Expression of LEA genes from *Brassica napus* in
transgenic *Arabidopsis thaliana* during seed
development and in seedlings under abiotic stress

By

Sarah M. Amer

B. Sc. (Honours), Carleton University, 2008

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in
partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University

Ottawa, Ontario

©2014

Sarah M. Amer

Abstract

Late embryogenesis abundant (LEA) proteins are small, hydrophilic, and largely unstructured proteins that are synthesized in seeds during late maturation and in vegetative tissues in response to abiotic stress. They have been associated with dehydration and desiccation tolerance in many plants, animal, and bacteria, and are documented to protect plants against desiccation and abiotic stress. In this study we characterize three *Brassica napus* LEA promoters; LEA 411 (Group1), ECP 66, and ECP248 from group 3. All three showed accumulation during seed development as well as in vegetative tissue in response to salt, drought, and cold stress. We show that the LEA proteins are being expressed as early as 10 days after flowering in seeds. The response to drought and salt stress occurred as quickly as 3 hours after stress treatment while that to cold took longer to detect. Promoter analysis revealed ABRE, DRE/CRT, DRE2, MYB and MYC motifs.

Acknowledgements

I would like to express my deepest gratitude to my Supervisors Dr. Johann Schernthaner and Dr. Owen Rowland for their continuous support and encouragement through this long road. Thank you for giving me the opportunity to be part of your team and for always providing me with the guidance I needed at the most difficult times and throughout all the obstacles I faced. Your help is greatly appreciated and it was a pleasure working for you.

I would like to thank the members of my examining committee Dr. Doug Johnson, Dr. Shelley Hepworth and Dr. Tim Xing for agreeing to be part of my journey and for providing me with your valuable time, I greatly appreciate it.

I would like to thank Denise Chabot, the microscopy specialist at Agriculture and Agri-food Canada, for her guidance and time. It was a thrill working on these microscopes.

There are a number of people without whom this might not have been written, and for whom I am greatly indebted.

To Natalie Labbe, words can't describe my gratitude to you. I owe you everything that I know, and the skills that I acquired and will use for my career to come. Thank you for the countless hours you spent with me and for always being there. Thank you for being a great friend when needed the most.

I would also like to thank the Suleiman Saloojee and the Hunt Club Pharmacy team for giving me the time to work on my thesis and for always covering for me when I couldn't be there.

To my family and friends (Saly, Laila, Karma, and Tammy), this wouldn't have been possible without you, thank you for your encouragement and support. Last, but not least, I would like to thank my Husband and my best friend, Aziz, for his continuous support through it all.

I would like to dedicate this to my parents Mounir and Afkar, thank you for believing in me and always encouraging me to be a better person. I love you so much and I hope to make you proud.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	v
List of Figures	viii
List of Tables	xii
List of Abbreviations	xiv
1 Introduction	1
1.1 Seed development	2
1.2 LEA genes.....	5
1.3 LEA Groups.....	6
1.3.1. Group 1	8
1.3.2. Group 2.....	9
1.3.3. Group 3.....	11
1.4 Transcription factors and regulatory elements	16
1.4.1 Dehydration responsive element binding (DREB)	16
1.4.2 Dehydration Responsive Element/C-repeat responsive element (DRE/CRT)	17
1.4.3 Absciscic Acid Responsive Element (ABRE)	18
1.4.4 RY element.....	18
1.5 LEA Function and Structure	19
1.6 Thesis rationale.....	22
1.7 Objective	22
2 Materials and Methods	23
2.1 Plant material and growth conditions	23
2.1.1 Germination of seeds in Petri dishes	23
2.1.2 Germination of seeds in soil muffins.....	24
2.1.3 Preparation of sterile Murashige and Skoog (MS) medium	24

2.2 Plant transformation	25
2.2.1 Transforming <i>Agrobacterium</i> GV3101-pMP90 electro-competent cells	25
2.2.2 <i>Agrobacterium tumefaciens</i> cultivation	25
2.2.3 <i>Arabidopsis</i> transformation	26
2.2.4 Flower tagging and seed collection	26
2.2.5 Testing <i>Arabidopsis</i> lines for transformation by protein analysis.....	27
2.3 Constructs	28
2.3.1 Plasmid vectors	28
2.3.2 LEA gene promoters, cDNAs and reporter genes	28
2.4 Cloning of promoter and cDNA fragments	29
2.4.1 LEApro::cDNA::eYFP constructs	29
2.4.2 LEApro::GUS constructs.....	30
2.4.3 Ligation of plasmids	30
2.4.4 Bacterial strains and <i>E. coli</i> transformation	31
2.4.5 Colony PCR	32
2.4.6 Plasmid minipreps	32
2.4.7 Agarose gel electrophoresis	33
2.5 Analysis of transformed plants	34
2.5.1 YFP Imaging	35
2.5.2 GUS imaging	35
2.5.3 Expression analysis by quantitative real-time PCR	36
3 Results	41
3.1 LEA expression during seed development.....	46
3.1.1 qPCR analysis of <i>Brassica</i> promoters-ORF-eYFP, of LEA 411, ECP 66, and ECP 248 in <i>Arabidopsis</i> seeds	46
3.1.2 Confocal analysis of <i>Brassica</i> promoters-ORF-eYFP, of LEA 411, ECP 66, and ECP 248 in <i>Arabidopsis</i> seeds	49
3.2 LEA expression under environmental stresses	55

3.2.1	qPCR analysis of <i>Brassica</i> promoters-ORF-eYFP of LEA 411, ECP 66, and ECP 248 constructs in response to salt, cold, and drought stresses in <i>Arabidopsis</i> seedlings	55
3.2.2	GUS analysis of <i>Brassica</i> promoters-GUS of LEA 411, ECP 66, and ECP 248 constructs in response to salt, cold, and drought stresses in <i>Arabidopsis</i> seedlings	64
3.3	Motif analysis in promoter sequence.....	70
4	Discussion	79
4.1	LEA gene expression during seed development.....	79
4.2	LEA gene expression under environmental stresses.....	80
4.2.1	Salt and drought stresses.....	80
4.2.2	Cold stress.....	82
4.3	Spatial localization of LEA genes.....	83
4.4	LEA promoter analysis.....	85
5	Future Directions	88
6	Conclusion	90
7	References:	91
8	Appendix	xcvi

List of Figures

FIGURE 1: OVERALL REPRESENTATION OF SEED DEVELOPMENT IN PLANTS BASED ON A DAYS AFTER FLOWERING TIMELINE.....	4
FIGURE 2: PHYLOGENETIC TREE SHOWING THE DISTANCE BETWEEN THE DIFFERENT LEA PROMOTERS CONSTRUCTED USING THE GENEIOUS SOFTWARE	41
FIGURE 3: CONSTRUCT DESIGN OF LEAPRO-YFP IN POLYLINKER OF PBI101 VECTOR USING HINDIII, XBAI, AND BAMHI RESTRICTION ENZYMES.....	43
FIGURE 4: CONSTRUCT DESIGN OF LEAPRO-GUS IN POLYLINKER OF PBI101 VECTOR USING XBAI, AND BAMHI RESTRICTION ENZYMES.	44
FIGURE 5: CONSTRUCT DESIGN OF LEAPRO-LEACDNA-YFP IN THE POLYLINKER OF PBI101 VECTOR USING XBAI RESTRICTION ENZYME. ..	45
FIGURE 6: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP66PRO-ORF-EYFP DURING SEED DEVELOPMENT OBTAINED FROM QPCR ANALYSIS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES.....	47
FIGURE 7: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP248PRO-ORF-EYFP DURING SEED DEVELOPMENT. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES.	47
FIGURE 8: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OFLEA411PR ORF-EYFP DURING SEED DEVELOPMENT. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES.	48
FIGURE 9: CONFOCAL MICROSCOPY DATA SHOWING THE EXPRESSION OF LEA GENES DURING THE DEVELOPMENTAL STAGES OF ARABIDOPSIS THIALIANA SEEDS (SEED COAT REMOVED) UNDER 10X OBJECTIVE. TOP	

IMAGE IS UNDER YFP FILTER, BOTTOM IMAGE IS BRIGHT FIELD. SCALE BAR 400 μ M.	51
FIGURE 10: CONFOCAL MICROSCOPY SHOWING THE EXPRESSION OF LEA GENES DURING THE DEVELOPMENTAL STAGES OF ARABIDOPSIS THALIANA SEEDS (SEED COAT REMOVED) UNDER DIFFERENT OBJECTIVES.	54
FIGURE 11: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP66PRO-ORF-EYFP UNDER SALT STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES.	57
FIGURE 12: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP66PRO-ORF-EYFP UNDER DROUGHT STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG THREE TECHNICAL REPLICATES	57
FIGURE 13: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP248PRO-ORF-EYFP UNDER SALT STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES	58
FIGURE 14: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP248PRO-ORF-EYFP UNDER DROUGHT STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES	58-59
FIGURE 15: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF LEA411PRO-ORF-EYFP UNDER SALT STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES	59
FIGURE 16: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF LEA411PRO-ORF-EYFP UNDER DROUGHT STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES	59

FIGURE 17: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP66PRO-ORF-EYFP UNDER COLD STRESS ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATE	62
FIGURE 18: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF ECP248PRO-ORF-EYFP UNDER COLD STRESS. ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES	62
FIGURE 19: ACCUMULATION PATTERN OF ONE BIOLOGICAL REPLICATE OF LEA411PRO-ORF-EYFP UNDER COLD STRESS ERROR BARS PRESENT THE STANDARD DEVIATION AMONG ITS THREE TECHNICAL REPLICATES	63
FIGURE 20 : GUS STAINING SHOWING THE EXPRESSION OF LEA GENE 14-DAY-OLD ARABIDOPSIS THALIANA SEEDLINGS IN RESPONSE TO SALT STRESS. SCALE BAR 4 MM	66
FIGURE 21: GUS STAINING SHOWING THE EXPRESSION OF LEA GENE 14-DAY-OLD ARABIDOPSIS THALIANA SEEDLINGS IN RESPONSE TO DROUGHT STRESS. SCALE BAR 4 MM	67
FIGURE 22: GUS STAINING SHOWING THE EXPRESSION OF LEA GENE 14-DAY-OLD ARABIDOPSIS THALIANA SEEDLINGS IN RESPONSE TO COLD STRESS. SCALE BAR 4MM	69-75
FIGURE 23: PROMOTER ALIGNMENT OF ECP 66 PROMOTER, ECP 248 PROMOTER AND LEA 411 PROMOTER SHOWING MOTIFS ASSOCIATED WITH DROUGHT/COLD STRESSES AS OUTLINED IN TABLE 17	77
FIGURE 24: ABA-DEPENDENT AND INDEPENDENT PATHWAY IN RESPONSE TO STRESS SHOWING THE TRANSCRIPTION FACTORS AND THEIR <i>CIS</i> -ACTING ELEMENTS INVOLVED IN COLD, DROUGHT AND SALINITY GENE EXPRESSION. MODIFIED FROM (ROYCHOUDHURY ET AL., 2013)	78
FIGURE A1: MAP OF BNECP66-ORF-EYFP SHOWING PRIMING SITE OF QPCR PRIMERS	CXXII-CXXV

FIGURE A2: MAP OF BNECP248-ORF-EYFP SHOWING PRIMING SITE OF QPCR PRIMERS	CXXV-CXXVII
FIGURE A3: MAP OF BNLEA411-ORF-EYFP SHOWING PRIMING SITE OF QPCR PRIMERS	CXXVIII
FIGURE A4: PROMOTER ALIGNMENT OF THE ECP 66 PROMOTER AND ECP 248 PROMOTER SHOWING SEQUENCES HOMOLOG.	CXXIX

List of Tables

TABLE 1: DIFFERENT NOMENCLATURE OF THE GIVEN LEA PROTEIN GROUPS (MODIFIED FROM HE ET AL., 2012)	7
TABLE 2A: SUMMARY OF LEA PROTEINS NUMBERING AND CHARACTERISTICS (BATTAGLIA ET AL., 2008)	14
TABLE 2B: SUMMARY OF AMINO ACID SEQUENCES OF DIFFERENT MOTIFS OF LEA GROUP (BATTAGLIA ET AL., 2008)	15
TABLE 3: SEQUENCES OF PRIMERS USED FOR RT-QPCR ANALYSIS.....	40
TABLE 4: KNOWN CHARACTERISTICS OF LEA AND LEA-LIKE GENES USED IN THIS STUDY MODIFIED FROM (HUNDERTMARK & HINCHA, 2008).....	43
TABLE A1: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP66 PRO-ORF-EYFP DURING SEED DEVELOPMENT	XCVI
TABLE A2: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP248 PRO-ORF-EYFP DURING SEED DEVELOPMENT	XCVII
TABLE A3: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED LEA411 PRO-ORF-EYFP DURING SEED DEVELOPMENT.....	XCVIII
TABLE A4: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP66 SAMPLES UNDER SALT STRESS	XCIX
TABLE A5: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP66 SAMPLES UNDER DROUGHT STRESS	C
TABLE A6: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP248 SAMPLES UNDER SALT STRESS	CI
TABLE A7: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP248 SAMPLES UNDER DROUGHT STRESS	CII
TABLE A8: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED LEA411 SAMPLES UNDER SALT STRESS.....	CIII
TABLE A9: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED LEA411 SAMPLES UNDER DROUGHT STRESS.....	CIV

TABLE A10: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP248 SAMPLES UNDER COLD STRESS	CV
TABLE A11: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED ECP248 SAMPLES UNDER COLD STRESS	CVI
TABLE A12: QPCR DATA COMPARING THE P VALUE AND RQ OF TESTED LEA 411 SAMPLES UNDER COLD STRESS.....	CVII
TABLE A13: MOTIFS IDENTIFIED USING PLACE IN ECP 66, ECP 248, AND LEA 411 PROMOTERS	CVIII

List of Abbreviations

ABA: Absciscic acid
ABF: ABRE-binding factor
ABI3: Absciscic acid insensitive3
ABRE: Absciscic acid response elements
AREB: ABRE-binding protein
ARF: Auxin response factor
bZIP: Basic leucine zipper
CBF: Cold-binding factor
CD: Circular dichroism
CE1: Coupling Element 1
CRT: C-repeat responsive element
CT: Threshold Cycle
DAF: Days After Flowering
DRE: Dehydration Responsive Element
DREB: Dehydration responsive element binding
DREB: Dehydration responsive element binding
DREB: Dehydration responsive element binding
ECP: Embryonic cell proteins
ELF2: Elongation factor 2
eYFP: Enhanced yellow fluorescent protein
GFP: Green fluorescent protein
Glu: Glutamine
Gly: Glycine
GUS: β -glucuronidase
IDPs: Intrinsically disordered proteins

LDH: Lactate dehydrogenase
LEA: Late embryogenesis abundant
MAT: Maturation
MS: Murashige and Skoog
MYB: Myeloblastosis oncogene
NMR: Nuclear magnetic resonance
NYC: Myelocytomatosis oncogene
ORF: open reading frame
polyA: polyalanine
polyQ: polyglutamine
RQ: Relative quantification
TAG: Triacylglycerols
TF: Transcription factors
VP1: Viviparous-1
ZF-HD: Zinc-finger homeodomain
G: Glycine
P: Proline
A: Alanine
V: Valine
L: Leucine
I: Isoleucine
M: Methionine
C: Cysteine
F: Phenylalanine
Y: Tyrosine
W: Tryptophan
H: Histidine
K: Lysine
R: Arginine
Q: Glutamine
N: Asparagine
E: Glutamic Acid
D: Aspartic Acid
S: Serine
T: Threonine

1 Introduction

Intrinsically disordered proteins (IDPs) occur in all organisms. In plants and animals, they are an essential part of signal transduction and transcriptional regulation. As the name suggests, they are unstructured proteins that have the ability to change between alternative conformations (Tompa & Kovacs, 2010). IDPs can function with or without direct molecular interactions. With no interaction they form entropic chains that either separate molecules in space (as is the case with microtubule associated proteins), or function by providing exclusion (such as spatial exclusion in the nuclear pore complexes where molecules are filtered by size). When interacting directly with the molecules, they can bind to different recognition sites, including post-translation modification sites, via short recognition elements (Tompa & Kovacs, 2010). As (Tompa & Kovacs, 2010) outlines, being disordered gives IDPs great advantages; they are able to rapidly recognize and bind their targets with great specificity while undergoing a completely irreversible interaction. That also allows them to carry out multiple functions simultaneously either using multiple interaction sites or through one specific site. They act as chaperones that assist the target protein into its proper structure; actively unfolding misfolded substrates, preventing aggregation, and solubilizing aggregates. Plants being stationary rely on a number of defense mechanisms to survive stresses. In plants, IDPs are involved in developmental stages and they play a major role in induced stress responses. A

protein family of particular interest to us are the late embryogenesis abundant (LEA) proteins.

During plant development or generally under environmental conditions that lead to reduced water potential, water loss leads to serious effects such as DNA damage, higher cytoplasmic viscosity, oxidative stress, and protein aggregation (Wu et al., 2011). Environmental stresses, such as dehydration, impose great limitations to crop productivity. Plants produce a wide array of stress responses to adapt to these stresses. These responses include physiological, biochemical, molecular, and cellular process. Many stress inducible genes are classified into two major groups; one involved in stress tolerance such as chaperones and LEA proteins and the other group in gene expression and signal transduction. For the purpose of this study, LEA proteins will be discussed further, in particular, with respect to the critical role LEA proteins play in plant seed development and under abiotic stress.

1.1 Seed development

Seeds are a key stage of the plant's development. They are the base of the plant that determines how successful an organism is and the connection point between all the different generations to come. Seeds, in general, undergo different stages during their development; the cotyledon stage, maturation phase, seed desiccation and finally dormancy, but are generally divided into early embryogenesis, where morphology is established, and late embryogenesis, where storage products are accumulated and seed desiccation occurs (Delseny et al., 2001). The cotyledon stage is characterized by rapid cell division (Bustros, 1989). This is followed by the

maturation (MAT) phase where the accumulation of proteins, lipids, and carbohydrates takes place (Bobb et. al, 1997). Early morphogenesis starts with numerous cell divisions starting with double fertilization in the embryo sac (Baud et. al, 2002). The triploid endosperm develops later in two separate stages; a growth stage followed by a differentiation stage before being completely disassembled during maturation. During that second stage of maturation and differentiation most storage proteins, including LEA proteins, and lipids start accumulating. During the last stage of seed development, the late maturation or desiccation stage, the embryo becomes metabolically quiescent and tolerant to desiccation. It is at this stage where orthodox seeds; seeds shed from parent plant at low moisture content and are able to withstand dehydration (Delahaie et al., 2013), acquire the ability to withstand extreme dehydration (Tunnacliffe & Wise, 2007).

Lipids in the developing embryo accumulate in the form of triacylglycerols (TAG), which are esters of glycerol and fatty acids. TAGs are used as a major source of carbon skeleton and energy as they are broken down for germination of seeds and growth of the young seedlings. They are mostly stored in cytosolic oil bodies, which occupy 60 % of the cotyledons as the embryo matures (Baud et al., 2002). The fatty acids pathway is well studied in plants.

Once the embryo has undergone some development, it is in the cellular expansion stage, where the dry mass increases and where the generation of the energy to be used for germination takes place. The final stage is usually maturation. This is where reduction in seed water occurs, and where gene expression, and protein profiles

change greatly (Shih et al., 2010). These changes are usually associated with achieving desiccation tolerance, increasing Absciscic acid (ABA) content, or developing the capacity for seed germination; which might not take place for years. Seed development is usually characterized based on days after flowering, where the flower is fully developed and the seed starts forming as outlined in figure 1 below.

Figure 1: Overall representation of seed development in plants based on a days after flowering timeline

1.2 LEA genes

Late embryogenesis abundant (LEA) proteins are one of the most important stress associated proteins in plants. LEA proteins have been shown to be associated with the water stress response as well as desiccation; they accumulate at the onset of seed desiccation. The first LEA protein was discovered in cotton, *Gossypium hirsutum*, back in the mid 1980s (Tunnacliffe & Wise, 2007). The name is based on its initial discovery at high levels in the later stages of embryo development in plant seeds when the embryo is developing desiccation tolerance (Tunnacliffe & Wise, 2007), (Hinch & Thalhacker, 2012). With respect to our current understanding of LEA proteins being associated with desiccation tolerance, seed maturation is the first stage to detect LEA proteins, as the seeds need the ability to withstand extreme dehydration. As further studies were carried out in plants and other organisms, LEA proteins were found to not be restricted to plants, seeds, or associated with desiccation. In fact, after their discovery in cotton and other plant species, they have been discovered in many desiccation-tolerant organisms including bacteria, microorganisms, invertebrates, and recently animals (Hand et. al, 2011; Hinch & Thalhacker, 2012) .

Very little light has been shed on how LEA proteins identify their targets or the diversity of these targets. It is not known whether different LEA proteins interact with each other or other molecules, how they provide the hydrogen bond, or organize the available water molecules to maintain the functional structure of their ligands.

Difficulties in understanding LEA proteins arise from their sequence diversity and the

formation of large families in unicellular to multicellular organisms (Wu et al., 2011). But what all LEAs have in common is their high glycine content, high hydrophilicity, and low secondary structure (Rorat, 2006). That lack of knowledge on the function of LEAs proteins reflects on the naming as well. If greater insight were evident on the role of LEA proteins, then a more appropriate naming system could be developed.

1.3 LEA Groups

Since their discovery in cotton seeds, many more LEA proteins were discovered and a nomenclature had to be developed. The first nomenclature for LEA proteins was based on their discovery in cottonseed prototypes and uses D-19, D-11, D-7 (Dure, 1993b). Common classifications, group LEAs based in their amino acid sequence as proposed by (Battaglia et. al, 2008). Table 1 compares classification systems proposed by (Dure, 1993b) and (Battaglia et. al, 2008). In this work we use the nomenclature introduced by (Battaglia et. al, 2008).

Table 1: Different nomenclature of the given LEA protein groups (modified from He et al., 2012)

Pfam no.	Pfam	Dure	Battaglia	Representative protein name
PF00477	LEA_5	D-19	Group 1	Wheat TaEm
PF00257	Dehydrin	D-11	Group 2	Tomato <i>Dhn1</i> , soybean <i>PM12</i>
PF02987	LEA_4	D-7 or D-29	Group 3	Carrot ECP63, barley <i>HVA1</i>
PF03760	LEA_1	D-113	Group 4	Tomato le25
PF04927	SMP	D-34	Group 5A	Carrot ECP31
PF03242	LEA_3	D-73	Group 5B	<i>Arabidopsis</i> ATD121
PF03168	LEA_2	D-95	Group 5C	Cotton LEA14-A
PF10714	LEA_6		Group 6	Kidney bean PvLEA18
PF02496	ABA_WDS		Group 7	Tomato Asr1, pine lp3-3

Group 1 contains a hydrophilic 20-amino acid motif; Group 2 had at least two of three distinct sequence motifs (Y, S, and/or K) (Close, 1997); and Group 3 contained multiple copies of an 11-amino acid motif. For the scope of this paper, only these major LEA groups are going to be discussed in details.

Most LEA proteins belong to the hydrophilins family; mostly composed of a group of highly hydrophilic IDPs with an amino acid composition enriched in Glycine (Gly) and other small residues that allows a more flexible conformation (Hinch et al., 2011). The only feature that does not enable the classification of all LEA protein as hydrophilins is their content of Gly, therefore only LEA proteins with a high Gly content are labelled as hydrophilins (Hinch & Thalhammer, 2012). Besides being hydrophilic, they are mostly cytosolic and contain random coils or α helices (Soulages

et al. 2002). They are non-globular, have low complexity and are capable of changing their confirmation according to the water content of their surrounding environment (Wu et al., 2011). Differences in amino acid sequences gives rise to at least 5-7 groups of LEA. The grouping and the nomenclature of LEAs is not consistent in the literature due to different numbering systems. LEA proteins that show a different expression profile to the original groups or that fall outside these groups are termed LEA-like proteins.

1.3.1. Group 1

LEA proteins with a hydrophilic twenty-mer sequence (TRKEQ{L/M}G{T/E}EGY{Q/K}EMGRKGG{L/E}) that can be repeated from one to four times are only found in plants but these repeats can occur up to eight times in other organisms and can contain a large number of charged residues (Battaglia et al., 2008) which contributes to its high hydrophilicity. They are characterized by regularly spaced Gly and Glu with a period of twenty (Tunnacliffe & Wise, 2007). Studies in plants also identified two other conserved motifs; an N-terminal motif (TVV PGG TGG KSL EAQ EH/N LAE) located upstream of the twenty-mer sequence and a C-terminal motif (DK/E SGG ERA A/EE/RE GIE/D IDE SKF/Y). Proteins within this group are unstructured in solution (Rorat, 2006) and exist largely as random coils (Battaglia et al., 2008). NMR analysis of Group 1 LEAs suggest that they have an unstable structure and are quite flexible (Eom, Baker, Kintanar, & Wurtele, 1996). To determine the function of Group 1 LEA protein, two proline residues were inserted in the predicted (by computer sequence analysis) N-terminus a helical domain of the

wheat EM protein; termed EM since its produced during Early germination among Methionine labeled proteins, the alteration severely effected the ability of rEm; recombinant EM, to protect Lactate dehydrogenase (LDH) activity while drying which concluded that group 1 plays a role in providing protection from drying as well as the importance of the N terminus in that function (Gilles, Hines, Manfre, & Marcotte, 2007). In another study, it was found that knock-out mutations of two Group 1 genes in Arabidopsis lead to premature seed dehydration and maturation thus suggesting a role during seed development (Manfre, Lanni, & Marcotte, 2006). They also determined that Group 1 LEA proteins have a role during normal seed/seedling development from studying vegetative tissue grown under normal conditions (Manfre, Lanni, & Marcotte, 2006). LEA proteins have also been shown to accumulate during embryo development especially in dry seed and have also been identified in organs that undergo dehydration such as pollen and grains (Battaglia et al., 2008).

1.3.2. Group 2

Dehydrins (D-11) are the most studied LEA group. Typically accumulating during the late stages of embryogenesis or in response to low temperature, ABA application, drought, salinity, or extracellular freezing (Close, 1997). They are characterized by their high hydrophilicity and thermostability. They contain a high portion of charged and polar amino acids, a low fraction of nonpolar hydrophobic residues and are completely lacking tryptophan and usually lack cysteine, suggesting that they act as structure stabilizers with properties similar to detergents and chaperones. As their sequences became available, dehydrins were defined as

proteins containing at least one copy of a conserved segment called the K segment, which is a highly conserved 15-mer Lysine rich amino acid sequence (EKKGIM{E/D}KIKEKLPG) present in 1-11 copies near the C terminus of all dehydrins (Close, 1997). They may also contain another conserved Y segment of tyrosine rich amino acids (V/T) D (E/Q) YGNP usually found in one to 35 copies near the N-terminus and a serine rich S segment made up of a stretch of 4-10 serines that are part of a conserved sequence LHR SGS₄₋₁₀(E/D)₃. When present, the S segment precedes or follows the Y segment and can be phosphorylated (Battaglia et al., 2008). These conserved segments can further be subdivided according to the combination of the K, S and Y segments into 5 structural subgroups: K_n, SK_n, K_nS, Y_nK_n, and Y_nSK_n, where n= any number of repeats. When hydrated in aqueous solution, they are found as random coils forming intermolecular hydrogen bonds with neighboring water molecules and intramolecular hydrogen bonds between the different amino acid residues (Hanin et. al, 2011). Due to the low proportion of intramolecular hydrogen bonds and high content of hydrophilic amino acids they are unstructured and similar to intrinsically disordered proteins. Being unstructured in solution, they are able to change their confirmation according to changes in their environment. Results measured using far-UV circular dichroism (CD) confirmed that changes in the hydration state of the solution leads to conformational changes (Hanin et al., 2011). Addition of glycerol, detergent, or salts into the solution all lead to conformational changes. The K segment adapts to the reduced hydration by forming an α -helical confirmation, where the negatively charged amino acids lie on

one side, the hydrophobic amino acid lie on the opposite side, and the positively charged amino acids lie on the polar-non polar interface. Thus the K-segment motif alone is thought to protect membranes as well (Battaglia et al., 2008). With conformational changes come changes in the function of the protein. It is believed that one of the methods by which dehydrins act is based on a molecular shield mechanism. They act by inhibiting the interaction between denatured protein molecules thus preventing their formation of protein aggregates (Szalainé Ágoston et. al, 2011). Dehydrins are believed to function by stabilizing hydrophobic interactions, such as the ones present in membrane structures and proteins. The conserved polar regions act as surfactants, preventing coagulation under cellular dehydration and low temperature (Ismail, Hall, & Close, 1999).

LEA Group 2 proteins localize to all the common tissues; root tips, root vascular system, stems, leaves and flowers (Battaglia et al., 2008). They are commonly localized in the nucleus, cytoplasm, and plasma membrane (Borovskii et. al, 2002), but some have also been found to accumulate in the vicinity of the plasma membrane, mitochondria, vacuole, and endoplasmic reticulum (Battaglia et al., 2008).

1.3.3. Group 3

Also known as D-7 and D-29 members are LEA proteins characterized by a repeated 11-mer amino acid motif, FF{E/Q}XFK{E/Q}KFX{E/D/Q}, where X is a hydrophobic residue, the remaining amino acids are mostly neutral to basic. The number of repeats of the motif determines the differences in molecular masses of

Group 3 proteins (Dure, 1993a). Other conserved groups that have been identified are motifs 1, 2 and 4 in subgroup D7 and motif 5 in subgroup D-29 as outlined in Table 2b. LEA proteins belonging to this group have been discovered in other organisms such as nematodes, algae and prokaryotes in addition to its initial discovery in plants and encompass all non-plant LEA proteins (Tunnacliffe & Wise, 2007).

As for other LEA proteins not belonging to these groups, they are also natively unfolded in solution but become obtain structure upon drying (Rorat, 2006). This is the one of the few similarities they have with other LEA groups. When compared with other LEA groups, Group 3 is more diverse; that diversity arises from the changes of sequence in different positions within its 11-mer motif which leads to sub-classification of Group 3 into 3A (D-7) and 3B (D-29) as outlined in Table 2b. Interestingly, studies showed that while Group 3 has a random coil conformation in solution, it responds differently to different conditions. In the presence of glycerol, ethylene glycol, methanol or after fast drying they adopt an α -helical conformation (Dure 2001), while in response to slow drying, α -helical and intermolecular extended β -sheets are formed. Thus the final structure depends on the drying rate or the stress. Soluble non-reducing sugars result in the formation of glass at low water concentration in mature seeds and pollen (Wolkers et al., 2001). Interestingly, when Group 3 LEAs are rehydrated they return to their random coil structure, thus the structural transition is fully reversible (Tollete et al., 2007). The diversity of this group

and the varieties in the conformations its members can adopt suggests that variety will arise in their intracellular locations and possibly their targets.

Table 2a: Summary of LEA proteins numbering and characteristics (Battaglia et al., 2008)

Group	Motif Name	Dure naming	Motif sequence	# of motif repeats	Size (a.a.)	MW (kDa)	Charge
1		D-10	GGQTRREQ LGEEGYSQ MGRK	1-4	83-153	6.8-20.3	Acidic to Neutral
2	Y	D-11 Dehydrins	DEYGNP	Y _n SK _y : n= 1-35, y =1-3 Y _n K _y : n=1-11, y=1-4	80-620	5.3-66.3	Neutral to Basic
	S		S _n	SK: 1-11 KS: 1-3			
	K		EKKGIMDKI KEKLPG	1-11			
3A		D-7	TAQAAKEK AXE	1-8	172-555	7.2-67.2	Neutral to Basic
3B		D-29					
4		D-113		1-2	80-124	12.6-18.8	Basic
5		D-34					
6		D-73 Lea5					
7		D-95 Lea14					

Table 2b: Summary of amino acid sequences of different motifs of LEA group
(Battaglia et al., 2008)

GROUP	MOTIF	CONSENSUS SEQUENCE
LEA 1 (D-19)	1	T V V P G G T G G K S L E A Q E H L A E N
PF00477	2	T R K E Q L G T E G Y Q E M G R K G G L M E K E
	3	D K S G G E R A A E E G I E I D E S K F E E R D Y
LEA 2 (D-11)	K	E K K G I M D K I K E K L P G L L E D I K R M K F
PF00257	S	L H R S G S W S S S S S D D D H S E E E
	Y	R T D E Y G N P V H Q V I Q
LEA 3 (D-7)	1	G G V L Q Q T G E Q V S F S A K
PF02987	2	A A D A V K H T L G M K E N F
	3 ^a	T A Q A A K D K T S E S E Q Q
	5 ^a	A T E A A K Q K A S E T A Q T E A
	4	S Y K A G E T K G R K T R A K A Q
LEA 3 (D-29)	1 ^a	T A E K A G E Y K D Y A
PF02987	4 ^a	T V E K A K E A K D T A Q T R T A
	2 ^a	A Y E K A G S A K D M D A
	3 ^a	A A Q K A K D Y A G D S E D S
	5	E S W T E W A K E K I A G D
LEA 4 (D-113)	1	A Q E K A E K A T A R D P x E K E M A H E K K E A K V E V R M T H T K Q I T Q R K E R
PF03760	2	M Q S A K E K A S N M A A S A K A G M E K T K A K E A T V A D V G S A E K V D I T
	3	E A E M D K H Q A K A H H A A E K Q Q K E L Q A E R E N R E A E
	4	P T G T H Q M S A L P G H G T G Q P T G H V V E G G M A T M A A T D
LEA 6 (LEA-18)	1	L E D Y K M Q G Y G T Q G H Q Q P K P G R G R K A A E L E V S D V P H
PF10714	2	G S T D A P T L S G G A V S G E I P A L A T
	3	T D A I N R H G V P T Q Y
	4	G L P T E T S P Y V Q V D D P T
LEA 7 (ASR)	1	A A G A Y A L H E K H K A K K D P E H A H R H K I F E
PF02496	2	E I A A A A A V G A G G F A F H E H H E K K E A K V V A S Y V Y Q D D H
	3	D Y K K E E K H H K H M E H L G E L G A V K R L Q Q I M T M
	4	H H H H H L F H H H K D W F R K Q Q
	5	E E E E E A H G K K H H L F K D Q H E F

The color in letters indicates the type of amino acid. Non-polar: violet = aliphatic (A, V, G, M, L, I, P); grey = aromatic (Y, W, F); Polar: green = uncharged (S, Q, N, T); blue = positively charged (R, K, H); red = negatively charged (D, E).

^a 11-mers as described by Dure (2001).

1.4 Transcription factors and regulatory elements

The deposition of LEA proteins within the cell requires a network of hundreds of response elements and transcription factors (TF) that regulate the expression of LEAs and other proteins involved in stress resistance. These response elements act as messengers between the response (gene) that needs to be expressed (or not) not only at the right time but at the right location as well. Studies have shown that plant response to abiotic stress is divided into two groups. One group encodes the genes that protect the cell from water stress effects. This includes reactive oxygen species that protect and stabilize the structure of the cell as well as LEA proteins, anti-freeze proteins and chaperones. The second group includes proteins that regulate stress signal transduction and modify gene expression and includes several transcription factors (TFs) such as myelocytomatosis oncogene (MYC), myeloblastosis oncogene (MYB), basic leucine zipper (bZIP), and dehydration responsive element binding (DREB) types (Lata & Prasad, 2011). These transcription factors interact with cis-regulatory elements in the promoter region of different stress genes to overcome the stress the plant is undergoing. LEA expression and other abiotic stress responses are regulated by abscisic acid (ABA) and thus, abscisic acid response elements (ABRE) play an important role as well as several other elements, some of which are outlined below.

1.4.1 Dehydration responsive element binding (DREB)

DREB are well-studied plant TFs and are divided into two groups, DREB1 and DREB2. DREB1 is involved in signal transduction under low temperature while

DREB2 is involved in dehydration response (Lata & Prasad, 2011). Overexpression of DREB2A in *Arabidopsis* resulted in increased drought tolerance and a slight increase in freezing tolerance (Huang et al., 2012). Microarray analysis of DREB2A overexpression lines, showed that it regulates expression of many water stress inducible genes, nonetheless some of the genes associated with DREB2A were not found with DREB1A; which is known to recognize DRE/CRT (outlined below). Therefore despite the ability of both DREB1 and DREB2 to bind to a DRE sequence, their DNA binding specificity to DRE core motifs differ, thus the activated genes downstream are also partially different (Huang et al., 2012).

1.4.2 Dehydration Responsive Element/C-repeat responsive element (DRE/CRT)

The DRE motif, TACCGACAT, is essential for gene expression in response to drought, high salinity, as well as low temperature (Lata & Prasad, 2011). The CRT motif, CCGAC, is closely related to the DRE motif, and has been identified in promoters of cold-inducible genes (Huang et al., 2012). It too is essential for the induction of several genes by low temperature. DRE/CRT is one of the many *cis*-acting elements that function in response to changes in ABA levels, but it is also one of the few that act independent of ABA signaling. Other elements that are independent of ABA are the cold-binding factor/dehydration responsive element binding (CBF/DREB) and the zinc-finger homeodomain (ZF-HD) (Lata & Prasad, 2011). DRE/CRT has been shown to be activated in an ABA-independent manner in

response to drought and cold stress in *Arabidopsis* mutants that were ABA-deficient and ABA-insensitive (Lata & Prasad, 2011).

1.4.3 Absciscic Acid Responsive Element (ABRE)

ABA is an important phytohormone during all stages of plant development. It has been shown to be associated with embryo maturation, seed development, seed dormancy, seed germination, root growth, fruit ripening, stomatal regulation, and the activation of stress response genes (Lata & Prasad, 2011). ABRE is a major *cis*-acting response element in the ABA-dependent pathway. ABRE starts a cascade effect when bound by bZIP transcription factor ABRE-binding protein (AREB) and ABRE-binding factor (ABF). Studies have shown that ABA-induced gene expression needs multiple ABREs or the combination of ABRE with a coupling element (Yoshida et al., 2010).

1.4.4 RY element

The RY element is a conserved sequence motif, CATGCA, usually found in repeats. It is the element that is recognized by ABI3 (Absciscic acid insensitive3), a member of the B3 domain protein family, which is a plant specific DNA-binding domain. ABI3 combined with FUS3 and LEC2 form a conserved transcriptional activation suite that regulates different stages of embryo maturation. ABI3 has been found to regulate both maturation (MAT) genes; the first phase of seed maturation where storage proteins are accumulated preceding the LEA stage, where the embryo is dormant and desiccation tolerant (Guerriero et al., 2009). The RY element is usually found in seed-specific promoters clustered with ABRE in upstream regions of

genes activated by a transcription factor *Viviparous-1* (VP1) and ABA. VP1 is a novel element but its binding site remains unknown, it is documented to transactivate the EM promoter in maize as well as its association with the Sph elements in the myb-like C1 gene and with ABRE elements in wheat and rice (Hollung et al., 1997). Mutants lacking ABI3 initially showed reduced sensitivity to ABA, maintained their water contents during maturation. Thus they were not desiccation tolerant and mature seeds lost their viability and did not germinate (Delahaie et al., 2013). All this suggests that ABI3 is an important player in desiccation tolerance of seeds.

1.5 LEA Function and Structure

In summary, despite being discovered 30 years ago, the exact functions of the LEA proteins are still unclear. Although some mechanisms have been proposed, the mode of action of LEA proteins remains unknown with respect to identifying their target and how interactions occur between LEAs and other molecules. LEAs stabilize membranes during freezing and desiccation. They also stabilize enzymes under stress conditions. Due to their hydrophilic and unstructured nature, they do not aggregate during drying, freezing, or boiling. Also it is proposed that in cells, LEAs establish desiccation tolerance, acting as osmoprotectants or chaperones (Delseny et al., 2001), but are not considered true chaperones as they do not prevent heat aggregation and heat-induced inactivation. LEA proteins are intrinsically disordered in the hydrated state and may function as disordered protein, but upon partial or complete dehydration they partially fold into α -helical structure. (Hinch & Thalhammer, 2012) proposes that induced folding during target binding is a

possibility when attempting to decipher LEA proteins. LEAs bind to RNA/DNA, display water/ion binding, antioxidative activity or sugar glass stabilization in the dry state. Based on their amino acid sequence, programs that determine the structure of protein based on their sequence, accurately predicted some of the secondary structure but not others (Hinch & Thalhammer, 2012). Thus determining LEA structure based on sequence is unlikely to be successful. In an attempt to understand the aggregation properties of LEAs in living cells, a study conducted by (Tunnacliffe & Wise, 2007) co-expressed a LEA protein with an aggregation-prone protein containing polyglutamine (polyQ) or polyalanine (poly A) stretches in mammalian cells. These polyQ or polyA containing proteins are associated with human neurodegenerative diseases (Chakrabortee et al., 2007). This not only suggests that LEAs can reduce aggregation *in vivo*, but most importantly that LEA may function in the hydrated state as well (Hinch & Thalhammer, 2012).

Another proposed mechanism of how LEAs function is termed a molecular shield mechanism. In this mechanism, it is believed that the unstructured LEA occupies volume in the cell that prevents partially denatured proteins from interacting together therefore reducing aggregation. A plausible hypothesis is that they function by hydrogen bonding polar groups on macromolecules or surrounding nuclear or cytoplasmic tissue thus coating the exposed hydrophobic surfaces and preventing coagulation of the macromolecules (Close, 1997).

LEA genes have also been discovered in the transcriptome of desiccation-tolerant bryophytes undergoing rehydration thus playing a role in recovery from desiccation as well (Oliver et al., 2004).

What is known is that LEA genes expression is associated with tolerance to water stress, or the anticipation of water stress; whether it is developmentally programmed (in the case with developing seeds) or in response to environmental stresses (Tunnacliffe & Wise, 2007).

1.6 Thesis rationale

Canola, also known as Rapeseed, is grown for the production of animal feed, vegetable oil for human consumption, and biodiesel. It is the third largest source of vegetable oil, and the world's second leading source of protein meal. Crop loss due to dehydration is a major hurdle that affects humans and livestock. LEA proteins are associated with cellular dehydration tolerance; which is induced by freezing, high salinity, and drying (Crowe, Hoekstra, & Crowe, 1992).

There are about 50 LEA genes found in *Arabidopsis thaliana* (Wu et al., 2011) that are divided into 9 different groups. Some are more widely studied than others. One of the most studied group is the dehydrins. Understanding the characteristics of different LEA proteins can help in protecting crop from drought and other environmental stresses.

1.7 Objective

This thesis is part of a project whose overall objective was to characterize LEA proteins in plants using LEA gene promoters isolated from *Brassica napus*. The objective for this thesis was to determine the spatial and temporal expression patterns of three *B. napus* LEA gene promoters fused to reporter genes in *Arabidopsis thaliana*, in the developing seeds as well as in response to environmental stresses. Analysis was carried forward using fluorescence microscopy as well as quantitative real-time PCR.

2 Materials and Methods

2.1 Plant material and growth conditions

All experiments involving plant transformation were carried out using *Arabidopsis thaliana*, ecotype Colombia-0 (Col-0). *Arabidopsis* plants were grown in pots and maintained in growth chambers at 22°C under 16 hours of light and 8 hours of dark. The plants were kept well watered at all times until seed germination. Once seeds started to dry out and before seeds were due to be harvested, the plants were covered with bread bags to decrease the chance of losing seeds caused by shattering of the drying silique. Plants undergoing cold stress treatment were placed at 4°C (16 hours of light, 8 hours of dark) at time 0 under stress and collected according to stress treatment times.

2.1.1 Germination of seeds in Petri dishes

Seeds of wild type and transformed plants that needed to be germinated in sterile conditions were first surface-sterilized. All work for seed surface sterilization was carried out in a flowhood that was wiped with ethanol before being used. Seeds were placed in Eppendorf tubes, then incubated in 95% ethanol for 1 minute. The ethanol was discarded and replaced by a freshly prepared solution of 3% sodium hypochlorite, 0.05% Tween 20. The seeds were soaked for 5 minutes followed by three rinses with distilled H₂O. After the water was removed, the seeds were suspended in 0.05% agarose (500 µL for each plate). Seeds were distributed evenly on to media dishes (about 20 seeds per plate) and the dishes were incubated in the

cold room at 4°C for three days before they were moved to the growth cabinet. After 10-14 days, when seedlings had developed sufficient roots, they were moved transferred to soil.

2.1.2 Germination of seeds in soil muffins

Small 5” pots were packed with soil until a little dome was created over the top of the pots thus resembling a muffin. 1mg of *Arabidopsis* seeds were suspended in 1ml of 0.05% agarose for each pot. 1mL was distributed evenly over the entire pot in a circular motion starting with the center of the pot and ending at its perimeter. Muffins were placed in the trays, water added, and the trays covered and sealed to prevent the escape of water vapor. Muffins were incubated at room temperature for 2-3 days until water condensation was evident on the covers of the trays. The trays were then moved to growth chambers and the lids were kept sealed. After another 2-3 days the cover seal was removed but trays were kept partially covered. Once seeds started to germinate the covers were removed.

2.1.3 Preparation of sterile Murashige and Skoog (MS) medium

2.15g of MS salt (purchased from PhytoTechnologies Inc., USA) were added to 800 mL of de-mineralized water and mixed, the pH was adjusted to 5.7 ± 0.1 with 1N KOH and the volume was brought to 1L. For each liter of MS solution 10 g (1%) of sucrose and 7 g agarose were added. The media were autoclaved for 20 minutes and allowed to cool to about 50°C after which 50µg/ml of kanamycin were added. The media was then poured into Petri dishes and allowed to solidify. Plates used for salt stress were prepared with 125mM NaCl, thus 7.31g/L NaCL were added to the

media before autoclaving. Drought stress effect was achieved using 250mM of Mannitol equaling 45.45g/L of MS media. Plates were stored in a sealed bag at 4°C.

2.2 Plant transformation

2.2.1 Transforming *Agrobacterium* GV3101-pMP90 electro-competent cells

Agrobacterium strain GV3101 was transformed by adding 1-5 μ g of pBI101 DNA to 40 μ L of GV3101-pMP90 electro-competent cells in a chilled cuvette.

Electroporation was carried out at 1.8KV, 200 Ω , and 25 μ F. Immediately after transformation 1000 μ L of SOC media were added to the cells and they were placed on ice. The reaction was then incubated at 28°C for 1 hour shaking at 225rpm. Once done, 100 μ L and 20 μ L of each transformation event were plated on LB Kan50 plates. The two sets of plates were incubated at 28°C overnight until colony formation was detected. Colony PCR of successful colonies was done as outlined below.

2.2.2 *Agrobacterium tumefaciens* cultivation

Arabidopsis plants were transformed using *Agrobacterium tumefaciens* as the bacterial vector. 1ml of overnight culture of *Agrobacterium tumefaciens* GV3101-pMP90 containing the appropriate DNA construct was added to 200 ml LB medium with kanamycin (50 mg/L) and gentamycin (50 mg/L) and then grown in a shaker overnight at 28°C. Once the culture was at an optical density OD_{600nm} $\approx$ 0.5, the bacteria was pelleted by centrifugation at 4000 rpm for 15 minutes at room temperature. Pellets were resuspended in 700 mL (700mL were used for each 2-3

small 5" pots) of 5% table sugar and 0.05% Silwet L77. 35g of sugar were added to each 700mL and 240 μ L of Silwet L77 were added immediately before the plants were dipped.

2.2.3 *Arabidopsis* transformation

Arabidopsis plants were transformed using the floral dip method according to (Clough and Bent.1998). Only flowering plants were used for transformation. Flowers that were already open were cut off. Plants were placed upside-down into the *Agrobacterium* solution ensuring that all bolts, above ground parts are submerged. Plants were then removed from the solution after a minute. Plants were kept on their sides to drain excess liquid, then covered and incubated overnight at room temperature to maintain high humidity. Plants were then staked and bagged with bread bags to prevent contamination with neighbouring transformed plants and moved to growing chambers. As seeds became dry, watering was stopped.

2.2.4 Flower tagging and seed collection

To collect *Arabidopsis* seeds at different developmental stages, plant growth was monitored. When plants neared flowering they were checked every morning (at approximately the same time) and the flowers were tagged with strips of coloured tape whereby every colour represented a different date since siliques were too small to hold tags that contained information. Initially some flowers were checked under a dissecting microscope to determine the right stage at which flowers needed to be tagged. The most optimal stage was after flowers opened and before pollination takes place. A jewellers' head magnifying glass was then used to determine the right

stage and flowers were tagged as 'Days After Flowering' (DAF); with day of tagging being day 0. Whole siliques (3-5 siliques) were cut off on the day of harvesting and immediately placed in liquid nitrogen; then moved to -80°C. Plants were staked and tied during the tagging process and covered with bread bags to eliminate cross pollination and entangling.

2.2.5 Testing Arabidopsis lines for transformation by protein analysis

To determine which lines to carry forward from the harvested T1 generation, a seed sample of each line from each transformation event was placed in an Eppendorf tube containing 100 μ L PBS and crushed using a hand-held homogenizer. Tubes were centrifuged at maximum speed for 15-20 minutes. Supernatant was transferred to new tubes and the protein concentration was measured using a Nanodrop spectrophotometer. All samples were diluted to the same protein concentration and placed in a Falcon 3072 96 plate suitable for the FLUOstar OPTIMA BMG LABTECH plate reader. The experiment required 3 chromatic number channels, all excitation filters were set at 485 nm and an emission filter of 520 nm and gains were set at 4095 for Chromatic No. 1, 3200 for 2 and 2900 for 3. There was a 0.2 s positioning delay as the readings were being collected in one direction. A Bradford protein concentration assay was used to construct a standard curve using a serial dilution of BSA stock and the protein concentration measured using a spectrophotometer at an absorbance of 595 nm (OD 595). From the equation derived from the standard curve, the concentrations of the proteins were calculated. A table of protein concentration for all T1 lines was constructed and the best expressing lines

within the normal range were chosen to be carried forward (Plants used in the analysis were from T3 seeds).

2.3 Constructs

2.3.1 Plasmid vectors

Plasmid pJET (Fermentas) was used to subclone DNA fragments obtained from PCR amplifications. The plasmid eYFP (GeneBank # CS538989, obtained from Prof. O. Rowland, Carleton University) was used for fusing promoters and cDNAs to the eYFP reporter gene. The plasmid peYFP is a binary vector based on plasmid pBI101 (Clontech) pBin19 sequence, containing a neomycin phosphotransferase gene driven by nos promoter" U12639.1. This vector was used as the receptacle for all final constructs used for *Agrobacterium*-mediated plant transformation.

2.3.2 LEA gene promoters, cDNAs and reporter genes

The LEA promoters ECP66, ECP248 and LEA411 and cDNAs were obtained from the lab of Dr. Schernthaner, AAFC/ECORC. These promoters had been isolated from *Brassica napus* by Genome Walking using primers derived from *Brassica napus* EST sequences. The group 3 LEA protein ECP63 (embryogenic cell protein) from *A. thaliana* has first been described by Yang et al. (1997). The *B. napus* cDNA sequence, BNECP63, is available from GenBank under accession # FJ147551. LEA411 corresponds to a group 1 LEA protein from *B. napus* (GenBank # AY572958). The promoters of ECP66 and ECP248 differ in sequence but drive the same cDNA, BNECP63, in the constructs used for promoter analysis.

2.4 Cloning of promoter and cDNA fragments

2.4.1 LEApro::cDNA::eYFP constructs

For subcloning, LEA promoters were amplified with primers containing a 5' *Hind*III and a 3' *Xba*I restriction site and inserted between the *Hind*III/*Xba*I sites in eYFP. The LEA cDNAs were in frame fused to the eYFP reporter gene in eYFP as *Xba*I fragments in between the LEA promoter and the YFP reporter gene and their correct orientation and sequence were verified (Figure 5).

To prepare pBI101vectors for subcloning, clones containing pBI101 LEA pro::eYFP constructs were grown overnight from glycerol stocks. Plasmids were extracted and 20 μ L of plasmid DNA was digested with 2 μ L of *Xba*I restriction enzyme in 5 μ L of 1X Tango buffer and 23 μ L of H₂O. The reaction was incubated at 37°C for 1 hour to allow for complete digestion. Dephosphorylation of the *Xba*I sites was carried out using 1 μ L FastAP thermosensitive Alkaline Phosphate (Fermentas) activated by incubating at 37°C for 10 minutes then inactivated by heating at 70°C for 5 minutes.

LEA cDNA was obtained by extracting RNA from *Brassica napus* seeds using the Trizol (Ambion) extraction method, followed by first-strand cDNA synthesis using the Phusion RT-PCR kit (New England Biolabs). First-strand cDNA was amplified by PCR using 1.5 μ L template cDNA, 0.2 μ L phusion enzyme, 4 μ L of 5x phusion HF buffer, 0.4 μ L of each 10mM dNTPs, forward primer, and reverse primer (with flanking *Xba*I sites) in a total volume of 20 μ L. DNA from PCR was run on agarose gels to confirm the size of the digest, cut out, and purified. The purified fragments were

ligated into plasmid pJET by using 0.5 μ L pJET1.2 blunt (50ng/ μ L), 0.5 μ L T4 DNA ligase, 1 μ L 10X ligase buffer, 6 μ L PCR product (calculated for each insert/vector combination), and 2 μ L of H₂O and incubated at 16°C overnight. The volume of DNA added was calculated as 3 x pJET capacity x (size of insert/size of volume). Positive colonies were identified and the inserts were sequenced using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Clones with the correct cDNA sequence were carried forward.

2.4.2 LEApro::GUS constructs

To remove YFP as a reporter gene, and use GUS, which is part of the original pBI101 vector, instead, LEApro::eYFP plasmids were digested with 1.5 μ L each of *Xba*I and *Sma*I restriction enzymes in 4.5 μ L 10X Tango buffer (Fermentas) with 25.5 μ L of H₂O and 12 μ L of sample DNA for a reaction volume of 45 μ L. Digestion reactions were incubated for 3 hours at 30°C (for *Sma*I) then for 1 hour at 37°C for *Xba*I. T4 DNA polymerase (Invitrogen) was used to blunt the *Xba*I site end by adding 1 μ L of 10mM dNTP and 0.5 μ L of T4 DNA polymerase directly to the digestion tubes. Reactions were incubated at 11°C for 20 minutes, then inactivated by incubating at 70°C for 10 minutes. DNA concentration was measured using a Nanodrop spectrometer.

2.4.3 Ligation of plasmids

25-50 ng of DNA were ligated with 2.5 μ L T4 DNA ligase (Fermentas) in 2.5 μ L 10X ligation buffer with H₂O for a final volume of 25 μ L. The reaction was incubated at 16°C overnight to allow sufficient time for complete ligation. The reaction was

inactivated by incubating at 65°C for 1 minute. An agarose gel was run (as outlined above) to confirm the ligation reaction. DNA concentration was measured using a Nanodrop spectrophotometer.

2.4.4 Bacterial strains and *E. coli* transformation

The *E. coli* strains DH5α (Invitrogen Cat. No. 18265-017) and DH10B (Invitrogen Cat. No.18290-015) were used to propagate plasmid constructs.

Plasmids were transferred into electrically competent DH10B (Invitrogen) cells by adding 1μL of DNA to 25μL of DH10B competent cells in a chilled cuvette. Electroporation was carried out at 1.8KV, 200 Ω, and 25 μF. Immediately after transformation, 1mL of SOC media was added to the cells and they were placed on ice. The cells were then incubated at 37°C for 1 hour with shaking at 225rpm. 100 μL and 20 μL of each transformation event were plated on LB Kan50 plates. The two sets of plates were incubated at 37°C overnight. Successfully transformed colonies(after confirmation by colony PCR as outlined below) were then picked with a sterile tip and placed in 3 mL LB Kan50 broth and incubated; shaking, overnight at 37°C.

When chemically competent DH5α were used, 1 μL (30-65 ng) of DNA was added to 60μL of DH5α chemically competent cells and stored on ice for 30 minutes. Once all samples were ready, all tubes were placed in a heating plate at 42°C for 90 seconds, then transferred to an ice bath and incubated for 1-2 minutes. 800 μL of SOC were added to each tube and warmed up to 37°C, then incubated at 37 °C with gentle shaking. Samples were then plated on LB Kan50 plates and incubated at

room temperature overnight then incubated at 37°C. Successfully transformed colonies (after confirmation by colony PCR as outlined below) were then picked with a sterile tip and placed in 3 mL LB Kan50 broth and incubated; shaking, overnight at 37°C.

2.4.5 Colony PCR

Colonies from each transformation event were picked using a sterile tip and placed in 25 μ L 0.5% TE, 0.1% Tween 20, incubated at 95 °C for 10 minutes, spun for 45 seconds to remove any debris. Then 2 μ L were used to perform the PCR amplification process. PCR was carried out using 0.2 μ L of an in-house Taq polymerase with 2 μ L of template in a 20 μ L reaction, using 0.4 μ L of each 10mM dNTP, forward primer and the reverse primer. PCR amplification were conducted on a thermocycler using the following program: 3 min at 94°C (denaturation step), followed by 33 cycles of 15 seconds at 94°C, 15 seconds at the optimized annealing temperature for each set of primers; usually between 55°C and 60°C (annealing temperature), and 1 min at 72°C (extension step), then one incubation of 7 minutes at 72°C. The reaction was kept at 10°C. Using the full volume of each reaction, 3 μ L of loading dye were added to each well of 1% agarose gel for electrophoresis and an image captured to confirm presence of DNA.

2.4.6 Plasmid minipreps

400 μ L of diatomaceous earth (Sigma) were dispensed into columns and left to settle while 1500 μ L of overnight bacterial cultures were spun down at maximum

speed for 1 min. The supernatant was removed and 175 μ L Tris/RNase buffer was added and the pellet gently suspended. 175 μ L of lysis buffer (1% SDS, 200 mM NaOH) was added and mixed by inverting the tubes until the lysate became clear (not more than 5 minutes). Another 250 μ L of binding buffer (6 M guanidine hydrochloride) was added and the contents mixed. The tubes were spun at maximum speed for 10 minutes. Before the lysates tubes finished spinning, the diatomaceous earth tubes were spun at 5K for 1 min to settle the matrix. Clear lysate was added to the diatomaceous earth columns and set to settle, then spun at 5K for 1 min, flow-through was removed and the columns washed with 600 μ L of ethanol wash solution (80% isopropanol). The columns were spun for another 1 min and flow-through removed, the spinning was repeated to ensure all residual ethanol has been removed. DNA was eluted from the columns in 30-50 μ L of TE buffer (10 mM Tris.HCL buffer, pH 8.0) or RNase free water. Five μ L of eluted DNA was loaded onto an agarose gel to confirm the presence of plasmid DNA.

2.4.7 Agarose gel electrophoresis

Agarose gel electrophoresis of plasmid mini preps and PCR-amplified DNA products were carried out using 1% (w/v) agarose gels. Gels were prepared by using 1g of agarose in 100mL of TAE buffer (40mM Tris acetate, 1mM EDTA (pH 8.0)). The solution was heated in a microwave for 1 minute, swirled gently, and reheated for another 1 minute. The solution was then poured into gel casts and left to cool and solidify. Samples were mixed with a dye-containing loading buffer before loading onto the gel. A DNA ladder (Fermentas) was used to determine fragment sizes. Gels

were usually run at 90-100V for approximately 1 hour. Gels were then incubated in an (10mg/ml) ethidium bromide solution on a shaker for 20 minutes, destained in H₂O, and the gel was imaged using a GelDoc (BioRad) imager system.

2.5 Analysis of transformed plants

For each construct transformed, 5-10 transgenic lines were carried forward, after initial analysis and when the transgenic lines behaved the same, it was narrowed down to three independent lines. The lines were numbered based on their previous generation, they were carried to the third generation. Thus the numbering was carried forward as 66-1-1-1-1 (66-T₀-T₁-T₂-T₃) for example. The lines that were carried forward were:

- 66-3-11-9-1
- 66-3-1-1-1
- 66-3-1-2-1
- 248-3-2-5-2
- 248-3-1-2-1
- 248-3-1-1-1
- 411-3-1-1-1
- 411-3-1-2-1
- 411-3-2-3-1

2.5.1 YFP Imaging

Transformed seeds/seedling containing YFP constructs required detailed preparation for microscope analysis. Siliques used for microscopy were opened and the seed coat removed (where it was present), Dry seeds were imbibed in water for 10-15 minutes to facilitate the removal of the seed coat. Seeds were placed on a slide with water and the coat was removed using fine forceps and a fine scalpel blade. Multiple seeds were prepared from each sample as the chances of damaging the seeds were very high. Seeds were mounted with fluoromount G mounted media and analyzed under a Zeiss; axio scope A1 with a colibri 2 light system under 555nm filter, fluorescent microscope as soon as possible after seed coat was removed. Where all samples showed the same pattern, one was chosen as a representative of the 3 lines.

2.5.2 GUS imaging

GUS staining solution was always prepared fresh before each staining event. The staining solution required 0.1 M of NaPO₄ pH 7.0, 10mM of EDTA, 0.1% Triton X-100, 1.0mM K₃Fe(CN)₆, 2.0mM X-Gluc (5-bromo-4-chloro-3-indolyl glucuronide), and H₂O. For a 1.0mL final volume of staining solution 100μL of 1.0M NaPO₄ stock solution, 20.0μL of 0.5M stock EDTA, 10.0μL of 10% Triton X-100, 20.0μL of 50mM K₃Fe(CN)₆ and 20.0μL X-Gluc were mixed. 50mM K₃Fe(CN)₆ was prepared by using 82.3mg of ferricyanide powder (MW 329.26) and H₂O to a final volume of 5mL. X-Gluc (Sigma) was prepared by adding 50mg of X-Gluc powder (MW 521.8) with

water up to 1.0mL. Both ferricyanide and X-Gluc were stored at -80°C in smaller aliquots and thawed immediately before preparing the final solution. Seedlings or seeds were fully immersed in staining solution in either 6 or 12 well plates depending on the size of samples and incubated, covered for 16-24 hours or until colour develops. Staining solution was removed and samples were washed with several changes of 50% ethanol, incubating 12 hours between ethanol washes. Samples were imaged under a Zeiss bright field microscope with a Zeiss camera on a slide with water to prevent dehydration. Where all seeds showed the same pattern, one was chosen as a representative of the 3 lines.

2.5.3 Expression analysis by quantitative real-time PCR

2.5.3.1 DNA-free RNA isolation from seeds and siliques

After trying several methods of RNA extraction from seeds, and due to the limited starting material, the only successful method that yielded reliable results was using Ambion's RNAqueous-Micro Kit (AM1931) as well as RNAqueous-96 Kit (AM1920). RNA extraction was carried out as outlined in the Ambion manual. Upon collecting in liquid nitrogen, samples were immediately stored at -80°C until needed. For analysis, seeds were placed in 1.5mL Eppendorf tubes and kept in liquid nitrogen. Using a pestle (blue polypropylene pellet pestles by Sigma) chilled in liquid nitrogen, the seeds were crushed on the side of the Eppendorf tube. Immediately after crushing the seeds 1 volume per unit mass of Plant Aid Solution (usually 50µL for 0.5g) and 200µL lysis/binding solution was added. Samples were homogenized using a hand-held homogenizer for 10 seconds. Tubes were centrifuged at 1320xg

for 5 minutes (to pellet seed coat and other insoluble material). Supernatants were transferred to new RNase free tubes and 75 μ L of 95% ethanol was added and vortexed briefly. Lysate/ethanol mixtures were then added into the microfilter cartridge assembly (provided with the kit) and centrifuged (bench top centrifuge) for 15 seconds at maximum speed (repeated when needed until all solution passed through filter). 180 μ L of wash solution 1 (provided with kit) was added, then tubes centrifuged for another 15 seconds. This was followed by another wash of solution 2/3 (provided with kit), then centrifuged for 15 seconds, repeated with a second wash of solution 2/3 and centrifuged for 15 seconds. Flow-through was removed from the collection tubes and the tubes were spun for 1 min to ensure all solution has been dried out. The filters were then transferred to new tubes and eluted in 10 μ L of heated elution buffer and left to set on the bench for 1 minute then centrifuged for 30 seconds, followed by another elution of 20 μ L. Eluted RNA was treated with DNase, to ensure the absence of DNA, by adding 1/10 volume of 10X DNase I buffer (3 μ L DNase I buffer + 1 μ L DNase I). The reaction was then incubated at 37°C for 20-30 minutes (at which time the DNase inactivation reagent was removed from -20°C and allowed to thaw at room temperature). The DNase inactivation reagent was vortexed and 1/10 volume was added to the solution and incubated at room temperature for 2 minutes, vortexing once to disperse the DNase inactivating reagent. It was then centrifuged at max speed for 1.5 minute to pellet the DNase inactivating reagent. The supernatant was transferred to new RNase free tubes, and the RNA concentration measured using a Nanodrop spectrophotometer. The RNA was stored at -80°C if not

used. For the RNaqueous-96 Kit, the same procedure was followed but sample preparation was carried out in 1.5 mL Eppendorf tubes where the lysis solution was added and samples homogenized. The solutions were transferred to 96 well plates (provided with the kit) and the extraction was carried as outlined above.

2.5.3.2 First-strand cDNA synthesis using Reverse Transcriptase

Once the concentration of the RNA had been determined, 1 μ g of mRNA, 1 μ L of oligo dT, 1 μ L dNTP and sterile dH₂O to a volume of 12 μ L were mixed together in an Eppendorf tube. The mixture was heated to 65°C for 5 minutes and quickly chilled on ice for 2 minutes, briefly centrifuged to collect contents and a master mix consisting of 4 μ L of 5X first-strand buffer, 2 μ L 0.1M DTT, and 1 μ L RNaseOUT added to the reaction. Contents were mixed gently and incubated at 37°C for 2 minutes. 1 μ L of SSIII RT was added and mixed by pipetting up and down, and incubated at 37°C for 50-60 minutes. The reaction was inactivated by heating at 70°C for 15 minutes. The cDNA was directly used as a template for PCR amplification.

2.5.3.3 cDNA amplification by PCR

Two μ L of first strand cDNA were mixed with 4 μ L 5X phusion HF buffer, 0.4 μ L 10mM dNTP, 0.4 μ L of 10 μ M forward primer, 0.4 μ L of 10 μ M reverse primer, 0.2 μ L Phusion DNA polymerase and 12.6 μ L sterile dH₂O to a total volume of 20 μ L. A negative control where no cDNA was added was included, as well as a positive control with actin primers. PCR was run with 30 seconds at 98°C (denaturation step), followed by 35 cycles of 10 seconds at 98°C, 30 seconds at the optimized annealing

temperature for each set of primers, usually between 55°C and 72°C, and 30 seconds at 72°C(extension step), followed by a final step of 5 minutes at 72°C.

2.5.3.4 Quantitative PCR

Two primers sets were designed for LEA 411, LEA 248(same sequence for LEA 66 primers), YFP, and tubulin (house keeping primer) using Primer Express software (provided with the qPCR machine) (Table 3). Once RNA was extracted as outlined above and converted to cDNA, their concentration was measured using a Nanodrop spectrophotometer and their concentration was adjusted to 1 μ g. 80 μ L of RNase free water was added to all tubes and labeled as a concentration of 1. 10 μ L of each template were pooled together and a 5-fold serial dilution was prepared. All primer sets, YFP 1, YFP 2, 248-1, 248-2, 411-1, 411-2, along with house keeping primer tubulin, and elongation factor (ELF2) were tested with the serial dilutions to determine the best combination of primer to dilution to carry out the experiment. 10 μ L of 2X SYBR green Master Mix, 0.4 μ M of both forward and reverse primers, 6.7 μ L of H₂O, and 2.5 μ L of template were added to each well, and the 96 well plate was run with standard curve setting. Choosing a CT (Threshold Cycle) value around 25 but less than 30 is ideal. Once the successful primers (248-2, 411-2, and tubulin) and dilutions were determined cDNA was diluted and run with the appropriate primers using the volumes outlined above but with the comparative CT cycles. For the LEAs being tested, ECP 66, ECP 248, or LEA 411, the samples were the different time points T3 - T48. Biological groups from separate plants of the same transgenic line were the repeated triplicate samples of each time point T3-1, T3-2

, T3-3, reference sample was T0 (Time 0 before stress started), endogenous control were the ELF2 samples and the passive reference was ROX from the PCR master mix. Once the qPCR cycle was completed results were analyzed by generating RQ vs. time under stress graphs (figures 6-8,11-19). Where RQ is relative quantification (relative to the untreated control gene transcript copies) was calculated as:

$$RQ = 2^{-\Delta\Delta C_t}, \text{ where } \Delta\Delta C_t = \Delta C_{t\text{sample}} - \Delta C_{t\text{Reference}}, \Delta C_t = C_{t\text{gene test}} - C_t$$

endogenous control (ELF2 values), and ΔC_t SD as Standard deviation. Graphs were designed to determine 1) The expression of LEA genes in developing seeds of *Arabidopsis* and 2) Expression of LEA genes in response to stress treatment.

Table 3: Sequences of primers used for RT-qPCR analysis

Name	Sequence	Tm	length
411LEApro-cDNA-yfp 37F	TCGCAGCTTCTGCCAAGTCT	59	70
411LEApro-cDNA-yfp 106 R	GTCATCTTCTCCGCCTTTTCC	59	
248ECPpro-cDNA-yfp 904F	CAAAAGACAAGAGAGAGTACTGATTCTG	58	95
248ECPpro-cDNA-yfp 998R	ATTGTCCCCTTCGCTTCATTAC	58	
Q YFP 582 F	CTGCTGCCCCGACAACCA	59	73
Q YFP 654 R	TGTGATCGCGCTTCTCGTT	59	

3 Results

Figure 2: phylogenetic tree showing the distance between the different LEA promoters constructed using the geneious software

The aim of this study was to determine the expression patterns of *Brassica napus* LEA genes by analyzing LEA promoter::reporter gene constructs in *Arabidopsis thaliana* during seed development as well as in response to abiotic stress. Initially, seven LEA promoters representing three different LEA groups were

identified from *Brassica napus*, using the genome walker protocol were *Brassica napus* libraries were constructed (by a previous student) using the GenomeWalker Universal Kit (Clontech Laboratories). The GenomeWalker is a method for finding unknown genomic DNA sequences flanking a known sequence; such as the *Brassica napus* cDNA (published data for LEA protein). The first step is to construct libraries. These libraries are pools of uncloned; adapter ligated genomic DNA fragments that have been digested by different restriction enzymes. These DNA fragments are then used as templates for PCR. The PCR products; which have a known sequence at the 3' end (identified by the cDNA) and where the region upstream of the nested, gene-specific primer can be sequenced to provide a complete promoter sequence that can be used for further analysis. In this case, the sequenced products are the seven LEA promoters, which were all the promoters isolated.

Their expression patterns were analyzed as shown above. Since some promoters gave similar responses, it was decided that only three promoter constructs would be carried forward, to focus more on LEA groups that have so far been described less in literature. As a dehydrin, LEA 212 promoter was dropped since dehydrins are the most studied LEA group and ample characteristics are already known. LEA 111, and LEA 314 were dropped due to their close proximity to LEA 212 (Figure 2). Initial LEA 411 and LEA 421 data showed very close similarity and since the two are related, LEA 411 was carried forward as a representation of the two LEA groups. Although ECP 66 and ECP 248 are closely related, very little is known about them, thus they were both carried forward for further analysis. Genomic sequence

analysis revealed that the cDNAs of ECP63 and ECP248 shared a high degree of homology but their respective promoters; ECP66pro and ECP248pro were quite different. In fact, overall homology between these two promoters is only 21.1% although the region close to the start of the cDNAs are not so different as shown in appendix figure A4.

Table 4: Known characteristics of LEA and LEA-like genes used in this study modified from (Hundertmark & Hinch, 2008)

Name	LEA	Accession #	AGI Code	Predicted localization
411	1	AY766378, AY572958	AT5g06760	Seed and under salt stress
ECP 66.1	3	FJ147551	AT2g36640 lea 4	Seed only
ECP 248	3	FJ147551	AT2g36640 lea 4	Seed only

Figure 3: construct design of LEApro-YFP in polylinker of pBI101 vector using HindIII, XbaI, and BamHI restriction enzymes.

Figure 4: construct design of LEApro-GUS in polylinker of pBI101 vector using XbaI, and BamHI restriction enzymes.

Identified *Brassica napus* LEA promoters were fused in frame fusion to a fluorescent marker gene enhanced yellow fluorescent protein (eYFP) (Figure 3) and their expression in *A. thaliana* was analyzed. LEApro::eYFP microscopy analysis (data not shown) showed very low fluorescence that were not distinguishable and it did not provide precise spatial localization of LEA, thus deemed LEApro-YFP constructs not suitable to demonstrate the developmental expression patterns of the LEA proteins. To remedy this, another strategy was employed consisting of 1) changing the reporter gene from LEApro::YFP to the more robust GUS reporter gene (Figure 4) and 2) applying constructs consisting of a LEA promoter with the corresponding LEA open reading frame joined in frame with the eYFP reporter gene (Figure 5). While the LEApro::GUS constructs were used to analyze expression patterns on a whole tissue scale, in particular with respect to environmental stress analysis, the LEApro-LEA-ORF-eYFP assembly was thought to be better suited to

show the patterns during seed development as well as the localization of LEA proteins within the cell whereby the placement of the LEA cDNA behind the LEA promoter should ensure the proper cell-type expression of the fusion proteins; ORF refers to open reading frame, which is a segment of the nucleotides that spans from the start codon to the stop codon but does not include the stop code, and is believed to contain important information that are needed for the correct translation of the protein followed in the coding sequence. With these two new approaches, the constructs for the three LEA promoters that were chosen for the task; LEA 411, ECP 66, and ECP 248 were constructed. In this study, since ECP 66 and ECP 248 cDNA shared a high degree of homology and the main purpose of the study the the promoter expression, ECP 66 cDNA was fused to both ECP66pro and ECP248pro to study the respective promoters, LEA 411 cDNA was fused to the LEA411promoter.

Figure 5: construct design of LEApro-LEAcDNA-YFP in the polylinker of pBI101 vector using XbaI restriction enzyme.

3.1 LEA expression during seed development

3.1.1 qPCR analysis of *Brassica* promoters-ORF-eYFP, of LEA 411, ECP 66, and ECP 248 in *Arabidopsis* seeds

ECP 66 showed a steady increase in accumulation starting from day 5 after flowering and up until the maturation stage, reaching its highest accumulation in mature seeds (Figure 6), it shows the highest accumulation levels compared to others. Although it is not known if expression would have been evident before 5 DAF, assuming that LEA protein accumulate late in seed development day 5 is considered towards the middle stage of seed development and picked as a good starting point. Figures presented below are showing the accumulation trend for one biological replicate with error bars representing the standard deviation along its three technical replicates. For a more detailed representation of all biological replicates refer to the complete tables in the appendix section.

Figure 6: accumulation pattern of one biological replicate of ECP66pro-ORF-eYFP during seed development obtained from qPCR analysis. Error bars present the standard deviation among its three technical replicates

Figure 7: accumulation pattern of one biological replicate of ECP248pro-ORF-eYFP during seed development. Error bars present the standard deviation among its three technical replicates.

Figure 8: accumulation pattern of one biological replicate of LEA411pr ORF-eYFP during seed development. Error bars present the standard deviation among its three technical replicates.

Despite belonging to the same group, ECP 248 promoter showed a weak accumulation at 5 DAF and increased through seed development reaching its strongest accumulation around 15 DAF but decreasing significantly in the mature seeds to its starting levels on 10 DAF flowering (figure 7). It should be noted that the graph is on an exponential scale, and therefore the changes in accumulation levels are large as the values increase.

Despite LEA 411 belonging to the group 1 LEA proteins, different from ECP66 group 3, it showed the same accumulation pattern as ECP66. LEA 411 accumulation started at 5 DAF and slowly increased as the seeds matured reaching its highest accumulation (relatively low compared to others) in the mature seed (figure 8).

3.1.2 Confocal analysis of *Brassica* promoters-ORF-eYFP, of LEA 411, ECP 66, and ECP 248 in *Arabidopsis* seeds

To obtain a better understanding of how LEA proteins are expressed in developing seeds, confocal microscopy analysis was carried using embryos of transgenic *Arabidopsis thaliana* with the seed coat removed. Although the qPCR analysis (Billinton & Knight, 2001) indicated that the expression of LEA proteins starts at 5 DAF, we did not observe any accumulation of YFP by confocal microscopy at that stage. Under fluorescent microscopy, auto-fluorescence of plants is a common observation where endogenous molecules or media that absorb light in the near-uv spectrum produce some untargeted fluorescence under the proper wavelength; these include chlorophyll, flavin, and structural components to name a few examples

(Billinton & Knight, 2001). Factoring in the green colour auto-fluorescence of embryo tissue, and the low levels of LEA fluorescence as the signal begins to accumulate, it was challenging to observe the fluorescence of LEAs at low levels.

Once the accumulation increased, it was easier to discern eYFP signal from the background signal. Judging by the fluorescence intensity from (figure 9), ECP66 was accumulating the most, showing accumulation in the whole cotyledon, followed by ECP248 that produced a faint signal at 10 DAF with the signal increasing as the seed matured. Similar to ECP66, expression of ECP248 was observed in the whole cotyledon. LEA411, on the other hand, displayed a unique accumulation pattern. LEA411 signal was observed around day 15 after flow but only in the radicle of the embryo and not in the whole cotyledon.

Figure 10 shows more confocal images of the same cotyledons in Figure 9, but with a higher magnification. In an effort to localize eYFP expression to sub-cellular structures, different developmental stages of seeds were studied under the 40x and 63x objectives of a confocal microscope. Despite using a high magnification with a very sensitive microscope, the analysis was challenging. When working with a 40x magnification lens, the camera is able to zoom in and these zoomed-in images shed some light on the expression patterns. From 15 days after flowering, different accumulation patterns arise.

Figure 9: Confocal Microscopy Data showing the expression of LEA genes during the developmental stages of *Arabidopsis thaliana* seeds (seed coat removed) under 10X objective. Top image is under YFP filter, bottom image is bright field. Scale bar 400μm.

40x Objective	WT	ECP 66	ECP 248	LEA 411
15 Days After Flowering 40 x
40x-zoom 4
20 Days After Flowering 40x

40x-zoom 4
Mature Seed 40x
Dry Seed 40x-zoom 4

Figure 10: confocal microscopy showing the expression of LEA genes during the developmental stages of *Arabidopsis thaliana* seeds (seed coat removed) under different objectives.

ECP66-YFP showed localization in the cytosol but not in the vacuoles. ECP 248-YFP was localized in the vacuoles and not in the cytosol. Similar to ECP 248, LEA 411-YFP was found in vacuoles not the cytosol. All three fusion proteins showed high accumulation levels in the mitochondria. Embryos imaged at 20 DAF under the 40x objective suggest that ECP 248 does not accumulate in the nucleus (small dark area in the cell figure 10). An interesting observation not directly relating to the localization of LEA protein is thickness of the cell wall, the

cell wall in ECP 248 and LEA 411 transgenic embryos is much thicker compared to that of ECP 66 and untransformed embryos (WT) as shown in figure 10 for the mature seed at 63x magnification. This might be due to other stress response the plants have in place, it might also explain why ECP66 accumulation is stronger since it doesn't have the aid of a thicker cell wall to help against stress.

Due to the small size of the embryo, whole embryos were used in the specimen unfortunately that resulted in some sections being out of focus since the whole embryo was not in the same plane, in future single sectioning should be used.

With the temporal expression of the studied LEA constructs outlined and with an idea of their spatial distribution formed, the next step was to analyze their expression under environmental stresses.

3.2 LEA expression under environmental stresses

3.2.1 qPCR analysis of *Brassica* promoters-ORF-eYFP of LEA 411, ECP 66, and ECP 248 constructs in response to salt, cold, and drought stresses in *Arabidopsis* seedlings

To obtain representative data of the effect of environmental stresses on the expression of LEA proteins, qPCR was carried out using 14-day-old seedlings under three different stresses; salt (NaCl), drought (Mannitol), and cold (incubation in cold chamber at 4°C). Figures presented below are showing the accumulation trend for one biological replicate with error bars representing the standard deviation along its three technical replicates. For a more detailed

representation of all biological replicates refer to the complete tables in the appendix section.

3.2.1.1 Salt and drought stress

ECP66 accumulation under salt stress started as soon as three hours under stress, reaching its minimum accumulation at 24 hours under stress before increasing to its highest accumulation at 48 hours under stress (figure 11). ECP66 expression under mannitol stress showed a different accumulation pattern, accumulation was low but steadily increased from three hours under stress reaching its maximum at 48 hours (figure 12).

ECP248 seedlings seemed to be less responsive to salt stress until 48 hours under stress (figure 13), it showed a small accumulation by three hours under stress that increased slightly by six hours only to decrease by twelve hours and reach its lowest accumulation level at 24 hours under stress before increasing by 48 hours under stress. This same pattern was not observed for ECP248 seedlings under mannitol stress, where accumulation was the highest at 3 hours, decreased rapidly by six hours and decreased even more by twelve hours before increasing slightly by 24 and 48 hours under stress (figure 14)

When exposed to NaCl LEA411-YFP did not accumulate in significant levels until 24 and 48 hours under stress (figure 15), this simply suggests that LEA 411 is not needed until prolonged stress periods. A similar pattern is observed from LEA411 under mannitol stress, despite the higher accumulation at 12 hours compared to its starting level, accumulation of LEA411 is relatively low under mannitol stress and does not last long as it decreases by 24 hours (figure 16)

Figure 11: accumulation pattern of one biological replicate of ECP66pro-ORF-eYFP under salt stress. Error bars present the standard deviation among its three technical replicates.

Figure 12: accumulation pattern of one biological replicate of ECP66pro-ORF-eYFP under Drought stress. Error bars present the standard deviation among three technical replicates

Figure 13: accumulation pattern of one biological replicate of ECP248pro-ORF-eYFP under salt stress. Error bars present the standard deviation among its three technical replicates

Figure 14: accumulation pattern of one biological replicate of ECP248pro-ORF-eYFP under Drought stress. Error bars present the standard deviation among its three technical replicates.

Figure 15: accumulation pattern of one biological replicate of LEA411pro-ORF-eYFP under salt stress. Error bars present the standard deviation among its three technical replicates

Figure 16: accumulation pattern of one biological replicate of LEA411pro-

ORF-eYFP under Drought stress. Error bars present the standard deviation among its three technical replicates.

3.2.1.2 Cold stress

The accumulation pattern of ECP66 under cold stress showed a steady increase reaching its maximum at 24 hours. Enough data were not recorded for the accumulation at 48 hours due to some seedlings not surviving prolonged exposure (figure 17).

The accumulation of the YFP in ECP248 seedlings under cold stress showed a pattern that was almost the opposite of the two other stresses. Here, expression increased gradually reaching its maximum level at twelve hours, but decreased again to approximately the starting levels (figure 18).

Data from LEA411 plants under cold stress (figure 19), it begins accumulation at three hours and increases slightly by six hours but reaches its maximum level at twelve hours before decreasing to levels lower than its initial accumulation level under prolong stress.

Figure 17: accumulation pattern of one biological replicate of ECP66pro-ORF-eYFP under cold stress Error bars present the standard deviation among its three technical replicate

Figure 18: accumulation pattern of one biological replicate of ECP248pro-ORF-eYFP under cold stress. Error bars present the standard deviation among its three technical replicates.

Figure 19: accumulation pattern of one biological replicate of LEA411pro-ORF-eYFP under cold stress Error bars present the standard deviation among its three technical replicates

3.2.2 GUS analysis of *Brassica* promoters-GUS of LEA 411, ECP 66, and ECP 248 constructs in response to salt, cold, and drought stresses in *Arabidopsis* seedlings

To have a better understanding of the expression pattern in the seedling, analysis was carried out using whole seedling. Due to the size of the seedling, confocal microscopy was ruled out and GUS analysis was carried out using a bright field microscope.

3.2.2.1 Salt stress

Not surprisingly, accumulation of GUS in ECP248 seedlings under salt stress is absent, which confirm the low expression levels observed from qPCR data, but the absence of expression at 48 hours cannot be explained.

The expression pattern observed for both ECP66 and LEA411 showed an initial increase of expression followed by a slight decrease at 12 hours before increasing again. After 3 hours under stress, ECP66 expression is only observed in the cotyledons as well as in the hypocotyl. Past the 12 hour mark, expression extends to all leaves.

Interestingly, LEA 411 seedlings show accumulation in the main root and all leaves after 3 hours and 6 hours under stress. Accumulation extended to the lateral roots as well. At the 12 hour time point, the signal was decreased, but increased again in all leaves and roots at 24 and 48 hours. Again, this expression pattern – with respect to intensity - is similar to the qPCR result.

3.2.2.2 Drought stress

Similar to what had been observed with the salt treatment, GUS expression in ECP248 seedlings was again significantly weaker compared to ECP66 and LEA411. GUS expression was apparent in the hypocotyl and the main root of the seedling at 3 hours and 6 hours under stress, but the signal decreased with prolonged incubation. Again, ECP66 showed strong expression at the 3 hour mark in the cotyledon, increasing in intensity and reaching all leaves by 24 hours under stress.

Expression of LEA411 after at 3 hours was visible in the hypocotyl and leaves and noticeably increased with time, reaching a maximum after 48 hours under stress. This pattern is again comparable to the qPCR result. Interestingly, the GUS stain in the cotyledons at the 48 hour mark appeared to be fading.

Figure 20 : GUS staining showing the expression of LEA gene 14-day-old *Arabidopsis thaliana* seedlings in response to Salt stress. Scale bar 4 mm

Figure 21: GUS staining showing the expression of LEA gene 14-day-old *Arabidopsis thaliana* seedlings in response to Drought stress. Scale bar 4 mm.

3.2.2.3 Cold stress

Under cold stress conditions, seedlings grew long roots, possibly to accommodate for the stress. Strong GUS staining could only be observed for the ECP66 construct whereby staining was mainly localized to the cotyledons and the hypocotyl. GUS staining was visible in the seedlings of ECP248 and LEA411 but overall staining was weak and a clear temporal expression pattern could not be discerned. As was the case with ECP66, GUS staining was mainly confined to cotyledons and the hypocotyl. It is possible that exposure to cold influences the activity of the GUS enzyme. Since analysis was carried out using whole seedlings under a dissection microscope, close ups of the root were not possible to enable the localization of the deposition to specific layers of the cotyledons, and root.

Figure 22: GUS staining showing the expression of LEA gene 14-day-old *Arabidopsis thaliana* seedlings in response to Cold stress. Scale bar 4mm.

3.3 Motif analysis in promoter sequence

For the purpose of this study, promoter analysis was approached in two ways; lining up the three promoter and detecting motifs that are similar or different between them to explain their expression and/or localization patterns, and the second approach was analyzing the motifs in each promoter individually to determine if it has unique motifs that aid in its function. Motif analysis was carried out using, Plant cis-acting regulatory DNA elements (PLACE) database, (updated on Jan.8.2007) (Higo et al., 1999) last accessed on Apr.1.2014

Although no Absciscic acid (ABA) was directly used as a stress factor, one of the major pathways plants respond to stress is via an ABA-dependent pathway as outlined in Figure 24. ABA regulates the expression of many genes functioning in dehydration tolerance in seeds as well as vegetative tissue. ABA induced genes contain a conserved sequence named ABA responsive element (ABRE) which function primly during late embryogenesis but has also been expressed in maturing seeds as well as dehydrated vegetative tissue (Higo et al., 1999). Table A13 (rows 4-9) indicates that all three promoters do contain ABREs, while their ABRE content differ slightly, they all are capable of ABA-dependent expression. Numerous binding site and coupling elements are also found in the promoters; such as ACGT-core (Table A13 row 11) with is a sequence requirement of motif A in ABRE in rice, LEA 411 contains a binding site for ABRE binding factors (Table A13 row 49). Another ABA-dependent pathway

that is carried out by all three promoters is the Adaptive MYB response (Table A13 rows 33,78-86). Both ECP66, and LEA 411 also contain the DRE2 (Table A13 row 43) core found in maize expressed during late embryogenesis, which is induced by ABA (Higo et al.,1999). Both also contain the DRE/CRT core with produces a rapid response through the ABA-independent-pathway (Table A13 row 44).

With respect to location, all promoters contain a site that is specific for storage protein in embryo and endosperm (Table A13 row 28) along with an E-box of napA storage-protein gene of *B. napus* (Higo et al.,1999) (Table A13 row 48). Seed storage protein expression is found to be regulated temporally during embryogenesis and has been shown to be restricted to seed tissues such as the cotyledons or the endosperm (Bustros, 1989). Another site found in all three promoters is Dof core site (Table A13 row 41) which is unique to cDNA encoding PBF; an endosperm specific Dof protein that binds to the prolamin box. A site unique to LEA 411 is L1BOXATPDF1 (Table A13 row 71) which is an L1 box involved in L1 layer-specific expression.

RY-repeats; found in all three promoters (Table A13:109-112), have been found necessary in seed tissue for positive regulation but once it is associated with the vegetative tissue it carries a negative regulation role (Bobb et al., 1997). They are mostly associated with maturation in embryos where accumulation of proteins, lipid, oil and carbohydrate reserves are synthesized, and are usually found upstream of dicot genes for storage proteins, lectins and oil-body proteins

specifically expressed in seeds (Bobb et al., 1997). Interacting with neighboring CCAC-box it is associated with seed-specific expression. CCAC is also present within the Coupling Element 1 (CE1); which is a cis-acting element involved in ABA regulation occurring on many plant promoters in seed and non-seed tissue. Published sequences of ABI3-regulated A.t. gene promoter included an RY/CB complex upstream of the TATA-box, indicating the ABI3 also regulates RY-containing genes in A.t. This strengthens the suggestion that LEA411, ECP 66, and ECP248 are associated with seed-storage proteins.

The presence of NAPINMOTIFBN (Table A13 row 88) in ECP 248 promoter suggests that it might be localized in the nucleus or the nuclear membrane, since the gene interacts with a protein that is found in the crude nuclear extracts from developing *Brassica napus* seed. In addition to seed localization, POLLENILELAT52 (Table A13 row 98) was present in all three promoters; it is one of the two co-dependent regulatory elements responsible for pollen specific activation of Tomato *lat52* gene. Interestingly, ECP 66 contains a GBOXPC (Table A13 row 57); A G-box that confers relatively high expression in roots.

ECP 248 carried unique motifs, CIACADIANLELHC (Table A13 row 37); a region needed for circadian expression of tomato *Lhc* gene was present in its promoter as well as an evening element EVENINGAT (Table A13 row 51) required for circadian control of gene expression (Higo et al., 1999). Indirectly related to the circadian clock are Light-regulated genes, a G-box (Table A13 row 55) conserved sequence usually found upstream of these light-regulated genes

was identified in LEA 411, it also contained IBOXCORENT (Table A13 row 67); an I-box core motif associated with light responsive promoter regions. ECP 66 contains a positive light regulatory element (Table A13 row 73), ECP 66 and LEA 411 also contained two shared sequences, one of the sequences is usually over-represent in light-induced cotyledons and root specific genes, and the other a computationally identified phyA-induced motifs, while all three promoter contained GT1CONSENSUS; a GT-1 binding site (Table A13 row 58) found in many light-regulated genes. LEA 411 promoter also contained a low temperature responsive element (Table A13 row 74), and a AP2 transcriptional activator of DNA binding regulated by Temperature (Table A13 row 39), in addition to a core of low temperature responsive element present in all three promoters (Table A13 row 75). All three promoter also carry MYC recognition site, where ICE1 binds, regulating the transcription of CBF/DREB1 genes in the cold.

Interestingly, all promoters carry the sequence for other hormone responses; such as presence of an auxin response factor (ARF) in LEA 411 promoter (Table A13 row 20), and ASF-1 binding site in ECP 248 and LEA 411 (Table A13 row 22). Other examples are a core of TGAC-containing W-box (Table A13 row 149), and an ethylene responsive element was also found in LEA 411 promoter (Table A13 row 50). They also contain a core of copper-response element (CuRE) (Table A13 row 40),

All promoters also contain a “CCAAT box” which are usually found in promoters of heat shock proteins (Table A13 row 34), ECP 66 contains a motif

usually found in up-regulated genes after main stem decapitation in A.t. (Table A13 row 142), and ECP 248 contains an unfolded protein response cis-acting element (Table A13 row 143).

ECP 66 promoter	-----CCACCGTTTTTTCCTAATG--GGCCTGTGATGATGACCCAATA--AATTAA-----ATAGGATTG-A--ATACGTT---TTAATCATATTTT	DRE2
ECP 248 promoter	-----TCCAAG---CTTAA-CAA-----GCCCCGAACATCACAACATACGAACCTCAACCAAGCCCATCTTCATG-AC-ACACG----CATG--TACTATTTT	
LEA 411 promoter	AATGCAGGACATGGGAGTCGAAGAACCTTAAGCAATGGTGGTCTAGGAGATTGCCAATGACTCAGCCAAGAAGAGACCATCTAGTAGCACAAAACGGTGGCGGAGCAAGCTTTT	DRE2 DRE2
ECP 66 promoter	120 130 140 150 160 170 180 190 200 210 220 230 CAGT-AGT-----CTG-----AG-----TG TGT-----TCA-AATGA---TG TCT-TTTT--ACG---TAGTACAGTC-----TCCAATGCCTCT-TAACT	DRE2
ECP 248 promoter	CTGT-AGTACTTCACGAAC TG---ACACAGAAATATACATCT-----TCAGAAATAT---TG TCTT--GTT--AAGATTAACTAAAATC---AAAACCAAGAAGTATATAAAG	MYB
LEA 411 promoter	CGATGAGA-----TCGAAC TGTTCAACACAGACTC-GCTTGTAAGAGGTAAGCATCTCTCTGTCTCTCCGTGGATGAGTTTAACTTTTGCTTAGTCTGAAGTGT TTTTGCGT	
ECP 66 promoter	240 250 260 270 280 290 300 310 320 330 340 GTCA T-----GCTAAATCAATAATTTAA-----CAGTGC---AGTATAATTGCTACAAACGATCTTGCAACGGCTCAGGCG-----AAGCTCAATGAGCCTGAGA	DRE2
ECP 248 promoter	TTGATCATAT-AACCCAAAATAAGTTTGATGTGGGTAACTCCATGGTTAGTTAAGTTTCAACAAAAAAAATCCACGGTTAGTTT-----ATATTTAA T----TTGAGT	ABRE
LEA 411 promoter	CGCATTAGGTGGAGAGAGTTTTTGATTCTAATCTTCTGACCC--GCACGAGCTGTAAGGCCAAGAAGCTACTCAAGGTCAGAGACGTAAACAAGACGCTTTATTA--TTTA-T	
ECP 66 promoter	350 360 370 380 390 400 410 422 430 440 450 460 ATCAAAGAGAA-AGATCACTAAATTCTACTTCTCTTC-TTCATAGACTTAAAGGTTAGATTCTTCACTCG-422---GTATTCT----TCARGGCAACCATGCTT---AAGCAT	ABRE
ECP 248 promoter	ATAAATGATAAGTTATCA-TATGCAAAACGTGACTGA-TTT-TAG----ATGAGTTGAGAT-CAATTACGCGTCATGGTGTTGTCGGAGAAATAGAAAAAGGAGTG---ACGTTT	DRE2
LEA 411 promoter	ATAAACGAT---TGATCA-AAAGTGAACTTGGTTGCTTTATTG---GATGATGCAGGAA-CATGAACAGGCGCTTGT-TTCT----GCCATTGCA---AGGCTGCGGATGCTT	
ECP 66 promoter	470 480 490 500 510 520 530 540 550 560 570 580 CTCAA---GAAA-----ATACATT--GTTTTATTAATGTGG--T TAGTTC-----ACTAGAAAAATAT--CAAA TGCAAAACCTA-----TATCATCATGATATGATGAAA	MVC
ECP 248 promoter	ATAAAATAGAGACGTTCAACATTAGAGCACTTTTAAAGTGA--TTGGTTTATGTGATTAATATAACACTAGTCTATGTGCGAA-CACAACTAA TGTA-----ATGTGATTAGC	
LEA 411 promoter	CTGATTTGAAAG-----CGGTA-AGAAACATTAAGGAGGGCTTTGTATATGACCTGAGAATAAGTTTATTGCCTATTCAATCACATTGTTTGTGACAGCACTCAATGCTT	DRE2

Figure 23: promoter alignment of ECP 66 promoter, ECP 248 promoter and LEA 411 promoter showing motifs associated with Drought/Cold stresses as outlined in table A13

Figure 24: ABA-dependent and independent pathway in response to stress showing the transcription factors and their *cis*-acting elements involved in cold, drought and salinity gene expression. modified from (Roychoudhury et al., 2013)

4 Discussion

Plants use various physiological and molecular mechanisms to cope when faced with stresses to minimize its damage and survive. These mechanisms include reactive oxygen species and scavengers and the production of putative osmoprotective proteins such as LEA proteins (Wu et al., 2014). The very first dehydration challenge the plant must undergo is during seed development to achieve maturation drying.

4.1 LEA gene expression during seed development

The seed is an important stage of a plant development; it is the connecting point between generations and without a functioning seed development, survival would be lost. The difference between the seeds becoming orthodox seeds or recalcitrant depends on their LEA content. Orthodox seeds undergo successful maturation drying and are shed from the parent plant at low moisture content allowing them to be stored for long periods without irreversible damage (Delahaie et al., 2013). Conversely, recalcitrant seeds (e.g. avocado, mango, cocoa to give some examples) do not have the ability to undergo maturation drying and are shed at a high moisture content, and thus lose viability during drying and storage as the moisture is lost. LEA expression in seeds starts in late stages of seed development, in general after the maturation stage in which storage proteins along with other protective proteins are deposited, (Guerriero et al., 2009). In my study, ECP 66, LEA 411 showed a steady exponential increase in expression

starting early in seed development and reaching its highest point in mature seeds. ECP 248, on the other hand, showed a different expression pattern, reaching its maximum around day 15 after flowering and decreasing to its initial expression levels as the seed matured. This expression pattern by ECP 248 suggests that this LEA-like protein is not associated with desiccation and plays a role in the maturation of the seeds.

4.2 LEA gene expression under environmental stresses

Although the expression of the three LEA genes differed only slightly during seed development, their response to stress is quite different. Drought and salinity are the main factors that affect the growth of crops and can drastically reduce yield.

4.2.1 Salt and drought stresses

Under salt stress, various key responsive genes are induced, the LEA genes that encode drought/desiccation proteins are first regulated during developmental stages, but also in response to ABA and dehydration signals (Wu et al., 2014). They are involved in the adaptive response to salt and dehydration (Tunnacliffe & Wise, 2007). It is thus expected, as shown by LEA 411, to have increased expression rates as the plant is experiencing more water loss. The high activity observed by ECP 66 at 3 hours, suggests that is it one of the early responses against salt stress; its activity decreased as other proteins were induced, but under prolonged stress (≥ 48 hours) its activity is increased again. With the exception of accumulation being higher at 6 hours under stress

compared to 3 hours, ECP 248 showed the same expression pattern as ECP 66. Exposing plants to high NaCl or mannitol levels reduces their ability to uptake water leading to slower growth (Wu et al., 2014). In plants overexpressing SmLEA under drought resistance their tolerance was enhanced and they showed a slower rate of water loss (Wu et al., 2014). Studies with LEA proteins on yeast desiccation tolerance also showed an increased survival rate after desiccation and rehydration (Dang et al., 2014).

LEA group 3 proteins; where ECP 66 and ECP 248 belong, are characterized by their motif of 11 amino acids as discussed above, which is predicted to form the amphipathic α -helix suggested to form structural interactions, these interactions are thought to be mainly associated with counteracting the irreversible damaging effect of increased ionic strength in the cytosol during stress (Gal, Glazer, & Koltai, 2004). This explains the delay in ECP66 and ECP248 accumulation under these stresses, suggesting they take part once plants are susceptible to ionic damage, this is also confirmed by their presence in the cytosol evident in the confocal analysis and the promoter analysis discussed below. To confirm that the enhanced tolerance of nematodes under dehydration is a result of LEA proteins, (Gal, Glazer, & Koltai, 2004) silenced the Ce-LEA proteins and the result are reduced survival under osmotic stress.

4.2.2 Cold stress

In overwintering plants, freezing tolerance is increased after a period of exposure to low, but non-freezing temperatures. This cold acclimation allows plants to develop tolerance for winter survival through alteration of gene expression (Sasaki et al., 2013). Interestingly, ECP 248 under cold stress showed a spike in expression at the 12 hours mark. These results require further investigation to determine the role of the circadian clock in response to LEA gene expression since the 12 hour mark corresponds to the only measurement during the night cycle.

The observation of different expression levels at different time intervals under stress was also documented by (Liu et al., 1998), where the expression of AtDREB1/CBF related genes were studied. AtDREB1 was expressed within 10 min in response to cold stress at 4 °C, while CBF was detected after 30 mins under stress with maximum levels at 1 hour (Lata & Prasad, 2011). Other studies that tested similar transcripts include CBF1/DREB1B and CBF3/DREB1A, which accumulated after 15 min of cold treatment while CBF2/DREB1C transcripts accumulated at a slower rate with maximum expression after 2.5 h of cold exposure and then gradually declined (Novillo et al., 2004).

There are several examples in the literature of animal and plant LEA proteins that are unstructured in the hydrated state that partially fold into α -helical structures during dehydration in response to freezing or drying. CD spectroscopy confirmed that proteins from LEA 4, LEA 5, and LEA 6 groups showed a reduced

degree of un-structure and suggested a secondary structure for some of these proteins upon dehydration, suggesting they form amphipathic class A α -helices; a motif that interacts with membranes (Hinch & Thalhammer, 2012). However, cold induced proteins COR47, LT129 and LT130 simulated crowding and did not induce folding (Hinch & Thalhammer, 2012). Thus induced folding of LEA proteins during target binding is a suggested possibility of how LEA proteins function.

4.3 Spatial localization of LEA genes

Experimental data as well as computer-based prediction about the sub-cellular localization of LEA proteins showed that they are not transmembrane proteins. In fact, different groups have found LEAs in a number of subcellular compartments. Group 2 LEAs were found to localize in the cytoplasm, nucleus, chloroplast, peroxisome, vacuoles and mitochondria, while group 3 LEAs were localized to the cytoplasm, chloroplast, endoplasmic reticulum, protein storage vacuoles and the mitochondria (Battaglia & Covarrubias, 2013) and distributed in the cytosol of all cell types in the embryo as well as in the cytoplasm and protein storage vacuoles or in mitochondrial matrix as in the case in pea seeds (Grelet et al., 2005). (Gu et al., 2012). Protein localization has been predicted by computer programs based on their amino acid sequence, and in the case of CarLEA4::GFP the localization was found to be in the nucleus by confocal microscopy in onion epidermal cells. (Grelet et al., 2005) identified a LEA protein that was named PsLEAm in pea seeds, sequence analysis of PsLEAm revealed that it belongs to

group 3 LEA proteins with high hydrophilicity, interestingly its N-terminal analysis revealed a succession of hydrophilic and hydrophobic stretches that resembled a transit peptide. Bioinformatics analysis of that sequence with subcellular programs predicted high probability of mitochondrial localization while others ranked first in plastidial location followed by mitochondrial. To confirm their findings, they constructed a plasmid driven by a 35S promoter and tagged with green fluorescent protein (GFP) in pea leaf. Their confocal microscopy analysis did indeed localize the first group 3 LEA protein to the mitochondria, they further localized the protein to the matrix space thus providing protection within the matrix space (Grelet et al., 2005).

The subcellular location of the proteins also assists in determining their function and their mode of action. The maintenance of the integrity of the mitochondria in the dry state can be explained as a reaction to the potential release of proapoptotic factors in the intermembrane space or to the disruption in oxidative phosphorylation (Hand et al., 2011). DHN1, a LEA protein suggested to concentrate near endomembrane cells, changes conformation when bound to lipid vesicles, suggesting its involvement in membrane stabilization under stress conditions (Koag et al., 2003)

This may explain why ECP66 pro and ECP248 pro exhibited different responses to abiotic stress. This is not atypical of LEA genes since very closely related genes behave differently in response to different stresses.

4.4 LEA promoter analysis

Stress response in plants is achieved through the activation of a number of different biochemical pathways. The response to abiotic stress, for example, occurs via ABA-dependent and independent pathways. ABA is produced under osmotic stress caused by drought or salinity and it plays an important role in regulating many genes under osmotic stress. It is important in regulating gene expression during seed development and germination (Nakashima & Yamaguchi-Shinozaki, 2013). The ABRE motif is the major *cis*-element for ABA-dependent gene expression regulated by the ABRE-binding protein (AREB) and the ABRE-binding factor (ABF) as well as basic leucine zipper (bZIP) transcription factors. ABRE motifs regulate the transcription of target genes such as RD29B (Nakashima et al., 2006) and RD20A leading to stress tolerance. A second group of transcription factors that are involved in regulating stress signal transduction and stress response in an ABA-dependent pathway include myelocytomatosis oncogene (MYC), myeloblastosis oncogene (MYB) and DREB (dehydration responsive element binding).

In addition to DREB working in an ABA-dependent manner it has also been shown to act in the ABA-independent response during drought and cold stress in ABA-deficient and ABA-insensitive mutants of *Arabidopsis* (Lata & Prasad, 2011). Despite DREB being activated in both the ABA-dependent and independent pathway, it works in correlation with ABRE as reported by (Roychoudhury et al., 2013) in the case of the RD29A promoter containing three

functional DREs and one ABRE involved in its gene expression, thus suggesting cross talk between DRE and ABRE. The cross talk theory is strengthened by work on the *fiery1* (*fry1*) mutation that is suggested to be a second messenger in ABA signaling (Roychoudhury et al., 2013). The mutation increases the degree and sensitivity of stress induction by ABA, but also shows increased sensitivity to salt, drought and cold stresses (Roychoudhury et al., 2013). That is not the only case of crosstalk, as (Guerriero et al., 2009) showed that the *napA* gene from the *Brassica napus* napin promoter required the two elements ABRE and RY for seed-specific expression and ABA responsive activation. Despite the importance of ABA in plant stress response, plants treated with methyl jasmonate, gibberellic acid and auxin also showed induced expression of CarLEA4 protein, a LEA gene from chickpea (Gu et al., 2012). Further analysis with LEA411, ECP 66, ECP 248 in response to other phytohormones could aid in determining their expression patterns.

Analysis of the alignment of the three promoters; figure 23, showed that their regions contains some putative *cis*-elements related to abiotic stress, such as; ABRE, DRE/CRT, MYB, MYC. These elements are more present in the 3'-end of the promoter with a conserved location, These results suggest that modification of the ABA signaling factors, including ABRE/ABF and MYCRs/MYBRs along with the overexpression of LEA proteins may contribute to enhanced drought tolerance leading to better crop production under drought and other abiotic stresses. Despite these findings, the gap between our

understanding of the molecular mechanisms of stress response and the improvement of stress resistance in the field still remains.

5 Future Directions

This study indicated that the three *B. napus* LEA genes analyzed here have different gene expression patterns during development and in response to abiotic stressors. To ensure that all response trends were measured, analysis of shorter time intervals starting at 10 min under the stresses should be collected, as well as more collection points at night to determine if time-of-day plays a role into LEA expression (influence of the circadian clock).

The various groups of LEA proteins differ in structure but serve very similar functions – the protection of cellular components against stress-induced damage. A critical element of the function of LEA proteins is the cellular guiding system that determines when and where LEA proteins are deposited. A thorough bioinformatics analyses of regulatory as well as structural elements may provide clues to subcellular location and functions. To determine the exact tissue distributions and subcellular locations of the LEA proteins, flash freezing and cryo-sectioning of the LEA seeds at different developmental stages in combination with *in situ* immunolocalization experiments should be performed. This can be done with antibodies raised against representatives of each group of LEA proteins. For this, LEA proteins will be expressed in *E.coli* using their respective cDNAs, purified, and then used for animal injections to obtain polyclonal antibodies.

Finally, there is also the possibility of obtaining *Arabidopsis* lea knock-out mutants which, when tested in various stress situations, could give clues to functions associated with different types of LEA proteins.

6 Conclusion

Understanding the spatial and temporal expression of LEA proteins as well as the molecular mechanisms of plant responses to abiotic stresses, such as drought, salinity, and cold is very important in helping to manipulate plants to improve stress tolerance and crop productivity. LEA 411, ECP 66 and ECP 248 were shown to accumulate during seed developments from 5-10 DAF and induced by drought, salinity, and cold. Expression of ECP66 and ECP248 was observed in the whole cotyledons during seed development while LEA 411 only showed expression in the radicle. A closer look revealed that ECP 66 localized primarily to the cytosol but not in vacuoles. ECP248 and LEA411 showed the opposite localizing to vacuoles and not cytoplasm. All three proteins showed high accumulation levels in the mitochondria. In response to these stresses, many genes are regulated mainly by TFs and their gene products function in providing stress tolerance to plants. The control of the expression of LEA proteins and other genes in response to abiotic stress may improve tolerance to environmental stresses.

7 References:

- Battaglia, M., & Covarrubias, A. A. (2013). Late embryogenesis abundant (LEA) proteins in legumes. *Frontiers in Plant Science*, 4, 190.
- Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., & Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiology*, 148(1), 6-24.
- Baud, S., Boutin, J. -P., Miquel, M., Lepiniec, L., & Rochat, C. (2002). An integrated overview of seed development in arabidopsis thaliana ecotype WS. *Plant Physiology and Biochemistry*, 40(2), 151-160.
- Billinton, N., & Knight, A. W. (2001). Seeing the wood through the trees: A review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical Biochemistry*, 291(2), 175-97.
- Bobb, A. J., Chern, M. S., & Bustos, M. M. (1997). Conserved ry-repeats mediate transactivation of seed-specific promoters by the developmental regulator pvalf. *Nucleic Acids Research*, 25(3), 641-7.
- Borovskii, G. B., Stupnikova, I. V., Antipina, A. I., Vladimirova, S. V., & Voinikov, V. K. (2002). Accumulation of dehydrin-like proteins in the mitochondria of cereals in response to cold, freezing, drought and ABA treatment. *BMC Plant Biology*, 2, 5.
- Bustros, M. M., Guiltinan, M. J., Jordano, D., Begum, F. A., & Hall, T. C. (1989). Regulation of β -glucuronidase expression in transgenic tobacco plants by an A/t-rich, cis-acting sequence found upstream of a french bean β -phaseolin gene. *The Plant Cell*, 1, 839-853.
- Carles, C., Bies-Etheve, N., Aspart, L., Léon-Kloosterziel, K. M., Koornneef, M., Echeverria, M., & Delseny, M. (2002). Regulation of arabidopsis thaliana em genes: Role of ABI5. *The Plant Journal*, 30(3), 373-383.
- Chakrabortee, S., Boschetti, C., Walton, L. J., Sarkar, S., Rubinsztein, D. C., & Tunnacliffe, A. (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), 18073-8.
- Close, T. J. (1997). Dehydrins: A commonalty in the response of plants to dehydration and low temperature. *Physiologia Plantarum*, 100(2), 291-296.

Crowe, J. H., Hoekstra, F. A., & Crowe, L. M. (1992). Anhydrobiosis. *Annual Review of Physiology*, 54(1), 579-599.

Dang, N. X., Popova, A. V., Hundertmark, M., & Hinch, D. K. (2014). Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and in vitro. *Planta*.

Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H., & Buitink, J. (2013). LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals abi3-regulated LEA protein abundance linked to desiccation tolerance. *Journal of Experimental Botany*, 64(14), 4559-73.

Delseny, M., Bies-Etheve, N., Carles, C., Hull, G., Vicent, C., Raynal, M., & Aspart, L. (2001). Late embryogenesis abundant (LEA) protein gene regulation during *Arabidopsis* seed maturation. *Journal of Plant Physiology*, 158(4), 419-427.

Dure, L. (1993). A repeating 11-mer amino acid motif and plant desiccation. *The Plant Journal : For Cell and Molecular Biology*, 3(3), 363-9.

Eom, J., Baker, W. R., Kintanar, A., & Wurtele, E. S. (1996). The embryo-specific EMB-1 protein of *Daucus carota* is flexible and unstructured in solution. *Plant Science*, 115(1), 17-24.

Gal, T. Z., Glazer, I., & Koltai, H. (2004). An LEA group 3 family member is involved in survival of *C. Elegans* during exposure to stress. *FEBS Letters*, 577(1-2), 21-6.

Gilles, G. J., Hines, K. M., Manfre, A. J., & Marcotte, W. R. (2007). A predicted n-terminal helical domain of a group 1 LEA protein is required for protection of enzyme activity from drying. *Plant Physiology and Biochemistry : PPB / Société Française De Physiologie Végétale*, 45(6-7), 389-99.

Grelet, J., Benamar, A., Teyssier, E., Avelange-Macherel, M. H., Grunwald, D., & Macherel, D. (2005). Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiology*, 137(1), 157-67.

Gu, H., Jia, Y., Wang, X., Chen, Q., Shi, S., Ma, L., & Ma, H. (2012). Identification and characterization of a LEA family gene *carlea4* from chickpea (*Cicer arietinum* L.). *Molecular Biology Reports*, 39(4), 3565-72.

Guerriero, G., Martin, N., Golovko, A., Sundström, J. F., Rask, L., & Ezcurra, I. (2009). The RY/sph element mediates transcriptional repression of maturation genes from late maturation to early seedling growth. *The New Phytologist*, 184(3), 552-65.

- Hand, S. C., Menze, M. A., Toner, M., Boswell, L., & Moore, D. (2011). LEA proteins during water stress: Not just for plants anymore. *Annual Review of Physiology*, 73, 115-34.
- Hanin, M., Brini, F., Ebel, C., Toda, Y., & Takeda, S. (2011). Plant dehydrins and stress tolerance: Versatile proteins for complex mechanisms. *Plant Signaling & Behavior*, 6(10), 1503-9.
- He, S., Tan, L., Hu, Z., Chen, G., Wang, G., & Hu, T. (2012). Molecular characterization and functional analysis by heterologous expression in *E. Coli* under diverse abiotic stresses for oslea5, the atypical hydrophobic LEA protein from *oryza sativa* L. *Molecular Genetics and Genomics : MGG*, 287(1), 39-54.
- Higo, K., Ugawa, M., Iwamoto, M. & Korenaga T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database:1999. *Nucleic Acids Research* Vol.27 No.1 pp. 297-300
- Hincha, D. K., & Thalhammer, A. (2012). LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochemical Society Transactions*, 40(5), 1000-3. d
- Hollung, K., Espelund, M., Schou, K., & Jakobsen, K. S. (1997). Developmental, stress and ABA modulation of mrna levels for bzip transcription factors and vp1 in barley embryos and embryo-derived suspension cultures. *Plant Molecular Biology*, 35(5), 561-71.
- Huang, G. -T., Ma, S. -L., Bai, L. -P., Zhang, L., Ma, H., Jia, P., & Guo, Z. -F. (2012). Signal transduction during cold, salt, and drought stresses in plants. *Molecular Biology Reports*, 39(2), 969-87.
- Hundertmark, M., & Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *arabidopsis thaliana*. *BMC Genomics*, 9, 118.
- Ismail, A. M., Hall, A. E., & Close, T. J. (1999). Purification and partial characterization of a dehydrin involved in chilling tolerance during seedling emergence of cowpea. *Plant Physiology*, 120(1), 237-44.
- Koag, M. -C., Fenton, R. D., Wilkens, S., & Close, T. J. (2003). The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiology*, 131(1), 309-16.
- Lata, C., & Prasad, M. (2011). Role of drebs in regulation of abiotic stress responses in plants. *Journal of Experimental Botany*, 62(14), 4731-48.
- Manfre, A. J., Lanni, L. M., & Marcotte, W. R. (2006). The arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiology*, 140(1), 140-9.

Nakashima, K., & Yamaguchi-Shinozaki, K. (2013). ABA signaling in stress-response and seed development. *Plant Cell Reports*, 32(7), 959-970.

Nakashima, K., Fujita, Y., Katsura, K., Maruyama, K., Narusaka, Y., Seki, M., & Yamaguchi-Shinozaki, K. (2006). Transcriptional regulation of ABI3- and aba-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of arabidopsis. *Plant Molecular Biology*, 60(1), 51-68.

Oliver, M. J., Dowd, S. E., Zaragoza, J., Mauget, S. A., & Payton, P. R. (2004). The rehydration transcriptome of the desiccation-tolerant bryophyte *tortula ruralis*: Transcript classification and analysis. *BMC Genomics*, 5, 89.

Olvera-Carrillo, Y., Luis Reyes, J., Covarrubias, A. A. (2011). Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments. *Plant Signaling & Behavior*, 6(4), 586-9.

Rorat, T. (2006). Plant dehydrins--tissue location, structure and function. *Cellular & Molecular Biology Letters*, 11(4), 536-56.

Roychoudhury, A., Paul, S., & Basu, S. (2013). Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. *Plant Cell Reports*, 32(7), 985-1006.

Sasaki, K., Christov, N. K., Tsuda, S., & Imai, R. (2013). Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant & Cell Physiology*.

Shih, M. D., Hsieh, T. Y., Lin, T. P., Hsing, Y. I., & Hoekstra, F. A. (2010). Characterization of two soybean (*glycine max* L.) LEA IV proteins by circular dichroism and fourier transform infrared spectrometry. *Plant & Cell Physiology*, 51(3), 395-407.

Soulages JL, Kim KM, Walters C, Cushman JC (2002) Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physiol* 128:822–832.

Szalaiiné Ágoston, B., Kovács, D., Tompa, P., & Perczel, A. (2011). Full backbone assignment and dynamics of the intrinsically disordered dehydrin ERD14. *Biomolecular NMR Assignments*, 5(2), 189-93.

Tolletier, D., Jaquinod, M., Mangavel, C., Passirani, C., Saulnier, P., Manon, S., . . . Macherel, D. (2007). Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *The Plant Cell*, 19(5), 1580-9.

Tompa, P., & Kovacs, D. (2010). Intrinsically disordered chaperones in plants and animals. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 88(2), 167-74.

Tunnacliffe, A., & Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Die Naturwissenschaften*, 94(10), 791-812.

Wolkers, W. F., McCready, S., Brandt, W. F., Lindsey, G. G., & Hoekstra, F. A. (2001). Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. *Biochimica Et Biophysica Acta*, 1544(1-2), 196-206.

Wu, G., Zhang, H., Sun, J., Liu, F., Ge, X., Chen, W. H., & Wang, W. (2011). Diverse LEA (late embryogenesis abundant) and lea-like genes and their responses to hypersaline stress in post-diapause embryonic development of *artemia franciscana*. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, 160(1), 32-9.

Wu, Y., Liu, C., Kuang, J., Ge, Q., Zhang, Y., & Wang, Z. (2014). Overexpression of smlea enhances salt and drought tolerance in *escherichia coli* and *salvia miltiorrhiza*. *Protoplasma*. doi:10.1007

Yoshida, T., Fujita, Y., Sayama, H., Kidokoro, S., Maruyama, K., Mizoi, J., & Yamaguchi-Shinozaki, K. (2010). AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate abscisic acid-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *The Plant Journal : For Cell and Molecular Biology*, 61(4), 672-85.

8 Appendix

Table A1: qPCR data comparing the p value and RQ of tested ECP66 pro-ORF-eYFP during seed development

DAF	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
5	30.12	1.18	0.04	1.00	32.06	1.15	0.04	2.23	30.07	2.06	0.07	6.00
10					31.56	4.24	0.13	15.29				
12	26.52	0.72	0.03	399.41	27.79	1.03	0.04	83.70	36.31	0.88	0.02	3.84
15	33.07	2.97	0.09	38.89	26.13	1.34	0.05	980.47	32.72	3.02	0.09	42.96
Mature	20.74	0.08	0.00	4894.26	27.11	0.59	0.02	1319.11	27.93	0.49	0.02	1702.10

Table A2: qPCR data comparing the p value and RQ of tested ECP248 pro-ORF-eYFP during seed development

Time under stress (hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
5	27.10	0.29	0.01	1.06	26.99	0.28	0.01	1.57	29.46	0.17	0.01	3.03
10	23.97	0.29	0.01	2.17	27.25	0.42	0.02	1.62	27.36	0.30	0.01	1.45
12	24.98	0.17	0.01	3.50	24.11	0.56	0.02	2.26				
15	25.01	0.18	0.01	19.22	23.29	0.29	0.01	1.54	29.68	0.63	0.02	1.58
Mature	32.47	1.02	0.03	1.52	35.39	1.53	0.04	1.37	30.47	0.68	0.02	2.34

Table A3: qPCR data comparing the p value and RQ of tested LEA411 pro-ORF-eYFP during seed development

Time under stress (hours)	Biological Replicates								p ≤ 0.05		0.05 ≥ p ≤0.1		p ≥ 0.1	
	A				B				C					
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ		
5	27.99	1.60	0.06	1.00	27.42	0.11	0.00	3.05	27.43	0.48	0.02	2.03		
10	33.08	0.39	0.01	10.43	33.85	1.10	0.03	6.17	34.26	1.00	0.03	0.42		
12	32.40	0.43	0.01	2.90	33.72	0.13	0.00	5.71	36.16	0.15	0.00	8.79		
15	35.53	0.29	0.01	8.07	30.14	0.49	0.02	20.22	33.78	0.38	0.01	76.97		
Mature	34.06	0.41	0.01	43.54	34.53	0.15	0.00	19.17	36.50	0.48	0.01	15.51		

Table A4: qPCR data comparing the p value and RQ of tested ECP66 samples under salt stress

Time under stress (hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	27.51	1.33	0.05	28399.8 2	20.17	0.18	0.01	350.87				
6	23.29	0.24	0.01	95.05	25.18	0.29	0.01	17.44				
12	20.38	0.14	0.01	350.55	23.11	0.42	0.02	11.76	24.52	0.22	0.01	9.27
24	26.70	0.24	0.01	0.29	23.94	0.37	0.02	2.08	25.92	0.45	0.02	2.30
48	24.80	1.06	0.04	828.33	24.35	0.83	0.03	2.48	23.57	0.13	0.01	595.56

Table A5: qPCR data comparing the p value and RQ of tested ECP66 samples under drought stress

Time under stress (hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	29.30	0.14	0.00	1.02	28.57	0.15	0.01	1.10				
6	26.82	0.26	0.01	1.10	27.31	0.06	0.00	1.32	27.86	0.12	0.00	1.08
12	32.16	0.60	0.02	2.00	29.29	0.58	0.02	1.03	28.69	0.26	0.01	1.13
24	29.57	0.05	0.00	1.01	24.72	0.13	0.01	1.67	27.41	0.29	0.01	1.07
48	24.41	0.25	0.01	2.18	25.44	0.31	0.01	1.07	27.35	0.21	0.01	1.01

Table A6: qPCR data comparing the p value and RQ of tested ECP248 samples under salt stress

Time under stress(hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1	p ≥ 0.1	
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	28.18	0.13	0.00	13.23	28.90	0.13	0.00	5.26	6.65	0.86	0.13	2.52
6	29.10	0.84	0.03	3.38	30.06	0.18	0.01	8.86	29.86	0.33	0.01	2.25
12	29.32	0.23	0.01	2.43	30.12	0.07	0.00	0.87	30.55	0.26	0.01	0.40
24	30.33	0.22	0.01	0.26	28.34	0.22	0.01	1.63	27.42	0.45	0.02	715.10
48	26.29	0.07	0.00	22.47	26.24	0.20	0.01	25905.84	23.48	0.21	0.01	92.35

Table A7: qPCR data comparing the p value and RQ of tested ECP248 samples under drought stress

Time under stress (hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	31.52	3.35	0.11	4976.61	29.09	0.04	0.00	6.29	28.89	0.07	0.00	18.84
6	29.79	0.04	0.00	8.19	25.30	0.05	0.00	922.11	30.80	0.43	0.01	14.61
12	31.25		0.00	9.93	30.56	0.29	0.01	10.38	31.57	0.18	0.01	2.19
24	28.37	0.07	0.00	22.42	25.07	0.49	0.02	327.37	26.93	0.54	0.02	63.42
48	28.75	0.58	0.02	3.97	29.76	0.45	0.02	28.07	28.58	0.02	0.00	56383.21

Table A8: qPCR data comparing the p value and RQ of tested LEA411 samples under salt stress

Time under stress (hours)	Biological Replicates								$p \leq 0.05$	$0.05 \geq p \geq 0.1$	$p \geq 0.1$	
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	31.61	0.06	0.00	1.29	33.29	0.52	0.02	1.00	31.83	0.25	0.01	1.33
6	30.39	0.21	0.01	1.60	31.13	0.66	0.02	2.72	33.20	0.14	0.00	1.36
12	31.65	1.13	0.04	0.32	31.93	0.41	0.01	0.17	29.44	0.21	0.01	1.65
24	32.52	0.10	0.00	2.62	34.93	0.05	0.00	4.42	31.62	0.23	0.01	0.25
48	33.99	0.65	0.02	1.03	35.42	1.65	0.05	327.79	31.58	0.45	0.01	16.28

Table A9: qPCR data comparing the p value and RQ of tested LEA411 samples under drought stress

Time under stress (hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	33.01	0.13	0.00	1.28	34.28	0.57	0.02	1.00	32.02	0.55	0.02	1.64
6	32.72	0.15	0.00	1.37	29.93	0.07	0.00	2.12	30.08	0.17	0.01	4.73
12	30.51	0.15	0.01	2.70	29.23	0.13	0.00	3.49	29.69	0.10	0.00	3.93
24	29.59	0.16	0.01	2.74	32.86	0.20	0.01	1.26	31.74	3.64	0.11	1.70
48	32.94	0.12	0.00	321.46	31.17	0.47	0.02	1.32	31.77	0.09	0.00	1.11

Table A10: qPCR data comparing the p value and RQ of tested ECP248 samples under cold stress

Time under stress (hour s)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	29.15	0.10	0.00	2822.3 3	31.03	0.05	0.00	1.00	26.96	0.10	0.00	9.56
6	27.65	0.05	0.00	10.41	28.11	0.16	0.01	6.29	27.16	0.23	0.01	307248. 78
12	29.64	0.22	0.01	29.50	29.86	0.53	0.02	458.83	27.85	0.06	0.00	2.80
24	28.55	0.70	0.02	1187.1 6	27.50	0.22	0.01	1671.8 2				
48	26.25	0.02	0.00	2.53								

Table A11: qPCR data comparing the p value and RQ of tested ECP248 samples under cold stress

Time under stress (hours)	Biological Replicates								$p \leq 0.05$	$0.05 \geq p \geq 0.1$	$p \geq 0.1$	
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	31.60	0.41	0.01	1.00	32.20	0.74	0.02	1.41	31.40	0.38	0.01	1.57
6	30.15	0.06	0.00	0.48	28.86	0.21	0.01	1.92	31.23	0.14	0.00	4.31
12	29.91	0.03	0.00	33.48	25.65	0.20	0.01	3285.81	29.31	0.51	0.02	1433.21
24	30.83	0.54	0.02	0.40	29.39	0.15	0.00	4.43	28.52	0.65	0.02	4.01
48	30.40	0.62	0.02	2.12	31.81	0.07	0.00	918.03				

Table A12: qPCR data comparing the p value and RQ of tested LEA 411 samples under cold stress

Time under stress (hours)	Biological Replicates								p ≤ 0.05	0.05 ≥ p ≤0.1		p ≥ 0.1
	A				B				C			
	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ	CT Mean	CT SD	p value	RQ
3	33.92	0.15	0.00	21.71	33.97	0.06	0.00	1.00	33.75	0.30	0.01	1.61
6	33.76	0.36	0.01	266.02	28.54	0.16	0.01	38.74	32.92	0.11	0.00	1.40
12	33.30	0.16	0.00	1.21	34.49	0.43	0.01	93.41	32.72	0.79	0.02	1.85
24	31.12	0.31	0.01	4.89	31.85	1.03	0.03	35.15	34.56	0.14	0.00	0.67
48	24.99	0.60	0.02	70669.29	32.19	1.00	0.03	0.78	32.53	0.15	0.00	9.40

Table A13: Motifs Identified using PLACE in ECP 66, ECP 248, and LEA 411 promoters (highlighted cells correspond in colour to cis-acting elements involved in stress responsive transcription as outlined in the figure below

	Promoter Site Name	ECP 66	ECP 248	LEA 411	Description
1	NEGATIVE 10PEHVPSBD	X	X	X	negative10 promoter element
2	NEGATIVE 300ELEMENT	X		X	Present upstream of the promoter from the B-hordein gene of barley and the alpha-gliadin
3	2SSEEDPROTBANAPA	X			
4	ABREATCONSENSUS			X	ABA-responsive elements
5	ABRELATERD1	X	X	X	ABRE-like sequence (from -199 to -195) required for etiolation-induced expression of erd1
6	ABREMOTIFAOSOSEM	X	X		motif A" ABRE-like sequence found in rice (O.s.) Osem gene promoter; Important for regulation by ABA
7	ABREOSRAB21			X	ABA responsive element
8	ABRERATCAL	X	X	X	ABRE-related sequence" or "Repeated sequence motifs" identified in the upstream regions of 162 Ca ²⁺ - responsive upregulated genes
9	ABREZMRAB28	X	X	X	Cat1 gene promoter responsible for induction of ABA
10	ACGTABOX	X		X	"A-box"
11	ACGTABRENOTIF2OSEM	X	X	X	seq. requirement of ACGT-core of motif A in ABRE rice

12	ACGTABREMOTIFOSOSEM	X	X		ABRE motif A" found in the promoter of the rice ACGT-containing ABRE; Required for ABA-responsiveness and VP1 activation; Binding site of TRAB1; Motif A & CE3
13	ACGTATERD1	X	X	X	ACGT sequence (from -155 to -152) required for etiolation-induced expression of erd1 in Arabidopsis;
14	ACGTTBOX	X		X	T-box
15	AMMORESIVDCRNIA1		X		Motif (IVD) found in the Chlamydomonas Nia1 gene promoter
16	AMYBOX2			X	amylase box; "amylase element"
17	ANAERO1CONSENSUS			X	One of 16 motifs found in silico in promoters of 13 anaerobic genes involved in the fermentative pathway
18	ANAERO2CONSENSUS			X	
19	ANAERO3CONSENSUS	X		X	
20	ARFAT			X	ARF(auxin response factor) binding site found in the promoters of primary/early auxin response genes of At
21	ARR1AT	X	X	X	"ARR1-binding element" found in Arabidopsis; ARR1 is a response regulator
22	ASF1MOTIFCAMV		X	X	ASF-1 binding site" in CaMV 35S promoter; involved in transcriptional activation of several genes by auxin and/or SA
23	BIHD1OS	X	X	X	Binding site of OsBIHD1
24	BOXIIPCCHS	X		X	Core of "Box II/G box"
25	CAATBOX1	X	X	X	CAAT promoter consensus sequence
26	CACGTGMOTIF	X	X	X	G-box

27	CACTFTPPCA1	X	X	X	Tetranucleotide (CACT) is a key component of Mem1
28	CANBNNAPA	X	X	X	Core of "n element" in storage protein genes in <i>Brasica napus</i> embryo- and endosperm-specific transcription of napin (storage protein) gene, napA; seed specificity; activator and repressor
29	CAREOSREP1		X		CAREs (CAACTC regulatory elements) found in the promoter region of a cystein proteinase rice (REP-1) gene
30	CARGCW8GAT	X	X	X	A variant of CArG motif
31	CATATGGMSAUR		X		Sequence found in NDE element in soybean (G.m.) SAUR(Small Auxin-Up RNA)
32	CBFHV	X		X	binding site for dehydration responsive elements binding proteins
33	CCA1ATLHCB1			X	CCA1 binding site (myb-related transcription factor)
34	CCAATBOX1	X	X	X	"CCAAT box" found in the promoter of heat shock protein genes; Located immediately upstream from the most distal HSE of the promoter
35	CGACGOSAMY3			X	found in the GC-rich regions of the rice may function as a coupling element for the G box element
36	CGCGBOXAT		X	X	Ca ⁺⁺ /calmodulin binds to all AtSRs
37	CIACADIANLELHC		X		Region necessary for circadian expression of tomato Lhc gene
38	CPBCSPOR	X		X	Cytokinin-enhanced Protein Binding
39	CRTDREHVCBF2			X	AP2 transcriptional activator DNA binding is regulated by temperature
40	CURECORECR	X	X	X	core of a CuRE (copper-response element)

41	DOFCOREZM	X	X	X	Core site required for binding of Dof proteins in maize which are DNA binding proteins, with only one zinc finger, and are unique to plants Four cDNAs encoding Dof proteins, Dof1, Dof2, Dof3 and PBF, PBF is an endosperm specific Dof protein that binds to prolamin box;
42	DPBFCOREDCDC3	X	X	X	novel class of bZIP transcription factors Dc3 expression is normally embryo-specific, and also can be induced by ABA;
43	DRE2COREZMRAB17	X		X	expressed during late embryogenesis, and is induced by ABA
44	DRECRTCOREAT	X		X	Core motif of DRE/CRT (dehydration-responsive element/C-repeat) cis-acting element
45	E2FCONSENSUS		X	X	E2F consensus sequence
46	EBOXBNNAPA	X	X	X	E-box of napA storage-protein gene of Brassica napus
47	EECCRCAH1	X	X	X	motif of the two enhancer elements, EE-1 and EE-2
48	ELRECOREPCR1		X		EIRE (Elicitor Responsive Element) core of parsley; Box W1 & W2 are the binding site of WRKY1 & WRKY2
49	EMBP1TAEM			X	Binding site of ABFs(ABRE binding factors)
50	ERELEE4			X	ERE (ethylene responsive element)" of tomato
51	EVENINGAT		X		"Evening element" found 46 times in promoters of 31 cycling genes in A.t. Required for circadian control of gene expression
52	GADOWNAT	X	X	X	Sequence present in 24 genes in the GA-down regulated d1 cluster (106 genes) found in Arabidopsis seed germination; This motif is similar to ABRE
53	GAREAT		X	X	GARE(GA-responsive element)

54	GATABOX	X	X	X	"GATA box"; GATA motif in CaMV 35S promoter
55	GBOXLERBC2			X	"G box"; Conserved sequence upstream of light-regulated genes
56	GBOXPC			X	G box cytosolic G-box binding factors
57	GBOXSORBCS1	X			G-box G-box trimer confers relatively high level expression in roots
58	GCN4OSGLUB1	X		X	"GCN4 motif" found in GluB-1 gene in rice(O.s.); Required for endosperm-specific expression AACA and ACGT motifs
59	GT1CONSENSUS	X	X	X	Consensus GT-1 binding site in many light-regulated genes
60	GT1CORE	X	X	X	Critical for GT-1 binding to box II of rbcS
61	GT1GMSCAM4	X	X	X	"GT-1 motif" found in the promoter of soybean (Glycine max) CaM isoform, SCaM-4; Plays a role in pathogen- and salt-induced SCaM-4 gene expression
62	GTGANTG10	X	X	X	found in the promoter of the tobacco late pollen gene
63	HEXAMERATH4			X	hexamer motif of Arabidopsis thaliana (A.t.) histone H4 promoter
64	HEXMOTIFTAH3H4		X	X	"hexamer motif" found in promoter of wheat
65	IBOX		X	X	"I box" Conserved sequence upstream of light-regulated genes
66	IBOXCORE	X	X	X	"I box" Conserved sequence upstream of light-regulated genes of monocots and dicots
67	IBOXCORENT			X	I-box core motif" in the CAMs (conserved DNA modular arrays) associated with light responsive promoter regions
68	INRNTPSADB	X		X	"Inr (initiator)" elements found in the tobacco psaDb gene promoter without TATA boxes; Light-responsive transcription of psaDb depends

					on Inr, but not TATA box
69	INTRONLOWER			X	"3' intron-exon splice junctions"; Plant intron lower sequence , Consensus sequence
70	IRO2OS	X	X	X	OsIRO2-binding core sequence; "G-box plus G"; Transcription factor OsIRO2 is induced exclusively by Fe deficiency
71	L1BOXATPDF1			X	"L1 box" found in promoter of At PROTODERMAL FACTOR1 (PDF1) gene; Involved in L1 layer-specific expression; L1-specific homeodomain
72	LECPLEACS2		X		Core element in LeCp (tomato Cys protease) binding cis-element
73	LRENPCABE	X			LRE"; A positive light regulatory element in tobacco
74	LTRE1HVBLT49			X	"LTRE-1"(low-temperature-responsive element) in barley new LTRE; A previously known LTRE is CCGAC
75	LTRECOREATCOR15	X	X	X	Core of low temperature responsive element
76	MARARS	X			ARS element
77	MARTBOX		X		"T-Box"; Motif found in SAR (scaffold attachment region;or matrix attachment region, MAR)
78	MYB1AT	X	X	X	MYB recognition site found in the promoters of the dehydration-responsive gene rd22
79	MYB1LEPR	X			Tomato Pti4(ERF) regulates defence-related gene expression via GCC box and non-GCC box cis elements (Myb1(GTTAGTT), G box (CACGTG))
80	MYB2AT	X	X	X	Binding site for ATMYB2, an Arabidopsis MYB homolog
81	MYB2CONSENSUSAT	X	X	X	MYB recognition site found in the promoters of the dehydration-responsive gene rd22 & other genes in A.t.

82	MYBATRD22	X		X	Binding site for MYB (ATMYB2) in dehydration-responsive gene
83	MYBCORE	X	X	X	Binding site for all animal MYB and at least two plant MYB proteins ATMYB1 and ATMYB2, both isolated from Arabidopsis
84	MYBCOREATCYCB1	X		X	"Myb core" in the 18 bp sequence which is able to activate reporter gene without leading to M-phase-specific expression
85	MYBPZM			X	Core of consensus maize P (myb homolog) binding site
86	MYBST1	X	X	X	Core motif of MybSt1(a potato MYB homolog) binding site
87	MYCCONSENSUSAT	X	X	X	MYC recognition site found in the promoters of the dehydration-responsive gene rd22 & other genes in At. Binding site of ICE1 (inducer of CBF expression 1)
88	NAPINMOTIFBN		X		Sequence found in 5' upstream region (-6, -95, -188) of napin (2S albumin) gene in Brassica napus (B.n.); Interact with a protein present in crude nuclear extracts from developing B. napus seeds;
89	NODCON1GM	X	X	X	One of two putative nodulin consensus sequences
90	NODCON2GM	X	X	X	One of two putative nodulin consensus sequences
91	NTBBF1ARROLB		X		NtBBF1(Dof protein from tobacco) binding site in Agrobacterium Required for tissue-specific expression and auxin induction
92	OSE1ROOTNODULE	X	X	X	One of the consensus sequence motifs of organ-specific elements (OSE) characteristic of the promoters activated in infected cells of root nodules
93	OSE2ROOTNODULE	X	X	X	

94	P1BS		X	X	PHR1-binding sequence found in the upstream regions of phosphate starvation responsive genes (PSR)
95	POLASIG1	X		X	poly A signal found in legA gene , near upstream elements (NUE) in At
96	POLASIG2	X	X	X	"PolyA signal"
97	POLASIG3		X	X	"Plant polyA signal"; Consensus sequence for plant polyadenylation signal
98	POLLEN1LELAT52	X	X	X	One of two co-dependent regulatory elements responsible for pollen specific activation of tomato (L.e.) lat52 gene
99	PREATPROD		X	X	PRE (Pro/hypoosmolarity-responsive element) found in the promoter region of proline dehydro genasegene in At.
100	PRECONSCHSP70A		X		Consensus sequence of PRE(plastid response element) in the promoters of HSP70A in Chlamydomonas;
101	PROXBBNNA	X			B-Box proximal portion, found in napA gene of B.n. Required for seed specific expression ABA responsiveness
102	PYRIMIDINEBOXOSRAMYIA	X	X		Pyrimidine box found in rice(O.s.) alpha-amylase (RAmy1A) gene; Gibberellin-respons cis-element of GARE and pyrimidine box are partially involved in sugar repression; Found in the promoter of barley alpha-amylase (Amy2/32b) gene which is induced in the aleurone layers in response to GA; BPBF protein binds specifically to this site
103	QARBNEXTA			X	QAR (quantitative activator region)" in promoter region of Bn
104	QELEMENTZMZM13			X	"Q(quantitative)-element" in maize
105	RAV1AAT	X	X	X	Binding consensus sequence of Arabidopsis
106	REALPHALGLHCB21		X	X	"REalpha" found in Lemna gibba Lhcb21 gene promoter

107	RHERPATEXPA7	X	X	X	"Right part of RHEs (Root Hair-specific cis-Elements)" conserved among the At A7 (AtEXPA7) orthologous and paralogous genes from diverse angiosperm species with different hair distribution patterns
108	ROOTMOTIFTAPOX1	X	X	X	Motif found both in promoters of rolD (<i>Agrobacterium rhizogenes</i>)
109	RYREPEATBNNAPA	X	X	X	"RY repeat" found in RY/G box (the complex containing the two RY repeats and the G-box) of napA gene in Bn Required for seed specific expression
110	RYREPEATGMGY2	X	X	X	"RY repeat motif (CATGCAT)"; Present in the 5' region of the soybean
111	RYREPEATLEGUMINBOX	X	X		RY repeat (CATGCAY)" or legumin box found in seed-storage protein genes in legume such as soybean
112	RYREPEATVFLEB4	X	X		RY repeat motif"; quantitative seed expression; Binding site of A.t. B3-domain-containing transcription factor mediates abscisic acid-induced transcription; FUS3 protein physically interact with two RY elements present in the AtGA3ox promoter
113	S1FBOXSORPS1L21	X		X	"S1F box"plays a role in downregulating RPS1&RPL2 encoding the plastid ribosomal protein S1& L21promoter a
114	SEBFCONSSTPR10A	X		X	Binding site of the potato silencing element binding factor(SEBF) Similar to the auxin response element
115	SEF1MOTIF		X		"SEF1(soybean embryo factor 1)" binding motif
116	SEF3MOTIFGM		X		SEF3 binding site
117	SEF4MOTIFGM7S	X	X	X	"SEF4 binding site"; Soybean consensus sequencein 5'upstream region of beta-conglycinin (7S globulin) gene
118	SGBFGMGMAUX28	X			bZIP proteins SGBF-1 and SGBF-2 binding site in soybean

119	SITEIIATCYTC	X	X		"Site II element"
120	SORLIP1AT	X		X	one of "Sequences Over-Represented in Light-Induced Promoters in At. Over-represented in light-induced cotyledon and root common genes and root-specific genes"
121	SORLIP2AT	X		X	one of "Sequences Over-Represented in Light-Induced Promoters in At Computationally identified phyA-induced motifs"
122	SP8BF1BSP8BIB		X		One of SPBF binding site(SP8b)
123	SPHCOREZMC1	X			Core motif of Sph element in maize C1 gene to which maize VP1 protein binds; VP1-responsive element specifically required for expression of the maturation program in seed development
124	SREATMSD	X	X	X	"sugar-repressive element (SRE)" found in 272 of the 1592 down-regulated genes after main stem decapitation in Arabidopsis
125	SURECOREATSULTR11	X	X	X	Core of sulfur-responsive element (SURE) found in the of SULTR1 in At.
126	SV40COREENHAN	X		X	"SV40 core enhancer"
127	T/GBOXATPIN2	X	X	X	"T/G-box" Involved in jasmonate (JA) induction of proteinase inhibitor II and leucine aminopeptidase genes
128	TAAAGSTKST1	X		X	TAAAG motif found in promoter of Solanum tuberosum
129	TATABOX2	X	X	X	"TATA box"; TATA box found in the 5'upstream region of pea legA gene
130	TATABOX3		X		TATA box
131	TATABOX4		X	X	"TATA box"; TATA box found in the 5'upstream region of sweet potato sporamin A gene

132	TATABOX5	X	X	X	"TATA box"; TATA box found in the 5'upstream region of pea
133	TATABOXOSPAL	X	X	X	binding site for OsTBP2
134	TATAPVTRNALEU		X	X	TATA-like motif" Frequently observed upstream of plant tRNA genes
135	TATCCACHVAL21			X	"TATCCAC box" is a part of the conserved cis-acting response complex(GARC)
136	TATCCAOSAMY			X	"TATCCA" element found in alpha-amylase promoters of rice
137	TATCCAYMOTIFOSRAMY3D			X	"TATCCAY motif" found in rice a GATA motif as its antisense sequence; TATCCAYmotif and G motif are responsible for sugar repression
138	TBOXATGAPB		X		"Tbox" found in the Arabidopsis thaliana (A.T.) GAPB gene promoter Mutations in the "Tbox" resulted in reductions of light-activated gene transcription; GAPB encodes the B subunit of chloroplast glyceraldehyde-3-phosphate dehydrogenase(GADPH) of A.T.;
139	TGACGTVMAMY		X	X	TGACGT motif" found in the Vigna mungo (V.m.) alpha-Amylase(Amy) gene promoter Required for high level expression of alpha-Amylase in the cotyledons of the germinated seeds
140	TGTCACACMCUCUMISIN	X		X	"TGTCACA motif" A novel enhancer element necessary for fruit-specific expression of cucumisin gene
141	TRANSINITMONOCOTS			X	Context sequence of translational initiation codon in monocots
142	UP1ATMSD	X			"Up1" motif found in 162 of the 1184 up-regulated genes after main stem decapitation in Arabidopsis
143	UPRMOTIFIIAT		X		Motif II" in the conserved UPR (unfolded protein response) cis-acting element in Arabidopsis genes coding for SAR1B, HSP-90, SBR-like

144	WBBOXPCWRKY1	X	X		"WB box"; WRKY proteins bind specifically to the DNA sequence motifTTTGAC(C/T), which is known as the W box; alpha-Amy2 genes in wheat, barley, and wild oat, and a transcription factor gene in Arabidopsis
145	WBOXATNPR1	X	X	X	"W-box" found in promoter of Arabidopsis thaliana (A.t.) NPR1 gene; Located between +70 and +79 in tandem; They were recognized specifically by salicylic acid (Sasaki, Christov, Tsuda, & Imai, 2013)-induced WRKY DNA binding proteins
146	WBOXHVISO1	X	X	X	SUSIBA2 bind to W-box element in barley iso1 (encoding isoamylase1) promoter
147	WBOXNTCHN48	X		X	"W box" possibly involved in elicitor-responsive transcription of defense genes in tobacco
148	WBOXNTERF3	X	X	X	"W box" found in the promoter region of a transcriptional repressor ERF3 gene in tobacco; May be involved in activation of ERF3 gene by wounding;
149	WRKY71OS	X	X	X	"A core of TGAC-containing W-box" a transcriptional repressor of the gibberellin signaling pathway
150	ZDNAFORMINGATCAB1	X	X		Z-DNA-forming sequence" found in the Arabidopsis(A.t.) chlorophyll a/b binding protein gene (cab1) promoter; Involved in light-dependent developmental expression of the gene; "Z-box"; Activation of Z-box containing promoters is regulated by downstream regulatory components

lBnECP63pro UTR -cDNA-yfp

3
5' ACTTGAAACCTTTCACAAGTAGAGAAGCGAAGCCatggcgtcggagaaacaacgaaagac 1140
o
o
o
5' ggagagagccgaggtcgcagcaaggctagcagctgaagacttgcatgacataaacaaca 1200
o
o
o
5' ccgtgacgatgacgtggcattgtataaggtaacggagagaaacagtcgaacatccaccaga 1260
o
o
o
5' gcaggagagccaggggtgataggatctatgttcagagccgtgcaaggcacatacgagca 1320
o
o
o
o qPCR primer F
5' cgccagagacgctgttgcggaaaaagccacgatgcggccgtggcgaccagtgagggagc 1380
o
o
o
o qPCR primer R
5' taagatggcctcagagaaaagcggctggagccaaagatgcaacccttgagaaagcgaaaga 1440
o
o
o
5' gacggcggattacacggctgacgagcggaaggaagccaaagataagacggcggagaagat 1500
o
o
o
5' gggagaatacaaaagactatacgggtggataaaggcgggtggaagcgaaagataagacggcaga 1560
o
o
o
5' gaaggcaaaaggagacggcgaaactatacagcggataaaagccaaagaggcgaaggacaagac 1620
o
4
5' tgctcagaaggtgggtgagtataaggattacacgggtggacaaggccaaagaggcgaagga 1680
o
o
o
5' ttacacggcggaaaaaggctattgaagcaaaagataaagacggcggagaaggcgggagagta 1740
o
o
o
5' taaggactacacgggtggagaaggcggctgaagggaagatggttgagtgaagcttgg 1800
o
o
o
5' agagttgaaggatagtgcataagatacggccaagagagctatgggtttcttgctctggaaa 1860
o
o
o
5' gacggaagaaaccaaacaacaaaagctattgagaccaaagatagtccaaggaaaaaatgga 1920
o
o
o
5' ggaagctggagaagaaacgagacgtaagatggaggagaagagattggaaggtaaagaact 1980
o
o
o
5' caaataccaagctggggagaaggcacgtgaggcatctcaaaagacaagagagagtactga 2040
o
o
o
5' ttcgcgagctgaaagagcccacgagacaaaagattctgctgcggttaggggtaataagac 2100
o
o
o
5' gaaggggacaatatattgggtgcactaggggaatgtgacggacgccataaagagcaaaactgac 2160
o
o
o
cDNA

Figure A1: map of BnECP66-ORF-eYFP showing priming site of qpcr primers

1BnECP248proUTR-cDNA-yfp

Figure A2: map of BnECP248-ORF-eYFP showing priming site of qpcr primers

ea411pro-cDNA-yfp

5' AATGCAGGACATGGGAGTCTGAAGAACCCTTAAGCAATGGTGGTCTAGGGAGATTGCCAATG 60
+ + + + +
Promoter

5' ACTCAGCCAAGAAGAGACCATCTAGTAGCACAAAACGGTGGCGGGAGGAAGCTTTTCGAT 120
+ + + + +
Promoter

5' GAGATCGAACTGTTC AACACAGACTCGCTTG TGAAGAGGTAAGCATCTCTCTGTCTCTC 180
+ + + + +
Promoter

5' CGTGGATGAGTTTTAACATTTTGCTTAGTCTGAAGTGTTTTTGCGTCGCATTAGGTGGAG 240
+ + + + +
Promoter

5' AGAGTTTTTGATTCTAATCTTCTGACCCGCACGAGCTTGATAAGGCCAAGAAGCTACTC 300
+ + + + +
Promoter

5' AAGGTCAGAGACGTAACAAGAACGCTTTATTATTTATATAAACGATTGATCAAAGTGAA 360
+ + + + +
Promoter

5' ACTTGGTTGCTTTTATTGGATGATGCAGGAACATGAACAGGCGCTTGTTTCTGCCATTGC 420
+ + + + +
Promoter

5' AAGGCTTGCGGATGCTTCTGATTTTCGAAAGCGGTAAGAAACATTAAAGGGAGGGCTTTGTA 480
+ + + + +
Promoter

5' TATGACCTGAGAATAAGTTTATGCCTATTCAAATCACATGTTTGT CAGAACGACTCAA 540
+ + + + +
Promoter

2

5' TGTCTTTGTTTCATTACAAGACCAGTGATGTTATGTTTTtTTTTATCTTGGTTTTGTGT 600
+ + + + +
Promoter

5' GTTGCTGCTGAACCAGATGGAGAAGAAGCGTATATGCATGAGCTTCCAGTGCATGAGGGA 660
+ + + + +
Promoter

5' TAACAGTGGATAAAGAAAAGAGATTGCCCCATCAAACCTTAGCAYACTCTTTTCATGTCC 720
+ + + + +
Promoter

5' tTCACGTATAGCAAACCGAACAGTTCTCCGAAAATGAAAAATTATGTATTGgTTTCA 780
+ + + + +
Promoter

5' AAATGTATTAAAATAACCGTTAGAAAtTTTGtTTGTAAAACAGTTAATTATTTTCAAA 840
+ + + + +
Promoter

5' TTTCTTAATAATTTATGTGACCATCAGGTCTATCCATCAAGATTATGCTTTACCTTTGTG 900
+ + + + +
Promoter

5' GTTAACGTTAAATTTGCAACATGTATCGAATCAGTGATGACGAACGTGCCaAAAGCCTAA 960
+ + + + +
Promoter

5' TGGTGTAGTTAGATTTTTGGTGTATACGCGGCGGACGTATAAAACAGCTATAACTAATAA 1020
+ + + + +
Promoter

5' CTCTTTCCACCACTAACCAAGCCTAACACGCGTCAACATGCGAACTGCTCTACGGCTCT 1080
+ + + + +
Promoter

Figure A3: map of BnLEA411-ORF-eYFP showing priming site of qpcr primers

ECP 248 promoter	1	TCCAAGCTTAACAAGCCCACGAACATCACAACATACGAACCTCAACCAAGCCCATCTTCA	60
ECP 66 promoter		-----	
ECP 248 promoter	61	TGACACACGCATGTACTATTTCTGTAGTACTTCACGAACTGACACAGAATATACATCTT	120
ECP 66 promoter		-----	
ECP 248 promoter	121	CAGAATATTGTCTTGTTAAGATTAATAAAATCAAAACCAAGAAGTATATAAAGTTGATC	180
ECP 66 promoter		-----	
ECP 248 promoter	181	ATATAACCCAAAACATAAGTTTGTATGTGGGTAATCCATGGTTTAGTTAAGTTTCAACAAA	240
ECP 66 promoter	1	-----CCACCGT-----TTTTTTCCTAATGGGCCTGT-GATGATGACCCA--ATAAAT	45
ECP 248 promoter	241	AAAAAATCCACGGTTTGTATTTAATTTGAGTATAAATGATAAGTTATCATATGC	300
ECP 66 promoter	46	TAAATAGGATTGA-----ATACGTTTAAATCATATTT---TCA-----GTAGTCTGAG-	90
ECP 248 promoter	301	AAAACGTGACTGATTTTAGATGAGTTGAGATCA-ATTACGCGTCATGGTGTGTCGGAGA	359
ECP 66 promoter	91	--TGTGTTCAAATGA-TGTC-TTTTTTACGTAGT-ACAGTCTC-CAAT---GC-CTCTTA	140
ECP 248 promoter	360	AATAAG-AAAAAGGAGTGACGTTTATAAAATAGAGACGTTACACATTAGAGCACTTTT	418
ECP 66 promoter	141	ACTGTCTATG---CTAAATCATAATTTTAACAGTGCAGTATA-ATTCGTACAACCGATCTT	196
ECP 248 promoter	419	AACGTGATTGGTTTATGTGATTAATAATAACACT--AGTCTATGTGCGAACA-----	467
ECP 66 promoter	197	GCAACGGCTCAGCGCAAGCTCAATGAGCCTGAGAATCAAAGAGAAAGATCACTAAATCT	256
ECP 248 promoter	468	-CAA---CTAATGTGAATGT-GATTAGCTTGTAAACGAAAGGTTAACGTGA-TAAATT--	519
ECP 66 promoter	257	ACTTCTCTTCTTCATAGA-CTTAAAGGTTAGATTCTTCACTCGGTATCT-TCARGGC	314
ECP 248 promoter	520	-----A-AGATCTTCAA-----TATATAGCGT-AGTC-----TCTGTCAA--T	553
ECP 66 promoter	315	AACCATGCTTAAGCATCTCAAG-AAAATACATTGTTTATTAAATGTGGTTAGTTCAACT	373
ECP 248 promoter	554	AA-CATGCTAACCTGAATCAAGTAAAAAATATGTGTAGATCGGATCCCCATAGT--AAAG	610
ECP 66 promoter	374	AGAAA-AATATCAAAT-----GCAAAACCTATATCATCATGA-----TATGATGA--AACT	421
ECP 248 promoter	611	AGAATCAAGTTAAATATGGGTCAAAA---ATATCTTTATAAAGTTTGTAGGTGAACAAAT	667
ECP 66 promoter	422	GTTCGCAAGCGT---CTTACTGACTCAACATCTTGT---ACTTTCT---TCGCCCT---	468
ECP 248 promoter	668	GT--GAATTAGTTAACTTTGTAAACAAATTATTTGTCCGACCTTGCTGGATCGCTTGGG	725
ECP 66 promoter	469	-CGACT--TATAAATTTTATAGCGTTTCACGTTTGTGGTGATTTTATTTTAGATTAATT	525
ECP 248 promoter	726	GCTACTAGCAGAAAATTCATA-CTTTTAAC-----TGCCTTAACGCAAAATGCAACT	775
ECP 66 promoter	526	AGTGTGCTAATTCTGTCCACCACTTCTCTGTGTATTACGGTGTGCGGTTTACTTTGT	585
ECP 248 promoter	776	A-TCTCGTTAA---GTTTCTCAACTAGCTGTTGT-TTTTAGAGTG--GTT---CT---	821
ECP 66 promoter	586	AGATAGCTCTCCTTGTA--AAGGTTAAGTGTGGCCTTATTTCTCTG--TAAACT--CT	639
ECP 248 promoter	822	--AAAACCTTTTTTACTCAAGCAGGTTTAAATT--CATCAACCTTTGATTAGCTAGAG	877
ECP 66 promoter	640	TGAACGTTTTTA---TTTTA---TGTTTAACATATAGGCTATATATCAATACTCAGTT	691
ECP 248 promoter	878	TTATCCTTGTGTAGTAATTTTAAATCTCATCAAGTTAATAGCTTCACATCAA---CATTT	933
ECP 66 promoter	692	AGATATAAAATTCCTCGCGCAGCTCAGAGGTGTTTGATACAACGACCGACCATCAATGCTT	751
ECP 248 promoter	934	AG-CATACATTCTCTGAGCAACTAA-----TT-----AA--AATTATCTTAAATGTT	978
ECP 66 promoter	752	GCTTATGCTTAAAGTCATGACCATGAACATGAAGATCCATGCATGAAAGGTCACGCTGT	811
ECP 248 promoter	979	GATACAGCC--AAGTCA--ACC--GGCCAATGACCAT-CATGCATGAAAGGACCACGCTGT	1031
ECP 66 promoter	812	CATGGTACATCCCAAGAGGATA--AACGGTCTATACGTGTACACGAAAGGCTACTTGGGA	869
ECP 248 promoter	1032	CGTGGTACATCAAGAGAGGATAAAACGATCTA-----CTCAAGATA---TACGTG--	1079
ECP 66 promoter	870	AGGCAAGAAGAAAAGATTAAAGAAGCTAATAACGTGGCAGCTCTTGCCTCTGCCTAT	929
ECP 248 promoter	1080	TCGCACG-----AAAGA--CTA---CTTGGCAGCTCCTACGCTGCCTAT	1119
ECP 66 promoter	930	CTACCACGTGGAGTTTCCGCATAAAAGGAGCATTCACTATAATAAGCAACAAAAA-CACA	988
ECP 248 promoter	1120	CTACCACGTGGGTCTCCACAGAAACAGAGCATTACTATAAGATGCAACAAAAAGCACT	1179
ECP 66 promoter	989	TCGACT--AATAACACAA---GTCAAAAATATTCAAAAAATCCACAAATCACTTGAA--	1040
ECP 248 promoter	1180	GTAACGCACTATCACAACCCGTAACCTTGTTCACGAATACTCAAGTAA--TTGAGCA	1238
ECP 66 promoter	1041	-ACCTTTCACAAGTAGAGAAGCGAAGCCATGGCGTC	1075
ECP 248 promoter	1239	GAGCATCTAGATGT-GAC-----	1255

Figure A4: Promoter alignment of the ECP 66 promoter and ECP 248 promoter showing sequences homolog.