Molecular characterisation of the "LEAome" in the resurrection plant *Xerophyta humilis* (Baker)

Robyn Waters

Supervisor: Professor Jill M. Farrant

Co-supervisors: Dr Suhail Rafudeen and Dr Nashied Peton

Submitted in fulfilment of the requirements for the degree of Master of Science in the Department of Molecular and Cell Biology, University of Cape Town, South Africa.

August 2015

Declaration

The experimental work described in this MSc thesis was undertaken in the Molecular and Cell Biology Department, University of Cape Town, South Africa under the direct supervision of Professor Jill M. Farrant, Dr Suhail Rafudeen and Dr Nashied Peton, from January 2014 to August 2015.

I hereby declare that this thesis entitled:

Molecular characterisation of the "LEAome" in the resurrection plant *Xerophyta humilis* (Baker)

is my own work and has not previously, in its entirety or in part, been submitted at any university for another degree, and that all the sources I have used and cited have been stipulated and acknowledged through complete referencing.

I know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Signed by candidate

Robyn Waters

August, 2015

Acknowledgments

Thank you endlessly to my supervisor Professor Jill Farrant and co-supervisor Dr Suhail Rafudeen for making this entire Masters project possible, and for ALWAYS being supportive and encouraging of me. Thank you for guiding me in the right direction and growing my confidence internally and in science. Thank you to Jill and L'Oréal for the funding for this project.

To Nashied Peton- for being the greatest co-supervisor of all time. Thank you for making lab work interesting, challenging me to answer all my own questions and gain the confidence I have needed all along. I appreciate all the effort, energy, and time that you devoted to my project and to helping me every step of the way. You have always been so incredibly helpful and for that, I am forever grateful.

Dr Amelia Reddy, Keren Cooper and Dr Amelia Hilgart, thank you for always being there for me and always making the time to help me troubleshoot or to just give advice, it has all been so greatly appreciated.

Thank you to Madhu Chauhan for all her help with the CD machine and patience with obtaining the data of my interesting and challenging proteins.

Thank you to Jacqueline Meyer and Dr Nicki Adams at CPGR for all your hard work and energy devoted to my gene expression analysis. Thank you to Dr Colin Anthony and Dr Joanna Evans of the Infectious Diseases and Immunology Department, Medical School, UCT, for their guidance, support and advice with the ddPCR™ protocols.

A special thank you to my lab colleagues for all the help, laughs, advice and support along the way- Rizqah Kamies, Tamaryn Ellick, Humaira Lambarey, Cynthia Fan, Ali Kiyaei, Hawwa Gabier, Khalil Malekzadeh, Kirshni Govender , Tim Dennis and especially to Zoe Gill.

Lastly and most importantly, thank you to my parents, Mark and Hilary, for your endless support and love. Thank you for making all of my educational dreams a reality and being behind me all the way. You have taught me the importance of hard work and always finding the balance. I love you both dearly and appreciate everything you have done for me throughout my life.

Abbreviations

ANOVA Analysis Of Variance

AWC Absolute Water Content

BLAST Basic Local Alignment Search Tool

CD Circular Dichroism

cDNA Complementary DNA

Ct Cycle threshold

CPGR Centre for Proteomic and Genomic Research

DW Dry Weight

ddPCR™ Droplet Digital PCR

EF1a Elongation Factor 1a

EM Early Maturation

FW Fresh Weight

FTIR Fourier Transform Infrared Spectroscopy

IDP Intrinsically Disordered Protein

IPTG Isopropyl β-D-1-thiogalactopyranoside

kDa Kilo Daltons

LD Leaf Dehydrin

LB Luria-Broth

LEA Late Embryogenesis Abundant

MCS Multiple Cloning Site

mRNA messenger RNA

NCBI National Center for Biotechnology Information

NMR Nuclear Magnetic Resonance

OD Optical Density

ORF Open Reading Frame

PAGE Polyacrylamide Gel Electrophoresis

PBS Phosphate Buffer Saline

PCR Polymerase Chain Reaction

Pfam Protein Families

pI Isoelectric point

PONDR Predictor of Naturally Disordered Regions

qPCR quantitative Real-Time PCR

RD Root Dehydrin

RWC Relative Water Content

TFE Tetrafluoroethylene

Xh *Xerophyta humilis*

Tables and Figures

Figures

Tables

Abstract

Studies on resurrection plants and other anhydrobiotic organisms, have shown that Late Embryogenesis Abundant (LEA) proteins are expressed upon the onset of desiccation and are therefore inferred to be associated with the desiccation tolerance response. To date, despite some 25 years of research on these proteins, there is still very little understanding of the physiological function(s) of the majority of LEAs. This is because they lack tertiary structure in the hydrated state, making assigning of physiological roles difficult.

This MSc study was undertaken to investigate the gene expression of a set of 21 putative LEAs during dehydration and subsequent rehydration stress, in the resurrection plant *Xerophyta humilis* (Baker). Recombinant proteins were expressed for 3 of the LEA genes from this set in order to perform structural studies and to ascertain their LEA status. These studies were conducted with the purpose of shedding light on the role of LEAs in desiccation tolerance, to add to the ever-growing transcriptomic and proteomic data, and to the current knowledge of these enigmatic proteins.

Quantitative real-time gene expression (qPCR) analysis was conducted on the set of 21 full length *X. humilis* cDNA clone nucleotide sequences, with similarities to late embryogenesis mRNA sequences, derived from a study conducted by Collett *et al*., (2004). Expression analysis was conducted in both leaves and roots, across a dehydration and rehydration profile of *X. humilis*. Of this total group of 21 full length cDNA clones, three LEAs; XhLEA2-3 and XhLEA2-6 (two putative Group 2 LEA genes) and XhLEA3-5 (a putative Group 3 LEA gene), were chosen for cloning and expression studies. cDNAs of these XhLEAs were cloned into a modified bacterial expression vector and recombinant protein expression was attempted in *E. coli*.

qPCR analysis of the 21 full length cDNAs was successful for leaf vegetative tissues across the dehydration and rehydration profile only, whereas root tissues proved hugely troublesome from the beginning stages of RNA extraction and ultimately through to the generation of qPCR data. Preliminary Droplet Digital PCR (ddPCR™) trials with root tissues, as an alternative to relative gene expression using qPCR analysis, proved inconclusive.

Western blot analysis resulted in the successful detection of the putative Group 2 LEA protein, XhLEA2-3, in both the crude and his-tagged purified fractions, whereas the other

two proteins chosen, namely XhLEA2-6 and XhLEA3-5, were not successfully purified and thus, not detected.

Preliminary structural analysis of the recombinant LEA proteins using Circular Dichroism (CD), pointed to a putative alpha-helix conformation for recombinant proteins of both XhLEA3-5 and XhLEA2-3, however, additional work needs to be conducted for more conclusive results, using methodologies of Ginbot (Phd thesis, 2011).

This is the first report in which the expression profiles of 19 LEA-like transcripts in the leaves of the resurrection plant *X. humilis*, identified by Collett *et al*., (2004), have been investigated using qPCR. This study includes the two Group 1 LEAs, XhLEA1-4S1 and XhLEA1-1S2, that have been previously studied in this regard by Ngubane (MSc thesis, 2008) and Ginbot (2011).

Table of Contents

Chapter 1

The Desiccation Tolerance Phenomenon in Angiosperm Resurrection Plants and the Enigmatic LEA Proteins: Literature review

1.1 Introduction

Aquatic algal ancestors gave rise to early land plants and the plants that inhabit our world today (Farrant and Moore, 2011). Successful adaptation to dry terrestrial environments required plants to overcome one of the first and most formidable obstacles – dehydration. The evolution of a plant cuticle, the stomata and xylem transport tissues, allowed the resistance and regulation of water loss, respectively. The acquisition of these evolutionary traits thrust angiosperm plants towards improved drought-avoidance mechanisms, and hence, dehydration-tolerance. These mechanisms can be extremely effective in preventing the development of injuriously low plant water contents.

Many higher plants have evolved to possess a single phase in their life cycle where their tissues are able to survive drying (Scott, 2000; Walters *et al*., 2002; Vicré *et al*., 2004; Moore *et al*., 2009; Farrant and Moore, 2011; Farrant *et al*., 2012). Species survival was ensured by the development of desiccation tolerance in reproductive structures, such as seeds, which allowed their survival in adverse environmental conditions and then germination and plant growth once favourable conditions were re-established (reviewed by Gaff and Oliver, 2013). Such seeds are termed 'orthodox' (Roberts, 1973).

Over time, higher plants have developed a suite of survival strategies for an ever-changing environment, including physiological, genetic, biochemical and developmental changes, amongst others, in response to adverse and stressful conditions, specifically to avoid excess water loss from their vegetative tissues (Hong-Bo *et al*., 2005). It is understood that the vegetative tissues of most plants are extremely sensitive to desiccation. There are however, a small proportion of plants, commonly termed 'resurrection plants', which are completely 'desiccation tolerant' (reviewed by Gaff, 1977; Oliver *et al*., 1984; Farrant *et al.,* 2007; 2012; Moore *et al*., 2009; Dinakar and Bartels, 2013; Gaff and Oliver, 2013). It has been proposed that desiccation tolerance in vegetative tissues of angiosperms, over time, has been co-opted from mechanisms present in orthodox seeds (reviewed in Farrant and Moore, 2011).

The term 'desiccation-tolerance' describes the ability of plant tissues to survive the loss of up to 95% of their protoplasmic water or the dehydration to tissue water concentrations of less than 0.1g $H_2O.g^{-1}$ where the hydration shell of molecules is gradually lost (Scott, 2000; Hoekstra *et al*., 2001; Farrant *et al*., 2007; Farrant *et al*., 2012). The rehydrating plant tissue can revive from this air-dry state or from equilibration with air of approximately 0% relative humidity (RH) (Gaff and Oliver, 2013).

Resurrection plants are able to survive with desiccated tissues for extended periods of time and the ability to recover full metabolic competence within these tissues, upon rehydration, is vital to this survival (Hoekstra *et al*., 2001; Farrant *et al*., 2012). The ability to survive desiccation in the vegetative tissues of plants is an uncommon, yet a widespread phenomenon in the plant kingdom ranging from algae, lichens, and bryophytes to 135 angiosperm species, but is completely absent in gymnosperms (Oliver *et al*., 1998; Oliver *et al.*, 2000; Vicré *et al*., 2004; Farrant *et al*., 2007; Farrant and Moore, 2011; Farrant *et al*., 2012; Gaff and Oliver, 2013).

Mechanisms of tolerance in angiosperm higher plants, differ from those in lower order species, and for the purpose of this thesis, only mechanisms present in angiosperms will be discussed.

The majority of angiosperm resurrection plants are found in tropical and sub-tropical zones in southern Africa (Scott, 2000; Gaff and Oliver, 2013). Various mechanisms, both biochemical and physiological, that mitigate desiccation stress, enable these resurrection plants to proliferate in, and be well adapted to, arid environments (Scott, 2000).

1.2 The importance of water in the life of a plant

Resurrection plants survive virtually complete water loss or 'anhydrobiosis'. The remarkable ability to enter anhydrobiosis, translating as 'life without water', a state characterised by limited intracellular water and almost no metabolic activity, requires a co-ordinated series of events during dehydration that enables these desiccation-tolerant plants to survive for many months or even years (Hoekstra *et al*., 2001; Kranner *et al*., 2002; Farrant *et al*., 2007; Farrant *et al*., 2012). In order to gain an understanding of desiccation tolerance and the resurrection phenomenon in plants, it is essential to recognise the importance of water in the life of a plant and the effects of the loss of water on the metabolic, physiological and physical functioning of plants (Farrant *et al.,* 2012). Every plant must ensure that there is a fine equilibrium between the uptake and loss of water over time, and this poses a challenge to land plants that

have been rendered immobile over the course of evolution (Hsiao, 1979; Dainty, 1976). Water plays a crucial role in the life of a plant. It is a major limiting factor in plant growth, development and reproduction (Mundree *et al*., 2002; Vicré *et al*., 2004; Farrant *et al.,* 2012). It plays a role at various levels of the physical and physiological make-up of the plant. At the cellular level, water plays a structural role, providing mechanical stabilization, by filling intracellular spaces resulting in turgor pressure (Walters *et al*., 2002; Farrant *et al.,* 2012). Hydrophobic and hydrophilic interactions as well as intermolecular distances are controlled by the presence of water at the molecular level. The provision of these interactions and controls, determine the conformation of proteins and macromolecules and their interactions and partitioning within plant organelles (Walters *et al*., 2002; Farrant *et al*., 2007; Farrant *et al.,* 2012). If water completely dissipates from these living plants, the driving force for cellular organization is essentially lost (Hoekstra *et al*., 2001). It is therefore expected that water deficit is accompanied by numerous stresses and strains (Walters *et al*., 2002; Farrant *et al.,* 2012).

Resurrection plants, nevertheless, manage to survive during periods of severe desiccation, indicating that multiple mechanisms and strategies have evolved over time, that allow the maintenance of important native cellular structures in the absence of water (Hoekstra *et al*., 2001).

1.3 Research into desiccation tolerance and the resurrection phenomenon in Angiosperm plants

Desiccation tolerance is a complex and diverse phenomenon and while there are many similarities amongst tolerant species, there are also differences in their acquisition of this tolerance. This highlights the need to identify and understand the varying levels of plant physiological responses to drying, as it is unrealistic to assume that there is a single "gene for desiccation tolerance" (Mundree *et al*., 2002). Through recent research, it is becoming clear that some generalizations may be made regarding resurrection plants (Scott, 2000; Farrant *et al.*, 2012; Dinakar and Bartels, 2013; Gaff and Oliver, 2013).

Many putative protectants have been proposed to be part of the process aiding the survival of resurrection plants during dehydration and rehydration. There are various proteins involved in these two processes that have, as yet, not been fully investigated in resurrection plants. Therefore the need for further studies on the adaptations of resurrection plants aiding survival during desiccation is highlighted. Molecular studies on resurrection plants, involving the

plant transcriptome and proteome, have indicated the up-regulation of chaperonin-like proteins, possibly functioning in protein folding, assembly, translocation and degradation in many normal cellular processes and re-establishing protein homeostasis in desiccation (Ingram and Bartels, 1996; Bockel *et al*., 1998; Vicré *et al*., 2004; Ingle *et al.,* 2007; Farrant *et al*., 2009; Farrant *et al.,* 2012). In addition, anti-oxidants, whose role is to eliminate the excess free radicals produced as a result of metabolism disruption, especially during the dehydration stages of these plants, have been found to be important (Vicré *et al*., 2004; Illing *et al*., 2005; Farrant *et al*., 2007). Recent studies have demonstrated, among other proteins, the up-regulation of mRNA transcripts of Late Embryogensis Abundant (LEA) proteins in response to desiccation and other abiotic stresses (Rabbani *et al*., 2003; Collett *et al*., 2004; Olave-concha *et al*., 2004; Illing *et al*., 2005; Cuming *et al*., 2007). Particularly, transcriptomic studies on *X. humilis* have revealed that a high proportion of the up-regulated genes, present in plant tissues in the dehydrated state, included those encoding protective proteins such as dehydrins and LEA's (Collett *et al*., 2004). Therefore, it is postulated that these LEA proteins may play a role in desiccation tolerance within resurrection plants. However, the role of LEA proteins remains enigmatic due to the difficulty in studying these proteins, which is attributed to their lack of structure under physiological conditions. A challenge facing researchers today when investigating these enigmatic proteins is to make sense of their various *in vitro* defined functions within living cells and hence answer the question: "Are the LEA proteins truly multi-talented, or are they still just misunderstood?" (Quoted directly from Tunnacliffe and Wise, 2007).

A number of proposed mechanisms and functions of these LEA proteins will be described in light of the biochemical, biophysical and structural information available. In addition to this, their classification will be summarised in this chapter.

1.4 The enigmatic Late Embryogenesis Abundant (LEA) proteins

The name "Late Embryogenesis Abundant proteins" reflects the origins that LEA proteins were first discovered and described as those abundantly up-regulated in the late stages of orthodox seed development (Dure *et al*., 1981; Galau and Dure, 1981; Galau *et al*., 1986; Illing *et al*., 2005; Goyal *et al*., 2005; Amara *et al.,* 2014) comprising 4% of total cellular protein (Roberts *et al*., 1993; Ingram and Bartels, 1996). LEA proteins were first identified as a number of small proteins shown to accumulate during the mature phase of cotton (*Gossypium hirsutum*) seed embryogenesis (Galau and Dure, 1981; Goyal *et al*., 2005;

Battaglia *et al*., 2008; Tunnacliffe and Wise, 2007) and this was coincident with the necessity of acquiring the ability to withstand desiccation, particularly at the point of post-abscission, when the seed is no longer hydrated by the mother plant. This led to the early proposal that these proteins may play a role in facilitating protection against extreme water deficit stress and that there may be a correlation between LEA gene expression and protein production, and desiccation tolerance (Hoekstra *et al*., 2001). LEA proteins have subsequently been discovered in a wide range of organisms, but mainly appear to be induced in response to abiotic stresses (drought, heat, freezing and salt stress *inter alia*) that result in subcellular water deficit. Importantly, they are apparently commonly produced in desiccation tolerant organisms such as some bacteria, tardigrades and nematodes, and are widely present in both orthodox seeds and vegetative tissues of angiosperm resurrection plants (reviewed by Farrant *et al*., 2012).

The LEA proteins are a very broad grouping of proteins with loose terms and have been separated into different classes and groups based on their occurrence and sequence variability (Table 1.1). Initially, the grouping of these proteins into LEA classes, stemmed from their discovery and abundance in mature seeds instead of their biochemistry, amino acid sequences or protein structure. They were defined solely on their presence, expression and accumulation during the late stages of seed embryogenesis and their clear association with the onset of desiccation tolerance in seeds.

Table 1.1. Correspondence between different nomenclatures of LEA protein groups (Dure, 1989; Battaglia *et al*., 2008; Bies-Ethève *et al*., 2008) **and their corresponding PFAM descriptions, numbers and general names** (extracted from Battaglia *et al.,* (2008) and edited).

Dure	Battaglia	Bies-Ethève	PFAM	PFAM No.	Name
D-19	1	1	LEA_5	PF00477	Em1, Em6
D-11	2	2	Dehydrin	PF00257	Dehydrin, RAB
D-7	3A	3	LEA_4	PF02987	ECP63, PAP240, PM27
D-29	3B	3*	LEA_4	PF02987	D-29
-	4A	4	LEA_1	PF03760	LE25_LYCES
D-113	4B	4	LEA_1	PF03760	PAP260,PAP051
D-34	5A	5	SMP	PF04927	PAP140
D-73	5B	6	LEA_3	PF03242	AtD121,Sag21, lea5
D-95	5C	7	LEA_2	PF03168	LEA14
-	6	8	LEA_6	PF10714	LEA18
-	7	-	ABA_WDS	PF02496	ASR

More recently, these proteins have been grouped according to their amino acid sequences largely contributing to their hydrophilicity. Most of them belong to the "hydrophilins" family, a group comprising of highly hydrophilic, intrinsically disordered/unstructured proteins (IDPs) that are characterised by a glycine enriched biased amino acid composition, also containing other residues enabling flexible conformational changes in the presence of aqueous solutions (Olvera-Carrillo *et al*., 2011). Newly discovered putative LEA proteins are required to share a level of sequence similarity with approved LEA proteins (Tunnacliffe and Wise, 2007). With LEA proteins no longer being restricted to vegetative and embryogenic plant tissues, due to their presence in different organisms, a definition of LEA proteins based solely on expression no longer holds true. Furthermore, it appears that some LEA proteins may be constitutively expressed and this has called into question what actually constitutes a LEA protein with respect to expression (Soulages *et al*., 2003; Wise, 2003; Jones and McQueen-Mason, 2004; Wise and Tunnacliffe, 2004; Battaglia *et al*., 2008). Due to these problems associated with the classification by LEA gene expression, common features and characteristics of LEA proteins have been reported, which are used to facilitate group definition, despite there being little similarity among different families. The unifying feature of LEA proteins is their low molecular weight, ranging from 10-30kDa in size (Hong-Bo *et al*., 2005). Other common features include their low complexity yet with the potential for structural flexibility, water deficit inducibility, unfolded nature in the presence of water and extreme hydrophilicity (with the exception of the D95 LEA family (Galau *et al*., 1993). All LEA proteins are unique in possessing an unusually high proportion of hydrophilic amino acids that are arranged in sequence clusters throughout the entire peptide sequence (Hong-Bo *et al.,* 2005; Illing *et al*., 2005). The extreme hydrophilicity of these proteins is attributed to the rich content of Glycine with a >6% content resulting in a hydrophobicity index of >1, indicating hyper-hydrophilicity (Illing *et al*., 2005; Hong-Bo *et al.,* 2005; Battaglia *et al*., 2008). On the basis of this Gly content, LEA proteins fall into the larger protein group of hydrophilins, largely inferred to be part of the dehydration tolerance response (Battaglia *et al*., 2008). In turn, the high hydrophilicity confers a very high stability to LEA proteins when they are in the presence of water. These proteins are also heat stable, have no structural domains, exhibit no catalytic activity, and many of them are characterised by an absence of cysteine and tryptophan residues (Ingram and Bartels, 1996; Bray, 1997; Battaglia *et al*., 2008; Amara *et al*., 2012). These characteristics, including the boiling-soluble nature of these LEA proteins indicating their hydrated and non-globular form, led to the suggestion that they are involved in protecting plant cells from dehydration induced damage (Zhang *et al*., 2000).

To date, despite some 25 years of dedicated research on these proteins, there is still very little understanding about the physiological function of the majority of LEAs. This is primarily because these proteins lack a tertiary structure in the hydrated state, making assigning of physical roles difficult. Some functions of these proteins have been postulated in reference to their amino acid composition, including: protection of subcellular milieu by acting *inter alia* as ion sequesters, acting as molecular chaperones and/or heat shields, cytoskeletal components and protein/membrane anti-aggregants, aiding water replacement and accommodating vitrification via glass formation of the cell cytoplasm under dehydration conditions (Bray, 1993; Bray, 1997; Hoekstra *et al*., 2001; Wise and Tunnacliffe, 2004; Goyal *et al*., 2005; Berjak, 2006; Tunnacliffe and Wise, 2007).

The expression of LEA proteins in seeds is known to be regulated by the plant stress hormone abscisic acid (ABA) coinciding with the onset of desiccation tolerance (Finkelstein *et al*., 2002). Subsequent transcriptomic studies have shown that LEA genes occur in a wide range of organisms in which expression is rarely constitutive, but rather they are induced in response to various abiotic stresses, the bulk of them involving some degree of subcellular water deficit (Dure, 1993; Rabbani *et al*., 2003; Olave-Concha *et al*., 2004; Goyal *et al*., 2005; Tunnacliffe *et al*., 2010). It has also been reported that most LEA genes are responsive to dehydration, temperature and osmotic (salt) stress, not only in seeds, but in vegetative tissues of plants too (Wise and Tunnacliffe, 2004; Illing *et al*., 2005). Examples of these include the CDT-1 from *Craterostigma plantagineum* (Bockel *et al*., 1998; Bartels and Salamini, 2001), responsive to dehydration stress and COR15α from *Arabidopsis thaliana* (Thomashow, 1999) and PpDHNA from *Physcomitrella patens* (Saavedra *et al.,* 2006), responsive to low temperatures and osmotic stress, respectively.

LEA genes are among the most differentially expressed and abundantly upregulated, as shown in transcriptomic studies conducted on both orthodox seeds and resurrection plants (Collett *et al.* 2004; Illing *et al*. 2005). The research has demonstrated considerable induction of transcripts coding for LEA-like proteins during the drying of resurrection plants, these disappearing shortly after rehydration (Collett *et al*., 2004; Illing *et al.,* 2005). Further research has indicated that some of these LEA genes are similarly induced in orthodox seeds during the maturation drying phase of embryological development and it has therefore been proposed that there is a subset of LEA genes that are specifically involved in the desiccation tolerance response (reviewed in Farrant *et al*., 2012).

A publicly available *Arabidopsis* genome and microarray data set analysed by Illing *et al*. (2005) has shown that out of 35 LEA genes analysed, 15 were exclusively expressed in seeds and a further 6 LEA genes were abundantly expressed during the phase of seed development where tolerance of desiccation is essential, but were also found to be expressed in response to various other abiotic stresses including osmotic, cold and salt stress (Illing *et al*., 2005). Furthermore, one of these LEA genes, a homologue of a group 6 LEA gene, was not expressed in the vegetative tissues of *A. thaliana*, however there was an orthologue present in the desiccating leaves of *X. humilis*. The activation of this "seed-specific" desiccation protection mechanism during drying of *X. humilis* points towards acquisition of desiccation tolerance from seeds. This could potentially have arisen from changes in the regulation of these genes that in desiccation-sensitive plants, such as the study plant *Arabidopsis*, are only abundantly expressed during the water limiting phase of seed maturation.

Adding to our knowledge on LEA genes, the largest set of LEA data from a single resurrection plant – *X. humilis* – was identified by Collett *et al*. (2004) using a mini-microarray screen of 424 cDNAs randomly selected from an 11K normalized library generated from leaf and root tissue of this resurrection plant. Sixteen of 55 genes shown to be upregulated during desiccation were annotated as LEAs. The study, together with the observation that the gene transcripts only became evident once the plant water content fell below 50% RWC, followed by the disappearance of the transcripts upon rehydration, indicates the potential importance of these transcripts in desiccation tolerance acquisition in this resurrection plant. Transcriptomic studies on many other species of resurrection plants have all reported the occurrence of LEA genes associated with the onset of desiccation (reviewed in Farrant *et al*., 2012). Regardless of the varying final numbers of LEA genes and the groups to which they belong, in all studies these genes form a high percentage of the genes tested, with LEA genes from groups 2 and 3 prevailing.

1.5 The classification of LEA proteins

With increasing information available on newly described LEA proteins, differences in expression profiles, description in organisms other than plants and especially with the new bioinformatics tools, the classification has been subjected to different rearrangements (Bies-Ethève *et al.,* 2008; Battaglia *et al*., 2008; Battaglia *et al.,* 2013; Amara *et al*., 2014). In this study, the classification scheme originally proposed by Bray (1993 &1994) and implemented

and refined by Battaglia *et al*., (2008), Hong-Bo *et al*., (2005) and Amara *et al*., (2014) has been adopted.

The group 1 LEA proteins (Pfam PF00477), originally represented by the D-19 and D-132 group of proteins from developing cotton seeds, are a highly conserved, in nucleotide and amino acid similarity, group of LEA proteins that possess a conservative sequence motif composed of 20 amino acid residues (GGQTRREQLGEEGYSQMGRK (Tunnacliffe and Wise, 2007))/ TRKEQ [L/M] G [T/E] EGY [Q/K] EMGRKGG [L/E] (Amara *et al*., 2014)) that may repeat up to four times throughout the protein sequence (Galau *et al*., 1992; Baker *et al.,* 1995; Close, 1996; Tunnacliffe and Wise, 2004; Hong-Bo *et al*., 2005; Battaglia *et al*., 2008; Ginbot, 2011, PhD thesis; Amara *et al*., 2014). This motif is proposed to result from genetic duplication processes followed by recombination or deletion (Tunnacliffe and Wise, 2007) and is highly hydrophilic, being capable of absorbing large quantities of bound water (Close, 1996). The large proportion of charged residues within these LEA proteins contributes to their high hydrophilicity (Battaglia *et al*., 2008). Group 1 proteins have been shown to be involved in the development of endosperm and within higher plants – playing a role in osmotic protection of vegetative organs (Swire-Clark and Marcotle, 1999). The wheat Em proteins and the maize EmB564 and EMB5 proteins belong to this group (Amara *et al*., 2012; Wu *et al*., 2013). Thus far, Group 1 LEA proteins have largely been described as plant specific, however, studies on bacteria (*Bacillus subtilis*) has revealed and reported similar proteins (Stacy *et al.,* 1999). In plants, these group 1 LEA proteins preferentially accumulate during embryonic development, as seen in dry seeds and have also been detected in pollen grains, those being organs that are confronted by water deficit and dehydration (Battaglia *et al*., 2008). Additionally, many of the characterised genes within this group of LEAs are ABA-responsive and also responsive under water-limiting conditions, mainly within embryos, and in a few cases, within the vegetative tissues of young seedlings (Battaglia *et al*., 2008).

Group 2 LEA proteins (Pfam PF00257), also referred to as dehydrins (dehydration induced proteins; Close *et al*., 1989) or the "D-11" family in developing cotton embryos (Dure, 1993; Battaglia *et al*., 2008; Hanin *et al*., 2011), are the most characterised group of LEA proteins. These LEAs are found in a variety of photosynthetically active organisms including cyanobacteria, algae, and lower and higher plants (Battaglia *et al*., 2008). They are also highly hydrophilic in nature, similar to the group one LEA proteins, containing a high proportion of charged, polar amino acids compared to a low fraction of non-polar,

hydrophobic residues (Battaglia *et al*., 2008). These proteins are identified by the presence of at least three distinct sequence motifs named Y, S and K. The K motif has a highly conserved region of 15 amino acids forming amphiphilic α-helix, consensus sequence: EKKGIMDKIKEKLPG, in its carboxyl terminal region (Soulages *et al.*, 2003; Battaglia *et al*., 2008; Hanin *et al*., 2011) and can exist in up to 11 repeats. The Y motif is a short conserved motif (consensus: (V/T)D(E/Q) YGNP) located in the N-terminal region, which can be repeated up to 35 times, and the S motif is a serine rich motif and is found to be a target of phosphorylation in some LEA proteins and is thought to participate in nuclear localization (Jiang and Wang, 2004; Soulages *et al.*, 2003; Hanin *et al*., 2011). The K motif is especially important because it is found in all dehydrins (Hanin *et al*., 2011). Another feature of this group is that they are intrinsically highly unstructured, remaining predominantly unfolded when analysed in aqueous solutions as shown by examples from studies conducted on maize, the resurrection plant *Craterostigma plantagineum, Citrus, Arabidopsis* and cowpea (Ceccardi *et al*., 1994; Lisse *et al*., 1996; Ismail *et al*., 1999; Hara *et al*., 2001; Mouillon *et al*., 2006). Selected dehydrins have been found to be preferentially induced under specific stresses, whereas many others are expressed constitutively (Soulages *et al.,* 2003; Hanin *et al*., 2011). Despite the many predicted roles of the group 2 LEA proteins, there have only been few *in vitro* functional analyses report and therefore the molecular mechanisms through which they can enhance stress tolerance, remains largely unknown.

Group 3 LEA proteins (eg: D-7) (Pfam PF02987) are those characterised by a repeating 11 amino acid sequence motif (TAQAAKEKAGE), repeated up to 13 times within the polypeptide, resulting in an amphipathic α-helical structure (Dure, 1993a; Hong-Bo *et al*., 2005; Battaglia *et al*., 2008; Amara *et al*., 2014). The variability in the 11-mer motif leads to a sub-classification of the group 3 LEA proteins into two subgroups: 3A, represented by the cotton D-7 LEA protein; and 3B, represented by the cotton D-29 LEA protein. The first subgroup is classified as highly conserved, whereas the other subgroup (3B) is more heterogeneous (Amara *et al*., 2014). Group 3 LEA proteins have been shown to be involved in the dehydration process of higher plants, functioning in sequestering ions (Hong-Bo *et al*., 2005). Examples of these include barley PMA1949, carrot Dc8 and soybean pGmPM2 proteins (Hong-Bo *et al*., 2005). Group 3 LEA proteins also remain unfolded in the hydrated state, however, structure-prediction computer based programmes predict that they adopt a high degree of folding (Ginbot, PhD thesis, 2011). Found in a wide range of higher plants, this group is also reported in non-plant anhydrobiotic organisms including crustaceans,

insects and nematodes (Ginbot, PhD thesis, 2011). The expression of the group 3 LEA proteins within these organisms was correlated with desiccation tolerance.

The group 4 LEA proteins (Pfam PF03760), including cotton D-113 and LEA 14, and *Craterostigma* PGC27-45, are of widespread occurrence in both higher plants (the vascular gymnosperms and angiosperms) and non-vascular plants (bryophytes) and were originally found highly accumulated in dry embryos (Dure, 1993; Battaglia *et al*., 2008; Amara *et al*., 2014). These proteins are devoid of repeated sequences motifs and are conserved in their N-terminal portion, which constitutes approximately 70 to 80 residues, predicted to form α-helical structures that are amphipathic (Hong-Bo *et al*., 2005; Battaglia *et al*., 2008). The less conserved C-terminal protein portion remains variable in size (Dure, 1993b). Four additional motifs can be distinguished in many group 4 LEA proteins. The presence or absence of these motifs within these proteins, allowed the subgrouping of this group. The first subgroup (group 4A) consists of small proteins that range from 80 – 124 amino acid residues long. The other subgroup (group 4B) has longer representatives that may range from 108 - 180 residues (Amara *et al*., 2014). In aqueous solutions, these proteins also remain disordered. Subsidiary structures that are able to be adaptive to conformational changes of other proteins, may also be formed by these proteins and may function in ultimately protecting the stability of membranes during times of desiccation (Hong-Bo *et al*., 2005).

Very little information is known regarding the group 5 and group 6 LEA proteins compared with the first four groups. The first proteins described for the Group 5 LEAs were D-34, D-73 and D-95 from cotton. They represent an atypical LEA group due to their higher proportion of hydrophobic amino acid residues within the peptides. These proteins have proven to be insoluble after boiling, suggesting they may adopt a globular conformation (Amara *et al*., 2014). Wise (1993) claims, substantiated by the use of bioinformatics tools, that the group 5 LEA proteins are a misclassified group and in fact share more similarity with group 3 LEA proteins, primarily on the basis of their shared sequence domains. Battaglia *et al*., (2008) suggested the designation of subgroups – 5a (D-34), 5b (D-73), 5c (D-95) - according to their sequence similarity and because D-34, D-73 and D-95 were the first proteins classified for this group. The first described protein in the group 6 LEA proteins was PvLEA18 from bean (*Phaseolus vulgaris*). To date, 36 members of this family have been identified from differing species of vascular plants (Battaglia *et al.,* 2008; Amara *et al*., 2014). Gene expression studies highlighting transcript and protein level abundance in PvLEA18, have been conducted and have shown high transcript and protein levels in pollen, seeds and desiccated vegetative

tissues, in response to water deficit (Battaglia *et al.,* 2008). An example from the resurrection plant family, *X. humilis* has shown excessive up-regulation of the LEA-6 protein, a seed maturation protein from *A. thaliana*, in its leaves during dehydration (Illing *et al*., 2005).

1.6 LEA protein structure and function

The classification of LEA protein groups is intimately tied to the structural analysis of these proteins and their possible proposed functions. In order to gain a better understanding of LEA protein function on a molecular level, many attempts have been made to relate the structure to function. The very first structural characterisation of a LEA protein recorded in 1985 (McCubbin *et al*., 1985), showed the stokes radius to far exceed that of similar molecular weight globular proteins, pointing towards a "natively unfolded" and "intrinsically disordered/unstructured" form exhibited by the LEA proteins, taking up no set conformation within the presence of water. This was supported by viscosity studies, suggesting that the protein has an asymmetrical and flexible conformation (Tunnacliffe and Wise, 2007). Structural modelling and structural prediction programmes suggest that some LEA proteins from particular families contained defined conformations (Amara *et al*., 2014). However, the common structural elements amongst the members of different families indicate that most of these proteins exist as randomly coiled structures when in solution (Battaglia *et al.,* 2008). The intrinsically disordered nature of these proteins is also supported by experimental studies that have shown that all hydrophilic LEAs studied have revealed a high degree of unordered structure in solution (Battaglia *et al.,* 2008; Amara *et al.,* 2014). The lack of structure when in the hydrated state, may confer advantages to these LEA proteins including: a larger interaction surface than globular proteins of a similar length, more conformational flexibility and the ability to scaffold and to interact with multiple proteins and lastly, that the regulation of their function and stability is facilitated by a diverse set of post-translational modifications (Babu *et al*., 2011). If we consider the highly hydrophilic nature and the disordered structure of these LEA proteins and the fact that they are grouped mainly by the presence of conserved sequence motifs, it is therefore likely to assume that their function may be closely related to their high amino acid affinity for water and the hydrated state, and to the recognition of multiple and different macromolecular targets (Battaglia *et al.,* 2008). In a paper submitted by Olvera-Carrillo *et al*., (2011), it was proposed that the "physiochemical properties" of these LEA proteins have potentially served as a driving force for the selection of proteins, that in the face of limiting water availability and the deleterious effects thereof, are able to, and capable of, preserving and maintaining the integrity of cell functioning. It is postulated that

they are able to maintain and preserve the integrity and function of the cellular structures by providing a hydrophilic surrounding when water is limited, substituting for the loss of water molecules (Battaglia *et al*., 2008). The conserved sequences that are innate to these proteins may be responsible for recognising a specific set of target molecules. It is assumed that these hydrophilic proteins mostly recognise their targets under stressful situations, because of their change from unstructured to structured forms, under conditions of water limitation (Battaglia *et al*., 2008).

The unstructured nature of these LEA proteins, at least in the hydrated state, renders these proteins difficult or virtually impossible to crystallise. Therefore, alternative techniques such as Circular Dichroism (CD), Nuclear Magnetic Resonance imaging (NMR) or Fourier Transfer Infrared (FTIR) spectroscopy, have been used to investigate LEA protein structure (Roberts *et al*., 1993; Tolleter *et al*., 2007; Shih *et al*., 2010; Ginbot (2011); Hundertmark *et al*., 2011; Hand *et al*., 2011). Group 1 and Group 4 LEA proteins have been predicted to exist as random coils (Soulages *et al.,* 2002) and may in turn confer stability during times of dehydration, by sharing their hydration shell composed of water, or possibly by acting through their amino acids as water replacement agents. Ginbot (2011), successfully showed the random coil formation of two recombinant Group 1 LEA proteins, XhLEA1-4S1 and XhLEA1-1S2, and the adoption of an alpha-helical structure upon the addition of helix-promoting co-solvent, trifluroethanol (TFE), through the use of CD analysis. Experimental structural analysis using Group 2 LEA proteins from a resurrection plant, cowpea, soybean and mouse-ear cress, described in Battaglia *et al*., (2008), pointed to the largely hydrated and unstructured conformation when in an aqueous solution. It is interesting to note here that it is predicted that the K-segment conserved motif, belonging to these Group 2 LEA proteins, is predicted to form amphipathic alpha-helical structures that are assumed to function in protecting membranes (Battaglia *et al*., 2008). *In silico* predictions of Group 3 LEA proteins, predict the 11-mer domain exists as amphipathic alpha-helices however, CD analysis and IR spectroscopy indicated that they are mostly devoid of secondary structure, remaining in a random coil formation when in solution (Battaglia *et al*., 2008). It has also been shown that when these proteins are in the presence of Suc, glycerol, methanol, ethylene or glycol, or after a process of fast drying, they tend to adopt an alpha-helical conformation. This has also been detected when these proteins are in the presence of TFE or SDS. As aforementioned, the Group 4 LEA proteins remain largely disordered within aqueous solutions, however, they, as with other groups, are able to be induced to take up a more ordered structure, generally

adopting an alpha-helical conformation. Little is known regarding the structure of the Group 5 LEA proteins, whereas the physicochemical characteristics and *in silico* analyses of the Group 6 proteins predict that they too are intrinsically unstructured, like the previous LEA groups.

1.7 This study

The objective of this MSc study was to characterize the "LEAome" of the resurrection plant *X. humilis,* a relatively well characterised monocotyledonous plant endemic to Southern Africa. A set of 35 LEA contigs, containing LEA's from classes 1, 2, 3, 4, 6, 7 and 8, was initially derived from the mRNA of dehydrated leaves and roots of this resurrection plant (Collett *et al*., 2004). A total of 21 LEA contigs, 19 contigs identified from the list of 35 above, and the two Group 1 LEA proteins identified in a study conducted by Ngubane (2008) and revised by Ginbot (2011), termed XhLEA1-4S1 and XhLEA1-1S2, were chosen for the purpose of this study. Changes in the transcript abundance of the cDNAs of these 21 LEA contigs, during drying and rehydration of leaf and root tissues of *X. humilis,* were investigated. After real time expression analysis, three LEA contigs of the set of 21 were chosen for further characterization. These three full length LEA cDNA transcripts, 2 putative Group 2's and one putative Group 3, were subsequently cloned into a bacterial expression vector and expressed using *E. coli.* Recombinant protein expression and purification of these three proteins were conducted. The in-solution preliminary structure of these proteins was investigated by circular dichroism (CD) and compared with results obtained by bioinformatics analysis tools (PONDR). A general discussion on the findings of this work, together with recommendations for future studies of this set of 21 LEA contigs or other similar proteins, will be presented as a final chapter.

Given that many studies postulate that LEA proteins may play an important role in the desiccation tolerance phenomenon within resurrection plants (as discussed), we have used the resurrection plant *X. humilis* here as a model organism. This study will hopefully shed some light on what is happening at the transcript level of this set of LEA-like genes under dehydration and rehydration stress of vegetative tissues (depicted in Figure 1.1). We can then postulate, after analysing gene expression profiles, whether these transcripts accumulate in response to dehydration stress and whether they may ultimately aid the survival of the plant under stress-induced conditions. It will also be interesting to assess whether, if after the plant is rehydrated, there is a marked down regulation of these LEA-like genes. This may indicate

that the plant no longer has a need for these transcripts, as the stress response is alleviated upon rehydration. The advantage of studying resurrection plants as a model for understanding desiccation tolerance, is two-fold. The first advantage is the overlapping of developmental processes inherent in seeds that minimizes complications within experimental data, and secondly, desiccation tolerance can be studied in a more physiologically and morphologically complex system - that of the whole photosynthetic plant, in comparison to seeds and lower plants (Leprince and Buitink, 2010). The main achievement of molecular studies with seeds has been the identification and characterization of the LEA proteins. LEA-protein mRNAs first appear at the onset of desiccation, dominate the mRNA population in dehydrated tissues and their levels gradually fall several hours after embryos begin to imbibe water. The question is: is the same pattern seen in our resurrection plant species?

Figure 1.1 *X. humilis* plants: Fully hydrated, fully dehydrated and rehydrated, respectively (reproduced with permission from Professor Jill Farrant).

Chapter 2

Bioinformatics analysis of the 21 target LEA genes.

2.1 Introduction

Bioinformatics tools are largely used today to gather useful information about biological data sets. This has become a crucial part of many areas in the science world, enabling the analysis of gene and protein expression and regulation and the comparison of genetic and genomic data, increasing our understanding of evolutionary aspects of molecular biology. In structural biology, it aids in the simulation and modelling of DNA, RNA and protein structures, as well as molecular interactions between these three and other components. For the purpose of this study, it was vital to analyse the given 21 LEA-like sequences, both at the nucleotide and amino acid levels, for homology and possible conserved domains, as 19 of these proteins were largely described as being "similar to LEA mRNA sequences" (Collett *et al*., 2004). Characteristics of LEA groups described by Hong-Bo *et al*., (2005), Tunnacliffe and Wise (2007), Battaglia *et al*., (2008) and Bies-Ethéve *et al*., (2008), have been used to place the given sequences into final groupings for subsequent use throughout the study.

2.2 Aim

To classify a given set of 21 LEA-like gene sequences according to sequence similarities and the presence of repeated conserved domains, and to provide revised characterisations and nomenclatures, confirming or validating the previous ones.

2.3 Origin of the 21 LEA genes under investigation

19 full length *X. humilis* cDNA clones (Collett *et al*., 2004), with putative similarity to LEA mRNA sequences, and two Group 1 LEA gene sequences (XhLEA1-4S1 and XhLEA1-1S2), which have been previously characterised by Ngubane (2008) and Ginbot (2011), were chosen for my research (Table 2.1). The 19 full length LEA-like gene sequences were originally grouped according to INTERPRO (http://www.ebi.ac.uk/interpro/) domains and Panther (http://www.pantherdb.org/). The two group 1 LEAs, XhLEA1-4S1 and XhLEA1-1S2, were identified and characterised in *X. humilis,* using degenerate oligonucleotide primers and cloning, and identified from a cDNA library of the same plant and annotated as seed specific, respectively. Whereas the number '1' following the XhLEA name indicates that these LEAs belong to Group 1, the next numbers ('1' and '4') indicate the number of times the specific Group 1 LEA conserved motif – GGQTRREQLGEEGYSQMGRK-

appears within the polypeptide sequence (Ginbot, 2011). The 'S' stands for 'small hydrophilic plant seed protein signature' (G-[EQ]-T-V-V-P-G-G-T), which is repeated once in XhLEA1-4S1 and twice in XhLEA1-1S2

Table 2.1. 21 Clone ID names and their corresponding Group/Class names described by Ngubane (2008) and Ginbot (2011) for the first two LEA genes (XhLEA1-4S1 and XhLEA1-1S2) and Collett *et al*., (2004) for the remaining 19 genes. To note: the first two Clone IDs were classified and named by Ngubane (MSc thesis, 2008) and renamed by Ginbot (PhD thesis, 2011). Xh refers to *Xerophyta humilis*; D indicates that the genes were present in libraries from desiccation tissues; L indicates their presence in leaf libraries and R, their presence in root libraries.

Clone ID :	Group/Class
XhLEA1-4S1	Group 1 LEA (Ginbot, 2011)
XhLEA1-1S2 (XHC00797a):	Group 1 LEA (Ginbot, 2011)
Xh_LD_39D01 / Xh_RD_27G12 (both identical)	Dhn9
Xh_RD_19A06	Dehydrin 7
Xh_LD_05G02 / Xh_RD_14C02 (both identical)	Dhn9
Xh_RD_30C12	dehydrin [Triticum aestivum]
Xh_RD_19H04	dhn9
Xh_LD_27A05 / Xh_RD_27D09 (both identical)	putative dehydrin [Xerophyta viscosa]
Xh_LD_51C01 / Xh_RD_08B08 (both identical)	CLASS 3
Xh_LD_12E05	CLASS 3
Xh_LD_38C09	GROUP/CLASS 3
Xh_LD_08G09 / NO RD	late embryogenesis abundant domain-containing protein CLASS 3
Xh_LD_44B08 / Xh_RD_28B12r (both identical)	lea-like protein CLASS 3
Xh_LD_10A10	group 3 late embryogenesis abundant protein
Xh_LD_12C12	Group 3 late embryogenesis
Xh_LD_39A02 / Xh_RD_29F06 (both identical)	late embryogenesis abundant (seed maturation protein) CLASS 4
Xh_RD_25D02	seed maturation protein CLASS 4
Xh_RD_27D08	late embryogenesis abundant protein d- CLASS 6
Xh_LD_08D06 / Xh_RD_25B08 (both identical)	late embryogenesis-abundant protein CLASS 7
Xh_LD_09F09 / Xh_RD_16B12 (both identical)	late embryogenesis-abundant protein CLASS 8
Xh_RD_27E09	late embryogenesis-abundant protein CLASS 8

2.4 Methods and materials

2.4.1 Analysis of amino acid sequences for conserved motifs characteristic of specific LEA groups

With the prior knowledge that both Group 1 LEA genes included in this study contained the specific Group 1 LEA conserved motif and the small hydrophilic plant seed protein signature (described in Section 2.3), the remaining 19 LEA-like proteins were investigated for the presence of any group-specific-domains or conserved repeated motifs, such as those described for each LEA group in Section 1.5 of the previous chapter.

2.4.2 Bioinformatics analysis of the amino acid sequences of each of the 21 LEA-like proteins

The deduced amino acid sequence and open reading frame (ORF) of each of the 21 LEA-like nucleotide sequences were obtained by translation of the cDNA sequence using the online ExPASy translate tool (http://web.expasy.org/translate/), and the ORFs of these proteins were confirmed by ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/). The ProtoParam tool (http://web.expasy.org/protparam/) was used to predict physiochemical properties of each of the 21 LEA-like proteins such as: molecular weight (kDa), pI, estimated half-life, as well as the stability properties of the proteins. The solubility of these LEA-like proteins was also calculated using the online Recombinant Protein Solubility Prediction tool (School of Chemical Engineering and Material Science, University of Oklahoma, http://biotech.ou.edu/).

2.4.3 Protein BLAST analysis of the 21 LEA-like proteins

Amino acid homology searches were carried out using the protein BLAST algorithms and National Centre for Biotechnology Information (NCBI) databases. Conserved domains or super-families that identified with each protein and the general homology across proteins in different species, were reported.

2.5 Results and discussion

2.5.1 Bioinformatics analysis of the amino acid sequences of each of the 21 LEA-like proteins

Results for the bioinformatics analysis of the cDNA fragments of each LEA-like gene are shown in Table 2.2. The longest ORF length, determined by the ORF Finder platform, and the predicted molecular weights (in kDa), determined by the ExPASy translate tool, were successfully reported for each of the 21 LEA-like gene and corresponding protein sequences.

The ProtoParam tool predicted the individual theoretical pI's of each of the LEA-like proteins, indicating the pH at which each protein carries no overall electrical charge. The theoretical pI value may affect the solubility of a protein at a particular pH. The half-life of the individual proteins were predicted in hours within *E. coli,,* and is an indication of the time it takes for half of the amount of protein within the cell to disappear post-synthesis. The instability index of each of the proteins reported indicated whether or not these proteins would be stable or not *in vitro*, with values of <40 indicating protein stability. This information regarding stability may be useful in subsequent protein expression studies and in the handling of proteins.

Table 2.2 Bioinformatics analysis of the 21 LEA-like genes with respect to their nucleotide and corresponding amino acid sequences, using ExPASy translate and ProtoParam tool.

Clone ID	Longest ORF (bp)	Number of amino acids encoded into a protein	Theoretical molecular weight (kDa)	Theoretical pI	Half-life (hours) in E. coli	Instability index	Solubility prediction when over-expressed in E. coli
XhLEA1-4S1	459	152	16,5	6,02	>10	51,57 (unstable)	0.0%
XhLEA1-1S2 (XHC00797a):	411	136	13,7	5,11	>10	52,68 (unstable)	0.0%
Xh_LD_39D01 / Xh_RD_27G12 (both identical)	453	150	15,9	8,13	>10	40,04 (unstable)	0.0%
Xh_RD_19A06	411	162	16,8	9,18	>10	31,27 (unstable)	0.0%
Xh_LD_05G02 / Xh_RD_14C02 (both identical)	429	142	14,9	9,82	>10	33,54 (stable)	0.0%
Xh_RD_30C12	420	139	14,8	6,64	>10	21,63 (stable)	0.0%
Xh_RD_19H04	417	138	14,4	9,6	>10	31,80 (stable)	0.0%
Xh_LD_27A05 / Xh_RD_27D09 (both identical)	306	101	11,2	6,97	>10	37,63 (stable)	0.0%
Xh_LD_51C01 / Xh_RD_08B08 (both identical)	447	148	15,6	9,06	>10	17,58 (stable)	100%
Xh_LD_12E05	477	158	16,8	7,79	>10	27,95 (stable)	0.0%
Xh_LD38C09	567	188	19,8	6,61	>10	31,57 (stable)	0.0%
Xh_LD_08G09 / NO RD	705	234	25,3	6,88	>10	32,66 (stable)	0.0%
Xh_LD_44B08 / Xh_RD_28B12r (both identical)	597	198	20,6	5,21	>10	31,25 (stable)	0.0%
Xh_LD_10A10	558	188	19,8	6,67	>10	31,85 (stable)	0.0%
Xh_LD_12C12	471	159	17,1	8,59	>10	29,32 (stable)	100%
Xh_LD_39A02 / Xh_RD_29F06 (both identical)	390	129	14	9,90	>10	40,44 (unstable)	0.0%

Xh_RD_25D02	339	112	11,3	9,34	>10	13,99 (stable)	0.0%
Xh_RD_27D08	372	123	13,2	4,62	>10	35,98 (stable)	0.0%
Xh_LD_08D06 / Xh_RD_25B08 (both identical)	303	100	10,6	9,59	>10	31,56 (stable)	100%
Xh_LD_09F09 / Xh_RD_16B12 (both identical)	456	151	16,6	8,56	>10	18,86 (stable)	100%
Xh_RD_27E09	465	151	16,5	5,57	>10	19,20 (stable)	100%

2.5.2 Protein BLAST analysis and conserved domain identification of the 21 LEA-like proteins

The results from the protein BLAST search, homology and conserved motifs/super-families, are reported in Table 2.3, along with the presence of conserved domains described by Tunnacliffe and Wise (2007), Battaglia *et al*., (2008) and Amara *et al*., (2014) (see Chapter 1, Section 1.5), identified within each LEA-like protein sequence. For the purpose of this study, and for ease of grouping, I ultimately chose to rearrange and revise the LEAs into the groups designated by Battaglia *et al*., (2008).

When analysing the two Group 1 LEA-like proteins, the 20-mer conserved domain, GGQTRREQLGEEGYSQMGRK (Tunnacliffe and Wise, 2007), appeared four times in XhLEA1-4S1, and once in XhLEA1-1S2, with some single amino acid deviations. The small hydrophilic plant seed protein signature- GETVVPGGT- appeared once in XhLEA1-4S1, and twice in XhLEA1-1S2, as previously described when introducing the two Group 1 LEA-like proteins (Section 2.3). General protein homology of both these LEA-like proteins was with Em-Like proteins, as predicted by Ngubane (2008) using multiple sequence alignment and reported by Ginbot (2011).

Of the three segments characteristic of Group 2 LEA proteins- Y, S and K- one, two or all of the motifs were recorded within the six Group 2 LEA-like proteins (with some single amino acid deviations) and half of them contained the motif considered to be conserved in the dehydrin superfamily (EKKGIMDKIKEKLPG). All of the LEA-like proteins described to be part the dehydrin family had homology with dehydrin proteins or putative/predicted dehydrins (in the case where no specific hits and superfamilies were detected) and in some cases RAB18 (proteins responsive to abscisic acid, 18).

The bioinformatics analysis of the LEA-like proteins falling into the Class 3 revealed interesting results, with several 'unusual' conserved domains and superfamilies being recorded. The conserved motif, PTZ00121 (NCBI CDD conserved protein domain family), was recorded for four out of the seven Group 3 LEA-like proteins. This domain is not conserved to any superfamily, and is classified as a model that may span more than one domain (http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=173412). Another interesting domain that was reportedly sharing homology with the putative Group 3 LEA protein Xh_LD_38C09 was the DUF883 domain (Domain of Unknown Function 883). This family consists of several hypothetical bacterial proteins of unknown function. The Xh_LD_08G09 LEA-like protein shared homology with three domains: the Tim44 superfamily (PF04280), PTZ00121 and Apolipoprotein. The Tim44 superfamily contains Tim44-like domains (Inner Membrane Translocase protein), deemed to be an essential component of the machinery that mediates the translocation of nuclear-encoded proteins across the mitochondrial inner membrane. The Apolipoprotein domain (PFAM01442), also recorded for this putative Group 3 LEA protein, represents a group of proteins that contain several 22 residue repeats which form a pair of alpha helices. This domain is also classified as a model that may span more than one domain and is not assigned to any domain superfamily. Two of the seven Class 3 LEA-like proteins contained a C-terminal conserved motif 1 – GGVLQQTGEQV- described by Battaglia *et al*., (2008) and Xh_LD_12E05 contained another conserved motif (TAQ[A/S]AK[D/E]KT[S/Q]E, with a single amino acid deviation at the end (an A instead of an E) (Battaglia *et al*., 2008). The protein BLAST homology overall revealed that most of these putative Class 3 LEA genes were similar to D-29-like proteins or Group-3-like proteins and will therefore be referred to as Group 3 LEAs.

The LEA_1 domain and superfamily were recorded for both the Class 4 LEA proteins - Xh_LD_39A02/Xh_RD_29F06 and Xh_RD_25D02- consistent with the characteristic PFAM domain of the Group 4 of LEA proteins (summarised by Battaglia *et al*., 2008; Hundertmark and Hincha, 2008; Amara *et al*., 2014). The protein blast for both these proteins revealed homology with the predicted LEA proteins D113 (*Phoenix dactylifera* (Date palm)) and *Elaeis guineensis* (African Oil Palm) and LEA Group 4 proteins (*E. guineensis*), respectively.

The putative Class 6 LEA gene, Xh_LD_27D08, shared homology with a predicted D-34 protein and the SMP (Seed Maturation Protein) superfamily and will therefore now be considered a Group 5a LEA-like protein (as described by Battaglia *et al*., 2008).

As the only putative Class 7 LEA-like protein, Xh_LD_08D06/Xh_RD_25D08 (both identical), this gene shared homology with a LEA-5 like protein and a LEA_2 superfamily, and will be considered as a Group 5b LEA protein (as described by Battaglia *et al*., 2008) throughout the rest of this study.

For the two remaining LEA's belonging to Class 8, sequence homology was shared with the LEA-14-like protein and the LEA_2 superfamily, specific to Group 5c LEAs (described by Battaglia *et al*., 2008). We will therefore include these two putative Class 8 LEAs as Group 5c LEAs. Overall, this analysis revealed a wide range of sequence diversity in and amongst the different groups.

Table 2.3. Protein Blast homology and conserved motifs designated for each LEA-like protein from each LEA Group and conserved domains identified visually. -, denotes that no conserved motifs/superfamilies were identified by the protein BLAST.

Clone ID	NCBI Protein blast homology	Conserved motifs (NCBI)	Conserved domains present in these proteins (Tunnacliffe and Wise, 2007; Battaglia *et al.*, 2008; Amara *et al.*, 2014)
XhLEA1-4S1	Em-Like protein Predicted LEA protein.	LEA_5 (Hundertmark and Hincha, 2008/PFAM) Small hydrophilic plant seed proteins signature G-[EQ]-T-V-V-P-G-G-T	GGQTRREQLGEEGYSQMGRK GETVVPGGT (protein signature)
XhLEA1-1S2 (XHC00797a):	Em-Like protein Predicted LEA protein.	2 LEA_5 domains (Hundertmark and Hincha, 2008/PFAM). Small hydrophilic plant seed proteins signature G-[EQ]-T-V-V-P-G-G-T	GGQTRREQLGEEGYSQMGRK GETVVPGGT protein signature)
Xh_LD_39D01 / Xh_RD_27G12 (both identical)	Dehydrin/LEA Group 2-like protein (RAB18)	Dehydrin superfamily	SSSSSSSS RKKGIKDKIKEKLPG
Xh_RD_19A06	Putative dehydrin, RAB18	-	SSSSSSSSS RKKKVKDKLKEKLPG EKKGMMEKIKEKLPG
Xh_LD_05G02 / Xh_RD_14C02 (both identical)	Dehydrin/ Predicted dehydrin	Dehydrin superfamily	SSSSSS RKKGIKDKIKEKMPG
Xh_RD_30C12	Dehydrin/ Predicted dehydrin	-	KKEGITEKIKEKLPG
Xh_RD_19H04	Dehydrin/predicted dehydrin/ RAB18	-	SSSSSSSSSS RKKGIKEKIKEKLPG EKKGMMEKIKEKLPG
Xh_LD_27A05 / Xh_RD_27D09 (both identical)	Putative dehydrin/dehydrin-like protein	Dehydrin superfamily	VDEYGNP SSSSS RKKGLKDRIMENLPG
Xh_LD_51C01 / Xh_RD_08B08 (both identical)	Homology with D29-like proteins	PTZ00121	-
Xh_LD_12E05	Predicted LEA protein	-	TAQAAKDKTSA
Xh_LD38C09	Predicted Group 3-like protein	DUF883 Superfamily, PTZ00121	GGVLQQAGDMI
Xh_LD_08G09 /	Hypothetical/uncharacterised protein/ desiccation-	Tim44 Superfamily , PTZ00121, Apolipoprotein	-

NO RD	related protein.		
Xh_LD_44B08 / Xh_RD_28B12r (both identical)	LEA-like protein/D-29-like protein.	PTZ00121	-
Xh_LD_10A10	LEA-protein/Group 3 LEA protein	-	GGVLQQAGDKV
Xh_LD_12C12	LEA-like protein/predicted D-29-like.	-	-
Xh_LD_39A02 / Xh_RD_29F06 (both identical)	Predicted LEA D-113 protein/ seed maturation protein-like.	LEA_1 superfamily (Group 4: Bray,1993; Tunnacliffe and Wise, 2007; Battaglia *et al*., 2008)	MQSAREKVKDMMSAAKAK AEEKAEKA
Xh_RD_25D02	LEA group 4/11kDa protein.	LEA_1 superfamily (Group 4, Bray,1993; Tunnacliffe and Wise, 2007; Battaglia *et al*., 2008)	-
Xh_RD_27D08	Predicted D-34 protein (-Dure, 1989).	SMP superfamily (Group 5a, Battaglia *et al*,. 2008; Group 6, Bray, 1993; Tunnacliffe and Wise, 2007)	-
Xh_LD_08D06 / Xh_RD_25B08 (both identical)	LEA-5-like protein	LEA_3 superfamily (Group 5b, Battaglia *et al*., 2008; Group 6, Bies Ethève et al., 2008)	-
Xh_LD_09F09 / Xh_RD_16B12 (both identical)	LEA-14A-like protein	LEA_2 superfamily (Group 5c, Battaglia *et al*., 2008; Group 7, Bies- Ethève et al., 2008).	-
Xh_RD_27E09	LEA-14A -like protein.	LEA_2 superfamily (Group 5c, Battaglia *et al.,* 2008; Group 7, Bies- Ethève et al., 2008).	-

In conclusion, the final table (Table 2.4) represents the grouping of the set of 21 LEA-like gene sequences along with the original groupings characterised by Dure *et al*., (1989), Bray (1993,1994), Illing *et al*., (2005), Battaglia *et al*., (2008), Bies-Ethève *et al*., (2008), Hundertmark and Hincha (2008) and their corresponding Pfam domains, Interpro classifications and general names, and finally, the novel nomenclature chosen for the set of 21 putative LEA proteins.

Table 2.4 A review and compilation of the LEA groups, motifs, corresponding Pfam families, Interpro terms, names and the genes falling into each category based on bioinformatics analysis, protein BLAST (NCBI) and the final novel nomenclature.

Dure *et al.*, 1989	Bray, 1993,1994	Illing *et al.*, 2005	Tunnacliffe and Wise, 2007	Battaglia *et al.*, 2008	Bies-Ethéve *et al.*, 2008	Hundertmark and Hincha, 2008// PFAM	PFAM	Interpro	Name	Our putative LEA proteins falling into groups based on protein blast (NCBI) (Battaglia *et al.*, 2008).	Novel LEA nomenclature
D-19	Group 1	LEA-1	Group 1	**Group 1**	Group 1	LEA_5	LEA_5, Small Hydrophilic Plant Seed Protein; PF00477	IPR000389	Em1, Em6	**XhLEA1-4S1** **XhLEA1-1S2**	**XhLEA1-4S1** **XhLEA1-1S2**
D-11	Group 2	LEA-2	Group 2	**Group 2**	Group 2	Dehydrin	Dehydrin, PF00257	IPR000167	Dehydrin/ Responsive to Absiscic Acid (RAB)	**Xh_LD_39D01/ Xh_RD_27G12** **Xh_RD_19A06** **Xh_LD_05G02 / Xh_RD_14C02** **Xh_RD_30C12** **Xh_RD_19H04** **Xh_LD_27A05 / Xh_RD_27D09**	**XhLEA2-1** **XhLEA2-2** **XhLEA2-3** **XhLEA2-4** **XhLEA2-5** **XhLEA2-6**
D-7 **D-29**	Group 3 Group 5	LEA-3 -	Group 3	**Group 3A** **Group 3B**	3 3* (Battaglia *et al.*, 2008) Classes 6 (Amara *et al.*, 2014)	LEA_4	LEA_4, PF02987 LEA_4, PF02987	IPR004238	ECP63, PAP240,PM 27 D-29	**Xh_LD_51C01 / Xh_RD_08B08** **Xh_LD_12E05** **Xh_LD38C09** **Xh_LD_08G09** **Xh_LD_44B08 /Xh_RD_28B12r** **Xh_LD_10A10** **Xh_LD_12C12**	**XhLEA3-1** **XhLEA3-2** **XhLEA3-3** **XhLEA3-4** **XhLEA3-5** **XhLEA3-6** **XhLEA3-7**
- **D-113**	Group 4 Group 4	LEA-4 LEA-4	Group 4	**Group 4A** **Group 4B**	Group 4	LEA_1 LEA_1	LEA_1;PF03760 LEA_1;PF03760	IPR005513	LE25_LYC ES PAP260,PA P051	**Xh_LD_39A02 / Xh_RD_29F06** **Xh_RD_25D02**	**XhLEA4-1** **XhLEA4-2**
D-34	Group 6	LEA-6	Group 6	**Group 5a**	Group 5	SMP	Seed Maturation Protein (SMP); PF04927	IPR007011	PAP140	**Xh_RD_27D08**	**XhLEA5a**
D-73	-	LEA-7	LEA_5	**Group 5b**	Group 6	LEA_3	LEA_3; PF03242	IPR004926	AtD121, Sag21, Lea5 (Galau et al., 1993)	**Xh_LD_08D06 / Xh_RD_25B08**	**XhLEA5b**
D-95	-	LEA-8	-	**Group 5c**	Group 7	LEA_2	LEA_2; PF03168	IPR004864	LEA14 (Galau et al., 1993)	**Xh_LD_09F09 / Xh_RD_16B12** **Xh_RD_27E09.**	**XhLEA5c-1** **XhLEA5c-2**
-	-		-	**Group 6**	Group 8	PvLEA18	LEA_6; PF10714		LEA18		

2.6 Conclusion

In conclusion, *in silico* analysis is a necessity when determining whether given genes used in a study have characteristic or unique sequences and physiochemical properties associated with them at the DNA and protein level, respectively, as this will better inform our experimental approach. Previously, LEA genes were classified according to overall sequence similarity, however, more recently, by identifying conserved domains. It was decided to combine these two approaches to ultimately produce a revised classification of the original set of 21 LEA-like genes, according to the simplified nomenclature and characteristics described by Battaglia *et al*., (2008). Our terminology derived from this is indicated in the final column of Table 2.4, as the novel nomenclature to be used throughout the study. Revising the nomenclature of this set of 21 LEA-like genes will allow us to investigate them with respect to general LEA group characteristics, both biochemical and structural, summarised by Hong-Bo *et al*., (2005), Tunnacliffe and Wise (2007), Battaglia *et al*., (2008), Ginbot (2011) and Amara *et al*., (2014).

The next chapter will focus on the changes in the gene expression patterns and mRNA transcript abundance of this grouped set of 21 LEA-like genes, in the resurrection plant *X. humilis,* when plants are exposed to dehydration stress and rehydration treatments.

Chapter 3

Gene expression analysis of the set of 21 LEA-like genes in leaves and roots of *X. humilis*

3.1 Introduction

Plants are able to cope with a range of abiotic stresses, including water deficit stress, by means of co-ordinating and regulating their gene expression. When studying plant responses towards stresses, genomic and proteomic techniques where cDNA or genomic libraries are generated from target tissues and screened for differentially expressed genes in response to a given stress, are frequently used. The expression of specific target genes may be tested at the mRNA level by techniques such as Northern blotting, quantitative Real-time Polymerase Chain Reaction (qPCR) and nuclear run-on and/or at the protein level by SDS-PAGE analysis and western blotting using antibodies specific to the target proteins or recombinant proteins. In this chapter, the changes in expression during dehydration and rehydration of 21 LEA-like genes, previously identified from cDNA libraries generated from leaf and root tissues of *X. humilis* (Collett *et al*., 2004; Ngubane, 2008), were analysed by qPCR.

To date, virtually all the research that has been reported on resurrection plants has been conducted on leaf tissues (Sherwin and Farrant, 1998; Farrant, 2000; Mowla *et al*., 2002; Balsamo *et al*., 2005; Farrant *et al*., 2007; Ingle *et al.,* 2007; Moore *et al*., 2007), while very little research has been reported on the root tissues of these plants (Farrant, 2007). This is largely due to the recalcitrant nature of root tissues of such species, and the problems associated with downstream standard molecular protocols used in transcriptome and proteome studies (Waisel, 2001; Mehta *et al*., 2008; Kamies, 2011). It is important to remember that when unearthing the roots of any plant that may be deeply buried within soil for observation, treatment or analysis, disturbances may occur and thus careful consideration needs to be practised. The removal of roots from increasingly dry soil, results in tissue breakage and thus, the induction of wounding responses. Soil particles cannot be removed by washing with water, as this would alter the water content of these roots at the time measurements were taken. This, in turn, would result in microbial contamination within the roots sampled causing inaccurate downstream analysis and incorrect recording of experimental results.

In this study, it was deemed important to understand the expression of LEAs in both roots and leaves, in order to gain insight into the tissue specificity of LEAs, and their overall role in

the desiccation tolerance response within whole *X. humilis* plants. Care was taken to minimize any form of root damage or soil contamination during the dehydration stress sampling experiment, and several protocols were utilised in an attempt to quantitatively measure changes in gene expression in root tissues.

To my knowledge, there have been no previous studies measuring the changes in gene expression across this specified set of the 19 LEA-like transcripts identified by Collett *et al*., (2004), excluding the expression patterns on leaves and roots for the two Group 1 LEAs that have been studied by Ngubane (2008) and Ginbot (2011). This analysis will therefore expose the relationship between LEA-like transcript abundance across a dehydration and rehydration series. Furthermore, this will serve as a reference point for future LEA studies and will help contextualise the parallel physiological and biochemical changes occurring during plant dehydration in a resurrection plant.

This section of the study encompasses the respective dehydration and rehydration treatments of **whole** *X. humilis* plants, followed by RNA extraction. For dehydration and rehydration treatments, the individual mRNA expression levels for each of the 21 LEA-like genes was assessed using qPCR, conducted by the Centre for Proteomic and Genomic Research (CPGR).

3.2 Aim

The aim of this chapter was to attempt to analyse the changes in transcript abundance and hence mRNA expression, in **both** leaves and roots of *X. humilis*, during dehydration and subsequent rehydration.

3.3 Methods and materials

3.3.1 Plant material

The *X. humilis* plants used in this study, were previously collected from Borakalalo National Park and Mpumulanga, Northwest Province, South Africa. These plants were potted, grown and maintained under greenhouse conditions previously described by Sherwin & Farrant (1996). Five weeks prior to stress treatments, plants were transferred to controlled growth chambers, otherwise referred to as convirons (Conviron Adaptis A350, Canada) and allowed to acclimatize to controlled environmental growth conditions of 60% relative humidity, light intensity of 150-200 µmol/m^{-2}/s^{-1}, and a photoperiod of 16 hour light and 8 hour dark, which

was maintained throughout the dehydration, and subsequent rehydration periods. The light phase began at a temperature of 15°C reaching a maximum temperature of 21°C after 30 minutes, increasing 2°C every 10 minutes. In contrast, the dark phase began with a temperature of 21°C and decreased to 15°C, as per above. During acclimation, for a period of one month, the plants were watered once a week, with due precaution taken to avoid over-watering. During this period of acclimatization, the senescent (dead) material at the tips of the leaves was carefully removed, and the leaves were allowed to restore to their healthy state prior to stress treatments. Due to limited convirons available for experimentation, dehydration and rehydration stress experiments were carried out on four of the healthiest trays of *X. humilis (*each comprising approximately 15-20 whole *X. humilis* plants), representing four biological replicates (adult, reproductive). Care was taken to choose trays with *X. humilis* plants of similar size and phenotype, as to minimize biological variation during the stress treatments. This entire treatment was repeated four times, each with a different set of plants.

3.3.2 Dehydration stress treatment of *X. humilis* plants

Plants were dried down by withholding water and rehydrated two weeks after the dehydration profile was completed. The dehydration process was monitored by determining the Relative Water Content (RWC) of leaf and root samples. Tissue samples were harvested from these plants prior to dehydration treatment (~100% RWC) representing time-zero, the state of the fully hydrated plant, at four points during dehydration (70-80%, 40-50%, 20-30% and <10% RWC) and three points post rehydration (representing approximately 2-10%, 50-60% and 90-100% RWC, respectively). Sampling was conducted at 10h00am every 'sample morning' to circumvent the potential differences in transcript and/or protein abundance caused by circadian or light-dark regulation (Ingle *et al*., 2007). At each sampling point, whole *X. humilis* plants were removed from trays and all leaf and root material was used. Three leaves and two roots, per sampling point, per tray, were set aside for RWC measurements. The remaining leaves and roots were pooled and flash frozen in liquid nitrogen and stored at -80°C for further downstream methodologies and analyses.

3.3.3 Determination of RWC

Due to the small size of leaf and root tissues, whole leaves and roots were used for RWC determination. The absolute water content of plants at full turgor (AWCft at 100% RWC) was determined at the very beginning of the treatment, after thoroughly watering the

biological replicates at dusk and harvesting tissues the following day at dawn. This measurement was subsequently used for the calculation of RWC (see below), the values thus obtained, being expressed relative to this full turgor condition. Absolute water contents, for each time of tissue sampling, were gravimetrically determined by oven drying at 70°C for 48 hours. Fresh and dry weight (FW and DW, respectively), were measured and AWC values were calculated using the following formula:

AWC= (FW-DW)/DW

AWCn (sample absolute water content) was calculated for each leaf and each root. For each plant, three AWC values were selected for the leaves and two for the roots, and the average of these values was recorded. RWC was calculated for each sampling point using the following formula (Farrant *et al*., 2000):

% RWC = (AWCn/AWCft) x 100

3.3.4 RNA extraction

3.3.4.1 RNA extraction from leaf tissues

All plastics, pestles and mortars used, were autoclaved twice. Five leaves per biological replicate plant, per sampling point, were used for subsequent RNA extractions. The leaves were ground into a fine powder using a pestle and mortar, in the presence of liquid nitrogen. In order to prevent degradation of RNA, the ground tissue was maintained on ice. The RNA isolation was performed using the RNeasy® Plant Mini Kit (QIAGEN, #74904), according to the manufacturer's instructions with the following modification included at the beginning of the protocol: after the buffer of choice (Buffer RLT) had been added to the ground tissue for homogenization, the sample was centrifuged at maximum speed (14, 000 x g) using a bench top centrifuge for 5 minutes, and only the supernatant was used for subsequent steps in the protocol.

3.3.4.2 DNase treatment and phenol:chloroform:iso-amyl alcohol clean-up of the isolated RNA from leaf tissue

Each RNA sample obtained in Section 3.3.4.1, was treated with deoxyribonuclease I (DNase I, Fermentas USA) to remove any genomic contamination. One microgram of RNA was treated with DNase I and the reaction was set up as per the manufacturer's guidelines. The samples were incubated at 37°C for 1 hour in a heating block (HB/01 digitally controlled heating block, OMEG Scientific, South Africa). An equal volume of phenol:chloroform:iso-amyl alcohol (ratio of 25:24:1) was then added to the reaction, mixed by inversion and

centrifuged at 14 000 x g at 4°C for 10 min. The upper aqueous was then carefully removed and placed in a sterile Eppendorf tube and a 0.1 volume of 3M sodium acetate (pH 5.2) was added to this and gently mixed by inversion. To this, a 5 x volume of 100% EtOH was added, mixed by inversion and subsequently centrifuged at 14 000 x g at 4°C for 10 min to precipitate the RNA. The supernatant was carefully discarded and the pellet was allowed to air-dry for 2 min. The pellet was then re-suspended in 20 µl of RNase-free water.

3.3.4.3 RNA extraction from root tissues

Due to the fibrous nature of the root material, an RNA extraction method needed to be optimised for this specific tissue, with careful precautionary steps. A protocol combining the use of TRI-Reagent®, and the DNAse treatment step, followed by a phenol:chloroform:iso-amyl alcohol clean-up of the isolated RNA, was implemented after trials with various combinations of extraction methods, including the methodology applied for leaf tissues in Section 3.3.4.1. Five roots were used per biological replicate plant, per sampling point. The root material was ground and maintained on ice prior to RNA extraction, as conducted for leaf tissues. 50-100mg of root tissue was homogenized in 1ml TRI Reagent®, using a vortex at maximum speed for 10 minutes. The homogenate was then stored for 5 minutes at room temperature, to permit the complete dissociation of nucleoprotein complexes. This storage step was followed by the supplementation with 200µl chloroform per 1ml TRI Reagent®. The samples were covered tightly and vortexed vigorously for 15-20 seconds. The resulting mixture was stored at room temperature for 10 minutes, after which, it was placed into a 4°C bench-top centrifuge at 12,000 x g for 15 minutes. Following centrifugation the mixture separated into a lower red phenol-chloroform phase, an interphase and a colourless upper aqueous phase where the isolated DNA, proteins and RNA respectively, were concentrated. The aqueous phase, containing exclusively the RNA, was transferred to a fresh Eppendorf and the interphase and organic phase were discarded. The RNA was precipitated with 500µl cold isopropanol and the samples were stored at room temperature for 10 minutes after which were placed into a 4°C bench-top centrifuge at 12,000 g for 8 minutes. An RNA wash step was then performed by removing the supernatant from the previous step and adding 1ml of cold 75% ethanol to the white pellet that had formed. The sample was first vortexed sufficiently for 15 seconds and then centrifuged at 7,500 – 12,000 g for 5 minutes at 4°C. The ethanol wash was then removed and the pellet was briefly air-dried for 3-5 minutes. The dry pellet was then re-suspended in 89µl RNase-free water (Qiagen, RNeasy Plant Mini kit) by pipetting and incubated at 55-60°C for 10 minutes.

3.3.4.4 DNase treatment and phenol: chloroform: iso-amyl alcohol clean-up of RNA extracted from root tissues

Each sample of RNA was treated with deoxyribonuclease I to remove any genomic contamination that would affect downstream qPCR applications as per previous protocols (Section 3.3.4.2) with the following modifications: 10µl 10X Reaction buffer with $MgCl_2$ and 10µl DNase 1, RNase-free was added to the 89 µl of RNA sample. Post incubation at 37ºC for 1 hour, 10µl of EDTA was added and the final RNA pellet was re-suspended in 30µl RNase-free water.

3.3.5 cDNA synthesis

The purified RNA obtained in Section 3.3.4, from both leaf and root tissues, was used for subsequent cDNA synthesis (Appendix A2).

3.3.6 qPCR primer design

The primer sets used for the analysis of gene expression changes using qPCR, were the forward and reverse primer sets per individual gene (Appendix A3, Table A3), and designed using combination of DNAMAN (Lynnon Corporation Bioinformatic Solutions) and Primer3 Plus. Care was taken to ensure that the primer sets for one LEA-like gene did not bind to another set for the list of 21 LEA-like genes. Primer amplification specificity and primer dimer formation was analysed by end point PCR (data not shown). These primer sets were designed to have annealing temperatures of approximately 60ºC and thus yield PCR products specific for each gene. 18S RNA and EF1a were included as the reference genes and primer pairs are represented along with those for the set of 21 LEA-like genes. Both reference genes yield a PCR product with an approximate size of 130bp. Primer efficiency (E), slope and R^2 value for each primer pair, calculated using qbase software (www.biogazelle.com) (Appendix A4, Table A4).

3.3.7 qPCR gene expression analysis

qPCR gene expression analysis studies were conducted on 64 RNA samples, extracted from plants displaying varying RWC's, analysed against 23 genes (21 targeted LEA-like genes and 2 reference genes) using the ABI 7900HT platform at the CPGR. The 64 RNA samples consisted of two sets of 32 samples, one set of *X. humilis* leaf RNA samples and the other set of *X. humilis* root RNA samples, isolated across 5 dehydration time points, and 3 rehydration

time points, of four biological replicate *X. humilis* plants. See appendix for methods and materials (Appendix A5).

The gene expression analysis using leaf cDNA samples was successful. However, root cDNA samples could not be successfully analysed using this platform. The individual LEA primer sets gave errors within the cycle threshold (Ct) technical replicate values. The reference genes utilised were successful in leaf tissue analyses, but gave inconclusive data in root studies (data not shown). It was thus decided to continue with the analysis of gene expression changes, using the methodologies described below, in leaf tissues only, as data points obtained were sufficient for downstream analysis. Alternative methods for root analysis were explored and attempted (see Section 3.4 below).

3.3.8 Standardization of real-time PCR gene expression data from leaf tissues from independent biological replicates

qPCR data obtained was standardized according to the simple, robust and powerful method developed by Willems *et al*., (2008). This method was created for a set of biological replicates that shows substantial variation between the replicates, even though a similar expression trend is detected. The data was analysed with qBase software (www.biogazelle.com) and further data mined with Excel 2013. The qPCR data set for each gene was log transformed, mean-centered and autoscaled, to ensure the cancelling out of high inter-experimental variation and for the correct statistical analysis to be used to assess significance of observed differences for the data obtained. The log transformation step in this method reduced the effect of outliers and attributed equal weight to all data points. Because log transformation did not correct any of the experimental differences observed between the biological replicates, the mean-centering step equalized the experimental averages and the autoscaling step equalized experimental standard deviations. An additional multiplication step by the mean exp. standard deviation, was included to generate fold changes that reflect those original experimental ones. The maximum allowed variability on technical replicates was set at 0.65. Relative quantification of the expression was achieved using the Pfaffl equation (Pfaffl, 2001) (Figure 3.1), and reaction efficiency E was calculated according to the equation E=10^[-1/slope] (Appendix A4, Table A4).

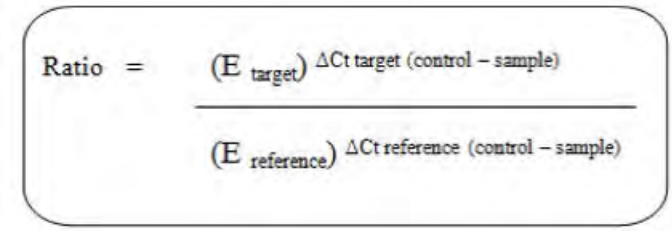


Figure 3.1. Equation representing the Pfaffl mathematical equation for calculating relative expression ratios for real-time PCR. The relative expression ratio (R) of a target gene expressed in a sample versus a calibrator sample (control) in comparison to a reference gene (reference). E_{target} and $E_{reference}$ represent the real-time PCR efficiencies of the target and reference gene transcripts, respectively. ΔCt target and ΔCt reference represents the deviations in the Ct values of the target gene in the control and sample and the reference gene in the control and sample, respectively. Real-time PCR efficiencies were calculated according to E=10^[-1/slope] (Pfaffl, 2001).

3.3.9 Statistical analysis

All statistical tests and graphs were performed and generated using GraphPad Prism software version 6 for Windows (GraphPad Software Incorporation, 1992-2007). One-way ANOVA with Dunnett's post-test was performed for dehydration-stress treatment analysis. All dehydration and rehydration RWCs for leaf tissues were compared to time-point zero (100% RWC).

3.4 Droplet Digital™ PCR (ddPCR™) protocol for root samples

Due to the failure of real-time gene expression analysis of root RNA samples, it was decided to attempt to quantify gene expression using ddPCR™. This method was developed to enable the absolute quantification of nucleic acid target sequences with high precision. The absolute quantity of target molecules is measured by counting nucleic acid molecules that are encapsulated in discrete water-in-oil droplet partitions that are volumetrically defined. The benefits for nucleic acid determination include: simplified quantification with no need for standard curve generation, increased signal–to-noise ratio which enriches for rare targets by means of reducing competition that arises from high-copy number templates, unparalleled precision enabling reliable measuring of small fold differences in target sequences due to massive sample partitioning, and the removal of PCR efficiency bias when the amplification efficiency reliance of PCR is removed thereby reducing error rates, ultimately enabling accurate target quantification.

Previously generated cDNA (High Capacity cDNA Reverse Transcription Kit, Applied Biosystems) from root RNA samples for qPCR analysis by CPGR was used for the ddPCR.

There was no need to regenerate the cDNA because the samples were maintained at the appropriate storage temperature of -80ºC, with minimal freeze-thaw cycles, and could therefore be used for further downstream analyses. ddPCR™ reactions were prepared using the intercalating dye EvaGreen® which forms part of the Qx200™ EvaGreen® ddPCR™ Supermix (200 reactions) (Bio-Rad, United States), specific primers and template cDNA. Experimental workflow using the QX200™ Droplet Digital PCR (ddPCR™) System situated at the Institute of Infectious Disease and Molecular Medicine, University of Cape Town, included the combination of water-oil emulsion droplet technology with microfluidics. The samples were placed into the QX200™ droplet generator which partitioned the individual samples into 20,0000 nanoliter-sized droplets. The droplets generated were then transferred to a 96-well plate after which PCR amplification was carried out within each droplet using the C1000 Touch™ thermal cycler. After the PCR reaction had reached completion, the plate containing the droplets was placed into and streamed in a single file on a QX200™ droplet reader, where the fluorescent positive and negative droplets were counted to calculate the target DNA concentration by using a two-colour detection system. Positive droplets, which contain at least one copy of the target cDNA molecule, should, in theory, exhibit increased fluorescence when compared with the negative droplets.

3.5 Results and discussion

3.5.1 Assessment of RNA quantity and quality

RNA quantity and purity were assessed spectrophotometrically using a NanoDrop® (ND1000, ThermoScientific, Delaware, USA) and were tabulated for leaf and root RNA samples (Appendix A1, Table A1a and A1b, respectively). RNA from leaf tissues was successfully visualised on a 1% agarose gel with additional ethidium bromide (EtBr) (Appendix A1, Figure A1). The 25S and 18S ribosomal RNA bands were clearly visible on the gel electrophoresis image, indicating good quality RNA. Many of RNA samples extracted from root tissues failed QC analysis (highlighted in red, Appendix A1, Table A1b). The A260/A280 and/or A260/A230 ratios of less than 1.5 for many of these root RNA samples indicated the likely presence of carbohydrates, phenol and possibly guanidine HCl (used in commercial DNA/RNA isolation kits). Co-purified phenol and carbohydrates in these samples, would have the potential to entrap RNA and their presence could have reduced the efficiency of their downstream enzymatic reactions. PCR inhibitors such as tannins, humic substances, polyphenols and polysaccharides need to be removed carefully during the process of RNA extraction from root tissues. The fibrous nature of root tissues, and the presence of

PCR inhibitors, such as those mentioned above, need to be addressed when preparing samples for downstream processing. This highlights the need to find an optimum RNA extraction protocol for root tissues from resurrection plants.

3.5.2 Analysis of the gene expression of the 21 target LEA genes in leaf tissues

LEA gene expression studies have primarily been conducted on seeds, as these plant reproductive structures were the first to be reported to accumulate LEA proteins during maturation drying and attainment of desiccation tolerance (as discussed in Chapter 1). Since subsequently being detected in vegetative tissues exposed to various stresses such as dehydration, osmotic and/or low-temperature stress, increased attention has been placed on measuring changes in LEA gene expression, to shed light on what is happening at the transcript levels in these tissues during abiotic stresses (Battaglia *et al.,* 2008; Bies- Ethève *et al*., 2008; Hundertmark and Hincha, 2008). Most of the work conducted on LEA gene and protein expression has focussed on LEAs falling into the first three groups, Groups 1, 2 and 3, and therefore we can only postulate what the functions of the other LEAs belonging to the other groups may be, if they are expressed in response to stresses, such as desiccation, as in this case.

In order to study the gene expression changes of these individual 21 LEA-like genes during dehydration and rehydration of *X. humilis* leaf tissues, the mRNA levels were monitored using quantitative real-time PCR. To correct for sample-to sample and PCR variations, the expression levels of the reference gene 18S RNA was used for normalisation. It has been previously reported that 18S RNA was successfully used as a reference gene in qPCR dehydration stress studies (Bresler, 2010; Maredza, 2007). In addition, the expression levels were also normalised to EF1A, chosen as an additional reference gene (data not shown). Gene expression results are displayed in both tabulated form (Table 3.1) and using graphical representations for each LEA gene (Figure 3.2-3.6, A-U).

Table 3.1 Data representing the mean transcription levels normalized to 100% RWC (fully hydrated) ± standard error of the mean (SE), across four biological replicates for each LEA gene arranged by their respective groupings. Significance is indicated by * and the darker the colour for that specified for each LEA group the more highly significant the p-value.

		Dehydration					Rehydration		
		100% RWC	70-80% RWC	40-50% RWC	20-30% RWC	<10% RWC	2-10% RWC	50-60% RWC	90-100% RWC
Group 1 LEAs	XhLEA1-4S1	0.0±0.0	ns 1.030±0.6098	* 1.552±0.5524	ns 1.323±0.1372	ns 0.7764±0.3783	ns 0.7980±0.4312	ns -0.2046±0.2590	ns 0.1443±0.3493
	XhLEA1-1S2	0.0±0.0	**** 1.866±0.4180	**** 2.338±0.07890	**** 2.642±0.1175	**** 2.114±0.07486	**** 1.865±0.1550	ns 0.3333±0.2776	ns 0.07990±0.1392
Group 2 LEAs (dehydrins)	XhLEA2-1	0.0±0.0	**** 2.375±0.1946	**** 2.600±0.1169	**** 2.593±0.08199	**** 2.344±0.06952	**** 2.098±0.07423	ns 0.5997±0.3992	ns 0.7434±0.2422
	XhLEA2-2	0.0±0.0	**** 2.411±0.2807	**** 2.504±0.2014	**** 2.710±0.08316	**** 2.426±0.1508	**** 2.359±0.07982	ns 0.4781±0.4136	* 1.045±0.2272
	XhLEA2-3	0.0±0.0	** 1.037±0.4242	*** 1.500±0.2853	**** 2.033±0.04208	**** 1.730±0.1243	**** 1.660±0.03656	ns 0.4742±0.1924	ns 0.3712±0.04097
	XhLEA2-4	0.0±0.0	* 1.055±0.4512	**** 1.852±0.09122	**** 2.503±0.07188	**** 2.324±0.01754	*** 1.739±0.2741	ns 0.5682±0.3497	ns 0.1806±0.06422
	XhLEA2-5	0.0±0.0	** 1.746±0.2325	* 1.367±0.2216	ns 0.6802±0.1930	ns 0.2093±0.3542	ns 0.1611±0.3458	ns -0.5253±0.4461	* 1.354±0.2309
	XhLEA2-6	0.0±0.0	* 0.7199±0.2713	* 0.7806±0.2430	**** 1.373±0.08939	**** 1.334±0.1789	**** 1.290±0.05098	ns 0.5076±0.1564	ns 0.2034±0.1481
Group 3 LEAs	XhLEA3-1	0.0±0.0	**** 2.351±0.2024	**** 2.471±0.06430	**** 2.417±0.09863	**** 1.945±0.09682	**** 1.775±0.1499	ns 0.4521±0.3454	ns 0.4213±0.1188
	XhLEA3-2	0.0±0.0	ns 1.704±0.6963	* 2.082±0.4764	** 2.367±0.4289	** 2.365±0.3868	** 2.262±0.2925	ns 0.7520±0.5094	ns 0.3326±0.3528
	XhLEA3-3	0.0±0.0	**** 2.484±0.3458	**** 2.509±0.2394	*** 1.752±0.3130	*** 2.036±0.1972	*** 2.061±0.2435	ns 0.4852±0.3789	ns 0.7265±0.2821
	XhLEA3-4	0.0±0.0	**** 2.786±0.1372	**** 2.921±0.0055995	**** 2.425±0.1141	**** 1.896±0.1627	**** 1.619±0.2116	ns 0.4316±0.3349	ns 0.6078±0.06376
	XhLEA3-5	0.0±0.0	**** 2.111±0.2890	**** 2.375±0.2663	**** 2.715±0.1976	**** 2.462±0.1430	**** 2.429±0.1456	ns 0.7051±0.4246	ns 0.7821±0.2203
	XhLEA3-6	0.0±0.0	**** 2.978±0.2708	**** 2.883±0.1922	**** 2.489±0.3013	**** 2.137±0.08310	**** 2.018±0.1598	ns 0.6989±0.3115	** 1.179±0.3159
	XhLEA3-7	0.0±0.0	**** 2.345±0.2390	**** 2.546±0.1756	**** 2.622±0.08695	**** 2.148±0.09147	**** 1.935±0.1747	ns 0.4215±0.4956	ns 0.7129±0.1381

Group 4 LEAs	XhLEA4-1	0.0±0.0	**** 1.789±0.2627	**** 1.893±0.3170	**** 2.514±0.06835	**** 2.128±0.06901	**** 1.936±0.1032	ns 0.4385±0.2171	ns 0.4879±0.1559
	XhLEA4-2	0.0±0.0	**** 2.376±0.2045	**** 2.509±0.1715	**** 2.932±0.1490	**** 2.617±0.06356	**** 2.269±0.1194	ns 0.7079±0.4158	* 0.8867±0.2675
Group 5a LEAs	XhLEA5a	0.0±0.0	* 1.386±0.5600	**** 2.472±0.1740	**** 2.780±0.07859	**** 2.361±0.1519	*** 2.053±0.2150	ns 0.5901±0.4298	ns 0.3406±0.2253
Group 5b LEAs	XhLEA5b	0.0±0.0	ns 0.004981±0.2321	ns 0.4954±0.3935	** 1.237±0.1042	** 1.133±0.1011	* 0.9476±0.06318	ns -0.2130±0.2334	ns -0.1974±0.1468
Group 5c LEAs	XhLEA5c-1	0.0±0.0	** 1.881±0.5172	*** 2.546±0.2219	**** 2.896±0.1928	**** 2.654±0.1486	*** 2.439±0.1616	ns 0.5747±0.6205	ns 0.3320±0.3284
	XhLEA5c-2	0.0±0.0	**** 1.969±0.3701	**** 2.172±0.2359	**** 2.118±0.06804	**** 1.860±0.1296	**** 1.937±0.1090	ns 0.3512±0.1160	ns 0.4786±0.1027

Symbol	Meaning
ns	$P > 0.05$
*	$P \leq 0.05$
**	$P \leq 0.01$
***	$P \leq 0.001$
****	$P \leq 0.0001$

qPCR gene expression analysis results from leaf tissues, showed the up-regulation of both putative Group 1 LEA genes, XhLEA1-4S1 and XhLEA1-1S2, during dehydration, and down-regulation during rehydration, with significance in XhLEA1-4S1 at 40-50% RWC ($p \leq 0.05$) during dehydration only and significance across all points of dehydration and the first point of rehydration for XhLEA1-1S2 ($p \leq 0.0001$) (Table 3.1, Group 1 LEAs, blue and Figure 3.2, A and B). Relative gene expression reached a maximum for XhLEA1-4S1 at 40-50% RWC and 20-30% RWC for XhLEA1-1S2, during dehydration. To date, studies investigating the expression of Group 1 LEA genes have been limited to seeds and the expression profiles in desiccation tolerant vegetative tissues are yet to be published (Illing *et al.,* 2005; Manfre *et al*., 2006). Ngubane (2008) reported similar results for both of these putative Group 1 LEA genes within the leaf tissues. While the author attempted analysis using root tissues, similar difficulties to those found in the current study, were experienced, and the results were inconclusive. Unfortunately, we were similarly unable to confirm the expression pattern of these Group 1 LEAs in root tissues. Ginbot (2011), however, was able to express both LEA proteins and successfully generate antibodies targeted to these LEAs. He was able to demonstrate through the use of western blotting that both LEAs were expressed in leaves at RWC below 55%, disappearing by 6 hours after rehydration. XhLEA1-4S1 did not accumulate in roots during drying, but XhLEA1-1S2 did. Furthermore, both LEAs were present in seeds of *X. humilis* and XhLEA1-4S1 was found to be present in the seeds of *A. thaliana*, whereas XhLEA1-1S2 was not. It was summarised in his study, that these two Group 1 LEA proteins were likely to be involved in membrane stabilization through the formation of a hydrogen bonding network and/or a glassy matrix that would prevent membrane adhesion and stabilize the cytoplasm in the plasma membrane-cell wall continuum in the dry state. The hydrophilic nature of these XhLEA proteins, their expression in response to desiccation and the fact that they are plasma membrane/cell wall associated, support the proposed function (Ginbot, 2011). It has been understood that in plants, many of the LEA genes belonging to groups 1 and 2, are preferentially accumulated during embryo development, especially in dry seeds, although they have been detected in other organs that undergo dehydration, such as pollen grains, vegetative and root tissues (Battaglia *et al*., 2008). Additionally, many of the characterised genes within the Group 1 LEA genes, are responsive during water-limiting conditions, mainly within the embryos and in a few cases, in the vegetative tissues of young plant seedlings (Gaubier *et al*., 1993; Vicient *et al*., 2000; Battaglia *et al*., 2008).

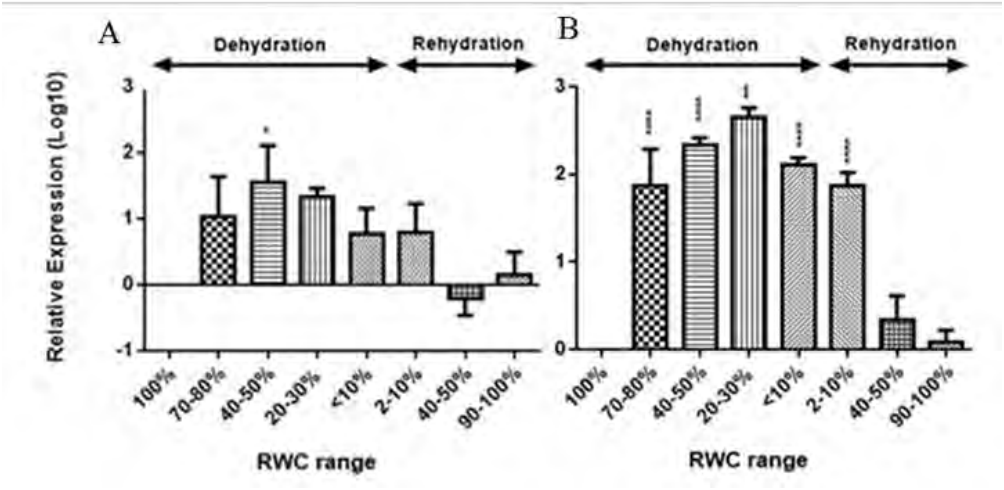


Figure 3.2 Bar graphs illustrating the relative gene expression levels of the two Group 1 LEA genes- XhLEA1-4S1 (A) and XhLEA1-1S2 (B)- during the dehydration and rehydration treatment of *X. humilis*. Each data point represents the mean and the error bars represent the standard error of the mean (SEM), of relative expression levels of each of the LEA-like genes (N = 4), compared to 18S ribosomal RNA, calculated according to Pfaffl (2001). The log10 change in expression levels is represented on the Y-axis and the % RWC values are represented on the X-axis. The level of significance in expression change is indicated by asterisks (*). The graph and SEM values were generated using the GraphPad Prism software (Version 6).

The putative Group 2 LEA genes (dehydrins) as a whole exhibited significant early up-regulation during dehydration and down-regulation during rehydration (Table 3.1, Group 2 LEAs, dehydrins, pink, Figure 3.3, C-H). Two interesting Group 2 LEAs were XhLEA2-2 (originally Xh_RD_19A06) and XhLEA2-5 (originally Xh_RD_19H04), both originally derived from root dehydrated (RD) tissues. In both of these LEAs there was significant change in relative gene expression in leaf tissues during rehydration, with a significant re-up-regulation at 90-100% RWC (72 hours post rehydration) ($p \leq 0.05$) (Figure 3.3, D and G, respectively). In almost all of the dehydrins within this group, with the exception of XhLEA2-5, a maximum increase in relative gene expression was reached at 20-30% RWC during dehydration stress treatment. Both transcript and protein abundance of Group 2 LEAs, the dehydrins, in general, accumulate to high amounts in vegetative tissues, following the exposure to various stresses (Hanin *et al*., 2011). As previously discussed, many of the dehydration-upregulated cDNAs identified by Collett *et al*., (2004), were homologous to LEAs and to dehydrins, suggesting that these dominate the mRNA population in desiccation *X. humilis* leaf tissues. At these stages of desiccation, the dehydrins may be acting as structural stabilizers or as putative chaperonin-like proteins, as described in Vicré *et al*., (2004), and/or they may be protecting and defending various protein structures, as described

by Hong-Bo *et al*., (2005). It is becoming increasingly evident that dehydrins have been useful as markers to detect for stress tolerances in many species (Hanin *et al*., 2011). This could be true too of *X. humilis*, given the large number of genes found encoding for dehydrins in this species (Collett *et al.,* 2004). In studies focussing on dehydration stress, correlations have been found between drought tolerance and the accumulation of dehydrin proteins in *Populis popularis*, a Dhn3 and a Dhn4 transcript accumulation in a set of Korean barley cultivars and between the levels of dehydrin transcript accumulation in two differently tolerant cultivars of Durum wheat (*T. turgidum* ssp. *durum*). Numerous transgenic studies have revealed a positive effect of dehydrin gene expression on plant stress tolerance. Therefore, with increasing information from various fields and studies, this group of dehydrins proves to be an incredibly versatile group of LEA proteins, that exhibit a variety of functions and possibly more functions that are yet to be unveiled (Hanin *et al*., 2011).

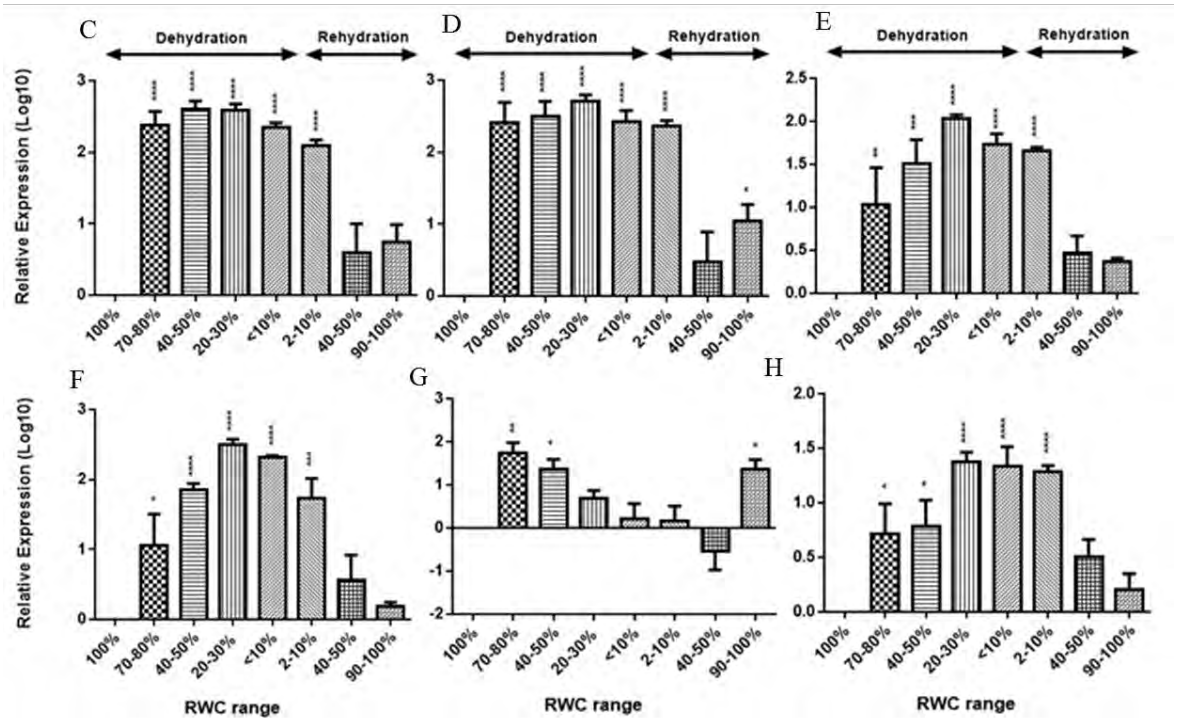


Figure 3.3 Bar graphs illustrating the relative gene expression levels of the six putative Group 2 LEA genes- XhLEA2-1 (C), XhLEA2-2 (D), XhLEA2-3 (E), XhLEA2-4 (F), XhLEA2-5 (G) and XhLEA2-6 (H)- during the dehydration and rehydration treatment of *X. humilis*. Each data point represents the mean and the error bars represent the standard error of the mean (SEM), of relative expression levels of each of the LEA-like genes (N = 4), compared to 18S ribosomal RNA, calculated according to Pfaffl (2001).). The log10 change in expression levels is represented on the Y-axis and the % RWC values are represented on the X-axis. The level of significance in expression change is indicated by asterisks (*). The graph and SEM values were generated using the GraphPad Prism software (Version 6).

Similar to the Group 2 LEA-like genes, the large number of the putative Group 3 LEA genes also exhibited significant early up-regulation during dehydration and down-regulation during rehydration, as a whole (Table 3.1, Group 3 LEAs, green and Figure 3.4, I - O). XhLEA3-6 (originally Xh_LD_10A10) was the only Group 3 LEA-like protein that showed significant re-up-regulation of relative gene expression at 90-100% RWC post rehydration ($p \leq 0.01$) (Figure 3.4, N). This result was similar to that obtained for the two putative dehydrins, XhLEA2-2 and XhLEA2-5. These Group 3 LEA gene representatives may be involved in enriching ions during the dehydration of the *X. humilis* plants used in this study (Zhang and Zhao, 2003; Yu, 2003; Hong-Bo *et al*., 2005). Microarray expression analysis and Northern blot procedures have both detected the upregulation of dehydrins, amongst other LEAs, and Group 3 LEAs, in *X. humilis,* in response to abiotic stresses, such as dehydration (Collett *et al*., 2004; Illing *et al*., 2005). The expression data obtained for the putative dehydrins and Group 3 LEAs, in our list of 21 genes, support the data obtained by Collett *et al*., (2004) and Illing *et al*., (2005).

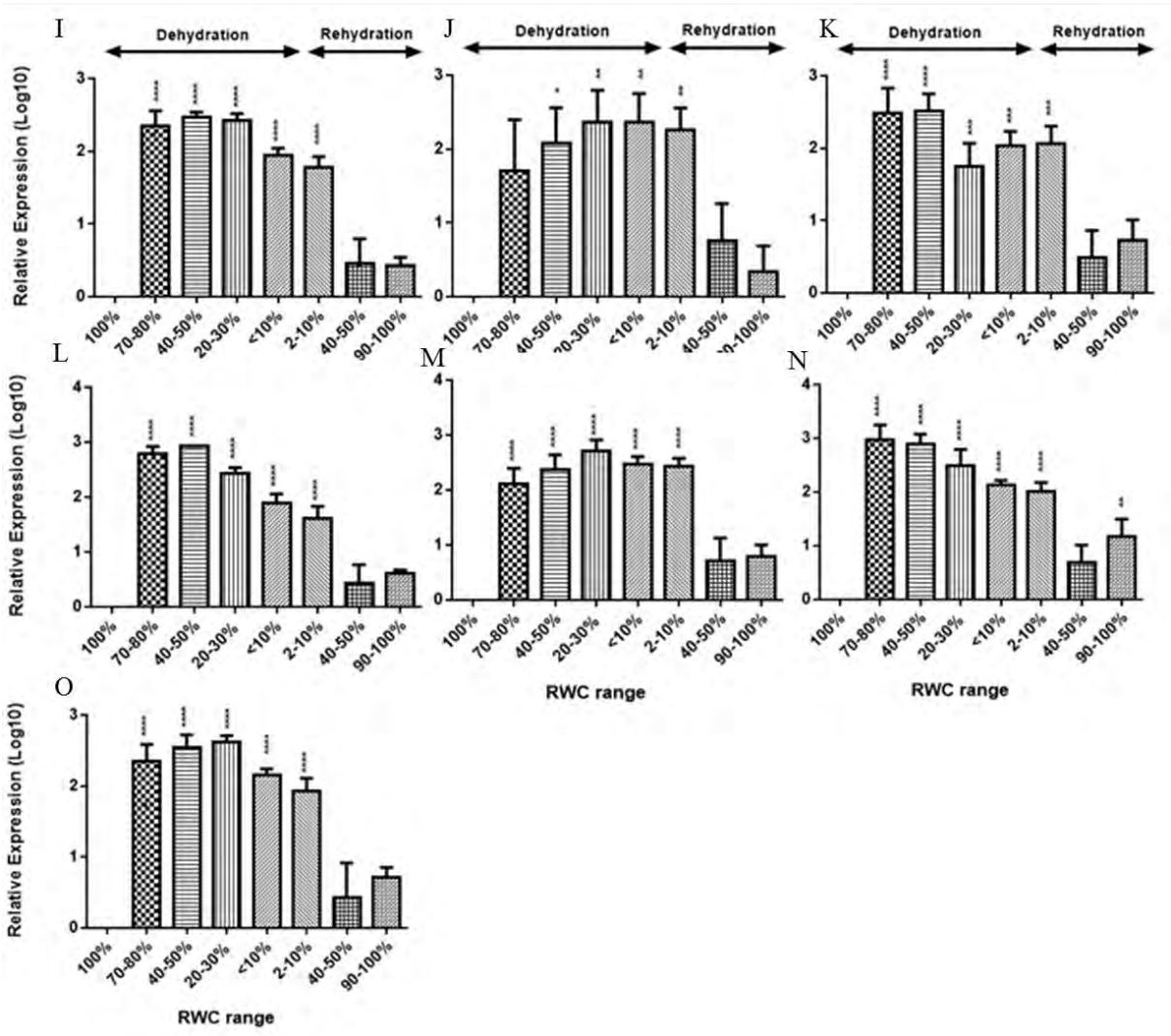


Figure 3.4 Bar graphs illustrating the relative gene expression levels of the seven putative Group 3 LEA genes- XhLEA3-1 (I), XhLEA3-2 (J), XhLEA3-3 (K), XhLEA3-4 (L), XhLEA3-5 (M), XhLEA3-6 (N) and XhLEA3-7 (O)- during the dehydration and rehydration treatment of *X. humilis*. Each data point represents the mean and the error bars represent the standard error of the mean (SEM), of relative expression levels of each of the LEA-like genes (N = 4), compared to 18S ribosomal RNA, calculated according to Pfaffl (2001). The log10 change in expression levels is represented on the Y-axis and the % RWC values are represented on the X-axis. The level of significance in expression change is indicated by asterisks (*). The graph and SEM values were generated using the GraphPad Prism software (Version 6).

The two putative Group 4 LEA genes within our set of 21 LEA-like mRNA gene transcripts showed significant up-regulation during dehydration and significant down-regulation after 6 hours of rehydration (2-10% RWC point) ($p \leq 0.0001$) (Table 3.1, Group 4 LEAs, orange and Figure 3.5, P and Q). The maximum point of up-regulation of relative gene expression in both these LEA genes occurred at 20-30% RWC during dehydration. The only point of significant deviation as seen in some other LEA genes from previous groups, in XhLEA4-2 (originally Xh_RD_25D02) is the significant re-up-regulation of the expression of this gene at 90-100%

RWC from rehydration ($p \leq 0.05$) (Figure 3.5, Q). Accumulation of Group 4 LEA transcripts, in response to water-deficit treatments, has also been recorded in *Arabidopsis* vegetative tissues and in tomato (*Solanum lycopersicum*) plants (Cohen *et al*., 1991; Delseny *et al*., 2001; Olvera-Carrillo *et al.,* 2010), supported by the pattern seen in the two putative Group 4 LEA genes, XhLEA4-1 and XhLEA4-2. The Group 4 LEA proteins appear to be associated with functions much like other LEA groups, protecting the plasma membrane during drying (Ginbot, 2011).

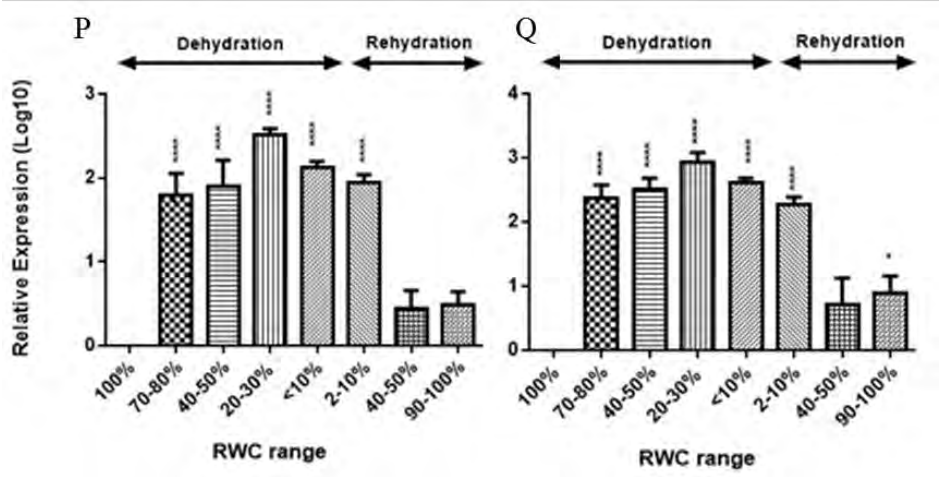


Figure 3.5 Bar graphs illustrating the relative gene expression levels of the two putative Group 4 LEA genes-XhLEA4-1 (P) and XhLEA4-2 (Q)-during the dehydration and rehydration treatment of *X. humilis*. Each data point represents the mean and the error bars represent the standard error of the mean (SEM), of relative expression levels of each of the LEA-like genes (N = 4), compared to 18S ribosomal RNA, calculated according to Pfaffl (2001). The log10 change in expression levels is represented on the Y-axis and the % RWC values are represented on the X-axis. The level of significance in expression change is indicated by asterisks (*). The graph and SEM values were generated using the GraphPad Prism software (Version 6).

The single LEA-like gene representative of the Group 5a LEAs, XhLEA5a, showed a significant increase in relative gene expression as dehydration commenced and down-regulation when water became available once again. The only point of significance during rehydration as compared to the fully hydrated (100% RWC) cDNA sample was the 2-10% RWC (6hrs post-rehydration) (Table 3.1, Group 5a LEAs, purple and Figure 3.6, R). Although little is known regarding this group of LEA genes and protein counterparts, the data that is available indicate that their transcripts accumulate during the late stage of seed development and as predicted, in response to stress conditions, including drought (Battaglia *et al*., 2008).

The only representative falling into the Group 5b LEAs, XhLEA5b, exhibited maximum change in relative gene expression and thus up-regulation at 40-50% RWC during dehydration. Upon rehydration, there seems to be a marked down-regulation of this gene, however statistically, this change is not significant as compared with 100% RWC (fully hydrated) (Table 3.1, Group 5b LEAs, red and Figure 3.6, S).

Both the gene representatives from the Group 5c LEAs, XhLEA5c-1 and XhLEA5c-2, showed similar trends in relative gene expression with significant up-regulation of expression during dehydration and a marked down-regulation during rehydration (however, statistically, the two final points of rehydration for both these genes was not significant when compared with the normalisation point) (Table 3.1, Group 5c LEAs, grey and Figure 3.6, T and U).

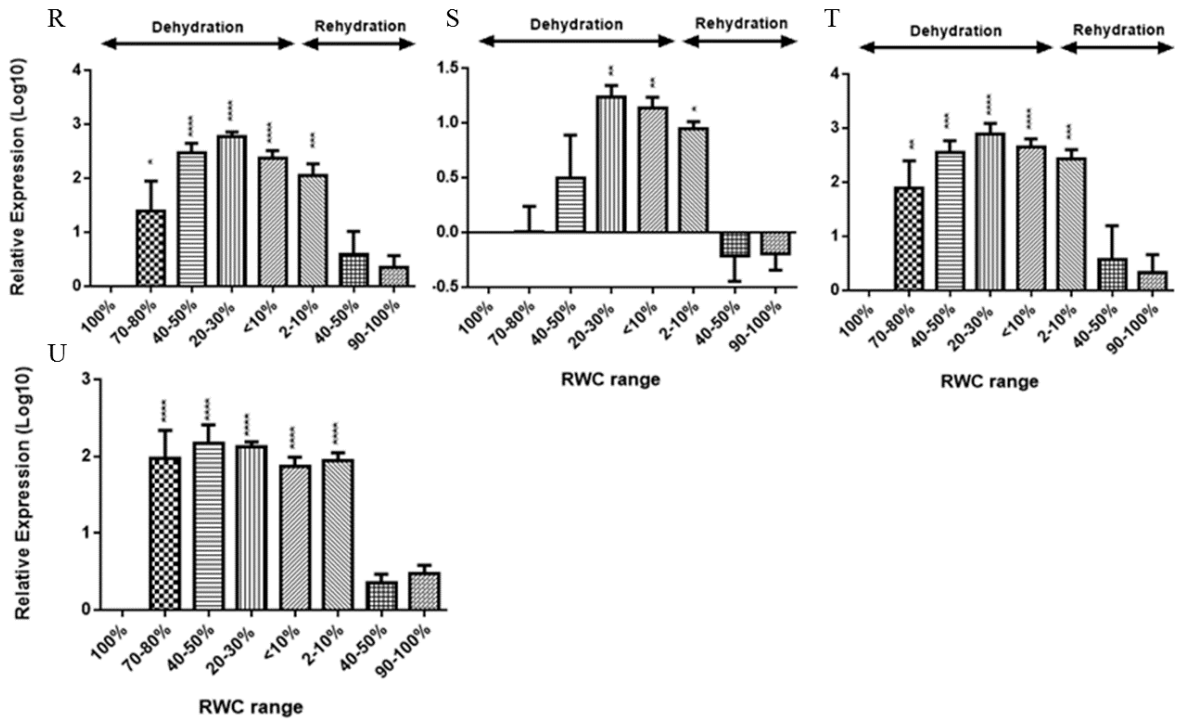


Figure 3.6 Bar graphs illustrating the relative gene expression levels of the 4 putative Group 5 LEA genes- XhLEA5a (R), XhLEA5b (S), XhLEA5c-1 (T) and XhLEA5c-2 (U)- during the dehydration and rehydration treatment of *X. humilis*. Each data point represents the mean and the error bars represent the standard error of the mean (SEM), of relative expression levels of each of the LEA-like genes (N = 4), compared to 18S ribosomal RNA, calculated using the Pfaffl method (Pfaffl, 2001). The log10 change in expression levels is represented on the Y-axis and the % RWC values are represented on the X-axis. The level of significance in expression change is indicated by asterisks (*). The graph and SEM values were generated using the GraphPad Prism software (Version 6).

Comparisons of expression patterns in the vegetative tissues of desiccation tolerant plants to those seen in the desiccation sensitive model organism *A. thaliana,* is made available as more

recent global transcriptomic and proteomic analyses have become technically possible. Bies-Ethève *et al*., (2008) have demonstrated that there are more than 50 LEA-like genes or LEA proteins within the genome of this species, and expression of a single gene or a whole suite of genes can be queried against different tissues and different developmental stages (of plants and other organisms), under varying conditions. It has been found, surprisingly, that there is very little overlap between those LEA gene sets specific to seeds versus vegetative tissues.

The results of this relative gene expression analysis point to the marked up-regulation of all of these genes during dehydration in the leaf vegetative tissues of *X. humilis,* and subsequent decline in their expression during rehydration. This correlation allows us to speculate that these proteins do contribute to the tolerance of this species to water deficit stress, although the mechanisms whereby they achieve this, is still misunderstood. Their universal decline during rehydration is most likely due to the degradation of these transcripts, suggesting that they are not required under hydrated conditions.

3.5.3 Droplet Digital™ PCR (ddPCR™) protocol for root samples: future optimisations and considerations

Due to the highly variable results generated for the root samples, the data are not shown. However, the protocols and future optimisations of the experiments are explained in the hope that these can be used in the future to analyse RNA and cDNA samples derived from problematic tissues, such as roots. Ultimately, the success of all downstream processing (eg: qPCR) and data analysis of RNA samples, begins with the initial good quality extraction of RNA and conversion of that RNA to cDNA. A well-designed and optimised ddPCR reaction yields highly reproducible and robust results. It is essential that amplification occurs for accurate gene expression results to be collected and understood. Optimisation of experimental protocols needs to take place in order to obtain both positive and negative droplets for the Poisson algorithm to be applied correctly, with a clear separation between the two for accurate results. In order to assess data obtained from ddPCR™ experiments, certain information needs to be known to draw conclusions. For example, when using EvaGreen®, fluorescence may vary with amplicon length as well as PCR efficiency. Longer amplicons will therefore bind to EvaGreen® dye with a brighter fluorescence. Parameters to try for future testing would include: ensuring good quality preparation of the nucleic acid from the sample of interest and removal of any PCR inhibitors prior to analysis (if known inhibitors

cannot be readily removed, consider reducing their impact on the PCR reaction by diluting the sample 1:10); optimising the DNA concentration used in the individual reactions, aiming to reduce the amount of starting material to a minimum in order to obtain unsaturated results; running a thermal gradient for target primers to assess the optimal annealing temperatures using ddPCR™ because these may differ to those used in qPCR, with the use of different machines. The final data obtained from these thermal gradients would be merged on a 1D plot to assess which annealing temperature is optimal for each primer pair. The annealing temperature that shows that clearest distinction between positive and negative droplets should be chosen; and lastly, a digestion of the template could be performed, to fragment the genome. This incorporates more upstream work, so the first two parameters should be optimised and then restriction digestions should be looked at as a final resort. This may be an option when dealing with a genome high in secondary structures.

With resurrection plants being our main focus in our laboratory, it is important to analyse gene expression patterns within these specially adapted plants, at both the leaf and root level, and ask and answer questions such as: Are the patterns we see in plants previously worked on, such as *Arabidopsis*, the same in resurrection species such as the *Xerophyta* group? If not, what could be the reason and could this mean different functions of these proteins or different responses? Are these LEA genes regulated differently in different plants of the same Genus eg: *X. humilis* vs *X. viscosa*? Intriguingly, in a paper by Illing *et al.,* (2005), at least one LEA gene expressed in seeds of *A. thaliana* has an orthologue that is up-regulated in desiccating leaves of study plant, *X. humilis.* This interesting observation is consistent with the hypothesis that resurrection plants acquired "systemic desiccation tolerance" by reprogramming sets of genes that are known to be seed-specific (Tunnacliffe and Wise, 2007).

3.6 Conclusion

In this chapter, the mRNA transcript expression of the 21 LEA-like genes in leaf tissues, in response to dehydration and subsequent rehydration, was investigated successfully through quantitative real time PCR (qPCR). A significant up-regulation of gene expression in response to dehydration was observed in all 21 putative LEA genes, particularly at 40-50% and 20-30% RWC during the stress treatment.

Root tissues proved problematic and many factors such as amplicon selection, primer and probe design and those specific to the ddPCR™ system, need to be optimised. As with any PCR-based technology, it is essential to plan out and design each ddPCR™ experiment well before execution and prepare samples well. Therefore reliable, robust and high quality data can be obtained.

In the next chapter, I will focus on protein expression of a few selected *X. humilis* LEA s used in the qPCR study. A number of studies have demonstrated a weak or moderate correlation between mRNA and protein levels (Gygi *et al*., 1999; Bajic, 2006). Often the activity of proteins is controlled totally independently of gene expression in response to stress. Resurrection plants have been shown to store various mRNA transcripts upon dehydration as part of a wider rehydration strategy. This highlights the importance of measuring both transcript and protein levels in response to stress and would therefore also be of future value to study the plant proteins. This important aspect needs to be investigated in the future, as this may provide further clarity into the results obtained from the qPCR gene expression analysis.

Chapter 4

Cloning of 3 LEA-like genes and expression of recombinant his-tagged proteins

4.1 Introduction

Before proceeding with the cloning and expression of any set of proteins*, in silico* analysis is necessary in order to determine whether the chosen proteins have characteristic or unique sequences and physiochemical properties at the DNA and the protein level. The information obtained will ultimately enhance and inform subsequent experimental approaches. The analysis of the putative amino acid sequences will also allow the determination of whether the proteins undergo any form of post-translational modifications, as these may imply regulatory roles. After gaining such insight, appropriate cloning procedures for subsequent protein expression, purification and downstream analysis can be attempted. For recombinant proteins to be studied at the biochemical and structural level, three factors regarding their production needs to be successful: protein expression, solubility and purification. It is essential that all factors within the cloning and expression protocols are optimised, ultimately to ensure the production of sufficient quantities of purified protein. Reliability of expression of these proteins in a heterologous system is essential and the choice of an expression system is also key, as expressed recombinant proteins should ideally maintain the integrity of being correctly folded and functionally active as they would in a natural *in vivo* state. *Escherichia coli* has been the preferred expression system for the bulk of laboratories committed to high-throughput cloning, expression and purification of recombinant proteins. The benefits of *E.coli* as an expression host include well-studied physiology, genetics and availability of a variety of protein expression vectors, accelerated growth, high yield protein production rates of up to 10-30% of total cellular protein, ease of handling in a standard molecular biology laboratory, minimal cost and the capability to multiplex both expression screening and protein preparation (Shivashanmugam *et al*., 2009).

Previously, Ginbot (2011) showed the successful production of the two Group 1 LEA recombinant proteins, XhLEA1-4S1 and XhLEA1-1S2, using the pET-21a(+) cloning vector and the T7 tag method of purification. For the purpose of this study, two Group 2 LEA-like genes and one Group 3 LEA-like gene, were chosen. As previously mentioned (Chapter 2), the unifying motifs characteristic of the Group 2 LEA proteins are the Lysine-rich 15-residue motif, EKKGIMDKIKEKLPG (K segment), the Y-segment, whose conserved consensus

sequence is VTD [E/Q] YGNP and the S-segment, forming a tract of Ser residues whereas for the group 3 LEA proteins a highly variable repeating motif of 11 amino acids, TAQ [A/S] AK [D/E] KT[S/Q] E. These two LEA groups are the most widely studied and described groups to date. The dehydrins are found in a huge variety of photosynthetic organisms, much like the group 3s, and there are several previous studies available to compare data obtained to, and enable the drawing of educated conclusions. Investigations in which the proteins themselves are studied remain few and far between, and this chapter aims to expand our understanding of these proteins.

4.2 Aim

The aim of this chapter was to

a) utilise bioinformatics to analyse the amino acid sequences of the three chosen LEA-like proteins: originally termed Xh_LD_05G02, Xh_LD_27A05 and Xh_LD_44B08 and now being referred to as XhLEA2-3, XhLEA2-6 and XhLEA3-5, respectively, and;
b) successfully clone, sequence, express and obtain soluble purified recombinant proteins of the above LEA-like clones using the pET-21a(+) vector system (Novagen, USA), after expression in competent *E. coli* BL21 (DE3) pLysS cells.

4.3 Methods and materials

4.3.1 Bioinformatics analysis of the amino acid sequences of the three target LEA-like proteins chosen for cloning and protein expression studies using BLAST and ExPAsy

The full length nucleotide sequences (from start to stop codon) for each LEA gene was obtained from the NCBI database using the LEA clone ID from Table 2.1 (see Appendix B2) as well as the corresponding amino acid sequences obtained by the ExPASy translation tool (http://www.expasy.org/). Amino acid homology searches were carried out using the protein BLAST algorithms and NCBI databases. The ProtoParam tool (www.expasy.org) was used to analyse the amino acid composition of each of the target proteins.

4.3.2 Analysis of the expression vector for cloning using restriction enzyme sites

The resistance marker, map and sequence of the pET-21a(+) expression vector to be used in the cloning procedure was analysed (Appendix B1). This enabled the choosing of restriction enzyme sites within the multiple cloning site (MCS) and two specific sites, *HindIII* and *XhoI*, were chosen for downstream cloning experiments. The full length nucleotide gene sequences for each of the three LEA genes were analysed with the internet based NEBcutter program (http://nc2.neb.com/NEBcutter2/), to confirm that the two restriction enzyme sites chosen would not cut into the gene of interest.

4.3.3 Primer design and synthesis

The full length gene sequences for each of the three LEA-like genes chosen, obtained through NCBI, were used as a basis for the design of all primers (Appendix B2). Primer sets for the cloning of XhLEA2-3, XhLEA2-6 and XhLEA3-5, into the cloning vector, were designed to amplify the full length cDNA sequences (Table 4.1). *HindIII* and *XhoI* restriction enzyme sites were designed into the primer sequences, to facilitate cloning of the PCR products into the MCS of the relevant plasmid vector (pET-21a(+)). Primers were synthesised using standard methods provided by the Synthetic DNA Laboratory, Molecular and Cell Biology Department, University of Cape Town.

Table 4.1 Primers used for cloning and sequencing of XhLEA2-3, XhLEA2-6 and XhLEA3-5, in pET-21a(+)(his) vector.

Name	Sequence 5'→ 3'
XhLEA2-3 Forward primer:	CGTCGACAAGCTTGC**ATG**GAGGGTTTCGGGAACCAGC
XhLEA2-3 Reverse primer:	TGGTGCTCGAGTCTAGTGGTTTCCCGAGATCTTGTCCTTAATCTTG
XhLEA2-6 Forward primer:	CAAGCTTGC**ATG**GAGGGATACGGGAAC
XhLEA2-6 Reverse primer:	TGCTCGAGTTCAGCGACGGCCAGG
XhLEA3-5 Forward primer:	CAAGCTTGC**ATG**GCGAGGATTGTG
XhLEA3-5 Reverse primer:	TGCTCGAGTTTACAGCTCCTCGGAC

NB. *HindIII* and *XhoI* restriction sites are underlined and translation start codon (ATG) is in **bold**.

4.3.4 Preparation of the pET-21a(+)(his) expression vector DNA

The pET-21a(+) expression vector, maintained inside DH5α *E. coli* competent cells, to be used later in the cloning procedure, was subjected to Inverse-PCR mutagenesis in order to incorporate a 6 x Histidine tag (a polyhistidine tag) in the N-terminal, while simultaneously

removing the T7-Tag for ease of downstream detection and purification of recombinant proteins (Appendix B1). This tag was removed, as it proved difficult to detect in previous western blots with several proteins chosen from the list of 21 (data not shown). The histidine tag was chosen for its ease of detection in western blots, as well as its ease of purification of recombinant proteins. The new vector, pET-21a(+)(his) was transformed into and maintained within a glycerol stock of DH5α *E. coli* cells and was streaked out onto an individual LB-agar plate containing the antibiotic ampicillin, in a final concentration of 100µg/ml. The plate was incubated inverted at 37°C overnight. Single colonies were used to inoculate several 5ml LB cultures containing Amp. The inoculated cultures were incubated overnight with shaking at 37°C. Plasmid DNA containing the pET21a(+)(his) vector was isolated from the cultures using the Wizard® Plus SV Minipreps DNA Purification System (quick protocol) with to the incorporation of the enzyme Alkaline Protease (an enzyme that would ensure no plasmid DNA damage by endonucleases) and pooled. The following amendments were made to the Wizard® protocol: The cultures were pelleted for 10 minutes for production of cleared lystae, 350ul clear cell lysis buffer (a less concentrated buffer than the blue Cell Lysis Buffer) was added to the 600µl cell re-suspension solution that the pellet had been re-suspended in and 400ul cold (4ºC) neutralization solution was added and mixed thoroughly by inversion. The final plasmid DNA pellet was re-suspended in 20µl 1 x TE buffer. The NanoDrop® ND-100 spectrophotometer (Thermo Scientific, USA) was used to quantitate the pooled purified plasmid DNA in ng/µl and to assess the DNA purity.

4.3.5 Preparation of the template LEA plasmid DNA

The individual LEA genes to be cloned and expressed were maintained in pBluescript vectors, within DH5α *E. coli* cells. Glycerol stocks of the *E.coli* containing the respective LEA genes were streaked onto three individual LB-Amp agar plates, containing Amp at a final concentration of 100µg/ml. These plates were incubated inverted at 37°C overnight. Single colonies from the individual plates were used to inoculate 5ml LB-Amp cultures. The newly inoculated cultures were incubated overnight with shaking at 37°C. Plasmid DNA from the individual 5ml LB cultures was isolated using the PureYield™ Plasmid Miniprep System (Promega, USA) according to manufacturer's guidelines and using the centrifugation protocol. The NanoDrop® ND-100 spectrophotometer was used as per above, to quantitate the purified DNA in ng/µl and to assess the DNA purity.

4.3.6 PCR amplifications

Traditional PCR was performed to amplify the individual LEA clones. Both small and large scale PCR analysis was used to confirm individual LEA gene amplification and prepare for restriction digestions, respectively. All PCR reactions consisted of 0.5µM of each primer, 100ng plasmid DNA, 0.4mM dNTP's, 2mM Mg^{2+} (already present in the KAPA HiFi fidelity buffer), 1U of the KAPA HiFi DNA Polymerase enzyme (Kapa Biosystems, USA), 1 x KAPA HiFi fidelity buffer and distilled water in a 50µl final reaction volume. A negative control (all PCR components without LEA plasmid DNA) was included to detect the presence of contamination, primer dimers or non-specific amplification. For the PCR reactions, increasing the melting temperature (T_M) by 1°C was necessary to reduce non-specific binding. The PCR amplification cycles were carried out according to standard KAPA HiFi guidelines, taking into account the relevant T_M of the primer pair and adjustment to the number of cycles and duration of extension and final extension post-optimisation (Table 4.2).

Table 4.2 PCR parameters for each target LEA-like gene.

Step	Temperature (ºC)	Duration	Cycles
Initial denaturation	95ºC	3 min	1
Denaturation	98ºC	20 sec	35 cycles
Annealing	T_A ºC*	15 sec	
Extension	72ºC	30 sec/kb	
Final extension	72ºC	5 min	1

*** Individual T_A's for XhLEA2-3, XhLEA2-6 and XhLEA3-5 were 65ºC, 59ºC and 57ºC and respectively.**

The PCR products were confirmed by agarose gel electrophoresis of 10µl of each completed PCR reaction tube, using 6 x loading dye at a final concentration of 1x, for 30 minutes at 90V. For all DNA gels, either a 0.8% or a 1% agarose containing 1µg/µl of the intercalating agent ethidium bromide (EtBr) was used, and a 1 x Tris-acetate buffer (TAE, pH 8.0) was used as the electrophoresis buffer. After the gel had electrophoresed to completion, it was visualised using the Molecular Imager ChemiDoc XRS+ system (Bio-Rad, Germany).

4.3.7 Restriction enzyme digestion

The PCR products were subjected to RE digestion with the relevant enzymes such that PCR products could be unidirectionally ligated into a similarly digested plasmid vector. Therefore, the remaining portions of the individual reactions for each LEA gene were PCR-cleaned- up using the Wizard® SV Gel and PCR Clean-up System kit (by centrifugation), pooled, quantified and subjected to restriction enzyme double digestion using FastDigest *HindIII* and

XhoI (Thermo Scientific), at 37ºC for 20 minutes. The pET-21a(+)(his) vector was then subjected to double digestion, as per above. Following double digestion, the products were confirmed by agarose gel electrophoresis for 45 minutes at 90V, and subsequently gel purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA) according to the manufacturer's instructions with the following modifications: for a large gel slice excised from the gel, 800µl ADB was added and dissolved for 10 minutes at 57ºC. A maximum of 850µl of the dissolved sample was added to the spin column and centrifuged for 40 seconds at maximum speed. The spin column was finally placed into a 1.5ml Eppendorf and to this 15µl 1 x TE buffer was added and stored at room temperature for one minute before centrifugation of the sample at 14, 000 x g for 6 minutes. DNA ladders in the relevant molecular weight (MW) range were included on all gels for size determination of the PCR/RE digested products (O'GeneRuler 1Kb DNA ladder, Thermo Scientific).

4.3.8 DNA ligations

The recovered DNA was ligated into the target pET-21a(+)(his) expression vector using a 3:1 insert:vector ratio and the T4 DNA ligase enzyme (Thermo Scientific, EL0011), incubated at 22ºC for 1 hour followed by incubation at 4ºC overnight. All DNA ligations consisted of approximately 1µg DNA, 2µl ligase enzyme, 2µl ligase buffer and water to make a total reaction volume of 20µl. A ligation reaction that contained the vector only was included to assess vector relegation as well as any possible resulting background colonies present on the agar plates.

4.3.9 Transformation protocol

Competent DH5α cells were prepared according to the Rubidium Chloride (RbCl) method (Promega Protocols and Applications guide (3rd edition) p.45-46) with the following modifications: 5ml LB was inoculated with a freshly plated colony of *E. coli* DH5α cells and incubated overnight at 37ºC with shaking. This pre-culture was used to inoculate 100ml LB/200ml media and incubated with shaking until the optical density (OD) at 600nm reached approximately 0.4-0.6. The culture was placed onto ice for approximately 15 minutes and this was following by centrifugation for 5 minutes at 4ºC (4,500 x g). The resulting pelleted culture was resuspended in 40/21ml of a pre-made, pre-cooled TFB1 solution (30mM KOAc, 100mM RbCl, 10mM $CaCl_2$, 50mM $MnCl_2$ and 15% (v/v) glycerol, adjusted to pH 5.8 with 1M acetic acid), placed on ice for 15 minutes and then pelleted as before. The new pellet was then resuspended in 4ml of pre-cooled TFB2 solution (10mM MOPS, 75mM $CaCl_2$, 10mM

RbCl and 15% (v/v) glycerol, adjusted to pH .6.5 with 1M KOH), placed on ice for 15 minutes. 100µl aliquots were made and stored at -80ºC for subsequent transformations. Competent cells of BL21 (DE3) pLysS, that were used for subsequent protein expression studies were also prepared as described above. For each transformation reaction, one Eppendorf containing 100µl competent cells was used and thawed on ice for 5 minutes. The ligation reaction (from section x) was used to transform the above competent DH5α *E. coli* cells, known for their ease of transformation. 15µl ligation mixture was carefully added to the 100µl fully thawed DH5α *E. coli* cells, vortexed briefly and incubated on ice for 30 minutes. The mixture was then subjected to a heat shock step at 42ºC for 90 seconds, after which it was placed on ice for 2 minutes. 900µl of LB was added to the mixture, vortexed briefly and incubated with shaking at 37ºC for a maximum of four hours. 100µl dilute and 100µl concentrated cells were plated onto the LB-Amp agar plates and incubated overnight at 37ºC.

4.3.10 Colony PCR screening

Colony PCR screening was conducted on single colonies grown overnight, using previous cycling parameters set for each individual target gene (PCR section and same gel composition etc). Single positive colonies confirmed by agarose gel electrophoresis were inoculated into new 5ml LB-Amp cultures and grown up overnight at 37ºC. DNA was extracted from the pET-21a(+)(his)-LEA cultures using the Wizard® kit (as described above), quantified using the NanoDrop and a portion was used to re-transform into competent BL21 (DE3) pLysS cells (as per above transformation protocol) to make 1ml glycerol stocks (500µl transformed cells and 500µl 50% Glycerol) while the other portion (20µl of a 100ng/µl DNA stock) was sent for sequencing confirmation. The use of host *E. coli* strains such as BL21 (DE3) pLysS provides high stringency and consistent expression (Pan and Malcolm, 2000) and most importantly, this strain contains the T7 RNA polymerase gene to minimize basal level expression of potentially toxic gene products. Since the LEA genes were successfully cloned into the pET-21a(+)(his) vector, the T7 promoter and T7 terminator primers were utilised for sequencing (see appendix B3).

4.3.11 Protein expression

3µl of the glycerol maintenance stocks containing each LEA gene cloned into the pET-21a(+)(his) was streaked out onto an Amp/Chloramp LB plate and incubated overnight at 37ºC. Starter cultures of 10ml LB containing 100µg/ml Amp and 34µg/ml Chloramp were inoculated each with a single BL21 (DE3) pLysS positive colony containing the recombinant

pET21a(+)(his) plasmid of either XhLEA2-3, XhLEA2-6 or XhLEA3-5, and incubated overnight at 37ºC with vigorous shaking to promote aeration and stimulate bacterial growth. The entire 10ml overnight culture was added to a 500ml LB (containing Amp and Chloramp), incubated at 37ºC with shaking for 2 hours and grown until the OD600 reached 0.6. Protein expression was then induced by the addition of isopropyl β- thiogalactoside (IPTG) to a final volume of 1.5mM, and the cultures were incubated with gentle shaking at 30ºC for an additional 4 hours. The cells within the induced 500ml culture were harvested and pelleted by centrifugation (4ºC, 10 000 x g for 10 minutes).

4.3.12 Protein extraction and purification

Extraction and purification of heat-stable His-tagged recombinant LEA proteins was conducted using the Protino® Nickel TED (Ni-TED) Histidine Tag affinity purification kit (Macherey-Nagel, Germany) as follows: 12ml of 1 x Lysis Equilibration Buffer (LEW) was prepared followed by the preparation of a 9ml extraction buffer (1 X LEW Buffer, 0.1% Triton X-100, BSA, Protease inhibitor tablet, 2mM MgCl2 and Benzonase). Immediately after the addition of the Benzonase, the total protein pellet was re-suspended in 8ml of the above extraction buffer (1ml buffer was stored for future protein quantification) and lysed by sonication with an output power set at 4 (40% cycle duty), using 3 to 4 cycles of alternating 20 seconds ON and 40 seconds OFF, to decrease foaming. The lysate was centrifuged to remove cell debris at 14 000 x g for 10 minutes and the supernatant was collected and pooled. 500µl of the pooled supernatant was separated and stored as the crude protein lysate and the remaining supernatant was heated at 80ºC for 10 minutes in order to separate heat-stable LEA proteins only. This was achieved by two centrifugation steps (14 000 x g for 10 minutes) after which the resulting supernatants containing the heat-stable LEA proteins, were collected. The heat treatment step was introduced as previously reported for the isolation of heat stable proteins in a related work (Pelah *et al*., 1995; Rudiger *et al*., 1995; Borovskii *et al*., 2002). Pure His-tagged recombinant LEA proteins were prepared from the heated protein extract by use of the Protino® Ni-TED kit according to the manufacturer's instructions with the following modifications: 8ml of a 1 x LEW buffer and 8ml of a 1 x Elution buffer were prepared. The Protino® Ni-TED Columns were equilibrated with 4ml 1 X LEW Buffer and this was immediately followed by the addition of the crude extract to the column, allowing it to pass through the column matrix by means of gravity flow. The unbound proteins were removed by a washing step with 4ml 1 x LEW buffer followed by the elution of the recombinant LEA proteins with 7ml of the 1 x Elution buffer. The eluted LEA protein

fraction was then concentrated using protein concentrator columns (Amicon ultra centrifugal filters, MWCO 3K, Merck Millipore). These columns concentrate a maximum to 15ml to a minimum of ~700µl in 1 hour. A washing step with autoclaved water was included to remove any glycine/glycerine in the filter. The 7ml eluate was then added to the filter and centrifuged at 4 800 x g for 50 minutes after which 3ml 1 x PBS buffer was added to buffer exchange and concentrate the proteins at 4 800 x g for 35 minutes to a final volume of 500-600µl. Protein concentration per µl for the crude protein lysate was determined using the Bio-Rad protein standard assay, whereas for the concentrated, purified protein the Bradford (1976) based microassay was used.

4.3.13 Protein analysis using SDS-PAGE and western blot

Proteins were subjected to SDS-PAGE analysis using a 12% protein gel. 20µg total and 4µg pure protein extracts (in 5 x SAB) were loaded and electrophoresed at 90V for 140 minutes. Gels were stained using a 2.5% Coomassie Brilliant Blue R-250 (Sigma Aldrich) solution placed in 45% methanol, 10% glacial acetic acid and 45% water for three hours with and de-stained over-night. Both staining and de-staining steps were conducted with gentle shaking. For Western blot analysis, 2µg total and 1µg pure protein extracts (in 5 x SAB) were separated on 12% SDS-PAGE gels as described above. After completion of electrophoresis, the gels were subjected to electro-blotting onto a nitrocellulose membrane at 100V for 1 hour, after which successful transfer of the proteins onto the membranes was indicated by PonceauS staining for 30 seconds followed by a de-staining step where the membrane was washed with 1 x TBST (1 minute). Following staining, the membranes were blocked with 1 x TBST containing 10% fat-free milk power (1 hour, room temperature). Membranes were then exposed to a 1:40 000 monoclonal anti-his tag antibody (HRP-conjugate) and incubated at room temperature for 1 hour, with gentle shaking. Post antibody incubation, membranes were washed four times with 1 x TBST (for 5 minutes each time) and detected using the Advansta WesternBright ECL HRP substrate (one part Stable Peroxide solution and one part Luminol enhancer solution; ratio 1:1) and visualised using the departmental Molecular Imager ChemiDoc XRS+ system (Bio-Rad, Germany). The recombinant protein of XhLEA5a (originally termed Xh_LD_27D08), successfully expressed, extracted and purified by Dr Nashied Peton, was used as a positive control in this study.

4.4 Results and discussion

4.4.1 Nucleotide and amino acid sequences of the three target proteins

The clone ID, and full length nucleotide and amino acid sequences obtained by NCBI and ExPASy, were those used in the subsequent cloning procedures (Appendix B2).

4.4.2 Amino acid sequence analysis for conserved domains (as described in Chapter 2) and homologies using BLAST

XhLEA2-3:

MEGFGNQQDQLRRNDRTNEHGAPGQTGYSAQHGVIGGQQHHQNKQQGLGSTGAGIKNK
LHRSNSSSSSSESDGEGGRRKKGIKDKIKEKMPGQHNQGQTGQGITGSHQSHGATGQQGY
GAAGQHGGKEGTMDKIKDKISGNH

XhLEA2-6:

MEGYGNHQHYNVDEYGNPLPSGYGENYGHQPMQRPGEYYGNQGYGGHQHGAYSGPGY
QQEYGSGPRLQRSGSSSSSEDDGYGGRRKKGLKDRIMENLPGRR-

Figure 4.1. Conserved Y segment (highlighted in yellow) and S segment (highlighted in pink) present within the protein sequences of the two putative group 2 (dehydrin) LEA proteins.

The putative group 3 LEA protein, XhLEA3-5, did not exhibit the characteristic highly variable sequence motif TAQ [A/S] AK [D/E] KT[S/Q] E and therefore it was decided to do a standard protein BLAST on the internet accessible NCBI website: (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) to assess homology to other proteins within the database. Protein BLAST results indicate that each of the three putative LEA genes chosen for cloning and expression studies have successful homology with other LEA-like proteins within the non-redundant protein sequences (nr) database. The putative dehydrin, XhLEA2-6, had the highest score of the single best aligned sequence with a *X. viscosa* putative dehydrin (unpublished data, Baker, Mundree and Thompson), whereas XhLEA2-3 shared a highest score with a dehydrin DH1 from *Coffea canephora* (Hinniger *et al*., 2006). XhLEA3-5 the putative Group 3 LEA protein, shared similarity with a LEA-like protein in *Lilium longiflorum* (Mogami *et al*., 2002). In the alignments list generated by NCBI protein BLAST, D-29 LEA proteins also shared some homology with this putative group 3 LEA. The D-29 classification, as previously explained, is that of a cotton LEA protein, and this falls under the

second subgroup of the Group 3 LEA group, 3B, this group being more heterogeneous than its counterpart group 3A (a highly conserved subgroup, with two characteristic motifs).

4.4.3 Amino acid composition of the three target proteins

The ExPASy ProtoParam tool was used to analyse the amino acid composition of each of the three LEA proteins chosen. Both the putative Group 2 (dehydrin) LEA proteins, XhLEA2-3 and XhLEA2-6, exhibited a high proportion of charged and polar amino acids (R, D, E, K, N, Q, H, M, S, T, Y) and a low proportion of nonpolar, hydrophobic amino acids and lacked Trp (W) and Cys (C) residues (Table 4.3), consistent with the typical characteristics explained by Battaglia *et al*., (2008). The preponderance of Gly (G) in both LEA proteins, contributing to 21.8% and 21.2% in the amino acid composition of XhLEA2-6 and XhLEA2-3, respectively, indicates their hydrophilic nature, a characteristic of LEA proteins (Battaglia *et al.,* 2008). XhLEA3-5 also exhibited a high fraction of polar, charged amino acids (predominantly A, E and K) and a low fraction of non-polar, hydrophobic amino acids, again consistent with general characteristics of LEA proteins (Battaglia *et al.,* 2008). Glycine (G) also made up a predominant portion of the amino acid composition (6.6%). All three LEA proteins have the characteristic unusual amino acid composition, with a high proportion of Gly, Glu and Gln residues, ultimately contributing to their largely unstructured nature (Tunnacliffe and Wise, 2007) (Table 4.3).

Table 4.3 Amino acid composition of XhLEA2-3 (A), XhLEA2-6 (B) and XhLEA3-5 (C) proteins.

A

Number of amino acids: 142

Molecular weight: 14947.0

Theoretical pI: 9.82

Amino acid composition:

Ala (A)	6	4.2%
Arg (R)	6	4.2%
Asn (N)	8	5.6%
Asp (D)	6	4.2%
Cys (C)	0	0.0%
Gln (Q)	18	12.7%
Glu (E)	6	4.2%
Gly (G)	30	21.1%
His (H)	10	7.0%
Ile (I)	7	4.9%
Leu (L)	3	2.1%
Lys (K)	13	9.2%
Met (M)	3	2.1%
Phe (F)	1	0.7%
Pro (P)	2	1.4%
Ser (S)	13	9.2%
Thr (T)	7	4.9%
Trp (W)	0	0.0%
Tyr (Y)	2	1.4%
Val (V)	1	0.7%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

B

Number of amino acids: 101

Molecular weight: 11217.9

Theoretical pI: 6.97

Amino acid composition:

Ala (A)	1	1.0%
Arg (R)	8	7.9%
Asn (N)	6	5.9%
Asp (D)	4	4.0%
Cys (C)	0	0.0%
Gln (Q)	8	7.9%
Glu (E)	7	6.9%
Gly (G)	22	21.8%
His (H)	5	5.0%
Ile (I)	1	1.0%
Leu (L)	4	4.0%
Lys (K)	3	3.0%
Met (M)	3	3.0%
Phe (F)	0	0.0%
Pro (P)	7	6.9%
Ser (S)	9	8.9%
Thr (T)	0	0.0%
Trp (W)	0	0.0%
Tyr (Y)	12	11.9%
Val (V)	1	1.0%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

C

Number of amino acids: 198

Molecular weight: 20638.3

Theoretical pI: 5.21

Amino acid composition:

Ala (A)	44	22.2%
Arg (R)	1	0.5%
Asn (N)	4	2.0%
Asp (D)	13	6.6%
Cys (C)	1	0.5%
Gln (Q)	3	1.5%
Glu (E)	23	11.6%
Gly (G)	13	6.6%
His (H)	2	1.0%
Ile (I)	2	1.0%
Leu (L)	10	5.1%
Lys (K)	29	14.6%
Met (M)	13	6.6%
Phe (F)	0	0.0%
Pro (P)	2	1.0%
Ser (S)	13	6.6%
Thr (T)	12	6.1%
Trp (W)	2	1.0%
Tyr (Y)	2	1.0%
Val (V)	9	4.5%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

4.4.4 PCR amplifications

PCR conducted with the primers designed for cloning XhLEA2-3, XhLEA2-6 and XhLEA3-5 cDNAs produced DNA bands with more or less the expected sizes of 429bp, 306bp and 597bp, respectively (Figure 4.2). The DNA of Xhlea3-5 migrated through the gel to a slightly higher position than anticipated (Figure 4.2, Lanes 16-19). The PCR products were successfully ligated, each into similarly restriction digested pET-21a(+)(his).

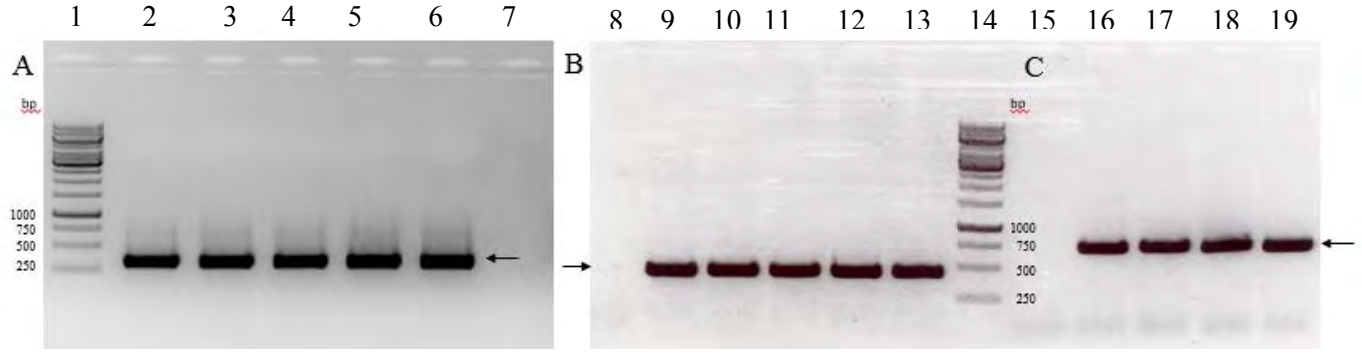


Figure 4.2 Confirmation of successful LEA gene amplification of XhLEA2-6 (A), XhLEA2-3 (B) and XhLEA3-5 (C) and insertion of *HindIII* and *XhoI* enzyme restriction sites by end-point PCR.
Lanes 1 and 14: O′GeneRuler 1 kb DNA Ladder (Thermo Scientific); Lanes 7, 8 and 15: no template controls; Lanes 2-6: XhLEA2-6 gene amplification (~306bp); Lanes 9-13: XhLEA2-3 gene amplification (~429bp); Lanes 16-19: XhLEA3-5 gene amplification (~597bp). The black arrows indicate the bands representing the entire amplified gene for each LEA clone.

4.4.5 Restriction enzyme digests using *HindIII* and *XhoI*

Restriction enzyme digests of all three LEA target genes and the pET-21a(+)(his) vector, proved successful after electrophoresis of digested products on an agarose gel (Figures 4.3 and 4.4, respectively). The digest using DNA from XhLEA3-5 was the first digestion conducted, tested overnight versus 1 hour, both times using the FastDigest enzymes, *HindIII* and *XhoI*. The digestion overnight depicts no product (Figure 4.3C, Lanes 11-13), possibly due to total digestion of the DNA, whereas digestion for 1 hour proved successful (Figure 4.3C, Lanes 14-16). Primer dimers remained in the lanes where the overnight digestion was loaded. This could be the remains of a semi-successful PCR product clean-up using a commercially available kit. It was therefore decided to digest the DNA of the remaining two target genes for 1 hour, and this proved successful For XhLEA2-6 and XhLEA2-3 (Figure 4.3, A and B, Lanes 2 and 3 and 4-8, respectively).

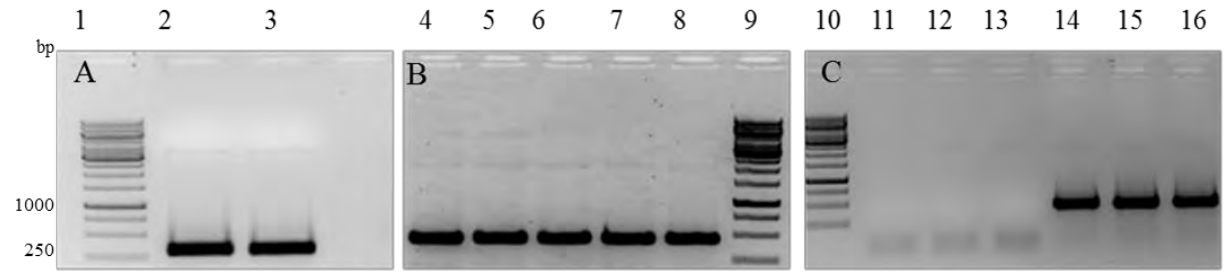


Figure 4.3 Restriction enzyme double digest of the individual PCR cleaned up reactions of amplified target genes XhLEA2-6 (A), XhLEA2-3 (B) and XhLEA3-5 (C).
Lanes 1, 9, 10: O′GeneRuler 1 kb DNA Ladder (Thermo Scientific); Lanes 2, 3: double digest of XhLEA2-6 PCR product; Lanes 4-8: double digest of XhLEA2-3 PCR product; Lanes 11-13: double digestion overnight and Lanes 14-16:1 hour double digestion, both reactions of XhLEA3-5 PCR product.

Single and double digestions of the pET-21a(+)(his) vector DNA proved successful and the individual bands from the double digestions (lower polypeptide bands in Figure 4.4, Lanes 5-7) were excised, pooled and gel purified for subsequent ligations.

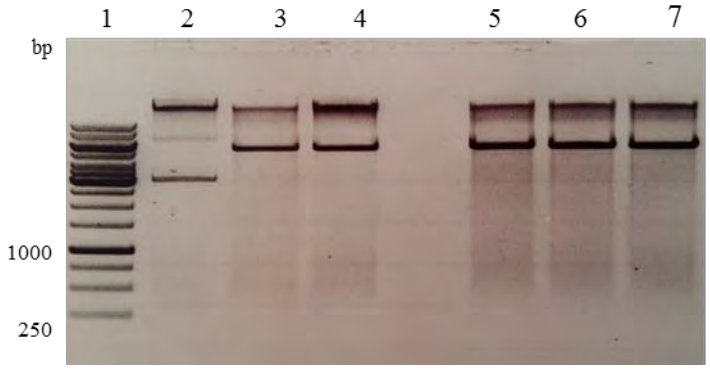


Figure 4.4 Single and double restriction enzyme digests of the vector pET-21a(+)(his) DNA.
Lane 1*:* O′GeneRuler 1 kb DNA Ladder (Thermo Scientific); Lane 2: undigested vector DNA; Lanes 3 and 4: 2.8µg vector DNA single digests using *HindIII* and *XhoI* restriction enzymes respectively and Lanes 5-7: 4µg vector DNA double digests using *HindIII* and *XhoI*.

4.4.6 Transformation of competent cells and colony PCR

Transformation of DH5α cells was successful. This was demonstrated by colony-PCR (Figure 4.5), from which the respective recombinant plasmids of XhLEA2-3, XhLEA2-6 and XhLEA3-5, were isolated. Distinct bands representing each gene were seen at the correct molecular weight (identified by the arrows for each colony PCR for each gene, Figure 4.5). The transformation of preferred expression host cells, BL21 (DE3) pLysS cells, was also successful when checked as above (data not shown). Sequence analysis of the constructs

corroborated these results and confirmed that XhLEA2-3, XhLEA2-6 and XhLEA3-5 were cloned in-frame into the pET-21a(+)(his) vector with no silent mutations (Appendix B3).

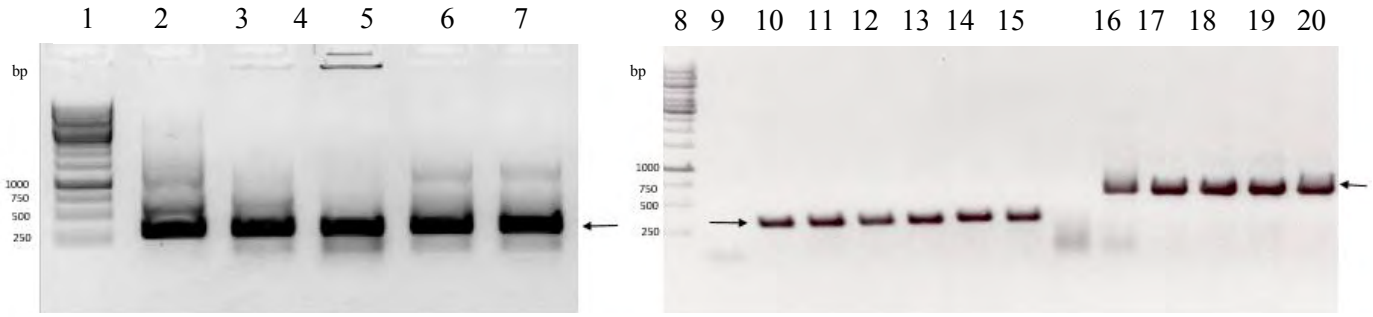


Figure 4.5 Colony screening by PCR for pET-21a(+)(his)- XhLEA2-6 (Lanes 2-7), pET-21a(+)(his)- XhLEA2-3 (Lanes 9-15) and pET-21a(+)(his)- XhLEA3-5 (Lanes 16-20) recombinant plasmids in DH5α cells using forward and reverse primers for each individual LEA gene. Lanes 1 and 8: O'GeneRuler 1 kb DNA Ladder (Thermo Scientific). The black arrow indicates the band representing the entire gene for each individual LEA

4.4.7 Expression and purification of recombinant his-tagged proteins of XhLEA2-3, XhLEA2-6 and XhLEA3-5.

When the crude and purified protein samples for the individual three target proteins were visualised on an SDS-PAGE gel, interestingly, the his-tagged recombinant protein of XhLEA2-3 was the only protein of the three that was evident in both its crude and purified form (Figure 4.6, Lanes 6 and 7). Western blot analysis using the anti-his-tag antibody (HRP conjugate) successfully detected the his-tagged recombinant XhLEA2-3 protein, again in both the crude and purified fraction, and the protein was present at the correct molecular weight (~15kDa, with the addition of the ~1kDa 6 x His tag) (Figure 4.7, Lanes 6 and 7). The positive control purified protein of XhLEA5a was confirmed by SDS-PAGE analysis and western blot, at the correct molecular weight (~13kDa) (Figure 4.6, Lanes 8 and 9; Figure 4.7, Lanes 8 and 9). There were no non-specific bands seen on the blot, which suggests efficient purification of the recombinant proteins of XhLEA2-3 and the positive control, XhLEA5a.

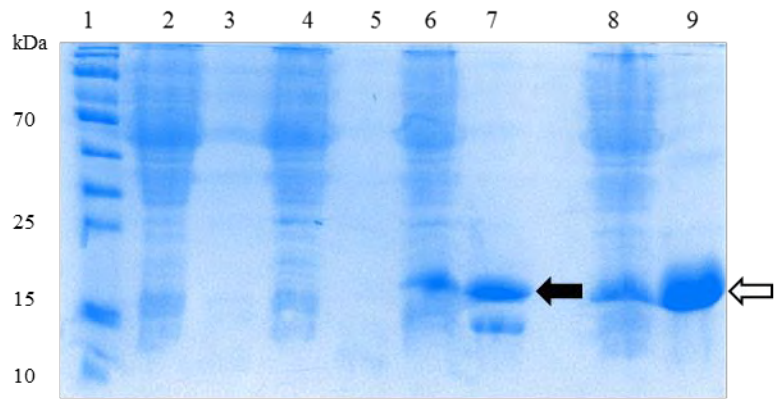


Figure 4.6 Purification of XhLEA2-6 (~12kDa), XhLEA3-5 (~22kDa) and XhLEA2-3 (~15kDa) recombinant proteins using the Protino® Nickel NTA Histidine Tag affinity purification kit. Protein samples from *E.coli* BL21 (pET-21a(+)(his)): XhLEA2-6, XhLEA3-5 and XhLEA2-3 respectively were separated using a 12% SDS-PAGE. Lane 1: PageRuler Prestained Protein Ladder (Thermo Scientific) (6µl); Lanes 2, 4 and 6: 20µg crude total protein from XhLEA2-6, XhLEA3-5 and XhLEA2-3, respectively and Lanes 3, 5, 7: 4µg concentrated, purified recombinant protein from XhLEA2-6, XhLEA3-5 and XhLEA2-3 in 1 x PBS, respectively; Lanes 8 and 9: Positive control, crude and purified form of the recombinant XhLEA5a (black outlined arrow). The solid black arrow represents the crude and purified recombinant protein of XhLEA2-3. The gel was stained with Coomassie blue staining solution.

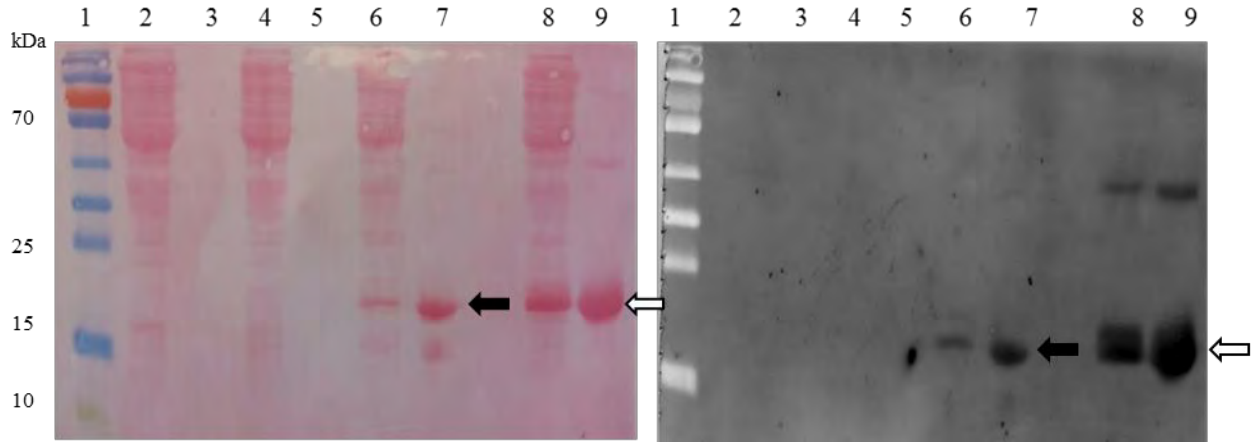


Figure 4.7 Confirmation of the presence of XhLEA2-6 (12kDa), XhLEA3-5 (~22kDa) and XhLEA2-3 (~15kDa) recombinant proteins by western blotting. Recombinant proteins expressed using pET-21a(+)(his) vector were separated on 12 % SDS gel and chemiluminescently identified after incubation with monoclonal anti-his tag antibody (1:40 000). The Ponceau S stain of the membrane post transfer (A) and the result of the western blot detection (B) are shown above. Lane 1: PageRuler Prestained Protein Ladder (Thermo Scientific) (6µl), Lanes 2, 4, 6: 2µg crude total protein from XhLEA2-6, XhLEA3-5 and XhLEA2-3, respectively and Lanes 3, 5, 7: 1µg concentrated, purified recombinant protein from XhLEA2-6, XhLEA3-5 and XhLEA2-3, respectively. Lanes 8 and 9: crude and purified recombinant protein of XhLEA5a (positive control).

It is worth recording here that LEA-recombinant proteins generated from two Group 1 LEAs (XhLEA1-4S1 and XhLEA1-1S2) were successfully generated using the T7 tag Affinity

Purification Kit (Ginbot, 2011). This method of purification proved troublesome in initial cloning work conducted by Dr Nashied Peton- on a subset of the 21 LEA-like proteins described in Chapter 2, whereas the use of the modified pET-21a+ vector: pET-21a(+)(his), was successful. It was for this reason, that the protocol involving the use of a modified Pet21a (+) vector would work for the three chosen target genes used in this study. Protein bands were evident in all three lanes of crude lysates for each target protein and thus the problem for the two remaining LEA-like proteins is likely to lie in the purification step.

Following the analysis of the proteins expressed, extracted and purified above, it was decided to attempt to optimise the protein expression protocol with regards to the length of IPTG induction and the purification step followed by SDS-PAGE and western blot analysis of the three target proteins produced. This optimisation protocol was included as a final step in this work in order to assess whether a longer induction of expression was needed and whether a more analytical protein purification protocol would assist in detecting where the problems were possibly occurring. A 6-hour-long and an overnight IPTG induction was completed, in contrast to the 4 hour induction time previously used. Recombinant his-tagged proteins of XhLEA2-3, XhLEA2-6 and XhLEA3-5 were re-expressed and extracted as per previous protocols (Section 4.3.12 and 4.3.13) and SDS-PAGE analysis with the crude extracts showed successful protein expression after both the 6 hour and overnight IPTG induction length, for each of the three target proteins (Figure 4.8).

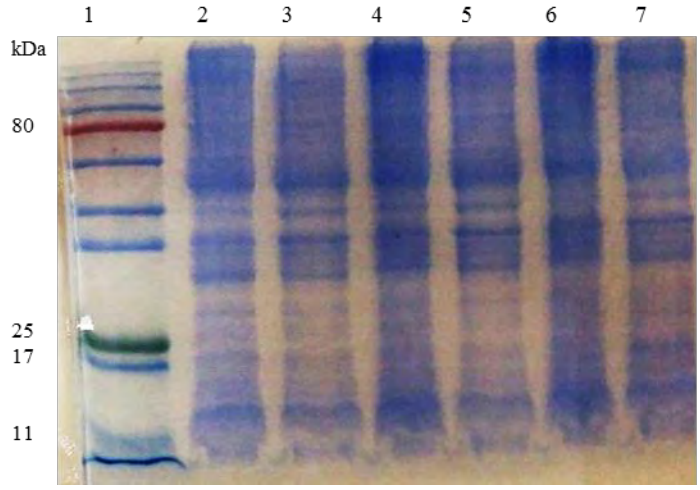


Figure 4.8 SDS-PAGE representing the crude protein extracts of his-tagged recombinant proteins produced during a 6 hour induction and an overnight induction with IPTG.
Lane 1: 6µl molecular weight marker (Colour Prestained Protein Standard, Broad Range (11-245kDa) (NEB, P7712S); Lanes 2-7: Crude extracts of each of three target proteins in order of XhLEA2-6, XhLEA3-5 and XhLEA2-3, after 6 hours and an overnight induction length with IPTG, respectively.

The amendments made to the protein purification protocol of the crude protein lysates shown in Figure 4.8 were as follows: post equilibration of the column, 2ml of the soluble crude extract was pipetted directly onto the Ni-TED column and the fractions of flow through were collected. Individual wash and eluted fractions were collected by means of gravity flow. 20µl of each purification step was used for analysis by SDS-PAGE. Proteins were retained in the 1 x elution buffer and the concentration and buffer exchanging step into 1 x PBS was not conducted in this optimisation step. The individual protein samples were heated at 95ºC in sample application buffer (SAB) prior to loading the SDS-PAGE gel to ensure complete dissociation of protein samples and successful downstream analysis using electrophoresis. Unfortunately, downstream analysis of the purified forms of XhLEA2-6, XhLEA3-5 and XhLEA2-3, using SDS-PAGE and western blot proved was not successful (data not shown).

In future studies attempting to successfully express and purify the three target proteins chosen here, as well as those that make up the rest of the 21 LEA-like proteins described in Chapter 2, careful consideration needs to take place prior to experimental planning and execution of protocols. Each individual target protein needs to be assessed with respect to molecular weight (kDa), pI, solubility and half-life, in order to optimise conditions specifically designated for that protein. Induction temperatures and times will need to be optimised for each individual LEA-like recombinant protein. It may be the case that some induce best at 37ºC and others at lower temperature ranges, such as 15-25 ºC. The purification of these LEA recombinant proteins can be adjusted in future studies by analysing the effect of different concentrations of imidazole on the elution of these proteins from the Ni-TED columns. It is a possibility that the buffers used may need to be optimised, with regard to imidazole concentration and/or the pH relating to the protein pI, for each of the three LEA-like proteins chosen. However, previous proteins expressed in this set of 21, with varying pI's, all showed successful purification with a single kit buffer pH of 8.0 (data not shown). Another challenge was possibly the presence of inclusion bodies. These bodies are known to accumulate during high level expression of many recombinant proteins in *E. coli*. If this was the case, future studies on these proteins need to focus on the pellet generated after the sonication and centrifugation steps of the extracted protein samples. A possible protein extraction method, incorporating Guanidine:HCl, and further purified by gel filtration protocols, as described by Palmer and Wingfield (2012), may need to be implemented. The boiling step introduced into the protein purification procedure was a modification of the Ni-TED protein purification protocol for the purification of his-tag proteins. This exposure to a high temperature has been

a common feature in the purification methodologies reported for LEA proteins (Russouw *et al.*, 1997). This step was intended to concentrate and maximise the abundance of the heat-stable recombinant LEA proteins, while minimizing the concentration of unnecessary, unwanted proteins. The heat stability of LEA proteins, owing to their high hydrophilicity, was tested here with the incorporation of the boiling step. This step can facilitate the purification protocol from bacterial lysates or plant tissue samples, for all future studies. Downstream affinity purification procedures will also be made easy by the removal of all heat-sensitive proteins. Another step of purification may need to be implemented and this could be in the form of size exclusion chromatography, ion exchange or high performance liquid chromatography (HPLC), but it may be essential to dialyse the samples to remove unwanted contaminants such as the imidazole prior to the forms of purification listed above. It will be essential to affinity purify these proteins using commercially available kits such as the Protino® kit used in this study, followed by dialysis to remove unwanted compounds and finally purification using a procedure such as HLPC, to identify, quantify and purify the individual components of the mixture. The implementation of all these aforementioned steps, will ensure the final purified form of the target protein is ready for any form of downstream analysis and success thereof. Further confirmation of affinity purified recombinant LEA-like proteins may be obtained by the use of Mass Spectrometry (MS) as was demonstrated by Ginbot (2011), who used this approach to confirm the identities of the recombinant proteins produced.

4.5 Conclusion

In conclusion, the full length cDNA nucleotide sequences of XhLEA2-6, XhLEA3-5 and XhLEA2-3 were successfully amplified by traditional PCR methods, using primers that were designed to contain unique restriction enzyme sites at their 5' ends. This enabled the successful restriction digestion of the amplified gene products, along with that of the plasmid vector, and ligation of each of the genes into the expression vector pET-21a(+)(his) for subsequent protein expression and purification. Potential recombinant colonies were screened using colony PCR and restriction enzyme mapping and those that produced positive bands were sent for sequencing. Sequencing analyses indicated successful, in-frame cloning of each of the individual LEA-like genes into the cloning vector pET-21a(+)(his). Protein expression and purification of the three LEA-like proteins proved challenging. Where using the original method of extraction, expression and purification and the conventional 4 hours of IPTG

induction, resulted in the successful purification of the his-tag recombinant protein, XhLEA2-3, the optimisation experiments using a longer induction time of 6 hours or overnight, removing the heat-stable step and using a different purification protocol, proved unsuccessful (data not shown).

Chapter 5

Preliminary structural analysis of XhLEA2-6, XhLEA3-5 and XhLEA2-3 using Circular Dichroism (CD)

5.1 Introduction

The rapid characterization of new proteins is of great importance for the fields of proteomics and structural genomics (Greenfield, 2009). Gaining knowledge and evidence of the structure of a biological molecule forms an important part of characterization, as structure gives a clue to biological function. Protein crystallization has been traditionally used to study protein structure, however, LEA proteins studied to date remain largely unstructured in the hydrated state and hence, crystallization approaches prove difficult. In recent years, alternate methods of structural characterisation, such as Circular Dichroism (CD), have been employed to determine LEA protein structure. CD is defined as the difference in the absorption of left-handed circularly polarised light and right-handed circularly polarised light. This occurs when a molecule or protein contains one or more light-absorbing groups known as chromophores in the far-UV CD spectra, in the range of 180-250nm. CD spectroscopy therefore measures the differences in this absorption, which is a result of structural asymmetry. For the analysis of proteins, a solution of 20-200µl containing highly purified protein of approximately 50µg/ml to 1mg/ml protein is required. In the far-UV range, the chromophore is the peptide bond, and a signal corresponding to particular structures arises when light interacts with a regular, folded environment. CD monitors the characteristic spectra displayed by three main protein structures: α-helices, β-sheets and random coils, across the wavelengths specified above (Figure 5.1). It is crucial that the utmost consideration is given to sample condition as many buffer components absorb in the requisite wavelengths and protein or DNA concentration greatly affects the signal to noise ratio.

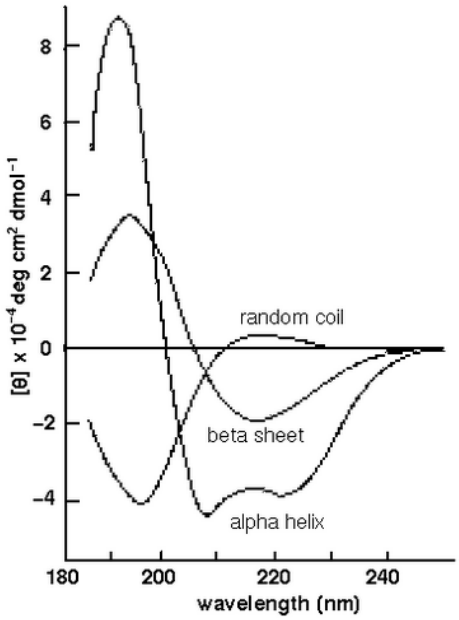


Figure 5.1 The far UV CD spectra characteristic of alpha-helix, beta-sheet and random coil structures within a protein sample. The x-axis represents the typical wavelengths in nm from 180-250nm. The y-axis represents molar ellipticity (Θ).
(http://www.cryst.bbk.ac.uk/PPS2/assignments/A1/secCD.gif)

Structural studies of LEA proteins using CD, have shown that these proteins remain largely unstructured when in conventional buffer solutions, and tend to display a spectrum similar to that of random coil (Soulages *et al*., 2002; Ginbot, 2011; Shih *et al*., 2010). There are also bioinformatics tools that are available online, such as PONDR (Prediction of Naturally Disordered Regions), enabling the prediction of the extent of disorder in LEA protein sequences. This programme is made available at the developer's website, http://www.pondr.com, and has been used successfully in numerous studies when predicting protein disorder (Obradovic *et al*., 2005; Peng *et al*., 2006; Fuxreiter *et al*., 2007, reviewed by Ginbot, 2011). Both CD analysis and online PONDR were used in this study to investigate the structure of XhLEA2-3, XhLEA2-6 and XhLEA3-5 proteins.

5.2 Aim

The aim of this chapter was to

i) assess the protein structure of the three target LEA proteins using PONDR and;

ii) to conduct preliminary structural analyses on the heat-stable purified fractions of the three target proteins using CD.

5.3 Methods and materials

5.3.1 Prediction of protein structure from the amino acid sequence using PONDR

PONDR was used to predict the overall structure of XhLEA2-3, XhLEA2-6 and XhLEA3-5 amino acid sequences. The VL-XT platform was used for analysis (a merger of three predictors, one trained on Variously characterised Long disordered regions and two trained on X-ray characterised Terminal disordered regions). FASTA formatted amino acid sequences of each of the three LEA proteins was entered into the analysis tool. The pictorial and descriptive data output was analysed for known LEA protein features and compared with published data on similar proteins.

5.3.2 Heat-stable purified recombinant proteins used for CD analysis

Affinity purified and column concentrated recombinant his-tagged proteins of XhLEA2-3, XhLEA2-6 and XhLEA3-5, described in Chapter 4 (Section 4.3.11 and 4.3.12), were used for subsequent preliminary structural analysis by CD.

5.3.3 In-solution structural investigation

To investigate the in-solution structure of the purified recombinant proteins of XhLEA2-3, XhLEA2-6 and XhLEA3-5, CD measurements were taken from aqueous solution of these proteins in 10 mM PBS (1 X PBS) using the JASCO J-810 Spectropolarimeter in the range of 185 to 260nm wavelength as described in the study of other LEA proteins (Soulages *et al*., 2003; Shih *et al*., 2004 and 2010; Mouillon *et al*., 2006). Measurements of millidegrees sensitivity were taken from sample concentrations of 0.7 - 1 mg/ml using a standard quartz cuvette with a path length of 1mm. Protein samples were optimised to dilution of 1 in 10, 1 in 5 and 1 in 5, for XhLEA2-6, XhLEA3-5 and XhLEA2-3, respectively, enabling the Spectropolarimeter to read into the lower wavelengths (nm). The generated CD data was exported to Excel worksheet and used to plot molar ellipticity against wavelength in nanometers (nm). The resulting CD spectra were compared with typical spectra of the three common forms of protein secondary structure, α- helical, β-sheet and random coil. The final exchange buffer (1 x PBS) and the original elution buffer (1 x elution buffer, Protino® Ni-TED Histidine Tag affinity purification kit (Macherey-Nagel, Germany)) were tested for absorbance in the far-UV wavelength and the original elution buffer was found to absorb, therefore care was taken to ensure the buffer exchange step into 1 x PBS minimized carry-over of the original elution buffer. A critical step in the analysis of proteins using CD is the calibration of the equipment on a regular basis. The use of crystallised CSA ((1S)-(+)-

Camphor-10-sulfonic acid) is commonly used as a calibration standard and will ensure that the ellipticity values and the wavelengths remain correct. This standard was run along with the albumin standards used in my protein runs, and it was concluded that this was slightly off the mark indicating the need for a calibration.

5.4 Results and discussion

5.4.1 Prediction of protein structure from the amino acid sequence using PONDR

Proteins with a PONDR score of above 0.5 are classified as disordered. The PONDR algorithm predicted that two of the three LEA-like proteins – XhLEA2-3, and XhLEA2-6, were likely to be disordered proteins (Figure 5.2, A and B). The third protein, XhLEA3-5, could not be classified as disordered (Figure 5.2, C). The disordered nature in all three LEA proteins was not continuous and the various disordered regions showed different disorder strength. XhLEA2-3, was found to have 113 residues out of a total of 142 predicted residues that were disordered. The longest disordered region for this protein was 70 residues resulting in an overall disorder percent of 79.58% with an average disorder score of 0.7627. For XhLEA2-6, 66 residues of the total 101 predicted residues comprised the longest disordered region resulting in an overall disorder percent of 74.26% with an average disorder score of 0.6469. 64 residues out of a total of 198 predicted residues for XhLEA3-5 were disordered, with a longest disordered region of 40 residues resulting in an overall percentage of disorder of 32.32% with an average disorder score of 0.3819. The different disordered regions displayed different disorder strengths for each individual LEA protein, with strengths ranging in XhLEA2-3 from 0.6212-0.8972, in XhLEA2-6 from 0.6123-0.8136 and in XhLEA3-5 from 0.5155-0.8718. The two putative dehydrins, XhLEA2-3 and XhLEA2-6 had similar structural disorders which is expected with proteins from the same structural and functional group. XhLEA3-5 displayed a more ordered nature with one large region of disorder in the middle of the protein.

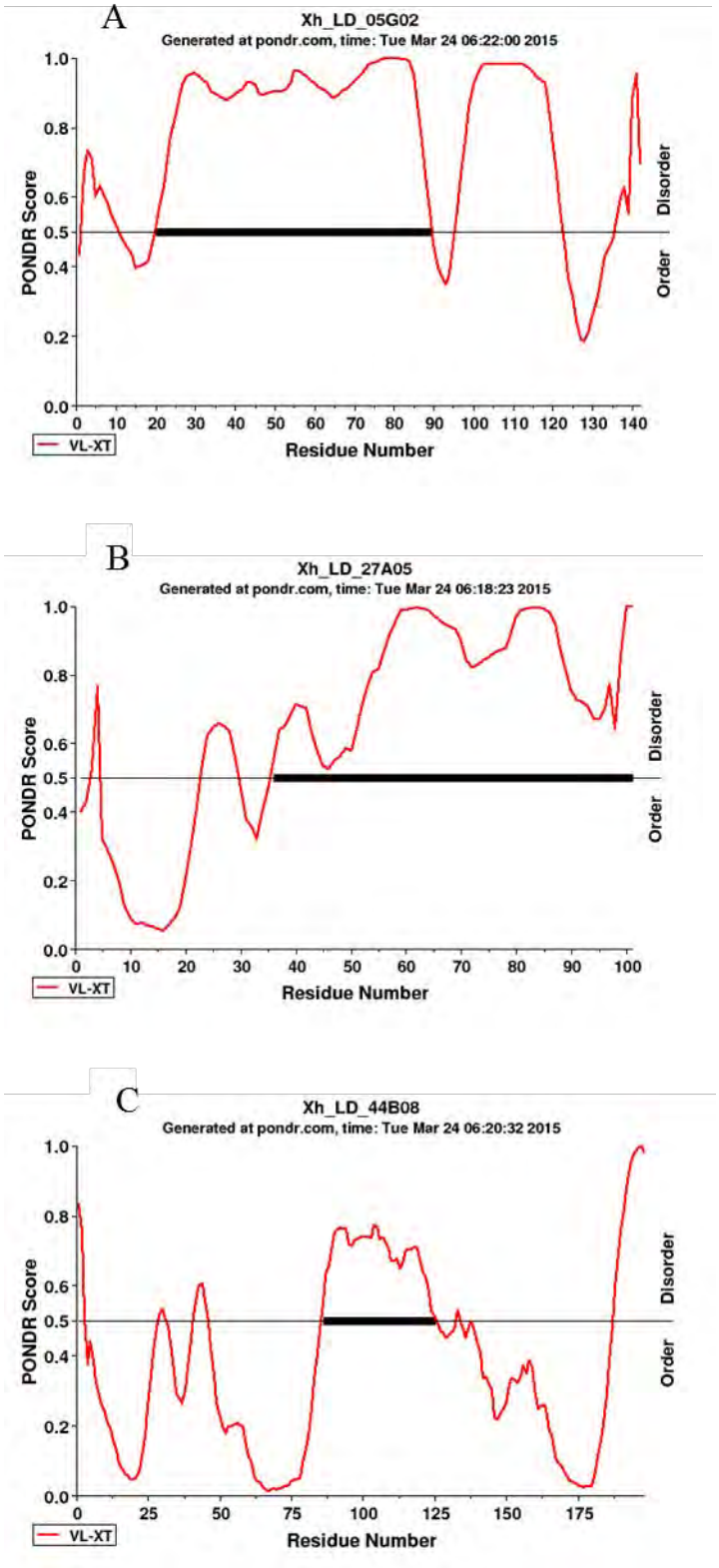


Figure 5.2 Prediction of structural disorder of XhLEA2-3 (originally Xh_LD_05G02) (A), XhLEA2-6 (originally Xh_LD_27A05) (B) and XhLEA3-5 (originally Xh_LD_44B08) (C) proteins using PONDR. PONDR scores of residues are indicated on the vertical axis against residue number on the horizontal axis. Residues making up a continuous disordered region are indicated by horizontal bold line.

PONDR has been used successfully to predict protein disorder, in many previous studies (reviewed by Ginbot, 2011). It is reported to be the best known tool for the prediction of the intrinsic disorder within given protein sequences (Oxford Protein Production Facility, 2007). As previously mentioned in Section 5.3.1, the PONDR-VL-XT platform and predictor was selected from the available options on this website as it integrates three feedforward neural networks trained with various settings (http://www.pondr.com/pondr-tut2.html). A previous study conducted on two dehydrins, ERD10 and ERD14 (Early Response to Dehydration), using PONDR demonstrated the prediction of their highly disordered state (Kovacs *et al*., 2008). This is in agreement with the results for the two dehydrins studied (Section 5.4.1) – XhLEA2-6 and XhLEA2-3. A study conducted on a multifunctional Group 3 LEA protein from maize, also demonstrated a highly disordered nature with the use of PONDR, supporting the disordered results obtained for XhLEA3-5, used in this study (Liu *et al.,* 2013). Unfortunately, CD analysis of this Group 3 LEA protein did not accompany the bioinformatics analysis using PONDR. It is with hope that, with any study, the PONDR structural prediction results are supported by CD spectroscopy results, where the latter, in the case of certain LEA proteins, will show the random coil disordered structure in these proteins.

5.4.2 Structural studies

It was decided to attempt to conduct CD analysis on the purified fractions of the three target proteins produced in Section 4.3.11 and 4.3.12, despite the unsuccessful detection of the crude and purified fractions of XhLEA2-6 and XhLEA3-5, with SDS-PAGE and subsequent western blot analysis. The preliminary CD analysis of the column purified and concentrated XhLEA2-3, XhLEA2-6 and XhLEA3-5 proteins in a 1 x PBS buffer, only showed a slight pattern for XhLEA3-5 and more so for XhLEA2-3, both similar to that of an alpha-helix structure in the range of 220 – 260 nm wavelength. For both the recombinant proteins of these two LEAs, the pattern was only able to be read until 220nm, indicating the possible presence of contaminating chromophores carried over from the original elution buffer, that ultimately limited the generation of ellipticity values in the lower wavelength range (185-220nm).

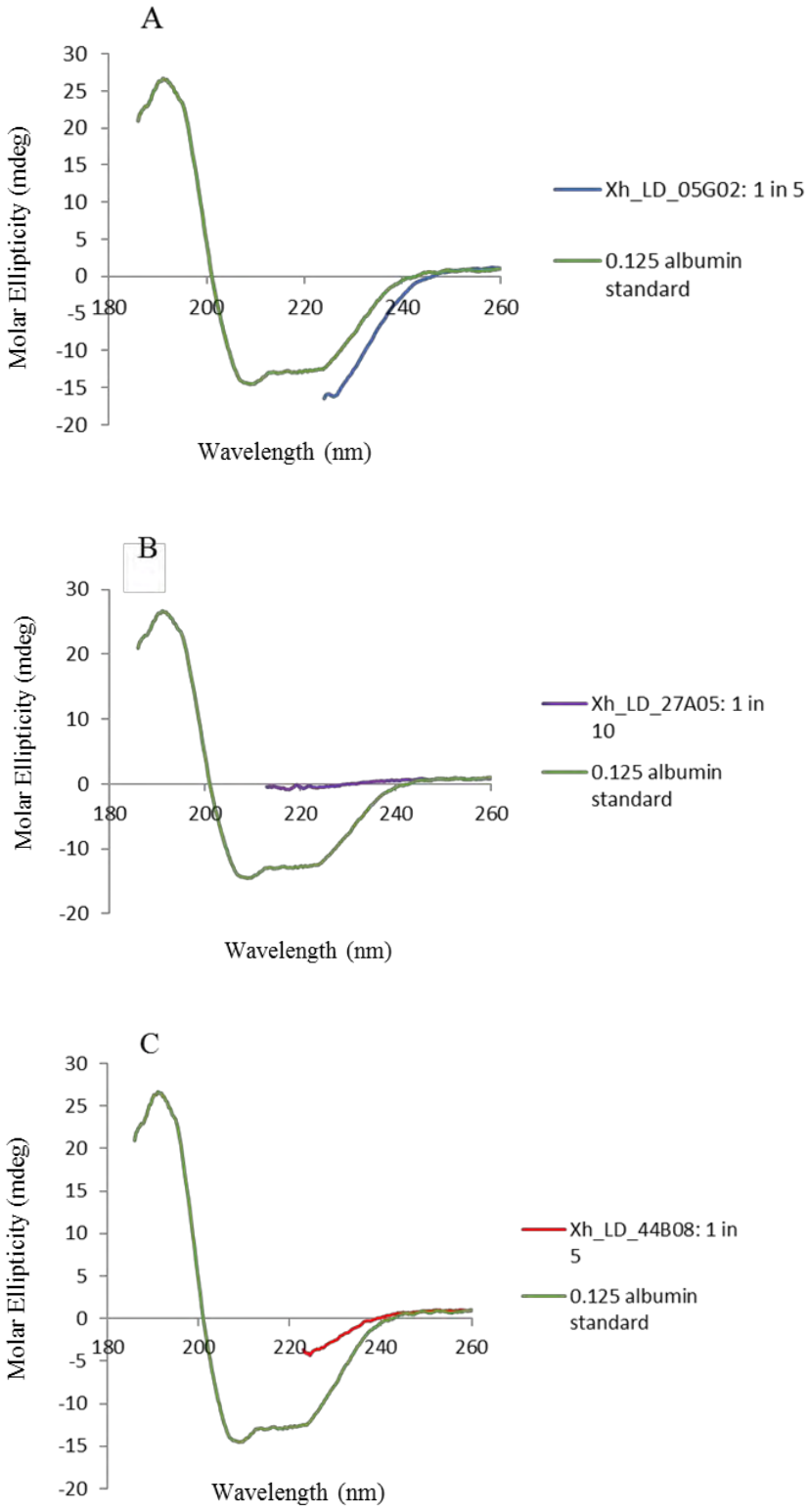


Figure 5.3 CD spectrophotometer readings of XhLEA2-3 (originally Xh_LD_05G02) (A), XhLEA2-6 (originally Xh_LD_27A05) (B) and XhLEA3-5 (originally Xh_LD_44B08) (C) in the range of 180 to 260nm wavelength. Blue, purple and red lines are the individual proteins in 10mM PBS (1 x PBS), respectively. Green line, 0.125 albumin protein standard. The CD signal and wavelength (nm) are indicated on the Y and X axis, respectively.

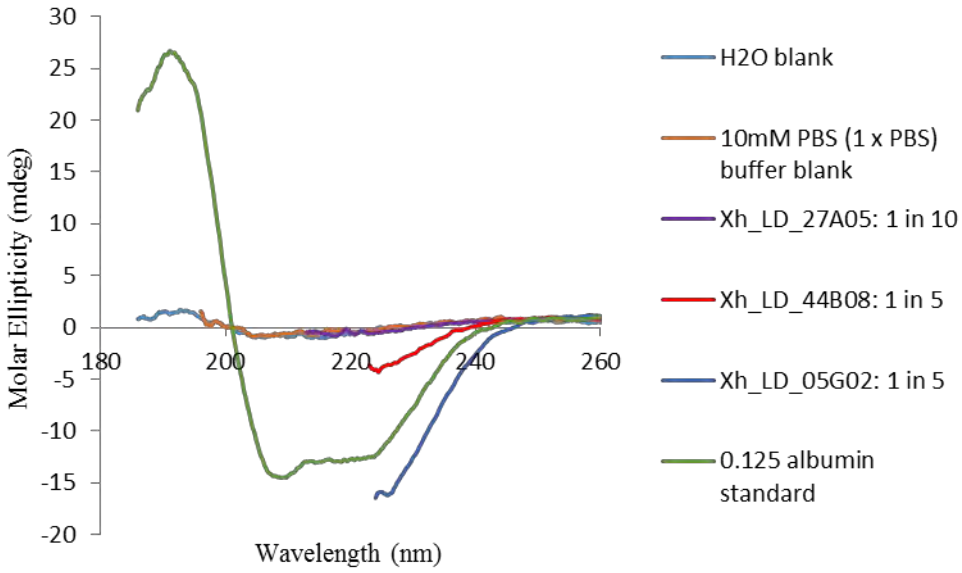


Figure 5.4 CD spectrophotometer readings of the three proteins combined in the range of 180 to 260nm wavelength. Included in this graph are the water (H_2O) blank (light blue line) and the 10mM PBS blank (orange line).

A previous study conducted by Mouillon *et al*., (2006), on *Arabidopsis* (*A. thaliana*) dehydrin proteins, described the structural nature of these proteins, using CD analysis, to be "poorly structured with a low content of secondary structure". This observation is consistent with the understanding that the predominantly hydrophilic nature of these dehydrins does not enable the formation of an extensive hydrophobic core within these proteins, which is typical for folded proteins. Experimental structural analysis of several other group 2 LEA proteins, has also highlighted the lack of a well-defined secondary structure and predominant random coil formation, when in solution (Battaglia *et al*., 2008). This was also recorded in studies of Dsp16 (YSK_2) from the resurrection plant *Craterostigma plantagineum* (Lisse *et al*., 1996), dehydrin proteins obtained from cowpea (*Vigna unguiculata* L. Walp.) seeds (Ismail *et al*., 1999), a DHN1 obtained from maize (*Zea mays*) kernels (Koag *et al*., 2003), rGmDHN1 (Y_2K) from soybean (Glycine max) (Soulages *et al*., 2003) and ERD10 (SK_3) from mouse-ear cress (*Arabidopsis*: Bokor *et al*., 2005). Interestingly and in contradiction, our preliminary data of the dehydrin, XhLEA2-3, indicated an alpha-helix structural trend for this protein (similar to the structure of the albumin protein standard).

CD analysis and IR spectroscopy of various Group 3 LEA proteins in previous studies, have indicated, as with the dehydrins, that these proteins are mostly devoid of secondary structure and remain largely in a random coil formation in solution (Battaglia *et al*., 2008). Similar to

the data obtained for XhLEA2-3, the data obtained for XhLEA3-5 seemed to be following the pattern of an alpha-helix structure. It is worth recording here, that many authors have supposed Group 3 LEA proteins to be largely alpha-helical structured. This has been based on the structural analyses performed using secondary-structure prediction programs (Tunnacliffe and Wise, 2007). However, these preliminary structural observations cannot be concluded with confidence and may even depict a random coil formation when the data is able to be fully visualised through the entire wavelength spectrum, up until 185nm.

Ginbot (2011), successfully demonstrated the random coil nature of two affinity purified Group 1 LEA recombinant proteins. This study demonstrated that these two proteins remained unstructured in a solution of 1 x PBS, whereas upon the replacement of the elution buffer with the co-solvent TFE, an increased α-helical content of 45 and 55%, for each protein, was observed. The CD results obtained were similar to the published data of unstructured LEA proteins (Shih *et al*., 2004 and 2010). Unfortunately, due to time constraints, we were unable to successfully purify two of the three target proteins used in this study, for analysis using CD with the use of affinity purification methods. Due to the probability of chromophore contamination from the elution buffer, HPLC purification would be the next logical step in obtaining a sufficiently pure protein sample for structural analysis, and will play a big role in the structural analysis of these, and other LEA-like proteins from the list of 21, in future studies.

It has been stated that environmental conditions can affect the folding of proteins, and several LEA proteins tend to take on a more structured form when dried (Tunnacliffe and Wise, 2007). Therefore, the buffer replacement with a co-solvent, such as TFE, to future purified protein solutions, will aid in reducing water availability within the sample and enabling the creation of a condition similar to that which would be expected of a dehydrating cytoplasm. This will hopefully shed light on the structural nature that these proteins adopt in solution, and under dehydration conditions. Overall, it is necessary to relate this structural information to LEA proteins in resurrection plants, and how the adoption of a more defined structure in water limiting scenarios may possibly play a role in aiding the survival of these plants during desiccation.

5.5 Future work considerations

Future optimisations for the structural analysis of recombinant LEA/LEA-like proteins using CD need to be addressed in this final future work considerations section. The preparation of protein samples for specific cuvette widths is crucial in obtaining high quality data. For typical measurements in a 0.1 cm cell, depending on the buffer (See Table 1, Greenfield, 2009) solutions of 0.05 to 0.2 mg/ml protein need to be made. For measurements in 0.01 to 0.02 cm cells 0.2 to 1 mg/ml protein concentrations should be used and for 1 cm cells 0.005 to 0.01 mg/ml protein should be used. As depicted in the results for the individual protein samples prepared, these concentrations may have not been prepared accurately for the sized cuvettes used in this study. Protein concentration needs to be determined in the most accurate way and according to Greenfield (2009), standard methods using Bradford dye or the Lowry method, produce different results for different proteins. The most accepted concentration determination method is therefore quantitative amino acid analysis, using the concentrations of the stable amino acids eg: alanine and lysine, in order to calculate the concentration of the entire intact protein. Another important factor is protein purity. The recombinant proteins generated, should always be of the highest purity. Commercially available kits used to purify proteins, such as the Protino® Nickel NTA Histidine Tag affinity purification kit used in this study, may contain compounds and reagents that are not compatible with downstream analysis by CD, as they have a high absorption tendency. Examples of these compounds include citrates, imidazole and sodium chloride, all contained within the reagents used in this kit for protein extraction and purification. Even though the individual protein samples were buffer exchanged into 1 x PBS (10mM PBS), left over kit elution buffer, originally used to elute the proteins, may still be present in the samples, hindering the correct structural analysis using CD. It is explained by Greenfield (2009) that samples used for CD analysis must be at least 95% pure by the criteria of HPLC, mass spectroscopy or gel electrophoresis. These samples were not HPLC purified, as discussed in section 5.4.2, and thus future experiments should always ensure that this step is included prior to analysis by CD.

5.6 Conclusion

PONDR was effectively used to predict the degree of disorder using the amino acid sequences of the LEA-like proteins XhLEA2-6, XhLEA3-5 and XhLEA2-3 obtained from the ExPASy translate tool. The two dehydrin-like proteins of the three, XhLEA2-6 and XhLEA2-3, were shown to have a high degree of disorder, whereas XhLEA3-5 adopted a lesser degree of disorder and was not considered a disordered protein, having a PONDR score

less than 0.5. Unfortunately, PONDR structure predictions were not supported by the preliminary analysis of protein structure using CD spectroscopy. The purification of the three target recombinant proteins proved unsuccessful for XhLEA2-6 and XhLEA3-5, whereas XhLEA2-3 was successfully purified and confirmed at the correct molecular weight (kDa) by SDS-PAGE analysis. Hence, the preliminary structural analysis for XhLEA2-3 gave the most data, indicating a possible alpha-helical structure when in solution. However, the data remains unclear and inconclusive. Future studies need to focus on the optimisation of purification protocols and subsequent structural analysis within a maintenance buffer such as 1 x PBS, and a co-solvent, such as TFE, to act as a secondary structure promoter.

Chapter 6

General concluding remarks and the future scope for this study

Transcriptomic and proteomic approaches, amongst other types of studies, have shown that LEA proteins are important in plants. Their importance is not only in the context of plant development and reproduction, but particularly in stress responses. Over the last two decades, increasing information regarding the association of LEA proteins with resurrection plants and anhydrobiotic organisms, has been made available through these reported studies. Although different functions have been predicted and proposed for these enigmatic proteins, experimental evidence remains limited and their precise roles are yet to be defined.

In the current study, aimed at adding to this ever-growing knowledge of the relationship between LEA proteins and resurrection plants, the gene expression of 21 putative LEAs in leaves of the resurrection plant *X. humilis* during dehydration and rehydration, has been described. In addition, cloning procedures of 3 chosen LEAs, expression of recombinant his-tagged LEA proteins and preliminary structural analysis using CD, was conducted. The results obtained indicated the upregulation of all 21 putative LEAs in response to dehydration in *X. humilis* leaf tissues and the subsequent decline in their expression during rehydration. This correlation allows us to speculate that these proteins do contribute to the tolerance of this species to water deficit stress, although the mechanisms whereby they achieve this is still not known. The initial fact alone, that this many LEA genes are present, and are induced by water deficit in *X. humilis*, points to their significance during the desiccation stress response in this resurrection plant.

An initial aim of the study was to determine the nature of gene expression in root tissues of *X. humilis*. However, despite the use of two different protocols measuring gene expression, namely the relatively standard method of qPCR analysis and an alternative method involving ddPCR™, inconclusive results were obtained. Further evidence of the gene expression profile changes in this set of LEAs in root tissues, in response to dehydration and rehydration stress, needs to be obtained and investigated. In the first instance, protocols used need to focus on obtaining the highest quality and quantity of RNA from these recalcitrant tissues. It is proposed here that subsequent use of ddPCR™ should be attempted. A future alternative to working on soil potted *X. humilis* plants, could be to make use of an aeroponic growth system that could stimulate dehydration without the interference of soil particles and associated microbes. This would further help in ensuring there is no damage inflicted on roots during

sampling. The use of seedlings in such a system could furthermore reduce the degree of secondary thickening associated with resurrection plant roots in older plants, facilitating easier extraction of RNA.

Recombinant his-tagged proteins of XhLEA2-3, XhLEA2-6 and XhLEA3-5, were successfully generated with the use of *E. coli.* However, purification protocols were completely successful for only the putative LEA- XhLEA2-3. Subsequent preliminary structural analysis of purified recombinant XhLEA2-3 and XhLEA3-5 proteins, using CD, indicated an alpha-helix like structure. In order to accurately assess the structure of these LEA proteins and further confirm their LEA status, CD protocols need to be modified and tested on an affinity and HPLC purified form of the individual protein in both an elution buffer, and a co-solvent, such as TFE, to assess secondary structure in normal and dehydration-mimicked environments.

Future production of antibodies specific to this set of 21 LEAs could facilitate our understanding of the relationship between mRNA abundance and protein accumulation in *X. humilis,* when this plant is subjected to dehydration and subsequent rehydration. Finally, localization studies of this set of LEAs, using immunocytochemical methods, would enable understanding of the subcellular locations of LEAs and possible changes in these during dehydration and recovery therefrom. Determining whether they are present in the nucleus, cytoplasm, plasma membrane, mitochondria or vacuoles is essential in understanding their roles within organisms. It would also be interesting to see whether different stresses, such as freezing/osmotic or heat stresses, cause the upregulation of these LEAs and proteins. Such studies would add to the understanding of particular roles of the various LEAs in relation to stress tolerance in general. An ultimate goal would be to clone these LEAs into a suitable plant based vector for transformation into drought sensitive plants, in order to ascertain their role in plant water deficit (and other stress) tolerance.

We began this study by referring to this set of LEAs as “putative”, but with increasing bioinformatics analyses including the identification of conserved group domains, and a preponderance of hydrophilic amino acids, gene expression patterns in response to desiccation stress and structural analyses similar to that reported for LEAs studied to date, we feel that we can now accept these LEAs as true LEAs. This study has highlighted the importance of treating each LEA as an individual, and the need to implement empirically determined methodologies for each one. The study has illustrated this point, where using the

same methodologies to assess gene expression and to express and purify a subset of these LEAs gave different results for each LEA tested. This again highlights the need for individual consideration with respect to bioinformatics, gene expression changes, protein expression and purification, structural studies and finally, functional studies. Once the functions of these LEAs have been determined by future studies, we may need to consider a different nomenclature and terminology altogether, instead of grouping indiscriminately according to bioinformatics. This study has established a base, onto which subsequent studies will build an understanding of the function of these LEAs and will ultimately help in unveiling just how these enigmatic LEA proteins play roles under stress scenarios. All the relevant information and results obtained will benefit future LEA protein characterization efforts and will contribute towards understanding the role of LEA proteins in (*inter alia*) desiccation tolerance.

References

Amara, I., Odena, A., Oliveira, E., Moreno, A., Masmoudi, K., Pagès, M., & Goday, A. (2012). Insights into Maize LEA proteins: from proteomics to functional approaches. *Plant & Cell Physiology*, *53*(2), 312–29.

Amara, I., Zaidi, I., Masmoudi, K., & Ludevid, M. D. (2014). Insights into Late Embryogenesis Abundant (LEA) Proteins in Plants : From Structure to the Functions. *American Journal of Plant Sciences*, *5*, 3440–3455.

Babu, M. M., van der Lee, R., de Groot, N. S., & Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology*, *21*(3), 432–40.

Bajic, J. (2006). Exploring the longevity of dry *Craterostigma wilmsii* (homoiochlorophyllous) and *Xerophyta humilis* (poikilochlorophyllous) under simulated field conditions. PhD Thesis, University of Cape Town.

Baker, E. H., Bradford, K.J., Bryant, J.A., Rost TL (1995) A comparison of desiccation-related proteins (dehydrin and QP47) in pea (*Pisum sativum*). *Seed Science Research*, *5*, 185–193.

Balsamo, R.A., Vander Willigen, C., Boyko, W., Farrant, J. (2005). Retention of mobile water during dehydration in the desiccation-tolerant grass *Eragrostis nindensis*. *Physiologia Plantarum*, *124*(3): 336-342.

Bartels, D., & Salamini, F. (2001). Desiccation tolerance in the resurrection plant *Craterostigma plantagineum*. A contribution to the study of drought tolerance at the molecular level. *Plant Physiology*, *127*(4), 1346–1353.

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.

Battaglia, M., & Covarrubias, A. a. (2013). Late Embryogenesis Abundant (LEA) proteins in legumes. *Frontiers in Plant Science*, *4*, 190.

Berjak, P. (2006) Unifying perspectives of some mechanism basic to desic- cation tolerance across life forms. *Seed Science Research, 16*, 1-15.

Bies-Ethève, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Delseny, M. (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Molecular Biology*, *67*(1-2), 107–124.

Bockel, C., Salamini, F., & Bartels, D. (1998). Isolation and characterization of genes expressed during early events of the dehydration process in the resurrection plant *Craterostigma plantagineum*. *Journal of Plant Physiology*, *152*(2-3), 158–166.

Bokor, M., Csizmók, V., Kovács, D., Bánki, P., Friedrich, P., Tompa, P., & Tompa, K. (2005). NMR Relaxation Studies on the Hydrate Layer of Intrinsically Unstructured Proteins. *Biophysical Journal*, *88*(3), 2030–2037.

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.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254.

Bray, E. A. (1993). Responses to Water Deficit. *Plant Physiology*, *103*, 1035–1040.

Bray, E. A. (1994) Alternation in gene expression in responsive to water deficit. In Stress-Induced Gene Expression in Plants. Harwood Academic Publishers. 1-23.

Bray, E. A. (1997). Plant responses to water deficit. *Trends in Plant Science*, *2*(2), 48–54.

Ceccardi, T.L., Meyer, N.C., Close, T.J. (1994) Purification of a maize dehydrin. *Protein Expression Purification*, *5*(3), 266–269.

Close, T.J., Kortt, A.A., Chandler, P.M. (1989) A cDNA-based comparison of dehydration induced proteins (dehydrins) in barley and corn. *Plant Molecular Biology*. *13*, 95–108.

Close, T.J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarium*, *97*, 795–803.

Cohen, A., Plant, A. L., Moses, M. S., & Bray, E. A. (1991). Organ-Specific and Environmentally Regulated Expression of Two Abscisic Acid-Induced Genes of Tomato. Nucleotide Sequence and Analysis of the Corresponding cDNAs. *Plant Physiology*, *97*, 1367–1374.

Collett, H., Shen, A., Gardner, M., Farrant, J. M., Denby, K. J., & Illing, N. (2004). Towards transcript profiling of desiccation tolerance in *Xerophyta humilis*: Construction of a normalized 11 k *X. humilis* cDNA set and microarray expression analysis of 424 cDNAs in response to dehydration. *Physiologia Plantarum*, *122*(1), 39–53.

Cuming, A.C., Cho, S.H., Kamisugi, Y., Graham, H., Quatrano, R.S., (2007). Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss *Physcomitrella patens*. *New Phytologist*, *176*, 275–287.

Dainty, J. (1976). Water relations of plant cells. In: Gottingen AP, Zimmermann MH, editors. Encyclopedia of Plant Physiology, Vol 2. New Series, Part A. Berlin: Springer-Verlag. 12–35.

Delseny, M., Carles, C., Hull, G., Vicient, C., Raynal, M., Grellet, F., Grellet, F. (2001). Late embryogenesis Abundant protein gene regulation during *Arabidopsis* maturation. *Journal of Plant Physiology*, *158*, 419–427.

Dinakar, C., & Bartels, D. (2013). Desiccation tolerance in resurrection plants: new insights from transcriptome, proteome and metabolome analysis. *Frontiers in Plant Science*, *4*, 482.

Dure, L., Greenway, S.C., and Galau, G.A. (1981) Developmental biochemistry of cotton seed embryogenesis and germination: Changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry-Us. 20*, 4162–4168.

Dure, L., Crouch, M., Harada, J.J., Ho, T., Mundy, J., Quatrano, R.S., Thomas,T.L., Sung, Z.R. (1989) Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Molecular Biology, 12*, 475–486.

Dure, L. (1993a) Structural motifs in LEA proteins. In TJ Close. EA Bray eds Plant Responses to Cellular Dehydration during Environmental Stress. American Society of Plant Physiologists Rockville MD, 91–103.

Dure, L. (1993b) A repeating 11-mer amino acid motif and plant desiccation. *Plant Journal. 3,* 363–369.

Farrant, J. M. (2000). A comparison of mechanisms of desiccation tolerance among three angiosperm resurrection plant species. *Plant Ecology*, *151*, 29–39.

Farrant, J. M., Brandt, W., & Lindsey, G. G. (2007). An Overview of Mechanisms of Desiccation Tolerance in Selected Angiosperm Resurrection Plants. *Plant Stress, Global Science Books*, *1*(1), 72-84.

Farrant, J. M., Lehner, A., Cooper, K., & Wiswedel, S. (2009). Desiccation tolerance in the vegetative tissues of the fern *Mohria caffrorum* is seasonally regulated. *The Plant Journal : For Cell and Molecular Biology*, *57*(1), 65–79.

Farrant, J. M., & Moore, J. P. (2011). Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Current Opinion in Plant Biology*, *14*(3), 340–5.

Farrant, J. M., Cooper, K., & Nell, H. (n.d.). Plant Stress Physiology, Desiccation Tolerance. *Plant Stress Physiology*, Book, Chapter 11, 238–265.

Finkelstein, R. R., Gampala, S. S. L., & Rock, C. D. (2002). Abscisic Acid Signaling in Seeds and Seedlings. *The Plant Cell*, 15–46.

Fuxreiter, M., Tompa, P., & Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics (Oxford, England)*, *23*(8), 950–6.

Gaff, D. F. (1977). Desiccation Tolerant Vascular plants of southern Africa. *Oecologia, 31*, 95–109.

Gaff, D. F. and Melvin, O. (2013). The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. *Functional Plant Biology, 40*, 315–328.

Galau, G.A., and Dure, L. III. (1981). Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as

shown by reciprocal heterologous complementary deoxyribonucleic acid-messenger ribonucleic acid hybridization. *Biochemistry*, *20*, 4169–4178.

Galau, G.A., Hugh, D.W., Dure, L. (1986) Abscisic acid induction of cloned cotton late embryogenesis-abundant (lea) mRNAs. *Plant Molecular Biology, 7,* 155–170.

Galau, G.A., Wang, H.Y-C., Hughes, D.W. (1993) Cotton Lea5 and Lea14 encode atypical late embryogenesis-abundant proteins. *Plant Physiology, 101,* 695–696.

Gaubier, P., Raynal, M., Hull, G., Huestis, G., Gellet, F., Arenas, C., Pages, M., Delseny, M. (1993) Two different Em-like genes are expressed in *Arabidopsis thaliana* seeds during maturation. *Molecular Genetics and Genomics, 238*, 409–418.

Ginbot, Z. G. (2011). Characterisation of two desiccation linked, Group 1 LEA proteins from the resurrection plant *Xerophyta humilis*, Thesis Presented for the Degree of University of Cape Town.

Goyal, K., Walton, L. J., & Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. *Biochemistry Journal*, *388*, 151–157.

Greenfield, N. J. (2009). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, *1*(6), 2876–2890.

Greenway, S. C., & Galau, G. A. (1981). Developmental Biochemistry of Cottonseed Embryogenesis and Germination: Changing Messenger Ribonucleic Acid Populations As Shown by in Vitro and in Vivo Protein Synthesis. *Biochemistry, 20*(14), 4162–4168.

Gygi, S,P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology, 19*(3), 1720-1730.

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., & Masmoudi, K. (2011). Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signaling & Behavior*, *6*(10), 1503–9.

Hara, M., Terashima, S., Kuboi, T. (2001) Characterization and cryoprotective activity of cold-responsive dehydrin from *Citrus unshiu*. *Journal of Plant Physiology, 58*, 1333–1339.

Hinniger, C., Caillet, V., Michoux, F., Ben Amor, M., Tanksley, S., Lin, C., & McCarthy, J. (2006). Isolation and characterization of cDNA encoding three dehydrins expressed during Coffea canephora (Robusta) grain development. *Annals of Botany*, *97*(5), 755–65.

Hoekstra, F. a, Golovina, E. a, & Buitink, J. (2001). Mechanisms of plant desiccation tolerance. *Trends in Plant Science*, *6*(9), 431–438.

Hong-Bo, S., Zong-Suo, L., & Ming-An, S. (2005). LEA proteins in higher plants: structure, function, gene expression and regulation. *Colloids and Surfaces. B, Biointerfaces*, *45*(3-4), 131–5.

Hsiao, T. C. (1979) Plant responses to water deficits, efficiency, and drought resistance. *Agriculture and Forest Meteorology*, *14*, 59–84.

Hundertmark, M., & Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics*, *9*, 118.

Hundertmark, M., Dimova, R., Lengefeld, J., Seckler, R., & Hincha, D. K. (2011). The intrinsically disordered late embryogenesis abundant protein LEA18 from Arabidopsis thaliana modulates membrane stability through binding and folding. *Biochimica et Biophysica Acta*, *1808*(1), 446–53.

Illing, N., Denby, K.J., Collett, H., Shen, A., Farrant, J.M. (2005). The Signature of Seeds in Resurrection Plants: A Molecular and Physiological Comparison of Desiccation Tolerance in Seeds and Vegetative Tissues. *Integrative and Comparative Biology, 45*, 771-787.

Ingle, R. a, Schmidt, U. G., Farrant, J. M., Thomson, J. a, & Mundree, S. G. (2007). Proteomic analysis of leaf proteins during dehydration of the resurrection plant Xerophyta viscosa. *Plant, Cell & Environment*, *30*(4), 435–446.

Ingram, J., & Bartels, D. (1996). the Molecular Basis of Dehydration Tolerance in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *47*, 377–403.

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 1. *Plant Physiology*, *120*, 237–244.

Jiang, X., Wang, Y. (2004) b-Elimination coupled with tandem mass spectrometry for the identification of in vivo and in vitro phosphorylation sites in maize dehydrin DHN1 protein. *Biochemistry, 43,* 15567–15576.

Jones, L., & McQueen-Mason, S. (2004). A role for expansins in dehydration and rehydration of the resurrection plant *Craterostigma plantagineum*. *FEBS Letters*, *559*(1-3), 61–5.

Kamies, R. (2011). A methodological investigation of the proteins present in the roots of the resurrection plant, *Xerophyta viscosa* , for further proteomic analyses. MSc. Thesis, University of Cape Town.

Koag, M., 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,* 309–316.

Kovacs, D., Kalmar, E., Torok, Z., & Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiology*, *147*(1), 381–90.

Kranner, I., Beckett, R. P., Wornik, S., Zorn, M., & Pfeifhofer, H. W. (2002). Revival of a resurrection plant correlates with its antioxidant status. *The Plant Journal : For Cell and Molecular Biology*, *31*(1), 13–24.

Leprince, O., & Buitink, J. (2010). Desiccation tolerance: From genomics to the field. *Plant Science*, *179*(6), 554–564.

Lisse, T., Bartels, D., Kalbitzer, H.R., and Jaenicke, R. (1996). The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Craterostigma plantagineum* displays no defined three- dimensional structure in its native state. *The Journal of Biological Chemistry*, *377*(9): 555–561.

Liu, Y., Wang, L., Xing, X., Sun, L., Pan, J., Kong, X., Li, D. (2013). ZmLEA3, a Multifunctional Group 3 LEA Protein from Maize (Zea mays L.), is Involved in Biotic and Abiotic Stresses. *Plant and Cell Physiology*, *54*(6), 944–959.

Manfre, A. J., Lanni, L. M., & Jr, W. R. M. (2006). The Arabidopsis Group 1 LATE EMBRYOGENESIS ABUNDANT Protein ATEM6 Is Required for Normal Seed Development. *Plant Physiology*, *140,* 140–149.

Maredza, A. T. (2007). Isolation of the aldose reductase gene (XvAld1) from the resurrection plant *Xerophyta viscosa*, and characterisation of the gene product and transgenic plants expressing the gene, PhD thesis, University of Cape Town.

McCubbin, W. D., Kay, C. M. & Lane, B. G. (1985) Hydrodynamic and optical properties of the wheat germ E,,, protein. Can. J. Biochem. Cell Biol. 63, 803-81 1

Mehta, A., Magalhães, B. S., Souza, D. S. L., Vasconcelos, E. A. R., Silva, L. P., Grossi-de-sa, M. F., Costa, P. H. A. (2008). Rooteomics : The Challenge of Discovering Plant Defense-Related Proteins in Roots. *Current Protein and Peptide Science, 9*, 108–116.

Mogami, N., Shiota, H., & Tanaka, I. (2002). The identification of a pollen-specific LEA-like protein in Lilium longiflorum. *Plant, Cell and Environment*, *25*(5), 653–663.

Moore, J. P., Ngoc, Tuan Le., Brandt, W.F., Driouch, A., Farrant, J.M. (2009). Towards a system- based understanding of plant desiccation tolerance. *Trends in Plant Science, 14*, 110-117.

Mouillon, J., Gustafsson, P., & Harryson, P. (2006). Structural Investigation of Disordered Stress Proteins . Comparison of Full-Length Dehydrins with Isolated Peptides of Their Conserved Segments 1. *Plant Physiology*, *141*, 638–650.

Mundree, S. G., Baker, B., Mowla, S., Peters, S., Marais, S., Vander Willigen, C., Thomson, J. a. (2002). Physiological and molecular insights into drought tolerance. *African Journal of Biotechnology*, *1*(2), 1–19.

Neumann, A. (2013). Characterisation of galactinol synthase II (XvGolSII) from the resurrection plant *Xerophyta viscosa* (Baker), MSc. Thesis, University of Cape Town.

Ngubane, A. (2008). Cloning and characterization of LEA1-EM genes in the resurrection plant *Xerophyta humilis*. MSc. Thesis, University of Cape Town.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., & Dunker, a K. (2005). Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins: Structure, function, and Bioinformatics Supplementary, 7*, 176–82.

Olave-concha, N., Ruiz-lara, S., Muñoz, X., & Bravo, L. A. (2004). Accumulation of dehydrin transcripts and proteins in response to abiotic stresses in *Deschampsia antarctica*. *Antarctic Science*, *16*(2), 175–184.

Oliver, M. J. and Bewley, J. D. (1984). Plant desiccation and protein synthesis. Changes in protein synthesis elicited by desiccation of the moss *Tortula ruralis* are effected at the translational level. *Plant Physiology, 74*, 923-927.

Oliver, M. J., Wood, A. J., O'Mahony, P. (1998). "To dryness and beyond "– preparation for the dried state and rehydration in vegetative desiccation-tolerant plants. *Plant Growth Regulation, 24*, 193-201.

Oliver, M. J., Tuba, Z., Mishler, B. D. (2000). The evolution of vegetative desiccation tolerance in land plants. *Plant Ecology, 151*, 85–100.

Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garciarrubio, A., & Covarrubias, A. a. (2010). Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in Arabidopsis. *Plant Physiology*, *154*(1), 373–90.

Olvera-Carrillo, Y., Luis Reyes, J., & Covarrubias, A. a. (2011). Late embryogenesis abundant proteins. *Plant Signaling & Behavior*, *6*(4), 586–589.

Palmer, I., & Wingfield, P. T. (2012). Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from *Escherichia coli*. *Current Protocols in Protein Science*, 1–25.

Pelah, D., Shoseyov, O., Altman, A. (1995). Characterization of BspA, a major boiling stable, water-stress-responsive protein in aspen (*Populus tremula*). *Tree Physiology*, *15*, 673-678.

Peng, K., Radivojac, P., Vucetic, S., Dunker, a K., & Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, *7*, 208.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), 2000–2007.

Rabbani, M. A., Maruyama, K., Abe, H., Khan, M. A., Katsura, K., Ito, Y., Yamaguchi-shinozaki, K. (2003). Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. *Plant Physiology*, *133*, 1755–1767.

Roberts, E. H. (1973). Predicting the storage life of seeds. *Seed Science and Technology, 1,* 499-514.

Roberts, E. H. (1993). Cellular Concentrations and Uniformity of Cell Type Accumulation of Two Lea Proteins in Cotton Embryos. *Plant Cell, 5*, 769–780.

Russouw, P.S., Farrant, J., Brandt, W., Lindsey, G.G. (1997). The most prevalent protein in a heat-treated extract of pea (*Pisum sativum*) embryos is an LEA group 1 protein; its conformation is not affected by exposure to high temperature. *Seed Science Research, 7,* 117–123.

Saavedra, L., Svensson, J., Carballo, V., & Izmendi, D. (2006). A dehydrin gene in *Physcomitrella patens* is required for salt and osmotic stress tolerance. *The Plant Journal*, *45*, 237–249.

Scott, P. (2000). Resurrection Plants and the Secrets of Eternal Leaf. *Annals of Botany, 85*(2), 159- 166.

Sherwin, H., & Farrant, J. M. (1996). Differences in Rehydration of Three Desiccation-tolerant Angiosperm Species. *Annals of Botany*, *78*(6), 703–710.

Sherwin, H. W., & Farrant, J. M. (1998). Protection mechanisms against excess light in the resurrection plants Craterostigma wilmsii and Xerophyta viscosa. *Plant Growth Regulation*, *24*(3), 203–210.

Shih M.D., Lin S.C., Hsieh J.S., Tsou C.H., Chow T.Y., Lin T.P., et al. (2004). Gene cloning and characterization of a soybean (Glycine max L.) LEA protein, GmPM16. *Plant Molecular Biology*, *56*, 689-703.

Shih, M., Hsieh, T.-Y., Lin, T.-P., Hsing, Y.-I. C., & 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.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., & Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science : A Publication of the Protein Society*, *18*, 936–948.

Soulages J.L., Kim K., Walters C., Cushman J.C. (2002). Temperature-induced extended helix/random coil transitions in a Group 1 late embryogenesis-abundant protein from soybean. *Plant Physiology*. *128*, 822–832.

Soulages, J. L., Kim, K., Arrese, E. L., Walters, C., & Cushman, J. C. (2003). Conformation of a Group 2 Late Embryogenesis Abundant Protein from Soybean. Evidence of Poly (L -Proline) -type II Structure 1. *Plant Physiology*, *131*, 963–975.

Stacy, R. A. P., Nordeng, T. W., & Culia, F. A. (1999). The dormancy-related peroxiredoxin anti-oxidant, PER1, is localized to the nucleus of barley embryo and aleurone cells. *The Plant Journal*, *19*(1), 1–8.

Swire-Clark, G.A., Marcotte, W.R. Jr. (1999). The wheat LEA protein Em functions as an osmoprotective moledule in *Saccharomyces cerevisiae*. *Plant Molecular Biology, 39*, 117–128.

Thomashow, M. F. (1999). PLANT COLD ACCLIMATION : Freezing Tolerance Genes and Regulatory Mechanisms. *Annual Review of Plant Physiology and Plant Molecular Biology*. *50*, 571-599.

Tolleter, 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.

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

Tunnacliffe, A., Hincha, D. K., Leprince, O., & Macherel, D. (2010). LEA proteins: Versatility of Form and Function. Dormancy and Resistance in Harsh Environments. *Topics in Current Genetics*, *21*, 91–108.

Vicient, C. M., Hull, G., Guilleminot, J., Devic, M., & Delseny, M. (2000). Differential expression of the Arabidopsis genes coding for Em-like proteins 1. *Journal of Experimental Botany*, *51*(348), 1211–1220.

Vicré, M., Farrant, J. M., & Driouich, a. (2004). Insights into the cellular mechanisms of desiccation tolerance among angiosperm resurrection plant species. *Plant, Cell & Environment*, *27*(11), 1329–1340.

Waisel, Y., Eshel, A., Kafkafi, U. (2001)."Plant Roots: The Hidden Half." 3rd edition. Marcel Dekker, Inc. New York, Base, 1-1136.

Walters, C., Farrant, J.M., Pammenter, N.W. and Berjak, P. (2002) Desiccation stress and damage. In Desiccation and Survival in Plants: Drying without Dying (Black, M. and Pritchard, H.W., eds). Wallingford, UK: CABI Publishing, 263–283.

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, *9*(5), 244–52.

Willems, E., Leyns, L., & Vandesompele, J. (2008). Standardization of real-time PCR gene expression data from independent biological replicates. *Analytical Biochemistry*, *379*(1), 127–9.

Wise, M. J. (2003). LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinformatics*, *4*, 52.

Wise, M. J., & Tunnacliffe, A. (2004). POPP the question : what do LEA proteins do ? *Trends in Plant Science*, *9*(1), 13–17.

Wu, X., Gong, F., & Wang, W. (2013). Functional assignment to maize group 1 LEA protein EMB564 within the cell nucleus using computational analysis. *Bioinformation*, *9*(6), 276–280.

Yu J.N. (2003). Cloning and characterization of tolerant stress genes from Triticum aestivum dissertation for Doctor's degree of Northwestern Sci-tech University of Agriculture and Forestry, Yangling, China.

Zhang, L., Ohta, A., Takagi, M., & Imai, R. (2000). Expression of Plant Group 2 and Group 3 lea Genes in *Saccharomyces cerevisiae* Revealed Functional Divergence among LEA Proteins. *Journal of Biochemistry*, *616*, 611–616.

Zhang, L.S., Zhao, W. M., (2003). LEA protein functions to tolerance drought of the plant. *Plant Physiology*. 39 (1), 61 – 66 [21].

Appendices

Appendix A: Protocols.

A1. Quality control of RNA samples:

The quality of RNA was measured using the NanoDrop-ND1000 by Dr Nicki Adams (CPGR). Results are tabulated (below). Several samples failed the QC analysis (Table A1a and A1b). The A260/A230 ratio indicated the likely presence of carbohydrates, phenol and possibly guanidine HCl used in commercial DNA isolation kits. Co-purified phenol and carbohydrates have the potential to entrap RNA and thus reduce the efficiency of downstream enzymatic reactions. To determine the concentration and purity of the RNA samples, 1.2µl of each RNA sample was analyzed on a NanoDrop (ND-1000, ThermoScientific). Nuclease-free water was used as the blank control (as conducted by CPGR).

Table A1a Nanodrop measurements of RNA from *X. humilis* leaf samples. The samples highlighted in red failed the QC requirements. This is indicated by A260/A230 ratios less than 1.5, which indicates poor RNA purity.

Sample	Sample ID	ng/ul	260/280	260/230
100% RWC Leaves-1	1	391.17	2.07	2.23
100% RWC Leaves-2	2	408.48	2.06	2.18
100% RWC Leaves-3	3	227.13	2.11	1.55
100% RWC Leaves-4	4	220.4	2.1	1.86
70-80% RWC Leaves-1	5	523.2	2.12	2.04
70-80% RWC Leaves-2	6	523.47	2.12	2.04
70-80% RWC Leaves-3	7	384.9	2.08	1.12
70-80% RWC Leaves-4	8	448.26	2.04	2.19
40-50% RWC Leaves-1	9	562.58	2.12	1.35
40-50% RWC Leaves-2	10	405.16	2.05	1.87
40-50% RWC Leaves-3	11	504.88	2.12	1.62
40-50% RWC Leaves-4	12	584.24	2.12	2.02
20-30% RWC Leaves-1	13	857.36	2.1	1.99
20-30% RWC Leaves-2	14	441.09	2.04	1.58
20-30% RWC Leaves-3	15	438.92	1.99	1.95
20-30% RWC Leaves-4	16	591.61	2.09	1.79
<10% RWC Leaves-1	17	345.11	2.04	1.7
<10% RWC Leaves-2	18	521.93	2.13	1.97
<10% RWC Leaves-3	19	882.28	2.12	2.15
<10% RWC Leaves-4	20	872.72	2.1	2.2
6H Post Leaves-1	21	438.36	2.05	1.02
6H Post Leaves-2	22	1101.52	2.1	2.15
6H Post Leaves-3	23	462.06	2.03	1.02
6H Post Leaves-4	24	696.79	2.11	2.12
24H Post Leaves-1	25	642.17	2.12	0.72
24H Post Leaves-2	26	666.22	2.13	1.53
24H Post Leaves-3	27	526.93	2.13	1.07
24H Post Leaves-4	28	1287.48	2.11	2.24
72H Post Leaves-1	29	875.35	2.13	1.84
72H Post Leaves-2	30	927.34	2.12	2.19
72H Post Leaves-3	31	1207.7	2.11	2.14
72H Post Leaves-4	32	836.18	2.11	2.01

Table A1b Nanodrop measurements of RNA from *X. humilis* root samples. The samples highlighted in red failed the QC requirements. This is evident in low A260/A280 and (or) A260/A230 ratios of less than 1.5, which indicates poor RNA purity.

Sample	Sample ID	ng/ul	260/280	260/230
100% RWC Roots-1	33	751.98	1.41	1.44
100% RWC Roots-2	34	710.86	1.41	1.5
100% RWC Roots-3	35	727.45	1.42	1.65
100% RWC Roots-4	36	761.53	1.41	1.27
70-80% RWC Roots-1	37	66.89	1.54	1.4
70-80% RWC Roots-2	38	82.74	2.06	1.21
70-80% RWC Roots-3	39	97.63	1.54	1.46
70-80% RWC Roots-4	40	63.94	2.02	0.4
40-50% RWC Roots-1	41	574.56	1.46	1.76
40-50% RWC Roots-2	42	596.55	1.45	1.91
40-50% RWC Roots-3	43	789.87	1.43	1.68
40-50% RWC Roots-4	44	545.92	1.48	1.74
20-30% RWC Roots-1	45	803.72	1.46	1.57
20-30% RWC Roots-2	46	646.09	1.53	1.56
20-30% RWC Roots-3	47	786.34	1.48	1.58
20-30% RWC Roots-4	48	914.07	1.52	1.46
<10% RWC Roots-1	49	947.25	1.52	1.62
<10% RWC Roots-2	50	1439.61	1.47	1.36
<10% RWC Roots-3	51	913.42	1.5	1.51
<10% RWC Roots-4	52	1004.05	1.49	1.44
6H Post Roots-1	53	748.57	1.47	1.61
6H Post Roots-2	54	757.89	1.42	1.58
6H Post Roots-3	55	795.38	1.44	1.34
6H Post Roots-4	56	808.28	1.45	1.55
24H Post Roots-1	57	467.18	1.85	2.14
24H Post Roots-2	58	908.37	1.42	1.49
24H Post Roots-3	59	711.82	1.41	1.71
24H Post Roots-4	60	900.76	1.41	1.5
72H Post Roots-1	61	634.06	1.44	1.61
72H Post Roots-2	62	827.49	1.42	1.46
72H Post Roots-3	63	845.44	1.45	1.42
72H Post Roots-4	64	633.3	1.46	1.55

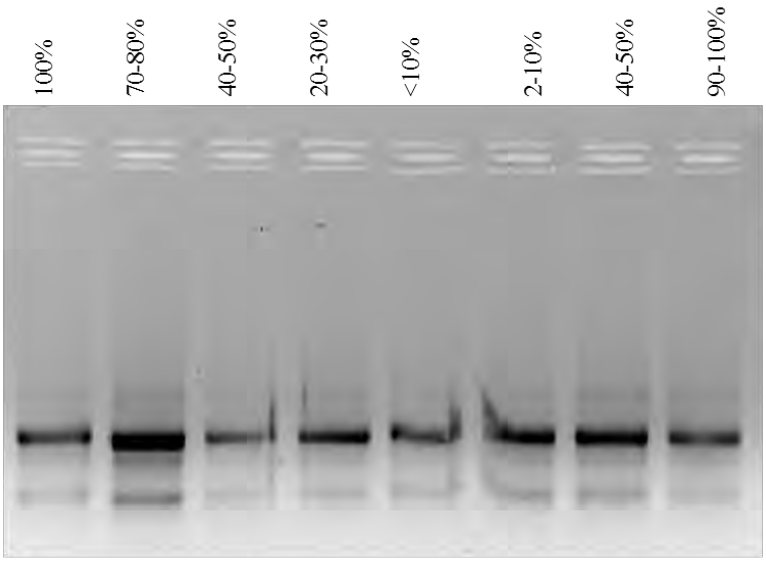


Figure A1. Assessment of RNA quality and integrity by electrophoresis of 1 µg total RNA through a 1% agarose/EtBr gel. This depicts the RNA extracted from the dehydration and rehydration time course of single biological replicate *X. humilis* plant, but is representative of all time points across all biological replicates for leaf tissue RNA.

A2. cDNA synthesis:

Two cDNA synthesis steps per normalised RNA sample (following the MIQE publishable requirements), using the High Capacity cDNA synthesis kit (Life Technologies) were conducted and the cDNA synthesised for each sample was pooled prior to gene expression analysis. cDNA synthesis of 500ng RNA was performed in duplicate for each RNA sample. The RNA was diluted to 50ng/µl and 10 µl of the diluted RNA sample was added to 10 µl of a cDNA synthesis master mix, comprised of reverse transcriptase buffer, RT Random Primers, dNTP mix, MultiScribe™ Reverse Transcriptase (High Capacity cDNA synthesis Kit, Life Technologies, Part # 4368814) and nuclease-free water. The components were thoroughly mixed and centrifuged using a bench top centrifuge to collect all the liquid. Cycling was performed on the GeneAmp® PCR System 9700 (Life Technologies) using the cycling parameters tabulated below:

Table A2. cDNA synthesis cycling parameters

Step	1	2	3	4
Temperature (°C)	25	37	85	4
Time (minutes)	10	120	5	Infinity

The experimental cDNA samples were made by combining 10µl of each duplicate. Before expression analysis, each experimental cDNA sample was diluted 1:4 or 1:16 with nuclease-free water. The original cDNA synthesis plate, which still had 10µl of each sample in duplicate, was kept at -80 °C.

A3. qPCR primer sets:

Table A3. The list of primer sets used in the qPCR gene expression studies on leaf vegetative tissues of *X. humilis*

LEA Clone Identity	Forward (F) and Reverse (R) primer sequences (5' to 3'):	Amplicon size (bp):
XhLEA1-4S1	F:GGACAGAGGGGTATCAGGAGATGGGC R:CTCTCTCCTCGCCTCCCCTCCTTTG	112
XhLEA1-1S2	F:CTCAAGAACACCTCGCTGACGGACG R:GTACGACCGTCTCCCCTGCTCC	179
XhLEA2-1	F:CAACACCAGCAGCACAAGGACCAGAG R:CTGTCCTTGGTTGTGTTGCCCTGGC	182
XhLEA2-2	F:AGAAGCAAGGAACCACCACC R:CCTTTCTTTTCCCCATGCGC	94
XhLEA2-3	F:AGGCCAGCATAACCAAGGTC R:TTGTCCATCGTCCCCTCCTT	126
XhLEA2-4	F:ACCAAGACCAGCACTACCGCACCAGC R:TCGTGAGCAGTTCCATGTCCCTGATGC	165
XhLEA2-5	F:CCATCAAACTCACCACCACGAGACAGG R:GATCCTTTCTATCATGTGGTTGCTGCAGGC	118
XhLEA2-6	F:GCCCTGGATATCAGCAGGAG R:TCCTTGAGCCCCTTCTTCCT	112
XhLEA3-1	F:AGTTCAACAGTGGAATGGGGATGTGGCG R:CCTTCCATTCCTCCGCCTTGCCG	183
XhLEA3-2	F:GGACAGGGCACAAGAACTCA R:GCATCCAGTTTCCGGTCTGA	131
XhLEA3-3	F:GCCCAGGAAGTGAAGGACAA R:CTTTCTCTTTCGCGCAGTCG	142
XhLEA3-4	F:TCTCAGGCACAGAAAGATGACGCAATGGAC R:CCAAGCCATCTTCCCCTGATCCTCAC	113
XhLEA3-5	F:AAGAGGCTTGGTCCGAGTGGGTGG R:GCCGACTTCGCCGCCTTATCTTTCACAG	207
XhLEA3-6	F:GGACAAGACGACAGGGACAG R:GACTCCATGTACCCGCCAGC	146
XhLEA3-7	F:GATGTGGCGCAACATGGGGAAGAATATAGC R:CTCCTTGCTCTCCCCTGCTTTCTCAG	132
XhLEA4-1	F:GGCGAAAGTAGACCTCCACC R:ATGCTTGTTATGGCGGTTGC	109
XhLEA4-2	F:ATGGACAGAGCCAAAGCCTC R:TCATCCTCTCAGCTGCGTTC	159
XhLEA5a	F:CGGGGAAGACGGAGGATGCAGTG R:CCAAGCACGCGATCCTCCTCAGC	123
XhLEA5b	F:CTGCAACTGGGGTGAGAAAGGTAGAGG R:GTGGTTGGCCGGCCGGTAATAGC	133
XhLEA5c-1	F:GGATCAATCAAGGGCAGTGAGACGACG R:CCAATCTCTTCCCACGTCCTTCATCAAGC	98
XhLEA5c-2	F:ACAAGATCGCCCACATCCAG R:GGGATTGGAGATGGCAAGCT	110
X.hum F 18S **X.hum R 18S**	F: CAGGCGCGCAAATTACCCAATCC R: CCTACCGTCCCGTCCCAAGGT	131
EF1A QRT F2 **EF1A QRT R2**	F: AGGGTTCGAGGGCGACAACAT R: ACCGATCTTGTACACGTCCTGGAG	131

A4. Primer pair characteristics:

Table A4 Primer efficiency (E), slope and R2 value for each primer pair as calculated using qBase software. An E value of 2 indicates 100% primer efficiency and an R2 value of 0.99 indicates a good correlation between Ct and sample concentration.

	E	R^2	Slope	Intercept
X. hum 18s	1.803	0.999	-3.906	11.923
EF1A	2.006	0.991	-3.307	25.103
XhLEA1-4S1	2.121	0.988	-3.063	28.996
XhLEA1-1S2	2.213	0.993	-2.899	27.124
XhLEA2-1	2.119	0.991	-3.491	19.905
XhLEA2-2	2.028	0.981	-3.257	19.784
XhLEA2-3	1.87	0.977	-3.678	22.224
XhLEA2-4	2.254	0.987	-2.833	25.435
XhLEA2-5	1.869	0.987	-3.68	22.213
XhLEA2-6	1.909	0.989	-3.562	23.906
XhLEA3-1	2.255	0.989	-2.832	22.094
XhLEA3-2	1.934	0.991	-3.491	19.905
XhLEA3-3	2.031	0.992	-3.25	23.513
XhLEA3-4	1.991	0.985	-3.343	28.78
XhLEA3-5	2.049	0.991	-3.209	20.402
XhLEA3-6	2.009	0.999	-3.301	21.346
XhLEA3-7	1.881	0.989	-3.645	28.408
XhLEA4-1	1.913	0.995	-3.549	25.931
XhLEA4-2	1.919	0.985	-3.532	29.991
XhLEA5a	2.156	0.996	-2.998	26.068
XhLEA5b	1.952	0.99	-3.442	25.481
XhLEA5c-1	1.9	0.993	-3.588	25.465
XhLEA5c-2	2.009	0.987	-3.301	26.559

A5. qPCR gene expression analysis:

The KAPA SYBR FAST qPCR Master Mix (2X) Universal was used for gene expression analysis on the ABI7900. The study also included the generation and optimisation of standard curves for sample analyses. An optimised standard curve was generated for each of the 23 genes that were investigated and the expression of the 21 LEA genes of interest and 2 reference genes was determined in each of the 64 samples in triplicate. The relative quantification was performed using qBase+ (BioGazelle) and the data was reported in an excel spread-sheet together with a formal report. To test the efficiency of the primers, a serial dilution of a cDNA pool was used as template for standard curve analysis. Three different serial dilutions were used as specified in Table A5a and Table A5c. The values shown are based on the input RNA of 25ng/µl.

Table A5a. Standard curves used for primer efficiency QC.

Standard curve	1	2	3
cDNA	Roots and Leaves	Roots and Leaves	Leaves only
Dilution	1:4	Combination 1:3 and 1:4	Combination 1:3 and 1:4
Dil 1	6.2500	8.3333	8.3333
Dil 2	1.5625	6.2500	6.2500
Dil 3	0.3906	2.7778	2.7778
Dil 4	0.0977	1.5625	1.5625
Dil 5	0.0244	0.9259	0.9259
Dil 6	0.0061	0.3906	0.3906
Dil 7	0.0015	0.3086	0.3086
Dil 8	0.000381	0.1029	0.1029

To test for DNA contamination, an RNA pool was made and diluted in accordance with the protocol for cDNA synthesis and the subsequent dilution. For gene expression analysis, each reaction (for standard curves and samples) consisted of 1 µl cDNA template; 0.1-0.2µl of each of the primers (final concentration of 100nM or 200nM, see Table A5c); 5 µl KAPA SYBR® FAST qPCR KIT MasterMix (2X) ABI Prism (Lasec, KK4604) and nuclease-free water up to 10 µl. Expression analysis was performed on the ABI 7900HT Fast Real Time PCR system using the following cycling parameters and followed by a dissociation (melt) curve analysis:

Table A5b. qPCR cycling parameters

Stage	1	2		3		
Cycles	1	40		1		
Temperature (°C)	95	95	60 or 62*	95	60	95
Time	3 mins	3 secs	1 min	15 sec	15 sec	15 sec

*See Table A5c

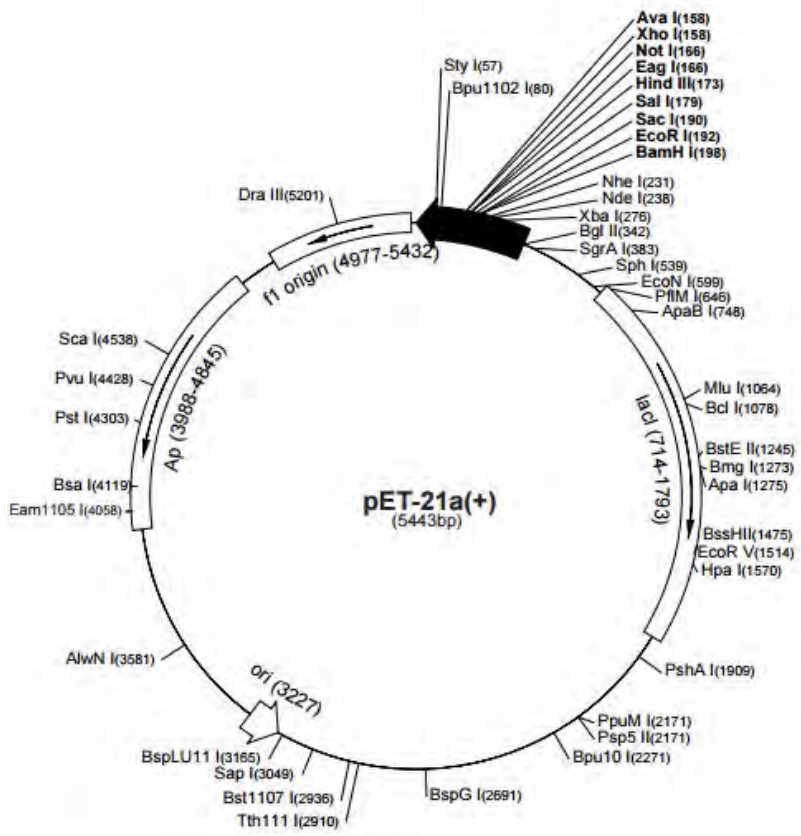
Post cycling, the data was analysed using the SDS v2.4 software (Life Technologies) and relative expression analysis performed using qBase+ (BioGazelle).

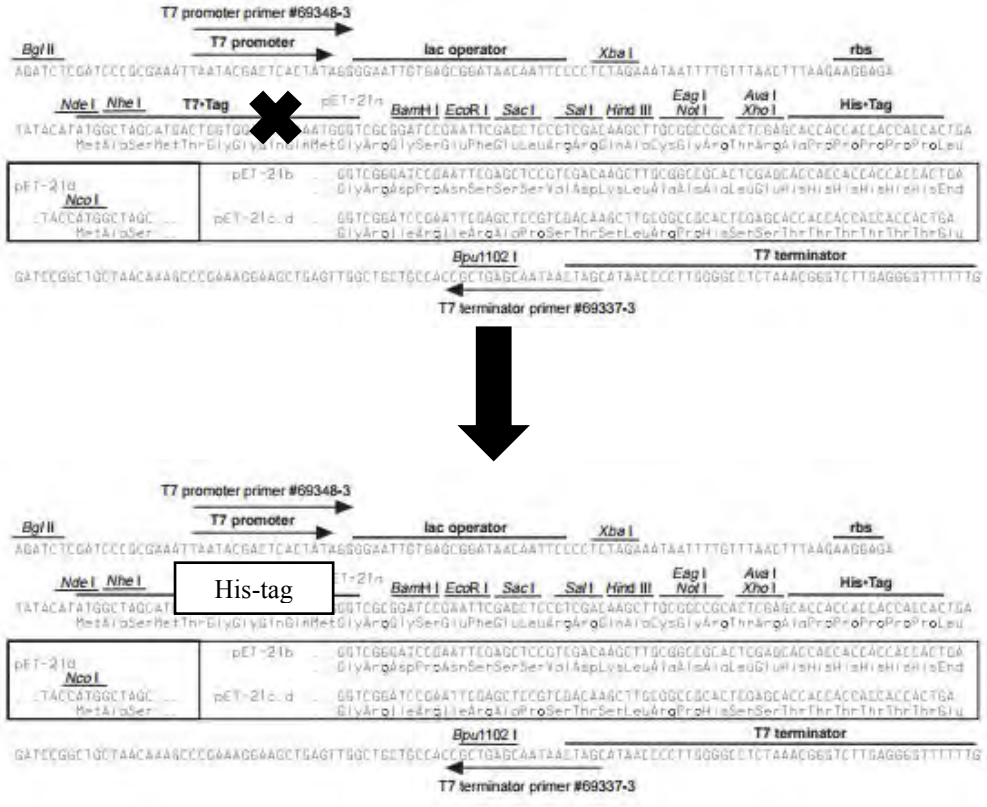
Table A5c. Variable parameters used for each individual target gene and the reference genes 18S and EF1A.

Gene	Sample dilution		Standard curve Refer to Table A5a	Primers (µM)	Annealing Tm (°C)
	Leaf	Root			
X. hum 18s	1:16	1:16	1	0.2	60
EF1A	1:16	1:16	1	0.2	60
XhLEA1-4S1	1:16	1:4	2	0.2	62
XhLEA1-1S2	1:16	1:4	2	0.2	60
XhLEA2-1	1:16	1:16	2	0.2	60
XhLEA2-2	1:16	N/A	3	0.2	60
XhLEA2-3	1:16	N/A	2	0.2	60
XhLEA2-4	1:16	1:4	2	0.2	60
XhLEA2-5	1:16	1:4	2	0.1	60
XhLEA2-6	1:16	N/A	2	0.2	60
XhLEA3-1	1:16	1:16	2	0.2	60
XhLEA3-2	1:16	N/A	3	0.2	60
XhLEA3-3	1:16	N/A	3	0.2	60
XhLEA3-4	1:16	1:4	2	0.2	60
XhLEA3-5	1:16	1:4	2	0.2	62
XhLEA3-6	1:16	1:16	1	0.2	60
XhLEA3-7	1:16	1:4	2	0.1	60
XhLEA4-1	1:16	N/A	3	0.2	60
XhLEA4-2	1:16	1:4	2	0.2	60
XhLEA5a	1:16	1:4	2	0.2	60
XhLEA5b	1:16	1:4	2	0.2	60
XhLEA5c-1	1:16	1:4	2	0.1	60
XhLEA5c-2	1:16	N/A	2	0.2	60

Appendix B: Vector map and nucleotide and amino acid sequence information

B1. pET-21a(+) vector map with an indication of modification to include an additional his-tag in the N terminal, and subsequent removal of the T7 tag in the sequence information





B2. Full length nucleotide and amino acid sequences of clone ID's XhLEA2-3, XhLEA2-6 and XhLEA3-5:

Clone ID:
Xh_LD_05G02/Xh_RD_14C02 [429 base pairs, ~15 kDa, Class 2, dehydrin]
Renamed as XhLEA2-3

Gene/Nucleotide sequence:
>EMBOSS_001
atggagggtttcgggaaccagcaagaccagctccgccgcaacgatcgcaccaacgaacac
ggcgctcccggtcagactggctacagcgctcagcatggagtaatcggaggccagcagcac
caccagaacaagcagcagggacttggcagtactggtgccggcattaagaacaagctccac
cgctccaacagctctagctctagctctgagagtgatggagaaggaggaaggaggaagaag
ggaattaaggacaagattaaggagaaaatgccaggccagcataaccaaggtcagaccggc
cagggcatcaccggcagccaccaaagccatggagccaccggtcagcaggggtatggagcc
gccggccagcacggagggaaggaggggacgatggacaagattaaggacaagatctcggga
aaccactag

Amino acid sequence:
MEGFGNQQDQLRRNDRTNEHGAPGQTGYSAQHGVIGGQQHHQNKQQGLGSTGAGIKNK
LHRSNSSSSSSESDGEGGRRKKGIKDKIKEKMPGQHNQGQTGQGITGSHQSHGATGQQGY
GAAGQHGGKEGTMDKIKDKISGNH

Clone ID:
Xh_LD_27A05/Xh_RD_27D09 [306 base pairs, ~12 kDa, putative Class 2, dehydrin]
Renamed as XhLEA2-6

Gene/Nucleotide sequence:
>EMBOSS_001
atggagggatacgggaaccatcagcactacaacgttgacgagtacggtaacccgctgccc
agtgggtatggtgaaaattacggccaccaacctatgcagcgccctggcgagtattatggt
aaccagggctacggcggccaccagcacggtgcgtacagcggccctggatatcagcaggag
tacggcagtggccctaggctacagcgctccggcagcagctccagctctgaggatgatggg
tatggtgggaggaggaagaaggggctcaaggataggatcatggagaacctgcctggccgt
cgctga

Amino acid sequence:
MEGYGNHQHYNVDEYGNPLPSGYGENYGHQPMQRPGEYYGNQGYGGHQHGAYSGPGY
QQEYGSGPRLQRSGSSSSSEDDGYGGRRKKGLKDRIMENLPGRR

Clone ID:
Xh_LD_44B08/Xh_RD_28B12r [597 base pairs, ~ 22 kDa, Class 3]
Renamed XhLEA3-5

Nucleotide/Gene sequence:
>EMBOSS_001
atggcgaggattgtgatgggtttgttatgggcggcgatgttgttggtggccgttcatgga
acgaccatggatgaagccaaagcaaaggcgaacgcggccgttccgagcatggatgaagcc
aaagcgaaggtgaacgctgccatcccgaccatggatgaaaccaaagagaagttgaatgct
gccatgcatgccgccaaggagaagaaagaggcttggtccgagtgggtggaagataagcta
gaaggattggggcttcagatggaaaataaaatggcggccgcagatgacactaccgataaa
gctgcagaagccgccatgaaggccaaggactctgcttcaggtgctgccgaaactacatcg
gagaaggcaggcgagagcaaggacactgtgaaagataaggcggcgaagtcggcggactgc
atgtccggcaaggccgaggatgtgaagggaaaaacctccgaaaccaccgggaaagcatcg
gagggagccggcggcatgtacgagagcgccaagcagaagatggaagaagcctacgctgct
gccaaggataaggtgtcgcaggcgaaggcggcggaggggaagtccgaggagctgtaa

Amino acid sequence:
MARIVMGLLLAAMLLVAVHGTTMDEAKAKANAAVPSMDEAKAKVNAAIPTMDETKEKL
NAAMHAAKEKKEAWSEWVEDKLEGLGLQMENKMAAADDTTDKAAEAAMKAKDSASG
AAETTSEKAGESKDTVKDKAAKSADCMSGKAEDVKGKTSETTGKASEGAGGMYESAKQ
KMEEAYAAAKDKVSQAKAAEGKSEEL

B3. Sequencing data (T7 promoter and T7 terminator): Nucleotide sequences and corresponding protein sequences (Pink) translated into compact form using ExPASy translate.

Xh_LD_05G02/XhLEA2-3:

>141203-30_K15_5A_T7promoter.ab1 1362
TTGGAAGTTACATTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGA
GATATACATATGCACCACCACCACCACCACCGCGGATCCGAATTCGAGCT
CCGTCGACAAGCTTGCATGGAGGGTTTCGGGAACCAGCAAGACCAGCTCC
GCCGCAACGATCGCACTAACGAACACGGCGCTCCCGGTCAGACTGGCTAC
AGCGCTCAGCATGGAGTAATCGGAGGCCAGCAGCACCACCAGAACAAGCA
GCAGGGACTTGGCAGTACTGGTGCCGGCATTAAGAACAAGCTCCACCGCT
CCAACAGCTCTAGCTCTAGCTCTGAGAGTGATGGAGAAGGAGGAAGGAGG
AAGAAGGGAATTAAGGACAAGATTAAGGAGAAAATGCCAGGCCAGCATAA
CCAAGGTCAGACCGGCCAGGGCATCACCGGCAGCCACCAAAGCCATGGAG
CCACCGGTCAGCAGGGGTATGGAGCCGCCGGCCAGCACGGAGGGAAGGAG
GGGACGATGGACAAGATTAAGGACAAGATCTCGGGAAACCACTAGACTCG
AGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAG
GAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT
TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTA
TATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGG
CGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTA
GCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGG
CTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTA
GTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCA
CGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA
ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCG
GCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT
TAACAAAAATTTAACGCTTACAATTTTAGGGGGGACTTTTTCCGGGGAAA
ATTGCCCGGGAACCCCTTATTTTGTTATTTTTTCTAAAAACAATTCAAAA
AATGTTTCCCGCTCCTGGAGACAAAAAACCCCCGGAAAAAAGGTTCTCTA
TATATTTTGAAAAAAGGAGAGAAATTTGGGTTTTTCACAAAATTTCCGGG
GGGCCCCTTTTT

MEGFGNQQDQLRRNDRTNEHGA
PGQTGYSAQHGVIGGQQHHQNKQQGLGSTGAGIKNKLHRSNSSSSSSESDGEGGRRKKGI
KDKIKEKMPGQHNQGQTGQGITGSHQSHGATGQQGYGAAGQHGGKEGTMDKIKDKISGNH

>141203-30_M15_5A_T7terminator.ab1 1305
GGGGGCATCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGT
GGTGGTGGTGGTGCTCGAGTCTAGTGGTTTCCCGAGATCTTGTCCTTAAT
CTTGTCCATCGTCCCCTCCTTCCCTCCGTGCTGGCCGGCGGCTCCATACC
CCTGCTGACCGGTGGCTCCATGGCTTTGGTGGCTGCCGGTGATGCCCTGG
CCGGTCTGACCTTGGTTATGCTGGCCTGGCATTTTCTCCTTAATCTTGTC
CTTAATTCCCTTCTTCCTCCTTCCTCCTTCTCCATCACTCTCAGAGCTAG
AGCTAGAGCTGTTGGAGCGGTGGAGCTTGTTCTTAATGCCGGCACCAGTA
CTGCCAAGTCCCTGCTGCTTGTTCTGGTGGTGCTGCTGGCCTCCGATTAC
TCCATGCTGAGCGCTGTAGCCAGTCTGACCGGGAGCGCCGTGTTCGTTAG
TGCGATCGTTGCGGCGGAGCTGGTCTTGCTGGTTCCCGAAACCCTCCATG
CAAGCTTGTCGACGGAGCTCGAATTCGGATCCGCGGTGGTGGTGGTGGTG
GTGCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG
GGAATTGTTATCCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCG
CGGGATCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCA
CCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGAT
GGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGT
GGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCT

TGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTA
CTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCC
CGGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGC
CCCGGAAGAGAGTCAATTCAGGGTGGGTGAATGTGAAAACCAGTAACCGT
TATACGAATGTCCCAAAAGTATGCCCGGTGTCTCTTTATCAAAACCGTTT
TCCCGCGTGGGGGAACCAGGCCCAGCCCCCGTTTCCTGCGAAAAACCCCG
GGAAAAAAATGGGAAACCGGCCAATGGCCGGAAACTTAAATTTCAATTTC
CCCAACCCGCGTGGGGCCAAACAAACTGGGCGGGGGGAAAAACAGTCCTT
TGTCTGGAATTGGGCGGTTTGCCCCCCCTCCAAATTTTGGGCCCTTGGAA
ACGGC

MEGFGNQQDQLRRNDRTNEHGAPGQTGYSAQH
GVIGGQQHHQNKQQGLGSTGAGIKNKLHRSNSSSSSSESDGEGGRRKKGIKDKIKEKMPG
QHNQGQTGQGITGSHQSHGATGQQGYGAAGQHGGKEGTMDKIKDKISGNH

Xh_LD_27A05/ XhLEA2-6:

>141218-14_J01_1A_T7promoter.ab1 635
GGGCCGGTGAAGGACACGGGGGGTGACGACATTCCCTCTAGAATAATTTT
GTTTACTTTAAGAAGGAGATATACATATGCACCACCACCACCACCATCGC
GGATCCGAATTCGAGCTCCGTCGACAAGCTTGCATGGAGGGATACGGGAA
CCATCAGCACTACAACGTTGACGAGTACGGTAACCCGCTGCCCAGTGGGT
ATGGTGAAAATTACGGCCACCAACCTATGCAGCGCCCTGGCGAGTATTAT
GGTAACCAGGGCTACGGCGGCCACCAGCACGGTGCGTACAGCGGCCCTGG
ATATCAGCAGGAGTACGGCAGTGGCCCTAGGCTACAGCGCTCCGGCAGCA
GCTCCAGCTCTGAGGATGATGGGTATGGTGGGAGGAGGAAGAAGGGGCTC
AAGGATAGGATCATGGAGAACCTGCCTGGCCGTCGCTGAACTCGAGCACC
ACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCT
GAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGC
CTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCG
GATTGGCGAATGGGACGCGCCCTGTAGCGGCGCAT

MEGYGNHQHYNVDEYGNPLPSGYGENYGHQPMQRPGEYYGNQGYGGHQHGAYSGPGYQQE
YGSGPRLQRSGSSSSSEDDGYGGRRKKGLKDRIMENLPGRR

>141218-14_K01_1A_T7terminator.ab1 1266
GGGGGGAGGGACTCGCCTTCTCTCGGGCTTTGTTAGCAGCCGGATCTCAG
TGGTGGTGGTGGTGGTGCTCGAGTTCAGCGCCCGTTAGGCAGGTTCTCCA
TGATCCTATCCTTGAGCCCCTTCTTCCTCCTCCCACCATACCCATCATCC
TCAGAGCTGGAGCTGCTGCCGGAGCGCTGTAGCCTAGGGCCACTGCCGTA
CTCCTGCTGATATCCAGGGCCGCTGTACGCACCGTGCTGGTGGCCGCCGT
AGCCCTGGTTACCATAATACTCGCCAGGGCGCTGCATAGGTTGGTGGCCG
TAATTTTCACCATACCCACTGGGCAGCGGGTTACCGTACTCGTCAACGTT
GTAGTGCTGATGGTTCCCGTATCCCTCCATGCAAGCTTGTCGACGGAGCT
CGAATTCGGATCCGCGGTGGTGGTGGTGGTGGTGCATATGTATATCTCCT
TCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTATCCGCTCAC
AATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATC
CTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGT
TGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCC
ACTTCTGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCC
GTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGC
GGTGGCGGTGCTGAACGGCCTCAACCTACTACTGGGCTGCTTGCTAATGC
AGGAGTCGCATAATGGAGAGCGTCGAGATCCCGGACACCATCGAATGGCG
CAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTTC
GGGTGGTGAATGTGAAACCGGTAACCTTATACGATGTCGCATAATATCCC
GGTGTCTCTTTTCGAACGTTTCCCGCGTGGGAAACCGGCCAGCCACGTTT

CTTGGAAAACGCGGGAAAAAAGTGGAAGTGGCCATTGGCGGAACTTAATT
AAATTCCCCACCCCGTGGGACAACCACCTGGGGGGGAAACCCCCCGTTGT
TGATTTGGGTTTGCCCCCTCTCAGGCGGGCCCCCGCCTCTCACCCTCCCC
AAATGGGTGGGGGGCGATCAAATTCTCGCGCCGAATTAAGGGGGGGGCCC
CAGACGGGATGGTTTGTCAATGGGGGAAAAACAAAATTGGGGGTTTCAAA
AGAGCTTGTTTAAGGG

MEGYGNHQHYNVDEYGNPLPSGYGENYGHQPMQRPGEYYGNQGYGGHQHGAYSGPGYQQEYGSGP
RLQRSGSSSSSEDDGYGGRRKKGLKDRIMENLPNGR

Xh_LD_44B08/XhLEA3-5:

>150129-32_I03_10A_T7promoter.ab1 1276
GGGTTGACGGGTACATTGCCCTCTAGAATAATTTTGTTTAACTTTAAGAA
GGAGATATACATATGCACCACCACCACCACCACCGCGGATCCGAATTCGA
GCTCCGTCGACAAGCTTGCATGGCGAGGATTGTGATGGGTTTGTTATTGG
CGGCGATGTTGTTGGTGGCCGTTCATGGAACGACCATGGATGAAGCCAAA
GCAAAGGCGAACGCTGCCGTTCCGAGCATGGATGAAGCCAAAGCGAAGGT
GAACGCTGCCATCCCGACCATGGATGAAACCAAAGAGAAGTTGAATGCTG
CCATGCATGCCGCCAAGGAGAAGAAAGAGGCTTGGTCCGAGTGGGTGGAA
GATAAGCTAGAAGGATTGGGGCTTCAGATGGAAAATAAAATGGCGGCCGC
AGATGACACTACCGATAAAGCTGCAGAAGCCGCCATGAAGGCCAAGGACT
CTGCTTCAGGTGCTGCCGAAACTACATCGGAGAAGGCAGGCGAGAGCAAG
GACACTGTGAAAGATAAGGCGGCGAAGTCGGCGGACTGCATGTCCGGCAA
GGCCGAGGATGTGAAGGGAAAAACCTCCGAAACCACCGGGAAAGCATCGG
AGGGAGCCGGCGGCATGTACGAGAGCGCCAAGCAGAAGATGGAAGAAGCC
TACGCTGCTGCCAAGGATAAGGTGTCGCAGGCCGAAGGCGGCGGAGGGGG
AAGTCCGAGGGAGCTGTAAACTCGAGCACCACCACCACCACCACTGAGAT
CCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGC
TGAGCAATAACTAGCATAACCCCCTTGGGGCCTCTAAACGGGGTCTTGAG
GGGGTTTTTTTGCTGGAAAGGAGGGAACTATATTCCGGGATTGGGCGGAA
TGGGGACGCCGCCCCTGTTAGCGGGCGCCTTTAAGGCGCGGGCGGGGTGT
GGGTGGGTTACCGCGCCAGCGGTGACCCGCTTACACTTTGCCCAGCGCCC
CTAACCGGCCCGGCTCCCTTTTCGCCTTTTCTTTCCCCTTTCCTTTTTCT
CCGCCCCCGTTTCCGCCGGGCTTTTTCCCCCGTTCAAAGCTTCTTAAAAT
CCGGGGGGGCTTCCCCTTTTAAAGGGTTTCCCGAATTTTAATTGGCTTTT
TACGGGGCCCCCTTCCAACCCCCAAAAAAAAACCCTTGGATTTAAGGGGG
GGGAATGGGGTTTCCACCGTTAAATGGGGGGCCCCTTTCCCCCCCTTGGT
TTTCGGCCGGGGTTTTTTTTCCGCCC

MARIVMGLLLAAMLLVAVHGT
TMDEAKAKANAAVPSMDEAKAKVNAAIPTMDETKEKLNAAMHAAKEKKEAWSEWVEDKLE
GLGLQMENKMAAADDTTDKAAEAAMKAKDSASGAAETTSEKAGESKDTVKDKAAKSADCM
SGKAEDVKGKTSETTGKASEGAGGMYESAKQKMEEAYAAAKDKVSQAEGGGGGSPREL

>150129-32_K03_10A_T7terminator.ab1 1304
GGGGGACCTCAGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGT
GGTGGTGGTGGTGCTCGAGTTTACAGCTCCTCGGACTTCCCCTCCGCCGC
CTTCGCCTGCGACACCTTATCCTTGGCAGCAGCGTAGGCTTCTTCCATCT
TCTGCTTGGCGCTCTCGTACATGCCGCCGGCTCCCTCCGATGCTTTCCCG
GTGGTTTCGGAGGTTTTTCCCTTCACATCCTCGGCCTTGCCGGACATGCA
GTCCGCCGACTTCGCCGCCTTATCTTTCACAGTGTCCTTGCTCTCGCCTG
CCTTCTCCGATGTAGTTTCGGCAGCACCTGAAGCAGAGTCCTTGGCCTTC
ATGGCGGCTTCTGCAGCTTTATCGGTAGTGTCATCTGCGGCCGCCATTTT
ATTTTCCATCTGAAGCCCCAATCCTTCTAGCTTATCTTCCACCCACTCGG
ACCAAGCCTCTTTCTTCTCCTTGGCGGCATGCATGGCAGCATTCAACTTC
TCTTTGGTTTCATCCATGGTCGGGATGGCAGCGTTCACCTTCGCTTTGGC
TTCATCCATGCTCGGAACGGCAGCGTTCGCCTTTGCTTTGGCTTCATCCA

TGGTCGTTCCATGAACGGCCACCAACAACATCGCCGCCAATAACAAACCC
ATCACAATCCTCGCCATGCAAGCTTGTCGACGGAGCTCGAATTCGGATCC
GCGGTGGTGGTGGTGGTGGTGCATATGTATATCTCCTTCTTAAAGTTAAA
CAAAATTATTTCTAGAGGGGAATTGTTATCCGCTCACAATTCCCCTATAG
TGAGTCGTATTAATTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGAC
GCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTAT
ATCGCCGACATCACCGATGGGGGAAAATCGGGCTCGCCAACTTCGGGCTC
ATGAGCGCTTGTTTCGGCGTGGGGTATGGTGGCAGGGCCCGGTGGCCGGG
GGAATGTTGGGCGCCATCTCCTTGCATGCACCAATTCCTTGCGGCGGCGG
TGCTTAACGGCCTCAACCTACTAATGGGGTTGTTTCTAAAGGCAGAGTCG
AAAAAGGGGAAACCTTGAAAATCCCGGACACCATCAAAGGGCGCAAAAAC
TTTTCCGGGGAATGGGGATTGATAGCCGCCCGGAAAGAAAATTCAATTTC
CGGGGGGGGGGGAAAGGTGTAAAAAACCGATAAACCTTATATTACAAAAT
GTTCGCCAAAAAAAAATGTTCCCCGGGGTGTTTTCTTCTTTTTACATAGA
AGAA

MARIVMGLLLAAMLLVAVHGTTMDEAKA
KANAAVPSMDEAKAKVNAAIPTMDETKEKLNAAMHAAKEKKEAWSEWVEDKLEGLGLQME
NKMAAADDTTDKAAEAAMKAKDSASGAAETTSEKAGESKDTVKDKAAKSADCMSGKAEDV
KGKTSETTGKASEGAGGMYESAKQKMEEAYAAAKDKVSQAKAAEGKSEEL

Turnitin Report:

Turnitin Originality Report
wtrrob005:editing_Chapter_3_and_4_for_gene_expression_Leaves_and_Roots.docx by Robyn Waters
From MSc Turnit in check (fed67e0f-8799-4204-910c-6b678d404d04)

- Processed on 14-Aug-2015 11:47 SAST
- ID: 561231816
- Word Count: 33046

Similarity Index
16%
Similarity by Source
Internet Sources:
13%
Publications:
11%
Student Papers:
4%

sources:

1

1% match (Internet from 29-Oct-2014)
http://www.plantphysiol.org/content/148/1/6.full

2

< 1% match (publications)
Amara Imen Ikram Zaidi, Khaled Masmoudi, M. Dolors Ludevid, Montserrat Pagès, Adela Goday, and Faical Brini. "Insights Into Late Embryogenesis Abundant (LEA) Proteins in Plants: From Structure to the Functions". American Journal of Plant Sciences, 2014.

3

< 1% match (Internet from 01-Jun-2014)
http://gene-quantification.net/bio-rad-ddpcr-app-guide-6407.pdf

4

< 1% match (Internet from 07-Feb-2013)
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2728378/

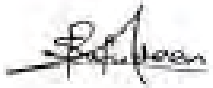
5

< 1% match (publications)
A. A. Covarrubias. "The Enigmatic LEA Proteins and Other Hydrophilins", PLANT PHYSIOLOGY, 07/16/2008

6

< 1% match (Internet from 02-Mar-2015)
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3256378/

7




turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Robyn Waters
Assignment title:	MSc Turnit in check
Submission title:	wtrrob005.editing_Chapter_3_and_ ...
File name:	diting_Chapter_3_and_4_for_gene ...
File size:	6.2M
Page count:	120
Word count:	33,046
Character count:	205,123
Submission date:	14-Aug-2015 11:35AM
Submission ID:	561231815