	5	44.00	22.00	2
60	NL-151	7	6	25	12	50	5	43.00	21.50	2
61	Shubhra	7	7	29	11	46	9	45.00	20.70	3
62	NL-126	6	6	20	12	54	1	42.00	22.68	1
63	NL-119	7	6	22	12	53	2	42.00	22.26	2
64	Sheela	6	5	26	11	51	4	41.00	20.91	3
65	Surbhi	7	3	24	13	52	3	42.00	21.84	2
66	Gaurav	8	4	22	17	48	7	40.00	19.20	5
67	Rashmi	7	7	25	12	47	8	36.05	16.94	7
68	Nagarkot	8	5	24	11	50	5	39.00	19.50	4
69	Jeevan	7	5	21	12	49	6	37.86	18.55	5
70	Dharwad local	8	6	19	14	52	3	41.00	21.32	3

Notes:

(i) *Enough seed material was not available for fatty acid analysis

(ii) The fatty acids are abbreviated as- PA: Palmetic acid, SA: Stearic acid, OA: Oleic acid, LA: Linoleic acid, ALA: α-Linolenic acid

The range of PIC scores obtained in this study was 0.03 to 0.49, with an average of 0.18. As ISSRs are usually dominant markers, 0.50 will be the highest PIC score for any locus (Li and Nelson, 2001). The ISSR primer index (SPI), which is the sum of PIC values of loci amplified by the same primer, ranged from 0.11 to 3.22 (Table 2.2). This index reveals the information content of the ISSR primers per assay. The probability (PI) of obtaining an identical match by chance using the same set of ISSR primers was very low at 5.1×10^{-8}. The intralocus gene diversity, h_i, for the polymorphic loci ranged from 0.01 to 0.49, whereas the average gene diversity, H_i, obtained using the ISSR markers was 0.15 across all the loci.

Table 2.2: The ISSR primers used, number of amplified & polymorphic loci and percentage polymorphism in evaluation of genetic diversity

Sr. No.	Primer number	Repeat motif *	No. of amplified loci	No. of polymorphic loci	Percentage polymorphism	SPI †
1	UBC 807	(AG)n C	12	7	58.3	1.62
2	UBC 808	(AG)n G	11	9	81.8	1.13
3	UBC 809	(GA)n T	9	1	11.1	0.11
4	UBC 810	(GA)n C	10	5	50.0	0.93
5	UBC 811	(GA)n A	12	9	75.0	1.93
6	UBC 815	(CA)n T	6	4	66.6	0.99
7	UBC 834	(AG)n YT	9	5	55.5	0.94
8	**UBC 835**	(AG)n YC	16	12	75.0	**2.92**
9	UBC 836	(AG)n YG	8	4	50.0	0.11
10	UBC 840	(GA)n YT	9	7	77.7	0.97
11	**UBC 841**	(GA)n YC	20	15	75.0	**3.22**
12	UBC 856	(AC)n YA	14	9	64.3	1.95
Average			**11.3**	**7.25**	**63.9**	**1.40**

* Y- Pyrimidine † SPI – ISSR primer index

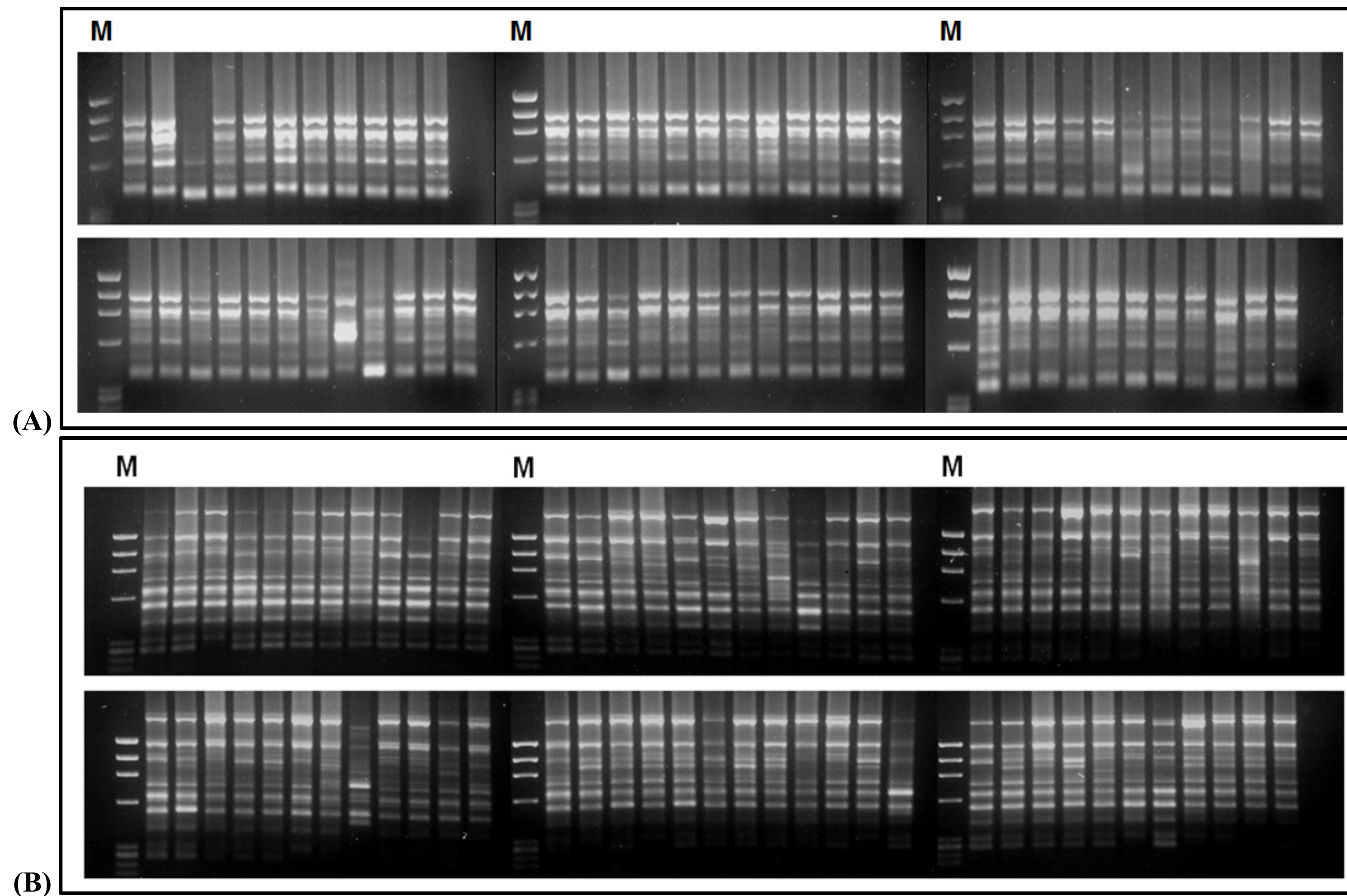


Figure 2.1: Amplification profiles of 70 genotypes using **(A)** UBC 836 primer and **(B)** UBC 841 primer. 'M' represents ΦX174/*Hae*III molecular marker

2.3.1.2 Analysis of intravarietal genetic variation

Genetic variation within the genotypes was assessed in 18 randomly selected genotypes by analyzing five plants each, with the 12 polymorphic ISSR primers. Of the 18 genotypes, nine showed monomorphic patterns with all the 12 primers, whereas the remaining nine showed polymorphism with 1 to 3 primers (different for each genotype). The percentage polymorphism ranged from 0.98 to 8.33. To test whether this variation was statistically significant, an Analysis of Molecular Variance (AMOVA) was performed, where it was observed that the intravarietal variation (7.38%) was significantly less than the intervarietal variation (92.62%). Therefore for further analysis, DNA derived from 1 g leaf tissue of each genotype was used.

2.3.1.3 ISSR based genetic relationships among the flax genotypes

The similarity coefficient for all the 70 genotypes ranged from 0.60 to 0.97, averaging 0.83. A dendrogram based on the UPGMA algorithm using Jaccard's similarity coefficient (Figure 2.2), distributed the 70 flax genotypes in five major clusters at various similarity values. The genotypes, 'Sheetal', 'SLS 50', 'Ayogi', 'EC 1424' and 'JRF 5' were the most distinct from others and did not fall in any cluster. The clusters II, III and IV comprised majority of the flax genotypes while the cluster I with 11 genotypes and Jaccard's similarity value of 0.82 appeared to be distinct from the other clusters. Two pairs of genotypes namely, 'KL 178' & 'LMS09-2K' and 'Rashmi' & 'Nagarkot' did not group well with the rest of the clusters. The former pair branched out from the node common to clusters III and IV and the latter from the node common to cluster I, II, III and IV. Cluster V was genetically the most diverse from the rest of the four clusters. The genotype 'IDSN-1' further joined at the node common to all five clusters.

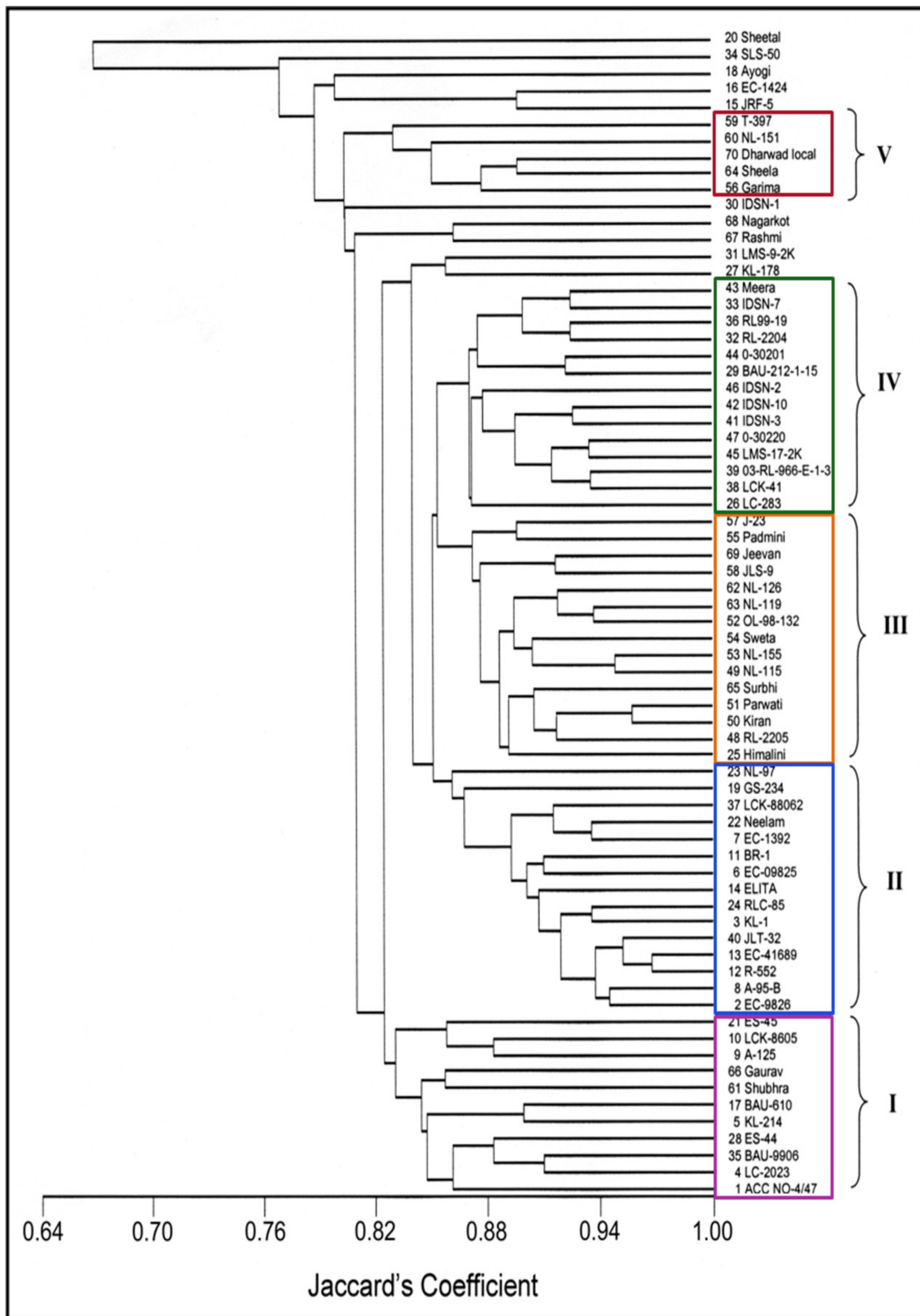


Figure 2.2: Dendrogram based on Jaccard's coefficient of genetic distances

2.3.1.4 Principal Component Analysis

The Principal Component Analysis (PCA) was performed to study the distribution of 70 genotypes according to ISSR molecular data, which revealed the first three most informative principal components with eigenvalues of 0.89, 0.77 and 0.66, respectively, which together accounted for 27.3% of the total genetic variation. PCA in combination with cluster analysis is a useful tool to extract maximum information from the molecular marker data, if the first two or three components explain >25% of the original variation (Messmer et al., 1992). Moreover all the 70 genotypes used for the analysis dispersed well into the four quadrants indicating good polymorphism in this set of flax varieties. The results of the PCA (Figure 2.3) were comparable to those obtained with cluster analysis. The genotypes 'Sheetal' (20) and 'Ayogi' (18) were placed away from the rest of the genotypes, indicating their diverse nature and confirming the results of clustering based on Jaccard's coefficient.

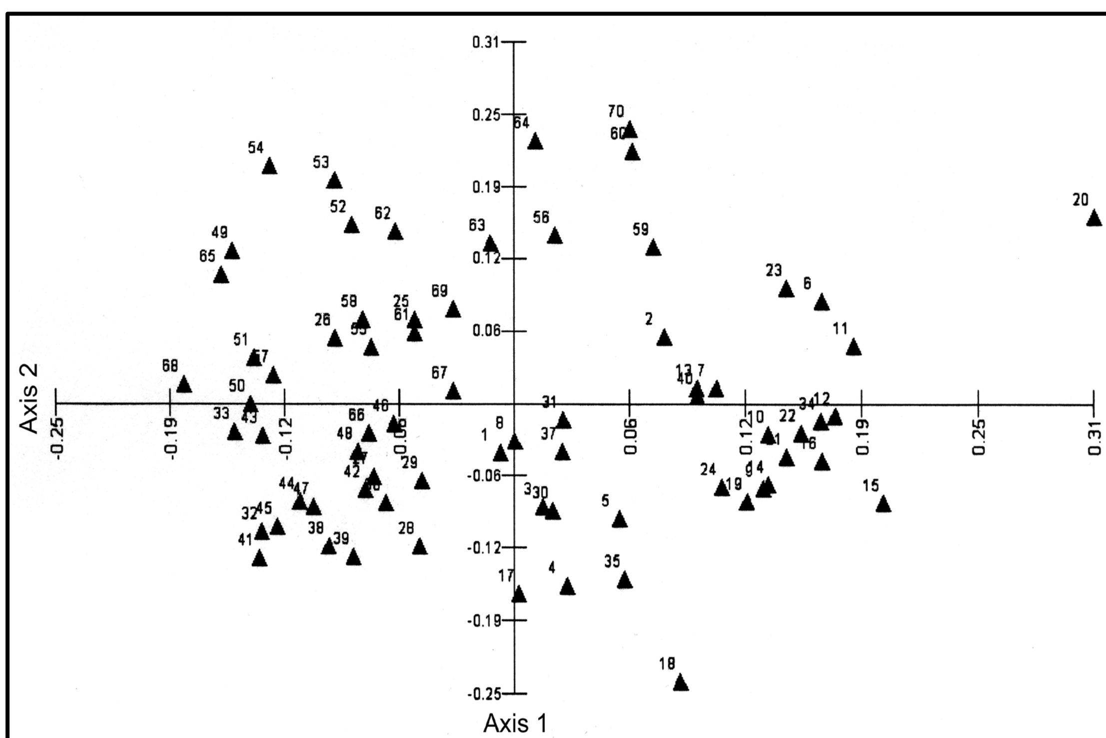


Figure 2.3: Principal Component Analysis (PCA) scatter plot of the flax genotypes (The numbers indicate genotypes as in Table 2.1)

2.3.2 Fatty acid analysis

Fatty acid analysis of the flax seeds showed the presence of five major fatty acids with predominance of 18 carbon species, namely, α-linolenic acid (ALA 18:3), linoleic acid (LA 18:2), oleic acid (OA 18:1) and stearic acid (SA 18:0) and 16 carbon species, namely palmitic acid (PA 16:0), all of which are members of the same pathway catalyzed by the elongase and desaturase enzymes. The ALA contents of the genotypes in the present study ranged from 30% ('BAU-212-1-15', 'Ayogi') to 54% ('NL155', 'Padmini', 'J-23', 'NL126') (Table 2.1). When the fatty acid data were arranged according to the clusters obtained from the molecular data, the high ALA containing genotypes were grouped together in the clusters III and V with less deviation while cluster I with high LA containing genotypes (Table 2.3). The clusters III and V contained cultivars with ranks 1-11 and 1-8 respectively, which correspond to ALA ranging 54-44% and 54-47%, respectively. Further, the average ALA contents of the clusters III and V were significantly different (higher) than that of other clusters, as indicated by the probability values (Table 2.3 and Table 2.4A). Average LA value of Cluster I was significantly different (higher) than all the remaining clusters (Table 2.3 and Table 2.4B) while, average OA values of cluster III and Cluster V were significantly different (lower) than that of remaining three clusters (Table 2.3 and Table 2.4C). Average SA values for all the five groups were low and showed very little variation among them (Table 2.3 and Table 2.4D). It can be clearly observed that the clusters III and V containing high ALA varieties had significantly lower OA and SA content (Table 2.3).

Table 2.3: Fatty acid content (%) among the groups of flax genotypes

Specific fatty acids*	Group I	Group II	Group III	Group IV	Group V
ALA	40.55 ±4.4	44.84 ±4.0	51.00 ±2.7	44.50 ±3.9	51.40 ± 1.1
LA	15.77 ±3.6	13.53 ±1.8	12.14 ±1.1	12.91 ±2.2	12.80 ± 1.1
OA	28.00 ±3.2	28.23 ±4.1	24.14 ±2.5	28.83 ±2.5	23.2 ± 2.3
SA	7.88 ± 2.1	6.61 ±1.19	5.78 ±1.12	6.83 ±1.4	5.6 ± 0.54
PA	7.77 ±0.6	6.76 ±0.8	6.92 ±0.4	6.91 ±0.8	7.0 ± 0.70

* For fatty acid abbreviations, please refer footnote of Table 2.1

Table 2.4A: Probability values among the groups of flax genotypes based on ALA content

	Group I	Group II	Group III	Group IV	Group V
Group I		0.0119 *	0.0000 **	0.0215*	0.0002**
Group II			0.0001 **	0.8144	0.0031**
Group III				0.0001 **	0.8326
Group IV					0.0025**
Group V					

Table 2.4B: Probability values among the groups of flax genotypes based on LA content

	Group I	Group II	Group III	Group IV	Group V
Group I		0.0126 *	0.0003 **	0.0015 **	0.0251*
Group II			0.0724	0.1731	0.4549
Group III				0.3154	0.1483
Group IV					0.2598
Group V					

Table 2.4C: Probability values among the groups of flax genotypes based on OA content

	Group I	Group II	Group III	Group IV	Group V
Group I		0.4584	0.0013 **	0.2858	0.0041**
Group II			0.0003 **	0.3063	0.0019 **
Group III				0.0001 **	0.4192
Group IV					0.0007 **
Group V					

Table 2.4D: Probability values among the groups of flax genotypes based on SA content

	Group I	Group II	Group III	Group IV	Group V
Group I		0.0245 *	0.0003 **	0.0404	0.0015 **
Group II			0.0331 *	0.4119	0.0621
Group III				0.0221	0.4608
Group IV					0.0455
Group V					

* Significant at 95% probability

** Significant at 99% probability

2.3.2.1 Total oil content

Total oil content of the genotypes ranged from 28 to 45% (Table 2.1), with an average of 40%. Ayogi had the lowest oil content (28%), while Shubhra yielded maximum of 45% oil. Based on the oil content, absolute ALA content was also calculated with maximum of 23.22 in Padmini and minimum of 8.68 in Ayogi (Table 2.1).

2.3.2.2 Principal Component Analysis

PCA was performed to study the distribution of 70 genotypes based on their FA composition (composed of five FAs *viz*., PA, SA, OA, LA and ALA). First two principal components accounted for 73.89% of the total observed variance (Figure2.4) in the FA composition among 70 genotypes.PC1 had positive loading of PA, LA and ALA and negative loading of SA and OA, while PC2 had positive loading of ALA and negative loading of PA, SA, OA and LA. These loadings separated ALA in the 1st quadrant, LA and PA in the 2nd quadrant and SA and OA in the 3th quadrant. The PCA scores distributed 70 genotypes in all the four quadrants. It was observed that ALA was the contributing FA for the distribution of varieties in the 1stand the 4thquadrant; LA and PA for the varieties in the 2ndquadrant while SA and OA for the varieties in the 3rdquadrant.

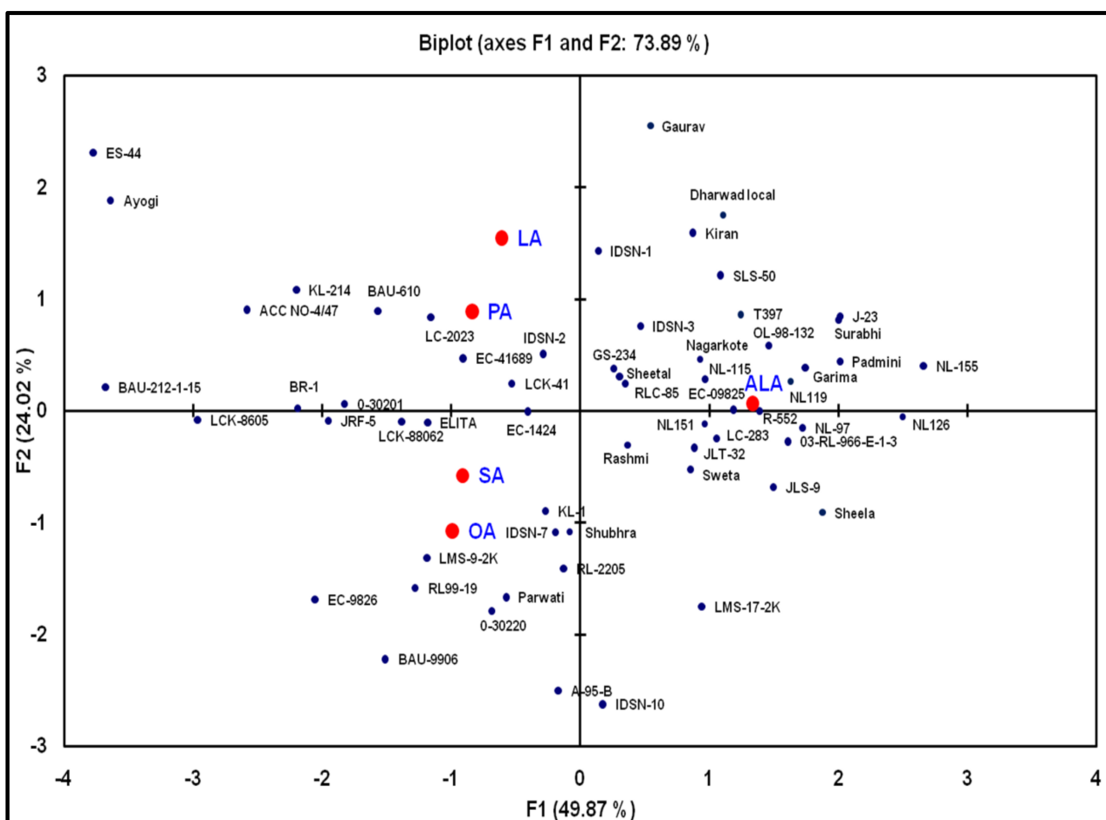


Figure 2.4: Principal Component Analysis (PCA) - distribution of genotypes based on the fatty acid composition (All the five FAs *viz*., PA, SA, OA, LA and ALA, were considered for the analysis)

2.3.2.3 Agglomerative hierarchical clustering (AHC)

Based on the standardized FA data of 70 flax genotypes, distance matrix using Euclidean distance method was generated, where the distance ranged from 0.24 to 7.56 with average of 2.65. Using agglomerative hierarchical clustering (AHC) with Ward's agglomeration method, five clusters were formed for the flax genotypes when all the five FAs as well as individual FAs were considered. Further, on performing the Mantel's test to find correlation between the distance matrices based on FA content and the molecular data for genotypes, it was observed that only the matrices based on LA and ALA data for the genotypes showed statistically significant (p<0.0001) correlation with the distance matrix based on the ISSR molecular variation.

2.4 Discussion

2.4.1 ISSR marker based genetic diversity

Information about the distribution of genetic diversity conserved in germplasm collections is important for unlocking that diversity for further use (Tanksley and McCouch, 1997). ISSR-PCR assays use Simple Sequence Repeat region anchored primers that bind to abundant microsatellite loci of the genome and show dominant inheritance. It is, therefore useful in detecting genetic diversity by amplifying a large number of loci to reveal polymorphism in wider portion of the genome simultaneously. Charters et al. (1996) reported 56 variable bands in oilseed rape cultivars with only two primers that could distinguish all the 20 cultivars. While Huang and Sun (2000) observed 2071 ISSR loci with 15 primers in 40 genotypes of Ipomea and 62% of the loci were polymorphic. Flax genome is relatively small in size; 370Mb/haploid genome divided in 15 similar appearing chromosomes; and 35% of the genome is highly repetitive with tandemly arrayed sequences (Cullis, 1990; Oh et al., 2000; Ragupathy et al., 2011). Wiesnerova and Wiesner (2004b) reported 72.6% polymorphism using nine anchored ISSR primers for 53 Czech flax cultivars, while in the present study; we detected 63.9% average polymorphism in 70 flax genotypes using 12 ISSR primers. The lower level of polymorphism indicated less diversity in the germplasm used in our analysis.

The average gene diversity observed among the genotypes was low (H_i=0.15), which could be because of the fact that most of the genotypes taken for this study were either breeding lines, released varieties for dual-purpose (oil and fiber) or elite varieties, which might be genetically similar. Low and non-significant intravarietal variation observed within few randomly selected genotypes indicated purity of the seeds of individual genotypes used in the study. The PIC value is a measure of variability at a specific locus. Higher the PIC value for a locus, higher is the probability that polymorphism will exist between any two randomly selected genotypes at that locus. The range of PIC scores in this study was 0.03 to 0.49, with an average of 0.18. The low average PIC value obtained in the present study again underscores the nature of the genotypes used. On the other hand, ISSR primer index (SPI) is indicative of the efficiency of the marker system. In our study, the SPI ranged from 0.11 to 3.22, with an average of 1.40. The primers UBC 841 and UBC 835, having the higher SPI values of 3.22 and 2.92, respectively, were found to be the two

best UBC-ISSR primers for detecting polymorphism in flax. The PI value (probability of two randomly selected genotypes exhibiting identical DNA fragment profiles using a set of primers) indicates the ability of a set of primers to uniquely distinguish two genotypes. The PI value using the 12 ISSR primers was 5.1×10^{-8}, which is very low, indicated the efficiency of these primers to uniquely identify different genotypes.

2.4.2 Fatty acid diversity and its correlation with the genetic diversity among the genotypes

A wide range of diversity for fatty acid composition was observed in the present study as well as among the Canadian flax cultivars (Diederichsen, 2001). The 70 flax genotypes in our study were grouped into five clusters based on the ISSR data. Correlation of FA contents with the clusters exhibited a meaningful segregation of clusters III and V with high ALA content while cluster I with that of LA (Table 2.3). The genotypes in these clusters suggested genetic differences associated with these chemotypes. To test if there was a true association between ALA content and genetic diversity, the genotypes were ranked according to ALA content and the ranks were placed against genotypes in each group. Although, the cluster III and V housed ranks 1-11 and 1-8 respectively, the correlation of the ranks with the other groups was not significant. However, Mantel's test revealed that the molecular variation statistically correlated well with the LA and ALA variation within the varieties analyzed. This indicated that even though the ISSR markers could not completely differentiate the flax varieties based on their FA composition, some of the loci amplified by them might be linked with the genes or QTLs involved in the FA biosynthetic pathway. This was augmented by the correlation between molecular variation and LA and ALA content in the genotypes.

All the fatty acids that we studied are from the same metabolic pathway; where SA is converted to OA, OA to LA and LA to ALA. Although, these conversions are mediated by $\Delta 9$, $\Delta 12$ and $\Delta 15$ desaturases, the influence of their variation in regulation of expression might not correlate them completely with the genetic groupings in the dendrogram. These results indicate that genic or functional markers for the fatty acid biosynthesis pathway and their expression patterns and levels might be necessary for elucidating the diversity of FA content within the flax genotypes.

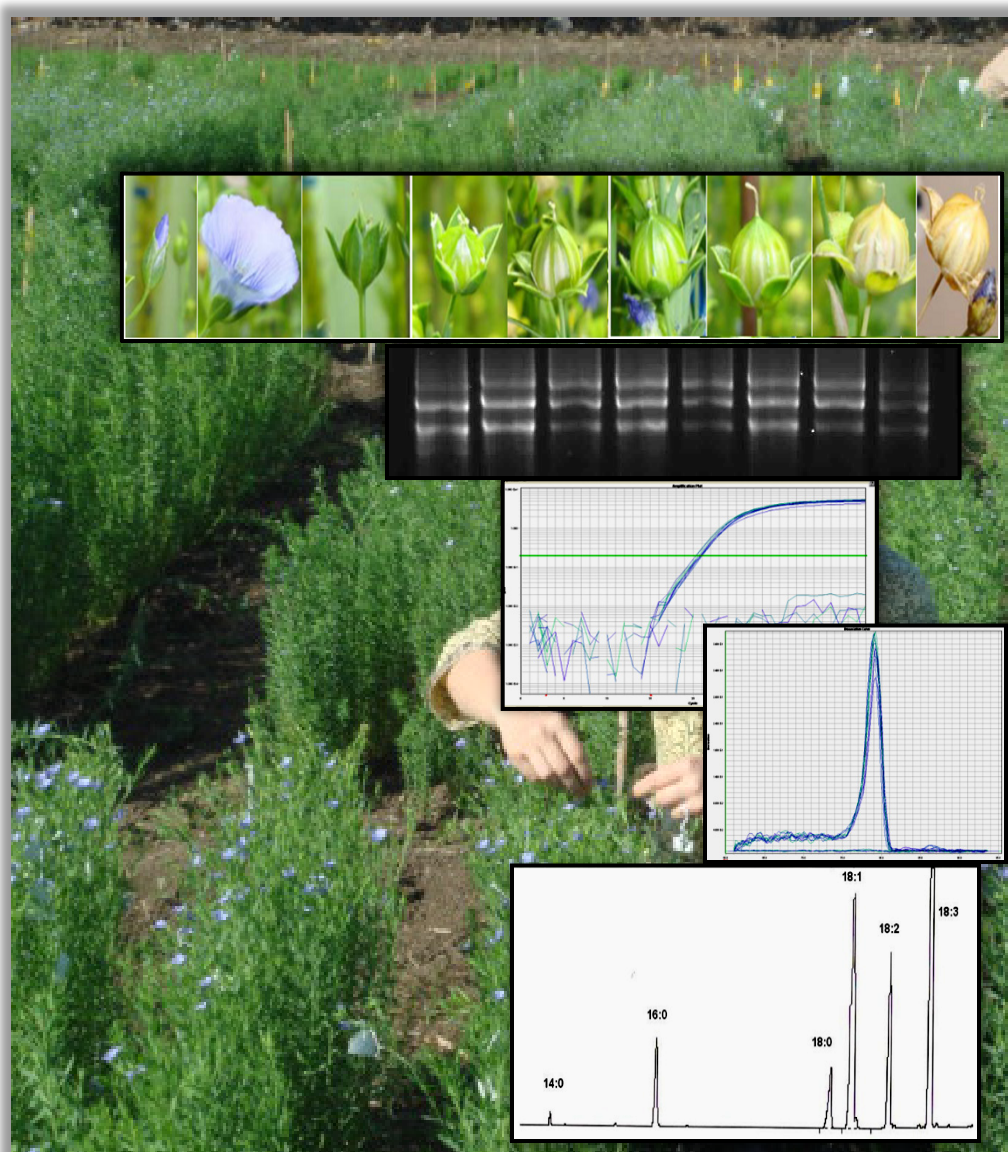
2.4.3 Implications of genetic diversity

The basis of crop improvement is selection operating on genetic variability, which provides adaptability to variables like environment, pest and disease incidences and market. The knowledge of genetic variation is therefore useful to select genotypes for breeding programs to develop new varieties for target traits, environments and markets. Several researchers have reported the use of molecular genetic distance to select diverse parents; which when crossed, gave higher heterosis in various crop plants (Xiao et al., 1996; Lanza et al., 1997; Paz and Veilleux, 1997; Chowdari et al., 1998; Sant et al., 1999; Joshi et al., 2001; Riaz et al., 2001). Riaz et al. (2001) showed that the crosses of *Brassica napus* lines located in different clusters had higher seed yield than those from the same clusters. In related studies, Paz and Veilleux et al. (1997) found that RAPD markers could facilitate identification of diverse parents to maximize the expression of heterosis in *Solanum phureja* hybrids. The parental genetic distance was successfully associated with tuber yield among their progenies. Similarly, Lanza et al. (1997) used RAPDs to determine the extent of genetic diversity among maize inbred lines, for allocating genotypes into various groups and also to select the parents from different clusters to develop superior hybrids.

This study identified the genotypes ‘Sheetal’, ‘SLS 50’, ‘Ayogi’, ‘EC-1424’ and ‘JRF5’to be genetically the most diverse from other genotypes. Their potential in containing unique alleles can be exploited by crossing them with elite lines from other genetic clusters. For example the crosses, ‘Ayogi’ (low ALA) X ‘NL155’ (high ALA), ‘Ayogi’ (low ALA) X ‘NL97’ (high ALA), ‘Neelam’ (susceptible to powdery mildew and bud fly) X ‘JRF5’(resistant to powdery mildew and bud fly), may be made, which might produce high yielding flax varieties, compared to other crosses involving parents from the same clusters. However, overall, the study revealed low genetic diversity in the 70 Indian flax genotypes that are actively being used in breeding programs. This emphasizes the need to widen and diversify the genetic base of Indian flax breeding material, by using genetically diverse lines from the flax germplasm collection. It is crucial to seek new sources of favorable alleles by collecting and utilizing the landraces, local and exotic varieties in the breeding programs. It is also vital to eliminate duplicates and to establish and catalog a core germplasm collection for better management and utilization in Indian flax breeding programs.

CHAPTER 3

Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content



This work has been communicated to
Phytochemistry

3.1 Introduction

Desaturation is an important biochemical process in the FA biosynthesis pathway. Fatty acid desaturases (FADs) are hence the key enzymes that convert saturated FAs to unsaturated FAs. Among these, the three desaturases, *viz*., SAD (Δ9), FAD2 (Δ12) and FAD3 (Δ15), drive the poly-unsaturated fatty acids (PUFA) synthesis pathway and appear to exist in multiple forms and locations within cells in different plant systems. In case of flax, two *SAD* genes *viz*., *SAD1* and *SAD2*, are reported (Singh et al., 1994; Jain et al., 1999). However, the expression profile of only *SAD1* gene isoform during seed development in flax variety AC McDuff has been analyzed by Fofana et al. (2006). The next desaturase in the series is membrane bound FAD2 and in flax, two full length *FAD2* genes *viz.*, *FAD2* and *FAD2-2* have been cloned and characterized (Krasowska et al., 2007; Khadake et al., 2009). Expression profile of *FAD2* gene during seed development was also studied in Canadian flax variety AC McDuff by Fofana et al. (2006). Lastly, three microsomal *FAD3* genes *viz*., *FAD3A, FAD3B* (Vrinten et al., 2005) and *FAD3C* (Banik et al., 2011) have been reported in flax. Transcript profiling of these three genes during seed development was performed by Banik et al. (2011) in four Canadian flax varieties varying in their ALA content.

Traditional Indian flax varieties are known to have high oil content (30-40%) with varying amounts of ALA. However, the molecular insight into ALA biosynthesis is not yet reported. In the present study, we performed FA profiling of various seed development stages of ten Indian flax varieties, classifying them in high and low ALA groups based on their ALA content at the mature stage. Further, transcript analysis of all the three desaturase genes was performed in the same tissues using real time PCR, so as to determine the role of each desaturase gene in the differential accumulation of ALA in these flax varieties.

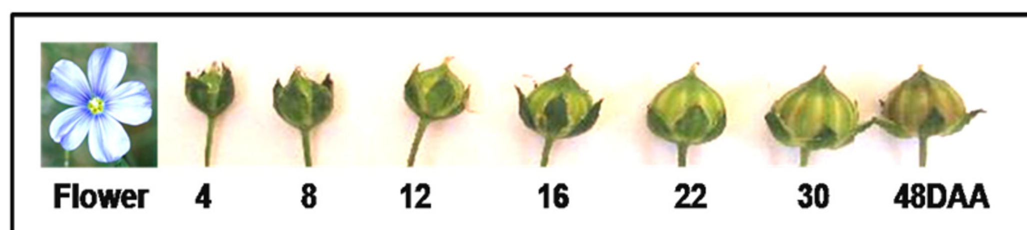
3.2 Materials and methods

3.2.1 Plant material

High and low ALA containing varieties of flax were selected based on earlier data obtained while studying flax genetic diversity. The varieties with more than 45% ALA (NL260, EC9825, Padmini, NL97 and Surabhi) were considered under high ALA group; while those with less than 45% ALA were grouped under low ALA group (JRF5, Acc. No. 4/47, Ayogi and ES44). TL23, a mutant variety with less than

2% ALA was also included in the study. These varieties were grown at the fields of All India Coordinated Research Project on Linseed (flax), College of Agriculture, Nagpur, Maharastra, India in triplicate sets. For fatty acid profiling by gas chromatography (GC) analysis and expression studies, flowers of each variety were individually tagged on the day of anthesis (Figure 3.1A). Some of the open flowers were collected on the same day and constituted the 0 days after anthesis (DAA) or flower stage of the sample set. Further, developing bolls were collected on 4, 8, 12, 16, 22, 30 and 48 days after anthesis (DAA), with a total of eight time-points (including flower stage). The stages 4-8 DAA represent early maturity, 12-16 DAA mid maturity, 22-30 DAA late maturity, while 48 DAA represents mature seeds (Figure 3.1B). The tissues were immediately frozen in liquid nitrogen and stored at -80°C till the extraction of total RNA and the fatty acid methyl esters (FAMEs).

(A)



(B)

Figure 3.1: (A) Flower tagging, **(B)** Eight developing seed stages of flax

3.2.2 Fatty acid analysis

Fatty acids methyl esters (FAMEs) were extracted from each of the eight seed developmental stages of the ten flax varieties as described in Chapter 2. The extractions were performed twice with three bolls per replicate (biological replicates). One µl of chloroform reconstituted extracts were injected in 6890N network GC system (Agilent Technologies, USA) with SP-2330 Supelco capillary column, 30 m long and 0.32 mm diameter. The GC was programmed as detailed in Chapter 2. Fatty acid peaks were identified and its percentage at each stage was also obtained as described there. FA profiling for each sample was repeated minimum twice (technical replicates). For obtaining the final percentage of FA content at each developmental stage, mean value of technical replicate of each biological replicate was obtained and further averaging of the mean values of all the biological replications was done. To analyze the distribution of ten flax varieties based on the FA data, PCA was carried out using Systat software (Version 11, Richmond, CA).

3.2.3 Transcript profiling

3.2.3.1 RNA extraction and cDNA synthesis

RNA from the eight seed developmental stages of ten flax varieties was extracted using Spectrum^{TM} Plant Total RNA Kit (Sigma-Aldrich, USA). The developing bolls were ground into a fine powder in liquid nitrogen using a precooled mortar and pestle. This tissue powder (100 mg) was further used to extract RNA according to the manufacturer's instructions. The isolated total RNA was quantified using Nanodrop 1000 spectrophotometer (Thermo Scientific, USA). The quality and relative quantities were also checked on 1% standard RNA formaldehyde agarose gel. The RNA extractions were performed in triplicates (three biological replicates) and each replicate consisted of 3-5 bolls. The extracted total RNA was treated with RNase-free DNase (Promega, USA) and 1 µg of this DNA-free total RNA was reverse-transcribed using oligo(dT) primer and MultiScribe^{TM} reverse transcriptase enzyme (Applied Biosystems, USA) in a 20 µl reaction volume. The cDNA synthesis was repeated for all the three biological replicates, which were used for further desaturase gene expression analysis by real-time PCR.

3.2.3.2 Real-Time PCR and analysis

Real-time PCR was performed on 7900HT Fast real-time PCR system (Applied Biosystems, USA) using FastStart universal SYBR green master mix (Roche, USA) and the gene-specific primers (designed using Primer3; Rozen and Skaletsky, (2000)) listed in Table 3.1. PCR conditions were optimized for annealing temperature, primer concentration and template dilutions/concentrations. Each 10 µl real-time PCR cocktail contained 0.3-0.4 µM (according to primer pair used) concentrations of both the forward and the reverse gene-specific primers, 4 µl of diluted first strand cDNA, 1X SYBR green master mix and autoclaved milliQ water to make up the reaction volume. Real time PCR amplification reactions were performed with the following conditions: 95°C denaturation for 10 min, followed by 40 cycles of 95°C for 3 s, with primer annealing and extension at 60°C for 30 s. Following amplification, a melting dissociation curve was generated using a 60–95°C ramp in order to monitor the specificity of each primer pair.

Three flax genes, *viz*. elongation factor 1-α (*EF1α*), eukaryotic translation initiation factor 5-α (*ETIF5α*) and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), were initially evaluated as reference or housekeeping genes (Huis et al., 2010) and the *ETIF5α* gene was selected for further analysis based on its stability of expression across the tissue types. The PCR efficiencies of the target and reference genes were determined using the LinReg software (Ramakers et al., 2003; Ruijter et al., 2009). Only those combinations of primer pairs and cDNA dilutions showing similar efficiencies (between 1.92 and 2.05) and R^2 values higher than 0.998, were considered for further study. Three technical replicates of each of the three biological replicates were performed. The mean of these replicates was used for relative gene expression analysis by $2^{-\Delta Ct}$ method (Livak and Schmittgen, 2001). The relative target gene expression was based on the variation in the number of thresh hold cycles (Ct) and was calculated in relation to the reference gene(ΔCt) in each of the developmental stages of all the ten varieties.

Table 3.1: Primers used for real-time PCR

Gene	Primer	Forward Primer Sequence (5' → 3')	Reverse Primer Sequence (5' → 3')
ETIF5α	ETIF-RT 1F/1R	CCCATCTGGACAGCTTCATT	TGCCACATGTGAACCGTACT
FAD3A	LUFAD3A-RT 1F/1R	CCGAACGAGAGGACATCGGTCA	AAGACCGGAAGGAAGCCGTAGATG
FAD3B	LUFAD3B-RT 2F/2R	CGTACCTTATCTGATATTCGTGGCA	CCCTCCACGTAGGTAGCTCCAT
FAD2-1	LUFAD2-1 RT 2F/2R	CACCGTGTGCTACATCCTATACAG	GTGATGAGGACTAGGAATCCATTC
FAD2-2	LUFAD2-2 RT 1F/1R	CGGTGTTCCATAACATCACG	TCGCCTTTGTAGCTTCCATT
SAD1	LUSAD1 RT 1F/1R	GGTTGGGCTGAGGATGTTCTATTA	CTCCTCGAACCCATCCGACTCA
SAD2	LUSAD2 RT 2F/2R	CCTGACGGTACAGTGTTGGCT	CCCCGATGCGTTGAGCGACC

3.3 Results

3.3.1 Fatty acid accumulation patterns in developing seeds

Six fatty acids were profiled at all the eight developing seed stages for ten varieties using GC and the profiles are depicted graphically (Figure 3.2 and 3.3) and as bar diagrams (Figure 3.4). Among these, myristic acid (MA; C14:0) was almost negligible throughout the developmental stages across the ten flax varieties. The palmitic acid (C16:0) content was the highest at flower stage and gradually decreased through boll development to maturity in the ten varieties. The proportion of stearic acid (SA; C18:0) remained nearly constant in all the developing stages, irrespective of the varieties. Oleic acid (OA; C18:1) content was low at the flower stage, but showed an increasing trend up to 8-16 DAA and later remained steady or declined slowly up to maturity in many of the varieties. The linoleic acid (LA; C18:2) content was high at the flower stage and gradually showed decline from 4 to 48 DAA in all the varieties except TL23, where it accumulated steadily till maturity. Conversely, the α-linolenic acid (ALA; C18:3) content was high at the flower stage, then declined at 4 DAA and again increased from 8 or 12 DAA till maturity in all the varieties except TL23.

There was a major difference observed in the accumulation trend of LA and ALA contents in TL23, when compared with the remaining nine varieties. In TL23, ALA content decreased steadily from 4 to 48 DAA; and at maturity, accounted only for 1.83%±0.07. On the contrary, LA content accounted for 63.29%±0.02 of the total FA composition.

Based on the FA accumulation at 48 DAA, the remaining nine varieties (other than TL23) showed more proportion of ALA content (35 to 53%) than any other FA. They revealed two ranges as 35.61±0.53% to 42.41±1.02% and 45.19±1.50% to 52.63±0.77% dividing them into two groups as high ALA (with >45% ALA) and low ALA (with <45% ALA) (Table 3.2). Student's T-test was performed and the difference in the mean ALA content between these two groups was statistically significant ($p<0.002$). In the high ALA group, the average ALA content was 48.76%±1.51, LA content was 13.56%±0.61 and OA content was 25.56%±1.53 of the total FA composition. While for low ALA group, the average ALA content was 39.53%±1.61, LA content was 15.56%±0.42 and OA content was 30.23%±1.66 of the total FA composition (Table 3.3). These results indicated much efficient conversation

of OA to LA and finally to ALA taking place in the high ALA group than the low ALA group. But in case of the mutant variety TL23, there was efficient conversion only up to LA and the conversion of LA to ALA was inefficient leading to very low ALA accumulation.

Table 3.2: ALA contents (%), oil content (%) and absolute ALA content of ten flax varieties

Variety	ALA content (%) ± SE (by GC)	Oil content (%) (by NMR)	Absolute ALA content
Group I- High ALA varieties			
NL260	52.63 ±0.77	28.94	15.23
EC9825	52.10 ±2.09	27.99	14.58
Padmini	47.27 ±0.96	32.79	15.50
Surabhi	46.62 ±0.92	30.08	14.02
NL97	45.19 ±1.50	31.26	14.13
Group II- Low ALA varieties			
Acc.No.4/47	42.41 ±1.02	25.40	10.77
JRF5	41.91 ±0.23	24.90	10.44
ES44	38.18 ±2.39	25.21	9.63
Ayogi	35.61 ±0.53	21.66	7.71
Mutant variety			
TL23	1.83 ±0.07	29.07	0.53

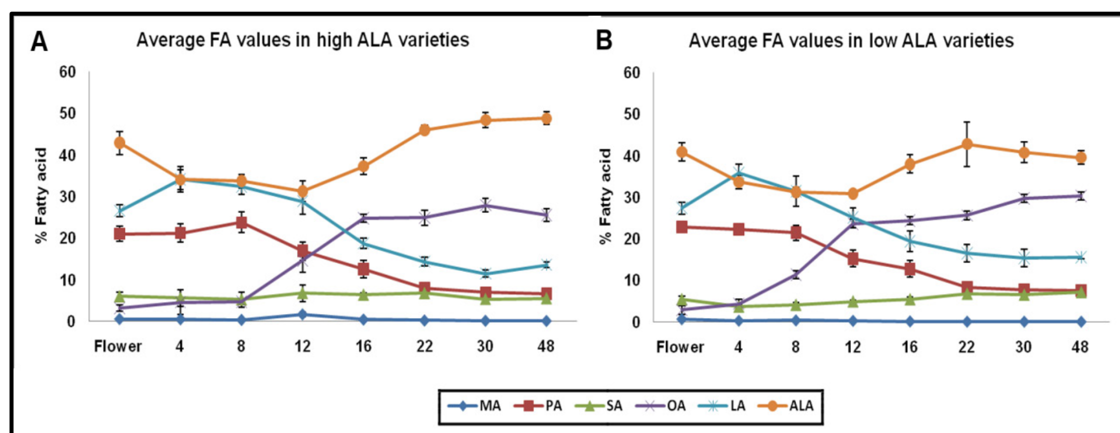


Figure 3.2: Trend of fatty acid accumulation in eight developing seed stages of **(A)** high and **(B)** low ALA containing varieties (considering average FA values of high and low ALA flax varieties)

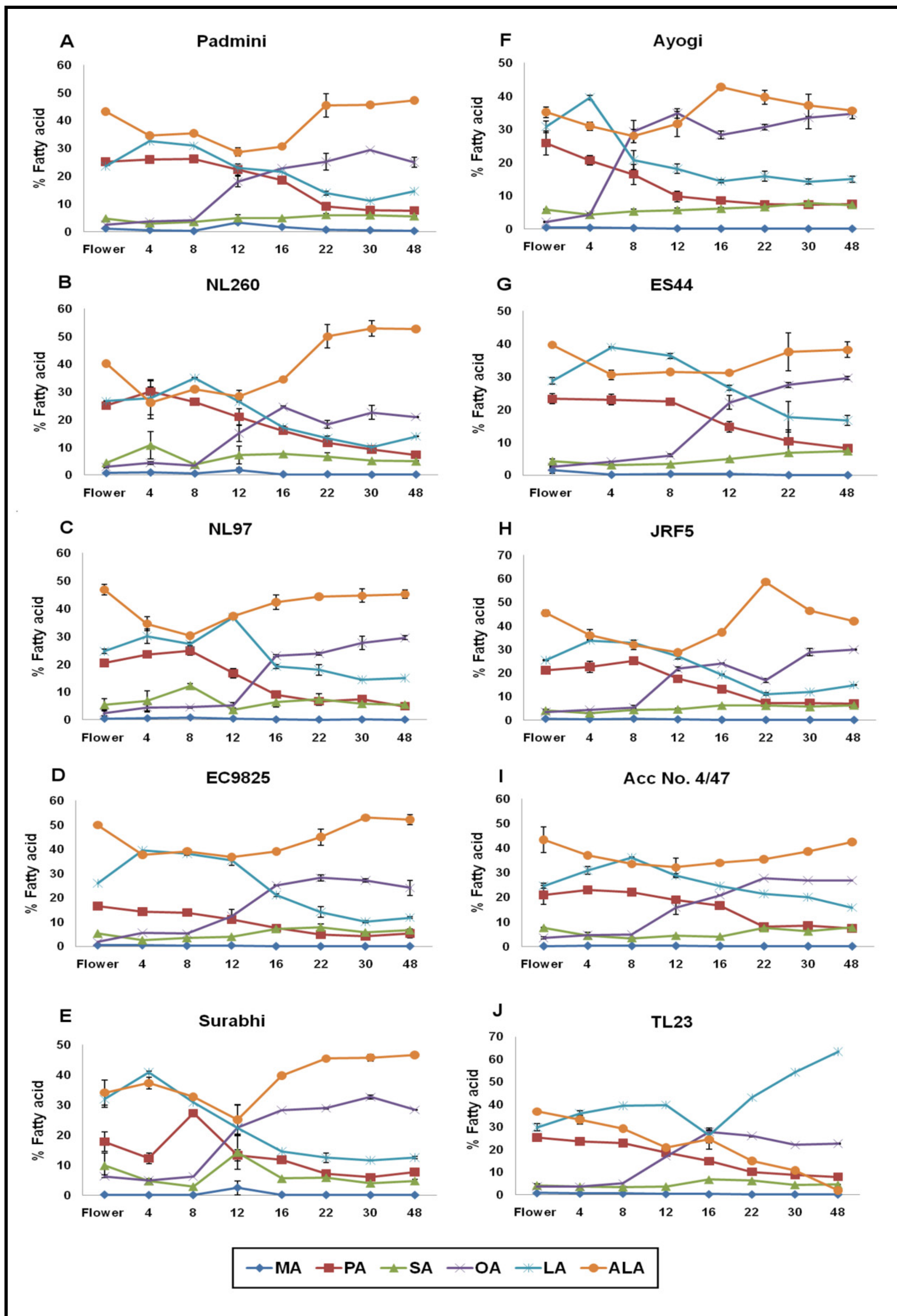


Figure 3.3: Fatty acid profiles of ten flax varieties in eight developing seed stages. **A-E:** high ALA varieties, **F-I:** low ALA varieties and **I:** Mutant variety.

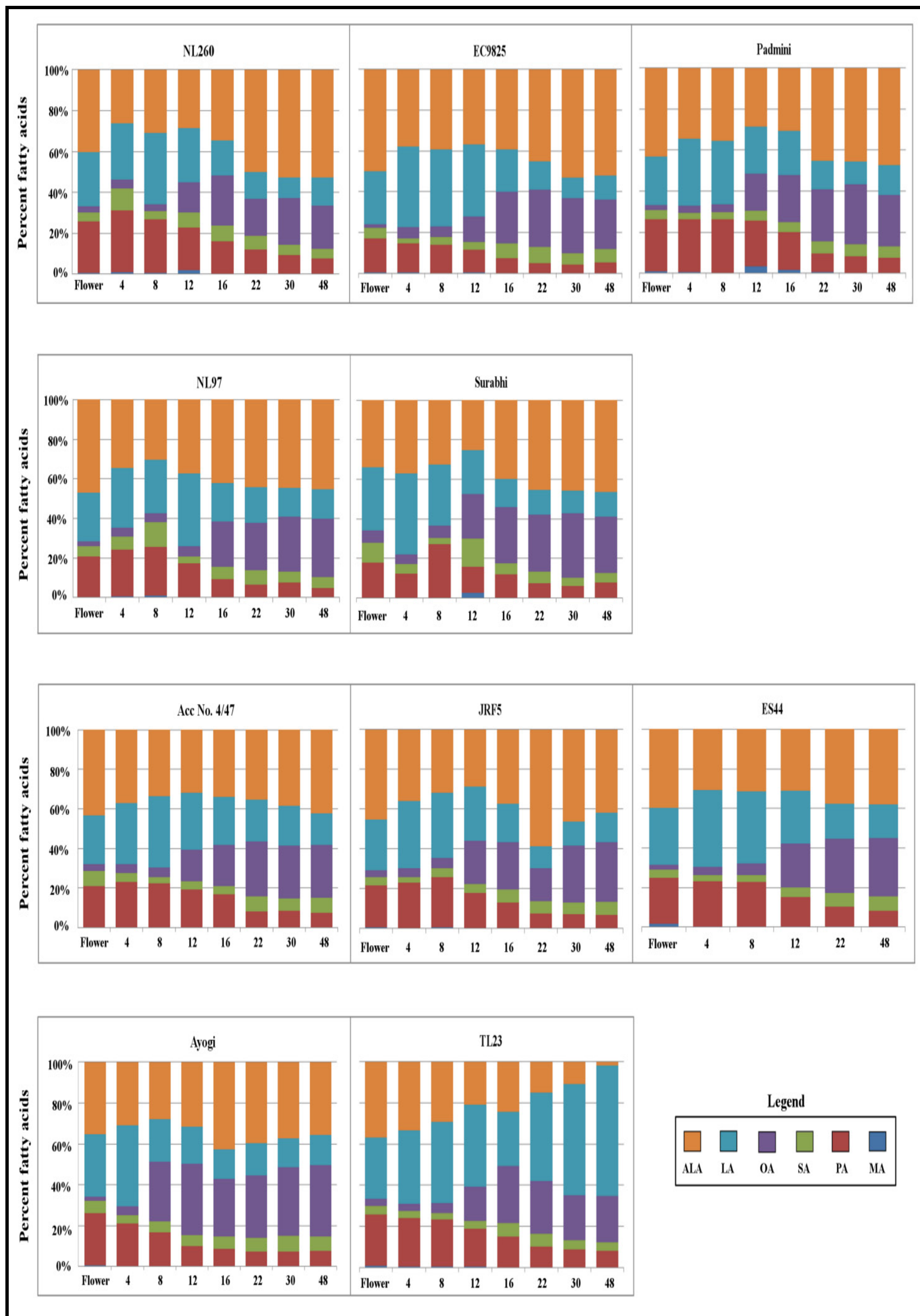


Figure 3.4: Fatty acid profiles of ten flax varieties in eight seed developmental stages. We analyzed six fatty acids [Myristic (MA), Palmitic (PA) Stearic (SA), Oleic (OA), Linoleic (LA) and Linolenic (ALA)] from eight seed developmental stages, from flower to mature stage (48 DAA) in flax

Table 3.3: Fatty acid (FA) levels in variable ALA groups. Average content of six FAs (% ± SE) *viz*., Myristic (MA), Palmitic (PA), Stearic (SA), Oleic (OA), Linoleic (LA) and Linolenic (ALA) in high and low ALA groups and TL23

	MA	PA	SA	OA	LA	ALA
High ALA group	0.10±0.03	6.54±0.060	5.48±0.33	25.56±1.53	13.56±0.61	48.76±1.51
Low ALA group	0.05±0.01	7.49±0.26	7.14±0.28	30.23±1.66	15.56±0.42	39.53±1.61
TL23	0.10±0.01	7.82±0.19	4.38±0.07	22.57±0.16	63.29±0.02	1.83±0.07

Principal Component Analysis (PCA) was carried out to analyze the distribution of flax varieties based on the contents of four major FAs, *viz*., SA, OA, LA and ALA in the mature seeds of the ten flax varieties. Four principal components (PC) could explain all the variation in the data, while the first two principal components accounted for 89% of the total observed variance (Figure 3.4). PC1 had positive loading of SA, OA and ALA and negative loading of LA, while PC2 had positive loading of SA, OA and LA and negative loading of ALA. These loadings separated SA and OA in the 1st quadrant, LA in the 2nd quadrant and ALA in the 4th quadrant. In the score plot of principal components, all the nine varieties could be clearly separated in to two groups as high and low ALA groups, while the mutant variety TL23 lay in the 2nd quadrant away from both the groups. This separation was mainly across PC2 with low ALA group (Ayogi, ES44, JRF5 and Acc No.4/47) having positive scores, while the high ALA group (NL260, EC9825, Padmini, NL97 and Surabhi) had negative scores. When the score plot and loading plot were compared, it was observed that OA and SA together separated the low ALA group in the 1st quadrant. On the contrary, ALA was the contributing FA for the presence of high ALA varieties mainly in the 4th quadrant. TL23 was placed in the 2nd quadrant because of the very high LA content.

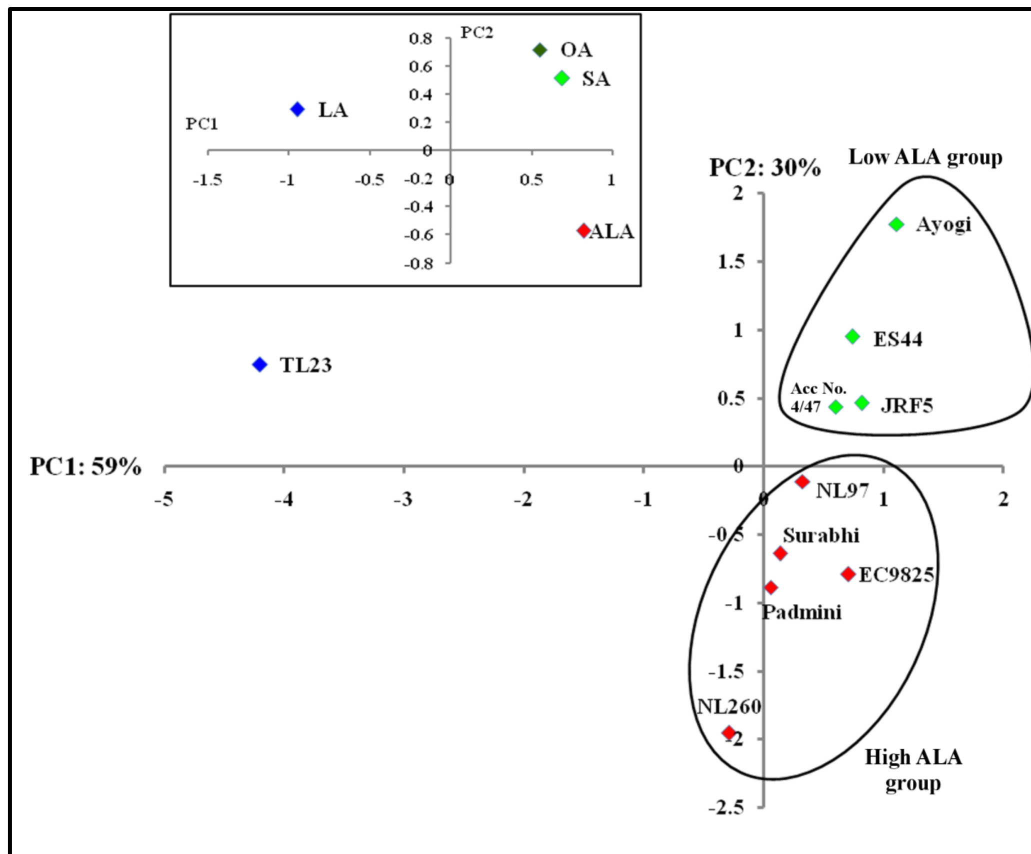


Figure 3.5: PCA plots. Score plot and loading plot (inset) of Principal Component Analysis (PCA) of four fatty acids *viz*., Stearic (SA), Oleic (OA), Linoleic (LA) and Linolenic (ALA) acids in the mature seeds of the ten flax varieties

3.3.2 Desaturase gene expression in developing seeds

The expression of *SAD* (*SAD1* and *SAD2*; Figures 3.6), *FAD2* (*FAD2* and *FAD2-2*; Figures 3.8) and *FAD3* (*FAD3A* and *FAD3B*; Figures 3.10) desaturase genes was quantified by real-time PCR at eight stages of seed development from flower to 48 DAA in the same ten flax varieties grown under identical environmental conditions. A two step reverse transcription method was used for relative quantification of the gene expression as detailed in the Materials and Methods section. Out of the three genes initially screened as reference genes, *ETIF5α* was selected for the further study, based on its stability of expression across all the genotypes and developmental stages. Expression level of a target gene for each developmental stage of each variety was calculated relative to the transcript accumulation of the *ETIF5α* reference gene. Amplification specificity of all the desaturase gene-specific and reference gene-specific primers was confirmed by observing a single dissociation curve for each pair of the primers. Similarly, it was ensured that the amplification

efficiencies of the target and reference genes were similar (Livak and Schmittgen, 2001) and close to 2.0 so as to use the ΔCt method to express the results of the target genes relative to the reference gene. We used the ΔCt method of calculation as it allowed relative quantification comparisons across ten different varieties, eight developmental stages and six genes (two genes each for three desaturases).

3.3.2.1 SAD gene expression

Expression of the *SAD1* and *SAD2* genes showed a definite pattern in the ten varieties and eight developmental stages in comparison to the reference gene *ETIF5α* (Figure 3.6A-D). In general, *SAD1* expression was low at the flower and 4 DAA stages (0.003-0.051-fold) and increased from 8 DAA onwards. It reached its maxima at 12 to 30 DAA (0.015-0.371) and decreased to a level as in flower or even less as the seeds reached maturity (Figures 3.6A & 3.6B). There was an average seven fold increase in the relative expression from minimum to maximum expressing stage with Padmini and TL23 showing the highest fold increase (16 and 11 fold, respectively). The range of expression of *SAD1* gene in groups of high and low ALA varieties and TL23 at their maximum expression stage is depicted in Figure 3.6A, while the average relative expression maxima of *SAD1* in the same groups are given in Table 3.4, indicating nearly threefold higher *SAD1* expression in the group of high ALA varieties.

The *SAD2* expression pattern was quite similar to *SAD1* expression with initial low expression in flower and 4 DAA (0.013-0.038 fold) (Figures 3.6C & 3.6D). From 8 DAA, the expression increased, reached its maximum (0.073-0.247) at 12-22 DAA and later decreased with maturity in the 10 varieties. There was an average 7.5 fold increase in the relative expression from minimum to maximum expressing stages with ES44 and TL23 showing maximum (12 and 11 fold, respectively) fold variation. However, the range of expression of *SAD2* gene as depicted in Figure 3.6B and average relative expression maxima as given in Table 3.4 for groups with high and low ALA varieties indicated only marginally higher *SAD2* gene expression in high ALA varieties. TL23, on the contrary, showed the highest expression among the groups. In general, the expression of *SAD2* gene in all the developing seed stages was higher in the ten varieties as compared to *SAD1* gene expression in the respective stages of respective varieties with the exception of Padmini, EC9825 and JRF5. In case of JRF5, *SAD2* expression was higher than that of *SAD1* expression only at 22 DAA stage.

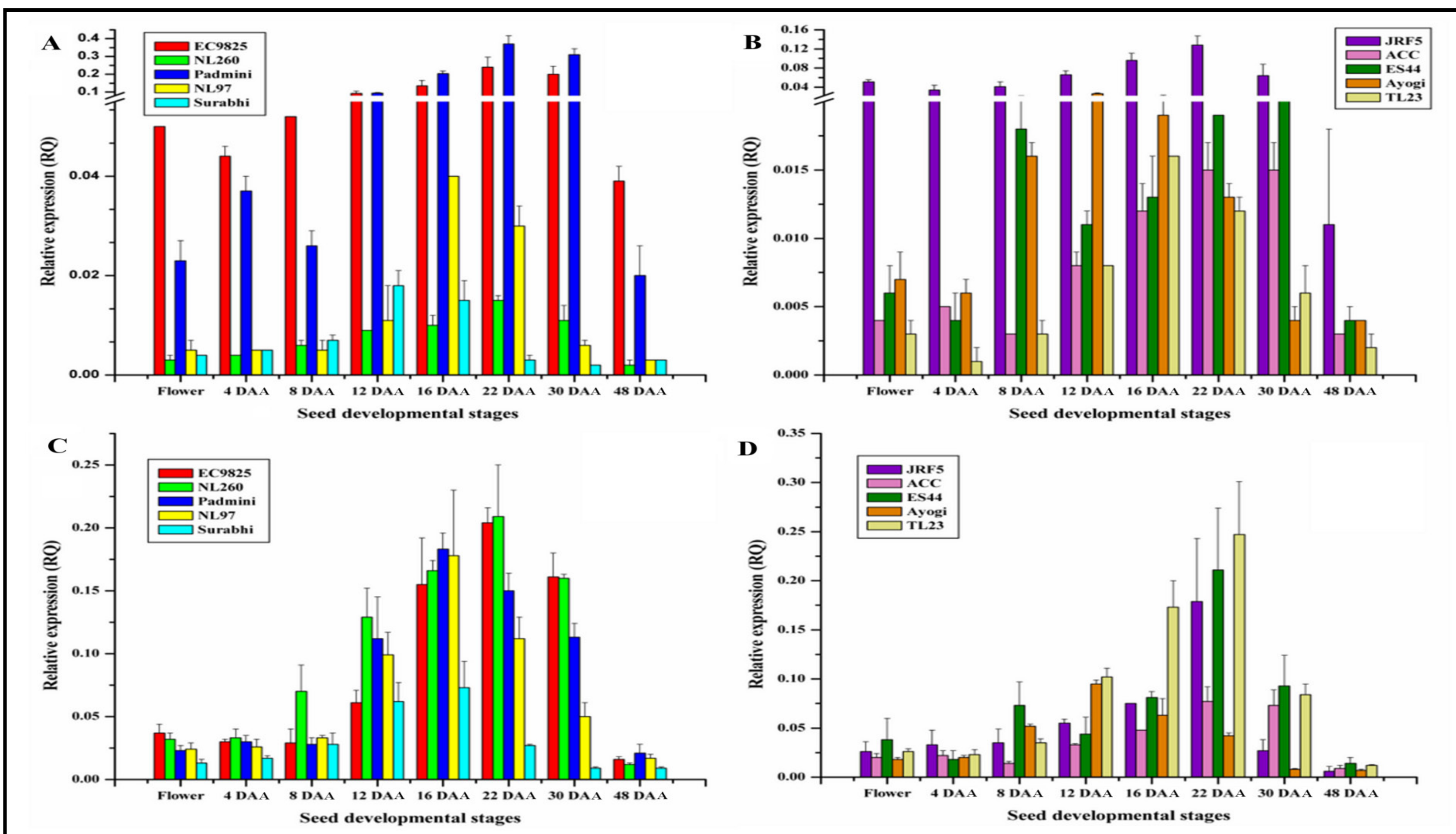


Figure 3.6: *SAD* gene expression. Relative expression of *SAD* genes in seed developmental stages of high and low ALA flax variety groups, **(A)** *SAD1* gene expression in high ALA group, **(B)** *SAD1* gene expression in low ALA group, **(C)** *SAD2* gene expression in high ALA group, **(D)** *SAD2* gene expression in low ALA group

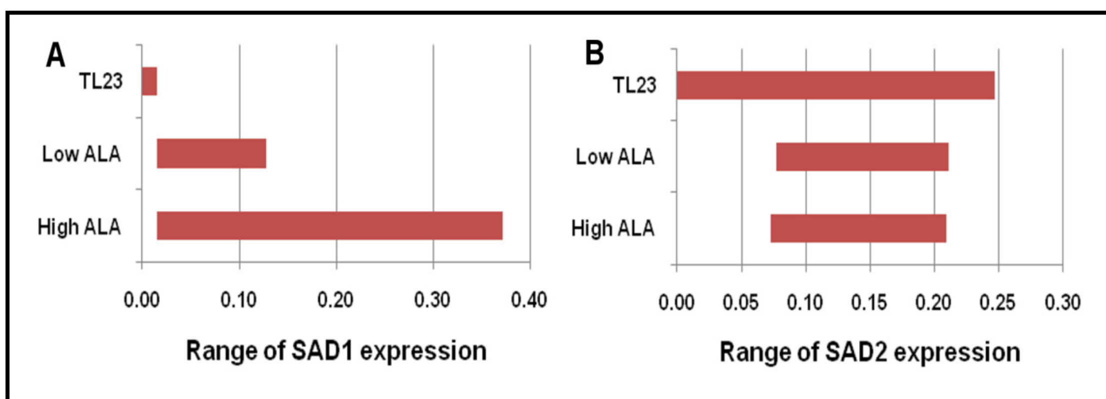


Figure 3.7: Expression range of *SAD* genes in high and low ALA groups and TL23. **(A)** Expression range of *SAD1*, **(B)** Expression range of *SAD2*

Table 3.4: Relative expression (RQ) maxima of *SAD* genes. Average RQ maxima (in folds) of SAD genes in high and low ALA groups and TL23

Gene	High ALA group	Low ALA group	TL23 (mutant variety)
SAD1	0.137±0.072	0.047±0.027	0.016±0.005
SAD2	0.169±0.025	0.141±0.032	0.247±0.094

3.3.2.2 FAD2 gene expression

The *FAD2* gene transcripts in the ten flax varieties were very low compared to the *ETIF5α* gene at flower and 4 DAA stage (Figures 3.8A & B). However, after this stage, the expression slowly increased to reach the peak levels (0.103-3.063 fold) by 16 to 30 DAA (in majority of the varieties at 22 DAA). Notably, there was an average 137 fold increase in the relative expression from minimum to maximum expressing stage with Padmini showing the highest (365 fold) fold increase. Further, Padmini and EC9825 were the only varieties, which showed higher transcript abundance relative to *ETIF5α* at their maximum expressing stage (22 DAA) and the increase was nearly 3 and 2 folds, respectively. It was also observed that in most of the varieties, the increase and later the decrease in the expression were sharp compared to the expression of other desaturase genes. The range of expression at their maximum expressing stage (Figure 3.8A) and average relative expression maxima (Table 3.5) of *FAD2* gene in groups of high and low ALA varieties indicated nearly 3.5 fold higher expression in high ALA group than in low ALA group varieties.

FAD2-2 gene transcripts also showed a similar pattern as *FAD2* gene. At the flower stage, the *FAD2-2* expression was 0.042-0.272 fold compared to the reference gene (Figures 3.8C & D). The expression marginally declined at 4 DAA and slowly increased from 8 to 22 DAA (0.15-0.73 fold) in most of the varieties. There was an average 5 fold increase in the relative expression from minimum to maximum expressing stages with ES44 showing the highest fold increase (9 fold). The range of expression at their maximum expressing stage (Figure 3.8B) and average relative expression maxima (Table 3.5) of *FAD2-2* gene in groups of high and low ALA varieties and TL23 indicated nearly two fold higher expression in high ALA group than in low ALA group varieties; while TL23 showed intermediate level of relative expression. Further, from the Table 3.5, it can also be observed that *FAD2* gene showed more than two fold higher relative expression compared to the *FAD2-2* gene in high ALA group; whereas in the low ALA group, the expression was higher by 1.3 folds.

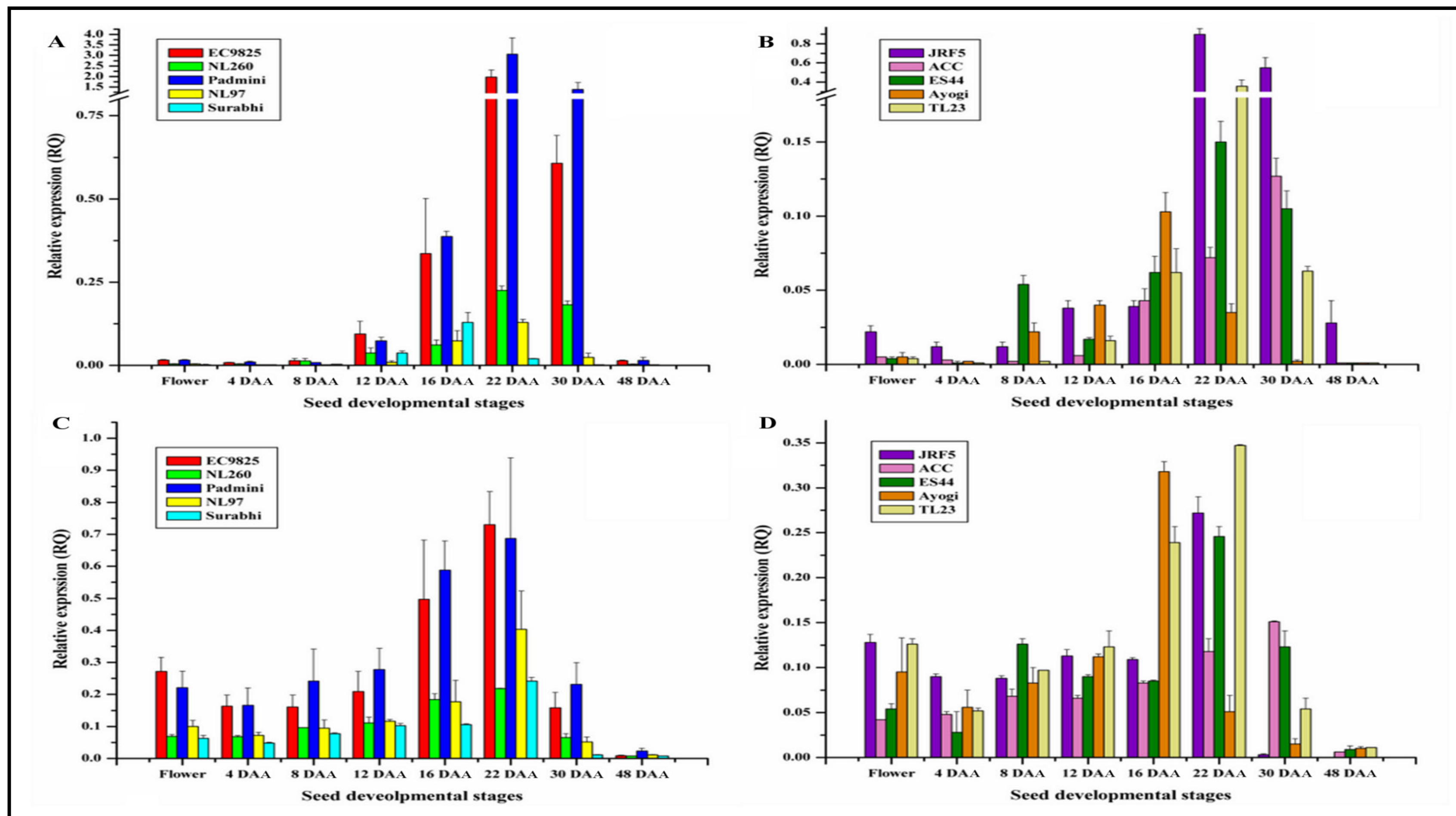


Figure 3.8: *FAD2* gene expression. Relative expression of *FAD2* genes in seed developmental stages of high and low ALA flax variety groups, **(A)** *FAD2* gene expression in high ALA group, **(B)** *FAD2* gene expression in low ALA group , **(C)** *FAD2-2* gene expression in high ALA group, **(D)** *FAD2-2* gene expression in low ALA group

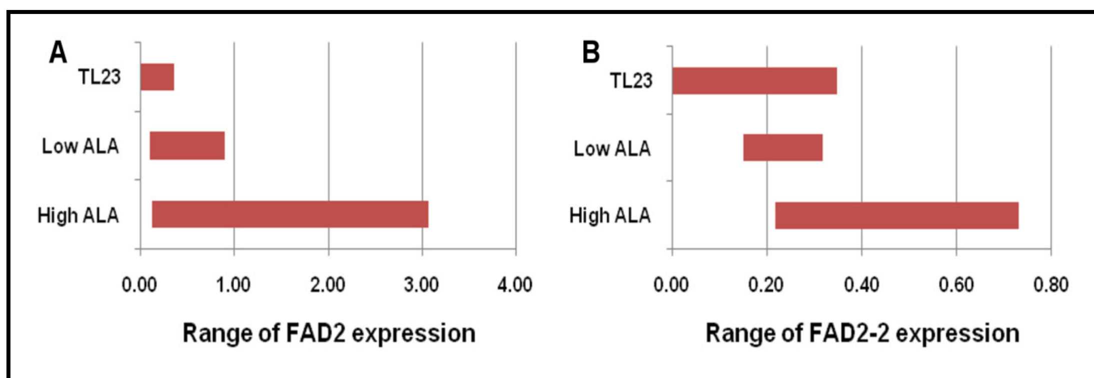


Figure 3.9: Expression range of *FAD2* genes in high and low ALA groups and TL23. **(A)** Expression range of *FAD2*, **(B)** Expression range of *FAD2-2*

Table 3.5: Relative expression (RQ) maxima of *FAD2* genes. Average RQ maxima (in folds) of *FAD2* genes in high and low ALA groups and TL23

Gene	High ALA group	Low ALA group	TL23 (mutant variety)
FAD2	1.104±0.603	0.319±0.172	0.356±0.066
FAD2-2	0.456±0.108	0.247±0.035	0.347±0.001

3.3.2.3 FAD3 gene expression

Like the earlier *SAD* and *FAD2* desaturase genes, *FAD3* (*FAD3A* and *FAD3B*) genes also showed a definite expression pattern in eight developing seed stages of the 10 varieties, although there was much variation in their peak expression (Figure 3.10A-D). *FAD3A* transcripts were very low (0.002-0.02) relative to reference gene in flowers. The expression further declined and was negligible at 4-8 DAA. It then gradually increased till 16 to 30 DAA (in majority of the varieties at 22 DAA). There was an average 2000 fold increase in the relative expression from minimum to maximum expressing stage with Surabhi and Padmini showing the highest (nearly 4000 fold) increase (Figure 3.10A). Again, for *FAD3A* gene, Padmini and EC9825 were the only varieties for which the transcript abundance at the peak expressing stage (22 DAA) was 1.8 and 1.6 fold higher relative to *ETIF5α* gene transcripts, respectively. In most of the varieties, the increase and then the decrease in the relative expression with the maturity was gradual. TL23 with the lowest ALA content (<2%) amongst the varieties under study, showed the lowest overall relative expression of *FAD3A* in all the eight developmental stages with expression maxima of 0.06 fold at 22 DAA.

FAD3B expression was the lowest at 4 DAA and gradually increased till 16 to 22 DAA and later decreased up to maturity in both, high and low ALA groups (Figures 3.10C & D). There was an average 500 fold increase in the relative expression from minimum to maximum expressing stage with EC9825 showing the highest fold increase (nearly 1500 fold). EC9825 (8.5 fold) and Padmini (7.7 fold) showed the highest transcript accumulation relative to *ETIF5α* gene at their expression maxima compared to the remaining varieties. Further, the range of expression of *FAD3A* and *FAD3B* genes at their maximum expressing stage (Figures 3.11A & B) and average relative expression maximum (Table 3.6) in high and low ALA varieties indicated nearly 5 and 3.5 fold higher expression, respectively. This observation was similar to the observation made for the expression of *FAD2* and *SAD* genes. However, the overall expression of *FAD3B* gene in all the developing stages was much higher in all the ten varieties as compared to *FAD3A* gene expression in the respective stages of the respective varieties.

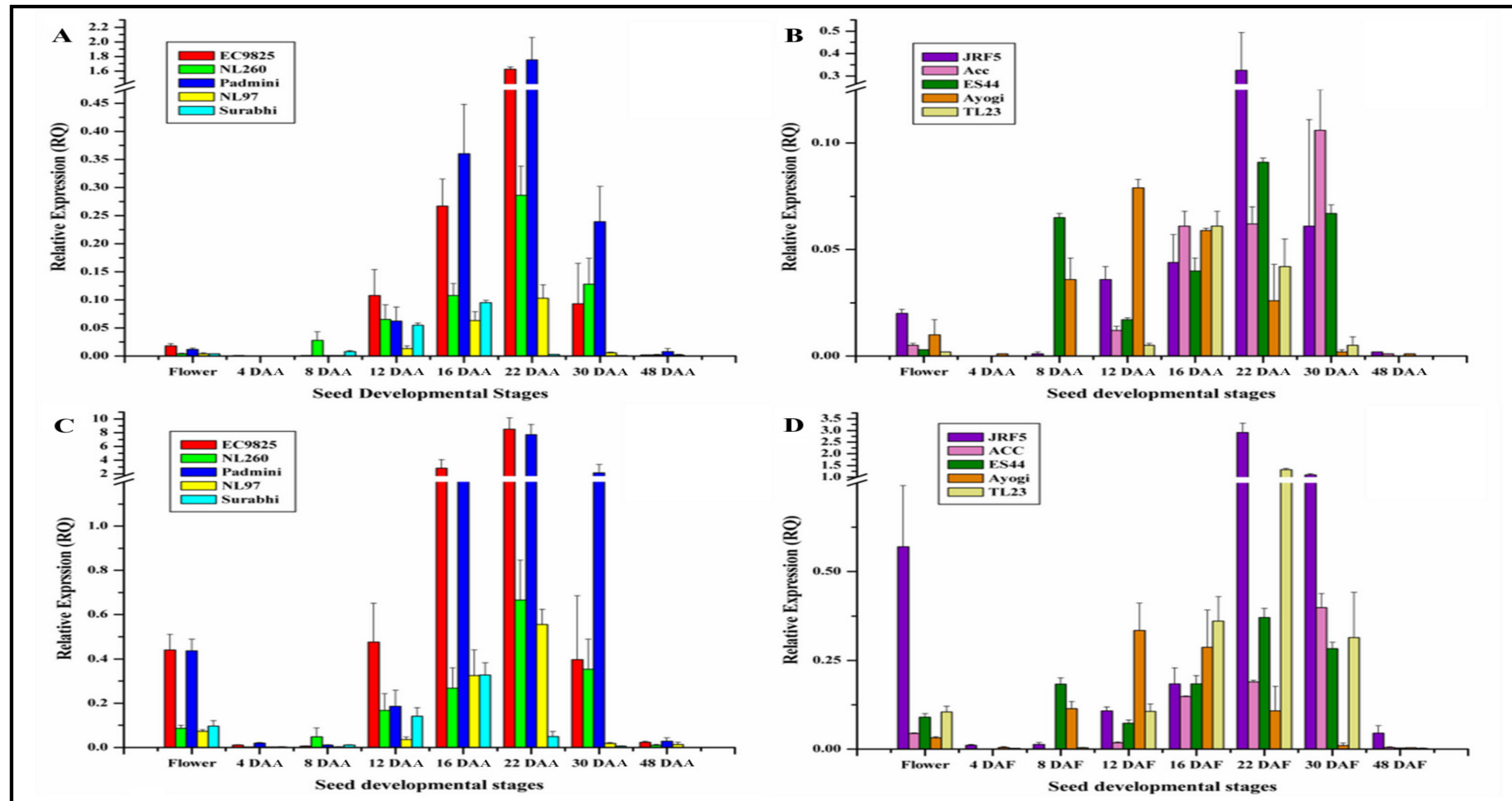


Figure 3.10: *FAD3* gene expression. Relative expression of *FAD3* genes in seed developmental stages of high and low ALA flax variety groups, **(A)** *FAD3A* gene expression in high ALA group, **(B)** *FAD3A* gene expression in low ALA group, **(C)** *FAD3B* gene expression in high ALA group, **(D)** *FAD3B* gene expression in low ALA group

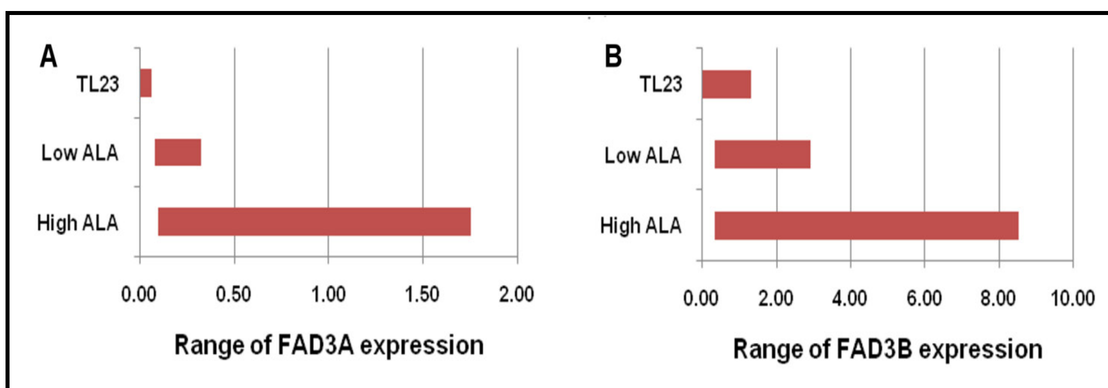


Figure 3.11: Expression range of *FAD3* genes in high and low ALA groups and TL23. **(A)** Expression range of *FAD3A*, **(B)** Expression range of *FAD3B*

Table 3.6: Relative expression (RQ) maxima of *FAD3* genes. Average RQ maxima (in folds) of *FAD3* genes in high and low ALA groups and TL23

Gene	High ALA group	Low ALA group	TL23 (mutant variety)
FAD3A	0.773±0.376	0.151±0.059	0.061±0.007
FAD3B	3.561±1.869	1.004±0.636	1.321±0.045

3.4 Discussion

Flax oil is one of the richest agricultural sources of ALA. It is known that sequential action of three microsomal fatty acid desaturases namely, *SAD*, *FAD2* and *FAD3*, is responsible for the desaturation of stearic to ALA through oleic and linoleic acid in the seeds (Fulco, 1974; Heinz, 1993). Variation in the content of these fatty acids in the seed can, therefore, be a result of differential activity of one or more desaturase enzymes in this chain. However, except for the *FAD3* gene (Banik et al., 2011), transcriptional correlation of these desaturases with ALA accumulation has not been studied in detail in flax varieties with differential ALA accumulation. In the present study, we selected ten Indian varieties of flax with varied ALA content at maturity (Chapter 2) and profiled their FAs contents from various seed development stages. Transcript profiling for all the microsomal desaturase genes (*SAD1* and *SAD2, FAD2* and *FAD2-2, FAD3A* and *FAD3B*) was performed in the same stages of seed

development of the ten varieties, so as to correlate the variation in ALA accumulation and the expression pattern of each desaturase gene in these varieties.

3.4.1 Expression dynamics of three desaturase genes in flax

Real-time PCR is considered as an accurate method for quantifying gene expression and is being routinely used for temporal and spatial expression profiling of genes (Charrier et al., 2002; Czechowski et al., 2004; Jain et al., 2007). However, accuracy in measurement of the transcript abundance depends on several factors such as RNA quality, cDNA quality and quantity, primer specificity and most important of all is the selection of a uniformly expressing stable reference gene for normalizing the transcript levels of the target genes (Gutierrez et al., 2006; Udvardi et al., 2008).

In flax, Huis et al. (2010) evaluated expression stabilities of 13 commonly used reference genes in 13 tissues of three developmental stages by computer algorithms geNorm and NormFinder. For all the tissues, geNorm identified three most stable genes as *EF1α*, *ETIF5α* and *UBI*, whereas NormFinder predicted *GAPDH* as the stable gene. In our study, initially we evaluated the efficiency and stability of *EF1α*, *ETIF5α* and *GAPDH* genes in the developing seed stages of the ten flax varieties. Both *EF1α* and *ETIF5α* showed stable expression across varieties and seed developmental stages. We used *ETIF5α* as the reference gene for further studies as the amplification efficiency of this gene was equivalent to the six target genes in order to use $2^{-\Delta ct}$ method for real-time PCR data analysis (Livak and Schmittgen, 2001; Pfaffl, 2005). This method gave us an advantage to make comparisons in the level of gene expression across developmental stages, varieties and genes.

The *SAD* gene is responsible for converting stearoyl-ACP to oleoyl-ACP, by introducing a double bond at C9 and thus, can increase the unsaturated fatty acid content of the plant (Ohlrogge and Jaworski, 1997). This gene is of commercial interest for the manipulation of unsaturated fatty acids in major crop plants (Knutzon et al., 1991) and has been well characterized (Shanklin and Somerville, 1991; Singh et al., 1994). In flax, the *SAD1* gene expression in developing seeds has already been reported (Fofana et al., 2006). We studied the expression of both, *SAD1* and *SAD2*, genes in the developing seeds and observed higher transcript levels of the *SAD2* gene compared to the *SAD1* gene in seven flax varieties and in differential ALA containing variety groups under consideration (Figure 3.5, Table 3.4). However, there was

exception in case of three varieties namely, EC9825, Padmini and JRF5, which showed marginally higher expression of the *SAD1* gene than the *SAD2* gene (Figure 3.5). TL23 showed the highest expression level of the *SAD2* isoform amongst the ten varieties (Table 3.4).

Overall expression of the *SAD2* gene appeared to be constitutive, whereas the *SAD1* gene appeared to be more temporal and genotype specific in expression as indicated by differential expression in low and high ALA containing groups (Figure 3.5). This observation can also be supported with studies of Jain et al. (1999) on flax *SAD1* and *SAD2* gene promoter activity in transgenic flax. The *SAD1* promoter was reported to be weaker and spatial, whereas *SAD2* promoter was stronger and constitutive in their analysis. The weaker activity of the *SAD1* promoter was suggested to be because of 368-bp region in the *SAD1* promoter containing negative regulatory elements and tissue specificity related elements.

The Δ12 desaturase/ FAD2 represents a diverse gene family in plants and is responsible for conversion of OA (18:1) to LA (18:2). LA production marks the synthesis of polyunsaturated fatty acid (PUFA) from monounsaturated OA and is a major factor in determining the quality of plant oils. Δ12 desaturation is also necessary for the synthesis of ALA (18:3) as LA, the product of this desaturation, can alone be accepted as the substrate for Δ15 desaturase (FAD3). Fofana et al. (2006) reported the seed specific *FAD2* (partially sequenced) expression in flax for the first time. However, two *FAD2* genes, namely *FAD2* and *FAD2-2* were later reported by Krasowska et al. (2007) and Khadake et al. (2009), respectively.

In this study, gene specific expression of *FAD2* and the *FAD2-2* in flax developing seeds was analyzed for the first time. *FAD2* expression increased sharply up to 22 DAA (in most of the varieties) to a level much above than at the flower stage and slowly decreased as the bolls matured, showing a bell shaped pattern. These findings were quite similar to the earlier studies in flax by Fofana et al. (2006) as well as in Arabidopsis by Ruuska et al. (2002). In case of *FAD2-2*, the expression pattern was similar but the rise in the transcript level was gradual till 22 DAA (in most of the varieties) and there was a sudden drop in the transcript abundance as the bolls matured.

Two *FAD2* genes namely, *FAD2* and *FAD2-2* have also been reported to express in developing seeds of soybean, wherein *FAD2-2* is constitutively expressed while *FAD2* is seed specific in its expression (Heppard et al., 1996). Similar to the soybean *FAD2-2* gene, it has also been observed in our studies that the *FAD2-2* gene expression was higher at the early seed development stages (flower to 12 DAA) than the *FAD2* expression. However, the *FAD2* transcripts increased rapidly and reached its peak at mid-maturation stages (16-22 DAA), and then gradually reduced as the seeds attained maturity. The transcripts for the *FAD2* increased significantly during embryo development, while the expression levels of the *FAD2-2* appeared to increase slightly. Further, it was observed that the average relative expression of the *FAD2* gene (0.72 ±0.32) was nearly twofold higher than the *FAD2-2* gene (0.36 ±0.06) in developing seeds of the ten flax varieties. All these observations are indicative of the predominance of the *FAD2* gene expression in determining the LA level of the seed storage oil. In case of TL23, the expression of both the genes was nearly the same (Table 3.5).

Δ15/FAD3 is the last desaturase required in the series of desaturation reactions leading to synthesis of ALA in plants. This gene has been shown to play a very important role in deciding the level of ALA in the total FA content of the plant (Yadav et al., 1993; Bilyeu et al., 2003; Vrinten et al., 2005; Banik et al., 2011). In flax, Vrinten et al. (2005) reported the presence of two *FAD3* genes, *FAD3A* & *FAD3B*, while later; Banik et al. (2011) reported the presence of a third less significant *FAD3C* gene. In our study, we analyzed the expression of only *FAD3A* and *FAD3B* genes, which are thought to play major roles in ALA accumulation compared to *FAD3C* (Banik et al., 2011). The gene expression pattern of *FAD3A* and the *FAD3B* in developing seed stages of the ten genotypes (divided in two groups according to ALA content) clearly showed higher *FAD3* (both *FAD3A* and *FAD3B*) gene expression in high ALA group than the expression in low ALA group (Table 3.6). Further, it was observed that the average expression maxima of the *FAD3B* (2.31±1.0) was nearly fivefold higher than average expression maxima of the *FAD3A* (0.45±0.2). These results were in accordance with the expression observed for these genes by Banik et al. (2011). They observed nearly threefold higher expression of *FAD3B* as compared to *FAD3A*. In the present study, even though the higher contribution in ALA accumulation was due to *FAD3B* than *FAD3A*, both the genes

contributed in the final differential accumulation of ALA content in mature seeds of high and low ALA groups with higher fold expression difference in the *FAD3A* than the *FAD3B* when high and low ALA containing groups were compared (Table 3.6).

3.4.2 Transcriptional correlation of desaturase genes with ALA content

The results of the expression analysis correlated well with the FA data for the groups based on the ALA content; where high ALA group showed higher expression of all the desaturase genes compared to that in the low ALA group; although, there were a few exceptions observed in individual varieties belonging to these groups. The FA values (Table 3.3) clearly indicated much efficient conversion of all the intermediate FAs, namely SA, OA and LA to the final ALA content due to efficient action of all the three desaturases acting sequentially in case of high ALA group. In case of low ALA group, even though the initial conversion up to OA by the *SAD2* was quite efficient, the later conversions up to ALA by consecutive action of the *FAD2* and *FAD3* enzymes seemed to be inefficient leading to higher accumulation of OA as well as LA instead of ALA in this group. However, differential expression of the *SAD1* in low and high ALA containing groups might be partly responsible for high and low ALA accumulation in their respective groups in addition to inefficient *FAD2* and *FAD3* activities in the low ALA group. These observations lead us to analyze the promoter region of all the three desaturase genes in high and low ALA containing flax groups to understand transcriptional regulation of expression of these genes. There are many earlier reports in various plant systems where promoter region analysis of a gene has revealed variation in transcription binding sites and corresponding variation in the transcription factors, leading to expression variation of the gene (de Meaux et al., 2005; McKhann et al., 2008; Yi et al., 2010).

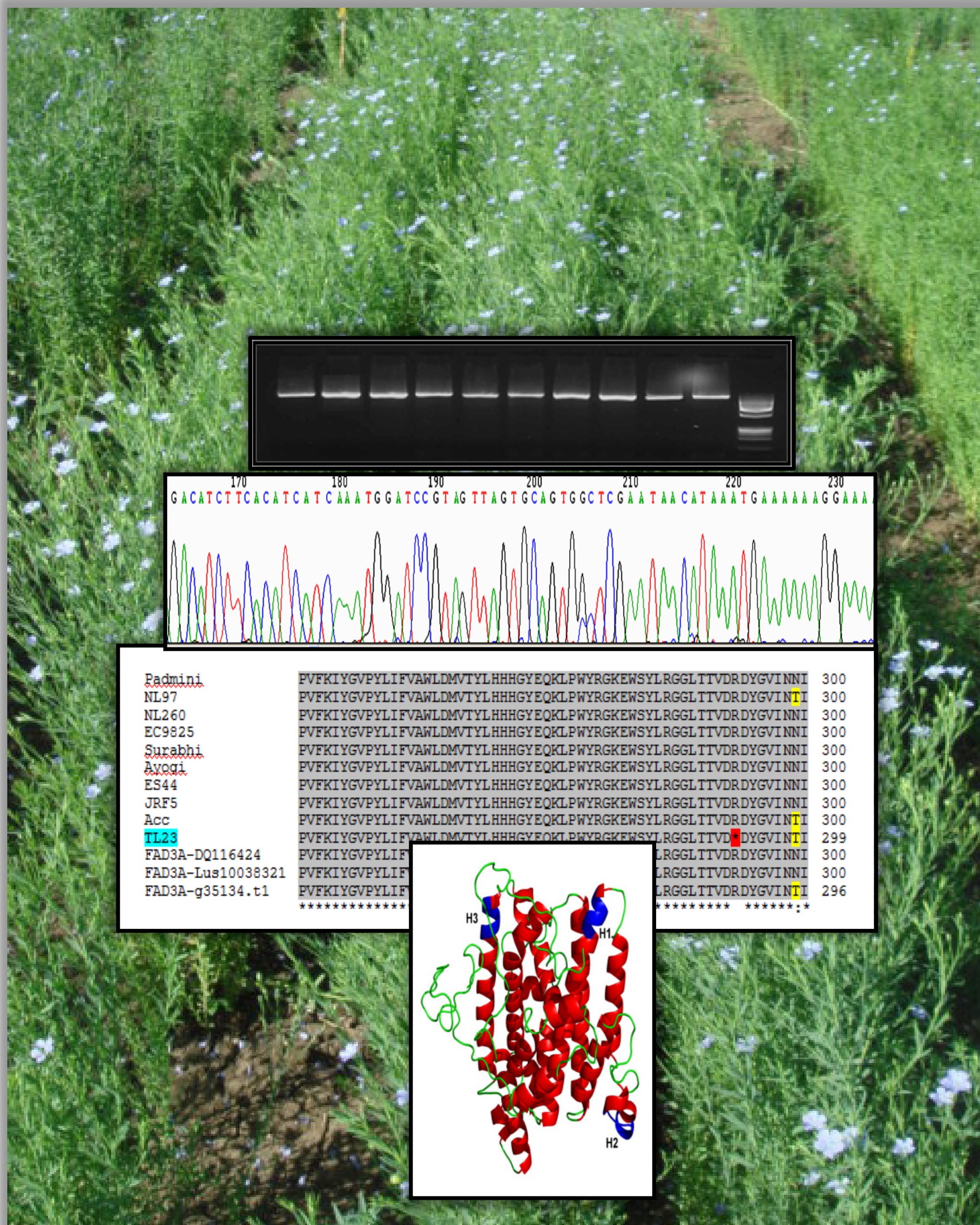
In case of NL260 and Padmini (belonging to high ALA group), we observed that NL260 showed the highest ALA accumulation among the ten flax varieties (Figures 3.3 & 3.4) but moderate transcript abundance, while Padmini revealed the highest expression of all the desaturase genes but moderate ALA accumulation within the high ALA group (Figures 3.6, 3.8 & 3.10). Though JRF5 belonged to low ALA group, high expression for most of the desaturases was observed. These deviations pointed out that the desaturase activity was not solely under the transcriptional control and there could be variety specific post-transcriptional, translational and post-translational regulations involved in deciding the total protein turnover or enzyme

efficiency. In soybean, Tang et al. (2005) showed the role of post-translational regulatory mechanisms in modulating the FAD2-1 enzyme activity, which supported the observation in flax. Recently, O'Quin et al. (2010) reported the post-translational regulation of plant FAD3 desaturases via ER associated degradation pathway as a response to temperature variation. This, however, is a physiological response, and it is not clear as yet if a similar way of controlling desaturases exist in plant seeds leading to synthesis of storage oils.

In conclusion, desaturation is an important biochemical process in the FA biosynthesis pathway mediated by FA desaturases. To decipher the contribution of different desaturases to the final levels of ALA, we performed fatty acid profiling and transcript profiling of six microsomal desaturase genes (*SAD1*, *SAD2*, *FAD2*, *FAD2-2*, *FAD3A* and *FAD3B*) from ten Indian flax varieties. We observed high overall correlation in expressions of these desaturase genes with the variation in the ALA accumulation in high and low ALA variety groups. Thus, the accumulation of the final product, ALA, was not entirely dependent on the activity of a single desaturase gene, but was a cumulative result of the activities of all the three desaturases acting sequentially.

CHAPTER 4

Sequence characterization and *in silico* structure prediction of desaturases from the flax varieties varying in α-linolenic acid content



4.1 Introduction

It has been observed in a number of studies in plant systems, especially in oil seeds like soybean, groundnut, canola etc., that certain changes in the nucleotide sequences of desaturases affected the accumulation of resultant fatty acid in the storage oil (Rahman et al., 1996; Byrum et al., 1997; Pantalone et al., 1997; Rahman and Takagi, 1997; Bilyeu et al., 2003; Patel et al., 2004; Hu et al., 2006; Belo et al., 2008; Pham et al., 2011). These allelic variants can extensively be used for candidate gene based molecular breeding, and have been used in soybean to develop genotypes with desired content of fatty acids in the oil (Hoshino et al., 2010; Pham et al., 2010). Further, important single nucleotide polymorphism (SNP) sites were also used to develop molecular marker based genotyping assays in soybean (Beuselinck et al., 2006; Chappell and Bilyeu, 2007).

In case of flax, nucleotide sequences of two *SAD* genes (*SAD1* & *SAD2*), two *FAD2* genes (*FAD2* & *FAD2-2*) and three *FAD3* genes (*FAD3A, FAD3B* & *FAD3C*) have been reported (Singh et al., 1994; Jain et al., 1999; Vrinten et al., 2005; Krasowska et al., 2007; Khadake et al., 2009; Banik et al., 2011). However, no studies have been performed to correlate the genic variants with differential fatty acid content in case of *SAD* and *FAD2* genes. While in case of *FAD3* genes, there are two reports from Canadian flax varieties where mutations leading to stop codons in the *FAD3A* and *FAD3B* genes resulted in very low ALA (2-3%) accumulation in the mutant lines (Vrinten et al., 2005; Banik et al., 2011). Interestingly, phylogenetic analysis of *SAD2* gene for a number of cultivated flax genotypes belonging to different infraspecific groups as well as pale flax, a wild progenitor of modern domesticated flax, was performed to understand the domestication process (Allaby et al., 2005; Fu et al., 2012). In this study, *SAD2* served as a candidate domestication locus associated with increased unsaturated fatty acid production in cultivated flax and lead to the conclusion that the flax domestication might have involved multiple paths but at least initially, it was domesticated for oil.

The amino acid sequence analysis of desaturases has shown two characteristic D/EXXH motifs involved in binding the di-iron complex in case of soluble acyl–acyl carrier protein (ACP) desaturases in plants, while three histidine boxes $H(X)_{3-4}H$, $H(X)_{2-3}HH$ and $H/Q(X)_{2-3}HH$ are required for binding the di-iron complex in membrane bound desaturases. Stearoyl-ACP desaturase (SAD) belongs to the former

class, while FAD2 and FAD3 belong to the later class (Shanklin and Cahoon, 1998). Hydropathy plots of membrane bound desaturases further reveal 4–6 membrane spanning helices, which account for nearly 30% of the amino acid sequence of these proteins. Although a significant diversity has been reported in sequences and expression patterns of the desaturases (Los and Murata, 1998), changes in the conserved His-boxes potentially alter their structure and may result in altered substrate specificity, regioselectivity and even loss of functionality (Shanklin et al., 1994; Avelange-Macherel et al., 1995).

The availability of crystal structures for acyl-ACP desaturases from two plant species, *Ricinus communis* (castor) and *Hedera helix* (English ivy) (Lindqvist et al., 1996; Whittle et al., 2005; Guy et al., 2007) can help in detailed structural and functional prediction for this enzyme from other plant species. Whereas for membrane bound desaturases, the crystal structures are not yet available, probably because of the difficulty in solubilization and purification of these enzymes due to their hydrophobicity and strong anchorage to membranes. Hence, it is difficult to understand their functionality. Acyl-ACP desaturases are homodimeric proteins, with each monomer folded into a compact single domain composed of nine helices. The di-iron active site of these enzymes is buried within a core four-helix bundle and is positioned along a deep, bent, narrow hydrophobic cavity in which the substrate is bound during catalysis.

In the present study, we sequenced all the currently known desaturase genes, *SAD* (*SAD1* & *SAD2*), *FAD2* (*FAD2* & *FAD2-2*) and *FAD3* (*FAD3A* & *FAD3B*) from the ten flax varieties which were transcriptionally characterized in Chapter 3. Further, these gene sequences were analyzed for variation in the exonic or intronic regions of the genes and if any changes in the exonic region lead to change at the amino acid (AA) level. Lastly, we checked the structural variations in these genes due to AA variation, the functional implications of these changes and their correlation with the final ALA accumulation if any, in these varieties.

4.2 Materials and methods

4.2.1 Plant material

Leaf tissue was harvested from ten flax varieties grown at the College of Agriculture, Nagpur, Maharashtra, India. The collected tissue was immediately frozen in liquid nitrogen and later stored at -80°C freezer, till the DNA was extracted.

4.2.2 DNA extraction and PCR amplification

Genomic DNA was extracted from the young leaf tissue of all the 10 flax varieties using the modified CTAB method as described by Ghosh et al. (2009). This method was slightly different from the DNA extraction method (Richards et al. 1994) described in Chapter 2. In this modified method, 100 mg (instead of 1 g) of leaf tissue was ground to fine powder in liquid nitrogen and was quickly transferred in a polypropylene tube containing 10 ml of pre-warmed extraction buffer (100mM Tris–HCl pH 8, 10mM EDTA pH 8, 1.4M NaCl, 2% CTAB and 0.2% β-Mercaptoethanol). This extraction buffer did not include sodium metabisulphite which was used in the buffer of earlier DNA extraction method. The steps following this were the same as the previous protocol. Besides the change in the buffer composition, the overnight precipitation step using the precipitation buffer was eliminated and instead, nucleic acids were precipitated using two-thirds volume of chilled *iso*-propanol. These were air dried and resuspended in 1 ml of 1M NaCl instead of dissolving it in high salt Tris–EDTA (HSTE) buffer. This was followed by RNase treatment at 37°C for 1 hr and RNase contamination was removed by giving a phenol:chloroform (1:1) wash followed by chloroform:*iso*-amylalcohol (24:1) wash. Further DNA precipitation and 70% alcohol wash steps were the same as described earlier. The quality of DNA obtained was suitable for gene specific amplification and DNA sequencing reactions.

Desaturase gene specific primers were designed based on the cDNA (for *FAD3*: DQ116424, DQ116425) or full length genomic sequences (for *SAD*: AJ006957, AJ006958 and *FAD2*: DQ222824, EU660501) available in the NCBI database. Two to five overlapping primers spanning the entire gene sequence were designed for each gene (Table 4.1) using the Primer3 software (Rozen and Skaletsky, 2000).

Amplification of 30-50 ng DNA was performed in 15 μl reaction volume containing 10× PCR buffer (10mM Tris HCl, pH 8.3; 50mM KCl, 1.5mM $MgCl_2$),

0.15mM dNTPs (Amersham Biosciences, USA), 0.3-0.5μM primer (varied according to the primer) and 0.6U *Taq* DNA polymerase (Bangalore Genei, India) using Veriti thermal cycler (Applied Biosystems, USA). A typical polymerase chain reaction consisted of initial denaturation at 94°C for 5 min, followed by 35 cycles, each comprising 1 min denaturation at 94°C, 45 sec annealing at 50°C to 55°C (varied according to the primer sequence) and 1 min extension at 72°C with 10 min final extension at 72°C at the end of 35 cycles. The amplification products were resolved along with Φ×174/*Hae*III marker (Bangalore Genei, India) on 1.5% agarose gel and electrophoresed in 0.5× TAE buffer. The gels were stained with GelRed^{TM} (Biotium, USA) and documented on Syngene Dyversity gel documentation system (Syngene, USA). PCR amplifications were repeated at least twice to establish reproducibility of the results.

4.2.3 Nucleotide sequencing and analysis of the fatty acid desaturase genes

The single band products with concentration, 75 to 100 ng/μl each, were used for bidirectional sequencing by dideoxy chain termination method using the DYEnamic ET DNA sequencing kit and the MegaBACE 1000 DNA Analysis System (GE Healthcare, USA). Each PCR product was sequenced at least twice and base calling was performed using three different basecallers. These sequences were blastn searched in the NCBI database to confirm their identity and the sequences with a confidence level of 90% and above were chosen for further analysis. Each gene was amplified and sequenced in multiple fragments using primer pairs detailed above. The sequences thus obtained were edited and then aligned using MEGA5 software (Tamura et al., 2011). Finally, the full-length gene sequences were assembled from these gene fragments. The sequence assembly was also validated using MIRA software (Chevreux et al., 2000). In all, six desaturase genes were sequenced from the ten flax varieties. Sequences of each of the six desaturase genes from the ten varieties were aligned with respective desaturase gene sequences from the NCBI and Phytozome databases (Goodstein et al., 2012). Amino acid sequences of all the six desaturases were deduced based on the gene sequence using MEGA5 software and important conserved motifs were determined in each case.

Table 4.1: Gene specific PCR primer sequences (5′→3′) used for amplification of *SAD*, *FAD2* and *FAD3* genes

Gene	Primer	Forward Primer Sequence	Reverse Primer Sequence
SAD1	LUSAD1 F1.2/R1	TTCCAATTTCCATTTCCTCATCTGC	AGCTTCTCAGCCTCCCTGTG
	LUSAD1 F2/R2	GACTAATGAGCTATGACATGATTACGA	CATCACGTGTATCCTTTCACAA
	LUSAD1 F3/R3	GTAATCACATACTTCATCCTTTTCT	TAAACCACCTACTTGTCCGATGG
	LUSAD1 F4/R4	TTAGCAGTTTGTCCTGGTATCTG	TCATACTCCAATCATTCCACACTCT
SAD2	LUSAD2 F1/R1	TTCCAACTTCCATCCTCATCTGCCTTAC	AACCAACAAACCACTGTCCCTCCG ACTC
	LUSAD2 F2/R2	TGTTTGGTGACAGGGAGGCTGAGAAG	ACCCAGGACAGAATAAAGTGATGAAGGA
	LUSAD2 F3/R3	GAACTGTTAAACAATACATGTGATGTC GTG	TCGGGCAAACTCCATTAGAGTACCAA
FAD2	LU FAD2 F1/R1	ATGGGTGCAGGTGGAAGAATGCCAGTG	TCATAACTTATTGTTGTACCAGAACACG
FAD2-2	LU FAD2-2 F1.1/R1.1	GCCGTGCCTCCATCGAATAAGGC	GTTGTAGATGGGGGATTTAGGAT
	LU FAD2-2 F2.1/R2.1	ATTGGCTGGCACTCAAAGTA	TCGACATAAACACACTCCTTAGC
FAD3A	LUFAD3A F1/R1	TTCAAAACTGTGGCTCTGCAC	GAGAGGGATGGTGAACCTCA
	LUFAD3A F2.1/R2	ATGAATCCTGGGTTCCACTG	CGTAGATTTTGAAGACCGGAAG
	LUFAD3A F2.1/R3	ATGAATCCTGGGTTCCACTG	CATTTGAGGGAAGAGATGGTGAATA
	LUFAD3A F4/R4	CGACATGGTGACCTACCTTCAC	AGAAAACGACGTCGCCTGTA
	LUFAD3A F5/R5.1	GTGCTGGGGAAGTACTACAGAGAA	TACTCAGAAAGAAAGCTGGGGT
FAD3B	LUFAD3B F1/R1.2	TTCAAAACTGTGGCTCTGCAG	AGGAATGGTGAACCTCATGAACTTA
	LUFAD3B F2/R2	ATGAATCCTGGGTTCCTCTA	TGGCTGTCCAGCACAATGTT
	LUFAD3B F3/R3	AAAGTGTACAAGAGCTTGGATACCA	AATCTCGATCAACGGTCGTC
	LUFAD3B F4/R4.1	CGACATGGTGACCTACCTTCAT	ACTGCTGGTACGGAACTTGG
	LUFAD3B F5/R5	TGCCACACTATCACCTTGTG	GCGGGAGAGACAGAGAAACA

4.2.4 Haplotype analysis

The gene sequences thus obtained were subjected to haplotype determination using DnaSP v5 software (Librado and Rozas, 2009). Various measures of sequence variation were obtained, such as number of variable sites, haplotype number, haplotype diversity H_d etc.

4.2.5 Three-dimensional protein modeling

Three dimensional (3D) models were predicted for all the three desaturase proteins from flax. For SAD (Δ9), homology modeling was performed using Modeller 9v7 (http://salilab.org/modeller/), where the reported crystal structure of Δ9 desaturase from castor (PDB ID: 1OQ4) was used as a template. Whereas, for FAD2 and FAD3 proteins, where the structures are not available, 3D models of the proteins were predicted using the I-TASSER server (http://zhang-lab.ccmb.med.umich.edu/I-TASSER/). The 3D models were thus obtained on the basis of multiple-threading alignments by LOMETS and iterative TASSER simulations, where it uses multiple PDB structures depending on its structural conservation, to model various parts of the proteins (Zhang, 2008; Roy et al., 2010). The models were viewed using PyMol software (The PyMol Molecular Graphics System, Version 1.2r3 pre, Schrödinger LLC).

4.3 Results

4.3.1 Molecular characterization of SAD, FAD2 and FAD3 genes of ten flax varieties

The gene fragments amplified for six desaturase genes namely, *SAD* (*SAD1* & *SAD2*), *FAD2* (*FAD2* & *FAD2-2*) and *FAD3* (*FAD3A* & *FAD3B*) in ten flax varieties were directly sequenced and the sequences were assembled to obtain full-length gene sequences for each gene. The gene sequences obtained in this study have been deposited in GenBank with the accession numbers JQ963139-JQ963148 (*SAD1*), JQ963149-JQ963158 (*SAD2*), JQ963109-JQ963118 (*FAD2*), JQ963159-JQ963168 (*FAD2-2*), JQ963119-JQ963128 (*FAD3A*) and JQ963129-JQ963138 (*FAD3B*) and are given in Annexure I.

4.3.1.1 SAD gene sequence analysis

The two *SAD* gene isoforms, *SAD1* and *SAD2* from the ten flax varieties were 2522 and 2523 bases long respectively, comprising three exons each and they were 7 and 4 bases shorter than the reported respective NCBI gene sequences (AJ006957, AJ006958). However, the open reading frame (ORF) for both the genes was 1191 bases in length, encoding 396 amino acids, similar to the reported NCBI sequences. For *SAD1* gene, as compared to its reported sequence in NCBI (AJ006957) there was a C267T transition in the 1st intron in all the varieties. Even in the 2nd intron, C to T transition was observed in all the varieties except Surabhi and Acc No. 4/47. Secondly, at the 2nd intron and 3rd exon junction, there was a loss of 7 bases (ATCCAAG) observed in all the ten varieties. However, identical sequence variations were also reported for the *SAD1* gene sequence of Phytozome database (Lus10027486). For *SAD2* gene, there were varietal variations observed in both intronic and exonic regions, when compared with the NCBI *SAD2* gene sequence (AJ006958) (Table 4.2). Most of the variations were present in the two introns. It was also observed that all the exonic variations did not result in AA variations (i.e. they were synonymous nucleotide substitutions), except for G2179A in TL23 and C2008T in all the ten varieties (Table 4.2). All the variations observed in the varieties were also observed in the *SAD2* gene sequence reported in Phytozome database (Lus10039241.g).

4.3.1.2 FAD2 gene sequence analysis

Both the *FAD2* genes, *FAD2* and *FAD2-2*, from all the ten flax varieties were intronless as reported in the earlier *FAD2* gene studies from flax (Krasowska et al. 2007; Khadake et al. 2009). The ORFs of the genes consisted of 1137 and 1149 bases, encoding proteins of 378 and 382 AA in length, respectively. The *FAD2* gene sequence from the ten flax varieties showed a few variations when compared with NCBI *FAD2* gene sequence (DQ222824). Most of the nucleotide variations were synonymous substitutions. The varietal variations for this gene are listed in Table 4.3.

Table 4.2: *SAD2* gene and amino acid (AA) variation in ten flax varieties

Intron/Exon	Nucleotide variation	AA variation	Variety	Remark
1st intron	A259C	-	JRF5 and TL23	Similar to sequence in Phytozome database (Lus10039241.g)
	A262G	-	JRF5 and TL23	Similar to Lus10039241.g
	T362C	-	All varieties	Similar to Lus10039241.g
	C629G	-	NL97,NL260, EC9825, Surabhi	New variation
2nd intron	A1241G	-	All varieties	Similar to Lus10039241.g
	C1260T	-	All varieties	Similar to Lus10039241.g
	TAA insertion at 1337	-	All varieties	Similar to Lus10039241.g
	T1435C	-	NL260, Ayogi, ES44 and TL23	Similar to Lus10039241.g
	T1457C	-	NL260, Ayogi, ES44 and TL23	Similar to Lus10039241.g
	C1662T	-	All varieties	Similar to Lus10039241.g
	A insertion at 1834	-	All varieties	Similar to Lus10039241.g
2nd intron and 3rd exon junction	ATCCAAG deletion	-	All varieties	Similar to Lus10039241.g
3rd exon	C2001T	P225S	All varieties	Similar to Lus10039241.g
	G2172A	G282S	TL23	Similar to Lus10039241.g
	C2225A	-	TL23	Similar to Lus10039241.g

In case of *FAD2-2* gene, we could obtain only partial sequences, from 106 bp to 1016 bp. When these partial gene sequences were compared, except for one nucleotide variation, C378T in variety ES44, all the remaining sequences from the rest of the varieties were identical to the NCBI *FAD2-2* sequence (EU660501.1). The variation in ES44 was observed to be a synonymous substitution.

Table 4.3: *FAD2* gene and amino acid (AA) variation in ten flax varieties

Nucleotide variation	AA variations	variety	Remark
C105T	-	TL23	Similar to another reported sequence EU660502.1 in NCBI (From variety NL97)
C117T	-	All varieties	Similar to EU660502.1
C126T	-	Surabhi	Similar to EU660502.1
C157T	H53Y	All varieties	Similar to EU660502.1
A350T	D117V	All varieties	Similar to EU660502.1
C624T	-	EC9823 and TL23	Similar to EU660502.1

4.3.1.3 FAD3 gene sequence analysis

Two *FAD3* genes, *FAD3A* and *FAD3B* analyzed in the present study, showed slight variation compared to the reported Phytozome (from variety CDC Bethun, Lus10038321.g) and NCBI gene sequences (from variety AC McDuff, HM991828.1) (Table 4.4 & 4.5). *FAD3A* gene from nine varieties except NL260 was 3208 bases in length, similar to the Phytozome gene sequence (Lus10038321.g), while two bases shorter than NCBI gene sequence (HM991828.1). In NL260, the *FAD3A* gene was 36 bases longer (3242 bases). *FAD3B* gene from all the varieties except NL260 and Padmini had a length of 2990 bases, which was one base longer than the reported NCBI (HM991832.1) and Phytozome (Lus10036184.g) gene sequence. While in case of NL260 and Padmini, the gene length was 2989 bp and similar to the reported gene. Both the genes consisted of six exons each and the ORF consisted of 1179 and 1176

bases, encoding proteins of 392 and 391 AAs, respectively, which is similar to the NCBI sequence for these genes (DQ116424, DQ116425) (Vrinten et al. 2005).

The Tables 4.4 and 4.5 summarize the variations observed in the intronic and exonic regions of these genes isolated from the ten varieties. It was interesting to note that, many of these variations in *FAD3A* gene sequence were observed in variety NL260, whereas in *FAD3B* gene sequence, NL260 and Padmini lacked most of the variations.

4.3.2 Amino acid sequence analysis for SAD, FAD2 and FAD3 among the ten varieties

The deduced amino acid (AA) sequence of *SAD1* gene was identical in the ten flax varieties. Even in case of *SAD2*, the deduced AA sequence including P225S substitution was similar among the ten varieties except for G282S substitution, which was observed only in TL23. Both these AA substitutions were similar to the *SAD2* sequence (Lus10039241.g) in the Phytozome database. AA sequences of both the desaturases also showed the presence of two consensus [(D/E) X2 H] domains, characteristic of soluble desaturases (Figure 4.1 & 4.2).

The deduced AA sequence of *FAD2* gene was identical in the ten flax varieties; however, two AA substitutions (H53V and D117V) were observed in all these varieties as compared to the reported *FAD2* sequence (DQ222824.1) in NCBI database. In case of FAD2-2 294 AA sequences representing 36th to 339th AA from the ten flax varieties were observed to be identical in all the ten varieties. All the sequences of FAD2 and FAD2-2 showed characteristic features of membrane-bound desaturases including the presence of three histidine boxes while, YNNKL motif was observed, only in FAD2 sequences, at the C terminus of the protein (Figures 4.3 & 4.4).

Table 4.4: *FAD3A* gene and amino acid (AA) variation in ten flax varieties

Intron/Exon	Nucleotide variation	AA variation	Variety
1st exon	C213T	-	NL260
	A229G	T77A	NL260
	G330T	-	NL260
	A362T	-	NL260
2nd exon	T429C	-	NL260
	C444T	-	NL260
4th intron	G864T	-	TL 23
	A1232C	-	Padmini
	C1578T	-	Ayogi
5th exon	G2226A	-	NL97, EC9825, Surabhi, JRF5, Acc No. 4/47, TL23
	C2263T	Stop codon at 292 AA	TL23
	A2285C	N299T	NL97, Acc No. 4/47, TL23
	A2352G	-	NL260
	A2416G	-	NL260
	T2420C	-	NL260
	A2427G	-	NL260
	G2431A	-	Acc No. 4/47
	From 2485Number of variations at different positions including additions, deletion and substitutions	-	NL260
6th exon	A3093C	H330P	NL260
	T 3277C	-	Padmini, NL97, NL260

Table 4.5: *FAD3B* gene and amino acid (AA) variation in ten flax varieties

Intron/Exon	Nucleotide variation	AA variation	Variety
1st exon	A225G	-	In all except Padmini, NL260 and TL23
	A366C	-	In all except Padmini, NL260 and TL23
2nd exon	T419 G	I112S	In all except Padmini and NL260
4th intron	C899T	-	In all except Padmini and NL260
	G907A	-	In all except Padmini and NL260
	T994C	-	In all except Padmini and NL260
	Additional T at 999 base	-	In all except Padmini and NL260
	T1088G	-	In all except Padmini and NL260
	A1104C	-	In all except Padmini and NL260
	T1124C	-	In all except Padmini and NL260
	G1234A	-	In all except Padmini and NL260
	C1413T	-	NL97, EC9825, Ayogi, ES44, Acc No. 4/47
	T1640A	-	EC9825, Acc No. 4/47
5th exon	A1739G	-	EC9825, Acc No. 4/47
	A1935C	-	EC9825
	G1963A	G283E	TL23
	C2629T	-	Acc No. 4/47

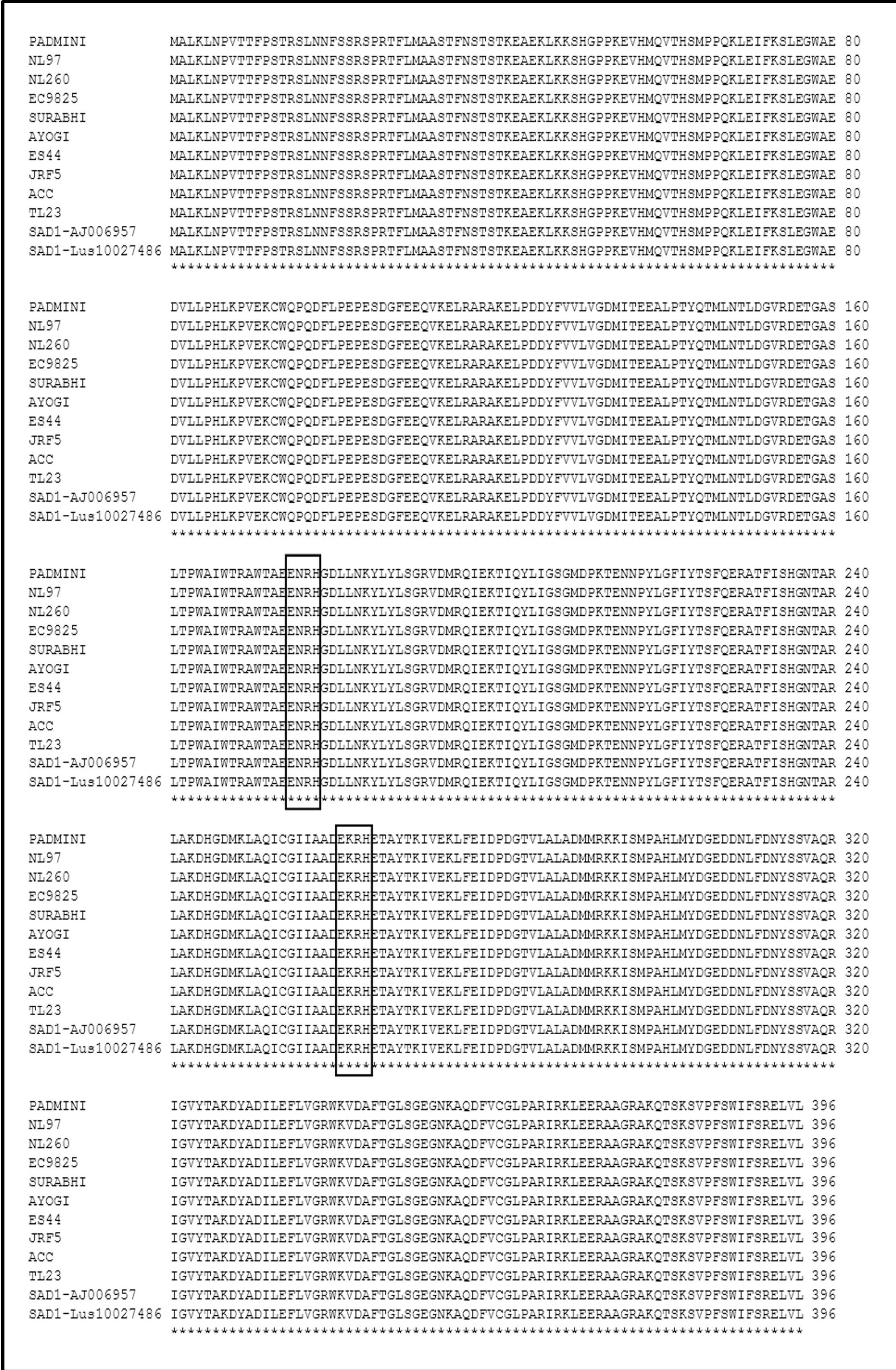


Figure 4.1: Alignment of deduced amino acid sequences of *SAD1* gene from ten flax varieties with *SAD1* sequence in NCBI and Phytozome databases. The boxes represent conserved di-iron binding motifs [(D/E) X2 H]

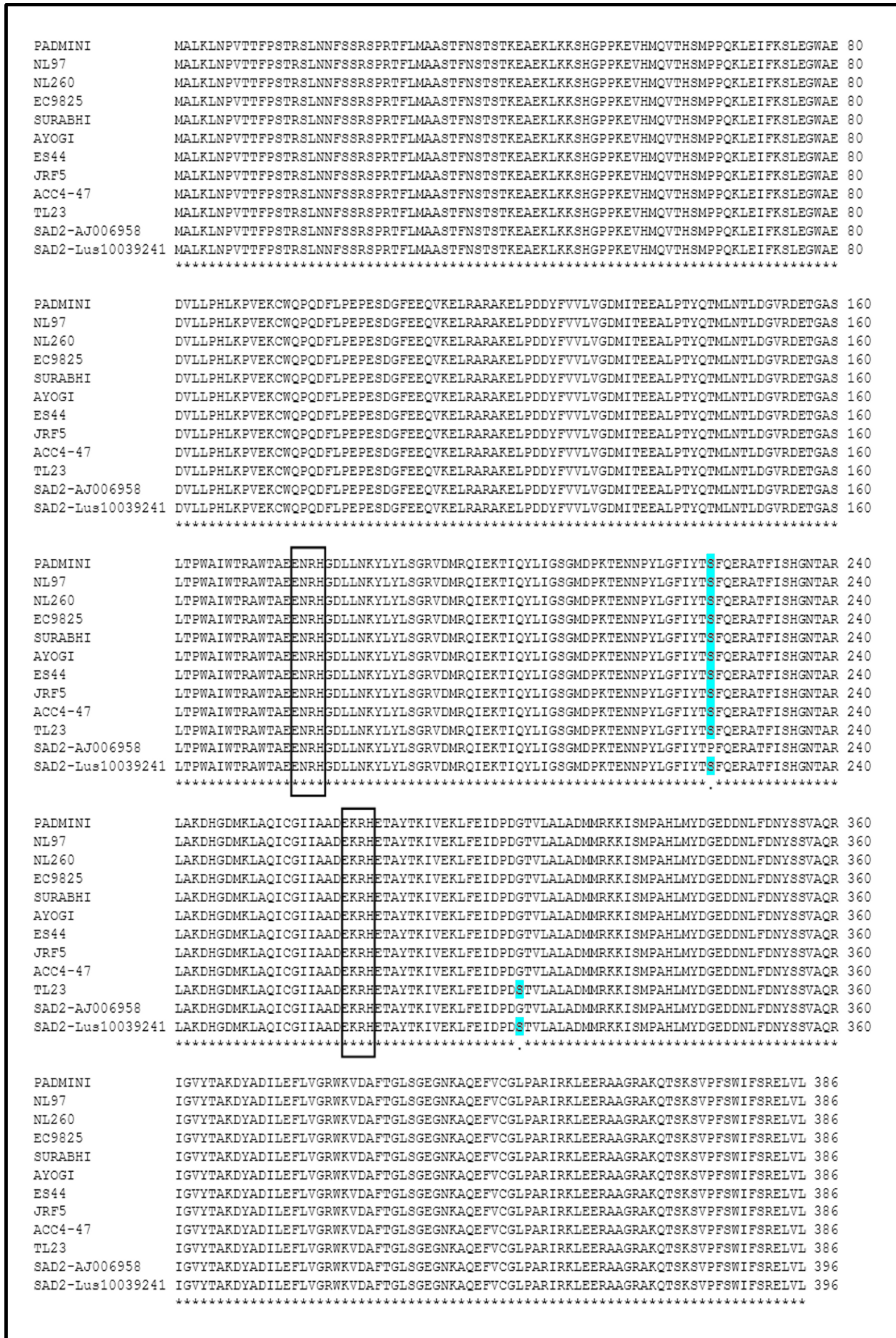


Figure 4.2 Alignment of deduced amino acid sequences of *SAD2* gene from ten flax varieties with *SAD2* sequence in NCBI and Phytozome databases. The boxes represent conserved di-iron binding motifs [(D/E) X2 H] and **(.)** represents variation in AA sequence and highlighted with turquoise colour

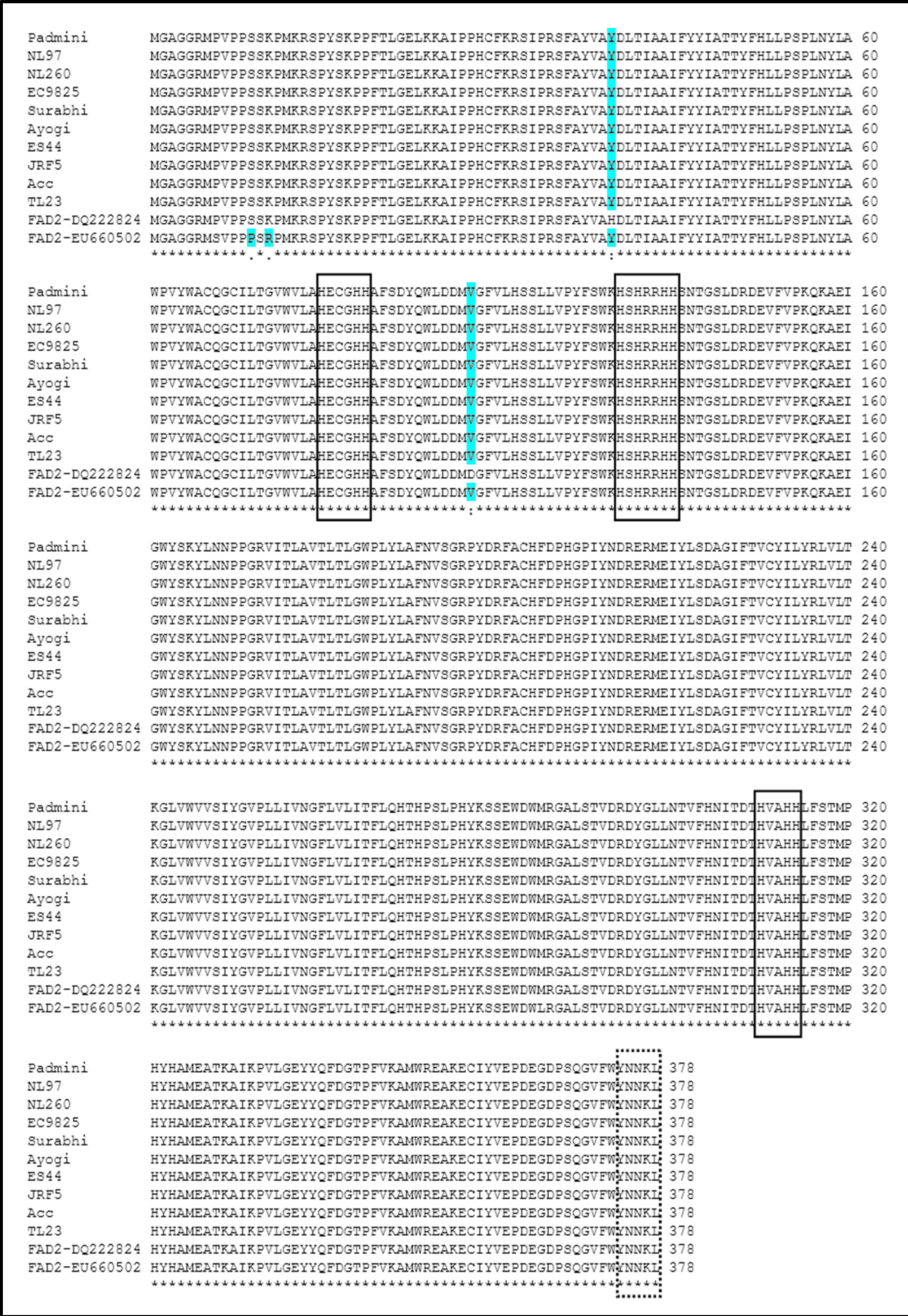


Figure 4.3: Alignment of deduced amino acid sequences of *FAD2* gene from ten flax varieties with two *FAD2* sequences in NCBI database. The solid boxes represent conserved histidine motifs and the dotted box represents ER retrieval motif. Variation in the AA sequences is represented by **(.)** and **(:)** and highlighted with turquoise colour

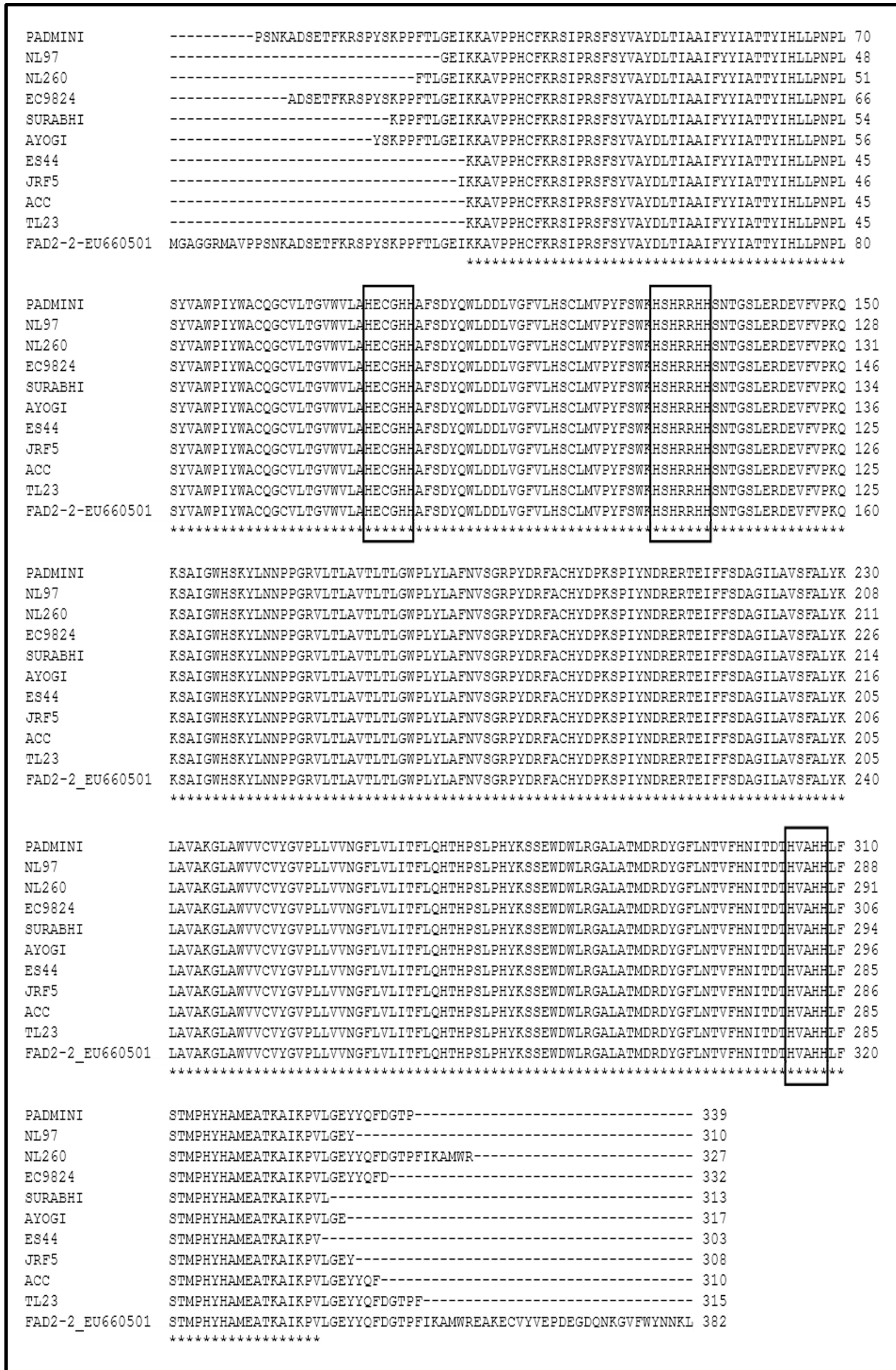


Figure 4.4 Alignment of deduced amino acid sequences of *FAD2-2* gene from ten flax varieties with *FAD2-2* sequence in NCBI database. The boxes represent conserved histidine motifs

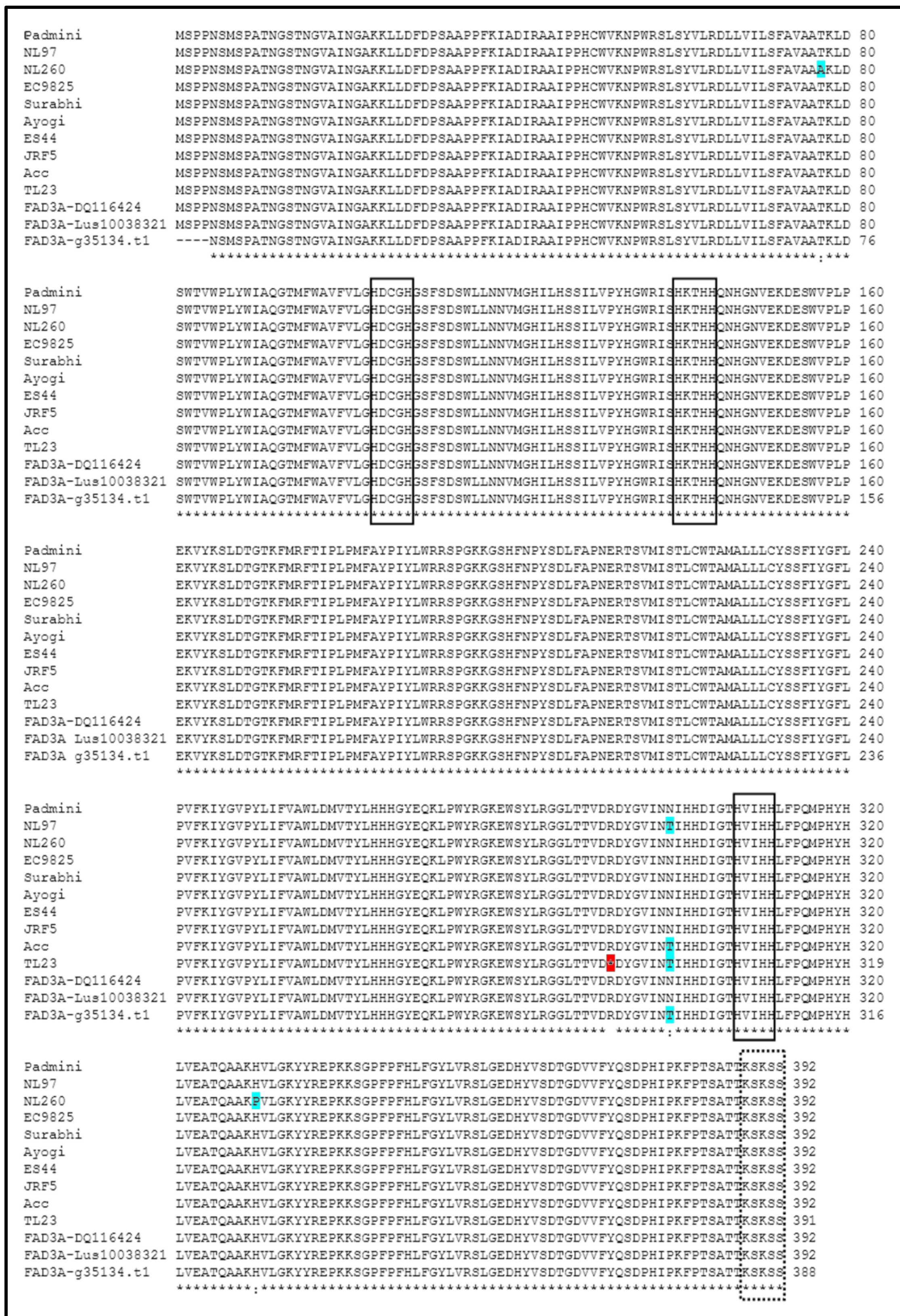


Figure 4.5: Alignment of deduced amino acid sequences of *FAD3A* gene from ten flax varieties with *FAD3A* sequence in NCBI, Phytozome and Linum.ca databases. The solid boxes represent conserved histidine motifs and dotted box represents di-lysine motif. Variation in the AA sequence is represented by **(:)** and highlighted with turquoise colour. Stop codon in TL23 is represented by * in the AA sequence

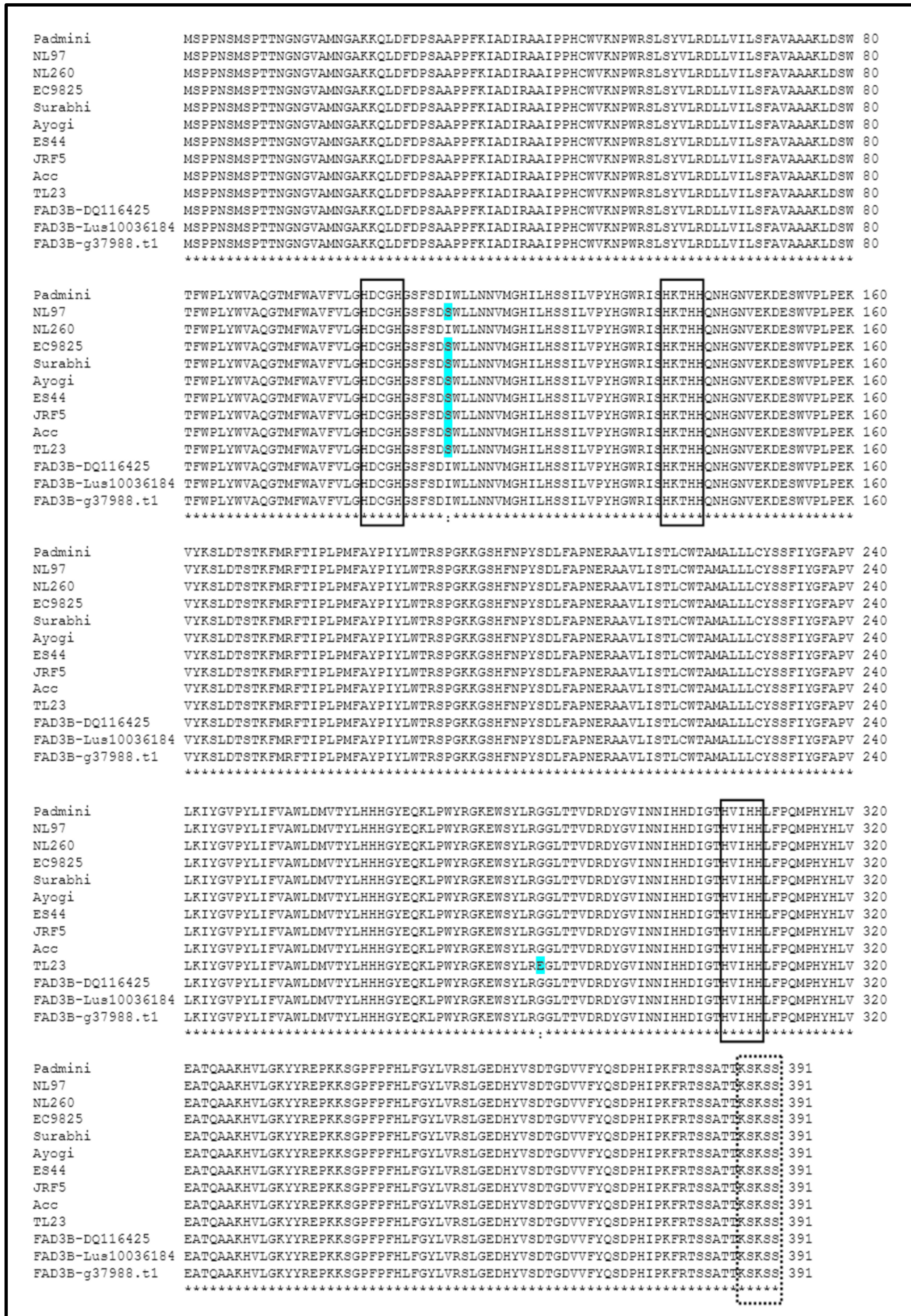


Figure 4.6: Alignment of deduced amino acid sequences of *FAD3B* gene from ten flax varieties with *FAD3B* sequence in NCBI, Phytozome and Linum.ca databases. The boxes represent conserved histidine motifs and dotted box represents di-lysine motif. Variation in the AA sequence is represented by **(:)** and highlighted with turquoise colour

The deduced AA sequences of *FAD3A* gene were identical in all the varieties except TL23 and NL260 (Figures 4.5 & 4.6). TL23 displayed a nonsense mutation (C to T) in the 5th exon causing a premature stop codon (TGA) that lead to a truncated protein with 291 AAs. NL260 showed two AA changes, T77A and H330P. Besides this, in three varieties (NL97, ACC No 4/47 and TL23) an additional substitution (N299T) was observed. For *FAD3B* gene, the deduced AA sequences were identical in the ten varieties, except for the TL23, where G283E change was observed. Similarly, in eight varieties except Padmini and NL260, I112S substitution was observed, which was identical to the *FAD3A* sequence (DQ116424.1) from NCBI (Figure 4.6). Both the FAD3 AA sequences carried the three consensus His-rich motifs required to bind the di-iron active site and for catalysis. Besides these, they also possessed a conserved dilysine ER-retrieval motif (KSKSS) near the C-terminal end (Figures 4.5 & 4.6).

4.3.3 Haplotype analysis

Sequence analysis of *SAD1, SAD2, FAD2, FAD2-2, FAD3A* and *FAD3B* genes from the ten flax varieties revealed varied number of haplotypes. A total of 2-9 alleles while, 1-4 protein isoforms were identified for the six genes (Table 4.6). Similarly, allelic frequency varied gene wise in flax varieties under present study (Table 4.7).

Table 4.6: Allelic variations and number of protein isoforms identified in the six desaturase genes among ten flax varieties

	SAD1	***SAD2***	***FAD2***	***FAD2-2***	***FAD3A***	***FAD3B***
Intronic and exonic variation	2	6	4	2	9	6
Exonic variation	1	2	4	2	7	5
Protein Isoforms	1	2	1	1	4	3

Table 4.7: Representation of number of haplotypes and their allelic frequency, for each desaturase gene, in the ten varieties. Values in parentheses represent allelic frequency

Gene	Haplotype number								
	1	2	3	4	5	6	7	8	9
SAD1	(**0.8**) Padmini, NL97, NL260, EC9825, Ayogi, ES44, JRF5, TL23	(**0.2**) Surabhi, Acc No. 4/47							
SAD2	(**0.2**) Padmini, Acc No. 4/47	(**0.3**) NL97, Surabhi, EC9825	(**0.1**) NL260	(**0.2**) Ayogi, ES44	(**0.1**) JRF5	(**0.1**) TL23			
FAD2	(**0.7**) Padmini, NL97, NL260, Ayogi, ES44, JRF5, Acc No. 4/47	(**0.1**) EC9825	(**0.1**) Surabhi	(**0.1**) TL23					
FAD2-2	(**0.9**) Padmini, NL97, NL260, EC9825, Surabhi, Ayogi, JRF5, Acc No. 4/47, TL23	(**0.1**) ES44							
FAD3A	(**0.1**) Padmini	(**0.1**) NL97	(**0.1**) NL260	(**0.1**) EC9825	(**0.2**) Surabhi, JRF5	(**0.1**) Ayogi	(**0.1**) ES44	(**0.1**) Acc No. 4/47	(**0.1**) TL23
FAD3B	(**0.2**) Padmini, NL260	(**0.3**) NL97, Ayogi, ES44	(**0.2**) EC9825	(**0.2**) Surabhi, JRF5	(**0.1**) Acc No. 4/47	(**0.1**) TL23			

4.3.3.1 SAD haplotypes

The number of variable sites in *SAD1* and *SAD2* were 1 and 7, respectively, and led to 2 and 6 alleles or haplotypes, respectively, in the ten flax varieties for these genes (Tables 4.8 & 4.9). The haplotype diversity H_d, for *SAD1* was 0.36, while for *SAD2* it was 0.89. Further, for *SAD1*, allelic frequency of haplotype 1 was 0.8 where all the varieties except Surabhi and Acc No. 4/47 belonged to this haplotype. For *SAD2*, out of the six haplotypes observed amongst the ten varieties, haplotype 3 was shared by maximum number of varieties (NL97, EC9825 and Surabhi) (Table 4.6). On the other hand, when only the coding (exonic) regions from these two genes were analyzed, *SAD1* showed only one haplotype and one protein isoform; while, *SAD2* revealed two haplotypes and two protein isoforms. In case of *SAD2,* TL23 represented one separate haplotype and the remaining nine varieties formed another haplotype (Table 4.6).

Table 4.8: Haplotype definition of *SAD1* gene

Haplotype	**Nucleotide position**	**Frequency**
	1811	
Hap 1	T	8
Hap 2	C	2

Table 4.9: Haplotype definition of *SAD2* gene

Haplotype	**Nucleotide position**							**Frequency**
	259	262	629	1435	1457	2172	2225	
Hap 1	A	A	C	T	T	G	C	2
Hap 2	.	.	G	.	.	.	.	3
Hap 3	.	.	G	C	C	.	.	1
Hap 4	.	.	.	C	C	.	.	2
Hap 5	C	G	.	.	.	.	.	1
Hap 6	C	G	.	C	C	A	A	1

4.3.3.2 FAD2 haplotypes

The *FAD2* and *FAD2-2* genes are intronless and we observed 3 and 1 variable sites respectively, among the ten varieties (Tables 4.10 & 4.11). These genes had 4 and 2 alleles or haplotypes, with the H_d values of 0.53 and 0.2, respectively. In *FAD2*, seven varieties represented one haplotype while, in *FAD2-2* except for ES44, the remaining nine varieties represented only a single haplotype (Table 4.7). However, all the varieties had only one form of the protein for both *FAD2* and *FAD2-2*.

Table 4.10: Haplotype definition of *FAD2* gene

Haplotype	**Nucleotide position**			**Frequency**
	105	126	624	
Hap 1	C	C	C	7
Hap 2	.	.	T	1
Hap 3	.	T	.	1
Hap 4	T	.	T	1

Table 4.11: Haplotype definition of *FAD2-2* gene

Haplotype	**Nucleotide position**	**Frequency**
	378	
Hap 1	C	9
Hap 2	T	1

4.3.3.3 FAD3 haplotypes

Complete *FAD3A* and *FAD3B* genes including intronic and exonic regions represented 178 and 16 variable sites, respectively among the ten varieties (Tables 4.12 & 4.13). These resulted in 9 and 6 alleles or haplotypes with the H_d values of 0.98 and 0.89, respectively. Only the haplotype 5 in *FAD3A* was shared by two varieties with an allelic frequency of 0.2, while the remaining haplotypes were represented by a single variety each (Table 4.7). When only the coding regions of both the genes were considered, the ten varieties showed 10 and 5 variable sites, which resulted in 7 and 5 alleles or haplotypes with H_d of 0.91 and 0.76, respectively

(Table 4.6). Most of the nucleotide substitutions were synonymous in nature; however, only 4 non-synonymous substitutions leading to 5 isoforms of *FAD3A* were observed. However, the FAD3A protein in TL23 was truncated and hence was not considered as an isoform (Table 4.6). In case of *FAD3B*, 2 non-synonymous substitutions leading to 3 protein isoforms were observed (Table 4.6).

4.3.4 Structure prediction and analysis of desaturases

The 3D structure of SAD (Δ9) protein was predicted by homology modeling, based on the 86% sequence identity of flax SAD protein with the crystal structure of Δ9 desaturase from castor (PDB ID: 1OQ4). The quality of the model produced was assessed by checking the protein stereology by Ramchandran plot analysis using RAMPAGE (Lovell et al., 2003) and the energy was checked by ProSA (Wiederstein and Sippl, 2007). Ramchandran plot of both SAD1 and SAD2 models showed 97.2% of the residues located in favored and allowed Φ and Ψ regions (94.2% and 94.9% most favorable, while 3% and 2.3% in additional allowed region for SAD1 and SAD2 proteins, respectively), whereas, 2.8% of the residues as outliers (Figures 4.7A & B). ProSA analysis yielded a Z-score of -7.06 and -5.29 for SAD1 and SAD2, respectively (Figures 4.8A & B), as well as the negative energy values for most of the residues (Figure 4.9A & B).

The root mean square deviation (RMSD) between the modeled structures of SAD1 and SAD2 with the template (castor SAD protein) was 0.25Å and 0.87Å. All these assessments indicated a good quality of model predicted for SAD from flax. The 3D structure of both the SAD genes, SAD1 and SAD2 showed a compact cylindrical structure, with the presence of 11 α-helices and one anti-parallel β-sheet at the C-terminus. Out of the 11 α-helices, 9 form a core bundle with the remaining 2 helices capping each end. Four out of the 9 core α-helices, $\alpha 3$, $\alpha 4$, $\alpha 6$ and $\alpha 7$ with 21, 26, 28 and 30 residues in case of SAD1 and 25, 27, 27 and 31 residues in case of SAD2 are longer than the rest as observed for the castor SAD structure (Figures 4.10A & B).

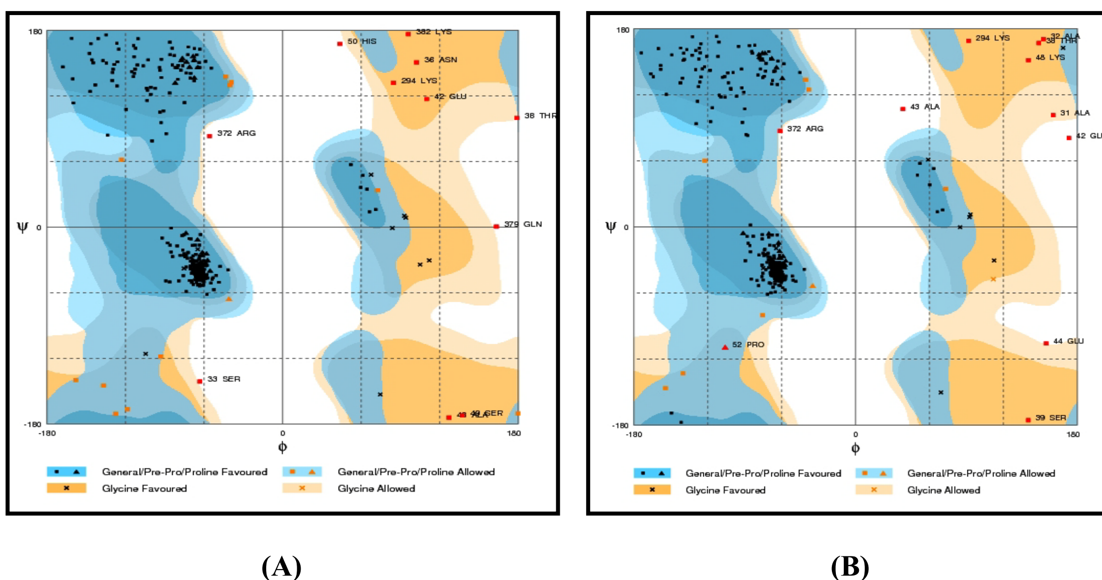
Table 4.12: Haplotype definition of *FAD3A* gene for nine varieties except NL260

Haplotype	Nucleotide position										Frequency
	864	1232	1578	2226	2263	2285	2431	2549	3000	3202	
Hap 1	G	C	C	G	C	A	G	T	A	C	1
Hap 2	.	A	.	A	.	C	.	C	.	.	1
Hap 3	.	A	.	A	.	.	.	C	.	T	1
Hap 4	.	A	.	A	.	.	.	.	.	T	2
Hap 5	.	A	T	.	.	.	.	C	.	T	1
Hap 6	.	A	.	.	.	.	.	C	.	T	1
Hap 7	.	A	.	A	.	C	A	.	.	T	1
Hap 8	T	A	.	A	T	C	.	.	C	T	1

Table 4.13: Haplotype definition of *FAD3B* gene

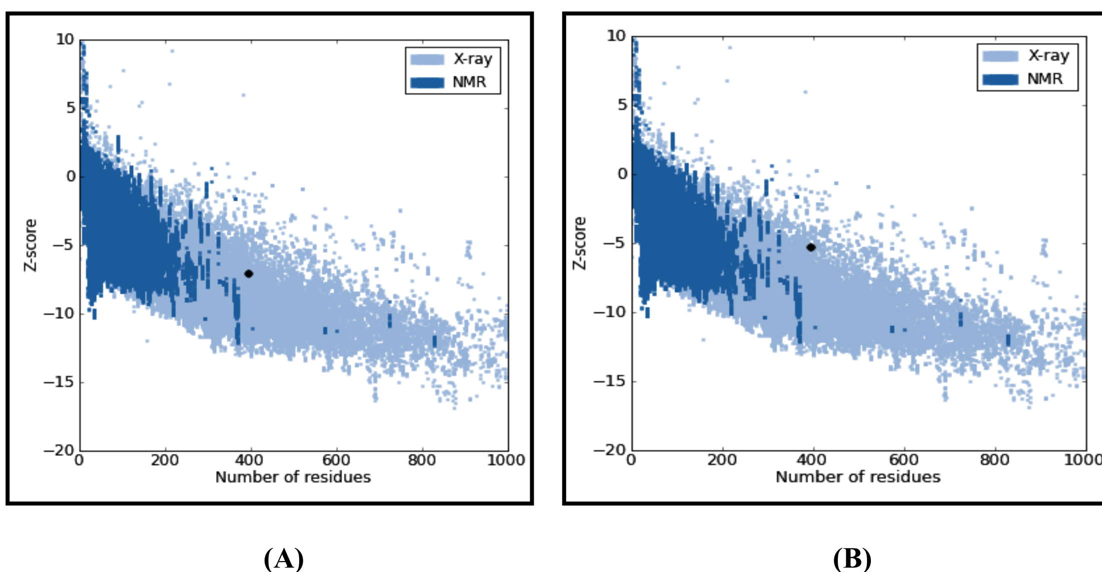
Haplotype	Nucleotide position																	Frequency
	225	366	419	899	907	994	999	1088	1104	1124	1234	1413	1640	1739	1935	1963	2629	
Hap 1	A	A	T	C	G	T	.	T	A	T	G	C	T	A	A	G	C	2
Hap 2	G	C	G	T	A	C	T	G	C	C	A	T	.	.	.	.	.	3
Hap 3	G	C	G	T	A	C	T	G	C	C	A	T	A	G	C	.	.	2
Hap 4	G	C	G	T	A	C	T	G	C	C	A	.	.	.	.	.	.	2
Hap 5	G	C	G	T	A	C	T	G	C	C	A	T	A	G	.	.	T	1
Hap 6	.	.	G	T	A	C	T	G	C	C	A	.	.	.	.	A	.	1

The di-iron binding conserved motifs for soluble desaturase were present on 4th and 7th core α-helices (Figures 4.10A & B). The variation P225S in all the ten varieties and G282S in TL23 SAD2 AA sequence did not affect the predicted protein structures of these varieties when compared with the predicted protein structure of NCBI (Figures 4.11A & B).



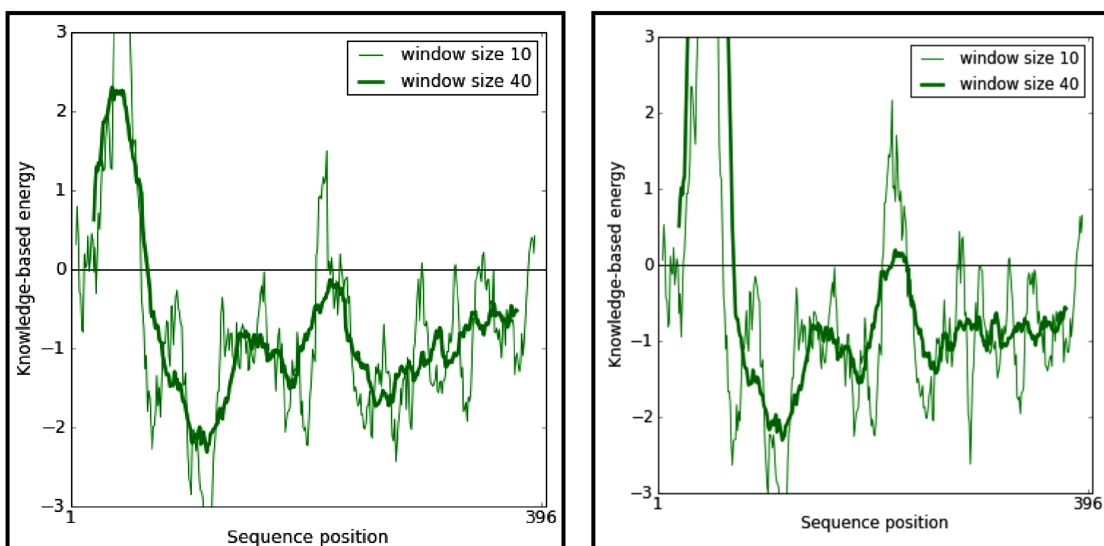
(A) **(B)**

Figure 4.7: Ramchandran plot of the model generated for **(A)** SAD1 and **(B)** SAD2 protein



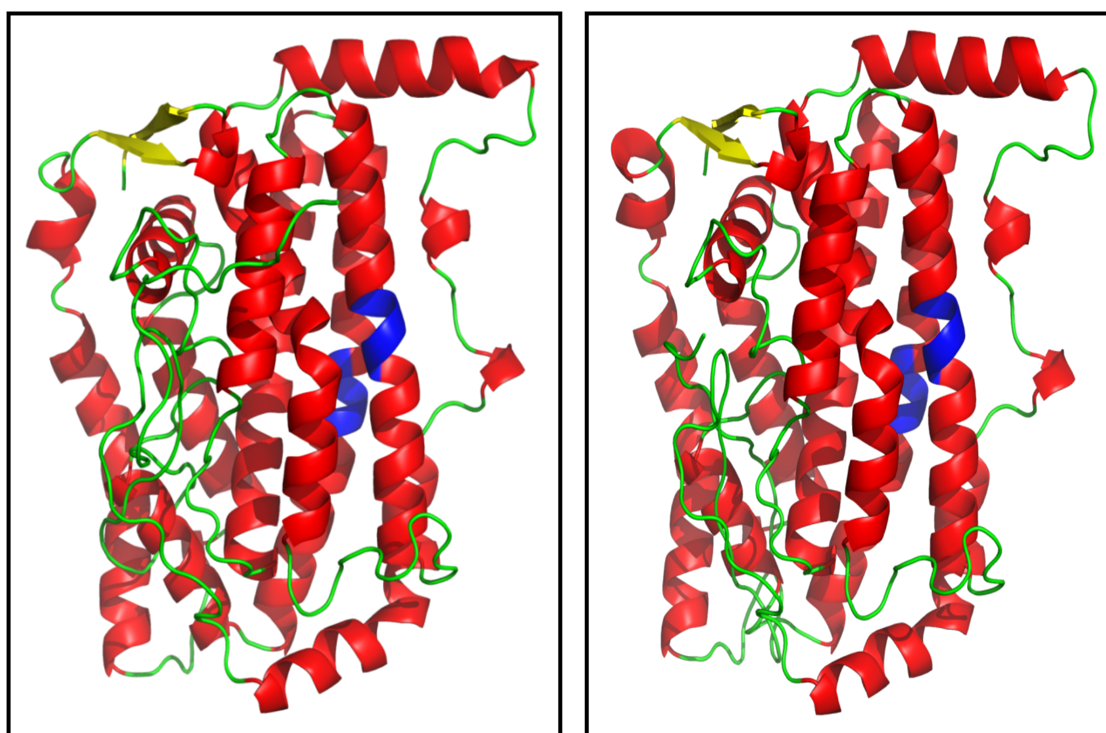
(A) **(B)**

Figure 4.8: Z-score plots of the model generated for **(A)** SAD1 and **(B)** SAD2 protein indicated by black filled circle (Z-score for SAD1: -7.06 and Z-score for SAD2: -5.29)



(A) **(B)**

Figure 4.9: ProSA based energy profiles of the model generated for **(A)** SAD1 and **(B)** SAD2 proteins



(A) **(B)**

Figure 4.10: **(A)** SAD1 and **(B)** SAD2 protein models (helix region in blue represents two conserved di-iron binding motifs)

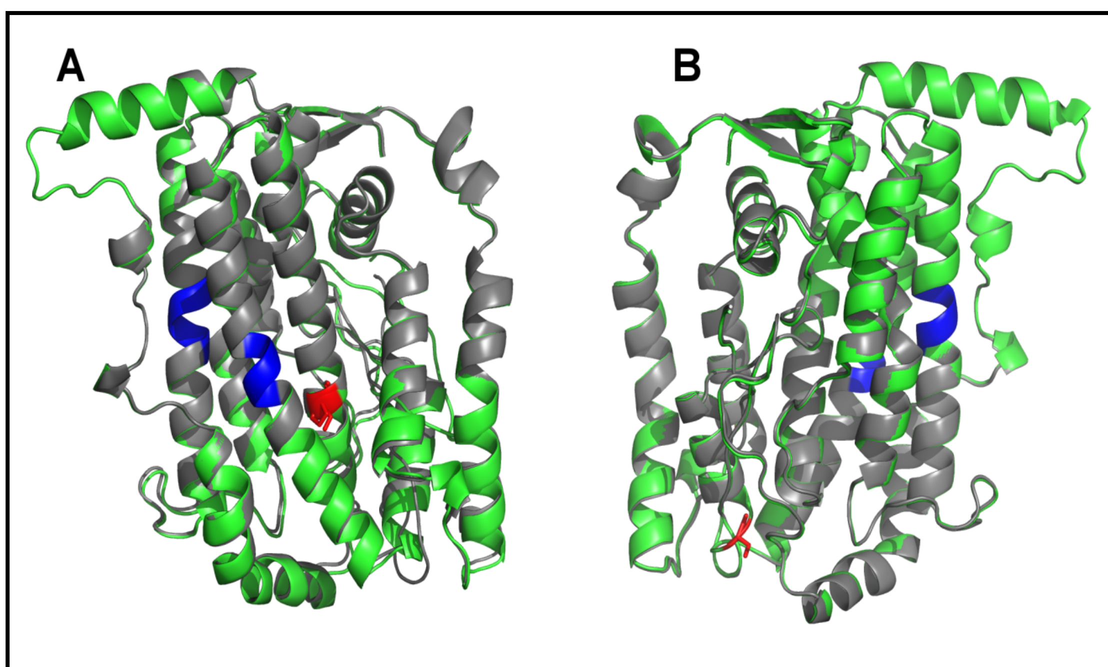


Figure 4.11: NCBI SAD2 protein model (grey) superimposed with **(A)** SAD2 protein model of NL260 (green) (representative of all the varieties), **(B)** SAD2 protein model of TL23 (green). The region in red indicate the position of the AA change while helix region in blue represents two conserved di-iron binding motifs

The 3D structures of FAD2 and FAD3 proteins were generated through the I-TASSER server (Zhang, 2008; Roy et al., 2010). The server initially predicted five models out of which the best model was selected from the output based on the C-score (C-score is a confidence score for estimating the quality of predicted models by I-TASSER). An initial structural model was generated and subjected to an energy minimization procedure with GROMOS96 as implemented in PyMol (The PyMol Molecular Graphics System, Version 1.2r3 pre, Schrödinger LLC) to reduce poor van der Waals contacts and correct the stereochemistry of the model. Further, the quality of the models produced (both FAD2 and FAD3) was also assessed by checking the protein stereology by Ramachandran plot analysis using the RAMPAGE and the energy was checked by ProSA. Ramachandran plot for predicted FAD2 model showed 92.3% of the residues located in favored and allowed Φ and Ψ regions (84.8% most favorable while 8.5% in additional allowed region), whereas, 6.6% residues were outliers (Figure 4.12A). ProSA analysis yielded a Z-score of -3.7 for FAD2 model and the energy values of the residues were negative in some region, while positive in other regions (Figure 4.12B & C). Similarly, Ramachandran plots of FAD3A and FAD3B model showed 95.9% and 95.6% of the residues located in

favored and allowed Φ and Ψ regions (88.7% and 84.6% most favorable while 7.2% and 8.7% in additional allowed region, respectively for FAD3A and FAD3B); whereas 4.1% and 4.4% residues were outliers, respectively (Figures 4.14A and B). While, the ProSA analysis yielded a Z-score of -2.56 and -2.79 for FAD3A and FAD3B models, respectively (Figure 4.15A & B). The energy values of the residues were negative in some regions, while positive in other regions (Figure 4.16A & B). These results indicated that the predicted FAD3 and FAD2 models are stable.

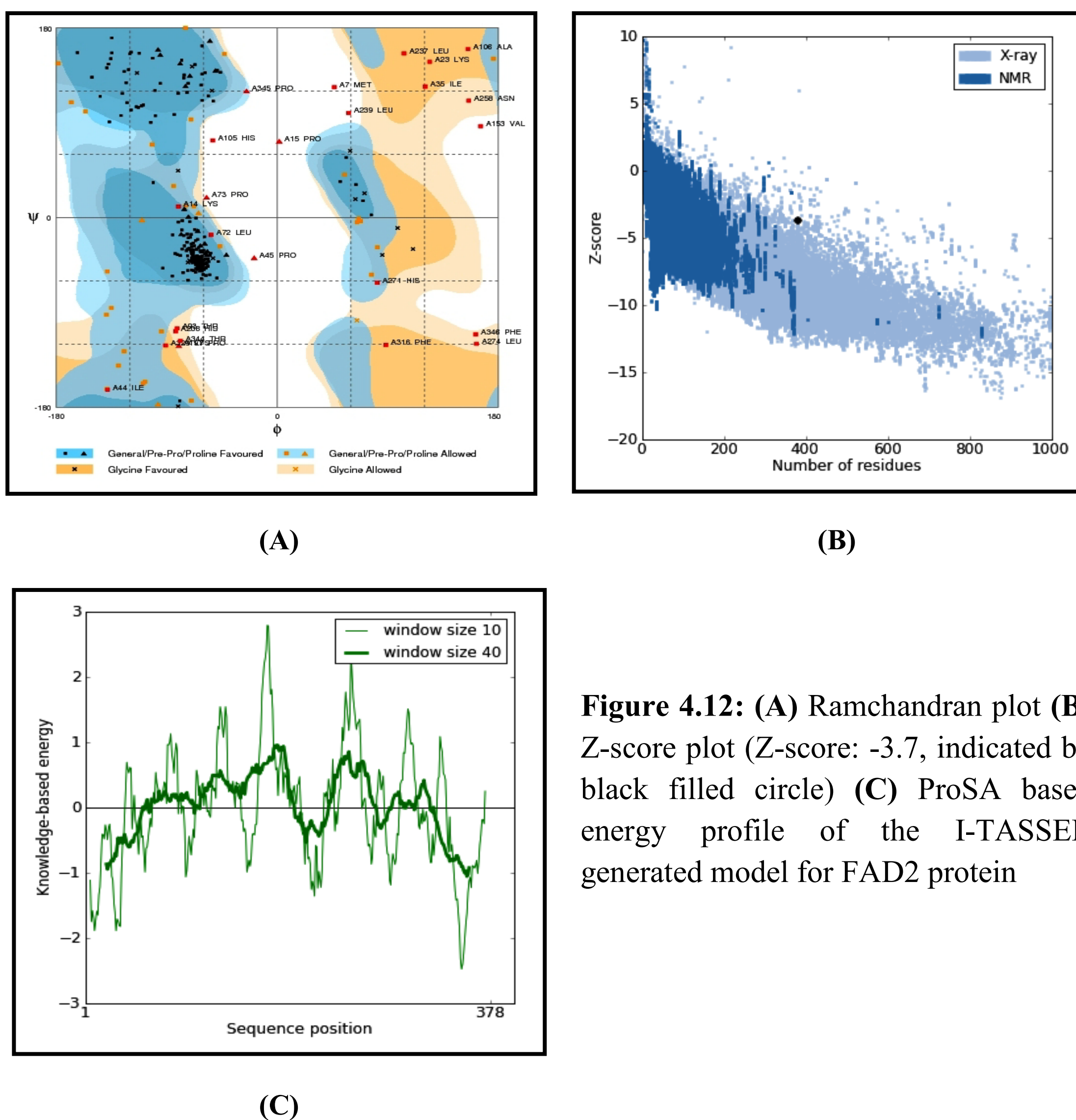


Figure 4.12: (A) Ramchandran plot **(B)** Z-score plot (Z-score: -3.7, indicated by black filled circle) **(C)** ProSA based energy profile of the I-TASSER generated model for FAD2 protein

The FAD2 protein model consisted of 12 α-helices and one anti-parallel β-sheet at the C-terminus, of which $\alpha1$ and $\alpha6$ were interrupted by 2-4 AA residues. Four of the α-helices, $\alpha2$, $\alpha3$, $\alpha5$ and $\alpha14$ were larger with 29, 23, 22 and 23 AA residues, respectively (Figure 4.13A). The predicted structure showed the presence of three conserved histidine boxes, namely H1 from 100-105 residues, H2 from 134-140

residues and H3 from 310-314 residues, respectively. AA sequences of all the ten varieties showed substitution at two positions (H53V and D117V) as compared to NCBI sequence, however there were no obvious structural changes observed when their protein models were superimposed on the NCBI model (Figure 4.13A).

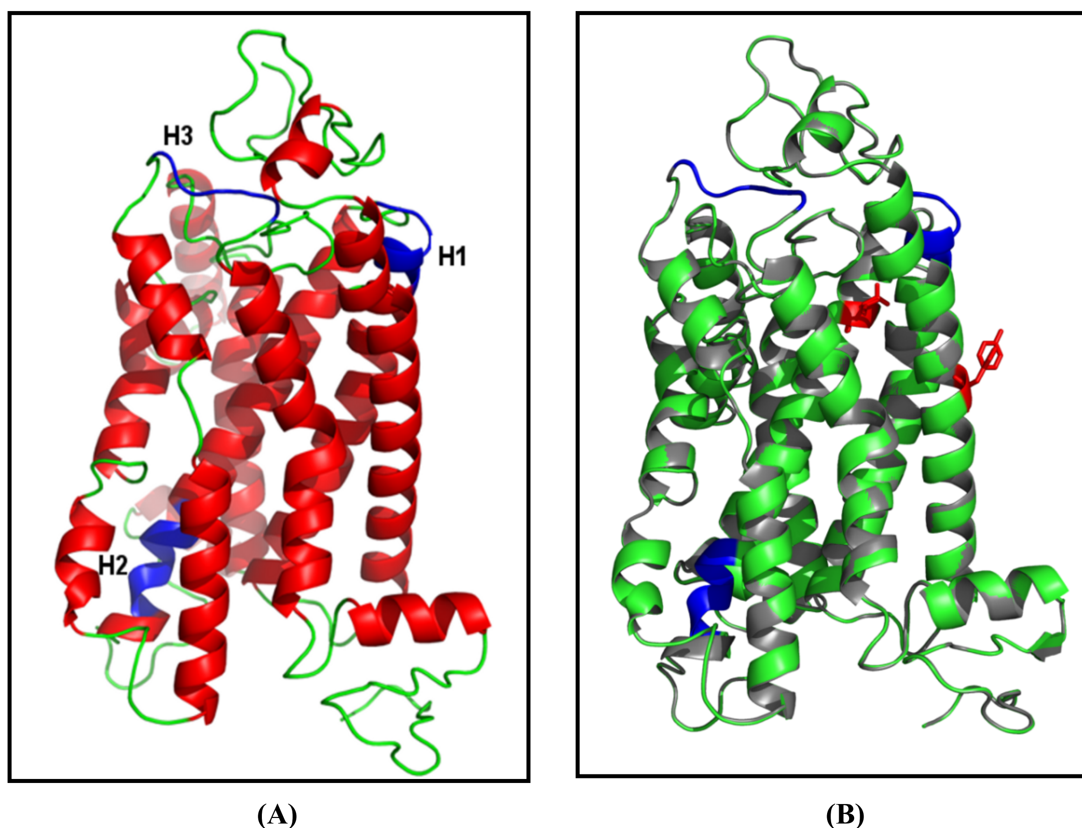
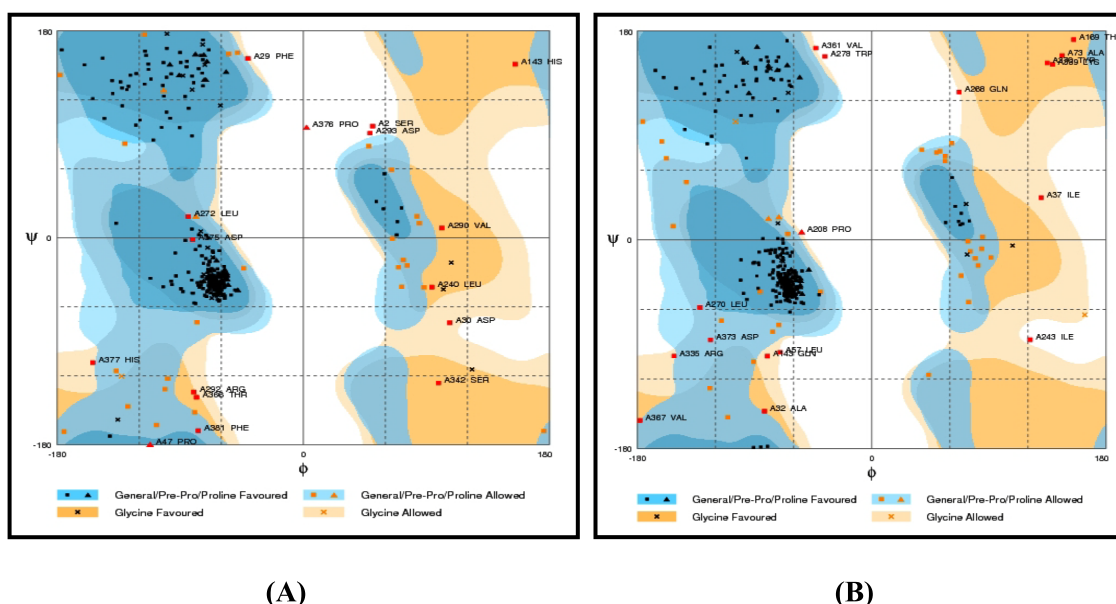


Figure 4.13: (A) FAD2 protein model, **(B)** NCBI FAD2 protein model (grey) superimposed with Padmini FAD2 protein model (green). The regions in red (in the superimposition model, **B**) indicate the positions of the AA change. Region in blue represents three conserved His-box motifs H1, H2 and H3 in both the structures **A** and **B**

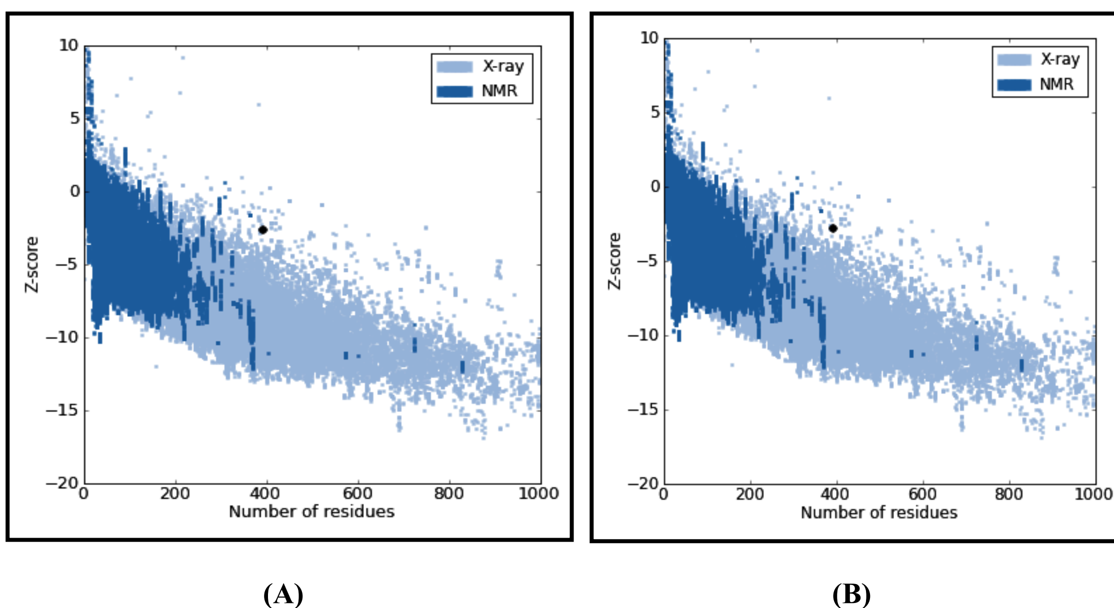
Both the FAD3 protein models consisted of 14 α-helices, of which α6, α9, and α14 in *FAD3A* while, α4, α7, α8 and α13 in FAD3B were interrupted by 1-3 non-helix forming AA residues. In FAD3A and FAD3B proteins, α10 with 36 AA and α7 with 41 AA were the largest helices, respectively (Figures 4.17 & 4.19). The predicted structure also showed the presence of three conserved histidine boxes, namely H1 from 104–108 residues in FAD3A and 102-106 residues in FAD3B, H2 from 140–144 residues in FAD3A and 138-142 residues in FAD3B, while H3 with 307–311 residues in FAD3A and 305-309 in FAD3B, respectively (Figures 4.17 & 4.19).

In the predicted FAD3A protein, due to the H330P substitution in variety NL260 and N299T substitution in varieties NL97 and Acc No.4/47, no structural variation observed when the NCBI and NL260, NL97 and Acc No.4/47 FAD3A predicted protein models were superimposed (Figure 4.19a, b and c). Further, in TL23 FAD3A gene the 3rd His-box was totally absent because of protein truncation at 291st AA residue (Figure 4.18A, B and C). However, For FAD3B gene of TL23, the G283E substitution was observed and no obvious structural variation was identified in the predicted protein model when superimposed over the NCBI model (Figure 4.20B). Another substitution, I112S, which was observed in the AA sequence of all the varieties except for Padmini and NL260, also did not result in any structural variations in these varieties (Figure 4.20B).



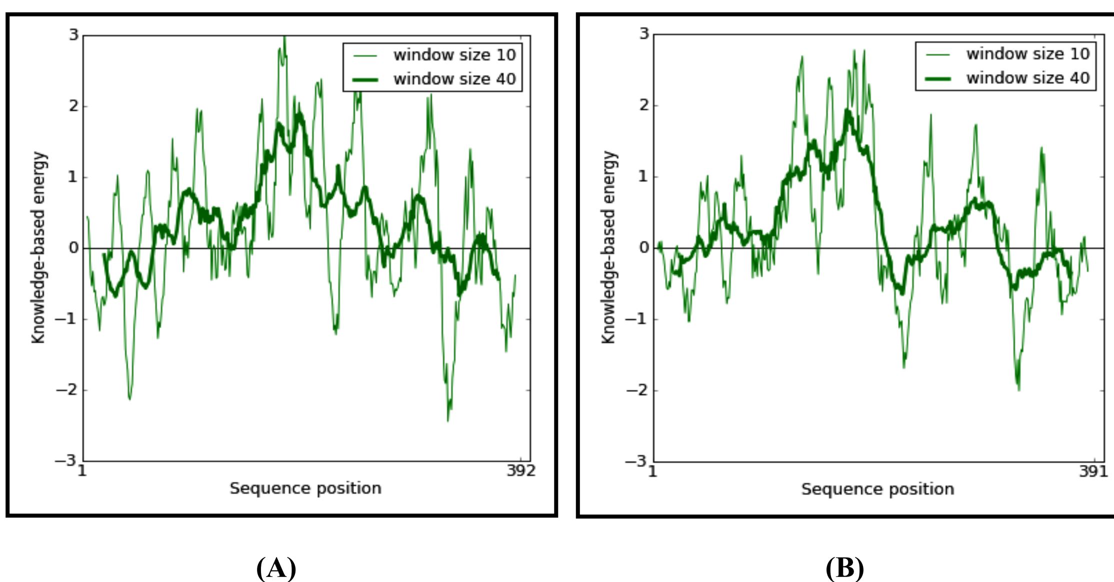
(A) **(B)**

Figure 4.14: Ramchandran plots of the I-TASSER generated model for **(A)** FAD3A and **(B)** FAD3B protein



(A) **(B)**

Figure 4.15: Z-score plots of the homology model generated for **(A)** FAD3A **(B)** FAD3B protein indicated by black filled circle (Z-score of FAD3A: -2.56 and FAD3B:-2.79)



(A) **(B)**

Figure 4.16: ProSA based energy profiles of the homology model generated for **(A)** FAD3A and **(B)** FAD3B protein

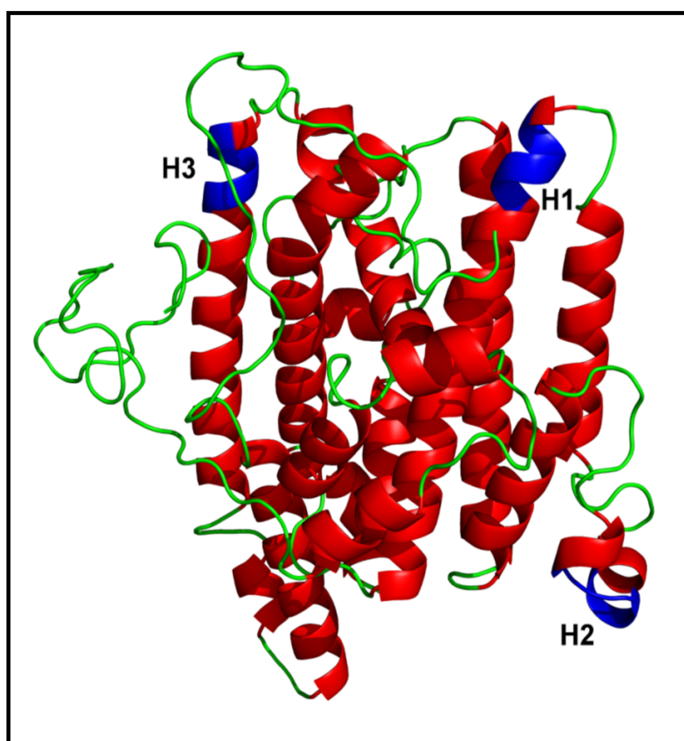


Figure 4.17: FAD3A protein model (helix region in blue represents three conserved His-box motifs, H1, H2 and H3)

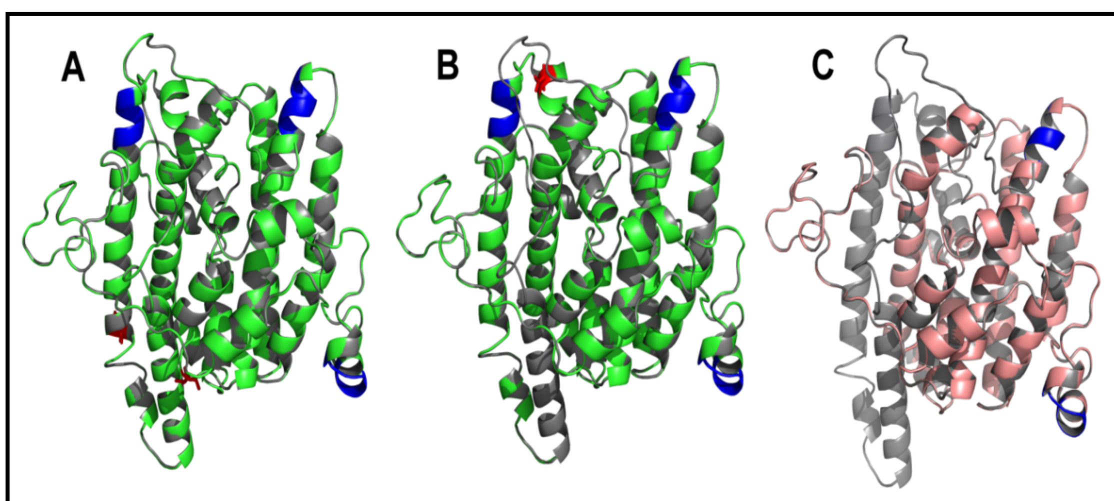


Figure 4.18: NCBI FAD3A protein model (grey) superimposed with **(A)** FAD3A protein model of NL260 (green), **(B)** FAD3A protein model of NL97 (green), **(C)** FAD3A protein model of TL23 (brick red). The regions in red indicate the position of AA changes while helix regions in blue represent three conserved His-box motifs. In TL23, the protein is truncated and the H3 motif is absent. The superimposition model shows incomplete alignment of TL23 model with the NCBI model (Only grey motifs are visible in the figure C from where the TL23 protein is truncated)

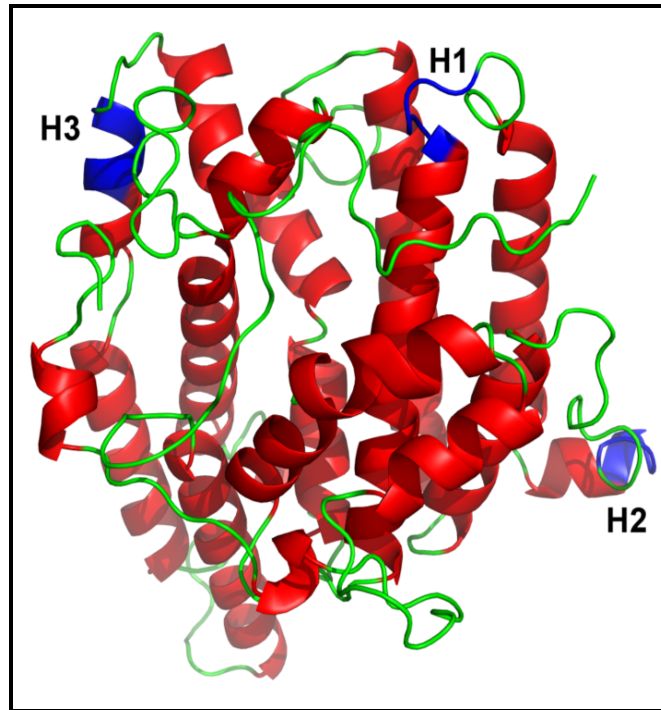


Figure 4.19: FAD3B protein model (region in blue represents three conserved His-box motifs, H1, H2 and H3)

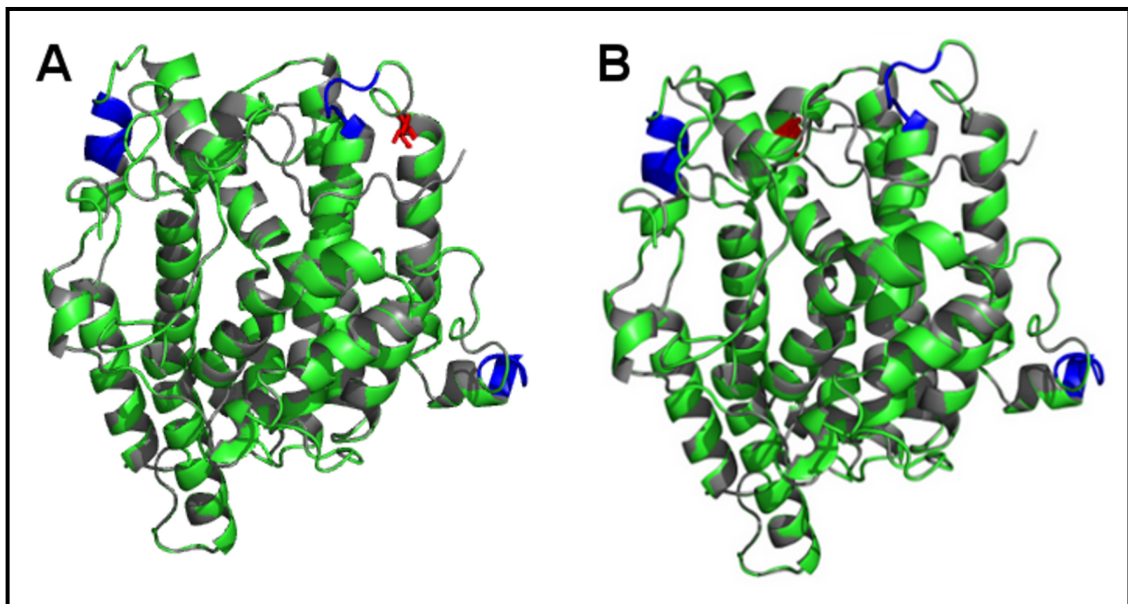


Figure 4.20: NCBI FAD3A protein model (grey) superimposed with **(A)** FAD3A protein model of NL97 (green) (representative of 8 varieties other than NL260 and Padmini), **(B)** FAD3A protein model of TL23 (green). The regions in red indicate the positions of AA change while the regions in blue represent three conserved His-box motifs

4.4 Discussion

4.4.1 Unique and distinct haplotypes in desaturase genes across flax varieties

In the present analysis, gene sequence comparison of six fatty acid desaturases from ten flax varieties differing in FA composition was performed with the respective reported gene sequences in the NCBI as well as the Phytozome databases. In this study, *SAD* and *FAD2* genes did not show much variation in the ten varieties compared to the *FAD3* genes. Further, it was also noticed that most of the observed variations were in the intronic region and hence, the number of protein isoforms detected were low except for *FAD3* genes (Table 4.6). The haplotypes predicted for all the six genes in our study, when compared with the desaturase gene sequences reported for flax in the NCBI database, showed new haplotypes. In case of *SAD2* gene, there are sequences from 101 flax genotypes reported in the database to reveal domestication history. However, even then, seven of our varieties showed four new haplotypes and only three, Padmini, Acc No. 4/47 and TL23 showed haplotypes similar to those in the database. For *FAD3A* and *FAD3B* genes, the ten genotypes in the present study revealed different haplotypes than the corresponding sequences reported for four Canadian flax varieties in the NCBI database. Overall, the number of haplotypes observed in the Indian genotypes for these desaturase genes was higher than those reported in the database. However, there are not many gene sequences reported in the database from varied geographic locations to comment upon the diversity observed in the Indian genotypes.

4.4.2 AA sequence comparison and structural features of desaturase proteins across flax varieties

The deduced AA sequences of all the desaturase genes (*SAD1, SAD2, FAD2, FAD2-2, FAD3A* and *FAD3B*) were identical in the ten flax varieties except for few changes observed in varieties TL23 (in *SAD2, FAD3A* and *FAD3B*) and NL260, NL97, Acc No. 4/47 and TL23 (in *FAD3A* gene), which were not reported earlier. Further, in the ten varieties we observed that the soluble desaturase, SAD showed two characteristic di-iron binding motifs while the membrane bound desaturases, FAD2 and FAD3 showed the presence of three di-iron binding histidine boxes (Shanklin et al., 1994; Shanklin and Cahoon, 1998; Teixeira et al., 2010) as well as specific motifs required for their steady state endoplasmic reticulum localization (Arondel et al., 1992; Reed et

al., 2000; McCartney et al., 2004) as reported for these desaturases in other plant species.

For SAD proteins (both SAD1 and SAD2) stereological stable models were obtained, which showed high similarity with the crystal structure of castor Δ9-desaturase protein and were identical in all the ten varieties in the present study. For FAD2, the structure was predicted only for one protein form (for FAD2 and not for FAD2-2) and all the ten varieties had identical structure as compared to NCBI protein model. To our knowledge, no other reports are available on the model of FAD2 proteins and for the first time it is being reported using any type of modeling approaches. Similarly, for FAD3 proteins (both FAD3A and FAD3B) except for predicted FAD3A protein from TL23 (truncated protein with only 291 AA acids), all the remaining varieties had identical models to that of predicted FAD3 protein models for the sequences reported in NCBI. However, our FAD3 models were different from earlier predicted models of FAD3A and FAD3B of the same sequence in NCBI by Khadake et al. (2011) using the same I-TASSER server. Both the models had different number of α-helices, the longest and the shortest helices also varied and absence of β-sheets was observed in our model. The percentage of residues in favoured and allowed Φ and Ψ regions analyzed by Ramchandran plots also varied between the two predicted models. Our FAD3A and FAD3B models with 95.6% and 94.6% residues respectively, showed higher percentage of residues in the favoured and allowed regions, with greater stereological stability than the models predicted for the same proteins by Khadake et al. (2011) with 88.5% of the residues in this region. Such discrepancies in the protein models cannot be further verified because of unavailability of crystal structure for any membrane bound desaturase in the protein databases. On the basis of the variations observed for the predicted FAD3 protein models from our analysis and that deduced by Khadake et al. (2011), it is clear that the bioinformatic studies are not enough to deduce the correct protein models and it just gives only a preliminary insight into the secondary and tertiary structure of these proteins. Thus, to confirm these models, we need to perform protein crystallographic studies. There is an urgent need to initiate such experiments for the membrane bound desaturases.

4.4.3 Correlation of desaturase gene mutation with ALA content

The loss of *FAD3A* and *FAD3B* gene activity, resulting into very low ALA content, due to truncated proteins has been reported in some Canadian varieties of flax (Vrinten et al., 2005; Banik et al., 2011). Similarly in Arabidopsis, it has been observed that the mutation of *FAD3* gene reduced seed ALA levels to 3% from wild-type level of 16% of total oil. It has further been shown that the mutation in any of the three microsomal ω-3 *FAD* genes contributing to the ALA content in soybean, lead to reduction in ALA in these mutant lines (Bilyeu et al., 2003). Such observations were also made for the other two desaturases. For example, in canola *fad2* gene, which is a major locus for high-oleic acid content, a nucleotide variation created a stop codon, which resulted in an inactive enzyme and, consequently, higher oleic acid content (Hu et al., 2006). In peanut, Miniature Inverted–Repeat Transposable Elements (MITE) insertions were responsible for creation of such stop codons in the coding sequence of the *fad2* genes (Patel et al., 2004) and hence, high OA content in it. In our study, TL23 is a mutant variety with only ~2% ALA. In this variety, similar observation was made in case of *FAD3A* gene. The nucleotide sequence of the *FAD3A* gene showed the presence of a nonsense mutation creating a premature stop codon in the fifth exon after amino acid 291, which was just before the 3rd His-box (HVIHH, starting at 307 base). This led to absence of the 3rd His-box from the mature protein and hence, the iron binding catalytic pocket, necessary for the action of desaturase was hampered. On the other hand, the *FAD3B* in TL23 was a full length gene expected to give higher production of ALA. However, in this *FAD3B* gene, G1963A substitution resulting in G283E AA variation was observed, which was unique and not observed in remaining nine varieties. The loss of complete activity of FAD3A and the substitution G283E in FAD3B indicated that even this change in FAD3B probably resulted in accumulation of such a low proportion of ALA and the pathway of ALA synthesis was interrupted at the production of LA, which was very high (63%).

Besides *FAD3A* gene in TL23, we observed that NL260, which showed the highest ALA accumulation amongst the ten varieties, formed a unique haplotype for this gene and also a protein isoform with two nonsynonymous substitutions. Further, even for *FAD3B*, Padmini and NL260 formed a different haplotype and protein isoform than the remaining eight varieties and interestingly, both were high ALA varieties. However, *in toto*, the other haplotypes were not observed to be specific to

variations in the ALA content of the varieties. On the basis of these observations, it can be suggested that in general, there was no specific haplotype that could be correlated with variation in the ALA content and therefore, probably not responsible for the high and low ALA accumulation in the groups except in varieties TL23, NL260 and Padmini. Alternatively, it is possible that such clear cut correlation of haplotype to variable ALA content could be masked by the variable sequential expression of all the three classes of the genes, *in vivo*, in the flax varieties studied. This needs to be confirmed by establishing recombinant enzymatic assays with individual haplotypes.

CHAPTER 5

Thesis summary and future directions

Summary and future prospects

The present work includes systematic molecular analysis of diversity in Indian flax germplasm and biochemical analysis of flax oil. Further, fatty acid profiling and expression analysis of microsomal desaturase genes from different seed developmental stages of flax varieties varying in ALA content was performed to determine the role of each desaturase gene in the differential accumulation of ALA in these flax varieties. Lastly, all the desaturase genes were isolated, sequences characterized and the deduced AA sequences were subjected to structural predictions of these proteins, from the same flax varieties.

5.1 Genetic and fatty acid diversity amongst Indian flax varieties

Lack of proper knowledge about the genetic diversity present in the Indian germplasm lead to the initiation of my work, involving molecular and biochemical approaches, to reveal the diversity present in a set of 70 genotypes which are prevalently used for flax breeding in India. Total of twelve polymorphic ISSR primers were selected for the genetic analysis yielding 136 loci, of which 87 were polymorphic. The percent loci polymorphism ranged from 11.1% to 81.8% with an average of 63.9% across all the genotypes. The range of PIC scores for various ISSR primers in this study was 0.03 to 0.49, with an average of 0.18. A dendrogram was generated based on the similarity matrix by the UPGMA, wherein the flax genotypes were grouped in five clusters. The Jaccard's similarity coefficient among the genotypes ranged from 0.60 to 0.97. Low genetic diversity detected in the 70 flax genotypes in the present study is in accordance with the other reports that the cultivated flax has low genetic diversity compared to its wild relatives. However, we could identify some genetically most diverse varieties such as, 'Sheetal', 'SLS 50' and 'Ayogi' amongst the set of analyzed genotypes. Unique alleles present in these varieties can be exploited by their probable use in performing crosses with individuals from different genetic clusters. Correlation of FA contents with these clusters exhibited a meaningful segregation of two clusters with high ALA content while one of the clusters with that of LA. The genotypes in these clusters suggested genetic differences associated with these chemotypes. Further, the Mantel's test, which analyses correlation between the distance matrices based on FA content and the molecular data for genotypes, revealed that the molecular variation statistically correlated well with the LA and ALA variation within the varieties analyzed. This is probably the first study where the FA variation, a

primary interest in flax, has been correlated with the molecular variation among the flax genotypes.

5.2 Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content

Indian flax germplasm showed variation in their FA contents. This observation created an interest in understanding the mechanism of differential accumulation of FAs in the germplasm, since such comprehensive studies have not been carried out earlier. Our study initiated with FA profiling of various seed developmental stages of ten Indian flax varieties including one mutant variety. Depending on their ALA content at the maturity, these varieties were grouped under high ALA and low ALA groups. Further, using real-time PCR, the transcript profiling of six microsomal desaturase genes (*SAD1*, *SAD2*, *FAD2*, *FAD2-2*, *FAD3A* and *FAD3B*) was undertaken in the same developing seeds of ten flax varieties, as the sequential action of desaturases decides the composition of various FAs in the storage oil of oilseed crops. We could clearly observe gene specific as well as temporal expression patterns of all the desaturases and the expression profiles correlated well with the FA data for the high and low ALA groups. There was only one exception, JRF5, observed which deviated from this general trend. Even though it belonged to low ALA group, high expression was observed for most of the desaturases in this variety. Our study indicated the efficient conversion of intermediate FAs [stearic (SA), oleic (OA) and linoleic acids (LA)] to the final ALA content due to efficient action of all the three desaturases in high ALA group. While in the low ALA group, even though the initial conversion up to OA was efficient, later conversions up to ALA seemed to be inefficient, leading to higher accumulation of OA and LA instead of ALA. From our analysis it can also be concluded that though all the desaturase genes are required for the final ALA accumulation, four genes (*SAD1, FAD2, FAD3A* & *FAD3B*) might be specially responsible for its differential accumulation in the individual flax variety. This study is not only unique for flax but also for other oil crops as this kind of transcriptional variation in the desaturase genes and its correlation with the FA content has not been analyzed for such a large number of varieties under their natural field growing conditions for any other oil seed crop.

5.3 Sequence characterization and *in silico* structure prediction of desaturases from the flax varieties varying in α-linolenic acid content

This study was further extended to sequence characterize all the desaturase genes from the same ten varieties and to find out if there is presence of any nucleotide sequence variation leading to AA variation and whether these AA sequence variants have any functional implications in the form of differential ALA accumulation. The *SAD* and *FAD2* genes did not show much sequence variations which would translate into nonsynonymous AA substitutions and hence, resulted into only one or two haplotypes and protein isoforms amongst the ten varieties. On the other hand, both *FAD3A* and *FAD3B* genes showed more nucleotide variations and few of them resulted into nonsynonymous AA substitutions, thus, showed more number of haplotypes and protein isoforms compared to the other two genes. *FAD3A* gene of TL23 had a premature stop codon which resulted in a truncated protein and consequently very low levels of ALA accumulation. Besides the mutant variety TL23, some of the high ALA group varieties, like NL260 and Padmini, formed different haplotypes and protein isoforms for both *FAD3A* and *FAD3B*. Further, there were no structural variations observed due to the AA changes observed in any of the desaturase genes. This further suggests probable involvement of transcriptional control of the desturases for ALA content variation.

5.4 Future prospects

Our studies on six variants of three desaturase genes in ten varieties varying in ALA content have shown transcriptional correlation of desaturase gene activities with that of FA accumulation. The nucleotide sequence as well as AA sequence of these gene variants and their haplotype analysis in our studies could not identify specific mutation or haplotype responsible for FA variation except for TL23 and probably NL260 and Padmini. In TL23, truncated FAD3A while in NL260, T77A and H330P substitutions in FAD3A as well as, absence of certain substitutions detected for remaining eight varieties for FAD3B except for Padmini and NL260, were observed. These results point at the possibility of transcriptional control of desaturase genes being primarily important in regulating their activity responsible for ALA content variation. These observations can be used to design future studies in the following areas:

5.4.1 Promoter region analysis

Promoter regions of all the three desaturase genes in the high and the low ALA containing flax groups need to be analyzed so as to understand the transcriptional regulation of expression of these genes. There are many earlier reports in various plant systems where promoter region analysis of a gene has revealed variation in transcription binding sites and corresponding variation in the transcription factors, leading to expression variation of the gene (de Meaux et al. 2005; McKhann et al. 2008; Yi et al. 2010).

5.4.2 Studies towards post-transcriptional and post-translational gene regulation

To explain certain exceptions, in case of NL260, Padmini and JRF5, where the the ALA accumulation did not completely correlate with the transcriptional activity of mainly the *FAD3* genes pointed towards the need to study desaturases gene regulation at the post-transcriptional and post-translational levels using proteomic and biochemical analysis.

5.4.3 Functional analysis of desaturase sequence variants obtained from this study

Isolation of gene sequence variants of different desaturases responsible for protein isoforms as observed in our study and their further cloning in yeast transformation system can be attempted. The functionality of the recombinant protein can be analyzed and compared on the basis of rate of conversion of substrate fatty acids to the final product.

5.4.4 Protein structure determination of membrane bound desaturases

For membrane bound desaturases, it is hard to obtain recombinant protein in large quantities because of difficulty in solubilisation and purification due to their hydrophobicity and strong anchorage to membranes. As a result, the crystal structure of the membrane bound desaturases from any of the species is not yet available. This makes it essential to attempt desaturase gene cloning in various host systems (bacterial, yeast or plant) and devise a better protocol to purify large amount of recombinant protein to establish protein structure using X-ray crystallographic studies.

5.4.5 Genetic resource development

Apart from these biochemical, molecular and structural studies, it is essential to develop genetic resources in case of flax. Our diversity studies can be further exploited to design crosses to be undertaken to map oil content, disease resistance, etc. in flax, for example, Ayogi' (low ALA) X 'NL155'/'NL115' (high ALA); 'Neelam' (susceptible to powdery mildew & Bud fly) X 'JRF5' (resistant to powdery mildew & Bud fly) etc. Association mapping for such complex traits can also be undertaken. The trait analysis in such biallelic or multiallelic cases under varied environmental conditions will throw light on genotype X environment interactions in control of the manifestation of the trait.

BIBLIOGRAPHY

Ahlawat IPS (2007) Agronomy - Rabi Crops, Linseed. Indian Agricultural Research Institute, New Delhi

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. Theoretical and Applied Genetics **112:** 58-65

Anonymous (2001) Nutritional profile of no. 1 Canada Western flaxseed and of yellow flaxseed samples. Canadian Grain Commission, Winnipeg, MB

Anonymous (2008) All India Coordinated Research Project on Linseed: Annual Report (2007-2008), Kanpur, UP, India

Anonymous (2010) Linseed: Annual Report, 2009-2010. Project Coordinating Unit (Linseed) C.S.A.U.A.& T., Kanpur, UP, India

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 residues in the delta 12 acyl-lipid desaturase of *Synechocystis*. FEBS Lett **361:** 111-114

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

Banilas G, Moressis A, Nikoloudakis N, Hatzopoulos P (2005) Spatial and temporal expressions of two distinct oleate desaturases from olive (*Olea europaea* L.). Plant Science **168:** 547-555

Barozai MYK (2012) In silico identification of micrornas and their targets in fiber and oil producing plant flax (*Linum Usitatissimum* L.). Pakistan Journal of Botany **44:** 1357-1362

Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Gupta VS (2012) Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns. BMC Genomics **13**

Behrouzian B, Savile CK, Dawson B, Buist PH, Shanklin J (2002) Exploring the hydroxylation-dehydrogenation connection: Novel catalytic activity of castor stearoyl-ACP Delta(9) desaturase. Journal of the American Chemical Society **124:** 3277-3283

Belo A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S, Rafalski A (2008) Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. Molecular Genetics and Genomics **279:** 1-10

Berberich T, Harada M, Sugawara K, Kodama H, Iba K, Kusano T (1998) Two maize genes encoding omega-3 fatty acid desaturase and their differential expression to temperature. Plant Molecular Biology **36:** 297-306

Berglund D (2002) Flax: New uses and demands. *In* J Janick, A Whipkey, eds, Trends in new crops and new uses. ASHS Press, Alexandria, VA, pp 358–360

Beuselinck PR, Sleper DA, Bilyeu KD (2006) An assessment of phenotype selection for linolenic acid using genetic markers. Crop Science **46:** 747-750

Bhatty R, Cherdkiatgumchai P (1990) Compositional analysis of laboratory prepared and commercial samples of linseed meal and of hull isolated from flax. Journal of the American Chemical Society **67:** 79-84

Bhella RS, Mackenzie SL (1995) Nucleotide-sequence of a cDNA from *Limnanthes Douglasii* L encoding a delta-15 linoleic-ccid desaturase. Plant Physiology **108:** 861-861

Bickel C, Gadani S, Lukacs M, Cullis C (2011) SSR markers developed for genetic mapping in flax (*Linum usitatissimum* L.). Res Rep Biol 2011**:** 23-29

Bilyeu KD, Palavalli L, Sleper DA, Beuselinck PR (2003) Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science **43:** 1833-1838

Borland VS (2003) Fiber focus. Textile World**:** 45

Browse J, McConn M, James D, Miquel M (1993) Mutants of *Arabidopsis* deficient in the dynthesis of alpha-linolenate - biochemical and genetic-characterization of the endoplasmic-reticulum linoleoyl desaturase. Journal of Biological Chemistry **268:** 16345-16351

Byfield GE, Xue H, Upchurch RG (2006) Two genes from soybean encoding soluble Delta 9 stearoyl-ACP desaturases. Crop Science **46:** 840-846

Byrum JR, Kinney AJ, Stecca KL, Grace DJ, Diers BW (1997) Alteration of the omega-3 fatty acid desaturase gene is associated with reduced linolenic acid in the A5 soybean genotype. Theoretical and Applied Genetics **94:** 356-359

Chappell AS, Bilyeu KD (2007) The low linolenic acid soybean line PI361088B contains a novel GmFAD3A mutation. Crop Science **47:** 1705-1710

Charrier B, Champion A, Henry Y, Kreis M (2002) Expression profiling of the whole *Arabidopsis* Shaggy-like kinase multigene family by real-time reverse transcriptase-polymerase chain reaction. Plant Physiology **130:** 577-590

Charters YM, Robertson A, Wilkinson MJ, Ramsay G (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L ssp *oleifera*) using 5'-anchored simple sequence repeat (SSR) primers. Theoretical and Applied Genetics **92:** 442-447

Chevreux B, Pfisterer T, Suhai S (2000) Automatic assembly and editing of genomic data. Genomics and Proteomics**:** 51-65

Chowdari KV, Venkatachalam SR, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998) Hybrid performance and genetic distance as revealed by the (GATA)(4) microsatellite and RAPD markers in pearl millet. Theoretical and Applied Genetics **97:** 163-169

Chung CH, Kim JL, Lee YC, Choi YL (1999) Cloning and characterization of a seed-specific omega-3 fatty acid desaturase cDNA from *Perilla frutescens*. Plant and Cell Physiology **40:** 114-118

Cloutier S, Miranda E, Ward K, Radovanovic N, Reimer E, Walichnowski A, Datla R, Rowland G, Duguid S, Ragupathy R (2012a) Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (*Linum usitatissimum* L.). Theoretical and Applied Genetics **125:** 685-694

Cloutier S, Niu Z, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theoretical and Applied Genetics **119:** 53-63

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

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

Cooke DT, Burden RS (1990) Lipid modulation of plasma membrane-bound ATPases. Physiologia Plantarum **78:** 153-159

Craig W, Lenzi P, Scotti N, De Palma M, Saggese P, Carbone V, Curran NM, Magee AM, Medgyesy P, Kavanagh TA, Dix PJ, Grillo S, Cardi T (2008) Transplastomic

tobacco plants expressing a fatty acid desaturase gene exhibit altered fatty acid profiles and improved cold tolerance. Transgenic Research **17:** 769-782

Creelman RA, Tierney ML, Mullet JE (1992) Jasmonic acid methyl jasmonate accumulate in wounded soybean hypocotyls and modulate wound gene-expression. Proceedings of the National Academy of Sciences of the United States of America **89:** 4938-4941

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

Cullis CA (1990) DNA rearrangements in response to environmental stress. Advances in Genetics **28:** 73–97

Cullis CA (2007) Flax *In* C Kole, ed, Genome mapping and molecular breeding in plants, oilseeds, Vol 2. Springer-Verlag, Berlin Heidelberg, pp 275-295

Cullis CA, Oh TJ, Gorman MB (1995) Genetic mapping in flax (*Linum usitatissimum*). *In* In: Proc3rd Meeting Int. Flax Breed Res Group, St Valery-en-caux, France, pp 161–169

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant Journal **38:** 366-379

Daun JK, DeClercq DR (1994) Sixty years of Canadian flaxseed quality surveys at the Grain Research Laboratory. Proc. Flax Inst.

Daun JK, Przybylski R (2000) Environmental effects on the composition of four Canadian flax cultivars. Proceedings of the 58th Flax Institute of the United States**:** 80-91

de Gomez D, Brenner R (1975) Oxidative desaturation of alpha-linolenic, linoleic, and stearic acids by human liver microsomes. Lipids **10:** 315-317

de Meaux J, Goebel U, Pop A, Mitchell-Olds T (2005) Allele-specific assay reveals functional variation in the chalcone synthase promoter of *Arabidiopsis thaliana* that is compatible with neutral evolution. Plant Cell **17:** 676-690

Deng X, Long SH, He DF, Li XA, Wang YF, Hao DM, Qiu CS, Chen XB (2011) Isolation and characterization of polymorphic microsatellite markers from flax (*Linum usitatissimum* L.). African Journal of Biotechnology **10:** 734-739

Deuticke B, Haest CWM (1987) Lipid modulation of transport proteins in vertebrate cell-membranes. Annual Review of Physiology **49:** 221-235

Diederichsen A (2001) Comparison of genetic diversity of flax (*Linum usitatissimum* L.) between Canadian cultivars and a world collection. Plant Breeding **120:** 360-362

Diederichsen A (2007) Ex situ collections of cultivated flax (*Linum usitatissimum* L.) and other species of the genus Linum L. Genetic Resources and Crop Evolution **54:** 661-678

Diederichsen A, Fu YB (2006) Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (*Linum usitatissimum* L. subsp usitatissimum). Genetic Resources and Crop Evolution **53:** 77-90

Diederichsen A, Hammer K (1995) Variation of cultivated flax (*Linum usitatissimum* L Subsp *Usitatissimum*) and Its Wild Progenitor Pale Flax (Subsp *Angustifolium* (Huds) Thell). Genetic Resources and Crop Evolution **42:** 263-272

Domier KW, Kerr N (2000) The potential for agricultural fibres. Proceedings of the 58th Flax Institute of the United States**:** 138-140

Durrant A (1976) Flax and linseed. *In* N Simmonds, ed, Evolution of crop plants. Longman, London, pp 190–193

Eaton SB, Konner M (1985) Paleolithic nutrition - a consideration of its nature and current implications. New England Journal of Medicine **312:** 283-289

Elladi V (1940) *Linum usitatissimum* (L.) Vav. consp. nov. – Len. (Russ.). *In* E Vul'f, N Vavilov, eds, Kul’turnaja flora SSSR, prjadil’nye [Flora of cultivated plants of the USSR, fiber plants], Vol 5. Sel'chozgiz, Leningrad, Moskva, pp 109-207

Everaert I, De Riek J, De Loose M, Van Waes J, Van Bockstaele E (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. Plant Varieties and Seeds **14:** 69-87

Falcone DL, Gibson S, Lemieux B, Somerville C (1994) Identification of a gene that complements an *Arabidopsis* mutant deficient in chloroplast omega-6 desaturase activity. Plant Physiology **106:** 1453-1459

Farmer EE (1994) Fatty-acid signaling in plants and their associated microorganisms. Plant Molecular Biology **26:** 1423-1437

Fenart S, Ndong YPA, Duarte J, Riviere N, Wilmer J, van Wuytswinkel O, Lucau A, Cariou E, Neutelings G, Gutierrez L, Chabbert B, Guillot X, Tavernier R, Hawkins S, Thomasset B (2010) Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray. BMC Genomics **11**

Fitzpatrick K (2006) Flax Facts, Ed 2

Fofana B, Cloutier S, Duguid S, Ching J, Rampitsch C (2006) Gene expression of stearoyl-ACP desaturase and delta 12 fatty acid desaturase 2 is modulated during seed development of flax (*Linum usitatissimum*). Lipids **41:** 705-712

Fofana B, Duguid S, Cloutier S (2004) Cloning of fatty acid biosynthetic genes beta-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 Science **166:** 1487-1496

Fox BG, Shanklin J, Somerville C, Munck E (1993) Stearoyl-acyl carrier protein delta-9 desaturase from *Ricinus-Communis* is a diiron-oxo protein. Proceedings of the National Academy of Sciences of the United States of America **90:** 2486-2490

Frankel OH, Brown AHD (1984) Plant genetic resources today: a critical appraisal. *In* JHW Holden, JT Williams, eds, Crop Genetic Resources: Conservation & Evaluation. George Allen & Urwin Ltd, London, pp 249-257

Fu YB (2005) Geographic patterns of RAPD variation in cultivated flax. Crop Science **45:** 1084-1091

Fu YB (2011) Genetic evidence for early flax domestication with capsular dehiscence. Genetic Resources and Crop Evolution **58:** 1119-1128

Fu YB, Allaby RG (2010) Phylogenetic network of Linum species as revealed by non-coding chloroplast DNA sequences. Genetic Resources and Crop Evolution **57:** 667-677

Fu YB, Diederichsen A, Allaby RG (2012) Locus-specific view of flax domestication history. Ecology and Evolution **2:** 139–152

Fu YB, Diederichsen A, Richards KW, Peterson G (2002a) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. Genetic Resources and Crop Evolution **49:** 167-174

Fu YB, Peterson G, Diederichsen A, Richards KW (2002b) RAPD analysis of genetic relationships of seven flax species in the genus *Linum* L. Genetic Resources and Crop Evolution **49:** 253-259

Fu YB, Peterson GW (2010) Characterization of expressed sequence tag-derived simple sequence repeat markers for 17 *Linum* species. Botany-Botanique **88:** 537-543

Fu YB, Rowland GG, Duguid SD, Richards KW (2003) RAPD analysis of 54 north American flax cultivars. Crop Science **43:** 1510-1515

Fukuchi-Mizutani M, Tasaka Y, Tanaka Y, Ashikari T, Kusumi T, Murata N (1998) Characterization of Delta 9 acyl-lipid desaturase homologues from *Arabidopsis thaliana*. Plant and Cell Physiology **39:** 247-253

Fulco AJ (1974) Metabolic alterations of fatty-acids. Annual Review of Biochemistry **43:** 215-241

Gassan J, Bledzki AK (1997) Dynamic-mechanical properties of natural fiber-reinforced plastics: The effect of coupling agents. Fourth International Conference on Woodfiber-Plastic Composites**:** 76-80

Ghislain M, Zhang D, Fajardo D, Huaman Z, Hijmans RJ (1999) Marker-assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. Genetic Resources and Crop Evolution **46:** 547-555

Ghosh R, Paul S, Ghosh SK, Roy A (2009) An improved method of DNA isolation suitable for PCR-based detection of begomoviruses from jute and other mucilaginous plants. Journal of Virological Methods **159:** 34-39

Gibson S, Arondel V, Iba K, Somerville C (1994) Cloning of a temperature-regulated gene encoding a chloroplast omega-3 desaturase from *Arabidopsis thaliana*. Plant Physiology **106:** 1615-1621

Gill KS (1987) Linseed. Indian Council of Agricultural Research, New Delhi, India

Gill KS, Yermanos DM (1967) Cytogenetic studies on genus *Linum* L. hybrids among taxa with 15 as haploid chromosome number. Crop Science **7:** 623-627

Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research **40:** D1178-D1186

Graber R, Sumida C, Nunez EA (1994) Fatty-acids and cell signal-transduction. Journal of Lipid Mediators and Cell Signalling **9:** 91-116

Gutierrez L, Conejero G, Castelain M, Guenin S, Verdeil JL, Thomasset B, Van Wuytswinkel O (2006) Identification of new gene expression regulators specifically expressed during plant seed maturation. Journal of Experimental Botany **57:** 1919-1932

Guy JE, Whittle E, Kumaran D, Lindqvist Y, Shanklin J (2007) The crystal structure of the ivy Delta4-16:0-ACP desaturase reveals structural details of the oxidized active site and potential determinants of regioselectivity. J Biol Chem **282:** 19863-19871

Hamada T, Kodama K, Nishimura M, Iba K (1994) Cloning of a cDNA-encoding tobacco omega-3-fatty-acid desaturase. Gene **147:** 293-294

Hamilton I (1986) Textiles 15, 30,

Hammer K (1986) Linaceae. *In* J Schultze-Motel, ed, Rudolf Mansfelds Verzeichnis landwirtschaftlicher und gärtnerischer Kulturpflanzen, Vol Bd. 2. Akademie-Verlag, Berlin, pp 710–713

Harlan JR (1971) Agricultural origins -centers and noncenters. Science **174:** 468-&

Harwood JL (1988) Fatty-Acid metabolism. Annual Review of Plant Physiology and Plant Molecular Biology **39:** 101-138

Heijenrath R, Peijs T (1996) Natural-fibre-mat-reinforced thermoplastic composites based on flax fibres and polypropylene. Advanced Composites Letters **5:** 81-85

Heinz E (1993) Biosynthesis of polyunsaturated fatty acids. *In* J T.S. Moore, ed, Lipid Metabolism in Plants. CRC Press, Boca Raton, FL, pp pp 33-90

Helbaek H (1959) Domestication of food plants in the old world. Science **130:** 365-372

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 Physiology **110:** 311-319

Hernandez ML, Mancha M, Martinez-Rivas JM (2005) Molecular cloning and characterization of genes encoding two microsomal oleate desaturases (FAD2) from olive. Phytochemistry **66:** 1417-1426

Hernandez ML, Padilla MN, Mancha M, Martinez-Rivas JM (2009) Expression analysis identifies FAD2-2 as the olive oleate desaturase gene mainly responsible for the linoleic

acid content in virgin olive oil. Journal of Agricultural and Food Chemistry **57:** 6199-6206

Hongtrakul V, Slabaugh MB, Knapp SJ (1998) A seed specific delta-12 oleate desaturase gene is duplicated, rearranged, and weakly expressed in high oleic acid sunflower lines. Crop Science **38:** 1245-1249

Horiguchi G, Kawakami N, Kusumi K, Kodama H, Iba K (1998) Developmental regulation of genes for microsome and plastid omega-3 fatty acid desaturases in wheat (*Triticum aestivum* L.). Plant and Cell Physiology **39:** 540-544

Horiguchi G, Kodama H, Nishimura M, Iba K (1996) Role of omega-3 fatty acid desaturases in the regulation of the level of trienoic fatty acids during leaf cell maturation. Planta **199:** 439-442

Horrobin DF (1990) Postviral fatigue syndrome, viral-infections in atopic eczema, and essential fatty-acids. Medical Hypotheses **32:** 211-217

Hoshino T, Takagi Y, Anai T (2010) Novel GmFAD2-1b mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with GmFAD2-1a mutant alleles. Breeding Science **60:** 419-425

Hourton-Cabassa C, Matos AR, Arrabaca J, Demandre C, Zachowski A, Moreau F (2009) Genetically modified *Arabidopsis thaliana* cells reveal the involvement of the mitochondrial fatty acid composition in membrane basal and uncoupling protein-mediated proton leaks. Plant and Cell Physiology **50:** 2084-2091

Hrádilova J, Rehulka P, Rehulkova H, Vrbova M, Griga M, Brzobohaty B (2010) Comparative analysis of proteomic changes in contrasting flax cultivars upon cadmium exposure. Electrophoresis **31:** 421-431

Hu XY, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of fad2 and fad3 allele-specific markers in canola (*Brassica napus* L.). Theoretical and Applied Genetics **113:** 497-507

Huang JC, Sun M (2000) Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. Theoretical and Applied Genetics **100:** 1050-1060

Huis R, Hawkins S, Neutelings G (2010) Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). BMC Plant Biology **10**

Iba K, Gibson S, Nishiuchi T, Fuse T, Nishimura M, Arondel V, Hugly S, Somerville C (1993) A gene encoding a chloroplast omega-3-fatty-acid desaturase complements alterations in fatty-acid desaturation and chloroplast copy number of the Fad7 mutant of *Arabidopsis thaliana*. Journal of Biological Chemistry **268:** 24099-24105

Im YJ, Han O, Chung GC, Cho BH (2002) Antisense expression of an *Arabidopsis* omega-3 fatty acid desaturase gene reduces salt/drought tolerance in transgenic tobacco plants. Molecules and Cells **13:** 264-271

Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiology **143:** 1467-1483

Jain RK, Thompson 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 Science **39:** 1696-1701

Jaworski JG, Clough RC, Barnum SR (1989) A cerulenin insensitive short chain 3-ketoacyl-acyl carrier protein synthase in *Spinacia oleracea* leaves. Plant Physiology **90:** 41-44

Jin UH, Lee JW, Chung YS, Lee JH, Yi YB, Kim YK, Hyung NI, Pyee JH, Chung CH (2001) Characterization and temporal expression of a omega-6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds. Plant Science **161:** 935-941

Joshi K, Lad S, Kale M, Patwardhan B, Mahadik SP, Patni B, Chaudhary A, Bhave S, Pandit A (2006) Supplementation with flax oil and vitamin C improves the outcome of Attention Deficit Hyperactivity Disorder (ADHD). Prostaglandins Leukotrienes and Essential Fatty Acids **74:** 17-21

Joshi SP, Bhave SG, Chowdari KV, Apte GS, Dhonukshe BL, Lalitha K, Ranjekar PK, Gupta VS (2001) Use of DNA markers in prediction of hybrid performance and heterosis for a three-line hybrid system in rice. Biochemical Genetics **39:** 179-200

Jung S, Swift D, Sengoku E, Patel M, Teule F, Powell G, Moore K, Abbott A (2000) The high oleate trait in the cultivated peanut [Arachis hypogaea L.]. I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. Molecular and General Genetics **263:** 796-805

Kachroo A, Kachroo P (2009) Fatty acid-derived signals in plant defense. Annual Review of Phytopathology **47:** 153-176

Kachroo P, Shanklin J, Shah J, Whittle EJ, Klessig DF (2001) A fatty acid desaturase modulates the activation of defense signaling pathways in plants. Proceedings of the National Academy of Sciences of the United States of America **98:** 9448-9453

Kale SM, Pardeshi VC, Kadoo NY, Ghorpade PB, Jana MM, Gupta VS (2012) Development of genomic simple sequence repeat markers for linseed using next-generation sequencing technology. Molecular Breeding **30:** 597-606

Kang JL, Snapp AR, Lu CF (2011) Identification of three genes encoding microsomal oleate desaturases (FAD2) from the oilseed crop *Camelina sativa*. Plant Physiology and Biochemistry **49:** 223-229

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

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

Khodakovskaya M, McAvoy R, Peters J, Wu H, Li Y (2006) Enhanced cold tolerance in transgenic tobacco expressing a chloroplast omega-3 fatty acid desaturase gene under the control of a cold-inducible promoter. Planta **223:** 1090-1100

Kitts DD, Yuan YV, Wijewickreme AN, Thompson LU (1999) Antioxidant activity of the flaxseed lignan secoisolariciresinol diglycoside and its mammalian lignan metabolites enterodiol and enterolactone. Molecular and Cellular Biochemistry **202:** 91-100

Klubicova K, Bercak M, Danchenko M, Skultety L, Rashydov NM, Berezhna VV, Miernyk JA, Hajduch M (2011a) Agricultural recovery of a formerly radioactive area: I. Establishment of high-resolution quantitative protein map of mature flax seeds harvested from the remediated Chernobyl area. Phytochemistry **72:** 1308-1315

Klubicova K, Danchenko M, Skultety L, Berezhna VV, Hricova A, Rashydav NM, Hajduch M (2011b) Agricultural recovery of a formerly radioactive area: II. Systematic proteomic characterization of flax seed development in the remediated Chernobyl area. Journal of Proteomics **74:** 1378-1384

Klubicova K, Danchenko M, Skultety L, Miernyk JA, Rashydov NM, Berezhna VV, Pret'Ova A, Hajduch M (2010) Proteomics analysis of flax grown in Chernobyl area suggests limited effect of contaminated environment on seed proteome. Environmental Science & Technology **44:** 6940-6946

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 Physiology **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. Proceedings of the National Academy of Sciences of the United States of America **89:** 2624-2628

Kodama H, Hamada T, Horiguchi G, Nishimura M, Iba K (1994) Genetic enhancement of cold tolerance by expression of a gene for chloroplast omega-3-fatty-acid desaturase in transgenic tobacco. Plant Physiology **105:** 601-605

Kodama H, Horiguchi G, Nishiuchi T, Nishimura M, Iba K (1995) Fatty-acid desaturation during chilling acclimation is one of the factors involved in conferring low-temperature tolerance to young tobacco-leaves. Plant Physiology **107:** 1177-1185

Kovach W (1998) Multivariate statistical package for Windows. *In*, Ed 3.0. Kovach company services, Wales, United Kingdom

Krasowska A, Dziadkowiec D, Polinceusz A, Plonka A, Lukaszewicz M (2007) Cloning of flax oleic fatty acid desaturase and its expression in yeast. Journal of the American Oil Chemists Society **84:** 809-816

Kris-Etherton PM, Pearson TA, Wan Y, Hargrove RL, Moriarty K, Fishell V, Etherton T (1999) High-monounsaturated fatty acid diets lower both plasma cholesterol and triacylglycerol concentrations. American Journal of Clinical Nutrition **70:** 1009-1015

Lands WEM (1991) Biosynthesis of prostaglandins. Annual Review of Nutrition **11:** 41-60

Lanza LLB, deSouza CL, Ottoboni LMM, Vieira MLC, deSouza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theoretical and Applied Genetics **94:** 1023-1030

Leaf A, Weber PC (1987) A new era for science in nutrition. American Journal of Clinical Nutrition **45:** 1048-1053

Lemay A, Dodin S, Kadri N, Jacques H, Forest JC (2002) Flaxseed dietary supplement versus hormone replacement therapy in hypercholesterolemic menopausal women. Obstetrics and Gynecology **100:** 495-504

Lepsch D, Horal J (1998) Development of an integrated modular plastic electrical carrier and flax/ polypropylene shelf panel for a vehicle rear shelf system. Plastics: Components, processes, and technology (SP-1340), SAE Technical Paper Series 980727.

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

Li ZL, Nelson RL (2001) Genetic diversity among soybean accessions from three countries measured by RAPDs. Crop Science **41:** 1337-1347

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:** 1451-1452

Lindqvist Y, Huang W, Schneider G, Shanklin J (1996) Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. EMBO J **15:** 4081-4092

Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the FAD2-1 fatty acid desaturase 5 ' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). American Journal of Botany **88:** 92-102

Liu XY, Yang JH, Li B, Yang XM, Meng QW (2006) Antisense-mediated depletion of tomato chloroplast omega-3 fatty acid desaturase enhances thermal tolerance. Journal of Integrative Plant Biology **48:** 1096-1107

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

Lopez Y, Nadaf HL, Smith OD, Connell JP, Reddy AS, Fritz AK (2000) Isolation and characterization of the delta(12)-fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. Theoretical and Applied Genetics **101:** 1131-1138

Los DA, Murata N (1998) Structure and expression of fatty acid desaturases. Biochimica Et Biophysica Acta-Lipids and Lipid Metabolism **1394:** 3-15

Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics **50:** 437-450

Manku MS, Horrobin DF, Huang YS, Morse N (1983) Fatty-acids in plasma and red-cell membranes in normal humans. Lipids **18:** 906-908

Mansby E, Diaz O, von Bothmer R (2000) Preliminary study of genetic diversity in Swedish flax (*Linum usitatissimum*). Genetic Resources and Crop Evolution **47:** 417-424

Martinez-Rivas JM, Sperling P, Luhs W, Heinz E (2001) Spatial and temporal regulation of three different microsomal oleate desaturase genes (FAD2) from normal-type and high-oleic varieties of sunflower (*Helianthus annuus* L.). Molecular Breeding **8:** 159-168

McCartney AW, Dyer JM, Dhanoa PK, Kim PK, Andrews DW, McNew JA, Mullen RT (2004) Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. Plant Journal **37:** 156-173

McConn M, Browse J (1996) The critical requirement for linolenic acid is pollen development, not photosynthesis, in an arabidopsis mutant. Plant Cell **8:** 403-416

McConn M, Browse J (1998) Polyunsaturated membranes are required for photosynthetic competence in a mutant of *Arabidopsis*. Plant Journal **15:** 521-530

McConn M, Hugly S, Browse J, Somerville C (1994) A mutation at the fad8 locus of *Arabidopsis* identifies a 2nd chloroplast omega-3 Desaturase. Plant Physiology **106:** 1609-1614

McKhann HI, Gery C, Berard A, Leveque S, Zuther E, Hincha DK, De Mita S, Brunel D, Teoule E (2008) Natural variation in CBF gene sequence, gene expression and freezing tolerance in the Versailles core collection of *Arabidopsis thaliana*. Bmc Plant Biology **8**

Messmer MM, Melchinger AE, Boppenmaier J, Herrmann RG, Brunklaus-Jung E (1992) RFLP analyses of early-maturing European maize germ plasm I. Genetic diversity among flint and dent inbreds. Theoretical and Applied Genetics **83:** 1003-1012

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

Miquel MF, Browse JA (1994) High-oleate oilseeds fail to develop at low-temperature. Plant Physiology **106:** 421-427

Mori TA (2006) Omega-3 fatty acids and hypertension in humans. Clin Exp Pharmacol Physiol **33:** 842-846

Morris DH (2007) Fax - A health and nutrition primer. *In*, Ed Fourth Edition. Flax Council of Canada

Moss T, Cullis C (2012) Computational prediction of candidate microRNAs and their targets from the completed *Linum ussitatissimum* genome and EST Database. Journal of Nucleic Acids Investigation volume **3:e2**

Mozaffarian D (2005) Does alpha-linolenic acid intake reduce the risk of coronary heart disease? A review of the evidence. Alternative Therapies in Health and Medicine **11:** 24-30

Muir A, Westcott N (2003) Flax, the genus *Linum*. Taylor and Francis, London

Murakami Y, Tsuyama M, Kobayashi Y, Kodama H, Iba K (2000) Trienoic fatty acids and plant tolerance of high temperature. Science **287:** 476-479

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

Muravenko OV, Lemesh VA, Samatadze TE, Amosova AV, Grushetskaya ZE, Popov KV, Semenova OY, Khotyuleva LV, Zelenin AV (2003) Genome comparisons with chromosomal and molecular markers for three closely related flax species and their hybrids. Russian Journal of Genetics **39:** 414-421

Murphy DJ (1999) Production of novel oils in plants. Current Opinion in Biotechnology **10:** 175-180

Napier JA, Michaelsont LV, Keith Stobart A (1999) Plant desaturases: harvesting the fat of the land. Current Opinion in Plant Biology **2:** 123-127

Narisawa T, Fukaura Y, Yazawa K, Ishikawa C, Isoda Y, Nishizawa Y (1994) Colon-cancer prevention with a small amount of dietary *Perilla* oil high in alpha-linolenic acid in an animal-Model. Cancer **73:** 2069-2075

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89:** 583-590

Neutelings G, Fenart S, Lucau-Danila A, Hawkins S (2012) Identification and characterization of miRNAs and their potential targets in flax. Journal of Plant Physiology **169:** 1754-1766

Nishida I, Beppu T, Matsuo T, Murata N (1992) Nucleotide-sequence of a cDNA clone encoding a precursor to stearoyl-(acyl-carrier-protein) desaturase from spinach, *Spinacia-Oleracea*. Plant Molecular Biology **19:** 711-713

Nishiuchi T, Iba K (1998) Roles of plastid omega-3 fatty acid desaturases in defense response of higher plants. Journal of Plant Research **111:** 481-486

Nishiuchi T, Nishimura M, Arondel V, Iba K (1994) Genomic nucleotide-sequence of a gene encoding a microsomal omega-3-fatty-acid desaturase from *Arabidopsis thaliana*. Plant Physiology **105:** 767-768

O'Quin JB, Bourassa L, Zhang D, Shockey JM, Gidda SK, Fosnot S, Chapman KD, Mullen RT, Dyer JM (2010) Temperature-sensitive post-translational regulation of plant omega-3 fatty-acid desaturases is mediated by the endoplasmic reticulum-associated degradation pathway. Journal of Biological Chemistry **285:** 21781-21796

Oh TJ, Gorman M, Cullis CA (2000) RFLP and RAPD mapping in flax (*Linum usitatissimum*). Theoretical and Applied Genetics **101:** 590-593

Ohlrogge J (1999) Plant metabolic engineering: are we ready for phase two? Current Opinion in Plant Biology **2:** 121-122

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

Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. Annual Review of Plant Physiology and Plant Molecular Biology **48:** 109-136

Okuley J, Lightner J, Feldmann K, Yadav N, Lark E, Browse J (1994) *Arabidopsis* fad2 gene encodes the enzyme that is essential for polyunsaturated lipid-synthesis. Plant Cell **6:** 147-158

Pan A, Yu DX, Demark-Wahnefried W, Franco OH, Lin X (2009) Meta-analysis of the effects of flaxseed interventions on blood lipids. American Journal of Clinical Nutrition **90:** 288-297

Pantalone VR, Rebetzke GJ, Burton JW, Wilson RF (1997) Genetic regulation of linolenic acid concentration in wild soybean *Glycine soja* accessions. Journal of the American Oil Chemists Society **74:** 159-163

Papadopoulos AN, Hague JRB (2003) The potential for using flax (*Linum usitatissimum* L.) shiv as a lignocellulosic raw material for particleboard. Industrial Crops and Products **17:** 143-147

Patel M, Jung S, Moore K, Powell G, Ainsworth C, Abbott A (2004) High-oleate peanut mutants result from a MITE insertion into the FAD2 gene. Theoretical and Applied Genetics **108:** 1492-1502

Paz MM, Veilleux RE (1997) Genetic diversity based on randomly amplified polymorphic DNA (RAPD) and its relationship with the performance of diploid potato hybrids. Journal of the American Society for Horticultural Science **122:** 740-747

Pellizzon MA, Billheimer JT, Bloedon LT, Szapary PO, Rader DJ (2007) Flaxseed reduces plasma cholesterol levels in hypercholesterolemic mouse models. Journal of the American College of Nutrition **26:** 66-75

Pena-Cortes H, Albrecht T, Prat S, Weiler EW, Willmitzer L (1993) Aspirin prevents wound-induced gene-expression in tomato leaves by blocking jasmonic acid biosynthesis. Planta **191:** 123-128

Pfaffl M (2005) Quantification strategies in real-time PCR. *In* B SA, ed, A-Z of quantitative PCR. International University Line, LaJolla, pp 87-112

Pham AT, Lee JD, Shannon JG, Bilyeu KD (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biol **10:** 195

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

Pirtle IL, Kongcharoensuntorn W, Nampaisansuk M, Knesek JE, Chapman KD, Pirtle RM (2001) Molecular cloning and functional expression of the gene for a cotton delta-12 fatty acid desaturase (FAD-2). Biochimica Et Biophysica Acta-Gene Structure and Expression **1522:** 122-129

POS (1994) Fatty acid analyses. POS Pilot Plant Corporation, Saskatoon, SK

Post-Beittenmiller D, Jaworski JG, Ohlrogge JB (1991) In vivo pools of free and acylated acyl carrier proteins in spinach. Evidence for sites of regulation of fatty acid biosynthesis. Journal of Biological Chemistry **266:** 1858-1865

Post-beittenmiller D, Roughan G, Ohlrogge JB (1992) Regulation of plant fatty-acid biosynthesis - analysis of acyl-coenzymeA and acyl-acyl carrier protein substrate pools in spinach and pea chloroplasts. Plant Physiology **100:** 923-930

Rachinskaya OA, Lemesh VA, Muravenko OV, Yurkevich OY, Guzenko EV, Bol'sheva NL, Bogdanova MV, Samatadze TE, Popov KV, Malyshev SV, Shostak NG, Heller K, Hotyleva LV, Zelenin AV (2011) Genetic polymorphism of flax *Linum usitatissimum* based on the use of molecular cytogenetic markers. Russian Journal of Genetics **47:** 56-65

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**

Rahman SM, Takagi Y (1997) Inheritance of reduced linolenic acid content in soybean seed oil. Theoretical and Applied Genetics **94:** 299-302

Rahman SM, Takagi Y, Kumamaru T (1996) Low linolenate sources at the fan locus in soybean lines M-5 and IL-8. Breeding Science **46:** 155-158

Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters **339:** 62-66

Reed DW, Schafer UA, Covello PS (2000) Characterization of the *Brassica napus* extraplastidial linoleate desaturase by expression in *Saccharomyces cerevisiae*. Plant Physiology **122:** 715-720

Riaz A, Li G, Quresh Z, Swati MS, Quiros CF (2001) Genetic diversity of oilseed *Brassica napus* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. Plant Breeding **120:** 411-415

Richards E, Reichardt M, Rogers S (1994) Preparation of genomic DNA from plant tissues. *In* F Ausubel, R Brent, R Kingston, D Moore, J Seidman, J Smith, K Struhl, eds, Current Protocols in Molecular Biology. John Wiley and Sons, New York, pp 2.3.1-2.3.7

Roach MJ, Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. Mol Genet Genomics **278:** 149-165

Roach MJ, Deyholos MK (2008) Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. Annals of Botany **102:** 317-330

Roose-Amsaleg C, Cariou-Pham E, Vautrin D, Tavernier R, Solignac M (2006) Polymorphic microsatellite loci in *Linum usitatissimum*. Molecular Ecology Notes **6:** 796-799

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols **5:** 725-738

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

Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, Moorman AFM (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Research **37**

Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. Plant Cell **14:** 1191-1206

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: A laboratory manual, Ed second. Cold Spring Harbor Laboratory Press Plainview, New York

Sant VJ, Patankar AG, Sarode ND, Mhase LB, Sainani MN, Deshmukh RB, Ranjekar PK, Gupta VS (1999) Potential of DNA markers in detecting divergence and in analysing heterosis in Indian elite chickpea cultivars. Theoretical and Applied Genetics **98:** 1217-1225

Scheffler JA, Sharpe AG, Schmidt H, Sperling P, Parkin IAP, Luhs W, Lydiate DJ, Heinz E (1997) Desaturase multigene families of *Brassica napus* arose through genome duplication. Theoretical and Applied Genetics **94:** 583-591

Scheideler S, Lewis N (1997) Omega eggs: A dietary source of N-3 fatty acids. Neb. Facts**:** NF97-354

Schlueter JA, Vaslenko-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 paleoplyploid genome. Crop Science **47:** S14-S26

Schmid KM, and Ohlrogge, J.B. (2002) Lipid metabolism in plants. *In* DE Vance, ed, Biochemistry of lipids, lipoproteins and membranes, Ed 4th Elsevier Sci. and Technol, pp 93- 126

Schmidt H, Dresselhaus T, Buck F, Heinz E (1994) Purification and PCR-based cDNA coning of a plastidial N-6 desaturase. Plant Molecular Biology **26:** 631-642

Schuler I, Milon A, Nakatani Y, Ourisson G, Albrecht AM, Benveniste P, Hartmann MA (1991) Differential-effects of plant sterols on water permeability and on acyl chain ordering of soybean phosphatidylcholine bilayers. Proceedings of the National Academy of Sciences of the United States of America **88:** 6926-6930

Serrano-Vega MJ, Venegas-Caleron M, Garces R, Martinez-Force E (2003) Cloning and expression of fatty acids biosynthesis key enzymes from sunflower (*Helianthus annuus* L.) in *Escherichia coli*. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **786:** 221-228

Shah J (2005) Lipids, lipases, and lipid-modifying enzymes in plant disease resistance. Annual Reviews of Phytopathology **43:** 229-260

Shanklin J, Cahoon EB (1998) Desaturation and related modifications of fatty acids. Annual Review of Plant Physiology and Plant Molecular Biology **49:** 611-641

Shanklin J, Guy JE, Mishra G, Lindqvist Y (2009) Desaturases: Emerging models for understanding functional diversification of diiron-containing enzymes. Journal of Biological Chemistry **284:** 18559-18563

Shanklin J, Somerville C (1991) Stearoyl-acyl-carrier-protein desaturase from higher-plants is structurally unrelated to the animal and fungal homologs. Proceedings of the National Academy of Sciences of the United States of America **88:** 2510-2514

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

Sharma HS (1992) The biology and processing of flax. M. Publications, Belfast, Northern Ireland

Shimakata T, Stumpf PK (1983b) Purification and characterization of beta-ketoacyl-ACP synthetase-I from *Spinacia oleracea* leaves. Archives of Biochemistry and Biophysics **220:** 39-45

Shimakata T, Stumpf PK (1983a) The purification and function of acetyl coenzyme-A - acyl carrier protein transacylase. Journal of Biological Chemistry **258:** 3592-3598

Simopoulos A (1995) Evolutionary aspects of diet: fatty acids, insulin resistance and obesity. *In* T VanItallie, A Simopoulos, eds, Obesity: new directions in assessment and management. Charles Press, Philadelphia, pp 241–261

Simopoulos AP (1991) Omega-3-fatty-acids in health and disease and in growth and development. American Journal of Clinical Nutrition **54:** 438-463

Simopoulos AP (2002) Omega-3 fatty acids in inflammation and autoimmune diseases. Journal of the American College of Nutrition **21:** 495-505

Singh S, Mckinney S, Green A (1994) Sequence of a cDNA from *Linum usitatissimum* encoding the stearoyl-acyl carrier protein desaturase. Plant Physiology **104:** 1075-1075

Slocombe SP, Piffanelli P, Fairbairn D, Bowra S, Hatzopoulos P, Tsiantis M, Murphy DJ (1994) Temporal and tissue-specific regulation of a Brassica napus stearoyl-acyl carrier protein desaturase gene. Plant Physiology **104:** 1167-1176

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L): comparisons with data from RFLPS and pedigree. Theoretical and Applied Genetics **95:** 163-173

Smýkal P, Bacova-Kerteszova N, Kalendar R, Corander J, Schulman AH, Pavelek M (2011) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. Theoretical and Applied Genetics **122:** 1385-1397

Somerville C, Browse J (1991) Plant lipids - metabolism, mutants, and membranes. Science **252:** 80-87

Soto-Cerda BJ, Saavedra HU, Navarro CN, Ortega PM (2011) Characterization of novel genic SSR markers in *Linum usitatissimum* (L.) and their transferability across eleven Linum species. Electronic Journal of Biotechnology **14**

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

Spielmeyer W, Green AG, Bittisnich D, Mendham N, Lagudah ES (1998) Identification of quantitative trait loci contributing to *Fusarium* wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). Theoretical and Applied Genetics **97:** 633-641

Sumida C, Graber R, Nunez E (1993) Role of fatty-acids in signal transduction -modulators and messengers. Prostaglandins Leukotrienes and Essential Fatty Acids **48:** 117-122

Tammes T (1928) The genetics of the genus *Linum*. Bibliographica Genetica **4:** 1–36

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28:** 2731-2739

Tang GQ, Novitzky WP, Griffin HC, Huber SC, Dewey RE (2005) Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. Plant Journal **44:** 433-446

Tang SY, Guan RZ, Zhang HS, Huang J (2007) Cloning and expression analysis of three cDNAs encoding omega-3 fatty acid desaturases from Descurainia sophia. Biotechnology Letters **29:** 1417-1424

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science **277:** 1063-1066

Teixeira MC, Carvalho IS, Brodelius M (2010) Omega-3 fatty acid desaturase genes isolated from purslane (*Portulaca oleracea* L.): Expression in different tissues and response to cold and wound stress. Journal of Agricultural and Food Chemistry **58:** 1870-1877

Thompson GA, Scherer DE, Aken SF, Kenny JW, Young HL, Shintani DK, Kridl JC, Knauf VC (1991) Primary structures of the precursor and mature forms of stearoyl-acyl carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme-activity. Proceedings of the National Academy of Sciences of the United States of America **88:** 2578-2582

Thompson L (2003) Analysis and bioavailability of lignans. *In* L Thompson, S Cunnane, eds, Flaxseed in Human Nutrition, Ed 2nd. AOCS Press, Champaign, IL, pp 92-116

Thompson LU, Rickard SE, Cheung F, Kenaschuk EO, Obermeyer WR (1997) Variability in anticancer lignan levels in flaxseed. Nutrition and Cancer-an International Journal **27:** 26-30

Tocher DR, Leaver MJ, Hodgson PA (1998) Recent advances in the biochemistry and molecular biology of fatty acyl desaturases. Progress in Lipid Research **37:** 73-117

Todoroki S, Hayashi T, Nagata T, Kanegae H, Mori M, Kikuchi S (1998) cDNA cloning of gamma-ray inducible genes encoding ω-3 fatty acid desaturase from potato tuber. Plant Cell Physiology **39:** s136

Tong L, Shu-Ming P, Wu-Yuan D, Dan-Wei M, Ying X, Meng X, Fang C (2006) Characterization of a new stearoyl-acyl carrier protein desaturase gene from *Jatropha curcas*. Biotechnology Letters **28:** 657-662

van Treuren R, van Soest LJM, van Hintum T (2001) Marker-assisted rationalisaton of genetic resource collections: a case study in flax using AFLPs. Theoretical and Applied Genetics **103:** 144-152

Udvardi MK, Czechowski T, Scheible WR (2008) Eleven golden rules of quantitative RT-PCR. Plant Cell **20:** 1736-1737

Uysal H, Fu YB, Kurt O, Peterson GW, Diederichsen A, Kusters P (2010) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) and its wild progenitor pale flax (*Linum bienne* Mill.) as revealed by ISSR markers. Genetic Resources and Crop Evolution **57:** 1109-1119

van Dam J, van Vilsteren G, Zomers F, Shannon W, Hamilton I (1994) Increased application of domestically produced plant fibers in textiles, pulp and paper production and composite materials. Directorate-General XII, Science, Research, and Development, European Commission.

van de Loo FJ, Somerville C (1994) Plasmid omega-3-fatty-acid desaturase cDNA from *Ricinus communis*. Plant Physiology **105:** 443-444

van Zeist W, Bakker-Heeres J (1975) Evidence of linseed cultivation before 6000 BC. Journal of Archaeological Science **2:** 215–219

Vavilov N (1926) Studies on the origin of cultivated plants. *In* Bull Appl Bot, Vol 26, USSR, Leningrad

Venglat P, Xiang DQ, Qiu SQ, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biology **11**

Vittal K, Kerkhi S, Chary GR, Sankar GM, Ramakrishna Y, Srijaya T, Samra J (2005) District based promising technologies for rainfed linseed based production system in India. . All India Co-ordinated Research Project for Dryland Agriculture, Central Research Institute for Dryland Agriculture, Indian Council of Agricultural Research, Hyderabad

von Kulpa W, Danert S (1962) Zur Systematik von *Linum usitatissimum* L. Die Kulturpflanze, Vol 3. Akademie Verlag, Berlin

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

Wada M, DeLong CJ, Hong YH, Rieke CJ, Song I, Sidhu RS, Yuan C, Warnock M, Schmaier AH, Yokoyama C, Smyth EM, Wilson SJ, FitzGerald GA, Garavito M, Sui DX, Regan JW, Smith WL (2007) Enzymes and receptors of prostaglandin pathways with arachidonic acid-derived versus eicosapentaenoic acid-derived substrates and products. Journal of Biological Chemistry **282:** 22254-22266

Wallis JG, Browse J (2002) Mutants of *Arabidopsi*s reveal many roles for membrane lipids. Prog Lipid Res **41:** 254-278

Wang JW, Ming F, Pittman J, Han YY, Hu J, Guo B, Shen DL (2006) Characterization of a rice (*Oryza sativa* L.) gene encoding a temperature-dependent chloroplast omega-3 fatty acid desaturase. Biochemical and Biophysical Research Communications **340:** 1209-1216

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 GKS, Wang J, Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. The Plant Journal **72:** 461-473

Watahiki MK, Yamamoto KT (1994) A new isozyme of plastid omega-3-fatty-acid desaturase in *Arabidopsis thaliana*. Plant Physiology **105:** 1451-1452

Wetton JH, Carter RE, Parkin DT, Walters D (1987) Demographic-study of a wild house sparrow population by DNA fingerprinting. Nature **327:** 147-149

Whittle E, Cahoon EB, Subrahmanyam S, Shanklin J (2005) A multifunctional acyl-acyl carrier protein desaturase from *Hedera helix* L. (English ivy) can synthesize 16-and 18-carbon monoene and diene products. Journal of Biological Chemistry **280:** 28169-28176

Whittle E, Shanklin J (2001) Engineering delta 9-16:0-acyl carrier protein (ACP) desaturase specificity based on combinatorial saturation mutagenesis and logical redesign of the castor delta 9-18:0-ACP desaturase. Journal of Biological Chemistry **276:** 21500-21505

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research **35:** 407-410

Wiesner I, Wiesnerova D (2003) Insertion of a reamplification round into the ISSR-PCR protocol gives new flax fingerprinting patterns. Cellular & Molecular Biology Letters **8:** 743-748

Wiesner I, Wiesnerova D (2004a) Statistical correlations of primer thermodynamic stability DeltaG degrees for enhanced flax ISSR-PCR cultivar authentication. Journal of Agricultural and Food Chemistry **52:** 2568-2571

Wiesnerova D, Wiesner I (2004b) ISSR-based clustering of cultivated flax germplasm is statistically correlated to thousand seed mass. Molecular Biotechnology **26:** 207-214

Williams D, Verghese M, Walker LT, Boateng J, Shackelford L, Chawan CB (2007) Flax seed oil and flax seed meal reduce the formation of aberrant crypt foci (ACF) in azoxymethane-induced colon cancer in Fisher 344 male rats. Food and Chemical Toxicology **45:** 153-159

Wolever T, Jenkins D (1993) Effects of fiber and foods on carbohydrate metabolism. *In* G Spiller, ed, Dietary fiber in human nutrition, Ed 2nd CRC Press, Boca Raton, FL, pp 111-152

Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD (1996) Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theoretical and Applied Genetics **92:** 637-643

Yadav NS, Wierzbicki A, Aegerter M, Caster CS, Perezgrau L, Kinney AJ, Hitz WD, Booth 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 omega-3-fatty-acid desaturases. Plant Physiology **103:** 467-476

Yap I, Nelson R (1996) WINBOOT: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. *In* IRRI Discussion Paper Series No. 14. International Rice Research Institute

Yi JX, Derynck MR, Chen L, Dhaubhadel S (2010) Differential expression of CHS7 and CHS8 genes in soybean. Planta **231:** 741-753

Zhang DY, 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 Physiology and Biochemistry **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**

Zhang M, Barg R, Yin MG, Gueta-Dahan Y, Leikin-Frenkel A, Salts Y, Shabtai S, Ben-Hayyim G (2005) Modulated fatty acid desaturation via overexpression of two distinct omega-3 desaturases differentially alters tolerance to various abiotic stresses in transgenic tobacco cells and plants. Plant Journal **44:** 361-371

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9:40**

Zhao G, Etherton TD, Martin KR, Gillies PJ, West SG, Kris-Etherton PM (2007) Dietary alpha-linolenic acid inhibits proinflammatory cytokine production by peripheral blood mononuclear cells in hypercholesterolemic subjects. American Journal of Clinical Nutrition **85:** 385-391

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored Polymerase Chain-Reaction amplification. Genomics **20:** 176-183

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford University Press, Oxford

Zohary D, Hopf M, Weiss E (2012) Domestication of plants in the old world: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin, Ed 4. Oxford University Press

ANNEXURE I

SAD1 gene sequences

> PADMINI (JQ963139)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> NL97 (JQ963140)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> NL260 (JQ963141)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> EC9825 (JQ963142)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT

CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> SURABHI (JQ963143)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATCTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> AYOGI (JQ963144)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> ES44 (JQ963145)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> JRF5 (JQ963146)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC

GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> ACC. NO. 4/47 (JQ963147)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATCTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

> TL23 (JQ963148)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCAACACGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACCTCC
ACCAAGTAAGCATCTCCTCCTCCTCGGAATCTCCGCCGATTTCTTTTAAGCGATTGATCGTAGATAAATTTGTCGGTTGCTTACCGTTCATCAAAATCTGCACGGTTCGTTTCTTCT
TCTGCGCCTAGATTGCATTATGTCATTGTTCGTTTTCCGATTTGACTGACCGACATAAATCAATTCCTTTGTGTTTCACGATTCTGGGTTTTGCGCTGTAATTGATTGTCAGTGTTT
GCACAGGTTTCCCCTTCTCCTCCTCCGTCCATCAAATGCATGTTATTACCATTTCAATTTCAGTTTCCTTCTCTGAAATATCCGTCTCTGGGAAAATAAGTCTCTGTATCTACTATC
CTATCAGCTTGTTTAGGAGAGGTTCGATATTCGTTTACATAAACCAATTGGCTTACAGTCCTTGAACGTTCTAAATGTTGGTCGCGGTGATAATAGGTTCTCAAAAGAGGTTTGTCT
ATGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTTAAGGTCCTTGGAAGAATGACTAATGAGCTATGACATGATTACGACGTAGTAGTTATTGAACTGCTGATAATTCAATAT
AGGGGTAACTTTGTTGATTGTTTGGTCACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGG
AGATATTTAAGTCTCTGGAAGGTTGGGCTGAGGATGTTCTATTACCGCACCTGAAGCCAGTTGAGAAATGCTGGCAGCCACAGGATTTCCTGCCCGAACCTGAGTCGGATGGGTTCG
AGGAGCAAGTGAAGGAGCTCAGGGCAAGGGCCAAAGAACTGCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTGCCGACTTACCAGACAATGCTCA
ACACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTAT
ACCTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTAATCACATACTTCATCCTTTTCTATTAATCTTTGCGTGAACAA
AATTCACTACACTGGTAGCAGCTGAAACTTTAGATGATTTTTTTTACTGCCTAGCTTCTATGAAACAAAACCACGTAAGTCAAATAGGGTTGACAATGAGTTCAAGTGGCAAAATTT
TTCTTATATACCAACTTCGAACCACTTTATATGACATACCAACTCCTAGTTCGGTTAAAATTCCTCCGTCGAAGATATAATACTTGGATTGGTTAAATGAATTGTGAAAGGATACAC
GTGATGTGGTCTGGAATTAATTTGTTTGAATGATCAGTTGGGTTCGGGGCGACAACTGTGAACTGGAACCACCCTAAGTAAATTTTCTTTCTGTCCTACAAATTTGAGGTTCTCCTT
GATCACCTTAGTCCATCTTAGGTTTGCCCGTTAGTAAGATCTGCATTTAGCAGTTTGTCCTGGTATTTGATATCACTAGTATCTTTGTTTGATTCCCTAGCATCTCTGAAACCATCG
GACAAGTAGGTGGTTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCCGTATCATCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAGCAAC
GGCTAGAAGAAGTTCTACATCTGATGCTTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTTCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTTTCAT
CTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAACACAGCCAGACTCGCCAAGGACCATGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAGACGA
GAAACGGCACGAAACCGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGCTGGCACTGGCGGACATGATGAGGAAGAAGATATCGATGCCCGCCCA
CTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCAGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCCGATATCCTGGAGTTCCTGGTGGGGAG
GTGGAAAGTGGATGCTTTTACGGGGCTTTCCGGGGAAGGGAACAAAGCTCAGGATTTTGTCTGCGGGCTTCCTGCGAGGATTCGAAAGTTGGAGGAGAGGGCTGCGGGGAGGGCAAA
GCAAACGTCGAAATCTGTCCCGTTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAATGGAGTTTGCTTGAGAGTTAGAGTGTGGAATGATTGGAGTATGA

SAD2 gene sequences

> PADMINI (JQ963149)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCCTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTATATGACATACCAACTTCTACTTTGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> NL97 (JQ963150)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCGTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG

GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTATATGACATACCAACTTCTACTTTGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> NL260 (JQ963151)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCGTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTAYATGACATACCAACTTCTACTTYGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> EC9825 (JQ963152)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCGTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTATATGACATACCAACTTCTACTTTGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> SURABHI (JQ963153)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCGTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTATATGACATACCAACTTCTACTTTGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> AYOGI (JQ963154)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCCTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG

ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTACATGACATACCAACTTCTACTTCGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> ES44 (JQ963155)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCCTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTACATGACATACCAACTTCTACTTCGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> JRF5 (JQ963156)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCCTCGTCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCCTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTATATGACATACCAACTTCTACTTTGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> ACC. NO. 4/47 (JQ963157)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCATCATCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCCTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG
GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTATATGACATACCAACTTCTACTTTGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACGGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CCCACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

> TL23 (JQ963158)

ATGGCTCTCAAGCTCAACCCAGTCACCACCTTCCCTTCGACCCGCTCCCTCAACAACTTCTCCTCCAGATCTCCTCGCACCTTTCTCATGGCTGCTTCCACTTTCAATTCCACTTCC
ACCAAGTAAGTTCCCGTCACCATCTCCTCTTCCTCGGAATCTCCGCCGTTTCATTTAAGCGATTGATCGTAGAAAATCTGTCGGTTGCTTAGCGTTCATTCAAATCTGCGCGGTTCG
TTTCTTTTTCTTTCTTCAGACTGCCTCGTCTGCATTATGTTATTGTTCGTTTCCGATTTGACTAACCTACATAATCAATTCCTTTGTGTTTCACGAGTCTGGATTTTGCGCTGTAAT
TGATTGTCAGCGTTTGCACAGGTTTCCATTTCTCCACCTCCGTCCATCAAATGCATGTTATTACCTACCAATTTCAGCGTCTTTCTCTGGAAATTTCTGTCTCTGTATCTACTATCC
TATTAGCTTGTTTGAGAGAGGTTCAATATTGGTTTGCATGAACCAAGTGGCTTACAATCCTTCAACGTTCTAAATGTTGGTCGCAGTAACAATAGGTTCTCAAAAGAGGTTTTTCTA
TGTTGTTTGGCAAAATCTTGTTTCTGTGAATCATGTTAAGGTCCTGGGAAGAATGATTAATGAGCTATGACATGATTAAGGCGTAGTAGTTATTGAACTGCTGATAATTCAATATAG

GGGTAACTTTGTTGGTTGTTTGGTGACAGGGAGGCTGAGAAGCTAAAGAAGTCACATGGACCACCAAAAGAGGTGCATATGCAAGTGACCCATTCCATGCCCCCACAGAAGCTGGAG
ATCTTTAAGTCCCTTGAAGGTTGGGCAGAGGACGTTCTGTTGCCGCACCTGAAGCCGGTTGAGAAATGCTGGCAGCCACAAGATTTCCTGCCCGAACCCGAGTCGGATGGGTTCGAG
GAGCAAGTGAAGGAGCTCAGGGCAAGGGCTAAAGAACTCCCCGATGACTATTTTGTTGTGCTGGTTGGGGATATGATCACCGAAGAAGCTCTACCGACTTACCAGACAATGCTCAAC
ACCCTTGACGGGGTGAGGGACGAGACTGGAGCCAGCCTTACGCCGTGGGCAATCTGGACAAGGGCGTGGACCGCTGAAGAGAATAGGCACGGTGACCTTCTCAACAAGTATCTTTAC
CTCTCTGGAAGGGTGGACATGAGGCAAATTGAAAAGACCATTCAGTATCTCATCGGCTCTGGAATGGTATGTACTCACATCCTATCTGCTCCTTTATCCTTTTCCATTAATCTTTGA
TTGAACAAAATTCAATAAACTGGTAGCTGAAACTTTAGATGATTTGTTATAACTGCCTAGCTTCTATGAGAAAACCACTGAAGTCAAATAGGTTTGACAATGGGTTTAAATGGAAAA
AGTTTCATATACCATCTTCCATCTATTTTACATGACATACCAACTTCTACTTCGGAGAAAATTCGCCGTGGATAATCATATTATTGAAGATATAGTACTTAGTAGATTGGTTAGATG
AACTGTTAAACAATACATGTGATGTCGTGTGCAATTAATTTGTGTAAATGATTAGCTGGGTTCGGGACGACAAATGTGAACTGGAACCCTAGTAAACTATGAATTGAGGTTGTCCTT
CATCACTTTATTCTGTCCTGGGTTTGTTTGCCTGTTTGCAAGATCTGCATGTAGCAGTTTGTCCTGGTATTTGCTACCAGTGGTATCTTTGTTTGATTCCCTAGCATCTCTGAAAAC
ATCGGACCAAGTATCTGGTTAGGACAAATTTGGTTCATTGCGGCATTTTTTGTTTGTATCGCTGTATCGTCTGGAAGAAGCAGACAGTTTTGCAAAGTGGCATCAAGCTCAAGAAAG
CAACGGCTAGAAGAAGTTCTACATCTGATGCGTTCCTTTTGTTTCTTTGTGTGCTTTTTGGACTTTGTTCTTTTTGCCTGTAGGATCCAAAAACAGAAAACAACCCCTACCTCGGTT
TCATCTACACCTCATTCCAAGAGAGGGCAACGTTCATCTCCCACGGAAATACGGCCAGACTCGCCAAGGACCACGGGGACATGAAGCTGGCGCAGATCTGCGGGATCATCGCAGCAG
ACGAGAAGCGGCACGAAACAGCATACACCAAGATCGTCGAGAAGCTCTTCGAGATCGACCCTGACAGTACAGTGTTGGCTCTGGCGGACATGATGAGGAAGAAGATATCGATGCCCG
CACACTTGATGTACGATGGAGAAGACGACAACCTCTTCGACAATTACTCGTCGGTCGCTCAACGCATCGGGGTGTATACTGCCAAGGATTATGCTGATATCCTGGAGTTCCTGGTGG
GGAGGTGGAAAGTGGATGCTTTTACGGGACTTTCCGGGGAAGGGAACAAAGCTCAGGAGTTTGTCTGTGGGCTTCCAGCGAGGATTCGAAAATTGGAGGAGAGGGCTGCGGGGAGGG
CAAAGCAAACGTCGAAATCTGTCCCATTCAGCTGGATCTTCAGCAGAGAATTGGTACTCTAA

FAD2 gene sequences

>PADMINI (JQ963109)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>NL97 (JQ963110)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>NL260 (JQ963111)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>EC9825 (JQ963112)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCATGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>SURABHI (JQ963113)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGTTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>AYOGI (JQ963114)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT

CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>ES44 (JQ963115)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>JRF5 (JQ963116)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>ACC. NO. 4/47 (JQ963117)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATCCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCACGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

>TL23 (JQ963118)

ATGGGTGCAGGTGGAAGAATGCCAGTGCCTCCATCATCCAAACCTATGAAGAGGTCTCCTTACTCAAAGCCACCATTCACGCTCGGTGAGCTCAAGAAGGCCATTCCTCCACACTGT
TTCAAACGCTCAATCCCCCGATCGTTCGCCTACGTGGCGTACGACCTCACCATTGCAGCAATCTTCTACTACATCGCCACCACTTACTTCCACCTCCTCCCTAGCCCTCTCAACTAC
CTCGCCTGGCCGGTCTACTGGGCCTGCCAGGGCTGCATCCTCACTGGAGTATGGGTGTTGGCTCACGAATGCGGTCACCATGCCTTCAGCGACTACCAGTGGCTCGACGACATGGTT
GGCTTCGTCCTCCATTCGTCCCTCCTTGTTCCTTACTTCTCCTGGAAGCACAGCCACCGCCGCCACCATTCCAACACGGGATCGCTTGATCGTGATGAGGTGTTTGTCCCCAAGCAG
AAGGCCGAAATCGGGTGGTACTCCAAGTACCTTAACAACCCACCTGGCCGTGTGATCACATTGGCCGTCACATTAACGCTCGGTTGGCCTCTGTACTTGGCATTCAACGTCTCCGGG
AGACCATATGACCGGTTCGCATGCCATTTTGACCCTCATGGTCCGATTTACAATGATCGCGAGCGTATGGAGATATACCTATCCGACGCAGGGATATTCACCGTGTGCTACATCCTA
TACAGACTCGTCCTCACGAAAGGACTCGTTTGGGTCGTGTCCATATACGGAGTCCCACTATTGATAGTGAATGGATTCCTAGTCCTCATCACTTTCTTGCAGCACACGCATCCTTCT
CTTCCGCACTACAAGTCCTCCGAATGGGACTGGATGCGAGGCGCCCTCTCGACCGTGGATCGAGACTACGGGTTACTCAACACCGTGTTCCACAACATCACCGATACACATGTCGCG
CACCATCTCTTCTCCACGATGCCTCATTACCACGCGATGGAGGCTACCAAGGCGATCAAGCCGGTTCTCGGGGAGTATTACCAGTTCGATGGGACTCCCTTTGTGAAGGCCATGTGG
AGGGAGGCAAAGGAGTGCATCTATGTCGAGCCGGATGAAGGCGACCCCAGCCAAGGCGTGTTCTGGTACAACAATAAGTTATGA

FAD2-2 gene sequences (Partial)

>PADMINI (JQ963159)

CCATCGAATAAGGCGGACTCCGAAACCTTTAAGCGGTCTCCTTACTCAAAACCTCCCTTCACTCTTGGTGAGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCC
CGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACTTACATCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTAC
TGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCCTTCAGCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCA
TGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCCCTCGAACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGG
CACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGCTGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTC
GCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCCGATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCC
AAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTCTTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCC
TCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACGGTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACG
ATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAGTACTACCAATTCGACGGGACTCCATT

>NL97 (JQ963160)

GGTGAGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACT
TACATCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCC
TTCAGCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCC
CTCGAACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGC
TGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCC
GATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTC
TTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACG
GTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAGTACT

>NL260 (JQ963161)

TTCACTCTTGGTGAGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATC
GCCACCACTTACATCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGT
CACCATGCCTTCAGCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAAT
ACTGGGTCCCTCGAACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTC
ACTCTCGGCTGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATA
TTCTTCTCCGATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGA
TTCCTTGTCTTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTT
CTGAACACGGTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAG
TACTACCAATTCGACGGGACTCCATTCATCAAGGCGATGTGGAGGG

>EC9824 (JQ963162)

GGCGGACTCCGAAACCTTTAAGCGGTCTCCTTACTCAAAACCTCCCTTCACTCTTGGTGAGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTC
CTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACTTACATCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCA
AGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCCTTCAGCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGT
ACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCCCTCGAACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTA
CCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGCTGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTA
CGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCCGATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGC
TTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTCTTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGA
CTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACGGTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTA
CCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAGTACTACCAATTCGACG

>SURABHI (JQ963163)

AAAACCTCCCTTCACTCTTGGTGAGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTT
CTACTACATCGCCACCACTTACATCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCA
CGAATGCGGTCACCATGCCTTCAGCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCA
CCACTCCAATACTGGGTCCCTCGAACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGC
AGTCACTCTCACTCTCGGCTGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCG
AACGGAGATATTCTTCTCCGATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGT
AGTGAATGGATTCCTTGTCTTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGA
CTACGGGTTTCTGAACACGGTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGT
ATTGGG

>AYOGI (JQ963164)

TACTCAAAACCTCCCTTCACTCTTGGTGAGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCC
ATCTTCTACTACATCGCCACCACTTACATCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTA
GCCCACGAATGCGGTCACCATGCCTTCAGCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGT
CGCCACCACTCCAATACTGGGTCCCTCGAACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACA
CTTGCAGTCACTCTCACTCTCGGCTGGCCTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGC
GAGCGAACGGAGATATTCTTCTCCGATGCTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTC
CTTGTAGTGAATGGATTCCTTGTCTTGATCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGAC
AGAGACTACGGGTTTCTGAACACGGTGTTCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAG
CCGGTATTGGGAGAGTA

>ES44 (JQ963165)

CAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACTTACATCCA
CCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCCTTCAGCGA
CTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCCCTCGAACG
AGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGCTGGCCTTT
GTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCCGATGCTGG
CATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTCTTGATCAC
TTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACGGTGTTCCA
TAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATT

>JRF5 (JQ963166)

AGATCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACTTACA
TCCACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCCTTCA
GCGACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCCCTCG
AACGAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGCTGGC
CTTTGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCCGATG
CTGGCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTCTTGA
TCACTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACGGTGT
TCCATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAGTAC

>ACC NO 4/47 (JQ963167)

AAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACTTACATCCAC
CTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCCTTCAGCGAC
TACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCCCTCGAACGA
GACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGCTGGCCTTTG
TACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCCGATGCTGGC
ATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTCTTGATCACT
TTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACGGTGTTCCAT
AACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAGTACTACCAATTC

>TL23 (JQ963168)

TCAAGAAAGCCGTCCCTCCACACTGCTTCAAAAGGTCCATCCCCCGCTCGTTCTCCTACGTGGCTTATGACCTCACCATAGCCGCCATCTTCTACTACATCGCCACCACTTACATCC
ACCTCCTCCCCAATCCTCTCTCCTACGTGGCGTGGCCGATCTACTGGGCCTGCCAAGGCTGCGTCCTCACTGGTGTCTGGGTCCTAGCCCACGAATGCGGTCACCATGCCTTCAGCG
ACTACCAATGGCTCGACGACTTGGTCGGCTTTGTCCTCCACTCATGCCTCATGGTACCCTACTTCTCGTGGAAGCACAGCCACCGTCGCCACCACTCCAATACTGGGTCCCTCGAAC
GAGACGAGGTTTTTGTCCCCAAGCAGAAATCAGCCATTGGCTGGCACTCAAAGTACCTCAACAACCCACCTGGCCGTGTGCTCACACTTGCAGTCACTCTCACTCTCGGCTGGCCTT
TGTACTTGGCATTCAACGTCTCTGGAAGGCCGTACGACCGGTTCGCCTGCCATTACGATCCTAAATCCCCCATCTACAACGACCGCGAGCGAACGGAGATATTCTTCTCCGATGCTG
GCATCCTTGCTGTGAGCTTTGCGCTCTACAAGCTTGCTGTCGCCAAGGGACTGGCTTGGGTGGTTTGTGTCTACGGAGTTCCACTCCTTGTAGTGAATGGATTCCTTGTCTTGATCA
CTTTCTTGCAGCACACCCACCCATCATTGCCGCACTACAAATCCTCCGAATGGGACTGGCTGAGAGGTGCTCTGGCGACCATGGACAGAGACTACGGGTTTCTGAACACGGTGTTCC
ATAACATCACGGATACCCACGTGGCGCACCACCTGTTCTCGACGATGCCTCATTACCATGCAATGGAAGCTACAAAGGCGATCAAGCCGGTATTGGGAGAGTACTACCAATTCGACG
GGACTCCATTCAT

FAD3A gene sequences

>PADMINI (JQ963119)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT

TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCCAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAGTGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTTCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAGTATATCGTATAAAGTTTCAAGTGATTGTAAGAGCAGAAAAGAAAAGAAGAAGAAGA
ATAATAAAGAGGATTGGCAACA

>NL97 (JQ963120)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAATGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACACCATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTCCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAGTATATCGTATAAAGTTTCAAGTGATTGTAAGAGCAGAAAAGAAAAGAAGAAGAAGA
ATAATAAAGAGGATTGGCAACA

>NL260 (JQ963121)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGTTTCGCCGTTGCGGCGGCAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATTATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTGTTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAGTGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTGGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGGTTCCGGAGTCGGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATAATCTAGATTAATCAGGGATAATGTGATTATATTCTTTTAAATATTTTTGTTGTATCATAATTTAGTCATCTTTGATCGCCAGACCACCA
TCTCCGGTCACCGGAAAATGTCCTGATCAGTTTTCCGGTCACTCTAACCTCCATTGACAAATTCTTTCACCCACGTGATCAGCCTCGCCGGCTTTACGTTTCTGGGAATTTTTCCCG
ATGACCGACTGTACAACTTTATATCGAGAGTTGTATGGATCATACAAATATGTATGTATTTGTACAAAAGTATATTCTAAGTACTGTACTAAGCATTATTTAGTATTACGTTGTACA
GACCTATAGAGAAATGTATATAATTTTATATAGAACTTAGTATACATATAGTTGTGAAATGTCAATTCCTCCGTATTTTAGTCGTTAGATCTAAGACGACATGATTTTAGATGGCTA
AAATTTATGCAAGCTTTTTTTTATCGGATAGAGAAGTCATTTCTCCAATTCGGACTGGATTATTAACTATATTATTCGTCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGG
CAGCGAAGCCCGTGCTGGGGAAGTACTACAGAGAACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATA
CAGGCGACGTCGTTTTCTATCAATCTGACCCACATATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGC
TGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAG

>EC9825 (JQ963122)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA

GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAATGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTCCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAG

>SURABHI (JQ963123)
ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAATGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAG

>AYOGI (JQ963124)
ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAATGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAGTGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTCCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAG

>ES44 (JQ963125)
ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT

GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAGTGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTCCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCACCCCAGCTTTCTTTCTGAG

>JRF5 (JQ963126)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAATGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCAGCTTTCTTTCTGAG

>ACC. NO. 4/47 (JQ963127)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAAGTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT
TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAATGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATCGAGATTACGGGGTCATCAACACCATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGAATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTTCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCAGCTTTCTTTCTGAG

>TL23 (JQ963128)

ATGAGCCCTCCAAACTCAATGAGTCCCGCCACCAACGGCAGCACCAATGGTGTGGCTATCAATGGGGCGAAGAAGCTACTCGATTTCGACCCGAGTGCTGCTCCCCCTTTCAAGATT
GCAGACATCCGTGCTGCAATCCCGCCGCATTGTTGGGTGAAGAACCCCTGGAGGTCACTCAGCTACGTCCTGAGAGACCTCCTGGTCATCCTCAGCTTCGCCGTTGCGGCGACAAAG
CTGGACAGCTGGACTGTCTGGCCTCTCTACTGGATTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAATTTCACATGATCTTTCTGGTAATGTGGGTT
TTCTTTTCTTATTGAAAAAGATTAAAACTTTTTATCTGGGCTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGTTGGTTGTTGAACAACGTGATGGGACATATACTCCATTCC
TCAATCCTCGTACCTTACCATGGATGGTATTGTAACTATTGTTCGATATTCGATTATGATTACTGTTCTTTCAGATGAAGAATCTGTACCCTAATTGTTTTTTGTTACCAGGAGAAT
TAGCCACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCAGTAAGTTGACATGCAGTTTGCTCTAAAATGCAGAGTCCTCTGTTTTTTGTGTGTTC
TTGTGCTTTAATGACGATGATAATGAAATTGAAATTTGTAATAGCTGCCGGAGAAGGTGTACAAGAGCTTGGATACCGGCACCAAGTTCATGAGGTTCACCATCCCTCTCCCAATGT
TTGCGTATCCTATCTACTTGGTAAGTAAACAGACTGACTCCAAATTAGGAACTAATGACAATTTTGGACCCGACCTGGTTTGGTTGACTCGGGTCGATATGTTTCGGGTGGGTAATT
ACCCGATCTGGCGATGGGTGTGCGGCGGACATTGTCTTGCTCGTGGTCCACCCCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTTCGGAATATGTATCAACAGAAAAATCTA
GTTTTTATGTTACTAGTTTTCTGTATTTCCATGTTTTTTCCTCAATTCTAGCCGGAATTTGAATTCAAACTGAAATCGGGTAATTCCGTCCATAACAAAACGGAATTGGGCAGCCGT
AATTAGTTGAAACTAGACCTCAATTTTGGCGGAATTGGACCCGGCCATTTTTTACGTTTGCAAAACGGAAAACGTTTTTCTTTTGTAAAGCGCAAAATGAAAAACGTATCTAGTGGA
ATTATTGGACCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGCTCCAATTCATGCCCGGAAAACACTACTGTCATGCATTTTAATCTTGTATGGTTTTACCCCAATGGATGCA
GCGATGGATCCGGACGATTTTTAAAATATTATCGGGTTAAATTTAAAAATATCTTAAAACTATAAGAAAAAAATAACCAATTTTAAAGAATAAAAGAACTGGACACATATGACGGGT
GTCGTGGATGGATGTACTTGTCCCGCTCTATTAAAGGCTGATAATATACAGGTCAACGGTGAATGAAGGTTAGATGCGCTATTGGATTTGAATCCGATATGAAATGATAATTTTGGA
CACGATCTGTTTTGGGTGGGTAATATTTGATCTAGGGATGGCTCGTGCTCCAAACCGCACCAAAACCGCCTAATTCTCGACCAAAAAGATTTTATGAATACATATCAACAGAAAAAT
CTAGTTTTCATGTTACTAGTTTTATGTACAACAATATTAGGTGTCGTTTTCCCAGCCTTTTTCTTCAATTCCGGCCGGAATTCGCATTCAAACCGGAATTGGATGGAATCGGTATAC
CTCGTCACGGATGCATTGTCAATTCCTAGTTAGTTTCATGGTTTTGAAACCAATCAATCTATTCTATATGGTTTTGATTAACAGTGGAGGAGAAGTCCGGGGAAGAAAGGGTCGCAT

TTCAACCCATACAGTGACCTGTTCGCACCGAACGAGAGGACATCGGTCATGATTTCGACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATCTACGGCTTC
CTTCCGGTCTTCAAAATCTACGGCGTCCCTTATCTAATATTCGTGGCGTGGCTCGACATGGTGACCTACCTTCACCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAA
GAATGGAGCTACCTACGTGGAGGGCTGACGACCGTCGATTGAGATTACGGGGTCATCAACACCATCCACCATGACATTGGCACCCATGTTATTCACCATCTCTTCCCTCAAATGCCA
CACTATCACCTAGTCGAAGCGGTAAGGAGGTCTTGATTATTAACTTAATGTTTTTGTTGTTATAATTTGAGTCCGATTCTGGAGTCAGGGGATTTCCTTCTTGGATCCGATCCAGGA
TCAAGCTGGTCCCTTGAATTTCTATATGATCTTATATTAATTAAGGATAATGTGGTCATATGTTTTTAAATATTTTTGTTTACCATCATTTTCGATCACCGGAAAATGTCCTGAGCA
GTTTTCCGGTCACTTTAACCTCCATTGACAAATTTTTTCACCCACATGATCACCCTAGCCGGGTTTACGTTTATTGAAAATTTTTATTTTTTTGAATTTTTTTTCGATGACCAACTG
TACAACTTTGTATTGAAAGTTGTATGGATCATACAAATGTGTATGTACAAAAGTATATTCTAAGTACTATACTAAGCATTACTTAGTATTACGTTTCTACAAACCTATAGAGAAATG
CATACAATTTTGTATAGAACTTAGTATACACGTAGCTGTGAAATGTCAATTTCCCTCCGTATTTTCAGAGACAAGACATGATTTTTAGACTGGCAGATTTTTTTTATCGGATAGATT
TCTCCAACTTCAGATTCGGACTGGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACCGACTCAGGCAGCGAAGCACGTGCTGGGGAAGTACTACAGAGA
ACCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTGAGGAGCCTGGGCGAGGATCACTACGTTAGCGATACAGGCGACGTCGTTTTCTATCAATCTGACCCACA
TATTCCCAAGTTCCCTACCAGTGCCACCACCAAGTCCAAATCTAGCTGATGATATTGGCTCTGATCTGATGTATGCTGCAGGCTGTTTTATTTTGTCCTTTGTTCGTTTCTTTCTGC
CAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCAGCTTTCTTTCTGAG

FAD3B gene sequences

>PADMINI (JQ963129)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCAAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTAAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACATCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTCCGACTCGGGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGTT
TTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGTCAGAAATTTGAATTCAAACTGACATTGGGTAATTCTGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAGA
CCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAGTTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTAT
TTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGCA
CGCTCTTCGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGTA
TAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTTA
TAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGGT
CTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGGC
ATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATTA
CGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTACT
GTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATGA
TCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTCC
GGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTGT
ACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGATA
CTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGGA
TCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAA
GTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATCA
GTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTTC
TTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>NL97 (JQ963130)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTTGTGTTCCAACCCGCACCCAAACGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>NL260 (JQ963131)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCAAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTAAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACATCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTCCGACTCGGGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTTGACGAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGTT
TTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGTCAGAAATTTGAATTCAAACTGACATTGGGTAATTCTGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAGA
CCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAGTTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTAT
TTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGCA
CGCTCTTCGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGTA
TAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTTA
TAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGGT
CTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGGC
ATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATTA
CGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTACT

GTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATGA
TCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTCC
GGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTGT
ACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGATA
CTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGGA
TCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGAA
GTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATCA
GTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTTC
TTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>EC9825 (JQ963132)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTTGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>SURABHI (JQ963133)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTCGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>AYOGI (JQ963134)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTTGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>ES44 (JQ963135)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTTGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>JRF5 (JQ963136)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTCGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>ACC. NO. 4/47 (JQ963137)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCGAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTCAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC
GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTTGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
AAAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCGAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGGAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTTAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTATCTTGTCCTTTGTTCGTTT
CTTTCTCCCAGAAACAAATTCTCTGTTTCTATGTTTCTCTGTCTCTCCCGCCCCCAGCTTTCTTTCTGAG

>TL23 (JQ963138)

ATGAGCCCTCCAAACTCAATGAGTCCCACCACCAACGGCAATGGTGTGGCTATGAATGGGGCGAAGAAGCAGCTCGATTTCGACCCGAGTGCTGCCCCCCCTTTCAAGATTGCAGAC
ATCCGTGCTGCAATTCCGCCGCATTGCTGGGTGAAGAACCCCTGGAGGTCGCTCAGCTACGTCCTGAGAGACCTCCTTGTCATCCTCAGCTTCGCCGTTGCGGCGGCAAAGCTGGAC
AGCTGGACTTTCTGGCCTCTTTACTGGGTTGCTCAAGGAACCATGTTCTGGGCAGTCTTTGTTCTTGGACATGATTGGTAAACTAATTTCACATTTTCTTTCTGGTAATGTGGGTTT
TATTGAAAAAGATTAAAACTTTTTATCTGGGTTGTTGCATGCAGTGGCCATGGGAGCTTCTCAGACAGCTGGTTGTTGAACAATGTGATGGGACATATACTCCATTCCTCAATCCTC
GTACCTTACCATGGATGGTATTGTAACTATTGTTCAATATTAGATTATTGCTAGTTCTTCAGCTGAAGAATCCAAACCCTAATTTTCTTTTTCTGAATATTGACCAGGAGAATTAGC
CACAAGACCCATCACCAGAATCACGGCAATGTGGAGAAAGATGAATCCTGGGTTCCTGTAAGTTGACATGCAGTTTGCTGTAAAAATGCAGAGTGCTCTGTTTTTTTGTGTTCTTGT
GCTTTAATGGTGATAATAATGAAATTGTTGAAATGTAACAGCTACCGGAGAAAGTGTACAAGAGCTTGGATACCAGCACTAAGTTCATGAGGTTCACCATTCCTCTCCCAATGTTTG
CTTATCCTATCTACTTGGTAAGTAAAGAGACTGATAAGACTCCAAAGTAGGAATTAATGACAATTTTGGACCCGAGCTTTCGACTCGAGTCGATTTATTTCGGGTGGGTACTTACCC

GATGCGGCGGACATTGTTTTGCTCGTGGTCCACCTCGCTCCCAACCCGCCCCATTCTCGACGTAAAAAGATTACGGAATATGTATCAACAGAAATATCTAGTTTTTATGTTACTAGT
TTTCTGTATTTCCGTGTTTTTCCCCTCAATTCCGGCAGAAATTTGAATTCCAACTGACATTGGGTAATTCCGTCCATAAGAGAACGGAATTGGGAAGCCGTAATTAGTTGGAATTAG
ACCTCGATTTCGGCGGAATTGGACCCGGCCATTTTTCGCGTCCGGAAAAGCGTTTCCAGTGGAATTAGACGCCCATCTAGAATGGGTCCAATTCCACCCCAATTTCGGGGCTACCTA
TTTTTAGTGTACATGCATTTTAGATCTTTTACGGTCTTACCCCGATGGATATTCTTGTCGCGCTCTATTAAAAACTGACAATATACAGGTTGATTGTGAATGAAGATTAGGTGCTGC
ACGCTCTTCGTGTTCCAACCCGCACCCAAACCGCCCCATTCTCGACCGAAAAGATTTTATGAATACATATCTACAGAAAAATCTAGTTGTCATGTCACTAGTTTAATGTACAACAGT
ATAGGTGTCGTTTTCCGCGTCTTTTTCTTCAATTCCGGCTGGAATTCGCATTCAAACAGGAATTGGATGGAATTGGTCCCGGATGCATAGTCATTTCCCAGGCAGTTTCATGGTTTT
ATAACCAATCAATCTAATCTTATGCTTTTGATAAACAGTGGACGAGAAGTCCGGGGAAGAAAGGGTCGCATTTCAACCCATACAGCGACCTATTCGCACCAAACGAGAGGGCAGCGG
TCTTGATTTCAACATTGTGCTGGACAGCCATGGCCTTACTCCTCTGCTACTCATCGTTCATATACGGCTTCGCTCCGGTCCTCAAAATCTACGGCGTACCTTATCTGATATTCGTGG
CATGGCTCGACATGGTGACCTACCTTCATCACCACGGGTACGAGCAGAAGCTGCCGTGGTACAGAGGCAAAGAATGGAGCTACCTACGTGAAGGGCTGACGACCGTTGATCGAGATT
ACGGGGTCATCAACAACATCCACCATGACATTGGCACCCATGTCATTCACCATCTCTTCCCTCAAATGCCACACTATCACCTTGTGGAAGCGGTAAACAATTTGATTATTAATTTAC
TGTTTTTGTTGTTATAATTTGAGTCGGGAGATTTCCTTCCTAAATCCGATCCCTGGTCAATCTTGGCCCTTGAATCTTCATATAATCTAAAAATCTAGATTAATCAGGAACAATATG
ATCATGTTGTTTAAACTAATTTTGTTGGACCATAACCTACCGCCAACTGATGGACCACCGTCTCTGGTTACCGGACCCATCATTTCCGGTTACCAAGAAGTTTCTCGATCAGTTTTC
CGGTTACTTTGACCTGCGTTGAGGAAAATTCTTTCACCCACGTAAACACTGTCGTCAACTTTACGTTTCTGGAAAGTTTTTCCGATGATTGGCCGTACAATTTTGTACGAAGAGTTG
TACGGATCATATAAATGTGTATAAGTTTCTAGAAATCCGTACTGAAATATATACATATTTGACTTTTGTATAAAGTGTAATACTAAATACTATACTAAGTGCTGTACTCAGTATGAT
ACTTAGTACACACATTTGTATGACTATGAAATGTCAATTTTGCCCTTATATTCTCAGCCGTTAGATCTAAGACACAGTTTTTATACGGCTGAAATTTGTGGGGGCTTTGTAGATCGG
ATCCATAAGTCATTTCTTGGCTCAAGATTCGGACTCGATTATTAACTATATTATTCATCAACTCTGACGTTTGATGTTGCATGTGACAGACTCAGGCAGCGAAGCACGTGCTGGGGA
AGTACTACAGAGAGCCGAAGAAATCAGGGCCTTTCCCATTCCACTTGTTTGGGTACTTGGTAAGGAGCCTGGGCGAGGATCACTACGTTAGCGACACAGGCGACGTCGTTTTCTATC
AGTCTGACCCACATATTCCCAAGTTCCGTACCAGCAGTGCCACCACCAAGTCCAAATCCAGCTGATGATATTTGGCTCTGATATATGCAGGCTGTTTAT

SYNOPSIS OF THE THESIS TO BE SUBMITTED TO THE UNIVERSITY OF MUMBAI FOR THE DEGREE OF PH.D. IN BIOTECHNOLOGY

Topic of research	Molecular and biochemical characterization of high and low α-linolenic acid containing Indian flax (*Linum usitatissimum* L.) varieties
Name of the student	Ashwini V. Rajwade
Name of the research guide	Dr. (Mrs.) Vidya S. Gupta, Chief Scientist and Chair, Biochemical Sciences Division, National Chemical Laboratory, Pune 411 008
Place of research	Plant Molecular Biology Group, Division of Biochemical Sciences, National Chemical Laboratory, Pune 411008, India
Reference number	Th/1947 of 2009
Registration number	13
Date of registration	20/2/2009

Signature of Research Guide **Signature of Student**

(Vidya S. Gupta) (Ashwini Rajwade)

Organization of the thesis:

TITLE: Molecular and biochemical characterization of high and low α-linolenic acid containing Indian flax (*Linum usitatissimum* L.) Varieties

Chapter 1: Introduction and review of literature

Chapter 2: Assessment of genetic and fatty acid diversity among Indian flax varieties

Chapter 3: Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content

Chapter 4: Sequence characterization of desaturase genes from the flax varieties varying in α-linolenic acid content

Chapter 5: Thesis summary and future prospects

Bibliography

Synopsis of the thesis entitled "Molecular and biochemical characterization of high and low α-linolenic acid containing Indian flax (*Linum usitatissimum* L.) varieties"

Introduction

Common flax or linseed *(Linum usitatissimum* L.) is a versatile crop grown since prehistoric times for seeds and fiber. More recently, flax seed oil has come into focus due to its fatty acid (FA) composition. It is the richest agricultural source of α-linolenic acid (ALA), an essential dietary polyunsaturated fatty acid of *ω*-3 class (Morris, 2007) which serves as a precursor for biologically active longer chain poly-unsaturated fatty acids (PUFA), mainly eicosapentanoic acid (EPA) and decosahexanoic acid (DHA). They are components of cell membranes and play an important metabolic role. Presence of high levels of ALA in flax seed oil has lead to its use in both, industrial as well as food and feed products including a wide variety of neutraceuticals and health foods.

Adequate intake and proper balance of essential fatty acids are necessary to ensure proper physiological functions of human body. However, in our diet, the levels of omega-3 fatty acids are decreasing in an alarming way. Since flax is the only agricultural source of high ALA (45-65 %) and there exists varietal variation in the biochemical parameters, oil content and quality and yield, it is essential to exploit it for omega-3 nutrition. Systematic efforts have not been undertaken in flax breeding from omega-3 fatty acid point of view and it is largely a neglected crop in India. Most of the traditional Indian varieties produce 30 to 40% oil of which 40 to 50% is ALA. On the higher side, varieties with 60 - 65% ALA are available in flax germplasm. Varieties with less than 3% ALA are developed and used in flax breeding targeting to produce oil with improved tolerance to rancidity (AICRP on linseed, Annual Report 2009-2010).

The basis of crop improvement is selection operating on genetic variability, which provides adaptability to variables like environment, pest and disease incidences and market. Genetic diversity is, therefore, essential for crop improvement. Similarly, the knowledge about diversity and genetic relationships among the primary germplasm and breeding material gives insight into crop improvement strategies.

Considering the nutritional importance and market requirement of flax, breeding for higher yielding, disease resistant varieties is imperative. It was therefore necessary to initiate systematic molecular and biochemical research in flax to improve its yield as well as nutritional quality. Agronomically superior and genetically most diverse genotypes are expected to contain unique alleles that can be exploited by crossing them. In recent years, several PCR based DNA marker technologies such as RAPD, ISSR, SSR, SSLP, AFLP etc., which are technically simple to use, time saving, informative and environment independent, have become the tools of choice for characterization of the genetic diversity in plants.

Desaturation is an important biochemical process in the FA biosynthesis pathway. Fatty acid desaturases (FADs) are hence the key enzymes that convert saturated FAs with single bond between two carbon atoms (C–C) to unsaturated FA with double bond (C=C) at specific location in fatty acyl chain. In the consensus pathway leading to formation of ALA, stearic acid (SA, 18:0) is converted to oleic acid (OA, 18:1) with the help of Δ9 desaturase (SAD) by introduction of the 1st double bond between 9th and 10th carbon from the carboxylic end (α-end) of the fatty acyle chain. Further, OA is converted to linoleic acid (LA, 18:2) by introduction of one more double bond at the 12th position from the α-end (or ω-6 carbon) with the help of Δ12 desaturase (FAD2). Similarly, third double bond is introduced at the 15th carbon from the α-end (or ω-3 carbon) of LA by Δ15 desaturase (FAD3) to form ALA (18:3) ((Fulco, 1974); Heinz, 1993).

The distribution of fatty acids and fatty acid desaturases is ubiquitous, observed in all aerobic organisms including algae, fungi, mosses, higher plants and mammals. They play a key role in maintaining proper structure and function of biological membranes (Browse et al., 1993; Ohlrogge and Browse, 1995; McConn and Browse, 1998); Schmid and Ohlrogge, 2002). Various studies have shown that the three desaturases, SAD (Δ9), FAD2 (Δ12) and FAD3 (Δ15), drive the poly-unsaturated fatty acids (PUFA) synthesis pathway and appear to exist in multiple forms and locations in various plant systems. These genes have been isolated and certain aspects have been studied from many different plant systems such as Arabidopsis, tobacco, soybean, safflower, sunflower, castor, cotton, mustard, rice, wheat etc. The genes for Δ9, Δ12 and Δ15 desaturases have also been cloned from flax (Singh et. al.1994, Fofana et. al., 2004; Vrinten et. al., 2005; Krasowska et. al.,

2007; Kadake et al., 2009; Banik et al., 2011) and expression studies were done, but little is known about genetic control of these enzymes, particularly whether their activity is controlled at the transcription, translation, or post–translation level. Till date, not many reports are available giving a comprehensive account about the expression profiles of all the three desaturases (SAD, FAD2 and FAD3) which act sequentially in the FA desaturation pathway during seed development of flax especially in high and low ALA containing varieties.

The present work includes systematic molecular analysis of diversity in Indian flax germplasm and biochemical analysis of flax oil. Further, fatty acid profiling and expression analysis of microsomal desaturase genes from developing seed stages of flax varieties varying in ALA content was performed so as to determine the role of each desaturase gene in the differential accumulation of ALA in flax varieties with high and low ALA content. Lastly, all the desaturase genes were isolated and characterized from the same flax varieties

Objectives

1. Assessment of genetic diversity among various Indian flax varieties using molecular markers and diversity in various fatty acid contents of these flax cultivars using gas chromatography.
2. Fatty acid profiling of seed developmental stages of ten flax varieties varying in ALA content (initial selection on the basis of above mentioned diversity studies). Classification of these varieties into groups of high and low ALA content depending on ALA accumulation at the mature stage.
3. Expression analysis of all the microsomal desaturase genes (*SAD*, *FAD2* and *FAD3*), which encode the three key enzymes of the PUFA biosynthetic pathway, from the same seed developmental stages of low and high ALA containing flax varieties.
4. Isolation and molecular characterization of complete desaturase genes from the same ten flax varieties belonging to high and low ALA groups.

Assessment of genetic and fatty acid diversity among Indian flax varieties

For genetic diversity assessment, 70 Indian flax genotypes actively used in breeding programs were selected from germplasm grown at All India Coordinated Research Project (AICRP) on linseed, at College of Agriculture, Nagpur. Leaf tissue was collected on a single plant basis and DNA was extracted. To assess the intravarietal genetic variation, five seedlings each from 18 randomly selected genotypes were raised at NCL and subjected to DNA extraction. A set of 100 ISSR primers (UBC 801-900) was used for the analysis. These primers were initially screened with a subset of flax genotypes and the primers that generated polymorphic and reproducible amplification profiles were further used to determine the intravarietal genetic variation in the set of 18 genotypes. Later, the same polymorphic primers were used with the full set of available 70 genotypes. Only consistently reproducible bands on the gel were scored as present (1) or absent (0) and a binary matrix for each marker was prepared. Similarity matrix within the population was constructed using the WINBOOT software (Yap and Nelson 1996). Cluster analysis was performed using Multi Variate Statistical Package (MVSP) (Kovach 1998) and a dendrogram was generated based on the similarity matrix by the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), using the Jaccard's coefficient as the measure of genetic similarity. Principal Component Analysis (PCA) was also carried out using MVSP. Diversity parameters such as heterozygosity per locus (or intralocus gene diversity) (h_i), average gene diversity (H_i), polymorphism information content (PIC) and ISSR primer index (SPI) have been calculated.

Fatty acid profiling of all the flax genotypes was done using gas chromatography. These data were used to determine correlations between the fatty acid profiles and the clusters formed on the basis of the diversity analysis.

After initial screening of 100 ISSR primers, twelve ISSR primers were selected for the final analysis yielding 136 loci, of which 87 were polymorphic. The average number of amplified loci and the average number of polymorphic loci per primer were 11.3 and 7.25, respectively, while the percent loci polymorphism ranged from 11.1 to 81.8 with an average of 63.9 across all the genotypes. The range of PIC scores in this study was 0.03 to 0.49, with an average of 0.18. A dendrogram was generated based on the similarity matrix by the UPGMA, wherein the flax genotypes

were grouped in five clusters. The Jaccard's similarity coefficient among the genotypes ranged from 0.60 to 0.97. When the omega-3 alpha linolenic acid (ALA) contents of the individual genotypes were correlated with the clusters in the dendrogram, the high ALA containing genotypes were grouped in two clusters, although, this could not establish a true relationship. This study identified most diverse genotypes and suggested their use in breeding programs and for developing mapping populations.

Transcriptional analysis of desaturase genes in developing seeds of flax varieties varying in α-linolenic acid content

Flax varieties differ markedly in their seed α-linolenic acid (ALA) levels. In the present study it is attempted to study the correlation of ALA content variation in flax varieties with that of transcriptional levels of fatty acid desaturases (FAD), which play key roles in deciding the level of unsaturated FAs in seed oil. Various omics approaches were used so as to decipher the role of each desaturase in the final accumulation of ALA in the seed storage oil. The developing seed tissue (eight stages representing early maturity, mid maturity and matured stages of seed development) of all the varieties required in the study was collected from Colleges of Agriculture, Nagpur.

a) Fatty Acid Analysis

Fatty acid (FA) profiling of various seed developmental stages of ten Indian flax varieties (selected based on earlier data obtained while studying flax genetic diversity) including one mutant variety, TL23, was performed using gas chromatography. Six fatty acids were profiled, of which myristic acid (MA; C14:0) was almost negligible throughout the developmental stages across the ten flax varieties. The palmitic acid (C16:0) content was the highest at flower stage and gradually decreased through boll development to maturity in the ten varieties. The proportion of stearic acid (SA; C18:0) remained nearly constant in all the developing stages, irrespective of the varieties. Oleic acid (OA; C18:1) content was low at the flower stage, but showed an increasing trend up to 8-16 DAF and later remained steady or declined slowly up to maturity in most of the varieties. The linoleic acid (LA; C18:2) content was high at the flower stage and gradually showed decline from 4 to 48 DAF in all the varieties except TL23, where it accumulated steadily till maturity. Conversely, ALA content

was high at the flower stage, then declined at 4 DAF and again started increasing from 8 or 12 DAF till maturity in all the varieties except TL23. There was a major difference observed in the accumulation trend of LA and ALA contents in mutant TL23, when compared with the remaining nine varieties. In the normal varieties as the seeds matured, LA content decreased with the increase in the ALA content, while in case of the TL23 the accumulation was vies a versa where LA levels increased and ALA content decreased steadily with the maturing seeds. Depending on their final ALA content at the mature stage, these varieties were grouped under high ALA and low ALA groups.

b) Desaturase gene expression

Transcript profiling of six microsomal desaturase genes (*SAD1*, *SAD2*, *FAD2*, *FAD2-2*, *FAD3A* and *FAD3B*), which act sequentially in the fatty acid desaturation pathway, was performed using real-time PCR in the same seed developing stages. Gene specific as well as temporal expression pattern of all the desaturases was observed.

The *SAD* gene is responsible for converting stearoyl-ACP to oleoyl-ACP, by introducing a double bond at C9, and thus, can increase the unsaturated fatty acid content of the plant (Ohlrogge and Jaworski, 1997). Expression of both, *SAD1* and *SAD2*, genes in the developing seeds was studied and higher transcript levels of the *SAD2* gene compared to the *SAD1* gene in differential ALA containing variety groups under consideration was observed. TL23 showed the highest expression level of the *SAD2* isoform amongst the ten varieties. Overall expression of the *SAD2* gene appeared to be constitutive where as the *SAD1* gene appeared to be more temporal and genotype specific in expression as indicated by differential expression in low and high ALA containing groups.

The Δ12 desaturase/*FAD2* represents a diverse gene family in plants and is responsible for conversion of OA (18:1) to LA (18:2). LA production marks the synthesis of polyunsaturated fatty acid (PUFA) from monounsaturated OA and is a major factor in determining the quality of plant oils. Δ12 desaturation is also necessary for the synthesis of ALA (18:3) as LA, the product of this desaturation, can alone be accepted as the substrate for Δ15 desaturase (FAD3). In my study, gene specific expression of *FAD2* and the *FAD2-2*, in flax developing seeds has been analyzed for the first time It was observed that the *FAD2-2* gene expression was higher at the early seed development stages (flower to 12DAF) than the *FAD2*

expression. However, the *FAD2* transcripts increased rapidly and reached its peak at mid-maturation stages (16-22 DAF), and then gradually reduced as the seeds attained maturity. The transcripts for the *FAD2* increased significantly during embryo development while the expression levels of the *FAD2-2* appeared to increase slightly. Further, it was observed that the average relative expression of the *FAD2* gene was nearly 2-fold higher than the *FAD2-2* gene in developing seeds of the ten flax varieties. All these observations are indicative of the predominance of the *FAD2* gene expression in determining the LA level of the seed storage oil. In case of TL23, expression of both the genes was nearly the same.

Δ15/FAD3 is the last desaturase required in the series of desaturation reactions leading to the synthesis of ALA in plants. This gene has been shown to play a very important role in deciding the level of ALA in the total FA content of the plant (Yadav et al., 1993; Bilyeu et al., 2003; Vrinten et al., 2005; Banik et al., 2011). The gene expression pattern for the *FAD3A* and the *FAD3B* in developing seed stages of the ten genotypes divided in two groups according to ALA content clearly showed higher *FAD3* (both *FAD3A* and *FAD3B*) gene expression in high ALA group than the expression in low ALA group. Further, it was observed that the average expression maxima of the *FAD3B* was nearly fivefold higher than average expression maxima of the *FAD3A*. In the present study, even though the higher contribution in ALA accumulation was due to the *FAD3B* than the *FAD3A*, both the genes contributed in the final differential accumulation of ALA content in the mature seeds of high and low ALA groups with higher fold expression difference in the *FAD3A* than the *FAD3B* when high and low ALA containing groups were compared.

Sequence characterization of desaturase genes from the flax varieties varying in α-linolenic acid content

Six fatty acid desaturase genes were isolated from each of the ten flax varieties and sequenced. These sequences were compared with the respective reported gene sequences in NCBI as well as Phytozome databases.

a) Nucleotide sequence analysis of desaturase genes from flax varieties

Sequence analysis of *FAD3A, FAD3B, FAD2, FAD2-2, SAD1* and *SAD2* genes from ten flax varieties revealed haplotypes in these genes. When the whole (including exonic and intronic regions) *SAD1* and *SAD2* genes were considered, there were 2 and

6 alleles, respectively, were identified. On the other hand when only the coding (exonic) regions from these two genes were analysed, *SAD1* showed only one allelic form with all the ten genotypes falling under one haplotype, while for *SAD2*, two alleles were found where except for one variety (TL23), all the remaining nine varieties were represented by one haplotype. The results indicated that most of the variation was present in the intronic region.

In case of *FAD2* and *FAD2-2* genes, which are intronless, 4 and 2 alleles, respectively, were observed. In *FAD2*, seven varieties were represented by one single haplotype while the other 3 haplotypes consisted of only one variety each. In *FAD2-2* except for one variety all the remaining nine varieties were represented by a single haplotype.

For *FAD3A* and *FAD3B*, 9 and 7 alleles, respectively, were observed when the whole genes were considered. Only one haplotype in *FAD3A* was shared by two varieties while all the remaining haplotypes were represented by single variety each. When only the coding regions of both the genes were considered the ten varieties showed presence of 7 and 5 alleles, respectively. The results indicated that the variation was mainly in the exonic region. Most of the substitutions were synonymous in nature while only 2 nonsynonymous substitutions leading to new isoforms of *FAD3A* and *FAD3B* were observed.

b) AA sequence comparison in desaturase genes across flax varieties

AA sequences were deduced from these sequences and variations were studied across the flax varieties. The deduced AA sequences of all the desaturase genes (*SAD1, SAD2, FAD2, FAD2-2, FAD3A* and *FAD3B* genes) were identical in the ten flax varieties except for few changes observed in varieties TL23 (in SAD2*,* FAD3A *and* FAD3B) and NL260 (in FAD3A), which were not reported earlier. Further, in the ten varieties we observed that the soluble desaturase, SAD showed the characteristic diiron binding motifs while the membrane bound desaturase, FAD2 and FAD3 showed the presence of three diiron binding histidine boxes as well as specific motifs required for their steady state endoplasmic reticulum localization.

The nucleotide sequence analysis of the *FAD3A* gene in TL23 showed the presence of a nonsense mutation creating a premature stop codon in the fifth exon after amino acid 291. However, the *FAD3B* in TL23 was predicted to produce a full

length FAD3B protein with single amino acid change at 283 AA (G to E transition) in the fifth exon when compared with the FAD3B of the other genotypes. TL23, which was the mutant variety with only 1.7% to 2% ALA, showed the lowest expression of the *FAD3A* gene throughout the seed developmental stages. This low accumulation of the *FAD3A* transcripts in TL23 could be attributed to the presence of premature translation termination codon (PTC) in the fifth exon of the *FAD3A* gene.

Thesis summary and future prospects

The varieties analyzed in the diversity study, like 'Sheetal', 'SLS 50', and 'Ayogi', are agronomically superior and genetically the most distant from the other genotypes, as revealed by the ISSR-PCR analysis. Their potential for containing unique alleles can be exploited by crossing them with other elite lines from different genetic clusters. These genotypes were suggested for the flax breeding programs and also for developing mapping populations.

Study of fatty acid desaturases in variable ALA containing varieties indicated high overall correlation between the expression of desaturase genes with the variation in the ALA accumulation in high and low ALA variety groups. Thus, the accumulation of the final product, ALA, is not entirely dependent on the activity of a single desaturase gene, but is a cumulative result of the activities of all the three desaturases acting sequentially. Further, it was also seen that the variation in the AA sequences of desaturases in these varieties was not consistent with either of the groups and therefore, probably not responsible for the high and low ALA accumulation in the groups except in mutant variety TL23. These results indicate the need to explain differential transcriptional control in the varietal groups of high and low ALA content by carrying out the promoter analysis of all the desaturase genes in these ten flax varieties. Secondly, our studies also indicate the importance of studies towards post-transcriptional and post-translational regulation of these three desaturase genes through proteomic and biochemical analysis to completely decipher variation in ALA content of flax varieties grown under identical environmental conditions.

Ashwini V. Rajwade

117/121/2, 'Nand-Nandan',
Lele Chowk, Near Kamla Nehru Park,
Erandawane, Pune – 411 004
Maharashtra, India

Phone no. (Res): +91 20 25675249
Mobile no.: +91 9822217448
Email: ashrajwade@yahoo.com
ash25rajwade@gmail.com

Academic qualifications

- **M.Sc.** (1999) in Zoology (Animal Physiology and Biotechnology), University of Mumbai, Mumbai, India. Score 64.70%
- **B.Sc. (Applied) Bio-Medical Techniques (B.M.Tech)** (1994), University of Pune, Pune, India. Score 69%
- **B.Sc.** (1993) in Zoology, University of Pune, Pune, India. Score 68.70%

Research Experience

Senior Research Fellow (2007-till date)

Pursuing PhD research on molecular and biochemical characterization of high and low ALA containing Indian flax (*Linum usitatissimum* L.) varieties, at Plant Molecular Biology group, Biochemical Sciences Division, National Chemical Laboratory, Pune, India.

Project Assistant

- Project entitled "Towards improvement of flax for oil and agronomic characters" (2006-2007), sponsored by Department of Biotechnology, New Delhi, India, at Interactive Research School of Health Affairs (IRSHA), Pune, India.
- Project entitled "Molecular tagging of rust resistance genes in wheat" (2005-2006), at NCL, Pune, India.
- Project entitled "Quality control of tissue culture raised plants-DNA based quality control of plants produced at MTP, NCL-A value addition to TC raised plants", (2002-2005) sponsored by DBT, New Delhi, India, at NCL, Pune, India.
- Project entitled "Prolific worm resistant meat sheep for Maharashtra, India", (2000) sponsored by ACIAR, Australia, at NCL, Pune, India.

Research Skills

- Biochemical and molecular techniques: DNA extraction, PCR, molecular marker analysis using RAPD, ISSR and SSR markers, agarose and polyacrylamide gel electrophoresis, cloning and transformation in bacterial system.
- Gene expression analysis: RNA extraction, semi-quantitative and quantitative qRT-PCR analysis (Real-Time PCR)
- Gas chromatographic analysis for fatty acids
- Knowledge of computational softwares: Various diversity and haplotype analysis softwares like MVSP, WINDIST, WINBOOT, DnaSP. Nucleotide and protein sequence database search using GenBank, EMBL, PDB; Nucleotide and amino acid sequence analysis using various Bioinformatics tools such as BLAST and

Megablast; Nucleotide and amino acid sequence alignments using MEGalign, ClustalX/W, BioEdit; primer designing, etc.

Awards

- Awarded **Senior Research Fellowship (SRF)** by CSIR, New Delhi, India in the year 2007.
- Awarded **First prize** for scoring highest marks in M.Sc. (Zoology) at The Ramnarain Ruia College, Matunga, Mumbai, India in the year 1999.

Publications

- Sujatha K, **Rajwade AV**, Gupta VS, Hazra S. (2010) "Assessment of *Pongamia pinnata* (L.) - a biodiesel producing tree species using ISSR markers". *Current Science*, **99**(10): 1327-1329.
- **Rajwade AV**, Arora RS, Kadoo NY, Harsulkar AM, Ghorpade PB, Gupta VS (2010) **Relatedness of Indian flax genotypes (*Linum usitatissimum* l.): An inter-simple sequence repeat (ISSR) primer assay**. *Molecular Biotechnology*, **45**: 161-170.
- Dholakia BB, **Rajwade AV,** Hosmani P, Chavan S, Khan RR, Lagu MD, Saini RG, Gupta VS (2012) **Molecular mapping of leaf rust resistance gene *Lr15* in hexaploid wheat**. *Molecular Breeding* (In Press)
- **Rajwade AV**, Kadoo NY, Borikar SP, Harsulkar AM, Ghorpade PB, Gupta VS (2012) **Differential transcriptional activity of SAD, FAD2 and FAD3 desaturase genes in developing flax seeds contributes to variation in α-linolenic acid content among flax varieties**. (Manuscript communicated to *Phytochemistry*)
- **Rajwade AV**, Kadoo NY, Gupta VS (2012) **Haplotype diversity based on the desaturase gene sequences from the flax varieties varying in α-linolenic acid content.** (Manuscript under preparation)
- Rupali K, **Rajwade A**, Ghorpade P, Gupta V, Ranjekar P and Harsulkar A. **Influence of seed germination on fatty acid content and desaturase gene expression in flax.** (Manuscript under preparation)

Conferences

- Attended the National Workshop on Biodiversity Resources Management and Sustainable Use; 11-15 October 2004, Center for Biodiversity and Forest Studies, Madurai Kamraj University, Madurai, India.
- Presented poster entitled **"Diversity analysis of Indian flax cultivars using ISSR markers". Rajwade AV,** Joshi TP, Arora RS, Harsulkar AM, Kadoo NY, Ghorpade PB, Gupta VS at International conference on Plant Genomics and Biotechnology: Challenges and Opportunities, October 26th-28th 2005, IGAU, Raipur, India.
- Poster entitled **"Influence of seed development, germination and genotype on Omega 3 ALA content in Flax".** Agrawal R, **Rajwade A**, Harsulkar A, Kadoo N., Gupta V, Hegde M, Ghorpade P and Ranjekar P. International Conference on Biotechnology Approaches for Alleviating Malnutrition and Human Health, 9th–

11th January 2006, University of Agricultural Sciences, GKVK Bangalore, India. **Received first prize for best poster**.

- Oral Presentation entitled **"Diversity of desaturase genes in flax varieties"**. Gurjar GS, **Rajwade AV**, Kadoo NY, Gupta VS at International Conference on Biodiversity and its Conservation, January 28th-30th, 2011 Modern College, Pune, Maharashtra, India.
- Participated in the Seminar on Food Safety Issues, with Special Emphasis on Genetically Modified Food Crops, November 14-16, 2010, NCL, Pune, India.
- Presented poster entitled **"Differential characterization of desaturase genes in variable ALA containing linseed (*Linum usitatissimum* L.) varieties". Rajwade AV,** Gurjar GS, Kadoo NY, Gupta VS as a part of National Science Day and International Year of Chemistry Celebration; February 24th-25th 2011, NCL, Pune, India. Awarded **best poster presentation** prize.
- Presented poster entitled **"Differential characterization of desaturase genes in variable ALA containing Indian Flax (*Linum usitatissimum* L.) varieties". Rajwade AV,** Gurjar GS, Banarjee R, Kadoo NY, Gupta VS at World Congress on Biotechnology, 21-23 March 2011, Hyderabad, India.
- Participated in the International Symposium on Proteomics Beyond IDs and 4th Annual Meeting of Proteomics Society (India), November 22nd-24th 2012, NCL, Pune, India.

Personal Information

Date of birth: 25th April 1973

Sex: Female

Marital status: Married

Nationality: Indian

References

Dr. Vidya S. Gupta
Chief Scientist and Chair,
Biochemical Sciences Division,
National Chemical Laboratory,
Pashan, Pune, India 411008
Phone: +91-20-25902237
Fax: +91-20-25902648
Email: vs.gupta@ncl.res.in

Dr. Narendra Y. Kadoo
Senior Scientist,
Biochemical Sciences Division,
National Chemical Laboratory,
Pashan, Pune, India 411008
Phone: +91-20-25902724
Fax: +91-20-25902648
Email: ny.kadoo@ncl.res.in

Dr. Abhay M. Harsulkar
Head, Cell and Molecular Biology Group,
Interactive Research School for Health Affairs,
Bharati Vidyapeeth (Deemed) University Campus,
Pune, India 411043.
Tel: +91-20-24362852
E-mail: aharsulkar@yahoo.com