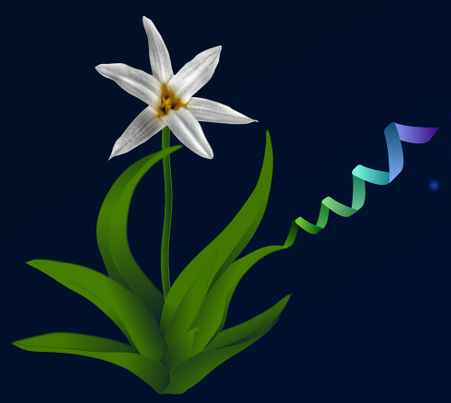
Unfolding Plant Desiccation Tolerance

Mariana Aline Silva Artur 2019

Unfolding Plant Desiccation Tolerance

Evolution, Structure & Function of LEA proteins



Mariana Aline Silva Artur

Propositions

1. The water content-driven coordination of metabolic shutdown and activation of cellular protection is an important contribution to a "footprint" of desiccation tolerance in *Xerophyta viscosa.*
   (this thesis)

2. The contribution of LEA proteins to desiccation tolerance is based on their individualities rather than on their presumed commonalities.
   (this thesis)

3. The lack of a systematic classification accounting for intrinsic disorder (Van der Lee *et al.* 2014. Chemical reviews, 114(13), 6589-6631) and structural plasticity hinders functional prediction and annotation of LEA proteins.

4. Protein intrinsic disorder and induced folding are sources of evolutionary novelties (Conant and Wolfe, 2008. Nature Reviews Genetics, 9(12), 938.).

5. All gene products that have survived millions of years of evolutionary selective pressure have biological significance.

6. Maintaining a natural hair texture is a subjective form of activism that contributes to the normalization of blackness in the society.

7. A true scientist allows big frustrations to last no longer than the excitement of small findings.

Propositions belonging to the thesis entitled:

"Unfolding plant desiccation tolerance: Evolution, structure and function of LEA proteins"

Mariana Aline Silva Artur
Wageningen, 11th January 2019

UNFOLDING PLANT DESICCATION TOLERANCE

Evolution, structure, and function of LEA proteins

Mariana Aline Silva Artur

Thesis committee

Promotor
Dr H. W. M. Hilhorst
Associate professor, Laboratory of Plant Physiology
Wageningen University & Research

Co-Promotor
Dr J. W. Ligterink
Assistant professor, Laboratory of Plant Physiology
Wageningen University & Research.
and Program Scientist Crop Innovation at KeyGene., Wageningen, The Netherlands

Other members
Prof. Dr G. C. Angenent, Wageningen University & Research
Dr J. Buitink, French National Institute for Agricultural Research, INRA, Angers, France
Prof. Dr J. C. M. Smeekens, Utrecht University
Dr R. Geurts, Wageningen University & Research

This thesis was conducted under the auspices of the Graduate School Experimental Plant Sciences.

UNFOLDING PLANT DESICCATION TOLERANCE

Evolution, structure, and function of LEA proteins

Mariana Aline Silva Artur

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 11 January 2019
at 4p.m. in the Aula.

Mariana Aline Silva Artur
Unfolding plant desiccation tolerance: Evolution, structure, and function of LEA proteins
176 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references and summary in English

ISBN: 978-94-6343-386-0
DOI: https://doi.org/10.18174/465557

Preface

"Living beings comprise a whole sequence of levels forming [such] a hierarchy. (...) Each level relies for its operations on all the levels below it. Each reduces the scope of the one immediately below it by imposing on it a boundary that harnesses it to the service of the next-higher level, and this control is transmitted stage by stage, down to the basic inanimate level. (...) Each separate level of existence is of course interesting in itself and can be studied in itself." (Polanyi, 1968)

"The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. (...) At each stage entirely new laws, concepts, and generalizations are necessary, requiring inspiration and creativity to just as great a degree as in the previous one. We expect to encounter fascinating and, I believe, very fundamental questions at each stage in fitting together less complicated pieces into the more complicated system and understanding the basically new types of behaviour which can result." (Anderson, 1972)

In the research field of desiccation tolerance (DT), we aim to 'unfold' the different underpinnings of life without water in tolerant plant species in order to 'unlock' this process in non-tolerant species. The idea of 'unlocking' DT arises in a context where food productivity and security are increasingly challenged by environmental changes such as increased drought, while the world population is predicted to steadily grow to over 8.3 billion of people in the year 2030 (FAO, 2017). In the development of new technologies to tackle these challenges, raising 'basic' questions and conducting fundamental research are of pivotal importance.

In this thesis I will present results of a study encompassing genomics, evolution, protein biochemistry, and molecular biology, and then go to a more general and speculative discussion of the main findings. Such a multilevel approach can aid to the comprehension of the complexity of the different levels that the process of DT involves, and can provide transferrable knowledge to produce more stress-tolerant crops.

Mariana Aline Silva Artur

TABLE OF CONTENTS

Chapter

1 General Introduction

Water to land transition: *Inception*

The colonization of land by plants occurred approximately 500 million years ago, and has massively impacted the terrestrial biosphere. Land plants (embryophytes) evolved from a diverse group of green algae known as streptophyte algae (charophycean algae), a group believed to pre-display features that were relevant for the successful conquest of land (Becker and Marin, 2009; Kenrick and Crane, 1997; Wodniok *et al.*, 2011). Some of the major features for survival on dry land included a series of major innovations in the plant body anatomy, physiology, and biochemistry, as well as adaptive traits such as the formation of an embryo, alternation of generations, cuticles, stomata, and vascularization (de Vries and Archibald, 2018). The algal ancestor of embryophytes is believed to have possessed the ability to equilibrate the cellular water status with that of the environment, which may, consequently, lead to desiccation under water-limiting conditions (de Vries and Archibald, 2018; Kenrick and Crane, 1997). Only a few organisms such as lichens, mosses and a few vascular plants, collectively termed **resurrection species**, still display such ability.

During evolution, duplication and adaptive specialization of pre-existing genes and gene families contributed to the evolution of novelties and plant diversity (Flagel and Wendel, 2009). In this sense, it is likely that genes and gene families, pre-existing in the ancestral green algae genome, gave rise to an essential set of genes necessary for the establishment and adaptation of resurrection species in harsh environments. **Desiccation tolerance** (**DT**), the tolerance to water removal from the cell and ability to re-establish normal metabolic activity after water is available again (Alpert, 2000; Leprince and Buitink, 2010; Oliver *et al.*, 2000), represents one of the key strategies for radiation and colonization of resurrection plants in extreme environments such as rock outcrops. In plants, DT is found in most seeds, pollen, spores, bryophyte gametophytes, and lichens (Smirnoff, 1993).

But, what are the mechanisms by which plants tolerate desiccation (Alpert, 2000)? Recent advances in genomic, molecular, and biochemical studies of desiccation tolerant species have shown that multiple mechanisms work in a concerted manner to allow **cellular protection** and repair against oxidation and loss of macromolecular that may occur during dehydration and rehydration. The players of the protective component of DT correspond to the final line of downstream responses that is initiated via a complex hormonal signalling pathway.

This general introduction puts the chapters of this thesis into context by reviewing the scientific literature on the evolution and molecular basis of DT with a focus on **Late Embryogenesis Abundant proteins** (**LEAs**) as a specific player of the cellular protective responses against desiccation-induced damage.

Survival in a dry world: *Provenance*

Plants may survive water deficit stress by avoiding removal of water from their structures through an active maintenance of cellular homeostasis (*homoiohydry*), or via passive equilibration of the cellular water with that of the environment (poikilohydry) (de Vries and Archibald, 2018). The latter is coupled with DT, which is achieved via remarkable physiological, anatomical, and molecular adaptations.

Desiccation is the complete loss of free water from the cellular structures to about 0.1g H_2O g^{-1} dry mass (Alpert, 2005). A variety of microbes, animals and plants can survive extremely intense and prolonged periods of desiccation. Amongst animals DT is commonly found in nematodes, rotifers, arthropods and tardigrades (Browne *et al.*, 2002; Gusev *et al.*, 2014; Kikawada *et al.*, 2006; Tunnacliffe *et al.*, 2005). In plants DT is rare in adult pteridophytes and spermatophytes, but common in their spores, pollen, and seeds, as well as in a few species of bryophytes and adult angiosperm species commonly named 'resurrection plants', which display variable degrees of DT (Oliver *et al.*, 2000).

The small number of angiosperm resurrection species is due to constraints such as size and geographic localization, since most of the desiccation tolerant species display slower growth and are rarely found in wetter environments (Alpert, 2005; Oliver *et al.*, 2000). DT can vary within populations, families, genera, species and even between body parts of the same individual. Often, DT species are also more tolerant to other stresses such as heat and cold, which is also related to their wide ecological, morphological, biogeographical and taxonomical range (for a review see Alpert, 2000).

The current hypothesis on the evolution of DT, also named ***the co-option hypothesis***, states that during the evolution and radiation of plants, vegetative DT commonly found in bryophytes was lost due to constraints such as plant size and evolution of morphological complexity and became confined to reproductive structures such as pollen and orthodox seeds (Oliver *et al.*, 2000). This hypothesis suggests that the DT displayed in seeds evolved from an ancestral form of vegetative DT, and later in evolution the mechanisms of DT could be successfully activated in vegetative structures of resurrection angiosperm species (Farrant and Moore, 2011; Oliver *et al.*, 2000).

Over the years, researchers in the DT field have been looking for evidences that support this co-option hypothesis. Some interesting reviews have already summarized the main findings on the evolution, physiology, molecular mechanisms, genetics, transcriptomics, proteomics, and metabolomics of DT in seeds and resurrection plants which will not be extensively discussed in this introduction (reviewed by Dinakar and

Bartels (2013), Farrant and Moore (2011), Gechev *et al.* (2012), Hilhorst *et al.* (2018), Illing *et al.* (2005), and Oliver *et al.* (2000).

Some of the major DT mechanisms commonly found in resurrection plants and orthodox seeds consist of arrested growth, inhibition of metabolic activity and activation of protective mechanisms (Costa *et al.*, 2017a; Hilhorst *et al.*, 2018). The latter has called particular attention in recent years due to its potential for improving stress tolerance, with emphasis on drought tolerance, in crops.

Decoding DT in resurrection genomes: *Distinction*

Since the discovery and characterization of its double-helix structure (by Franklin and Gosling (1953) and Watson and Crick (1953)), DNA is undeniably known to hold the major genetic information that is transferred through generations, offering the source of genetic variability on which evolutionary forces work. A genome is the conjunct of genetic information present in the DNA, whether or not translated into RNA, that has been evolutionarily selected to build and maintain the life of an organism.

The development of DNA sequencing technologies has revolutionized the way we study genomes and organisms, allowing a faster advance on studies of evolution, comparative genomics, molecular biology, and genetics. From massively parallel DNA sequencing technologies, it is now possible to perform real-time single-molecule sequencing. In combination, such techniques may facilitate the study of genome diversity in populations, individuals, as well as different cell types within one organism, and allow better assessment of evolutionary history, ancestry, and the fate-decision of cells during development (Shendure *et al.*, 2017). The advances in these techniques leave us with an enormous amount of data that, well analysed, will have an impact on our knowledge of life.

Genome sequencing techniques in combination with transcriptome analysis have been applied to comprehend plant evolution and physiology. In the field of DT, the availability of genome sequences of basal (bryophyte and lycophyte) and angiosperm resurrection species (monocots and eudicots) is enabling the evolutionary assessment of the genomic signatures of DT (Table 1).

Table 1 - DT species with sequenced genome.

Species	Clade	Genome size	PCG*	Reference
Physcomitrella patens	Bryophyte	~500Mb	35,938	Rensing *et al.*, 2007
Selaginella lepidophylla	Lycopodiophyta	~109Mb	27,204	VanBuren *et al.*, 2018
Selaginella tamariscina	Lycopodiophyta	~301Mb	27,761	Xu *et al.*, 2018
Xerophyta viscosa	Monocots	~295.5Mb	25,425	Costa *et al.*, 2017
Oropetium thomaeum	Monocots	~245Mb	28,466	VanBuren *et al.*, 2015; 2017
Boea hygrometrica	Eudicots	~1,691Mb	49,374	Xiao *et al.*, 2015

*PCG – Predicted protein-coding genes.

Many DT features already described at the physiological level are now being demonstrated at the genomic level, and new features have been discovered thanks to the assessment of whole genome sequences. This information enables us to zoom in on the key genomic features responsible for the DT phenotype, facilitating a better assessment of this trait in resurrection species and the comparison with closely-related non-resurrection species, the so-called sister-lineage approach.

The genomes and transcriptomes of basal resurrection plants have revealed the loss of genes associated with the aquatic lifestyle, the development of hormonal signalling pathways (such as auxin and abscisic acid-ABA), and acquisition of genes and expansion of gene families necessary for light, osmotic- and thermo-regulation in the terrestrial lifestyle, such as early-light induced proteins (ELIPs) and late embryogenesis abundant proteins (LEAs) (Khraiwesh *et al.*, 2015; Rensing *et al.*, 2008; VanBuren *et al.*, 2018; Xu *et al.*, 2018).

In angiosperm resurrection plants the co-option hypothesis has been supported by the observation of expansion and increased activity of gene families typically associated with DT in seeds. The molecular and genomic signatures of seeds in vegetative tissues of resurrection plants have been demonstrated by the involvement of similar genes, gene families and pathways that work downstream of the ABI3 (*ABSCISIC ACID-INSENSITIVE3*) transcription factor, the so-called ABI3-regulon, which is responsible for the shutdown of photosynthesis and activation of cellular protective processes (Monke *et al.*, 2012). ABI3 has been demonstrated to be one of the master regulators of the acquisition of DT and longevity in orthodox seeds during embryo development (Delahaie *et al.*, 2013; Monke *et al.*, 2012). In seeds, ABI3 is one of the major regulators of LEA protein expression (Delahaie *et al.*, 2013; Giarola *et al.*, 2017), and, interestingly, the same was observed in the moss *Physcomitrella patens* (Shinde *et al.*, 2012; Yotsui *et al.*, 2016), suggesting that protective mechanisms of DT, mainly the ones related to the functioning of LEAs, are conserved across basal resurrection plants, orthodox seeds and vegetative tissues of angiosperm resurrection species.

Unfolding plant DT: Protection

As plants are sessile organisms, molecular mechanisms that enable protection of subcellular structures in conditions of water limitation seem an obvious adaptive trait that has greatly contributed to the successful colonization of land.

In most plants, the complex responses to water stress are initiated by a signal transduction pathway mediated by increased endogenous levels of ABA, which triggers physiological responses such as stomatal closure, resulting in decreased transpiration and limitation of photosynthesis and plant growth (Hsiao, 1973; Osakabe *et al.*, 2014). As a consequence, a subsequent increase in the generation of reactive oxygen species (ROS) may cause oxidative damage of membranes and proteins (Smirnoff, 1993). Oxidative damage is a result of metabolic imbalances which occur due to changes in metabolic rates under stress. Often, water stress is accompanied by high light intensity and heat, which enhances the misregulation of metabolic reactions and ROS formation that can lead to irreversible cellular damages and subsequent plant death.

In basal resurrection plants such as the bryophytes, the primitive mechanisms of DT involve predominantly molecular mechanisms of protection and repair during desiccation and rehydration, which is partially similar to those of orthodox seeds and vegetative tissues of angiosperm resurrection plants (Oliver *et al.*, 2000). These mechanisms are also conserved in the other kingdoms of life that display DT (Vitt *et al.*, 2014).

In angiosperm resurrection plants the protective mechanisms of DT include several components, distributed between morpho-anatomical and molecular strategies. Morpho-anatomical strategies include leaf curling, rolling or folding, in order to provide protection to the photosynthetic machinery (*'homoichlorophylly'*). A more drastic strategy can occur at the molecular level and consists of the breakdown of chlorophyll, disassembly of chloroplasts and synthesis of anthocyanin (*'poikilochlorophylly'*) (Alpert, 2000). Both strategies are accompanied by the activation of antioxidant systems and accumulation of protective molecules such as sugars and molecular chaperone proteins, such as heat shock-proteins (HSPs) and LEA proteins that work as osmoprotectants (Moore *et al.*, 2007).

Under low water availability, sugars and LEA proteins can form intra-cellular glasses that help to stabilize membranes, avoid solute leakage, control water and ion uptake, and slow down oxidative reactions; conferring long-term stability to the cytoplasm in the dry state (Buitink and Leprince, 2004; Crowe *et al.*, 1992; Wise and Tunnacliffe, 2004). Thus, the ability to synthesize protective molecules at a right

proportion and at the correct moment is a successful ability of both seeds and resurrection plants to achieve DT, and to re-establish normal metabolic activity post-rehydration.

Despite the importance for survival of resurrection plants and orthodox seeds under desiccation, the accumulation of sugars seems not always to be necessary for DT outside of the plant kingdom (Hengherr *et al.*, 2008; Lapinski and Tunnacliffe, 2003). On the other hand, the accumulation of LEA proteins (or LEA-like proteins) is also commonly observed in kingdoms other than *Plantae* (Wise, 2003). These observations might suggest that LEAs are indispensable for DT establishment in living organisms.

A brief history of LEA proteins

In plants, *LEAs* were first discovered in cotton seeds, as distinct sets of mRNAs accumulating during the desiccation phase in the later stages of embryo development (Dure *et al.*, 1981b; Galau *et al.*, 1986). The expression of some of these mRNAs was found to be induced by endogenous ABA which led to differential protein accumulation patterns, sometimes hours after the detection of their respective mRNA, revealing that their regulation occurs at both the transcriptional and posttranscriptional levels during seed maturation (Chatelain *et al.*, 2012; Espelund *et al.*, 1992; Galau *et al.*, 1987; Hughes and Galau, 1987; Hughes and Galau, 1991; Verdier *et al.*, 2013).

Seeds that display a prolonged desiccation phase during maturation are called orthodox seeds and, contrary to recalcitrant seeds, can be stored for long periods of time in the dry state without losing their germination ability (Dekkers *et al.*, 2015; Marques *et al.*, 2018). During the development of orthodox seeds, embryogenesis and maturation constitute two important phases that are, respectively, important for the formation of the basic root-shoot body pattern and the acquisition of important traits such as dormancy, longevity and storability (Angelovici *et al.*, 2010; Bewley *et al.*, 2012; Leprince *et al.*, 2017; Ooms *et al.*, 1993b; Vicente-Carbajosa and Carbonero, 2005).

Expression of *LEAs* was found to be affected in mutants of ABI3 (Finkelstein, 1993; Nambara *et al.*, 1995; Ooms *et al.*, 1993b). Later, it was found that LEAs work downstream of ABI3, being part of the ABI3 regulon of essential genes for seed DT (Delahaie *et al.*, 2013; Monke *et al.*, 2012). Interestingly, it was found that the expression of a subset of *LEA* genes is controlled by the transcription factor ABI5, which physically interacts with ABI3, and also regulates seed maturation and longevity (Bies-Etheve *et al.*, 2008; Carles *et al.*, 2002; Lopez-Molina *et al.*, 2002; Zinsmeister *et al.*, 2016). Even more interesting was the finding that a protein encoded by a gene controlling seed dormancy, *DELAY OF GERMINATION 1* (*DOG1*), physically interacts with ABI3 and affects the expression of the *ABI5* gene, thus leading to altered LEA expression patterns

(Bentsink *et al.*, 2006; Dekkers *et al.*, 2016). These findings suggests that LEAs may be similarly important for seed DT, longevity, and dormancy.

Other studies have shown that LEA expression may also be induced by exogenous application of ABA and by other water-related stresses such as salinity, cold, heat and drought (Amara, 2014; Delseny *et al.*, 2001; Espelund *et al.*, 1995; Hughes and Galau, 1991; Maia *et al.*, 2011). *LEA* mRNA expression was also found to be reactivated in desiccation sensitive tissues of germinated seeds treated with an osmoticum such as polyethylene glycol (PEG) (Maia *et al.*, 2011) and of seedlings of *Xerophyta viscosa* treated with ABA (Costa *et al.*, 2017a). These findings suggest an universal role for these proteins in modulating cellular homeostasis under stress and corroborates the hypothesis that *LEA* genes are essential players of the protective component of DT.

Due to the large number of *LEA* genes expressed during desiccation in seeds of several species and vegetative tissues of resurrection plants, several attempts have been made over the years to classify LEAs in order to infer knowledge about common biological functions. Suggested classifications have been based on temporal patterns of expression and inducibility by ABA (Hughes and Galau, 1989), amino acid sequence (Bies-Etheve *et al.*, 2008; Dure, 1993; Dure *et al.*, 1989a), peptide composition using a computational approach (Tunnacliffe and Wise, 2007; Wise, 2003; Wise and Tunnacliffe, 2004) or motif composition (Battaglia *et al.*, 2008; Hunault and Jaspard, 2010; Hundertmark and Hincha, 2008) (Table 2).

Table 2 – Main classification of LEA proteins.

PFAM	Dure *et al.*, 1989	Bray 1993	Tunnacliffe and Wise, 2007	Battaglia *et al.*, 2008	Bies-Ethève *et al.*, 2008	Hundertmark and Hincha, 2008	LEAPdb 2010
PF00257	D11	Group2	Group2	Group2	Group2	Dehydrin	Class1 to 4
PF00477	D19&D132	Group1	Group1	Group1	Group1	LEA_5	Class5
	D7	Group3	Group3	Group3A	Group6		
PF02987	D29	Group5	...	Group3B	...	LEA_4	Class6
PF03168	D95	...	...	Group5C	Group7	LEA_2	Class7&8
PF03242	D73	...	LEA5	Group5B	Grou6	LEA_3	Class9
	...			Group4A			
PF03760	D113	Group4	Group4	Group4B	Group4	LEA_1	Class10
PF04927	D34	Group6	Group6	Group5A	Group5	SMP	Class11
				Group6			
PF10714	...	...	...	Group7	Group8	PvLEA18	Class12

Retrieved from LEAPdb (http://forge.info.univ-angers.fr/~gh/Leadb/index.php) (Hunault and Jaspard, 2010).

Despite extensive efforts, different authors prioritize distinct classifications, revealing that a consensus nomenclature is still far from being adopted. Most of the difficulties in classifying LEAs have their origin in the high variability of gene expression

patterns and protein accumulation profiles, as well as low amino acid sequence conservation.

In order to infer commonalities among LEA protein functions, recent studies have attempted to characterize LEAs of different plant species, not only at a molecular and functional level, but also at the genomic level. Genome-wide identification of *LEA* genes has been performed in several species, including rice (*Oryza sativa*) (Wang *et al.*, 2007), *Arabidopsis thaliana* (Bies-Etheve *et al.*, 2008; Hundertmark and Hincha, 2008), potato (*Solanum tuberosum*) (Charfeddine *et al.*, 2015), black cottonwood (*Populus trichocarpa*) (Lan *et al.*, 2013), chinese plum (*Prunus mume*) (Du *et al.*, 2013), legumes (Battaglia and Covarrubias, 2013), tomato (*Solanum lycopersicum*) (Cao and Li, 2015), sweet orange (*Citrus sinensis L.* Osb.) (Pedrosa *et al.*, 2015), maize (*Zea mays*) (Li and Cao, 2016), rapeseed (*Brassica napus*) (Liang *et al.*, 2016), melon (*Cucumis melo*) and water melon (*Citrullus lanatus*) (Altunoglu *et al.*, 2017), and three cotton genotypes (Magwanga *et al.*, 2018). In this thesis we propose that the variable distribution of LEA families in angiosperm genomes, and the distinct origins and evolutionary patterns are one of the sources of the high protein variability and difficulties in their classification.

LEA protective function: *One fold matters*

It is now becoming clear that the appearance of LEA protein families have contributed to the evolution of DT in the different kingdoms of life. The question now is: What do LEA proteins do? This question was already raised by (Wise and Tunnacliffe, 2004) in their *'POPP the question'* publication, wherein the authors developed a computational approach to investigate LEA protein structure and infer common functions.

Most LEA proteins are hydrophilic and small (Battaglia *et al.*, 2008; Dure *et al.*, 1989b; Garay-Arroyo *et al.*, 2000). These characteristics render heat-stability and enable anti-aggregation and protein stabilizing properties (Amara, 2014; Chakrabortee *et al.*, 2007; Goyal *et al.*, 2005c). Furthermore, several LEAs are also intrinsically disordered proteins (IDPs) as they lack secondary structure in solution (Tompa, 2012; Uversky, 2002; Uversky, 2009). As IDPs, LEAs display local folding transition induced by changes in their environment such as alteration in temperature, pH, presence of binding targets, osmolites, and variable solvent concentrations (Bremer *et al.*, 2017; Cuevas-Velazquez *et al.*, 2016; Hundertmark *et al.*, 2012; Mouillon *et al.*, 2006; Popova *et al.*, 2011; Rivera-Najera *et al.*, 2014; Shih *et al.*, 2012; Shih *et al.*, 2010; Soulages *et al.*, 2002; Uversky, 2009). Under desiccation, some LEAs will progressively adopt secondary structure, which has implications for their physiological role (**Figure 1**) (Olvera-Carrillo *et al.*, 2011).

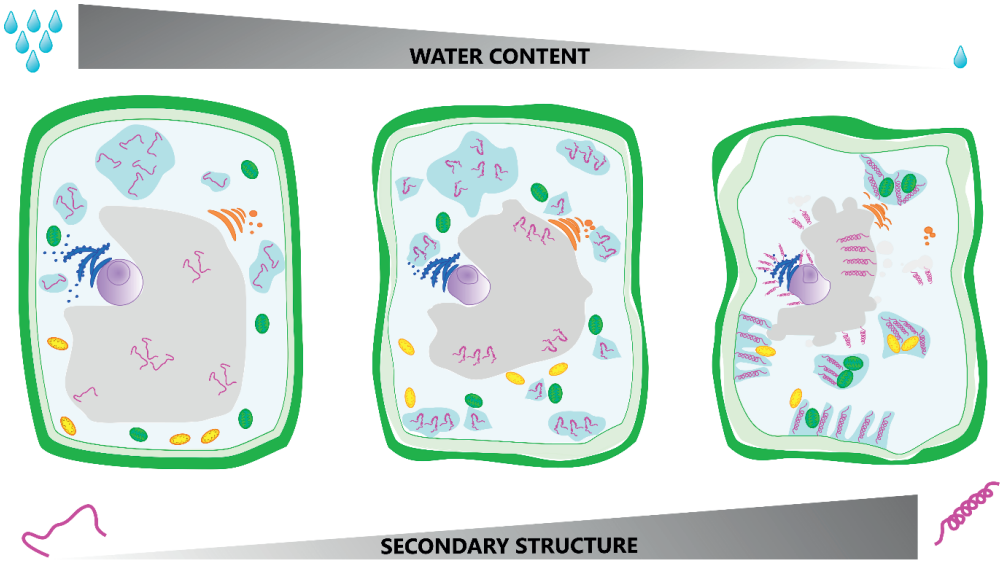


Figure 1 – **Desiccation-induced folding of intrinsically disordered LEA proteins promotes cellular stability**. During desiccation, the water content of the cell decreases which can cause loss of structures of macromolecules. Some LEA proteins acquire higher degree of secondary structure in lower cellular water content, being able to protect the intracellular milieu. The cell wall is represented in dark green, the plasma membrane in light green, and the hydration shell indicated in blue surrounding LEA proteins (in pink). Grey – vacuole, purple - nucleus and nucleolus, dark blue - endoplasmic reticulum and ribosomes, dark orange - Golgi apparatus and Golgi vesicles, yellow - mitochondria, green – chloroplasts.

The presence of intrinsically disordered regions (IDRs) seems to facilitate the formation of homo- and heterodimers, and the interaction of functional binding motifs with multiple targets (Cuevas-Velazquez *et al.*, 2016; Hernandez-Sanchez *et al.*, 2017; Popova *et al.*, 2015; Tolleter *et al.*, 2007). The IDP characteristic is thought to facilitate glass formation in the cytosol, thus trapping cellular macromolecules into an amorphous matrix and, hence, avoiding aggregation (Boothby *et al.*, 2017; Chakrabortee *et al.*, 2010). Some LEAs are indeed able to protect enzymes from thermal and chemical inactivation and aggregation, acting as 'molecular shields' by forming a physical barrier, or as molecular chaperones by interacting and modulating enzyme conformation under stress (Agarwal *et al.*, 2017; Furuki and Sakurai, 2016; Goyal *et al.*, 2005c; Halder *et al.*, 2017; Kovacs *et al.*, 2008; Liu *et al.*, 2016b; Nakayama *et al.*, 2007).

However, not all LEA proteins are predicted to display an IDP character (Wise and Tunnacliffe, 2004) and despite the development of several IDP and IDR prediction tools, the ability to undergo disorder-to-order transitions can only be proven with *in*

vitro analysis. Thus, their actual *in vivo* appearance and possible functions are still regarded as putative.

Nowadays, the most common ways used to investigate if a protein displays an IDP structure are performed with a combination of *in silico* predictions together with *in vitro* experimental analyses using techniques such as X-Ray crystallography, Nuclear Magnetic Resonance (NMR) and Circular-Dichroism (CD) spectroscopy. X-ray crystallography and NMR are the two primary methods to study protein structure at high resolution which, in combination, increases the power of solving protein structures (Yee *et al.*, 2005). Despite limitations in its power, acquisition of CD-spectra is a quicker method, requires lower amounts of purified proteins and less extensive data processing and gives insight into environment-induced folding of IDPs, being of advantage as a preliminary method to pre-select candidates before using more extensive methods (Receveur-Brechot *et al.*, 2006).

Heterologous expression of IDPs in *Escherichia coli* can give insights into their function *in vivo*. Several LEA proteins displaying IDP characteristics were shown to enhance bacterial growth and survival under stressful conditions (Boothby *et al.*, 2017; Drira *et al.*, 2015; Gao and Lan, 2016; Hu *et al.*, 2016; Jiang *et al.*, 2017; Ling *et al.*, 2016; Liu *et al.*, 2016a; Rakhra *et al.*, 2017; Saucedo *et al.*, 2017; Shi *et al.*, 2016; Wang *et al.*, 2017; Zhang *et al.*, 2014; Zhou *et al.*, 2017). Also, ectopic (over-) expression of individual LEAs in plants has been shown to be beneficial for plant survival (Babu, 2004; Chen *et al.*, 2015; Fu *et al.*, 2007; Xiao *et al.*, 2007).

The scope of this thesis

This thesis encompasses genomic, transcriptomic, gene family evolution, molecular, biochemical, and physiological principles and tools in order to understand the basis of DT, with a focus on the role of LEA proteins in seeds and resurrection plants (**Figure 2**).

In **Chapter 2** I provide an overview of the genomic and transcriptomic features of the monocot resurrection species *Xerophyta viscosa*, and provide insight into the contribution of molecular mechanisms, such as the expression of LEA proteins, to the DT phenotype. In **Chapter 3** I explore the myriad of *LEA* genes in the genome of 60 species using comparative genomics and phylogeny, in order to investigate LEA evolution, conservation, and diversification in the plant lineage. In **Chapter 4** I performed a comprehensive structural analysis of six *X. viscosa* LEA proteins using CD-spectroscopy combined with *in silico*, *in vitro*, and *in vivo* analysis of their putative function. One of the *X. viscosa* LEA proteins was further investigated in an heterologous

expression system in the model plant species *Arabidopsis thaliana*, confirming its role in plant stress tolerance.

Knockout and knockdown mutants of *LEA* genes of *A. thaliana* were investigated in **Chapter 5** to understand the contribution of specific LEA proteins to seed traits such as DT, longevity, dormancy, and germinability.

In **Chapter 6** I place all the chapters into a global context, and discuss how this study improves our understanding about the contribution of LEAs to plant DT. Also, I provide examples for the applicability of this knowledge in the scientific and public communities, and future perspectives for this research.

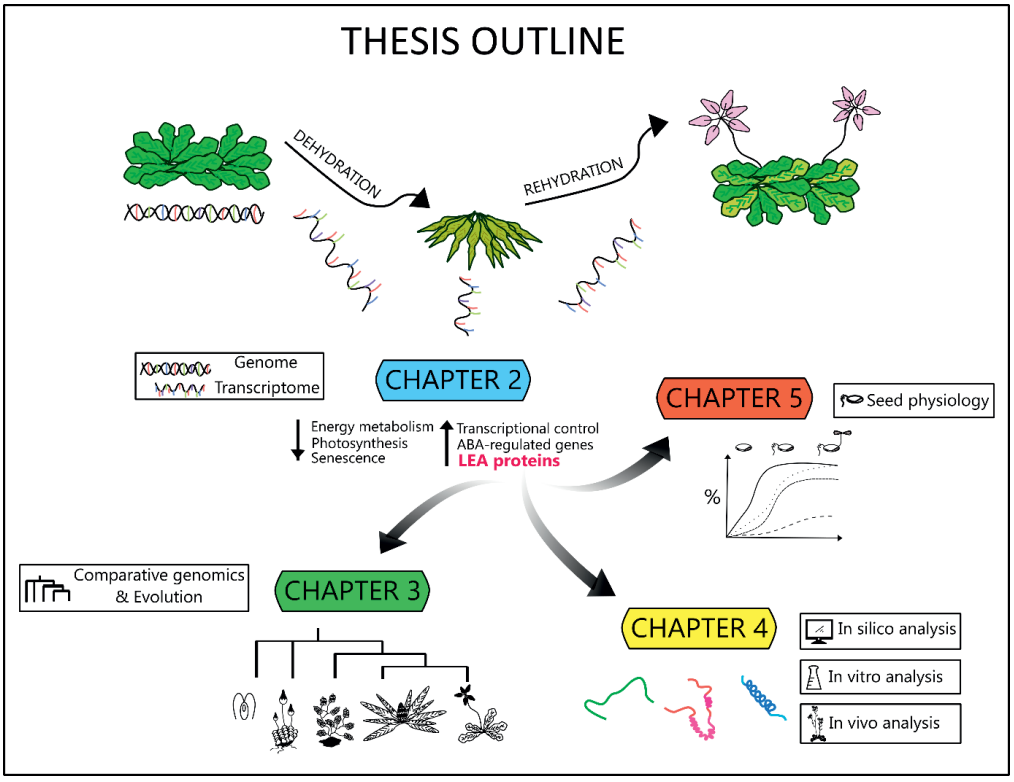


Figure 2 – Schematic representation of the chapters and the main analysis performed in this thesis.

Note added in proof. The official alteration of the scientific name of *Xerophyta viscosa* (Baker) to *Xerophyta schlechteri* (Baker) N.L.Meneses is currently under discussion.

Chapter

2 A 'footprint' of desiccation tolerance in the genome of *Xerophyta viscosa*

Maria-Cecília D. Costa[1,7*], **Mariana A. S. Artur**[1*], Julio Maia[1], Eef Jonkheer[1,2], Martijn F.L. Derks[1,2], Harm Nijveen[1,2], Brett Williams[3], Sagadevan G. Mundree[3], José M. Jiménez-Gómez[4], Thamara Hesselink[5], Elio G.W.M. Schijlen[5], Wilco Ligterink[1], Melvin J. Oliver[6], Jill M. Farrant[7], Henk W.M. Hilhorst[1]

Nature Plants v. 3, n. 17038 (2017)

[1]Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands.
[2]Bioinformatics Group, Wageningen University, Wageningen, The Netherlands.
[3]Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, Australia.
[4]Max Planck Institute for Plant Breeding; Research, Department of Plant Breeding and Genetics, Cologne, Germany.
[5]Bioscience, Wageningen Plant Research, Wageningen, The Netherlands.
[6]USDA-ARS-MWA-PGRU, University of Missouri, Columbia, MO 65211 USA.
[7]Dept of Molecular and Cell Biology, University of Cape Town, Cape Town, South-Africa.
**These authors contributed equally to this work.*

Summary

Desiccation tolerance is common in seeds and various other organisms but only a few Angiosperm species possess vegetative desiccation tolerance. These 'resurrection species' may serve as ideal models for the ultimate design of crops with enhanced drought tolerance. To understand the molecular and genetic mechanisms enabling vegetative desiccation tolerance we produced a high-quality whole-genome sequence for the resurrection plant *Xerophyta viscos*a and assessed transcriptome changes during its dehydration. Data revealed induction of transcripts typically associated with desiccation tolerance in seeds, and involvement of orthologues of ABI3 and ABI5, both key regulators of seed maturation. Dehydration resulted in both increased, but predominantly reduced transcript abundance of genomic 'clusters of desiccation-associated genes' (CoDAGs), reflecting the cessation of growth that allows for the expression of desiccation tolerance. Vegetative desiccation tolerance in *X. viscosa* was found to be uncoupled from drought-induced senescence. We provide strong support for the hypothesis that vegetative desiccation tolerance arose by redirection of genetic information from desiccation tolerant seeds.

Main

The grass family (Poaceae) is arguably the most important contributor to global food security. However, poaceous staple crops, such as maize, corn, rice, and wheat do not survive the extreme water loss that is inevitably brought about by extended periods of drought. In contrast, their seeds are desiccation tolerant, and they can be dried to water contents as low as 1-5% on a fresh weight basis without losing viability (Kermode, 1997).There are some 135 angiosperm species, termed "resurrection plants", which produce desiccation tolerant seeds and possess desiccation tolerant vegetative tissues (Black and Pritchard, 2002; Gaff, 1971; Porembski, 2011). Vegetative desiccation tolerance (DT) first arose with the transition from aquatic to terrestrial life forms, when both the probability of experiencing adverse conditions and the survival cost of such conditions were high. As plants expanded into terrestrial habitats and developed tracheids to move water from the substrate to their aerial parts and more complex ecosystems were established, the slow growth characteristic of desiccation tolerant plants was limiting their competitive ability. This favoured the loss of DT in vegetative tissues, the development of mechanisms to prevent water loss (physiological and morphological) and the confinement of DT to seeds, spores and pollen grains where it was required for dispersal and preservation of genetic resources (Alpert, 2006; Oliver *et al.*, 2005). It has been proposed that vegetative DT reappeared in the angiosperms, presumably in response to colonization of environmentally demanding habitats, in at least thirteen separate lineages to evolve the present day resurrection plants (Oliver *et al.*, 2000; Oliver *et al.*, 2005; Porembski, 2011).

The myriad of genetic changes experienced in the evolution of resurrection plants that enabled vegetative DT are not completely understood. An improved understanding of these changes will aid the development of crop improvement strategies for tolerance of water loss and survival of extreme drought conditions (Farrant *et al.*, 2015). The pressure to develop such crop varieties has been intensified by predictions of a near future with increased drought and declining water resources in the world's main agricultural areas (Dai, 2013). This pressure is strongest in developing countries, which struggle to maintain robust breeding capabilities and need support to develop stress tolerant crops critical for food security (Farrant *et al.*, 2015). In this context, the study of resurrection plants, such as *Xerophyta viscosa* (Velloziaceae), will bring valuable information to bear on improving stress tolerance in crops. *X. viscosa* (Figure 1) is a monocotyledonous plant species and, thus, phylogenetically closely related to staple cereal crops. It is thus an ideal model for understanding plant requisites to tolerate extreme dehydration.

Figure 1 - *X. viscosa* phenotypes. The chasmophytic nature of growth in the natural environment (a), a typical fully dehydrated plant (withholding water from the whole plant over a period of 25 days until < 5% RWC) (b) and rehydrated plant (at full turgor after 5 days of watering) (c).

We have produced a high-quality whole-genome sequence and assembly for *X. viscosa* from a mixed data set of Illumina and Pac-Bio reads along with a full assessment of the transcriptomal changes that occur in young seedlings in response to exogenous abscisic acid (ABA) and in adult plants during desiccation and rehydration.

A high-quality octoploid genome assembly

The source of genomic DNA consisted of individuals (five for Illumina and one for PacBio sequencing) grown from seeds harvested from a population of *X. viscosa* collected from the Buffelskloof Nature Reserve (Mpumalanga Province, South Africa). *X. viscosa* has a high level of heterozygosity due to self-incompatibility. High levels of

heterozygosity could cause several haplotypes to be assembled in different contigs causing duplicates in the final haploid assembly. However, we used Falcon (Chin *et al.*, 2016) which is a tool designed to deal with high levels of heterozygosity and, thus, no measures were taken to reduce genome complexity. The genome was sequenced and assembled (Boetzer and Pirovano, 2014; Song *et al.*, 2014; Ye *et al.*, 2016; Ye *et al.*, 2012) using a whole-genome shotgun approach which combined 23 Gb of raw paired-end Illumina reads (~77× coverage) and 17.5 Gb of PacBio long reads (~58× coverage). This high-quality assembly covers 99.8% of the genome and consists of 1,811 contigs and 896 scaffolds (Table 1).

Table 1. Properties of the *X. viscosa* genome. N50 – scaffold size above which 50% of the total length of the sequence assembly can be found. L90 – number of contigs whose summed length contains at least 90% of the sum of the total length of the sequence assembly.

Assembly	Number	N50 (Mb)	L90	Total length (Mb)	Alignment rate (%)
Scaffolds	896	1.67	238	295.5	96.7
Contigs	1,811	1.11	448	293.6	96.3
Additional haplotype contigs	6103	0.03		125	

Annotation	Number	Mean length (bp)	Density	Genome percentage (%)
BUSCO (missing/total)	57/956			
BUSCO* (missing/fragmented)	51/31			
GC content				36.51
Protein coding genes	25,425	4,444.5	-	38.2
Exons	149,027	265.9	5.9 exons/gene	13.4
Introns	120,004	610.7	4.7 exons/gene	24.8
rRNA	165	429.5	-	<0.01
snRNA	140	86.69	-	<0.01
tRNA	289	74.91	-	<0.01
Transposable elements	116,932	-	-	18.3
miRNA	165	126.6	-	<0.01
Orphan genes**	1,372	143.4	-	0.066 (5.4% of genes)

Polymorphisms	Number	Density
SNPs	1,384,518	4.7 kb^{-1}
INDELS	375,931	1.3 kb^{-1}
Multi-allelic sites	65,241	0.2 kb^{-1}

Repeat class	Number of elements	Genome percentage (%)
Retrotransposon	63,029	0.12
Long Terminal Repeat (LTR)	45,072	0.10
Gypsy (RLG)	30,293	0.08
Copia (RLC)	12,725	0.02
Penelope (RPX)	119	<0.01
Unknown LTR (RLX)	288	<0.01
LINE (RIL)	15,055	0.01
SINE (RSX)	2,902	<0.01
DIRS (RYD)	127	<0.01
Unknown retrotransposon (RXX)	292	<0.01

DNA transposon	56,603	0.07
Maverick (DMX)	584	<0.01
Unknown DNA transposon (DXX)	1,776	<0.01
No category	154,855	0.18

***** after MAKER annotation

**no hits in Swissprot, TrEMBLwith cut off $1e^{-10}$ and PFAM cut off $1e^{-6}$

A k-mer analysis of PacBio data indicated a (haploid) genome size of 295.5 Mb. Based on allele frequencies (Yoshida *et al.*, 2013), we determined the *X. viscosa* genome to be an octoploid (Supplementary figure 1). Polyploidy has been associated with vigour and size in crops, has been proposed to facilitate better environmental adaptability, and is regarded as one of the major drivers of speciation, and thus evolution. It follows that polyploid species may be more competitive and invasive in new environments (deMelo *et al.*, 1997). The African Velloziaceae species are all polyploid and assumed to have originated on the South-American continent, where family members are largely diploid (deMelo *et al.*, 1997). Thus, polyploidization may have contributed to successful invasion of the Velloziaceae in southern Africa. Our data indicates that members of the Velloziaceae may be octoploids and not hexaploids, which is the current consensus of opinion.

X. viscosa has 48 small chromosomes (estimated length of 1.5-2.0 microns, Supplementary figure 1). We assembled a 295.5 Mb haploid genome and 90% of the assembly is contained in 238 scaffolds. The scaffold N50 (the scaffold size above which 50% of the total length of the sequence assembly can be found) is 1.67 Mb which is very high for an octoploid genome. It approaches the N50 of the diploid resurrection species *Oropetium thomaeum* (N50 2.4 Mb, genome size 245 Mb) (VanBuren *et al.*, 2015) and is much greater than the N50 of the diploid resurrection species *Boea hygrometrica* (N50 110 kb, ~1,691-Mb sequenced genome)(Xiao *et al.*, 2015) or the tetraploid *Chenopodium quinoa* (N50 87 kb, genome size 1,448 Mb). In addition to the 295 Mb final assembly, we assembled 125 Mb of additional haplotypes in 6,103 contigs.

The GC content is 36.5% across the genome (Table 1). Greater GC contents have been associated with species that grow in seasonally cold and/or dry climates, perhaps suggesting an advantage of GC-rich DNA during cell freezing and desiccation (Smarda *et al.*, 2014). Whereas *B. hygrometrica* has a relatively high GC content of 42.30%, it is only 34.86% in *O. thomaeum*, thus not supporting a general positive correlation between DT and GC content.

A total of 25,425 protein-coding genes were annotated, 97% of the encoded proteins exhibit high sequence similarity to proteins in the TrEMBL database (Boeckmann *et al.*, 2003), 85% in Swiss-Prot (Boeckmann *et al.*, 2003) and 84% in InterPro (Mitchell *et al.*, 2015). The percentage of orphan genes (5.4 %, Table 1), or genes that do not share any similarity with genes in other species, is considerably lower

than expected for eukaryotic genomes (10-20%). This low percentage may suggest that the acquisition of vegetative DT by *X. viscosa* relied more on the redirection of genetic information than on the genesis of novel genes.

Transposable elements (TEs) account for 18% of the genome, which is a surprisingly low number (Table 1) considering that TEs in the genomes of *O. thomaeum* and *B. hygrometrica* account for 75% and 43%, respectively. Our high-quality data rules out collapsed or incomplete sequence assemblies. Plant TEs have been proposed to regulate DT in the dicot resurrection species *C. plantagineum* (Hilbricht *et al.*, 2008). Taken together, the proportion of TEs in a plant genome does not appear to be related to DT in general.

Orthologous clustering of the *X. viscosa* proteome with 15 plant genomes identified 15,450 orthologous groups (OGs) in common, with 20 OGs present specifically in *X. viscosa* (Supplementary table 1). The OGs were used to map signatures of expansions and contractions of gene families (Supplementary figure 2). Overall, *X. viscosa* did not undergo extensive expansions or contractions of gene families. While many gene families were shared between *X. viscosa* and other monocots, such as the desiccation sensitive but relatively drought tolerant *Eragrostis tef* (Cannarozzi *et al.*, 2014) (*X. viscosa* and *E. tef* share 70% of OGs) and the resurrection plant *O. thomaeum* (*X. viscosa* and *O. thomaeum* share 72% of OGs), clearly there are several expansions and contractions unique to *X. viscosa* (Supplementary figure 2). The *X. viscosa* exclusive expansions include several gene families associated with cellular stress and metabolic regulation, such as putative late embryogenesis abundant (LEA) proteins, heat shock domain-containing proteins, putative WRKY transcription-factors and MADS-box domain-containing proteins. The contractions unique to *X. viscosa* include gene families related to control of gene expression, such as plastid transcriptionally active chromosome proteins and DNA helicases. The low overlap in expanded or contracted gene families between *X. viscosa* and *O. thomaeum* may indicate different genetic architectures underlying the resurrection phenotype observed in both species and hint at independent evolution to reacquire vegetative DT(Gaff and Oliver, 2013).

Resurrection physiology

X. viscosa seeds and adult plants are desiccation tolerant. However, individuals are not continuously desiccation tolerant from seed to adult plant. They lose DT briefly upon germination, which recovers gradually during seedling development, first in shoots and later in roots. This recovery of DT can be induced earlier in shoots, but not in roots, by application of ABA (Supplementary figure 3). In shoots, exogenous ABA induced the

accumulation of transcripts of genes involved in chlorophyll degradation, translational control of gene expression, and transport, whereas it decreased transcript abundance of genes related to photosynthesis, energy metabolism, and ABA biosynthesis (Supplementary table 2). In roots, transcript abundance of genes involved in chlorophyll degradation and responsiveness to brassinosteroids increased whereas transcripts of genes related to energy metabolism and plant responses to external signals (mainly by modifying cell walls) were depleted (Supplementary table 2).

In adult plants, stomata closure and carbon gain from photosynthesis ceases when water content drops below 55% relative water content (RWC) (Farrant *et al.*, 2016; Mundree and Farrant, 2000). In desiccation sensitive species, cessation of carbon gain coupled with continued water loss results in metabolic stress, whereas resurrection plants redirect their metabolism towards subcellular protection and ultimate quiescence (Farrant *et al.*, 2016; Gaff and Loveys, 1993). In *X. viscosa*, the molecular signature of this metabolic redirection is reflected in the higher number of genes that exhibit differential expression at 40% RWC (1.0 gH_2O g^{-1} dwt) compared to other hydrated states (Supplementary figure 4). At 40% RWC, energy metabolism is strongly repressed and transcripts of genes related to chlorophyll degradation significantly accumulated. Significant accumulation of transcripts related to chlorophyll synthesis and chloroplast and thylakoid formation were also observed during dehydration at 40%RWC (Supplementary table 3) reflecting the strategy of poikilochlorophylly adopted by *X. viscosa*. Poikilochlorophyllous resurrection plants dismantle their photosynthetic apparatus (thylakoids and chlorophyll) in a controlled manner during dehydration, and reconstitute this upon rehydration utilizing transcripts stably stored in desiccated tissues (Bewley, 1979; Csintalan *et al.*, 1996; Farrant *et al.*, 2015; Tuba *et al.*, 1998).

This metabolic redirection is also evident from the grouping of the differentially expressed genes in four distinct self-organizing maps (SOMs, Figure 2).

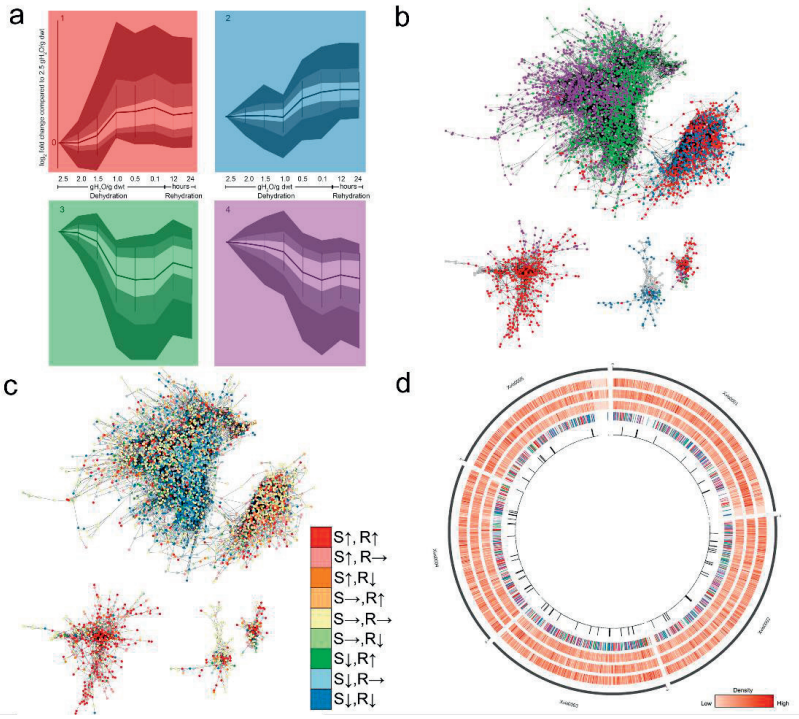


Figure 2 - Genomic organization of *X. viscosa*. a - Self-organizing maps of gene expression (SOMs) displaying log_2 fold change of transcripts relative to full hydration; b - Gene co-expression network with nodes colored according to SOMs; c - Co-expression network of differentially expressed genes after ABA application to shoots (S) and roots (R) of seedlings; d - Circos map with features of the five longest assembled scaffolds of the *X. viscosa* genome. The tracks from outside to inside display gene density, transposable element density, domain density, fold change of gene expression to hydrated state (color codes as in SOMs) and presence of putative CoDAGs.

From these, it is evident that two major changes in gene expression occur: between 60% RWC (1.5 $gH_2O\ g^{-1}$ dwt) and 40% RWC (1.0 $gH_2O\ g^{-1}$ dwt, SOMs 1 and 3) and between 40% RWC and 20% RWC (0.5 $gH_2O\ g^{-1}$ dwt, SOMs 2 and 4). Genes related to protein folding, protection and translation control are enriched exclusively in SOM 1, while genes related to nuclear import and control of gene expression are enriched exclusively in SOM 2 (Supplementary table 2). Likewise, a number of transcription factor-encoding genes and co-regulators were highly induced in desiccated leaves of the resurrection species *Haberlea rhodopensis* (Gechev *et al.*, 2013b), reflecting the massive transcriptional reprogramming behind the metabolic redirection. Genes related to lipid metabolism, nucleotide metabolism and protection against oxidative stress are enriched exclusively in SOM 3, while genes related to energy metabolism, photosynthesis, water transport and genetic information processing are enriched exclusively in SOM 4. Decline of transcripts related to energy metabolism and

photosynthesis was also observed in the resurrection species *H. rhodopensis* and *Craterostigma plantagineum* as decreasing metabolic activity is a general primary target during dehydration (Dinakar and Bartels, 2013; Gechev *et al.*, 2013b; Rodriguez *et al.*, 2010).

A gene co-expression network approach revealed a large overlap in the patterns of gene activation and -repression in ABA-treated seedlings and drying mature leaves (Figure 2). Young seedlings of *X. viscosa* are desiccation sensitive and obtain DT only later in development, but can be made tolerant by application of ABA (Supplementary Figure 3). Thus, this overlap points to similarities between seed- and vegetative-DT (Costa *et al.*, 2016). In both systems, metabolism is shut down in an orderly manner and protection mechanisms activated (Farrant *et al.*, 2016; VanBuren *et al.*, 2015; Williams *et al.*, 2015). Moreover, upon dehydration, both in seeds and in poikilochlorophyllous resurrection species, chloroplasts are disassembled and chlorophyll degraded. All these (ABA-controlled) events have so far been considered seed-specific but here we show that vegetative-DT in Angiosperms is also regulated by ABA, employing the same genes as in seeds. This supports a hypothesis that DT was reactivated in vegetative tissues, presumably utilizing genes associated with DT in the seeds of the species or family as a 'blueprint', and then later modified to result in species or family specific DT mechanisms (Gaff and Oliver, 2013).

In the context of the gene co-expression network, the genes in SOMs 3 and 4 are more tightly co-regulated than the genes in SOMs 1 and 2, highlighting the importance of a coordinated shutdown of metabolic and cellular processes during dehydration. Accordingly, most of the network's hubs (nodes with high number of connections to other nodes, or high degree) represent genes related to energy metabolism (Supplementary figure 4 and Supplementary Data Table).

The network aided the identification of ABA-independent gene expression (Supplementary figure 5). Although ABA-independent signalling pathways mediated by DREB2 and DREB1/CBF have been described (Todaka *et al.*, 2015), other regulators not described yet may also be acting. Around 20% of the nodes in the network represent genes differentially expressed in response to desiccation but not in response to exogenous ABA. The GO categories enriched in this group of genes are related to signalling (*consequence of signal transduction* and *generation of a signal involved in cell-cell signalling*) and transcriptional regulation (*positive regulation of transcription regulator activity, regulation of sequence-specific DNA binding transcription factor activity* and *regulation of transcription regulator activity*). These categories suggest the presence of ABA-independent regulators not described yet.

Clusters of Desiccation-Associated genes (CoDAGs)

Anhydrobiosis related gene islands (ARIds) are defined as genomic regions hosting compact clusters of genes which are anhydrobiosis-related and accumulate transcripts upon desiccation (VanBuren *et al.*, 2015). In analogy with ARIds, we define clusters of desiccation-associated genes (CoDAGs) as genomic regions hosting compact clusters of genes associated with desiccation and differentially expressed upon desiccation. These regions are of interest owing to their potential role in the evolution of DT. We performed a genome-wide search for CoDAGs in *X. viscosa* considering the same criteria as for ARIds (VanBuren *et al.*, 2015): (a) they host a paralogous set of genes (containing at least a pair of genes); (b) their localization in the genome is not necessarily related to that of the potential ancestor of the expanded set of genes; and (c) all genes located within CoDAGs are differentially expressed in response to desiccation. CoDAGs are abundant in the genome of *X. viscosa* (277 CoDAGs and 600 genes involved) and are mainly composed of genes with declining transcript abundance upon dehydration (Figure 2, Supplementary table 3). The expression pattern of the genes in the same CoDAG is not necessarily the same. Genes involved in various environmental interactions form CoDAGs, such as the ABA receptor PYL (Pyrabacin Resistance-Like), ABC transporters, ethylene-responsive transcription factors, intracellular ribonucleases, LEA proteins, pathogenesis-related proteins, polyphenol oxidases and uncharacterized proteins (Supplementary table 4). That the majority of CoDAGs consists of transcripts that decline upon dehydration appears contradictory, but may be related to mechanisms that suppress growth and development or energy metabolism in vegetative tissues during dehydration, and may trigger or inhibit the induction of the mechanisms that lead to vegetative DT. A less parsimonious possibility is that the genes in these CoDAGs may supress the expression of DT in vegetative tissues, allowing the induction of DT only at critical levels of water loss, and that these genes have to be suppressed in order to acquire DT.

Late Embryogenesis Abundant (LEA) Proteins

LEA proteins have been implicated in plant embryo development and maturation, response to high salinity, freezing and DT (Alpert, 2006; Tunnacliffe and Wise, 2007). They accumulate in seeds, pollen, fungal spores, yeast cells, nematodes, rotifers and embryos of some crustaceans(Alpert, 2006; Tunnacliffe and Wise, 2007). In resurrection plants, desiccation-induced expression of LEA transcripts is a common response as

shown in *H. rhodopensis* and *C. plantagineum* (Dinakar and Bartels, 2013; Gechev *et al.*, 2013a; Rodriguez *et al.*, 2010).

HMM (Hidden Markov Model) profiles provided by PFAM were used to search for LEAs in the genome of *X. viscosa*. We found 126 putative LEA motif-containing proteins divided into eight families (Figure 3a), of which 90 were differentially expressed during drying and rehydration of adult plants (Figure 3b).

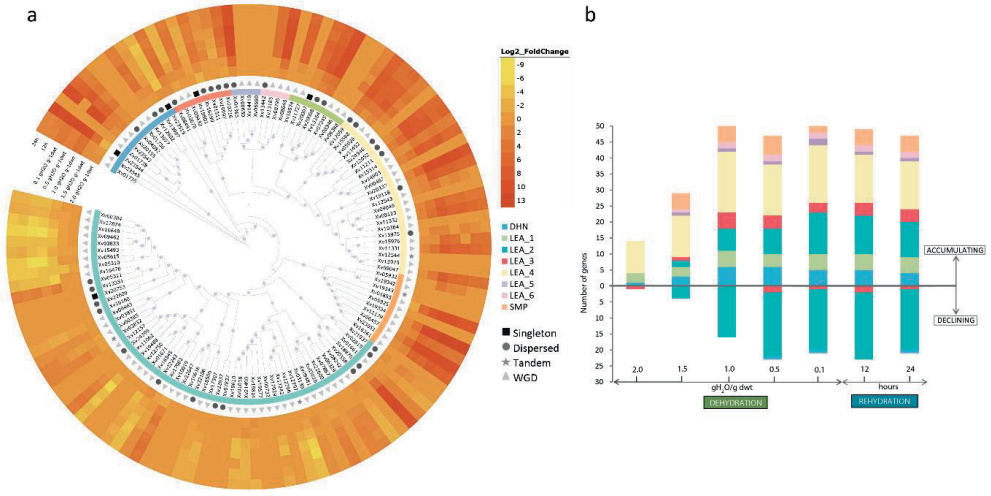


Figure 3 - LEAs transcript expression and accumulation patterns during dehydration and rehydration. a - The maximum likelihood tree shows the phylogenetic relationship between the 126 LEA proteins along with the gene duplication mode, and a heat map showing the $\log_2$ fold change of the transcripts in each point relative to the hydrated state (2.5 $gH_2O\ g^{-1}$ dwt). The dots on the branches represent bootstrap support values. b - Number of LEA transcript that accumulate and decline relative to the hydrated state. The colours on the chart represent each LEA family shown in the legend.

Most of the duplicated gene pairs have a dispersed origin, which are paralogs that do not show conserved synteny or are not near each other on chromosomes (Wang *et al.*, 2012) (Supplementary table 5). By performing a BLASTP search to assess the similarities between each of the LEA proteins (Supplementary table 5), we found that, at the protein level, there are no repeated sequences in the genome. Thus, each *LEA* gene that we identified is unique, and the high number of LEAs found is due to small (punctual) duplications, and not due to the octoploid nature of the genome. The number of LEAs is significantly higher than identified in the genomes of *B. hygrometrica* (67 *LEA* genes) (Xiao *et al.*, 2015) and *O. thomaeum* (94 *LEA* genes; http://resources.oropetium.org/Oropetium_v01/)(VanBuren *et al.*, 2015), as well as in a phylogenetic cross-section of 25 plant species (Supplementary table 5). Whole genome

duplications (WGD) played critical roles in the expansion of LEA families in *X. viscosa* (Figure 3a). In contrast, in *O. thomaeum* and *B. hygrometrica* dispersed duplications were more evident (Supplementary figure 6). A comparative genomic analysis with 25 plant species spanning the land plant phylogeny indicated that two of the eight LEA families, namely LEA_4 and LEA_6, are significantly expanded in the genome of *X. viscosa* (Supplementary figure 6, Supplementary table 5). These differences might relate to the poikilochlorophyllous nature of *X. viscosa,* in contrast to the homoiochlorophyllous species *B. hygrometrica* and *O. thomaeum*. This, in turn, might relate to longevity in the dry state. Although longevity of *X. viscosa* has not been systematically assessed, two other poikilochlorophyllous species of *Xerophyta* have demonstrated longevity. Dried excised leaves of *Xerophyta scabrida* can survive up to two years under laboratory conditions (Tuba *et al.*, 1993). In the only study reporting on longevity of whole plants in the dry state, (Bajic, 2006) demonstrated that *Xerophyta humilis* retains viability for more than two years, whereas the homoiochlorophyllous *Craterostigma wilmsii* lost viability within 3 months. Furthermore, *X. viscosa* (unlike *X. humilis*) is a chasmophyte (Figure 1a) and thus tolerates not only frequent periods of desiccation, but also extremes of temperature ranging from sub-zero to above + 50°C (Farrant *et al.*, 2015).

The response to desiccation appears LEA-family specific. Members of dehydrin (DHN), LEA_1 and LEA_4 families accumulate transcripts during drying and rehydration. It is likely that members of these LEA families are involved in a rapid primary response to dehydration and as transcripts continue to accumulate these LEA families may guarantee homeostasis in the desiccating cells. Moreover, as LEA proteins decline shortly after full rehydration, they might serve as a readily available nitrogen source to support resumption of photosynthesis and metabolism upon rehydration. Most members of LEA_2 family and a few members of DHN and LEA_3 families significantly decrease transcript abundance during drying and rehydration suggesting that these LEAs are not involved in the protection mechanism activated by desiccation.

The LEA_2 family is the most diverse in the genome of *X. viscosa*, with 57 members. A promoter enrichment analysis showed that the LEA_2 family is activated by a more diverse range of transcription factors than other LEA families (Supplementary figure 6). The ABI5 motif is prominent in the promoters of the LEA_4 family-members. The ABI5 transcription factor is mostly commonly associated with the regulation of seed maturation, and regulates transcription of *LEA* genes and degradation of chlorophyll essential to the establishment of seed longevity (measured by a moderate artificial aging method) in *Medicago truncatula* and *Pisum sativum* (Verdier *et al.*, 2013; Zinsmeister *et al.*, 2016).

Predicted subcellular localization shows that LEAs can be found in seven subcellular compartments (Supplementary figure 6), supporting the diversity of LEA expression in adult leaves. Most LEA families are overrepresented in the nucleus. SMP family-members are overrepresented in the cytosol. LEA_2 and LEA_3 families are overrepresented in the plastids and decrease transcript abundance during drying, supporting their poikilochlorophyllous origin.

During desiccation, the most profound change in LEA expression occurs between 1.5 $gH_2O\ g^{-1}$ dwt and 1.0 $gH_2O\ g^{-1}$ dwt, with a 43% increase in LEA expression, which was maintained during further desiccation and rehydration (Figure 3). Thus, these LEAs are largely represented by the early responses of SOMs 1 and 3 (Figure 2). These observations indicate an activation of regulatory pathways essential for the desiccation response around 1.0 $gH_2O\ g^{-1}$ dwt (40% RWC). This activation is likely caused by a regulatory 'switch' between the dehydration response (to initial water loss) and cellular preparation response (protection) for the desiccated state.

The ABI3 regulon

ABI3 is a transcription factor that mediates DT in seeds through a highly conserved gene regulatory network (Costa *et al.*, 2016; Delahaie *et al.*, 2013; Khandelwal *et al.*, 2010; Monke *et al.*, 2012). *X. viscosa* has two structural orthologs of ABI3. Due to different gene contexts (different neighbouring genes), these paralogs are not considered alleles of the same gene. They are expressed, but their expression did not change significantly in response to exogenous ABA or drying and rehydration of adult plants. We identified in *X. viscosa* 139 structural orthologs of the 98 target genes of the ABI3 regulon from *Arabidopsis thaliana* (Monke *et al.*, 2012). The majority displayed an increase in transcript abundance in response to exogenous ABA (62% in shoots of seedlings) and to drying (45% in SOM 1). Fifty of these structural orthologs were located in our co-expression network. Although the nodes representing these genes are not directly connected, they share first neighbours, forming tightly connected sub-networks (Supplementary figure 5). The GO categories enriched in these genes are primarily related to energy metabolism (*cellular carbohydrate metabolic process, generation of precursor metabolites and energy*) and photosynthesis (*plastid organization, regulation of photosynthesis, stomatal complex morphogenesis* and *cofactor metabolic process*). Structural orthologs of seed-specific members of the ABI3 regulon were also found in our analysis, such as embryonic proteins DC-8, oleosins, seed maturation proteins and 1-Cys peroxiredoxin. Overall, these analyses shed light

on the conserved mechanisms of the gene regulatory network orchestrated by ABI3 in seeds and vegetative tissues of resurrection plants.

Autophagy and Senescence

Autophagy is a catabolic cellular process that attempts to restore homeostasis during severe stress (Williams *et al.*, 2015). It involves the recycling of nutrients and removal of damaged and potentially harmful cellular material (Williams *et al.*, 2015). In resurrection plants, effective regulation of autophagy aids survival under extreme stress conditions (Williams *et al.*, 2015). Although drought-induced leaf senescence is thought to be an efficient strategy for reducing transpiration and allowing remobilization of water and nutrients, it does not occur in resurrection plants in the tissues undergoing induction of DT(Li *et al.*, 2012b; Williams *et al.*, 2015). It is intriguing to investigate how resurrection plants regulate autophagy and avoid senescence during dehydration. On the basis of sequence similarity to proteins predicted to be involved in autophagy and senescence in the resurrection grass *Tripogon loliiformis*, we found a number of similar proteins in *X. viscosa* (Supplementary table 6). *X. viscosa* homologs of *A. thaliana AUTOPHAGY* (e.g. *ATG2, ATG8, ATG9, ATG12* and *ATG18*), known to delay senescence in *A. thaliana* (Li *et al.*, 2012b), cluster in SOMs 1 and 2, whereas *X. viscosa* homologs of *A.* thaliana SENESCENCE-ASSOCIATED GENE and SENESCENCE-ASSOCIATED RECEPTOR-LIKE KINASE, known to promote senescence in *A. thaliana* (Li *et al.*, 2012b), cluster in SOM4. Overall, both in *X. viscosa* and in *T. lolliformis* (Williams *et al.*, 2015) transcripts associated with senescence and pro-apoptosis exhibit a reduction in accumulation, while transcripts associated with the delay of senescence and anti-apoptosis accumulated during drying. Despite different strategies concerning chlorophyll retention during drying (*X. viscosa* is poikilochlorophyllous and *T. lolliiformis* is homoiochlorophyllous), processes related to autophagy and senescence appear to operate in similar ways in *X. viscosa* and *T. lolliiformis*.

Stress signalling in the endoplasmic reticulum (ER) is transduced through the unfolded protein response (UPR) pathway (Reis *et al.*, 2011). The UPR activates the ER-located molecular chaperone binding protein BiP, that assists in folding newly synthesized proteins, acts in ER stress signalling and prevents senescence (Griffiths *et al.*, 2014; Reis *et al.*, 2011). We identified ten putative BiP genes in *X. viscosa*. Most of them clustered in SOM 2, indicating activation when the stress becomes more severe. Thus, apparently, *X. viscosa* engages a mechanism to prevent activation of the UPR cell death response during severe dehydration.

Conclusions

Combining genome sequence information with genome wide gene expression data is a powerful approach to gain insight into the 'footprint' of desiccation tolerance in a resurrection species. Among the currently known resurrection plant species, *X. viscosa* is probably the most resilient (Farrant *et al.*, 2015). Vegetative desiccation tolerance was reactivated in at least 13 families of Angiosperm plant species (Gaff and Oliver, 2013) as adaptation to environmental demands, and we propose that the basal DT mechanism was further modified according to specific environmental and habitat (niche) requirements. Reactivation of vegetative DT must have been based on the presence of genes associated with DT in reproductive structures, such as seeds, and hence the genomic information for DT was redirected towards vegetative tissues. Further genome modifications to deliver DT may have included polyploidization, as in the African Velloziaceae, during the establishment of this family on the African continent. Poikilochlorophyllous resurrection species have a clear seed imprint in their phenotype. For example, the degradation of chlorophyll during the maturation stage is common among orthodox seeds, presumably to avoid the generation of reactive oxygen species during subsequent long periods of storage in the dry state. During dismantling of photosynthetic machinery, *X. viscosa* engages anti-senescence mechanisms, giving evidence that in this species vegetative DT is uncoupled from common drought-induced senescence; again similar to seeds. In conjunction, the bZIP transcription factor ABI5 is strongly associated with the acquisition of longevity in orthodox seeds, assessed by artificial aging, as well as expression of several LEAs (Verdier *et al.*, 2013). Here we show that ABI5 may be a regulator of expression of the LEA_4 family which, consequently, may be an important factor in the longevity of *X. viscosa* in the dry state. Furthermore, we identified two structural orthologs of ABI3, a major regulator of seed maturation and DT along with the majority of the ABI3 regulon expressed in leaves. Again, this exemplifies the strong 'seed character' of vegetative DT in this species.

The CoDAGs that are downregulated upon dehydration may represent genes which are related to mechanisms that suppress growth and development or energy metabolism in vegetative tissues and may trigger or inhibit the induction of the mechanisms that lead to vegetative DT. The genes in these CoDAGs may also supress the expression of DT in vegetative tissues, allowing the induction of DT only at critical levels of water loss. Their down regulation may thus be part of the reactivated mechanism that evolved to deliver DT in vegetative tissues. Further study of possibly higher order regulation of this set of genes may identify targets for modification of crops towards desiccation tolerance.

Methods

DNA sequencing, de novo assembly, and validation

The genome of *Xerophyta viscosa* was sequenced using raw paired-end Illumina (~77× coverage) and PacBio (~58× coverage) technologies. Reads originating for contaminants (insects, bacterial, fungal, and human), as well as chloroplast and mitochondrial genomes, were removed from all sequence data prior to assembly. Illumina reads were error corrected using Lighter (Song *et al.*, 2014) and assembled using SparseAssembler (Ye *et al.*, 2012). PacBio reads were assembled using the diploid aware FALCON assembler (Chin *et al.*, 2016). A hybrid assembly was produced with DBG2OLC (Ye *et al.*, 2016) and the contigs were reordered and connected into scaffolds using SSPACE-LongRead (Boetzer and Pirovano, 2014). The assembly was polished using Sparc (Ye *et al.*, 2012) and Pilon (Walker *et al.*, 2014). PBJelly2 (English *et al.*, 2012) was used for gap closure and genome improvement. Alignments due to gene duplication and repeats were filtered out using the delta-filter utility of the MUMmer package (English *et al.*, 2012). The assembly was validated by mapping the available RNA and DNA libraries to the genome with Bowtie2 (Langmead and Salzberg, 2012) and Blasr. Assembly statistics were calculated using QUAST (Gurevich *et al.*, 2013). Gene space completeness was measured by using BUSCO (Benchmarking Universal Single-Copy Orthologs). The genome assembly and validation pipeline is summarized in Supplementary Data Figure.

Annotation

The *ab initio* predictors AUGUSTUS (Hoff *et al.*, 2016) and SNAP (Korf, 2004) were trained on the transcriptome data using BRAKER1 (Hoff *et al.*, 2016). The MAKER2 annotation pipeline was applied for gene prediction and repeat annotation. Predicted genes were functionally annotated by a consensus approach using InterProScan(Mitchell *et al.*, 2015), Gene Ontology (Harris *et al.*, 2004), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot (Boeckmann *et al.*, 2003), Translated EMBL Nucleotide Sequence Data Library (TrEMBL,(Boeckmann *et al.*, 2003), and BLAST2GO (Conesa *et al.*, 2005).

Repeatmodeler (Smit *et al.*, 2008)was applied to build a *de novo* repeat library and identify ribosomal and small nuclear RNAs. Repetitive sequences in the assembly were soft masked using RepeatMasker (Smit *et al.*, 2008). Transfer RNAs were annotated using tRNA-scan-SE (Lowe and Eddy, 1997). miRNAs were predicted using

BLAST and INFERNAL (Nawrocki and Eddy, 2013)against the RFAM database (Nawrocki *et al.*, 2015).

SNPs and INDELs were called relative to the genome using Freebayes (Garrison and Marth, 2012). Detected polymorphisms from short-read alignments were discarded when the quality was below 20.

Orthologous group inference

Orthology between 15 plant species (*Amborella tricopoda*, *Arabidopsis thaliana, Elaeis guineensis, Eragrostis tef, Hordeum vulgare, Musa acuminata, Oropetium thomaeum, Oryza sativa, Phalaenopsis equestris, Phoenix dactylifera, Physcomitrella patens*, *Spirodela polyrhiza, X. viscosa, Zea mays* and *Zostera marina*) was determined with OrthoFinder (Emms and Kelly, 2015). The proteomes were downloaded from Phytozome V11.0 (https://phytozome.jgi.doe.gov/pz/portal.html).

Gene family expansions and contractions in comparison with other sequenced genomes were calculated for all OGs (excluding singletons). The z-score was calculated for each OG, those with z-score ≥ 2 represent significantly expanded gene families and those with z-score ≤ -2 represent significantly contracted gene families. The number of genes per species for each OG was transformed into a matrix of z-scores. The z-score profile was hierarchically clustered using Pearson correlation as a distance measure. The functional annotation of each OG was predicted based on sequence similarity to entries in the Swiss-Prot database (Boeckmann *et al.*, 2003) and InterPro protein families database (Mitchell *et al.*, 2015) where more than 50% of proteins in the family share the same protein annotation.

The ancestral gene content at key nodes in the phylogeny of the 15 plant species was reconstructed by Wagner parsimony with a 1.2 gain penalty using COUNT (Csűös, 2010).

Cytogenetics

Chromosome preparations were performed as described by Szinay et al. (Szinay *et al.*, 2008). The chromosome preparations were screened under a phase-contrast microscope and late-pachytene cells with little or no cytoplasm, good chromosome spreading and well-differentiated chromatin morphology were selected (Supplementary figure 2).

The assessment of DNA content of *X. viscosa* was performed by Iribov (Enkhuizen, The Netherlands). Genome size was measured by flow cytometry on a CA-II cell analyzer (Partec). Tetraploid *Brassica oleraceae* was used as a reference for which the DNA content was calculated as 2.8 pg.

Desiccation tolerance in X. viscosa seedlings

Seeds were harvested from *X. viscosa* plants collected in the Buffelskloof Nature Reserve and cultivated under glasshouse conditions at the University of Cape Town (South Africa). Seeds were sown on two layers of blue filter paper (Blue Blotter Paper, Anchor Paper Company) and 50 ml of distilled water and incubated in germination cabinets with constant white light at 25˚C for ~10 days.

To assess desiccation tolerance in *X. viscosa* seedlings, we used seedlings with the first leaf between 1 and 2 mm length. Three replicates of 20 seedlings were incubated for 3 days in 6-cm Petri dishes containing 1.3 ml of solutions of distilled water (control) or 50 µM ABA on two sheets of white filter paper (grade 3hw, Biolab Products, Sartorius Stedim Biotec) in the dark at 20˚C. After incubation, seedlings were rinsed in distilled water, transferred to new Petri dishes with one sheet of white filter paper, dried and rehydrated. Drying was achieved by drying under an atmosphere of 32% relative humidity (RH) for 3 days at 20˚C, resulting in final water contents (WC) as low as 0.126 $gH_2O\ g^{-1}$ dwt. Water contents were assessed gravimetrically by determination of the fresh weight and subsequently dry weight after 17 h at 105˚C. After dehydration, seedlings were pre-humidified in humid air (100% RH) for 24 h at 22˚C and rehydrated in H_2O at 22˚C on a Copenhagen Table under a 12/12 h dark/light regime. Desiccation tolerance was quantified as percentage of seedlings that exhibited growth resumption with both green leaves and development of a root system (Supplementary figure 3).

For RNA extraction, seedlings treated with ABA and non-treated were dissected in three parts: seed coat, shoots and roots. The seed coat was discarded and shoots and roots were used for RNA extraction. Total RNA was extracted following a modified hot borate protocol (Wan and Wilkins, 1994).

The samples were sequenced (150nt, single-end) with Illumina HighSeq 2000 and 55.2 Gb were obtained after trimming (12 libraries). A *de novo* transcriptome assembly was constructed using Trinity (Grabherr *et al.*, 2011). The raw RNA-seq data have been deposited in NCBI SRA database under accession SRS1071017.

Desiccation tolerance in adult plants

A population of 80 clonally propagated five-month old *X. viscosa* plants were grown in Premier Pro-Mix BX (Premier Tech Horticulture Pennsylvania, USA) under the following greenhouse conditions: 16-hour days at 27°C during the day and 18°C at night. Prior to the initiation of the drying period well-watered plants were covered with a plastic bag overnight to maximize the hydrated water contents prior to sampling. Each of the eighty plants were treated as individual biological reps for sampling. Dehydration was achieved by withholding water from the pots. Young green leaf tissue was harvested

from 16 individual plants which were sampled daily to monitor their total water content. A portion of the tissue from each of the 16 plants was flash-frozen in liquid nitrogen and stored at -80°C. Total water content (TWC), expressed as grams of water per gram of dry weight, was determined gravimetrically using the remainder of the sample and calculated as TWC = (FWT-DWT)/DWT. Dry weight (DWT) of each sample was determined gravimetrically after exposure to 70°C in a convection oven until constant weight, which took four hours (Oliver *et al.*, 2011). Plants were rehydrated by fully saturating the soil and aerially by misting with water.

Triplicate samples were selected with TWCs of 2.5 (fully hydrated TWC), 2, 1.5, 1, 0.5, and 0.1 $gH_2O\ g^{-1}$ dwt and rehydrated tissue at 12 hours and 24 hours post-re-watering were selected for RNA extraction. RNA was extracted using the RNeasy (Qiagen, Hilden, Germany) kit with the RLC buffer following the manufacturer's recommended protocol. The RNA isolates were treated with DNase1 and cleaned using the DNA-free RNA Kit (Zymo Technologies, Irvine, CA). RNA quality was assessed by use of a fragment analyser (Advanced Analytical Technologies, Ankeny, IA) and concentration determined by use of a Nanodrop Spectrophotometer (ThermoFisher, Waltham, Massachussetts).

RNA libraries were created and individually bar-coded from 2.7µg of template total RNA utilizing the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) as described in the manufacturer's recommended protocol. Libraries were pooled in groups of six and sequenced (six samples per lane) on an Illumina HiSeq 2500 ultra high-throughput DNA sequencing platform (Illumina, San Diego, CA) at the DNA Core facility at the University of Missouri, Columbia, MO. USA (http://dnacore.missouri.edu//HiSeq.html).

Gene expression analyses

Gene expression was calculated using TopHat and Cufflinks from the Tuxedo suite (Trapnell *et al.*, 2012). Differential expression was then computed using Cuffdiff, also from the Tuxedo suite (Trapnell *et al.*, 2012). The expression of 12 genes was analysed by qPCR in shoots in order to verify the accuracy of the data. Candidate reference genes were chosen based on the RNA-Seq data applying a cut-off of ≥ 0.9 for *q*-value. These genes were further checked using qBase+ (Biogazelle) applying a cut-off of ≤ 0.5 for M-value (gene stability value) and ≤ 0.15 for CV-value (coefficient of variation), generating a shortlist of seven reference genes. The three reference genes with most stable expression in the qPCR were used for expression normalization of the target genes. Both RNA-Seq and qPCR showed comparable trends.

The annotated genome was used for an over-representation analysis (ORA) to recover over-represented biological processes (using Benjamini & Hochberg False

Discovery Rate (FDR) correction, *P*-value ≤ 0.05) based on gene ontologies using the plugin BiNGO for Cytoscape. InterProScan and protein BLAST (using *A. thaliana*, *O. sativa*, and *Z. mays* as a reference) were used to assign GO-terms to the total set of annotated genes. The whole annotation was used as reference set. The term's semantic distance with respect to other semantically close terms (''Dispensability'') was calculated using the online tool ReviGO (http://revigo.irb.hr/) and used to remove redundant terms applying a cut-off of ≤ 0.05.

Self-organizing maps (SOMs) were calculated using GeneMaths software (version 2.1, Applied Maths BVBA, Sint-Martens-Latem, Belgium) by importing gene expression data and mapping them into four groups (2 x 2 node format) that provided optimal representation of gene expression patterns in a small number of independent bins.

Orthologs were defined as hits with lowest Expect value (E-value), with a threshold of $\leq 10^{-10}$. Multiple hits were considered orthologs when the difference between their E-values and the lowest hit's E-value was smaller than 10^{-10}.

Identification of Late Embryogenesis Abundant (LEA) proteins

To identify members of LEA protein families, all Hidden Markov Models (HMM) from the PFAM database (Finn *et al.*, 2016) (http://pfam.xfam.org) were uploaded and used to generate HMM profile matrix using the program 'hmmbuild' of the HMMER3.0 package (Finn *et al.*, 2011). The HMM profiles were used to identify members of the eight LEA families (Dehydrin - PF00257, LEA_1 - PF03760, LEA_2 - PF03168, LEA_3 - PF03242, LEA_4 - PF02987, LEA_5 - PF00477, LEA_6 - PF10714, and SMP - PF04927) in the genome of *X. viscosa* using the program 'hmmscan'. All proteins with significant hits (E-value < 0.01) were collected.

Expansion of X. viscosa LEA families

We collected protein sequences from of 24 species (*Amborella trichopoda*, *Arabidopsis thaliana*, *Boea hygrometrica, Brachypodium distachyon*, *Brachypodium stacei*, *Chlamydomonas reinhardtii*, *Elaeis guineensis, Eragrostis tef*, *Hordeum vulgare, Musa acuminata*, *Oropetium thomaeum, Oryza sativa*, *Panicum hallii*, *Phalaenopsis equestris, Phoenix dactylifera, Physcomitrella patens*, *Selaginella moellendorffii*, *Setaria italica*, *Setaria viridis*, *Sorghum bicolor, Sphagnum fallax*, *Spirodela polyrhiza*, *Zea mays* and *Zostera marina*). Protein sequences were retrieved in the same way as described above for *X. viscosa*, using HMM from PFAM and searching with HMMER3.0 program.

LEA family expansions were calculated as described above. The families that presented z-score ≥ 2 were considered significantly expanded.

Phylogenetic and gene duplication analysis

Multiple sequence alignment of the 126 putative LEAs was generated using MAFFT version 7 (http://mafft.cbrc.jp/alignment/server/) with default parameters. The test for the best amino acid substitution model and maximum likelihood (ML) tree were performed by IQ-Tree with default parameters, and the model WAG+I+G4 was selected according to BIC and 1,000 ultra-rapid bootstraps. The phylogenetic tree was edited and displayed using the online application iTOL v3.0 (Interactive Tree of Life, http://itol.embl.de/).

For the comparative analysis with 24 species, multiple sequence alignments of the significantly expanded LEA_4 and LEA_6 families were performed using MAFFT with default parameters and trimmed with the online tool TrimAl v1.3 (http://trimal.cgenomics.org/) implemented in Phylemon v2.0 with settings: conserving at least 50% of the positions and gap threshold 0.7. ML analysis and amino acid substitution model tests were performed with IQ-Tree with default parameters and 1,000 ultra-rapid bootstraps. The model JTT+F+G4 was selected for LEA_4 family, and JTTDCMut+G4 for LEA_6 according to Bayesian Information Criterion (BIC). iTOL v3.0 was used for editing and displaying the phylogenetic trees. We used the 'duplicate_gene_classifier' command in MCScanX package to determine the origins of duplicate genes for the LEA families in *X. viscosa* genome.

Promoter motif analysis

The MEME program available within the web tool MEME-Lab (http://meme-suite.org/tools/meme) was used for discovery of motifs at 1kb region upstream each LEA family. The parameters used were: occurrence per sequence 1, motif count 10, motif width between 6 wise and 15 wide (inclusive). Subsequently, the motifs found were compared with annotated motifs in the Jaspar Core 2016 Plants database by Tomtom, which is part of the MEME suite platform, with default parameters.

Protein structure and subcellular localization

We used PSI program (Plant Subcellular-localization Integrative predictor, http://bis.zju.edu.cn/psi/) to predict the subcellular localization of LEAs of *X. viscosa* proteins. PSI was chosen because it integrates the main predictors such as Cello, mPloc, Predotar, mitoProt, MultiLoc, TargetP, Wolf PSORT, subcellPredict, iPsort, Yloc, PTS1. The grand average of hydropathy (GRAVY) value for protein sequences calculation was performed with GRAVY calculator (http://www.gravy-calculator.de/).

Gene co-expression network

Genes differentially expressed in at least one comparison between a time-point in the dehydration/rehydration curve and the hydrated state were used. Pearson correlation coefficients between all pairs of differentially expressed genes were calculated. A table with correlation coefficient values was exported to Cytoscape v.2.8.2 and correlation coefficients above a threshold of 0.96 (determined according to Freeman *et al.* (2007)) were used to filter the connections and determine the edges between nodes in the network. The resulting network was displayed with a yGraph Organic layout. The Cytoscape built-in app NetworkAnalyzer was used to compute network parameters.

**Correspondence and requests for materials should be addressed to H.W.M.H. (henk.hilhorst@wur.nl).

Acknowledgments

We thank Elizabeth Parker (owner) and John Burrows (manager) of Buffelskloof Nature Reserve Mphumulanga for allowing collection of Xerophyta viscosa plants. We thank all members of the Wageningen Seed Lab for discussions. MCDC received financial support from CNPq – National Council for Scientific and Technological Development (201007/2011-8). MSA received financial support from CAPES – Brazilian Federal Agency for Support and Evaluation of Graduate Education (BEX 0428/09-04, BEX 0857/14-9). JMF acknowledges use of funding supplied by the South African Research Chairs Initiative of the DST and NRF of SA (Grant No 98406).

Author contributions

M.C.D.C. and M.A.S.A. wrote the article; M.C.D.C., H.N., E.J. and M.F.L.D. performed bioinformatics; J.M. and W.L. contributed to genome and transcriptome analysis; J.M.J.-G. and M.O. performed and analysed transcriptomics; B.W. and S.G.M. provided the autophagy/anti-senescence dataset and performed blasting; T.H. and E.G.W.M.S. prepared libraries and performed PacBio sequencing and initial genome analysis; J.M.F. and H.W.M.H. initiated and coordinated the work and directed preparation of the article.

Competing Financial Interests statement

The authors declare no competing financial interests.

Supplementary Information

The supplementary material of this chapter can be downloaded from the online version of the manuscript at https://www.nature.com/articles/nplants201738.

Data availability

Sequence and transcriptome data are available from the National Center for Biotechnology Information (NCBI) under BioSample ID PRJNA291133 and BioProject ID SAMN03940242.

Chapter

3 Dissecting the genomic diversification of *LATE EMBRYOGENESIS ABUNDANT* proteins in plants

Mariana A. Silva Artur[1*], Tao Zhao[2*], Wilco Ligterink[1], M. Eric Schranz[2] & Henk W. M. Hilhorst[1]

Accepted in Genome Biology & Evolution

[1]Laboratory of Plant Physiology, Wageningen University, Wageningen The Netherlands.
[2]Biosystematics Group, Wageningen University, Wageningen, The Netherlands.
**These authors contributed equally to this work.*

Abstract

Late Embryogenesis Abundant proteins (LEAs) include eight multi-gene families that are expressed in response to water loss during seed maturation and in vegetative tissues of desiccation tolerant species. To elucidate LEA evolution and diversification, we performed a comprehensive synteny and phylogenetic analyses of the eight gene families across 60 complete plant genomes. Our integrated comparative genomic approach revealed that synteny conservation and diversification contributed to LEA family expansion and functional diversification in plants. We provide examples that: 1) the genomic diversification of the Dehydrin family contributed to differential evolution of amino acid sequences, protein biochemical properties, and gene expression patterns, and led to the appearance of a novel functional motif in angiosperms; 2) ancient genomic diversification contributed to the evolution of distinct intrinsically disordered regions of LEA_1 proteins; 3) recurrent tandem-duplications contributed to the large expansion of LEA_2; and, 4) dynamic synteny diversification played a role In the evolution of LEA_4 and its function in plant desiccation tolerance. Taken together, these results show that multiple evolutionary mechanisms have not only led to genomic diversification, but also to structural and functional plasticity among LEA proteins which have jointly contributed to the adaptation of plants to water-limiting environments.

Key words: Abiotic stress adaptation, desiccation tolerance, gene family evolution, intrinsic disorder, LEA proteins.

Introduction

When plants colonized land 450Ma ago, they developed a wide range of adaptations including physiological, structural and regulatory mechanisms to cope with variable environments. Land plants (embryophytes) evolved from streptophyte algae, a paraphyletic group of green algae believed to be physiologically pre-adapted to terrestrial environments due to their fresh water origin (Becker and Marin, 2009; Kenrick and Crane, 1997; Wodniok *et al.*, 2011).

As they colonized land, plants also developed desiccation tolerance (DT). DT is the ability to survive the removal of almost all cellular water without irreparable damage and is recurrent in reproductive structures of most vascular plants (e.g. during embryogenesis), in the vegetative body of non-vascular plants and in a few angiosperms species commonly known as 'resurrection plants' (Farrant and Moore, 2011; Gaff and Oliver, 2013; Illing *et al.*, 2005; Leprince and Buitink, 2010; Oliver *et al.*, 2000). Several genes that are thought to be important for DT are common amongst non-vascular and vascular plants, and are also present in their ancestral streptophyte algae (Rensing *et al.*, 2008; Wodniok *et al.*, 2011).

Within the conserved mechanisms of cellular protection involved in DT, a common group named Late Embryogenesis Abundant (LEA) proteins, has received considerable attention. LEAs were originally associated with the acquisition of DT in plant embryos due to the high gene expression and protein accumulation in the later stages of seed maturation (Delahaie *et al.*, 2013; Dure *et al.*, 1989b; Espelund *et al.*, 1992; Galau *et al.*, 1986; Manfre *et al.*, 2006). In vegetative tissues, LEAs were found to accumulate under abiotic stresses such as drought, salinity, heat and freezing, and during desiccation in resurrection plants (Amara, 2014; Cuming *et al.*, 2007; Hoekstra *et al.*, 2001; Stevenson *et al.*, 2016). Interestingly, *LEA* genes are also found outside the plant kingdom, suggesting a common mechanism of DT across distinct life forms (Browne *et al.*, 2002; Gusev *et al.*, 2014; Kikawada *et al.*, 2006; Tunnacliffe *et al.*, 2005).

LEA proteins exhibit peculiar biochemical properties such as a high proportion of polar amino acids, high hydrophilicity, and the presence of intrinsically disordered regions (IDRs) (Battaglia *et al.*, 2008; Dure *et al.*, 1989b; Garay-Arroyo *et al.*, 2000; Goyal *et al.*, 2005c). Intrinsically disordered proteins (IDPs) have been proposed as critical for plant adaptation in new environments because of their ability to perform more than one function, the so called 'moonlighting' activity (Covarrubias *et al.*, 2017). This property allows LEAs to perform anti-aggregation, protein stabilization, as well as molecular chaperone-like activities (Battaglia *et al.*, 2008; Chakrabortee *et al.*, 2007; Chakrabortee *et al.*, 2012; Cuevas-Velazquez *et al.*, 2017; Hincha and Thalhammer, 2012; Kovacs *et al.*, 2008).

Several studies have attempted to identify, classify, and assess LEAs function in plants (Amara, 2014; Battaglia *et al.*, 2008; Hundertmark and Hincha, 2008; Shih *et al.*, 2008), however, a comprehensive understanding of the evolutionary history and its relationship with the high diversification of protein sequence and function of LEAs in plants is still elusive.

With the increasing number of plant genomes available, a comprehensive analysis of the evolution and functional diversification of *LEA* gene families is now possible. The reconstruction of the evolutionary history of a protein family in an entire lineage involves homology identification by comparative genome analysis among different taxa, and provide a deeper understanding of the evolution of genomic complexity and lineage-specific adaptations (Koonin, 2005). Phylogenomic analysis (i.e. phylogenetic analysis at the genome scale) has often been employed in order to identify cross-species homologs and predict gene function by reconstructing the evolutionary history (Eisen, 1998).

In this study, we performed a large-scale phylogenomic analysis across 60 complete genomes, combining synteny network and phylogenetic analysis, in order to identify LEAs and investigate their origin and evolution in plants. Our synteny analysis reveals independent evolutionary patterns that shaped synteny diversification of LEAs in plants, and illustrates resultant functional novelties related to water-stress adaptation. Our work provides compelling opportunities for further functional classification and discovery of new LEA functions in plants.

Results

Distinct origins of LEA families in plants

We performed a genome-wide sequence homology search to identify the complete repertoires of LEAs across 60 genomes of diverse plant species (Figure 1). For that we used the most widely employed classification of LEAs that defines eight multi-gene protein families (PFAM): Dehydrin (DHN), LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6 and Seed Maturation Protein (SMP) (Hundertmark and Hincha, 2008).

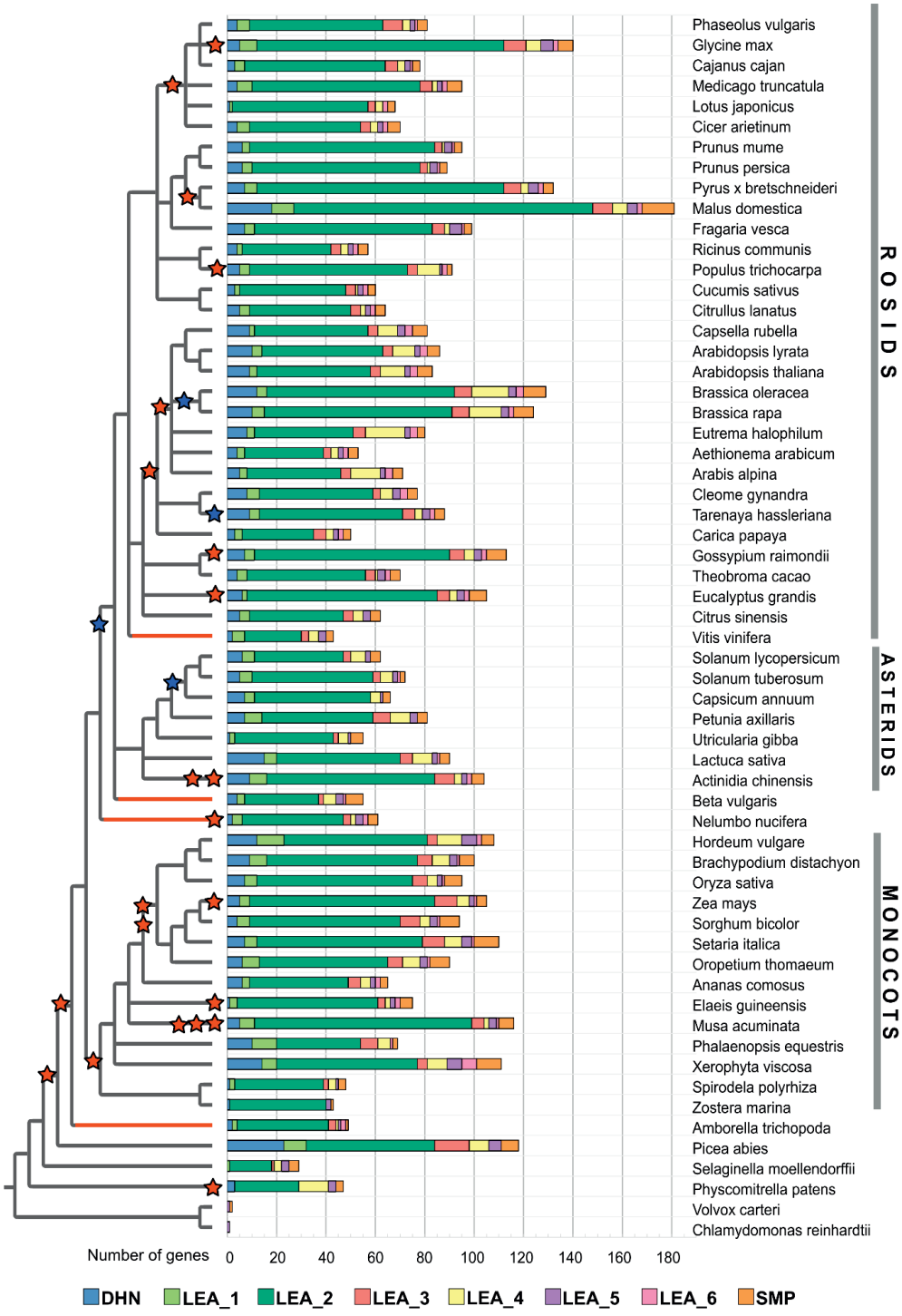


Figure 1- Species phylogeny and number of *LEA* genes identified in plant genomes. The species tree was inferred using NCBI Taxonomy Browser

(https://www.ncbi.nlm.nih.gov/taxonomy). Each LEA family is represented by a specific colour (See also Supplementary table 1). The red branches in the phylogenetic tree indicate the basal rosid *Vitis vinifera*, the basal eudicots *Beta vulgaris* and *Nelumbo nucifera*, and the basal angiosperm *Amborella trichopoda*. The red and blue stars on the phylogenetic tree indicate whole-genome duplication (WGD), and whole-genome triplication (WGT), respectively.

Based on the conservation of Hidden Markov Model (HMM) profiles for the eight LEA protein families we identified a total of 4,836 genes, with variable copy number distribution among the LEA families and the genomes investigated (Figure 1, Supplementary table 1). Only single genes belonging to SMP and LEA_5 were found in algal genomes, suggesting an ancestral origin of these families. The Dehydrin, LEA_2 and LEA_4 families were identified in the bryophyte clade (*Physcomitrella patens*) and LEA_1 and LEA_3 families appeared in the lycophyte lineage (*Selaginella moelendorffii*). The LEA_6 family only emerged in early angiosperms (*Amborella trichopoda*), likely representing the most recent LEA family in plants. Overall, the LEA_2 family was the most abundant with 3,126 genes, which are multi-copy in genomes of both angiosperms and lower plants. LEA_6, on the other hand, represents the smallest family with a total of 89 identified genes, with copy-number varying from 0 to 3, with the exception of the resurrection plant *Xerophyta viscosa* in which six LEA_6 genes were identified. The variable copy-number between different taxa suggests independent loss or duplication of genes in individual genomes. The under-representation of *LEA* genes in *Zostera marina* and *Spirodela polyrhiza* (Olsen *et al.*, 2016), and the over-representation in *X. viscosa* (Costa *et al.*, 2017a) have already been reported and correlated with the respective desiccation-sensitivity and -tolerance of these species, suggesting that the evolution of LEAs contributed to water stress adaptation in plants.

Differential conservation of LEAs in angiosperms

We used a synteny-based method to identify homology between the proteins and to explore the evolutionary history of LEAs in plants. Homologous genes comprise orthologs and paralogs, which are corresponding genes in different species that evolved from the same ancestral gene, and to genes duplicated within the same genome, respectively (Gabaldon and Koonin, 2013; Koonin, 2005). Generally, orthologs have equivalent functions in different taxa, while paralogs are prone to perform biologically distinct functions (Gabaldon and Koonin, 2013; Koonin, 2005). Synteny homologs (syntelogs) have similar genomic context and likely evolved from a common ancestor gene (Zhao *et al.*, 2017; Zhao and Schranz, 2017). Syntelogs were inferred with the Synteny Network (Synets) method (Zhao *et al.*, 2017; Zhao and Schranz, 2017) which enables detection of homologs in corresponding chromosomes in different species, as well as paralogs within a species. The output is a network in which the nodes

represent anchor genes in a syntenic block and the edges indicate synteny similarity (Supplementary figure 1). Synteny communities can be detected in synteny networks using community detection methods (Zhao *et al.*, 2017). Table 1 summarizes the percentage of syntelogs identified per LEA family as well as the number of synteny communities detected in each network (detailed information in Supplementary table 2).

Table 1 - Summary of syntenic genes and synteny communities identified per LEA protein family.

PFAM	Total genes	Syntelogs (%)	Synteny communities
DHN	365	62.2	12
LEA_1	251	63.3	10
LEA_2	3126	76.0	130
LEA_3	274	79.9	16
LEA_4	298	77.2	18
LEA_5	153	67.3	4
LEA_6	89	76.4	8
SMP	280	59.3	11
Total	4836		209

The variable percentage of syntenic genes and number of synteny communities suggest independent evolution between and within the LEA protein families. Genes not incorporated in synteny communities by our clustering method are likely to be species-specific singletons. No syntelogs were identified between angiosperm and non-angiosperm species, only a few 'in-paralogs' (paralogs from the same species) were detected in the basal species *Sellaginela moellendorffii* and *Physcomitrella patens* (Supplementary table 2). Considering the large evolutionary distance between the species analysed, we hypothesize that ancient and independent synteny diversification between LEA families have functional implications.

Phylogenetic profiling reveals angiosperm-wide and lineage-specific LEAs

We further analysed the origin of the synteny communities detected with Synets in order to obtain information on the evolutionary conservation and diversification of LEAs in angiosperms. Presence or absence of a species syntelog in a community of the synteny network can be visualized as a phylogenetic profile, enabling inference of the origin, expansions, and contractions of the gene family in each clade of the phylogenetic tree (Figure 2a).

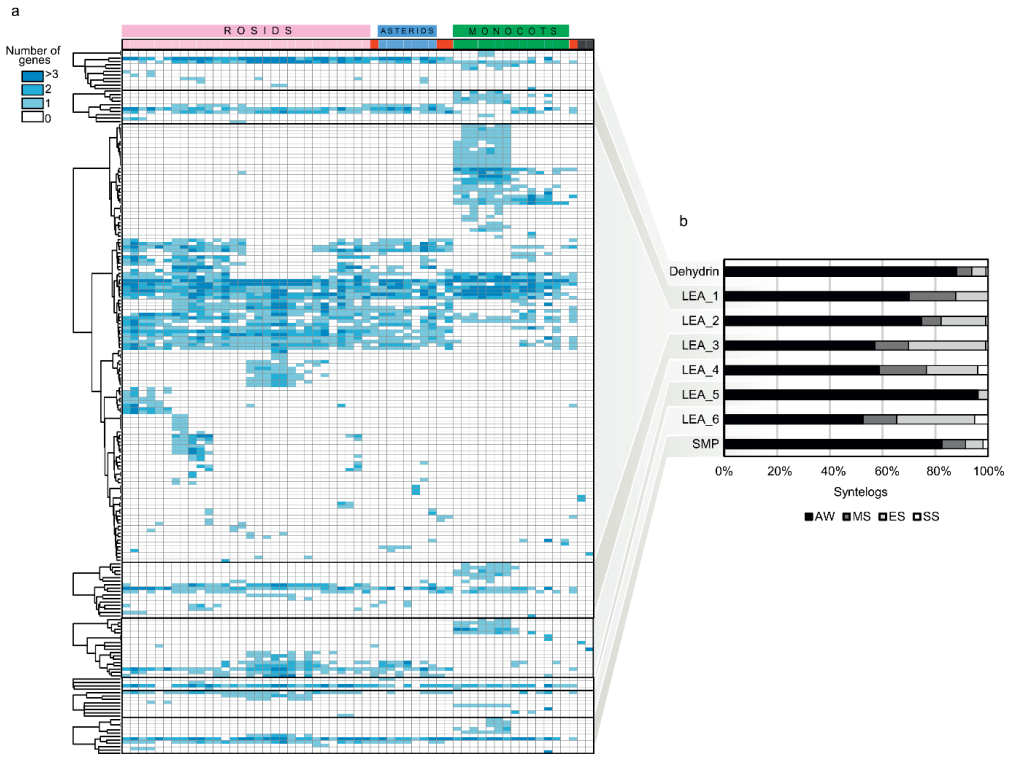


Figure 2 - Phylogenetic profile and evolutionary categorization of syntenic LEAs in the genomes analysed. a - Phylogenetic profile showing the number and distribution of syntenic *LEA* genes in plants. Rows represent synteny communities and columns indicate species. The colours on top of the profile indicate rosids (pink), asterids (blue), monocots (green), basal angiosperm species (red) and *Physcomitrella patens* and *Selaginella moelendorffii* (dark grey). The species were ordered from the most recent to the most ancient, from the left to the right. b - Distribution of syntenic genes in each evolutionary category. AW – angiosperm-wide, MS – monocot-specific, ES – eudicot-specific, SS – species-specific (see also Supplementary table 3).

We subdivided the synteny communities into four main evolutionary categories: angiosperm-wide (AW), monocot-specific (MS), eudicot-specific (ES), and species-specific (SS) (Figure 2b, Supplementary table 3). Angiosperm-wide are synteny communities that contain genes of at least one monocot and one eudicot species. Monocot-specific includes synteny communities containing only monocot genes, and eudicot-specific includes communities comprising eudicot genes only. Species-specific correspond to paralogs duplicated in an individual genome, also named ohnologs.

AW communities were found in all LEA families and encompasses the largest number of the syntelogs identified (Figure 2b), indicating that the majority of *LEA* genes have a common origin in angiosperms and are likely located in a more ancestral

genomic context. The angiosperm-wide conservation of LEAs is particularly observed in the families DHN, LEA_5 and SMP, where more than 80% of the syntelogs identified are shared amongst angiosperm species. Lineage-specific duplications (MS and ES) have also significantly contributed to the repertoire of LEAs in plants, especially in LEA_3 and LEA_6 families, where more than 40% of the syntelogs are distributed over these two categories. SS paralogs were overall underrepresented or absent in the genomes investigated, likely due to low frequency of local gene duplications, or the duplicated copies were more likely to be lost in individual genomes. The finding of lineage-specific and species-specific synteny suggests that duplication events other than whole genome duplications (WGD) have significantly contributed to the expansion of LEA families in plant genomes.

The fact that LEA_5 has the smallest number of synteny communities and that the majority of the genes belong to AW conserved genomic context indicate that this is the most conserved LEA family in plants. On the other hand, the large number of LEA_2 syntelogs in AW communities indicates that this is the most diverse LEA family in the plant lineage.

Structural and functional diversification contributed to LEA evolution

Duplication events may introduce a gene copy into a new regulatory context, leading to differential evolutionary and regulatory constraints, which is one of the main sources driving functional innovation within a gene family (Conant and Wolfe, 2008; Flagel and Wendel, 2009). Therefore, in the next sections we provide a few examples from our synteny analysis of remarkable structural and functional innovations within LEA families resulting from differential evolution of the genomic context.

Dehydrin: Biochemical, structural, and expression pattern innovations during angiosperm evolution

Dehydrin (DHN) is classified as a LEA family due to the gene expression during late seed embryogenesis and ability to perform 'classical' chaperone activity, preventing heat-induced protein aggregation and inactivation *in vitro* (Kovacs *et al.*, 2008; Liu *et al.*, 2017a). In our dataset, we found that DHN genes are distributed across two main angiosperm-wide synteny communities and a maximum likelihood tree supports the phylogenetic separation of these communities in angiosperms (Figure 3a).

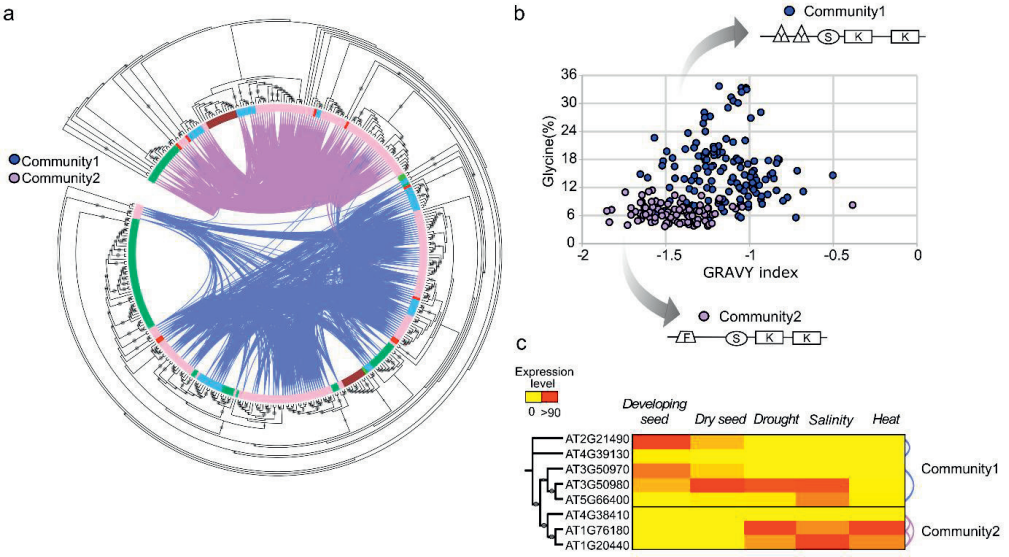


Figure 3 - Characteristics of Dehydrin synteny communities. a - Maximum likelihood tree of all *DHN* genes found in the genomes of 60 species. The inner circle indicates species belonging to monocots (green), rosids (pink), asterids (blue), basal species (red), the gymnosperm *Picea abies* (brown), and the bryophyte *Physcomitrella patens* (light green). The connections between the branches indicate synteny between the gene pairs. Synteny communities 1 and 2 are indicated (blue and pink connections, respectively), dots on the branches represent bootstrap support values (>85). The larger the dots the higher the bootstrap values. b - Glycine (Gly) content and GRAVY index plot (Gly/GRAVY plot) showing the distribution of hydrophylins between community 1 and 2. The arrows indicate a schematic representation of the consensus sequence of proteins of community 1 and 2, respectively. The F-, Y-, S-, and K- protein segments are indicated according to their position in the protein sequences. c - Absolute expression values of *DHN* genes in *Arabidopsis thaliana*. The expression data was retrieved from the Bio-Array Resource for Arabidopsis Functional Genomics (http://bar.utoronto.ca/) and from Hundertmark and Hincha (2008). The dots on the branches of the phylogenetic tree indicate bootstrap support values (>75). Connections between the rows represent synteny relationships.

Some of the DHNs are called 'hydrophylins' because of their specific response to osmotic stress (Garay-Arroyo *et al.*, 2000; Jaspard and Hunault, 2014). Hydrophylins play important roles in protecting cell components from the adverse effects caused by low water availability due to their biochemical properties such as high Glycine (Gly) content (> 6%) and low grand average hydropathy (GRAVY) (< -1) (Battaglia *et al.*, 2008; Garay-Arroyo *et al.*, 2000; Reyes *et al.*, 2008). In order to investigate the distribution of hydrophylins in angiosperms, we analysed the Gly content and GRAVY index of each protein within the two largest angiosperm-wide DHN communities (Figure 3b). Although both communities contain proteins with hydrophylin properties, community 1 contains proteins with more variable Gly/GRAVY composition than community 2

proteins which have a more homogeneous Gly/GRAVY distribution. These findings indicate that, even though hydrophylin-type proteins do not form an isolated synteny community, there is a clear biochemical divergence between proteins that evolved in distinct genomic contexts.

DHN proteins have been functionally subdivided into four to five main architectures based on the presence and organization of specific motifs called Y-, S- or K- segments (Banerjee and Roychoudhury, 2016; Close, 1996; Hunault and Jaspard, 2010; Malik *et al.*, 2017). We performed multiple sequence alignments of proteins from the DHN synteny communities 1 and 2 in order to investigate the diversification of the different functional motifs (Supplementary figure 2a-b). Our data indicate that the majority of proteins of community 1 comprises Y(n)SK(n) types (Figure 3b, Supplementary figure 2a), while community 2 contains mainly SK(n)-type proteins, lacking the Y-segment at the N-terminus (Supplementary figure 2b). While lacking the Y-segment, proteins from community 2 possess a new conserved segment at the N-terminus (DRGLFDFLGKK). This motif is named F-segment, and it was recently characterized as an overlooked motif in angiosperms and gymnosperms, with potential functional roles in membrane and protein binding (Strimbeck, 2017). Interestingly, genes encoding proteins belonging to community 1 are expressed mainly during seed development, and some of the genes can be induced by abiotic stress (Figure 3c). On the other hand, genes encoding the F-type DHN proteins of community 2 seem to be specifically induced by abiotic stresses such as drought, heat and salinity. The combined results indicate that the ancient synteny diversification of DHN in angiosperms has resulted in biochemical and sequence innovations, as well as changes in expression patterns that may be related to functional specificity within this protein family. To date, this is the first documentation of the evolution and diversification of the F-segment in angiosperms, and its association with abiotic stress.

LEA_1: Ancient diversification of Intrinsically Disordered Proteins (IDPs) in angiosperms

LEA_1 proteins, also known as Group 4, accumulate in the plant cell in response to water stress and have been proposed as model to study intrinsically disordered proteins (IDPs) in plants (Cuevas-Velazquez *et al.*, 2017; Olvera-Carrillo *et al.*, 2010). This family has been subdivided into two main subclasses based on protein sequence features (Battaglia *et al.*, 2008). One of the subgroups, named group 4A, comprises smaller proteins (80-124 residues) and the second group, 4B, has longer representatives (108-180 residues). Both subclasses possess a conserved portion at the N-terminal region, and a disordered C-terminal region predicted to form alpha-helices under water limiting conditions (Cuevas-Velazquez *et al.*, 2017).

Our data indicates that LEA_1 members are distributed in 10 synteny communities, and 70% of the homologs identified with Synets belong to two angiosperm-wide (AW) communities (Figure 2b, 4a).

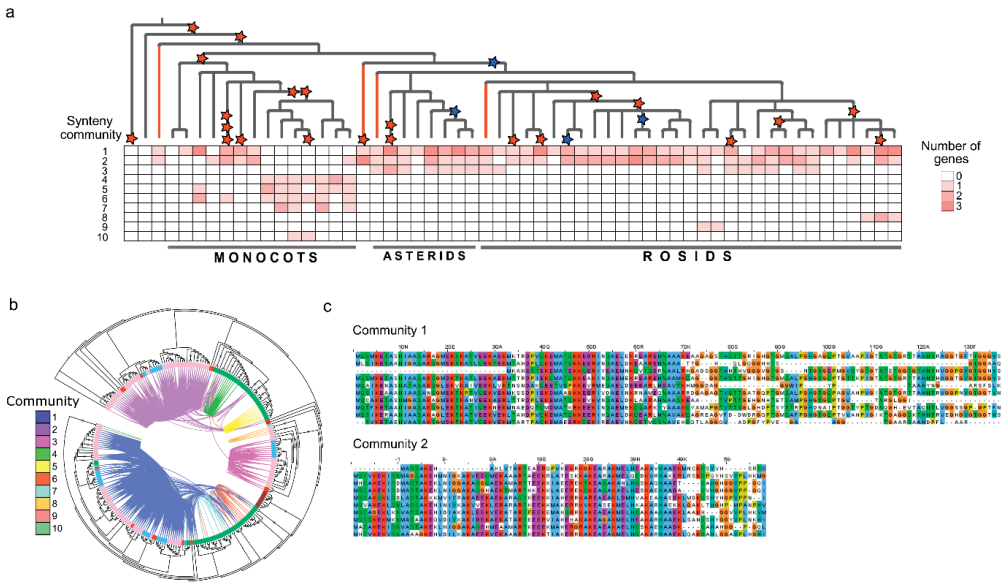


Figure 4 - Phylogenetic and synteny characteristics of LEA_1. a - Phylogenetic profile of LEA_1 indicating the distribution of the synteny communities detected in the species phylogenetic tree. The red and blue stars indicate whole-genome duplication (WGD) and whole-genome triplication (WGT), respectively. b - Maximum likelihood tree of the LEA_1 family. The circle inside the tree indicates species belonging to monocots (green), rosids (pink), asterids (blue), basal species (red), the gymnosperm *Picea abies* (brown), the bryophyte *Physcomitrella patens* (light green), and the lycophyte *Selaginella moellendorffii* (olive green). The connections between the branches indicate synteny between the gene pairs, and dots on the branches represent bootstrap support values (>85).The larger the dots the higher the bootstrap values. c - Partial representation of the multiple sequence alignments of amino acid sequences of the communities 1 and 2 (top 10 sequences).

The absence of clear synteny and phylogenetic separation in the phylogenetic tree suggests that some of the ES and MS communities have originated through duplication or transposition of genes from AW communities (Figure 4b). We found differences between the consensus sizes of the multiple sequence alignments of protein from the two AW communities (Figure 4c, Supplementary figures 3a-b), what indicates that AW community 1 represents the subclass 4B of longer protein sequences, whereas community 2 contains members of subclass 4A of smaller proteins. This suggests that the diversification of intrinsically disordered regions (IDRs) in LEA_1

occurred before the origin of monocots and eudicots, and that these protein types have been conserved in angiosperm genomes during evolution.

LEA_2: Expansion and diversification through recurrent tandem duplications

LEA_2 is the largest LEA family, and has been considered atypical because it contains proteins with more hydrophobic amino acids and more defined secondary structure in solution compared with the other LEA families (Hundertmark and Hincha, 2008; Singh *et al.*, 2005). Members of this family have been associated with the hypersensitive response (HR) after microbial and parasitic nematode infection, which also differs from the other families (Ciccarelli and Bork, 2005; Escobar *et al.*, 1999; VanderEycken *et al.*, 1996). However, functions associated with salinity, freezing, heat, UV radiation, osmotic, and oxidative stress *in vitro* have also been documented for LEA_2 proteins (He *et al.*, 2012; Jia *et al.*, 2014; Jiang *et al.*, 2017).

Despite the large number of members, in general, synteny and phylogeny of the LEA_2 are in agreement, with highly supported branches in the phylogenetic tree connecting genes that belong to the same synteny community (Figure 5a).

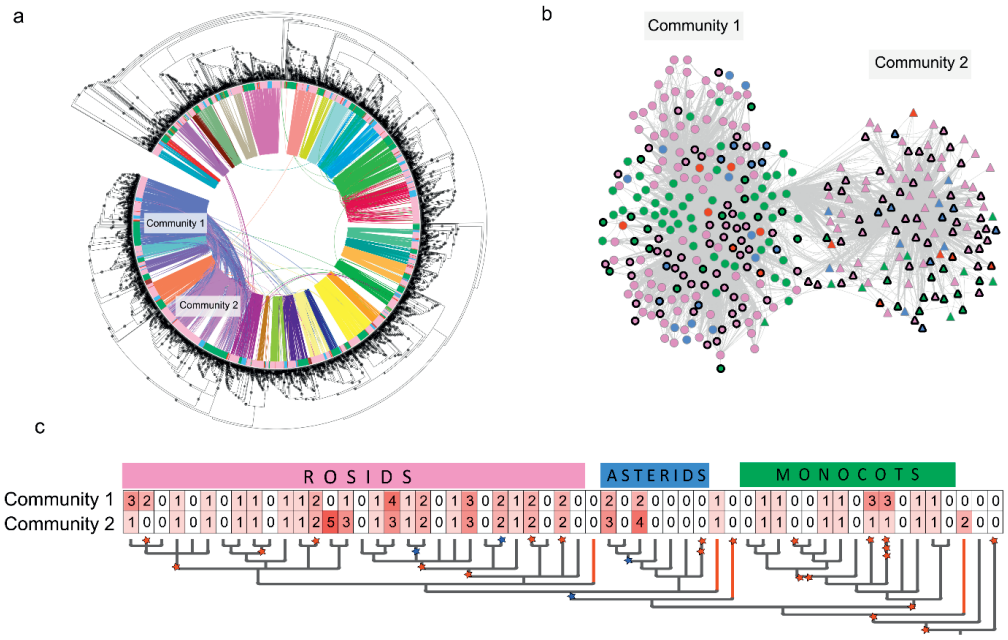


Figure 5 - Tandem duplications of the LEA_2 family. a - Maximum likelihood tree containing all *LEA_2* genes identified. The colours displayed in the inner circle indicate genes belonging to monocots (green), rosids (pink), asterids (blue), basal species (red), the gymnosperm *Picea abies* (brown), the bryophyte *Physcomitrella patens* (light green), and the lycophyte *Selaginella moellendorffii* (olive green). The connections between the branches indicate synteny between the gene pairs, and all the communities with at least 100 syntenic genes are displayed in

different colours. Synteny communities 1 and 2 are indicated. The dots on the branches indicate bootstrap support values (>85). The larger the dots the higher the bootstrap values. b - Synteny network of genes belonging to community 1 (circles) and community 2 (triangles). The colours displayed in the nodes represent the clades as indicated in (a). Tandem genes are indicated by a thicker black border. c - Summary of the number of tandem duplicates in the synteny communities 1 and 2 (see also Supplementary table 4). The tree is a simplified version of the species tree presented in figure 1. Red stairs indicate WGD and blue stars indicate WGT.

Interestingly, there is an evident interconnection between two of the largest LEA_2 synteny communities (Figure 5a-b). We found that these communities contain several tandem duplicates widespread in monocots and eudicots (Figure 5b-c). In fact, we also found several other tandem duplicates across other LEA_2 communities containing monocots and eudicots genes (Supplementary table 4). These results indicate that tandem duplications have significantly contributed to the expansion and diversification of the large LEA_2 family in angiosperms, and may be one of the causes of the diversified functionality of this atypical LEA family.

LEA_4: Dynamic synteny in plant desiccation tolerance

LEA_4 genes, also known as group 3, are also found in non-plant organisms that display DT such as rotifers, arthropods, nematodes, and tardigrades (Browne *et al.*, 2002; Gusev *et al.*, 2014; Kikawada *et al.*, 2006; Tunnacliffe *et al.*, 2005) suggesting an association with the evolution of DT. In plants, LEA_4 is strongly associated with DT in basal and angiosperm resurrection species via an ancient conserved ABA signalling pathway (Cuming *et al.*, 2007; Delahaie *et al.*, 2013; Hundertmark and Hincha, 2008; Shinde *et al.*, 2012; Stevenson *et al.*, 2016). Our species set contained two desiccation tolerant species, the bryophyte *P. patens* and the monocot *X. viscosa*; however, synteny cannot be detected between these species due to the large evolutionary distance. Nevertheless, we found that LEA_4 genes are distributed across several AW, MS, and ES synteny communities that are phylogenetically separated, suggesting a dynamic evolutionary history of this gene family in angiosperms (Figure 2b, Supplementary figure 4). Interestingly, only one of the eight LEA_4 genes identified in *X. viscosa* shares synteny with other angiosperm species, all the other duplicates are singletons or in-paralogs (Supplementary table 5). In *X. viscosa*, the LEA_4 family has expanded as compared with other monocot species (Costa *et al.*, 2017a), which was correlated with a stronger desiccation response. Altogether, it seems that LEA_4 has evolved dynamically in angiosperms, and loss of synteny may have resulted in fixation of these genes in the genome, resulting in improved contribution to DT in resurrection plants.

Discussion

How does the plant genome adapt to environmental stress? This question has been addressed frequently in recent years. It has been proposed that adaptation to novel or stressful environments is correlated with the retention of duplicated genes (Flagel and Wendel, 2009; Jiao *et al.*, 2011; Kondrashov, 2012; Panchy *et al.*, 2016). Among the many models for duplicated gene evolution (Conant and Wolfe, 2008; Innan and Kondrashov, 2010), it is suggested that genes that should be rapidly or constantly produced in response to environmental stress might be more prone to selection after duplication (Kondrashov, 2012).

In plants, the group of Late Embryogenesis Abundant (LEA) proteins, composed of eight multi-gene families (Dehydrin, LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6 and SMP), have been shown to play roles in water stress tolerance, and may represent a conserved and indispensable component of networks involved in environmental stress adaptation that allowed plants to endure the constraints associated with land adaptation (Amara, 2014; Hincha and Thalhammer, 2012; Shih *et al.*, 2008). Evidence suggests that there is functional variability between and within each of the eight families (Hundertmark and Hincha, 2008), which raises questions pertaining to the sources of functional variations, the precise biological functions of each family, if and how LEA families work as one entity, and which *LEA* genes are involved in plant development and stress tolerance.

To address these questions we interrogated 60 whole genomes, ranging from green algae to angiosperms and analysed the ancestry, conservation, and diversification of LEAs in plants. We found that LEA proteins belonging to LEA_5 and SMP families have arisen early in the plant lineage, while the other families appeared at later instants during plant evolution (Figure 1). Previous studies have already shown the presence and expression of ancient LEAs in algal genomes (Joh *et al.*, 1995; Wodniok *et al.*, 2011), corroborating the hypothesis that the ancestral fresh water lineages were pre-adapted to terrestrial environments, and the evolution of pre-existing and new gene families, including LEAs, may have facilitated the colonization of land (Becker and Marin, 2009; Rensing *et al.*, 2008). It seems possible that later LEA families expanded and diversified in embryophytes as a result of the evolution of more specialized cells, tissues and organs such as spores and seeds, that required a better control of water retention and protection against desiccation and other stresses.

Synteny homology analysis indicated that the majority of *LEA* genes are located in angiosperm-wide conserved genomic regions, while the finding of clade-specific as well as species-specific gene copies indicates that the continuing expansion and diversification of angiosperm genomes contributed to the evolution of *LEA* gene

families (Figure 2). Stress-regulated genes retained after duplication events are more likely to neofunctionalize instead of inheriting the ancestral function, which might be in part related to changes in biochemical function and in cis-regulatory regions (Arsovski *et al.*, 2015; Conant and Wolfe, 2008; Zou *et al.*, 2009). As a result of these changes, complete or partial diversification of the interaction and regulatory networks in which the duplicated genes are involved might also occur. It is likely that the genes belonging to the same synteny community (positional homologs) display similar functions, and genes in different communities are likely to display functional innovations (Dewey, 2011).

We identified highly conserved synteny between *LEA_5* genes in most genomes investigated, suggesting evolutionary constraints on maintaining the stability of their genomic context. These constraints may include the correct functioning of the maturation-induced desiccation program, where LEA_5 genes of *A. thaliana* were shown to play important roles (Manfre *et al.*, 2006), and may be conserved across all orthodox angiosperm species.

We also found several examples of correlation between synteny diversification and functional innovations. Genes from the largely studied Dehydrin (DHN) family are localized in two distinct synteny communities across the angiosperm lineage (Figure 3a). Presumably, new regulatory elements were acquired in the duplicated copies, and differential evolutionary forces may have driven protein diversification, resulting in distinct biochemical properties (Figure 3b). The consequent differential gene expression (developmental or stress induced) may have allowed the preservation of duplicated copies in the different genomes, and amplified the stress tolerance response. The finding of functionally diverse Dehydrin types in *P. patens* suggests that the colonization of land was one of the forces driving Dehydrin evolution (Agarwal *et al.*, 2017; Ruibal *et al.*, 2012). Similarly, LEA_1 have evolved into two angiosperm-wide synteny communities composed by two protein types containing distinct intrinsically disordered regions IDRs (Figure 4c). Our findings point toward an ancient functional divergence among LEA_1 members, which would explain their structural plasticity and 'moonlighting' properties associated with multiple abiotic stresses (Covarrubias *et al.*, 2017; Cuevas-Velazquez *et al.*, 2017).

Another source of evolutionary adaptations to environmental stress is gene family expansion via recurrent tandem duplications (Cannon *et al.*, 2004; Hanada *et al.*, 2008). Tandem duplications offers a pool of targets for evolutionary selection contributing to the maintenance of large gene families. These large gene families are enriched with genes important for rapid environmental adaptation such as biotic stress-responsive genes (Cannon *et al.*, 2004; Hanada *et al.*, 2008). We found several tandem duplicates in the synteny network of LEA_2 distributed across the whole

angiosperm lineage (Figure 5). This supports the atypical structured and hydrophobic nature of LEA_2 proteins and its broader spectra of gene expression in response to biotic and abiotic stresses (Ciccarelli and Bork, 2005; Hundertmark and Hincha, 2008; Singh *et al.*, 2005).

Most of the *LEA* gene expression during seed development and environmental stresses is regulated via abscisic acid (ABA)-signalling pathways (Delahaie *et al.*, 2013; Espelund *et al.*, 1992; Galau *et al.*, 1986; Shinde *et al.*, 2012; Stevenson *et al.*, 2016). The desiccation-induced *LEA* gene expression via ABA-responsive pathways is conserved across basal and angiosperm resurrection species (Cuming *et al.*, 2007; Shinde *et al.*, 2012; Stevenson *et al.*, 2016). It seems that the acquisition of new genomic contexts by desiccation-related LEAs of the resurrection monocot *X. viscosa* is an important footprint of DT, and suggests a conserved regulation of these duplicates in order to assure cellular protection under desiccation conditions.

Resurrection plants are species adapted to live in environments with low water availability, displaying specific molecular and genomic adaptations of DT (Farrant and Moore, 2011; Gaff and Oliver, 2013; Illing *et al.*, 2005; Mundree, 2002; Oliver *et al.*, 2000). The concept of DT is different from drought tolerance because drought tolerance refers to the tolerance to moderate water removal without removal of the bulk of cytoplasmic water (Shih *et al.*, 2008), while DT refers to the tolerance to a further dehydration with an increased removal of the water shell and the capacity to survive long periods in the dry state (Hoekstra *et al.*, 2001). Understanding the mechanisms underlying DT can help to improve drought tolerance in crops (Costa *et al.*, 2017a; Leprince and Buitink, 2010; Mundree, 2002). Several crops from the grass family (Poaceae) constitute major contributors to global food security that have become targets of genomic programs aiming at improved drought tolerance. In grasses, overexpression of *LEAs* has already been shown to enhance tolerance to drought and other stresses (Babu, 2004; Chen *et al.*, 2015; Fu *et al.*, 2007; Xiao *et al.*, 2007). We believe that comprehending the impact of synteny diversification in functional innovations in the LEA families may offer an extra powerful tool to select candidates for engineering drought and desiccation tolerant crops.

This data also opens several opportunities for hypothesis-driven fundamental and experimental characterization of the myriad of functions of LEA proteins, and the role of the diversification of the genomic context in plant evolution and adaptation to environmental stresses. Deciphering the evolution of eight gene families, with variable protein structure and diversified expression patterns over billions of years, is a challenging task. Despite the general association of LEAs with water stress response, our work provides strong examples of a clear evolutionary divergence resulting in differential protein evolution. The diversity of LEA families in angiosperms is a result of

extensive and dynamic synteny evolution, which indicates that the complexity of these gene families goes beyond their protein sequences.

Materials and Methods

Identification of LEA proteins in 60 genomes

We used 60 fully sequenced genomes available in Phytozome (Goodstein *et al.*, 2012) (https://phytozome.jgi.doe.gov/), and the recently published genome of *Xerophyta viscosa* (Costa *et al.*, 2017a). Our species list includes representative species belonging to green algae, mosses, lycophytes, gymnosperms, early angiosperms, monocots, early eudicots, asterids and rosids (Figure 1, Supplementary table 1).

Several classifications have been proposed for LEA proteins (for a review, see (Battaglia *et al.*, 2008). Here we used the PFAM annotation for protein families (Bateman *et al.*, 2002) based on conserved protein domains (Hundertmark and Hincha, 2008). This annotation classifies LEAs into eight PFAMs: Dehydrin (DHN) (PF00257), LEA_1 (PF03760), LEA_2 (PF03168), LEA_3 (PF03242), LEA_4 (PF02987), LEA_5 (PF00477), LEA_6 (PF10714), and Seed Maturation Protein (SMP) (PF04927). Hidden Markov Models (HMM) retrieved from the PFAM 3.0 database (http://Pfam.xfam.org) were queried against the 60 plant genomes to identify LEA proteins for each family using the program 'hmmscan' of the HMMER3.0 package (Finn *et al.*, 2011). All proteins with significant hits (e-value < 0.001) were used in this analysis.

Synteny network construction and community detection

We used the Synets method (Zhao *et al.*, 2017; Zhao and Schranz, 2017) for syntenic block calculations, network construction and community detection (https://github.com/zhaotao1987/SynNet-Pipeline). In summary, pairwise all-against-all comparisons were performed using RAPSearch (Zhao *et al.*, 2012). Synteny block detection was performed with MCScanX software (Wang *et al.*, 2012) with default parameters (minimum collinear block size = 5 genes, maximum gaps = 25 genes). The syntenic blocks containing the identified LEA sequences were used to build synteny networks (Synets) that were visualized and edited with Cytoscape 3.3.0 (Shannon *et al.*, 2003) and Gephi 0.9.1 (Bastian, 2009). Infomap (Rosvall and Bergstrom, 2008) was used to find communities within the synteny networks, which is implemented under "igraph" package in R (http://igraph.org/r/doc/cluster_infomap.html). All synteny communities were numbered according to the largest to the smallest number of genes, and later renamed per LEA family accordingly (Supplementary table 2). The synteny communities were further analysed with a phylogenetic profiling. Phylogenetic profiling allows the

visualization of the synteny communities that are lineage-specific or shared amongst different species. All synteny communities were decomposed into numbers of involved syntenic gene copies in each genome. Dissimilarity index of all clusters was calculated using the "Jaccard" method of the vegan package (Dixon, 2003), hierarchically clustered by "ward.D", and visualized by "pheatmap".

Phylogenetic analysis

Multiple sequence alignments (MSAs) were built for each of the eight LEA families using MAFFT v.7 (Katoh *et al.*, 2002). We used the automated method for the PFAM LEA_2 due to the large number of sequences, and the method G-INS-I for all other LEA PFAMs. Phyutility 2.2.6 (Smith and Dunn, 2008) was used to trim gaps and maintain 75% of the consensus protein alignment. The final MSAs were edited and displayed with Jalview 2.10.3 (Waterhouse *et al.*, 2009). IQ-TREE v.1.5.1 (Nguyen *et al.*, 2015) was used to infer Maximum Likelihood (ML) trees with 1000 bootstraps for each alignment. All phylogenetic trees were edited and displayed with the online tool iTOL (Letunic and Bork, 2016).

Physicochemical properties and expression data of Dehydrin

The hydrophilicity index of Dehydrin proteins was calculated with the online GRAVY calculator (http://www.gravy-calculator.de/). More hydrophilic proteins have a more negative GRAVY score, and more hydrophobic proteins have a more positive GRAVY score. In order to reveal hydrophylin-type proteins (GRAVY < -1 and Gly > 6%), individual GRAVY scores were plotted against the percentage of Glycine (Gly) per protein sequence (Battaglia *et al.*, 2008; Garay-Arroyo *et al.*, 2000). Absolute gene expression values were retrieved from the e-Northern tool provided by the Bio-Array Resource for Arabidopsis Functional

Genomics (http://bar.utoronto.ca/) as well as from the datasets of seed and silique development, dry seed, drought and heat shock of Hundertmark and Hincha (2008).

Acknowledgements

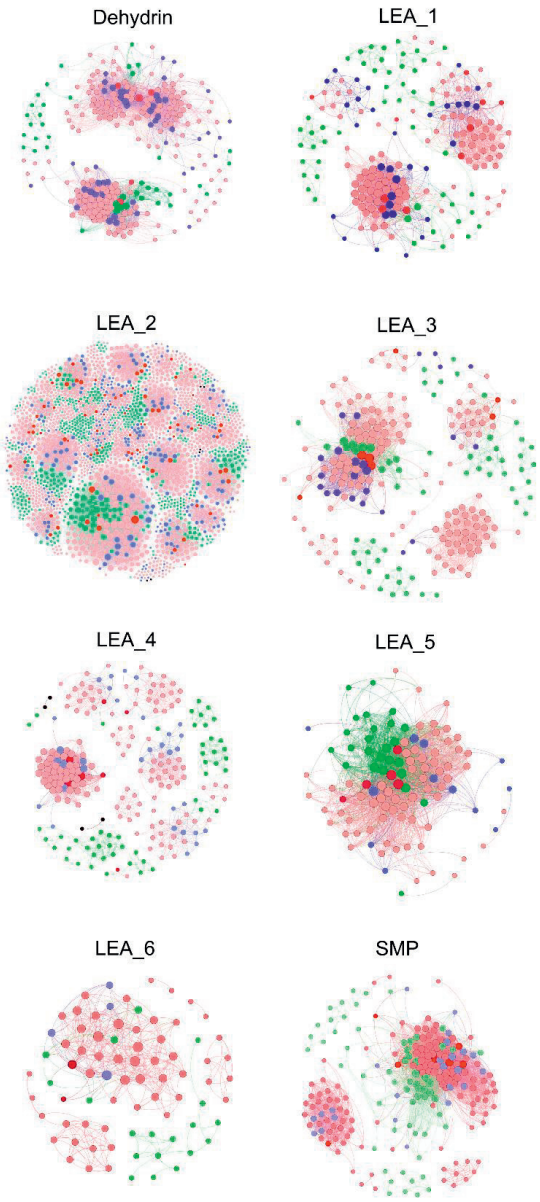
M.A.S.A. received financial support from CAPES – Brazilian Federal Agency for Support and Evaluation of Graduate Education (BEX 0857/14-9). T.Z. was supported by China Scholarship Council.

Author Contributions

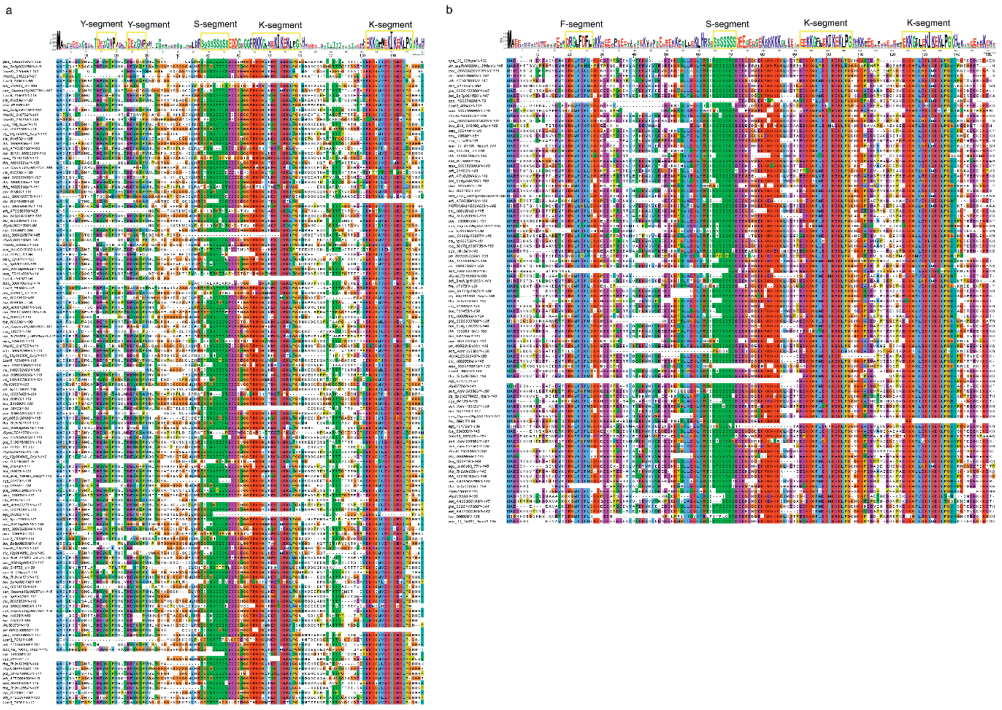
M.A.S.A, T.Z., W.L., M.E.S. and H.W.M.H. planned and designed the research. M.A.S.A and T.Z. performed the research and analysed the data. M.A.S.A interpreted the data and wrote the manuscript with contributions of T.Z., W.L., M.E.S. and H.W.M.H. All authors edited and commented on the manuscript.

Supplementary Information

Supplementary Figures

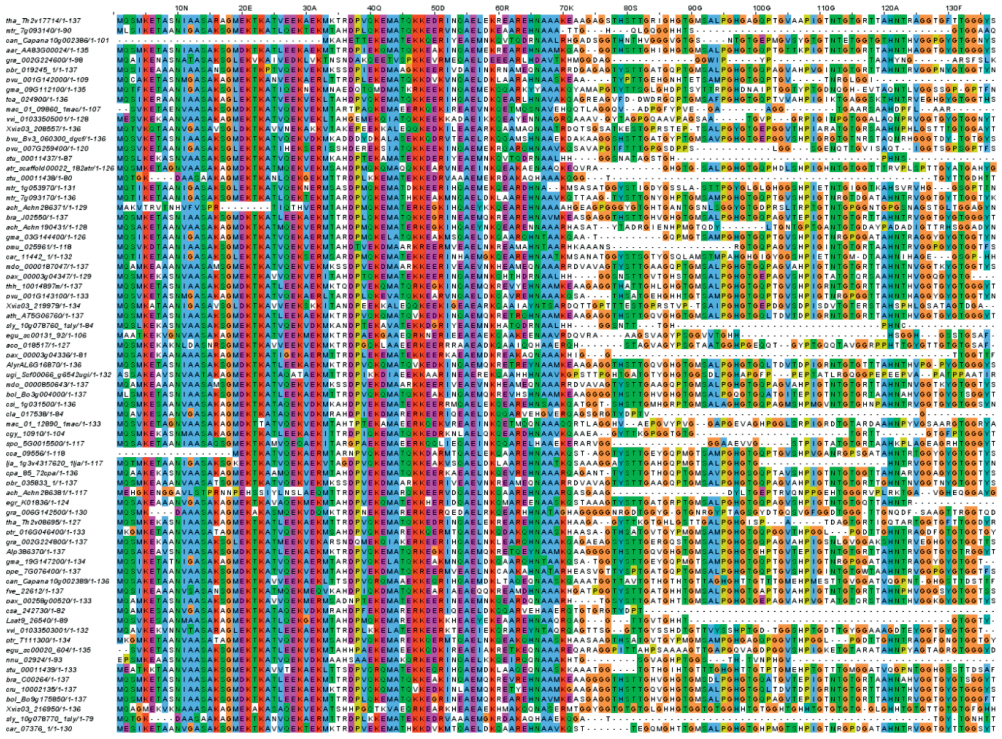


Supplementary figure 1 - Synteny networks of *LEA* genes. Synteny networks were built with the synteny information of syntelogs identified with Synets. Node colours indicate species belonging to monocots (green), rosids (pink), asterids (blue), the basal rosid *Vitis vinifera*, the basal eudicots *Beta vulgaris* and *Nelumbo nucifera*, and the basal angiosperm *Amborella trichopoda* (red). *Physcomitrella patens* and *Selaginella moelendorffii* syntelogs are indicated as black nodes. Node size indicates the number of connections: bigger nodes have more connections (stronger synteny relationships).

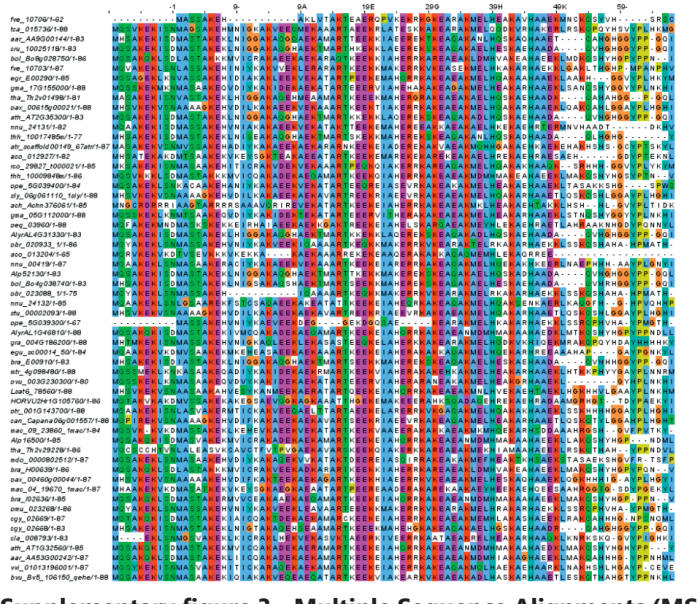


Supplementary figure 2 - Multiple Sequence Alignments (MSA) of proteins belonging to DHN. MSA of proteins belonging to DHN synteny community 1 (a) and 2 (b). The consensus sequence is shown at the top of each alignment. The F-, Y-, S-, and K- protein segments are indicated.

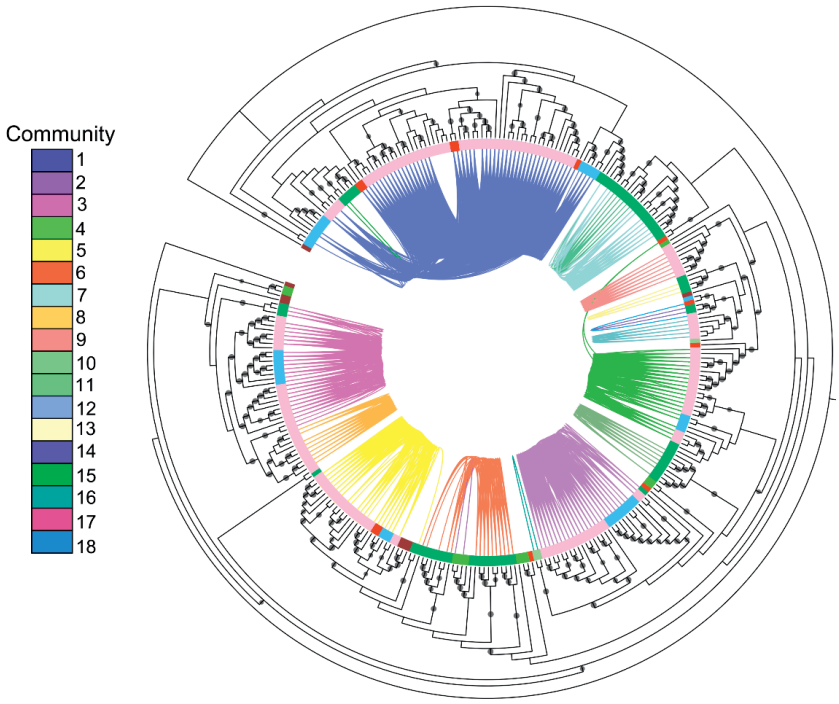
a



b



Supplementary figure 3 - Multiple Sequence Alignments (MSA) of proteins belonging to LEA_1. MSA of proteins from synteny community 1 (a) and 2 (b).



Supplementary figure 4 - Maximum likelihood trees and synteny relationships of LEA_4. The circle inside the tree represents species belonging to monocots (green), rosids (pink), asterids (blue), basal species (red), the gymnosperm *Picea abies* (brown), the bryophyte *Physcomitrella patens* (light green), and the lycophyte *Selaginella moellendorffii* (olive green). The connection between the branches indicates synteny between the gene pairs and the dots on the branches indicates bootstrap support values (>85). The larger the dot the higher the bootstrap value. The colour scale for the different communities is indicated at the left.

Supplementary Tables

The supplementary material of this chapter can be downloaded from
http://www.wageningenseedlab.nl/thesis/masilvaartur/SI/

Chapter

4 Structural and functional characterization of LEA proteins from *Xerophyta viscosa*

Mariana A. Silva Artur[1], Juriaan Rienstra[1], Tim J. Dennis[2], Jill M. Farrant[2], Wilco Ligterink[1] & Henk Hilhorst[1]

In preparation for publication

[1]Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands.
[2]Department of Molecular and Cell Biology, University of Cape Town, Cape Town, South Africa.

Abstract

Late Embryogenesis Abundant proteins (LEAs) are essential players of the desiccation tolerance (DT) response of resurrection plants due to their ability to protect the subcellular milieu against irreversible damage. In this work we characterized the structure and function of six LEAs expressed during desiccation in the monocot resurrection species *Xerophyta viscosa* (XvLEAs). *In silico* analyses suggested that XvLEAs are hydrophilic intrinsically disordered proteins (IDPs) and circular dichroism (CD) analysis indicated that the proteins are mostly unstructured in water, while acquisition of secondary structure may occur in hydrophobic environment. We show the protective properties of XvLEAs by their ability to preserve the activity of the enzyme lactate dehydrogenase (LDH) against desiccation, heat and oxidative stress, as well as growth of *Escherichia coli* upon osmotic and salt stress. Subcellular localization analysis indicated that XvLEA recombinant proteins are differentially distributed in the cytoplasm, membranes and nucleus of *Nicotiana benthamiana* leaves. Interestingly, a LEA_1 family protein (XvLEA1-8) showing the highest disorder-to-order propensity and protective ability *in vitro* and *in vivo* was also able to enhance salt and drought stress tolerance in *Arabidopsis thaliana*. Together, our results suggest that the structural plasticity of XvLEAs is essential for their protective activity to avoid damage of various subcellular components caused by water deficit stress. XvLEA1-8 constitutes a potential model protein for engineering structural stability *in vitro* and improvement of water-deficit stress tolerance in crops.

Key-words: circular dichroism, intrinsic disorder, Late Embryogenesis Abundant proteins, water-deficit stress tolerance.

Introduction

Water availability is one of the major environmental factors that affect plant growth, development and productivity. During their life cycle plants may endure periods of environmental drought and depending on the duration of such periods, it may lead to irreversible structural damages affecting plant development and survival. Most higher plants undergo programmed water loss during their life cycle, which may occur in organs such as pollen, spores and seeds (Bewley, 1995; Scott, 2000). A group of about 135 angiosperm plant species have been described as 'resurrection plants' for the ability of their leaves to tolerate the loss of nearly 80% to 95% of cellular water and resume growth and photosynthetic activity within a few hours after rehydration (Farrant *et al.*, 2015; Oliver *et al.*, 2000; Porembski, 2011; Scott, 2000). It is likely that desiccation tolerance (DT) appeared in the plant lineage during the transition from water to land, became confined to reproductive structures as plants evolved, and reappeared in vegetative structures of angiosperm resurrection plants by redirecting pre-existing genes and pathways in order to survive in the dry state (Farrant and Moore, 2011; Gaff and Oliver, 2013; Ingram and Bartels, 1996; Oliver *et al.*, 2000; Rensing *et al.*, 2008; Wodniok *et al.*, 2011).

Recently the genome of the resurrection species *Xerophyta viscosa* has been sequenced (Costa *et al.*, 2017b). *X. viscosa* is a monocot species belonging to the Velloziaceae family, and it is distributed mainly in southern African regions and inhabits rocky terrain or inselbergs in exposed grasslands (Farrant *et al.*, 2015; Mello-Silva *et al.*, 2011; Porembski and Barthlott, 2000). It is a poikilochlorophyllous plant, meaning that the photosynthetic apparatus is dismantled and chlorophyll is degraded below 55% relative water content (RWC) (Farrant *et al.*, 2015; Mundree, 2002; Sherwin and Farrant, 1998). The fact that *X. viscosa* is a resurrection monocot species and phylogenetically related to most important grass crops from the Poaceae family, opens several possibilities to study DT and apply this knowledge to improve drought tolerance in crops (Costa *et al.*, 2017b; Farrant *et al.*, 2015).

Several mechanisms of responses to desiccation in *X. viscosa* leaves have been correlated with its resurrection phenotype (Farrant *et al.*, 2015). Among the essential adaptive mechanisms to survive loss of water, the accumulation of protective molecules has been shown to be an essential component of DT at the subcellular level. These molecules include sucrose, raffinose family oligosaccharides (RFOs) and Late Embryogenesis Abundant (LEA) proteins which are thought to act, *inter alia*, as osmoprotectants in the formation and stability of the so-called cytoplasmic glassy state (Buitink and Leprince, 2004; Farrant *et al.*, 2012; Hoekstra *et al.*, 2001; Koster, 1991; Leprince *et al.*, 1993; Vicré, 2004).

LEA proteins were first discovered due to their accumulation during late stages of embryo development of cotton seeds, coinciding with their acquisition of DT (Dure *et al.*, 1989a; Dure *et al.*, 1981a; Galau *et al.*, 1986). *LEA* transcripts were later found to accumulate in leaves of resurrection plants upon drying, and their characteristic stress-induced expression has led to the hypothesis that these proteins are involved in stress-responses mediated by abscisic acid (ABA) (Galau *et al.*, 1986). LEA transcription appears to be inducible by ABA and osmotic stress and is evident upon drying below 60-40% RWC, a common pattern observed in both seeds and resurrection plants (Costa *et al.*, 2017b; Illing *et al.*, 2005; Leprince and Buitink, 2010). Interestingly, the translation of some LEAs has been shown to occur a few hours or even days after their transcription, suggesting regulation at the transcriptional and translational levels (Chatelain *et al.*, 2012; Espelund *et al.*, 1992; Galau *et al.*, 1987; Hughes and Galau, 1991; Verdier *et al.*, 2013). In addition to transcriptional and translational regulation, some LEA proteins were shown to undergo posttranslational modifications, reflecting the complexity of the regulation of these proteins at various levels (Boudet *et al.*, 2006; Riera *et al.*, 2004).

LEAs have been classified into eight protein families (PFAM), Dehydrin (DHN), LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, and seed maturation protein (SMP), based primarily on the nature of conserved motifs together with predictive modelling of domain composition (Finn *et al.*, 2011; Hundertmark and Hincha, 2008). This bioinformatics classification does not account for functional differences among the families and indeed little is known about the precise functions of most LEAs. They do have biochemical properties in common that enable them to perform anti-aggregant-, chaperone- and antioxidant activities *in vitro* (Chakrabortee *et al.*, 2007; Goyal *et al.*, 2005b; Hincha and Thalhammer, 2012).

Several LEA proteins are assumed to be intrinsically disordered proteins (IDPs) considering their ability to undergo order to disorder transitions in different *in vitro* environments (Bremer *et al.*, 2017; Cuevas-Velazquez *et al.*, 2017; Cuevas-Velazquez *et al.*, 2016; Hundertmark *et al.*, 2012; Mouillon *et al.*, 2006; Popova *et al.*, 2011; Rivera-Najera *et al.*, 2014; Shih *et al.*, 2012; Shih *et al.*, 2010; Soulages *et al.*, 2002). This interesting physicochemical characteristic allows LEAs to form homo- and heterodimers and to interact with multiple targets (Cuevas-Velazquez *et al.*, 2016; Hernandez-Sanchez *et al.*, 2017; Popova *et al.*, 2015; Tolleter *et al.*, 2007). The ubiquitous distribution of LEA-like proteins in bacteria and invertebrates suggests that similar protective mechanisms of DT involving LEAs have evolved across different life forms (Costa *et al.*, 2016; Tunnacliffe and Wise, 2007).

The myriad of secondary structures enables LEAs to play multiple roles in abiotic stress tolerance, constituting an essential footprint of desiccation tolerance in seeds

and resurrection plants (Costa *et al.*, 2017b; Maia *et al.*, 2011). Due to the fundamental role of LEAs in stress tolerance in organisms submitted to loss of water, these proteins have an enormous biochemical potential for engineering biostability in the dry state, and to be used as conserving and anti-aggregant agents of organic materials under desiccation-, heat- and oxidation stresses, as well as applications in the pharmaceutical or cosmetic industry for moisture retention due to their hydrophilic nature (Goyal *et al.*, 2005c; Li *et al.*, 2012a).

To further characterize and explore the biochemical, structural and functional properties of LEAs we cloned the coding sequences of six *LEA* genes expressed in *X. viscosa* leaves upon desiccation: *XvDHN12*, *XvLEA1-8, XvLEA4-8*, *XvLEA4-12, XvLEA6-2,* and *XvSMP4*. We examined the secondary structure of these LEA proteins by circular dichroism (CD) and monitored their folding dynamics in hydrophobic solution. *In vitro* and *in vivo* experiments demonstrated that these XvLEAs perform protective roles under desiccation-, heat-, oxidative-, salt- and osmotic stress. The protective activity of XvLEAs seems protein type-specific, with a strong correlation between the ability to form defined secondary structures *in vitro* and the extent of protection both *in vitro* and *in vivo*. XvLEAs are localized in multiple subcellular compartments such as nucleus, membrane and cytoplasm, supporting the idea of universal cell protection by LEAs. Heterologous expression of *XvLEA1-8* in *Arabidopsis thaliana* leads to higher tolerance to salt and osmotic stress in seedlings, and drought in adult plants. Our work sheds new light on the biochemical properties of these stress responsive proteins, and highlights characteristics of LEAs that can be useful for bioengineering protein stability *in vitro* and improving abiotic stress tolerance in crops.

Results

XvLEAs are predicted to be intrinsically disordered

The expression of the six *XvLEAs* used in our analysis has been reported to be increased in *X. viscosa* leaves upon desiccation between 60% RWC (1.5gH_2O g^{-1} dwt) and 40% RWC (1.0gH_2O g^{-1} dwt) (Supplementary figure 1), coinciding with the activation of the molecular signature of the resurrection physiology. In parallel, genes related to protein folding, protection and translational control were also enriched (Costa *et al.*, 2017b).

The cDNA of the six *XvLEAs* contained an open reading frame (CDS) ranging from 309 to 1149 nucleotides encoding proteins with 102 to 382 amino acids with molecular weights between 10.73 and 38.92kDa (Table 1). The theoretical isoelectric point (pI) ranged from 4.3 to 8.6, indicating that *XvLEAs* expressed upon desiccation

constitute acidic, neutral and basic proteins. All six XvLEAs had a negative GRAVY index, a common characteristic of hydrophilic LEA proteins (Battaglia *et al.*, 2008).

Table 1 – Characteristics of *X. viscosa* LEAs.

Gene ID	PFAM	Name	CDS size	Number of amino acids	Molecular weight (kDA)	Theoretical pI	GRAVY
Xvis02_06457	pf04927	XvSMP4	918	305	31.68	4.3	-0.437
Xvis02_11331	pf02987	XvLEA4-8	441	146	15.55	6.6	-1.316
Xvis02_12059	pf02987	XvLEA4-12	1149	382	38.92	5.9	-0.872
Xvis02_20008	pf03760	XvLEA1-8	321	106	10.73	8.6	-0.916
Xvis02_08790	pf10714	XvLEA6-2	309	102	11.11	5.0	-1.095
Xvis02_23545	pf00257	XvDHN12	372	123	13.44	7.2	-1.389

The percentage of polar residues was relatively high compared to non-polar residues in all XvLEAs, with the exception of XvSMP4 with 52.79% non-polar residues (Supplementary table 1). The disorder promoting amino acids Alanine (A), Lysine (K), Glycine (G), Glutamine (Q) and Glutamic acid (E) were enriched in at least four of the six sequences. These characteristics indicate that the majority of the studied XvLEAs are likely to be intrinsically disordered proteins (IDPs) (Dunker *et al.*, 2001; Tompa, 2002). In order to verify the IDP properties of the XvLEAs, we performed several *in silico* analyses. All XvLEAs were predicted to belong to the category of proteins with extended disorderedness (Supplementary figure 2a). With the exception of XvSMP4, all Xv LEAs protein sequences displayed high disorder tendency (Supplementary figure 2b). XvSMP4 seems to be the least disordered from the proteins herein investigated, because it displayed higher hydropathy and GRAVY (Table 1), as well as higher percentage of non-polar residues (Supplementary table 1).

In silico structural analysis indicated that four proteins (XvLEA1-8, XvLEA4-8, XvLEA6-2 and XvDHN12) are predicted as Janus sequences, which are proteins that undergo environment-dependent conformational transitions (Supplementary table 2). XvSMP4 is predicted to form tadpole and globular structures, in agreement with the less disordered nature of this protein. XvLEA4-12 is predicted to form coils, hairpins and chimeras, representing a structurally more heterogeneous protein type. It is important to highlight that despite *in silico* analysis that allows the prediction of structural propensities in IDPs, these proteins exist as ensembles of conformations in natural conditions, which can be a balance between more ordered and disordered structures (Das *et al.*, 2015).

Disorder binding regions (DBRs) are common amongst IDPs, and may contain short fragments of about 5 to 25 residues named Molecular Recognition Features (MoRFs) that are prone to undergo disorder-to-order transitions in the presence of

binding partners (Mohan *et al.*, 2006). We found that the number of DBRs as well as the number of MoRFs were variable among the six XvLEAs, and that the position of MoRFs not always correlated with those of DBRs (Supplementary table 2, Supplementary figure 3). XvDHN12 and XvLEA1-8 presented longer MoRFs within DBRs in the C-terminus, and XvLEA6-2 presented two longer MoRFs within DBRs located in the N- and C-terminus. Interestingly, the three proteins presenting larger number of smaller DBRs (XvLEA4-8, XvLEA4-12 and XvSMP4), also presented no or smaller MoRFs within DBRs (Supplementary figure 3). These observations indicate that there is high variability regarding the number, size and locations of DBRs and MoRFs between the different XvLEA proteins, and suggest that XvDHN12, XvLEA1-8 and XvLEA6-2 may undergo higher conformational changes in the presence of binding partners.

The binding affinity of MoRFs is believed to be modulated by phosphorylation of MoRF residues (Mohan *et al.*, 2006). In this sense, we investigated the percentage and location of phosphorylation residues (Ser, Thr and Tyr) of the six XvLEAs (Supplementary figure 3, Supplementary table 2). The percentage of phosphorylation sites was variable between the six LEAs, and XvLEA4-12 shows a strikingly higher percentage of phosphorylation sites (20.7%) followed by XvLEA4-8 (13.2%). Interestingly, in most of the proteins analysed, the predicted phosphorylation sites were in general located outside the MoRF regions (Supplementary figure 3). This suggests that phosphorylation may not be the major modulator of binding affinity in XvLEAs via MoRF recognition.

XvLEAs are heat stable and intrinsically disordered in aqueous solution

We produced and purified His-tagged recombinant XvLEAs in order to investigate their structure *in solution*. Similar to other LEA proteins (Hundertmark *et al.*, 2011; Kovacs *et al.*, 2008) all the XvLEAs showed higher molecular masses in the SDS-gel than the predicted mass (Table 1, Supplementary figure 4). This low mobility in gel might be due to the 6xHis-tag or due to the hydrophilic character of IDPs (Kovacs *et al.*, 2008; Tompa, 2002). All XvLEAs used in our analysis were confirmed to be heat-stable, a common characteristic of LEA proteins which are known to be part of the heat stable proteome associated with desiccation tolerance in seeds (Boudet *et al.*, 2006; Chatelain *et al.*, 2012; Kovacs *et al.*, 2008).

We investigated the disordered nature and folding behaviour of the six XvLEAs in aqueous solution by circular dichroism (CD) spectroscopy. We used BSA as a positive control protein with predominantly α-helical conformation (Reed et al., 1975). In general, the spectra of the six analysed XvLEAs were characterized by a negative minimum between 198nm and 201nm in water (Supplementary figure 5a-f), which is characteristic of natively unfolded proteins (Lopes *et al.*, 2014; Uversky, 2009). On the

other hand, BSA spectra were characterised by a positive maximum at ~190nm and negative minimum at 208nm and 220nm (Reed *et al.*, 1975) (Supplementary figure 5g). A residual alpha-helix content was commonly observed in all XvLEAs in aqueous solution, and the presence of turns or strands was variable between the six proteins (Supplementary figure 5h). Taken together, these results confirm that the six investigated XvLEAs are proteins with intrinsically disordered regions (IDRs).

To test the effect of changes in the pH and composition of the solvent we analysed the spectra of the six XvLEAs in aqueous solutions adjusted with HCl to pH 4.0 or pH 2.3 , and in the presence of 20mM NaCl. The decrease in the pH or addition of 20mM of NaCl did not significantly affect the spectra of BSA when compared to aqueous solution, and had little effect on the secondary structure content (Supplementary figure 5a-h). The most noticeable secondary structure variations in these conditions were observed for XvLEA1-8, with an increase in helix content at pH 2.3, and for XvLEA4-8, with an increased strand formation in NaCl solution (Supplementary figure 5h). On the other hand, a remarkable effect of high concentration of the organic solvent acetonitrile (ACN) on the secondary structure of XvLEAs was observed. ACN is an organic water miscible solvent that can increase protein chemical potential leading to conformational changes, acting as a denaturant, as well as an alpha-helix-promoting agent (Gekko *et al.*, 1998). With the exception of XvLEA4-12, a decrease of the negative minima (between 197-200nm) converting it to a positive signal and an increased negative signal between 210 and 240nm was commonly observed in the spectra of all XvLEAs in 80% ACN solution (Figure 1 a-f).

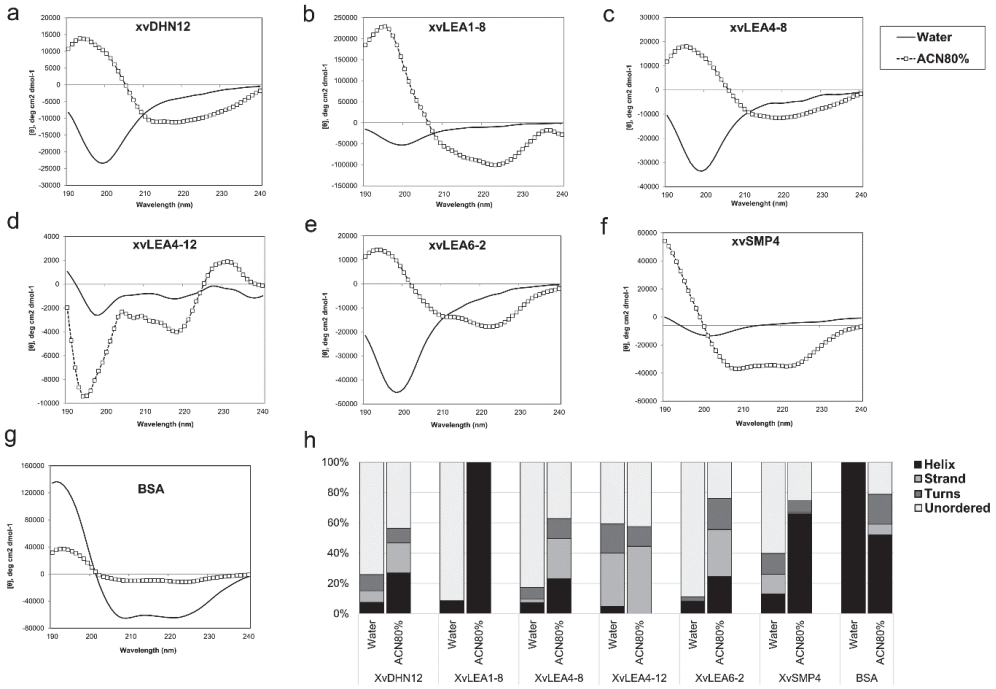


Figure 1 - Normalized CD spectra of *Xerophyta viscosa* LEA proteins and BSA. The CD spectra were obtained in water and 80% acetonitrile (ACN). All the spectra were analysed at room temperature (a-g). The graphs show the spectra obtained after subtracting the reads of a blank sample containing water only. h - Secondary structure content of XvLEAs. Predictions of the content of helix, strand, turns and unordered regions were performed with Dichroweb.

The protein XvLEA4-12 showed atypical behaviour in 80% ACN, with increased negative signal around 195nm and 200nm, and a positive signal near 230nm. The spectrum of this protein between 195nm and 220nm resembles that of a denatured protein (Greenfield, 2006; Venyaminov *et al.*, 1993). Different from what was observed for XvLEAs, BSA showed a decrease in helix content and appearance of disordered, turns and strand conformations in 80% ACN (Figure 1g-h). Interestingly, XvLEA1-8 underwent a conformational change from about 92% disordered in aqueous solution up to 100% alpha-helix in 80% ACN (Figure 1h). In summary, our results suggest that XvDHN12, XvLEA1-8, XvLEA4-8, XvLEA6-2 and XvSMP4 have conformational-inducible ability, and underscore the plasticity of XvLEA1-8 to acquire high degrees of secondary structure in a highly hydrophobic environment.

XvLEAs stabilize enzyme activity upon stress in a protein-specific manner

We investigated the *in vitro* protective functions of the different XvLEAs on the activity of the enzyme lactate dehydrogenase (LDH) during desiccation, heat and oxidative stress (Figure 2).

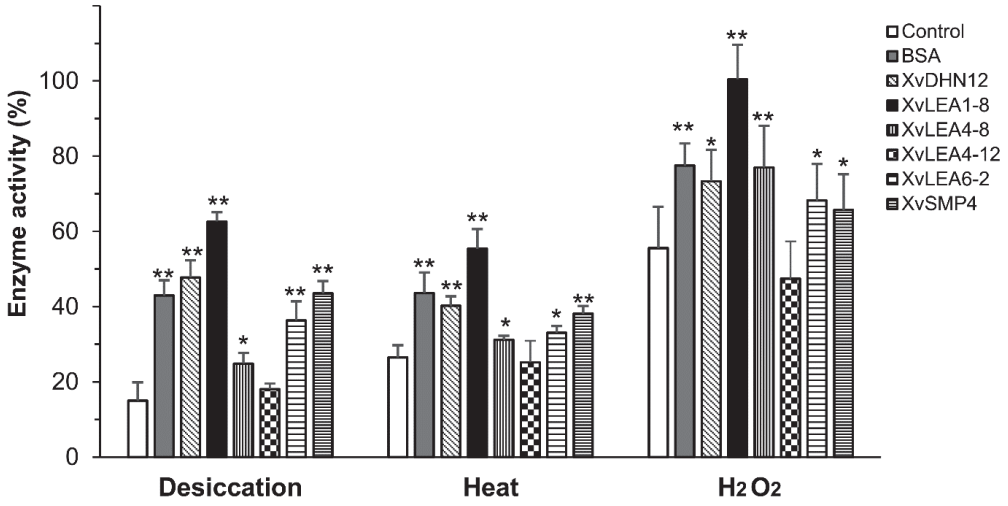


Figure 2 - Protective function of XvLEAs on LDH activity under stressful conditions. LDH was submitted alone (negative control) or in the presence of the six purified XvLEAs or BSA (positive control) at a molar ratio of 1:1 and dehydrated for 1 hour in a speed-vac, incubated at 40°C for 20 min, or incubated with 200µM H_2O_2 for 1 hour at room temperature. BSA was used as a positive control globular protein with anti-aggregant properties (Amara et al,. 2012; Finn et al., 2012). The experiments were repeated three times with three technical replicates in each experiment. Statistically significant differences as compared to control were analysed using Student's *t*-test (*p<0.05 or **p<0.01). The bars represent SD from nine replicates (n = 9).

Under desiccation we observed a reduction of LDH activity to about 15%. With the exception of XvLEA4-12, all XvLEAs showed protective functionality against desiccation on the activity of LDH, with a protection of enzymatic activity up to 63% with XvLEA1-8. We found that heat stress reduced LDH activity to about 27% in the absence of XvLEAs, and XvLEA4-12 was unable to protect LDH activity. However, all other XvLEA proteins were able to protect LDH activity against heat, and again XvLEA1-8 showed the highest protective ability. The heat-protection activity of isolated LEA proteins has already been demonstrated in *in vitro* assays before (Goyal *et al.*, 2005c; Halder *et al.*, 2017; Kovacs *et al.*, 2008) indicating that thermal anti-aggregation is a common feature of several LEA proteins.

The oxidative treatment imposed by H_2O_2 was less stressful as compared to desiccation and heat treatments. LDH activity dropped to 56% when treated with H_2O_2, and with the exception of XvLEA4-12 all proteins tested were able to protect enzyme activity. Taken together these results suggest that XvDHN2, XvLEA1-8, XvLEA4-8, XvLEA6-2 and XvSMP4 have chaperone-like activities on LDH activity upon desiccation,

heat and oxidative stress, and that XvLEA4-12 is ineffective in the conditions herein tested.

In vivo expression of XvLEAs enhances E. coli viability under salt, osmotic and heat stress

To investigate effects of *XvLEA* expression on *E. coli* survival, we analysed the colony growth in plates with high concentrations of NaCl and mannitol, or when subjected to a heat shock (Figure 3a).

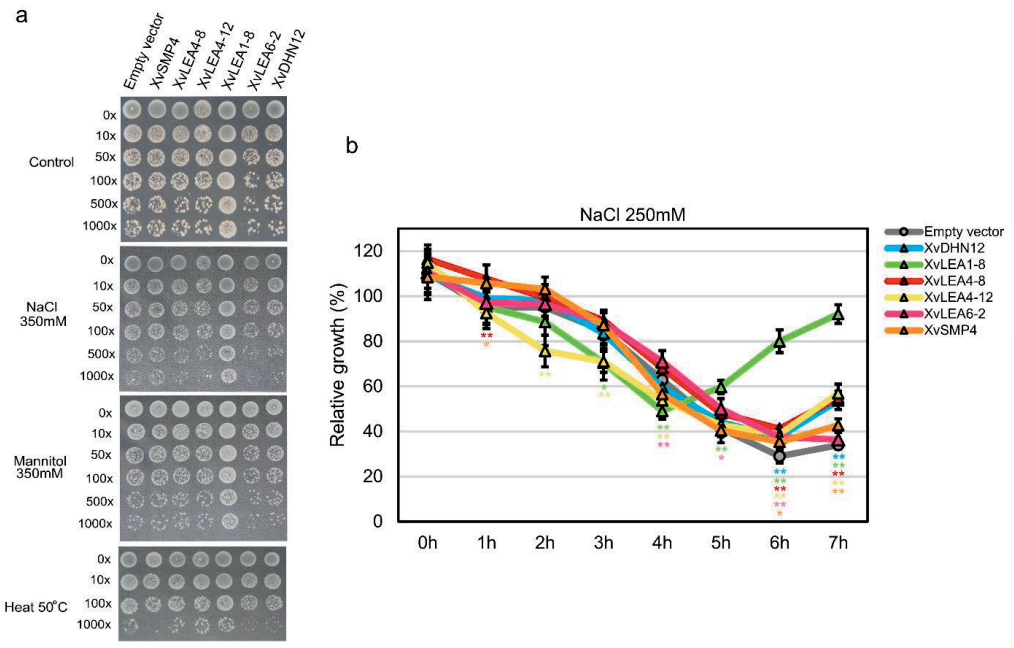


Figure 3 – *In vivo* protective role of XvLEAs. a – For salt and osmotic stresses *E. coli* was spotted on LB media supplemented with NaCl or Mannitol (350mM), respectively. For heat stress cultures were heated at 50°C for 30 minutes, cooled down for 10 min and spotted on LB media. b - Relative growth of *E. coli* in liquid media with 250mM NaCl. Relative growth was calculated as $(OD_{treatment}/OD_{control})*100$. The experiment was repeated twice with three replicates per construct. Statistically significant differences were analysed using Student's *t*-test, and the bars indicates SD (* $p<0.05$ or ** $p<0.01$).

Under control condition the strain expressing *XvLEA6-2* displayed less growth than the empty vector-carrying strain, while the strain expressing *XvLEA1-8* displayed significant better growth than all other strains. Interestingly, the *XvLEA1-8* strain also displayed better growth in all stressful conditions, suggesting that this protein has a higher protective function in *E. coli* when compared to the other XvLEAs. The strain with *XvLEA6-2* and *XvDHN12* presented a slower colony growth in all tested stresses,

and after heat treatment the strain expressing *XvSMP4* also presented slower growth compared to the empty-vector. These results suggest that not all XvLEAs have a protective function during *E. coli* growth under the tested conditions.

When analysing relative growth of the strains in liquid media with a lower concentration of NaCl, we observed that after 6 hours, all strains expressing *XvLEAs* displayed better growth than the empty vector (Figure 3b). Once more, the bacterial cells expressing *XvLEA1-8* showed a remarkable better growth recovery at the end of the experiment, indicating that this protein may perform a better protective function of *E. coli* upon salt stress. Interestingly, in the first 4 hours of stress a few strains displayed slower growth when compared to the empty vector. This observation might be due to leakiness of the expression vector, because the activation of the T7 promoter during bacterial growth prior induction will lead to production of recombinant protein which may limit bacterial growth (Briand *et al.*, 2016; Grossman *et al.*, 1998; Zhang *et al.*, 2015). We hypothesize that the production of some XvLEAs during bacterial growth prior to the stress treatments might have detrimental effects on growth, likely due to unspecific binding to other molecules, however, this hypothesis needs further experimentation in order to be confirmed. Taken together, our results point towards a potential role of XvLEA1-8 in osmoprotection *in vivo*.

Multiple localization of XvLEAs in plant cells

LEA proteins can be localized in various cell compartments, including the nucleus, cytosol, plasma membrane, mitochondria, plastids, endoplasmic reticulum (ER) and vacuole (Candat *et al.*, 2014; Hundertmark and Hincha, 2008). Prediction of subcellular localization indicates that XvDHN12, XvLEA1-8, XvLEA6-2 and XvSMP4 localize mainly in the nucleus and cytoplasm (Supplementary table 3). The nuclear and cytoplasmic localization of proteins from families DHN, LEA_1, LEA_6 and SMP have been shown *in vivo* for *A. thaliana* (Candat *et al.*, 2014), indicating that *in silico* analysis may have strong correlation with *in vivo* analysis for these LEA families. XvLEA4-8 was predicted to be localized in the cytoplasm, nucleus, cell wall and plastids, and XvLEA4-12 in the nucleus, cytoplasm and cell wall, corroborating the experimental data of multilocalization of *A. thaliana* LEA_4 proteins (Candat *et al.*, 2014). Despite limitations of *in silico* predictions, these tools seem to be useful to design experiments for studying *in vivo* subcellular localization of LEA proteins.

In order to verify the *in vivo* subcellular localization of XvLEAs and to compare *in vivo* and *in silico* analysis, we expressed GFP fusions of XvLEA proteins in tobacco leaf epidermal cells under control of a 35S promotor (Supplementary figure 6). The GFP-XvLEA proteins were localized throughout the cell compartments, especially in the cytoplasm, nucleus and membranes. GFP-XvLEA1-8, GFP-XvLEA4-8 and GFP-XvLEA6-2

accumulated mainly in the cytoplasm, membranes and nucleus, in the same way as the GFP empty vector. GFP-XvLEA4-12 showed a signal around the cells and in aggregate-like structures inside the cells which suggest that this protein may be secreted. GFP-XvSMP4 and GFP-XvDHN12 accumulated mainly in the membranes, and GFP-XvDHN12 also in the nucleus. In summary, our results indicate that the six *X. viscosa* LEA proteins involved in DT in leaves are heterogeneously localized throughout various subcellular compartments. Further analyses combining C-terminal fusions and organellar specific markers are necessary in order to draw stronger conclusions on the specific localization of each of these proteins and hypothesize about their cellular role.

Heterologous expression of XvLEA1-8 enhances A. thaliana stress tolerance

A. thaliana ecotype Columbia-0 plants were transformed with a *35S::XvLEA1-8* constructs, which led to a constitutive expression of the *XvLEA1-8* gene, including in the dry seed (Supplementary figure 7). The phenotypic analysis of seeds expressing the *35S::XvLEA1-8* indicated that the germination percentages of WT and five independent lines did not differ significantly (Supplementary figure 8). During germination under salt, osmotic and heat shock stress a mild protective response was observed in two independent lines, indicating that XvLEA1-8 may not play a significant role in seed germination under the conditions tested here. We also investigated the growth ability of seedlings of *35S::XvLEA1-8* in media containing salt and mannitol (Figure 4).

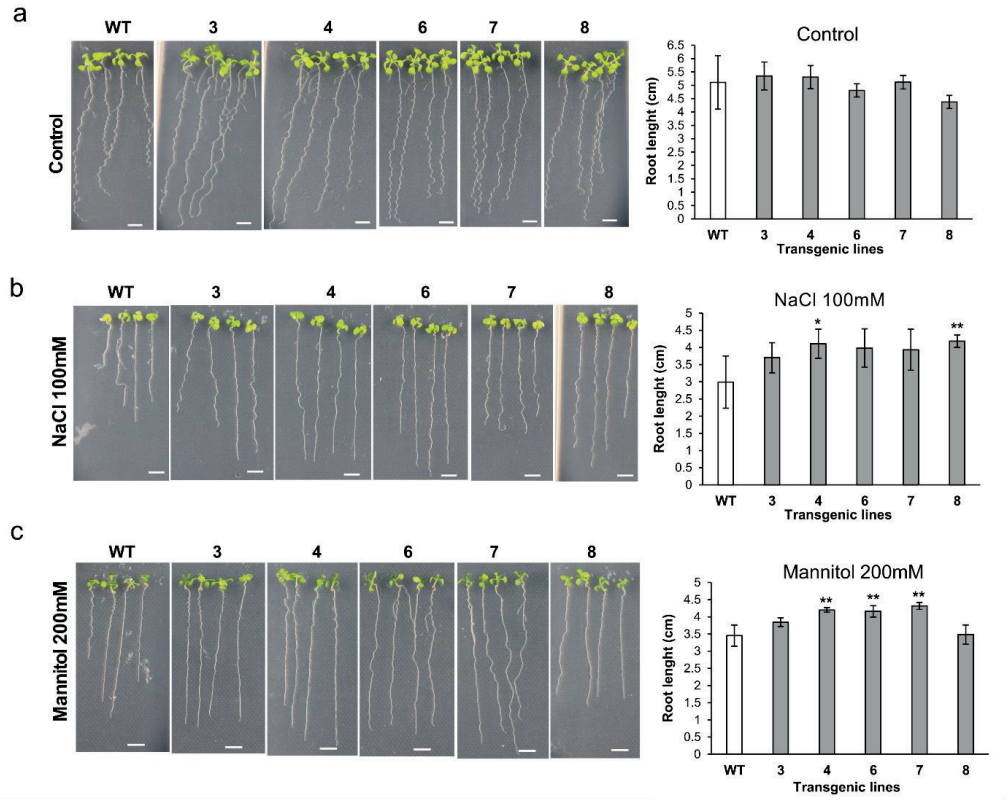


Figure 4 – Phenotypic analysis of *A. thaliana* seedlings expressing 35S::*XvLEA1-8*. Seedlings grown in plates under control conditions (a), 100mM NaCl (b), and 200mM mannitol (c) are shown at the left. Data of one representative experiment is shown. Bars showing average root length ±SD (n=3) are shown at the right. Statistically significant differences between the transgenic lines and the wild type (WT) were tested using Student's *t*-test (* $p<0.05$ or ** $p<0.01$). Scale bars = 0.5cm

Under salt stress the independent lines XvLEA1-8.4 and XvLEA1-8.8 displayed a better growth as indicated by the length of the primary root. Under osmotic stress imposed by mannitol three independent lines (4, 6 and 7) displayed significantly higher growth of the primary root when compared to the wild-type. The phenotypic analysis of transgenic *A. thaliana* adult plants under drought stress showed that three independent lines (3, 7 and 8) displayed significantly higher relative water content at 12 days after withholding water (Supplementary figure 9a), however, at this time point, no significant difference was observed in the dry weight of the transgenic lines when compared to the wild-type (Supplementary figure 9b). These results suggest that the transgenic lines expressing *XvLEA1-8* display a better control of water loss during drying, which may enhance their survival (Supplementary figure 9c). Together, these

findings confirm that XvLEA1-8 is potentially involved in enhancing osmotic stress tolerance in plants.

Discussion

Since their discovery as accumulating during the later stages of seed embryogenesis, increasing evidence suggests a protective function of LEA proteins against desiccation and other stresses, leading to great interest in the structural dynamics of these proteins in the subcellular environment. As in seeds, LEA proteins have been shown to be an essential footprint of vegetative desiccation tolerance in resurrection plants. *Xerophyta viscosa* is a monocot resurrection species of which the genome has recently become available, enabling genomic comparative analysis with other monocots and resurrection species, as well as the understanding of the evolution and functional diversification of LEA proteins and their role in these organisms (Costa *et al.*, 2017b). *LEA* gene expression is upregulated in leaves of *X. viscosa* between 60% to 40% RWC (1.5-1.0gW/gDW) (Supplementary figure 1), concomitant with the activation of the molecular signature of DT in seeds and resurrection plants (Costa *et al.*, 2017a; Illing *et al.*, 2005; Leprince and Buitink, 2010).

In silico analysis of six *X. viscosa* LEA proteins upregulated during desiccation indicated that these are typical intrinsically disordered proteins (or IDPs), as they have intrinsically disordered regions (IDRs) in water, and properties such as a high percentage of polar residues, low GRAVY scores, and extended disordered regions (this study (Battaglia *et al.*, 2008; Dunker *et al.*, 2001; Tompa, 2002). Predictions of biochemical properties may give insight into the lack of structure of a protein; however, it is important to highlight that isolated parameters such as length of polypeptide, net charge, or pI cannot be used as a signature of unfolded or disordered structure (Uversky *et al.*, 2000), revealing the necessity for more extensive *in silico* analysis to better predict the structural nature of such proteins. Using a specific predictor for Classification of Intrinsically Disordered Ensemble Regions (CIDER) (Holehouse *et al.*, 2017), we found that XvLEAs belong to different IDP categories such as environmentally determined disordered, globular, coiled, and chimeras of globular and coiled proteins. Two XvLEAs (XvSMP4 and XvLEA4-12) were predicted as containing structured regions, which is also a common feature of some IDPs able to form globules or chimeras of globules and coils, and are able to undergo folding upon binding (Das *et al.*, 2015; Dyson and Wright, 2002; Dyson and Wright, 2005). In fact, IDPs are able to interact with other molecules via Molecular Recognition Features (MoRFs) that undergo disorder-to-order transitions in the presence of binding partners (Mohan *et al.*, 2006). We found

that MoRFs are present in low number within XvLEA sequences and, in general, they localize within predicted disordered binding regions (DBRs) of the XvLEAs. MoRFs are thought to play important functions in protein-protein interactions related to signalling (Disfani *et al.*, 2012; Mohan *et al.*, 2006; Oldfield *et al.*, 2005). In this way, the investigation of mutations of specific amino acids contributing to MoRFs will provide further insights into the specific role of these regions as well as interactions of LEAs with client proteins.

Reversible phosphorylation of IDPs is one of the major post-translational modifications (PTMs) responsible for functional regulation of binding affinity of MoRFs (Iakoucheva *et al.*, 2004; Mohan *et al.*, 2006) and has already been shown to affect the function of LEA proteins in plants (Alsheikh *et al.*, 2003; Alsheikh *et al.*, 2005; Brini *et al.*, 2007; Heyen *et al.*, 2002; Liu *et al.*, 2017b). *In silico* analysis indicated that all the six XvLEAs have phosphorylation sites, with a remarkable high number in XvLEA4-12. In general, we found that the phosphorylation residues did not coincide with the predicted MoRF contributing residues in XvLEAs, what suggests that phosphorylation may not play a strong role in regulating binding affinity of MoRFs in the studied proteins.

In this study we also investigated the *in vitro* folding dynamics of the six XvLEAs using circular dichroism (CD). Conformational changes of IDPs *in vitro* can be induced by changes in their environment such as pH, temperature, and presence of osmolites or binding targets (Uversky, 2002; Uversky, 2009). Several studies have shown that LEA proteins are mainly disordered in aqueous solution, and are able to acquire secondary structures, mainly alpha-helices, upon desiccation and solute perturbations (Cuevas-Velazquez *et al.*, 2016; Hundertmark *et al.*, 2012; Mouillon *et al.*, 2006; Popova *et al.*, 2011; Shih *et al.*, 2012; Shih *et al.*, 2010). Our CD spectra corroborate the *in silico* predictions of the disordered (or unstructured) nature of the XvLEAs in aqueous solution (Figure 1). XvLEAs also possess residual structured regions which is also commonly found in IDPs (Tompa, 2012). Similar characteristics can be found in polypeptides containing local order such as helix and beta-sheet like structures, and in these proteins secondary structure can be induced by variations in temperature, pH, presence of binding targets, osmolites, and variable solvent concentrations (Bremer *et al.*, 2017; Cuevas-Velazquez *et al.*, 2016; Rivera-Najera *et al.*, 2014; Shi *et al.*, 2002; Soulages *et al.*, 2002; Uversky, 2009). In fact, we observed that XvLEAs are able to acquire higher secondary structure in an hydrophobic solution of acetonitrile (ACN). A highlight was the protein XvLEA1-8, which became fully alpha-helical in a solution of 80% ACN. ACN-mediated folding is triggered by significant reduction of hydration layers in direct contact with the protein surface, leading to conformational changes such as formation of helices and sheets (Gekko *et al.*, 1998; Nelson *et al.*, 1997; Simon

et al., 2001). This indicates that under highly hydrophobic conditions XvLEA proteins may acquire secondary structure, which may also be a factor regulating their functional activity.

It has already been shown that IDPs are able to preserve enzyme activity and avoid protein aggregation upon cellular stress by the ability to vitrify and trap cellular macromolecules into an amorphous matrix avoiding aggregation (Boothby *et al.*, 2017; Chakrabortee *et al.*, 2007). Several LEA proteins are able to protect enzymes from thermal and chemical inactivation and aggregation *in vitro* (Agarwal *et al.*, 2017; Furuki and Sakurai, 2016; Goyal *et al.*, 2005b; Halder *et al.*, 2017; Kovacs *et al.*, 2008; Liu *et al.*, 2016a; Nakayama *et al.*, 2007). We performed *in vitro* and *in vivo* assays in order to investigate if the ability to undergo conformational changes and acquire higher secondary structure under stress correlates with the protective ability of XvLEAs. With the exception of XvLEA4-12, all studied proteins were able to preserve enzymatic activity upon desiccation, heat and oxidative stress *in vitro* (Figure 2), which supports the hypothesis that the high conformational changing ability correlates with protective abilities against aggregation and denaturation. Furthermore, our results indicates that the *in vitro* protective properties of LEAs are, to some extent, conserved in different organisms. *In vivo* assays using *E. coli* have successfully demonstrated the protective role of LEAs upon stress (Boothby *et al.*, 2017; Drira *et al.*, 2015; Gao and Lan, 2016; Hu *et al.*, 2016; Jiang *et al.*, 2017; Ling *et al.*, 2016; Liu *et al.*, 2016a; Rakhra *et al.*, 2017; Saucedo *et al.*, 2017; Shi *et al.*, 2016; Wang *et al.*, 2017; Zhang *et al.*, 2014; Zhou *et al.*, 2017). We found that XvLEAs are able to improve bacterial growth under salt and osmotic stress, and the strains expressing XvLEA1-8 presented a faster stress recovery. Combined, these results indicate a correlation between *in vitro* and *in vivo* protective functions, and point towards a potential application of the properties of XvLEA1-8 for engineering stability *in vitro* and *in vivo*.

Subcellular localization analyses enable to infer the actual function of the proteins in plants. *A. thaliana* members of DHN, LEA_1, LEA_6, and SMP were shown to localize majorly in the nucleus and cytoplasm, while LEA_4 members are multilocalized across chloroplasts, mitochondria, ER and pexophagosomes (Candat *et al.*, 2014). *In silico* analysis of the six XvLEAs investigated in this study confirms the expected nucleo-cytoplasmic localization of XvDHN12, XvLEA1-8, XvLEA6-2 and XvSMP4, while XvLEA4-8 and XvLEA4-12 were also predicted to localize to the plastids and cell wall. In some cases, *in silico* and *in vivo* analysis may result in different outputs, as the example of AdLEA protein from wild peanut, of which *in silico* analysis indicated localization mainly in the cytoplasm, while the GFP-fused protein was localized in the nucleus and cytoplasm (Sharma *et al.*, 2016). In our *in vivo* analysis using GFP-fusions similar results as the *in silico* predictions were observed for most XvLEAs, suggesting that *in silico*

predictions can be useful for preliminary characterization of subcellular localization. In order to investigate the transferability of the protective properties shown by *in vitro* and *in vivo* assays into enhancing plant stress tolerance, we developed transgenic *A. thaliana* plants constitutively expression the XvLEA1-8 gene. The expression of XvLEA1-8 did not enhance stress tolerance during seed germination, but was able to enhance primary root growth under salt and osmotic stress in seedlings. Transgenic adult plants expressing XvLEA1-8 displayed a higher relative water content (RWC) after 12 days of drought, indicating that this gene may play a role in plant osmoregulation.

Our study provides evidence for the structure-function relationship of LEA proteins expressed during desiccation in *X. viscosa* plants. Our data reveals that *in silico* and *in vitro* analysis can provide useful information about LEA protein functions. We hypothesize that XvLEAs have been evolutionarily selected to be able to adopt diversified conformations driven by variations in their cellular environment. The conformational plasticity and multilocalization of XvLEAs may enable binding to essential cellular components (such as enzymes), and to regulate loss of water from the cells, resulting in enhanced osmotic stress tolerance in *X. viscosa* leaf cells. Furthermore, we believe that the high conformational plasticity and protective abilities of XvLEA1-8 make it a potential candidate for engineering biostability *in vitro* by serving as a model for synthetic chaperons, as well as for enhancing drought tolerance in crop species.

Conclusion

We presented the structural and functional characterization of six LEA proteins of the resurrection monocot *X. viscosa*. XvLEAs are IDPs able to acquire different levels of secondary structure in hydrophobic solutions, which correlates with their ability to preserve LDH activity upon desiccation, heat, and oxidative stress, as well as to enhance *E. coli* cell survival upon salt and osmotic stress. Multilocalization of XvLEAs in *N. benthamiana* leaves suggests that they may offer general cellular protection upon stress. The high conformational plasticity and protective functions of XvLEA1-8 indicate that this protein is a good candidate for enhancing desiccation tolerance *in vitro* and *in vivo*.

Material and Methods

In silico analysis

Protein Grand Average Hydropathy was calculated using the GRAVY calculator (http://www.gravy-calculator.de/). Molecular mass and isoelectric point (pI) were calculated with the Isoelectric Point Calculator (IPC) (Kozlowski, 2016). Protparam (http://web.expasy.org/protparam/) was used to analise amino acid composition and predict protein stability. The percentage of polar residues was calculated with EMBOSS PEPSTATS (https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/). Prediction of the degree of protein disorder was performed using IUpred (Dosztányi *et al.*, 2005) and PONDR (http://www.pondr.com/) with default parameters (VLXT predictor). Sequence-specific parameters and prediction of structural qualities of the proteins were predicted with CIDER (http://pappulab.wustl.edu/CIDER/analysis/) (Holehouse *et al.*, 2017). ANCHOR web server (http://anchor.enzim.hu/) was used to predict the number of disordered binding regions (DBRs), i.e. regions with propensity to undergo folding upon partner-binding with a probability higher than 50%, and MoRFpred (http://biomine.cs.vcu.edu/servers/MoRFpred/) was used to predict the number of residues of Molecular Recognition Features (MoRFs). DEPP (Disorder Enhanced Phosphorylation Predictor), also known as DisPhos, available from http://www.pondr.com/ was used to predict phosphorylation sites within the protein sequences (Iakoucheva et al., 2004). Subcellular localization was predicted with PSI (http://bis.zju.edu.cn/psi/), CELLO2GO (http://cello.life.nctu.edu.tw/) and Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/). Signal peptide predictions were performed with SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/), PrediSi (http://www.predisi.de/) and TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/).

Plant materials and growth conditions

Xerophyta viscosa plants were grown in a greenhouse with 16h at 27°C during the day and 8h at 18°C during the night. Dehydration was achieved by withholding water from the pots. Young green leaf tissue was harvested from 4 individual plants on a daily basis, frozen in liquid nitrogen, and stored at -80°C. *Nicotiana benthamiana* plants used for agro infiltration and *Arabidopsis thaliana* plants ecotype Columbia-0 used for floral dipping and drought treatments were grown on a mix of 50% vermiculite and 50% soil watered three times a week with Hyponex solution, in a greenhouse with long-day photoperiod cycles (16 h light/8 h dark) at 22°C ± 2°C for three weeks. *A. thaliana* plants used for seed collection were grown on Rockwool blocks (Grodan, the Netherlands) in Hyponex solution under greenhouse conditions (16h light/8h dark).

RNA extraction, cDNA synthesis and cloning of XvLEAs

Total RNA was extracted from leaves of adult *X. viscosa* plants nine days after dehydration using the hot borate protocol (Wan and Wilkins, 1994). cDNA was synthesized using the iScript™ cDNA Synthesis Kit (Bio-Rad, Laboratories B.V., The Netherlands) according to the manufacturer's protocol. Primers were designed to include flanking restriction sites, which were in turn flanked with Gateway AttB1 and AttB2 sites (Supplementary table 4). PCR was performed on *X. viscosa* cDNA with Q5 high-fidelity DNA polymerase (New England Biolabs) and purified from gel using Nucleospin Gel and PCR clean-up (Machery-Nagel). The purified amplicons were cloned into the entry vector pDONr201 using BP Clonase II according to the manufacturer's instructions (Thermo Fisher Scientific) to create pENTr201-LEA and sequenced to confirm the gene sequence.

Construction of plant expression vectors

For *N. benthamiana* transient gene expression, subcloning from pENTr201-LEA into pGWB606 (Nakamura *et al.*, 2010) was performed with a Gateway LR reaction in order to produce recombinant *p35S::GFP-XvLEAs*. Correct reading frame was confirmed with sequencing. For *A. thaliana* heterologous expression, subcloning of the *XvLEAs* CDS from pENTr201-LEA into the expression vector pB7WG2RS was also performed with a Gateway LR reaction. The expression vector pB7WG2RS was made by cloning the RedSeed selection marker (*pNAP::DsRed*) from pKGW-RedSeed between the left T-DNA border and the *Bar* resistance gene of pB7WG2 (Karimi *et al.*, 2002) using XbaI and KpnI. Insertion of the RedSeed marker was confirmed by sequencing.

Construction of bacterial expression vectors

For protein expression and purification, *XvLEAs* CDS were cloned from pENTr201-LEA into bacterial expression vector pDEST17 with a Gateway LR reaction in order to produce recombinant N-terminal His-tagged proteins (pEXPR17-LEA). The correct reading frame was checked by sequencing. For *E. coli* stress assays, the *XvLEAs* were subcloned from pENTr201-LEA to the bacterial expression vector pCDF-Duet (Novagen), which contains two multiple cloning sites (MCSs), using traditional cloning. *XvLEA6-2* and *XvDHN12* were subcloned to MCS1 using NcoI and EcoRI, while *XvSMP4*, *XvLEA4-8*, *XvLEA4-12* and *XvLEA1-8* were subcloned to MSC2 using NdeI and AvrII (New England Biolabds; supplementary table 1). In all cases, both pENTr201-LEA and pCDF-Duet were digested with the corresponding enzymes. The fragment containing the *XvLEA* CDS from the digested pENTr201-LEA was purified by gel extraction (Machery-Nagel). The digested pCDF-Duet vector was subjected to dephosphorylation

(shrimp alkaline phosphatase (rSAP) from New England Biolabs) and purified with a silica column (Machery-Nagel). The *XvLEA* fragment was ligated into pCDF-Duet using T4 DNA Ligase (Promega) to create pCDF-LEA. Correct orientation was confirmed by PCR.

Expression and purification of the recombinant XvLEAs

BL21 (DE3) pLysS (Novagen) *E. coli* expression strain carrying the different pDEST17-LEA vectors were grown overnight in 5mL sterile LB medium at 37°C (200rpm). The cultures were inoculated in 1L of sterile LB medium in a 5L Erlenmeyer flask and incubated at 37 °C with shaking (200 rpm). When an optical density at 600 nm (OD_{600}) of 0.6 was reached (± 3 hrs), isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1mM and incubation continued for 2 hours at 37°C (100rpm) to induce protein expression. The cells were then harvested by centrifugation at 4°C, 10,000*g* for 10 minutes and the cell pellets were frozen overnight at -20°C. The cell pellets were then thawed an resuspended in 15mL Lysis-Equilibration-Wash Buffer (LEW; Machery-Nagel) containing 50mM NaH_2PO_4 and 300mM NaCl with a pH of 8.0 using NaOH. One mg/ml of lysozyme (Sigma-Aldrich) was added and followed by incubation on ice for 30 minutes. The cells were lysed by sonication and the lysate was centrifuged for 30 minutes at 4°C, 10,000*g*. For purification of the His-tagged recombinant LEA proteins from the crude lysate, the Protino® Nickel TED (Ni-TED) Histidine Tag Affinity Purification Kit (Macherey-Nagel, Germany) was used with standard protocol. The purified His-tagged Xv-LEAs were subjected to concentration and buffer exchange into dH_2O using the Amicon Ultra Centrifugal Filters (3K for XvLEA4-8, XvLEA1-8, XvLEA6-2 and XvDHN12, and 30K for XvSMP4 and XvLEA4-12, MWCO, Merck Millipore). Centrifugation was performed at 8,000*g* for 1 hour or 14,000*g* for 15 minutes for the 3K and 30K filters respectively, at 4°C to concentrate the proteins, followed by three washing steps with 12mL of dH_2O (the same centrifugation parameters where used). Total protein was quantified using the Bradford BioRad Microassay (BioRad USA) according to manufacturer's instructions with Bovine Serum Albumin (BSA; Sigma Aldrich, USA) as a standard. As an additional purification step, making use of the heat stability of most LEA proteins (Boudet *et al.*, 2006), 10-15µg of the recombinant proteins were submitted to a 97°C treatment for 10 minutes, centrifuged, and the supernatant was analysed on a 12% SDS-Page at 90V for 2 hours with a Colour Prestained Protein Standard Ladder (New England Biolabs, USA). After electrophoresis, gels were stained with a Coomassie Blue solution (2.5g/L Coomassie Blue (Sigma-Aldrich), 50% v/v methanol, 10% v/v acetic acid) for 1 hr at room temperature and destained overnight in a solution of 45% v/v methanol and 10% v/v acetic acid prior to visualization.

In-solution circular dichroism analysis for secondary structure

Circular dichroism (CD) spectra of the six recombinant His-tagged LEAs were obtained using a JASCO J-810 Spectropolarimeter (JASCO Analytical Instruments, Japan) in 1mm path-length quartz cuvettes. Spectral data were recorded from 240nm to 190nm, with 10 accumulations per run using a 0.2nm data pitch, 100nm/min scanning speed, 1 second response time and 1nm of band width. To simulate changes in secondary structure we used 80% acetonitrile (ACN), 20mM NaCl, and MiliQ water adjusted with HCl to pH 2.3 or pH 4.0 as protein solvents. BSA (Sigma-Aldrich, USA) was used as a control for the spectra of an alpha-helical structure, at a concentration of 0.075mg/ml. Measurements of millidegrees obtained from the results were subsequently converted into mean residue (θ) and plotted against the wavelength range (nm). CD data were fitted using Dichroweb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) (Contin, dataset 7) (Whitmore et al., 2004) to estimate the secondary structure content.

In vitro lactate dehydrogenase assays

Lactate dehydrogenase (LDH) assays were adapted from Reyes et al. (2008). In short, LDH from rabbit muscle (Sigma-Aldrich) was diluted to a final concentration of 200nM in 25mM Tris pH 7.5. The purified XvLEAs or BSA were also diluted to a final concentration of 200nM in Tris-HCl 25mM. The mixture of LDH and XvLEA at a molar ratio of 1:1 was submitted to desiccation, heat or oxidative stress. After each treatment the enzyme and protein mixture was added to a reaction buffer containing 25mM Tris-HCl pH 7.5, 2mM of pyruvate (Sigma-Aldrich) and 0.15mM NADH (Roche) to a final volume of 1mL in 2mL plastic cuvettes and the initial absorbance was measured at 340nm. The rate of the decrease in absorbance due to the conversion of NADH to NAD^+ was determined every 5 seconds for 1.5 minutes at 25°C. For the desiccation-induced aggregation assay, LDH in the presence of each of the purified XvLEAs or BSA were submitted to dehydration in a centrifugal evaporator (Savant™ SpeedVac™ Plus SC210A) for 1 hour at room temperature. After dehydration, the initial volume was restored by adding 25mM Tris-HCl pH 7.5 and enzyme activity was measured as mentioned above. For the thermal inactivation assays, the samples containing each of the six purified XvLEAs or BSA together with LDH were heated at 42°C for 20 minutes, and cooled down at room temperature for 10 minutes. Oxidative stress was imposed by incubating the LDH enzyme and XvLEAs or BSA mixtures in 200mM of H_2O_2 (Sigma-Aldrich) at room temperature for 1 hour. Each assay was repeated three times with three technical replicates each and statistically significant differences were analysed using Student's *t*-test.

Abiotic stress tolerance assays of E. coli transformants

E. coli strain BL21(DE3)RIL (Agilent) was transformed with the pCDF-Duet-LEA vectors. LEA expression was induced with 1mM IPTG for 2 hours. Cells were diluted to OD_{600}=1.0 and a serial dilution of 5µL was spotted onto LB media containing 250mM or 350mM of NaCl or mannitol. To assess *in vivo* heat protective function 1mL of LB containing IPTG-induced *E. coli* cells (OD_{600}=1.0) was incubated in a water bath at 50°C for 30 minutes, cooled for 10 minutes at room temperature and then spotted onto LB control plates supplemented with 1.5% Daishin agar (Duchefa). Serial dilutions at 10, 50, 100, 500 and 1000 times were used for salt and osmotic stresses, and 0, 10, 100 and 1000 times for heat stress assays. The plates were incubated at 37°C for 16 hours. For liquid media growth assay 1mL of cells at OD_{600}=1.0 was diluted 10 times with LB liquid media supplemented with 250mM NaCl. The cells were kept at 37°C and 100 µl aliquots were taken every hour for measuring the OD_{600}. Due to the possibility of leaky vector activation, resulting in expression of proteins before induction with IPTG (Briand *et al.*, 2016; Grossman *et al.*, 1998; Zhang *et al.*, 2015), we analysed and compared the relative growth percentage of the XvLEA strains with the empty vector-carrying strain. The difference in OD_{600} at time point *x* (t*x*) between treated and non-treated (control) cultures where used to calculate the relative growth, expressed as $[OD_{treatment(tx)}/OD_{Control(tx)}]$x100. Each experiment was repeated twice and statistical differences were determined by using the Student's *t*-test.

In vivo localization of XvLEAs in N. benthamiana leaves

Agrobacterium tumefaciens strain AGL0 carrying constructs of pGWB606 *p35S::GFP-LEAs* were grown in LB medium containing antibiotics and harvested by centrifugation at 4,000g for 5 minutes at room temperature. The bacteria culture was resuspended in infiltration buffer (10mM $MgCl_2$, 10mM MES pH 5.6 and 200µM acetosyringone) and adjusted to OD_{600}=0.6. The bacterial suspension was incubated at room temperature on a rocking platform for 1 hour. Leaves of three to four week-old *N. benthamiana* plants were infiltrated with *A. tumefaciens* using a needleless syringe and a minimum of three independent agro infiltrations was performed ($n \geq 3$). Three days after infiltration, two leaves from three independently transformed plants were analysed for GFP fluorescence under a Leica TCS SP8 HyD confocal microscope (Leica) with an excitation wavelength of 488nm and the spectral detection was set between 500–557nm for GFP, and 642–747 nm for chlorophyll fluorescence. The objective used was 40x in water immersion.

A. thaliana transformation and RT-qPCR gene expression confirmation

The pB7WG2RS *p35S::LEA1-8* construct was introduced into *A. tumefaciens* strain AGL0 and transformed into *A. thaliana* Colombia-0 (Col-0) using the floral dip method (Clough and Bent, 1998). Using the RedSeed marker transgenic T1 plants were selected and single T-DNA inserts were identified in the T2. RNA of T3 dry seeds was extracted and 700ng of RNA was reverse transcribed as described above. RT-qPCR reactions were run on a CFX machine (Bio-Rad). Three technical replicates were used per sample. The reference genes used for data normalization were At4g12590 and At4g34270 (Dekkers et al., 2012). The primers used for RT-qPCR and the expression results are presented in Supplementary table 5. After confirmation of the expression of *XvLEA* genes in T3 seeds, plants were grown in a complete randomized design containing three biological replicates of at least four plants. Seeds were harvested and used for further experiments.

Seed, seedling and plant stress phenotyping

Germination was performed 10 days after harvest. For control conditions, seeds were sown in square trays on two layers of filter papers saturated with dH_2O, in accordance with Joosen *et al.* (2010). For salt and osmotic stress the filter papers were saturated with -0.3MPa NaCl or -0.6MPa mannitol respectively. After sowing, the seeds were stratified at 4°C in darkness for 48 hours and after it the trays were moved to 22°C continuous light. To access tolerance to deterioration conditions dry seeds were incubated at 40°C at approximately 85% relative humidity for three days, and then transferred to germination conditions at 22°C and continuous light. Germination percentages were scored during five days using the Germinator program (Joosen *et al.*, 2010). For seedling stress assays, seeds were stratified at 4°C in darkness for 72 h on square 14cm Petri dishes on ½ Murashige-Skoog (MS) medium supplemented with 0.5% sucrose, 0.1% MES monohydrate and 1% Daishin agar, pH 5.8 (KOH). After 4 days, seedlings were transferred to the same medium supplemented with 100mM NaCl or 200mM Mannitol. Each plate contained four seedlings of three genotypes, and three biological replicates per genotype were used following a complete randomized design. The plates were vertically placed at 21°C, under continuous light. After 10 days the primary root length was scored. Drought stress was assessed by withholding water from 3 weeks-old *A. thaliana* plants grown on soil in the greenhouse. After 12 days relative water content (RWC) from the leaves was measured, and plants were rewatered. Final survival was analysed 7 days after rewatering. Three biological replicates with three plants were used per time point. Leaves and soil fresh weight (FW) were measured immediately after harvest, and the dry weight (DW) were measured after 48 hours at 60°C. RWC (%) from the plants and soil were measured as (FW–DW)/FW×100.

Acknowledgements

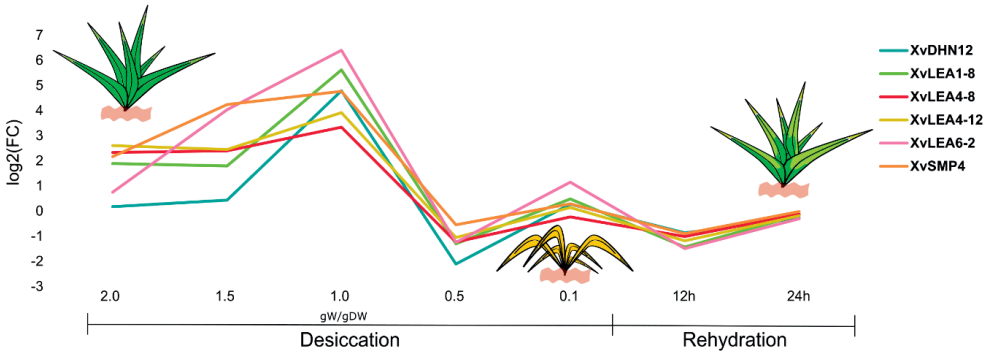
M.A.S.A received financial support from CAPES – Brazilian Federal Agency for Support and Evaluation of Graduate Education (BEX 0857/14-9). The authors thank Esmer Jongedijk (Wageningen University) for providing the pCDF-Duet empty vector, Jeanne Korsmann (Univesity of Cape Town) for providing competent cells of BL21 (DE3) pLysS, Mahdu Chauan (University of Cape Town) for the technical assistance with the CD-spectra analysis, and Dr. Eric Limpens (Wagenigen University) for the technical support with the confocal microscopy.

Author contributions

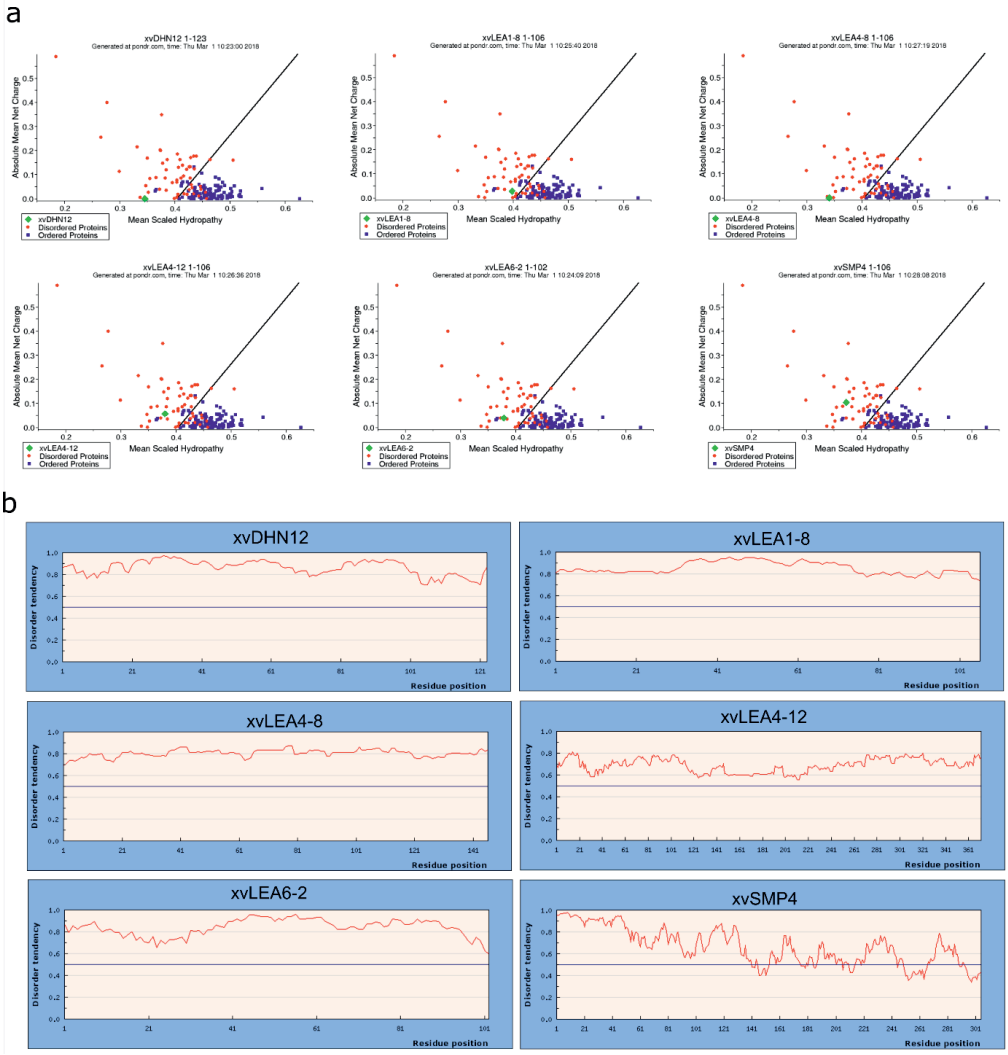
M.A.S.A performed the *in silico* analysis and data interpretation. M.A.S.A and J.R. performed gene cloning for bacterial assays. J.R. performed gene coning for plant heterologous expression. M.A.S.A. performed the *in vitro* assays and data analysis with significant contribution of T.J.D.. H.W.M.H., W.L. and J.M.F. participated in the design of the study and supervised the project. M.A.S.A. wrote the manuscript. All authors contributed with comments and revision of the manuscript.

Supplementary Information

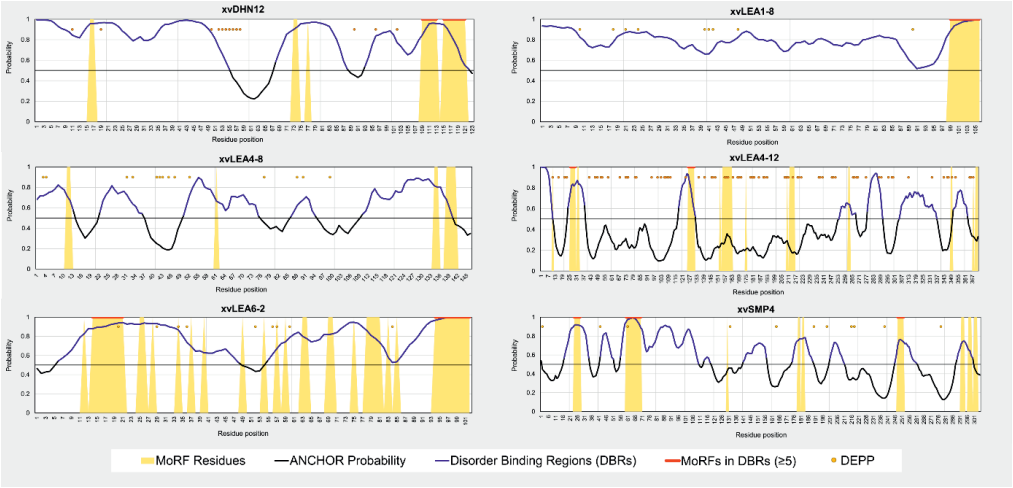
Supplementary Figures



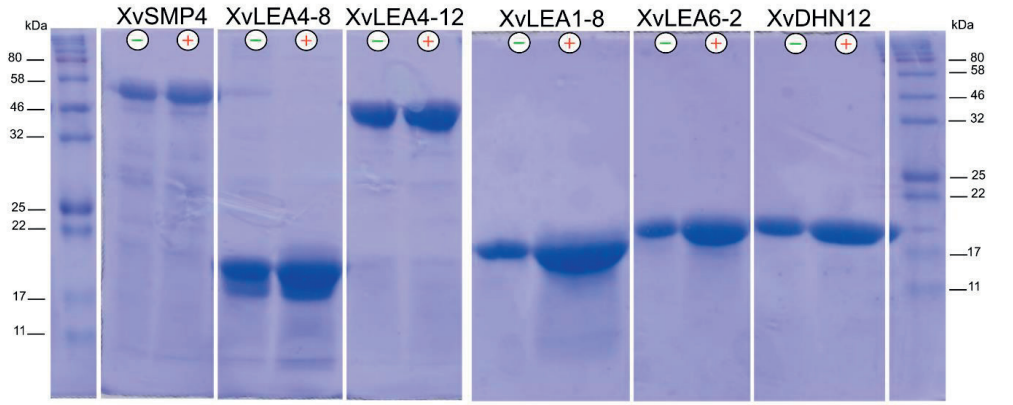
Supplementary figure 1 – Expression patterns of *X. viscosa LEAs*. The log2fold change was calculated relative to the expression in the previous time-point. Data from Costa et al. (2017).



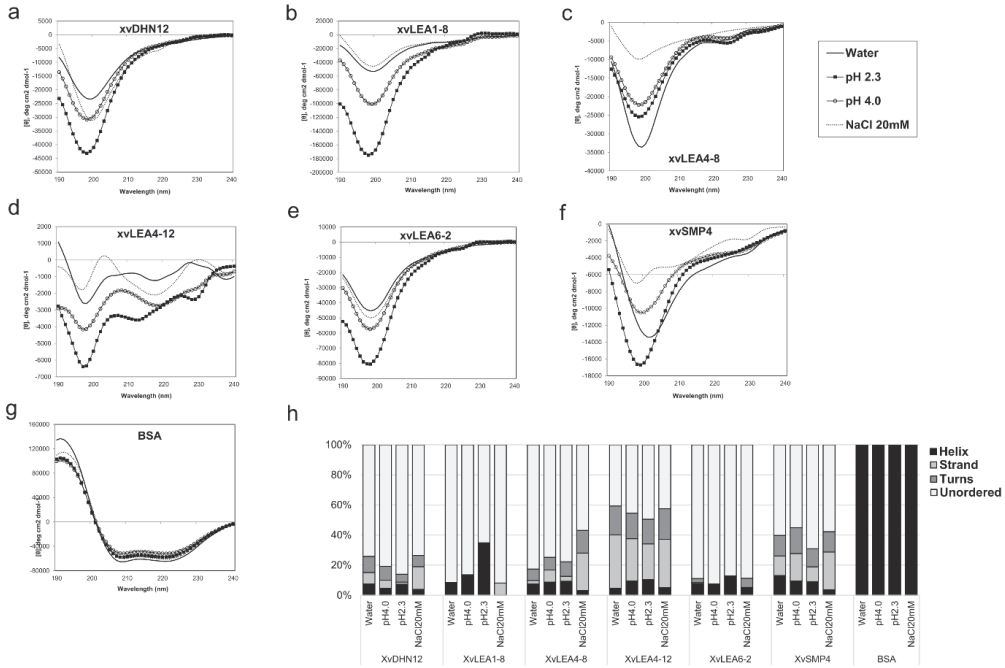
Supplementary figure 2 – *In silico* predictions of intrinsic disorderedness of six XvLEAs. a – Charge-hydropathy (C-H) plot including LEAs from *X. viscosa* (green diamonds) that are plotted together with known globular proteins (blue squares) and known intrinsically disordered proteins (red circles). The prediction was performed with PONDR (http://www.pondr.com/). b – Disorder tendency of the six XvLEAs predicted with IUPRED (Dosztanyi et al., 2005). Residues above 0.5 are predicted to be disordered, while those below 0.5 are considered ordered.



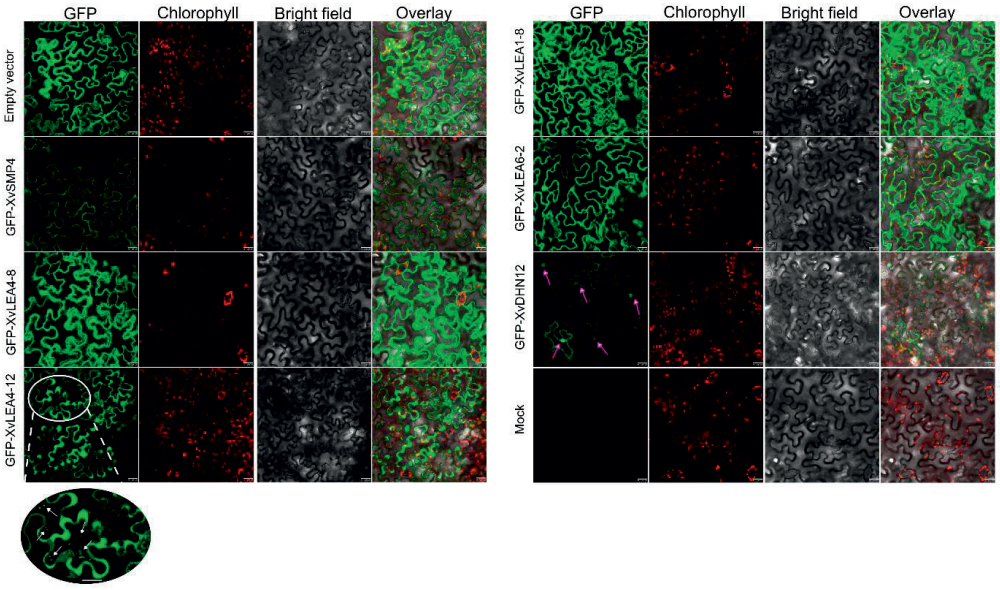
Supplementary figure 3 – *In silico* prediction of intrinsic disorder properties of six XvLEAs. The plots show the ANCHOR probability of disorder binding regions (line), the positions of Molecular Recognition Feature (MoRF) residues (yellow blocks) and the disordered binding regions (DBRs) with a probability above 0.5 of the ANCHOR probability (blue highlighted line). Positions containing more than 5 MoRF residues within DBRs are indicated by a red line and phosphorylation prediction of residues (Y,S,T) is shown by orange dots.



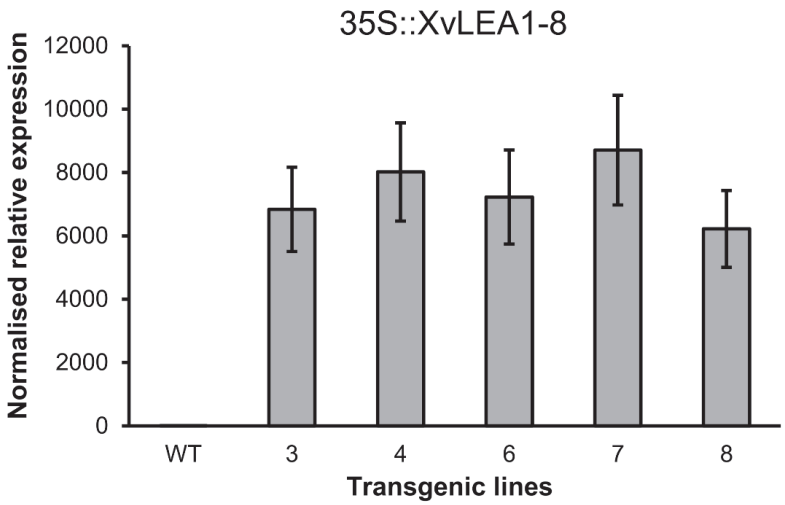
Supplementary figure 4 – Coomassie stained 12% SDS-PAGE gel showing purified recombinant XvLEA proteins. (-) Indicates 10µg of purified recombinant proteins loaded on gel without heating and (+) indicates 15µg of purified recombinant proteins loaded on gel after an extra heating step at 97°C for 10 minutes.



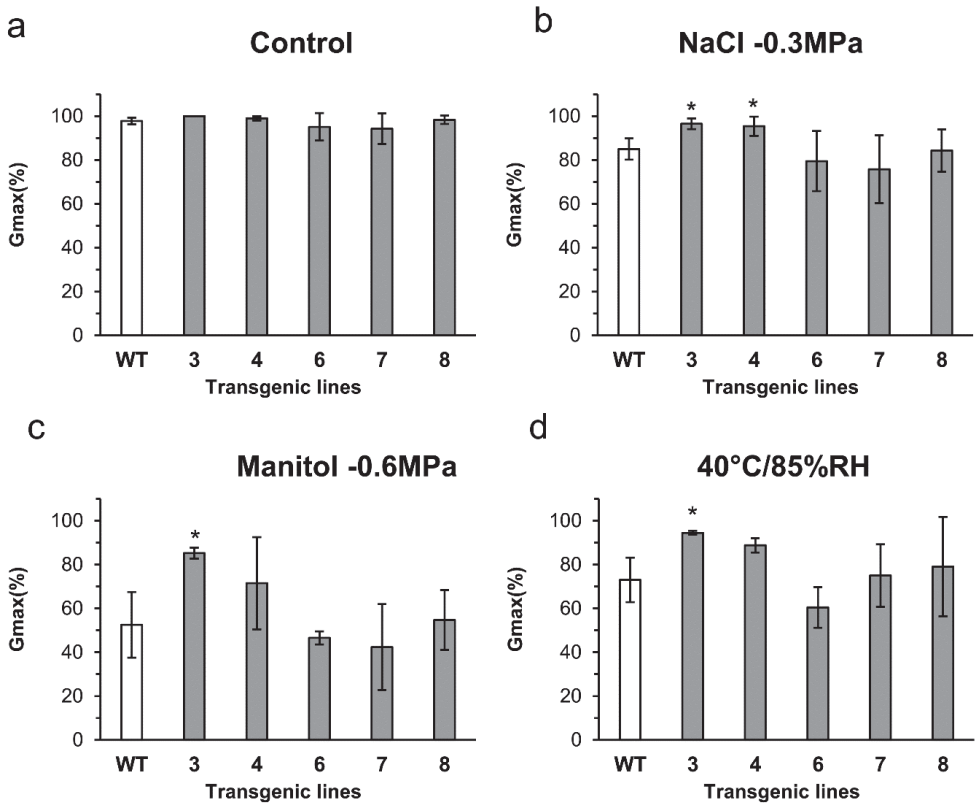
Supplementary figure 5 - Circular dichroism analysis of XvLEAs. CD spectra of *X. viscosa* LEA proteins and BSA (a-g). The CD spectra was obtained in water, water adjusted with 1M HCl to pH 2.3 and pH 4.0, and in 20mM NaCl. All the spectra were analysed at room temperature. The solutions were prepared about 1 hour before acquisition of the spectra. The graphs show the spectra obtained after subtracting the reads of a blank sample (water only). h – Secondary structure content. Predictions of the content of helix, strand, turns and unordered regions for the different proteins were performed with Dichroweb.



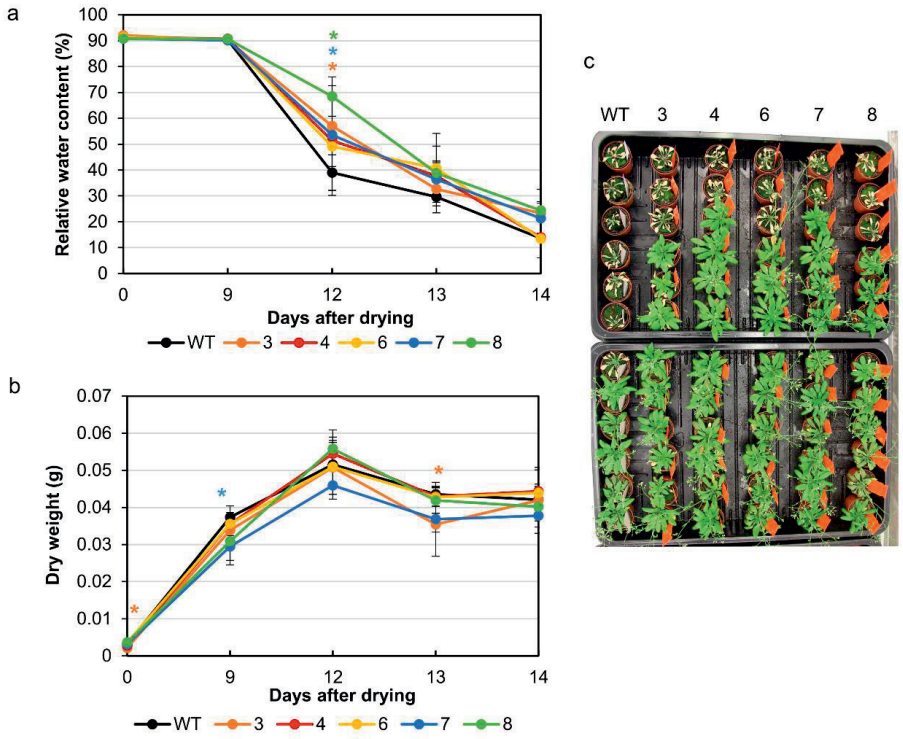
Supplementary figure 6 – **Subcellular localization of the GFP fusion proteins in transgenic *N. benthamiana* plants**. The subcellular distribution of the GFP fusions and control GFP in the epidermal leave cells of transgenic tobacco leaves were visualized using laser confocal scanning microscopy. Red indicates the auto fluorescence of chlorophyll and GFP fluorescence is presented in green. Pink arrows in GFP-XvDHN12 indicate the nucleus. The white arrows in the circle snapshot of GFP-XvLEA4-12 show vesicle-like structures. Scale bars = 25μM.



Supplementary figure 7 – Expression of 35S:*XvLEA1-8* in *A. thaliana*. RT-qPCR analysis was performed in cDNA produced from RNA extracted from *A. thaliana* dry seeds in order to confirm the expression of 35S::*XvLEA1-8*. Data was analysed according to the ΔΔCT method. Bars indicate means ± SD (*n* = 3).



Supplementary figure 8 – Germination of transgenic 35S:*XvLEA1-8 A. thaliana* seeds under stress. a – Seeds were stratified at 4°C for 48 hours and germinated for 5 days at 22°C and continuous light. b – Stratification and germination in the presence -0.3MPa of NaCl. c - Stratification and germination in the presence -0.6MPa of Mannitol. d – Heat stress at 40°C and 85% relative humidity (RH) prior germination at 22°C and continuous light. Statistically significant differences were analysed using Student's *t*-test (* $p<0.05$).



Supplementary figure 9 – Phenotypic analysis of *A. thaliana* adult plants expressing 35S::*XvLEA1-8* under drought. a – Relative water content and b – Dry weight of adult plants well-irrigated (day 0), or after withholding water for 9, 12, 13 and 14 days. Student's *t*-test showed significant differences between transgenic independent lines (3, 4, 6, 7, and 8) and the wild-type (WT). Bars indicate means ± SD (* $p<0.05$). c – Plants rehydrated for 7 days after 12 days of drought.

Supplementary Tables

The supplementary material of this chapter can be downloaded from
http://www.wageningenseedlab.nl/thesis/masilvaartur/SI/

Chapter

5 Physiological characterization of *Arabidopsis thaliana LATE EMBRYOGENESIS ABUNDANT* protein mutants

Mariana A. Silva Artur[1], Francis Okot[1], Harma Drent[1], Wilco Ligterink[1] & Henk Hilhorst[1]

In preparation for publication

[1]Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands.

Abstract

Late Embryogenesis Abundant (LEA) proteins are known as protective proteins which are involved in the acquisition of desiccation tolerance (DT) at the later stages of embryo development. In *Arabidopsis thaliana*, *LEA* mRNAs were found in dry mature seeds and in germinated seeds treated with polyethylene glycol (PEG), suggesting that they may contribute to other seed traits such as dormancy, longevity, and germination under stressful conditions. To investigate this hypothesis and the specific role of individual *LEA* genes, we identified and functionally characterized T-DNA lines of *A. thaliana LEA* genes (*AtLEAs*). We found that the downregulation of gene expression for individual LEAs leads to minimal effects on acquisition of DT in seeds of most tested lines, and that the effect of single mutations has little effect on seed dormancy and germinability. This suggests a tightly interconnected regulatory control and likely high functional redundancy between LEA proteins. Interestingly, a knockout mutant of At3g17520 (a*tlea30-1*) showed lower DT during seed development, and displayed a higher dormancy level in freshly harvested mature seeds compared to Columbia (Col-0) wild-type seeds. The *atlea30-1* mutant seeds also displayed lower longevity and increased sensitivity to stress. However, the ability to re-induce DT with PEG in these mutant seeds was not affected. These findings suggest that *AtLEA30* is downstream of the common network of DT, longevity and dormancy, and that this gene may play important but not essential roles in post-embryonic processes.

Key words: germination, LEA proteins, longevity, seed maturation.

Introduction

One of the key factors for the successful colonization of land by angiosperms was their ability to protect their embryos for long periods in a dry and dormant state until the environmental conditions become favourable. Orthodox seeds constitute an anhydrobiotic propagule in which desiccation tolerance (DT) is acquired during embryo development (Dekkers *et al.*, 2015). During orthodox seed development two main phases can be distinguished: embryogenesis and maturation (Angelovici *et al.*, 2010; Bewley *et al.*, 2012; Vicente-Carbajosa and Carbonero, 2005).

During early embryogenesis the basic root-shoot body pattern is established, and the primary organs – axis and cotyledons – follow distinct developmental fates (Goldberg *et al.*, 1994). During late embryogenesis there is extensive activation of gene sets such as those of the Late Embryogenesis Abundant proteins (LEAs), which encode proteins necessary for posterior desiccation and postembryonic life (Delseny *et al.*, 2001; Goldberg *et al.*, 1994; Leprince *et al.*, 2017). During the maturation phase, cells start to lose water, and processes such as embryo degreening, dormancy induction, and accumulation of reserves are initiated (Angelovici *et al.*, 2010; Leprince *et al.*, 2017; Ooms *et al.*, 1993a).

Genetic approaches, both by chemical- as well as insertional mutagenesis, have contributed to a fast comprehension of processes affecting embryogenesis and maturation (Koornneef *et al.*, 1982; Nambara *et al.*, 1998; Nambara *et al.*, 1994; Nambara *et al.*, 1992). The identification of the Arabidopsis *aba insensitive* (*abi*) loci lead to the identification of a B3 domain transcription factor (ABSCISIC ACID INSENSITIVE3; ABI3) which mutation leads to defective chlorophyll degradation, DT, dormancy and longevity (Delahaie *et al.*, 2013; Nambara *et al.*, 1994; Nambara *et al.*, 1992; Ooms *et al.*, 1993a). ABI3, together with the master transcriptional regulators, FUSCA3 (FUS3), LEAFY COTYLEDON1 (LEC1) and LEAFY COTYLEDON2 (LEC2), control the major processes involved in embryogenesis and maturation in seeds and, likely, spores of basal plant species (Bewley *et al.*, 2012; Carbonero *et al.*, 2017; Devic and Roscoe, 2016; Han *et al.*, 2017; Santos-Mendoza *et al.*, 2008; Vicente-Carbajosa and Carbonero, 2005, and references therein).

The Arabidopsis *ABI3* gene controls the expression of at least 97 maturation-specific mRNAs (Monke *et al.*, 2012). Among the genes regulated downstream from ABI3 several belong to the eight different LEA protein families (Bedi *et al.*, 2016; Monke *et al.*, 2012). *LEA* genes encode proteins that are known to perform protective roles in the dry state due to their specific biochemical properties (Amara, 2014; Battaglia *et al.*, 2008; Shih *et al.*, 2008; Tunnacliffe *et al.*, 2010). LEA proteins have been implicated in intracellular glass formation in the dry state together with non-reducing sugars, which

enable water replacement, ion sequestration, protein stabilization and antioxidant and molecular chaperone activities (Battaglia *et al.*, 2008; Close, 1996; Hoekstra *et al.*, 2001; Tunnacliffe *et al.*, 2010; Tunnacliffe and Wise, 2007). *In vitro* studies have already confirmed many of these roles, suggesting that LEAs are essential water-stress inducible protective proteins (Hincha and Thalhammer, 2012).

During maturation the expression of *LEAs* is induced by abscisic acid (ABA), and during the first hours of imbibition transient accumulation of the different *LEA* mRNAs may occur (Espelund *et al.*, 1992; Galau *et al.*, 1987). Expression of *LEAs* may also be achieved by exogenous application of ABA and by water-related stresses (Delseny *et al.*, 2001; Espelund *et al.*, 1992; Hughes and Galau, 1991; Maia *et al.*, 2011). Seed-specific *LEAs* show different patterns of gene expression and protein accumulation during embryo development and, in some cases, proteins may accumulate weeks after the detection of their transcript upregulation, indicating that different transcriptional and posttranscriptional regulatory pathways control LEA function, and suggesting different physiological roles during the progress of seed development (Chatelain *et al.*, 2012; Galau *et al.*, 1987; Hughes and Galau, 1991; Verdier *et al.*, 2013). The storage of *LEA* mRNAs in dry seeds also suggests a translational regulation during seed germination, and raises several questions about the role of LEAs during the imbibition phase (Bai *et al.*, 2017).

In order to the investigate the phenotype of individual LEA mutants during seed development and germination, we selected five previously uncharacterized candidate genes which are expressed during the later stages of embryo development, and which expression is induced by PEG treatment in germinated seeds, correlating with re-induction of DT (Maia *et al.*, 2011). We found that three of the single LEA mutations had minor effects during seed maturation and dormancy. A mutant line with a knockout of At3g17520 (*AtLEA30*) displayed lower acquisition of DT during maturation and higher dormancy in freshly harvested seeds. *atlea30-1* seeds also show a mild longevity phenotype and lowered germination under stressful conditions. Taken together, our results suggest that LEA function is tightly regulated in seeds, and that there is high functional redundancy between LEAs. Despite that, our work demonstrates the importance of AtLEA30 for seed maturation and germination under stress. Future analysis on the role of multiple LEA knockouts may be useful to access the specific role of LEAs as well as their regulation during the different phases of maturation and post maturation life.

Results

LEA expression in seeds

To select candidates for physiological characterization in *A. thaliana* we investigated publicly available expression datasets from seed development and during re-induction of DT in germinated seeds treated with PEG (Maia *et al.*, 2011). We selected five candidate genes that were expressed at later stages of embryo development, and which expression was upregulated in the re-induction of DT in germinated seeds (Supplementary figure 1). We also investigated a co-expression network that was produced with the five LEA candidates as query (Supplementary table 1). The five LEA candidates show co-expression with 11 other *LEA* genes, suggesting a tight regulation of the expression of LEA family members. The GO enrichment analysis of the co-expression network of the candidate genes shows an enrichment for seed-related processes such as seed maturation (GO:0010431), response to water deprivation (GO:0009414), embryo development ending with seed dormancy (GO:0009793), and response to abscisic acid (GO:0009737) (Supplementary table 2). We wondered what would be the contribution of individual *LEA* genes for seed maturation, despite the indication of their strong co-regulation. For that, we analysed several available T-DNA insertional mutants. After a screening of several T-DNA lines, we were able to confirm five knock-out and knock-down lines corresponding to four of the five selected LEA candidate genes (Supplementary figure 2).

Screening for LEA proteins affecting dormancy, germinability and acquisition of DT during embryo development

For physiological characterization of the five confirmed T-DNA knock-out and knock-down lines we tested the effects of dry storage during development (30% RH, 48h) on the germination of the mutant seeds compared to the wild-type (Col-0) (Figure 1). In this analysis we also included a severe mutant allele (*abi3-6*) of the *ABI3* gene which is already known to act as a regulator of seed developmental processes such as DT and dormancy (Bedi *et al.*, 2016; Dekkers *et al.*, 2016; Delmas *et al.*, 2013; Monke *et al.*, 2012).

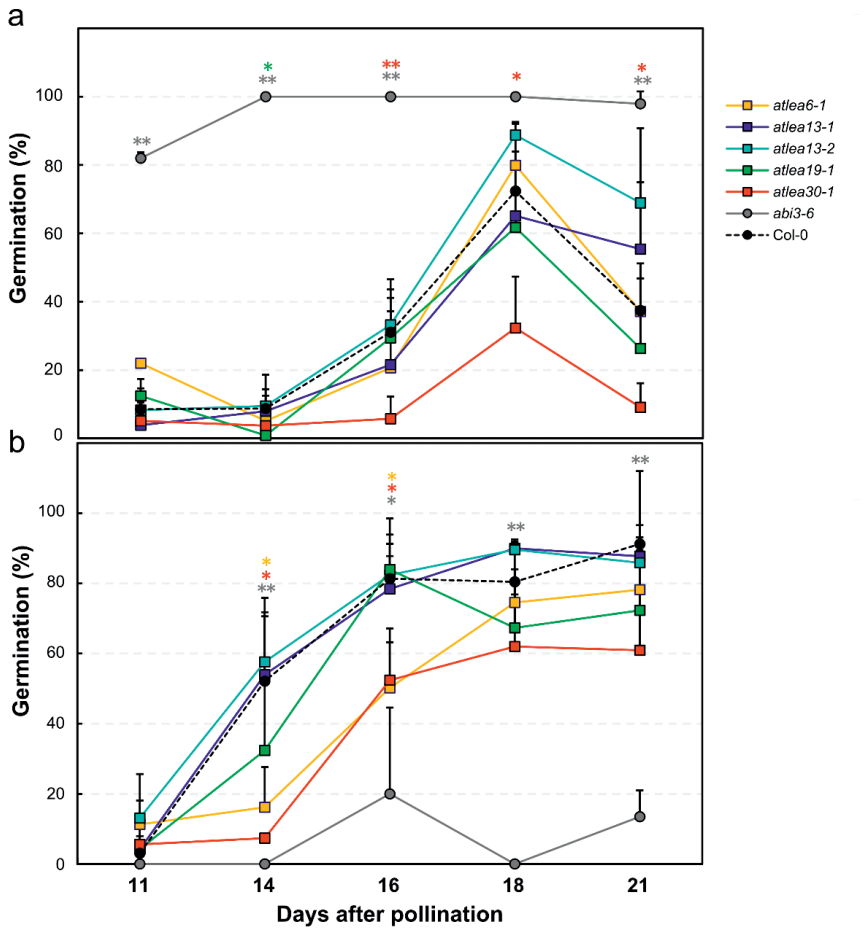


Figure 1 - Germination ability of developing *A. thaliana* seeds. Freshly harvested seeds at various days after pollination were germinated on water at 22°C (a) or were dried for 48 hours at 25°C and 30% RH and germinated in water (b). Data is the average of three biological replicates and the bars indicate standard deviation (SD). Asterisks denote significant difference by Student's *t*-test between the mutants and Col-0 seeds (* $p<0.05$, ** $p<0.01$).

The *abi3-6* seeds were fully germinable at 11 DAP onwards (Figure 1a), as previously observed for other alleles of the *ABI3* gene (Koornneef *et al.*, 1989; Ooms *et al.*, 1993a). The acquisition of full germinability of mutants deficient in or insensitive to ABA has been correlated with the absence of dormancy and precocious germination (Karssen *et al.*, 1983; Koornneef *et al.*, 1989; Ooms *et al.*, 1993a). In contrast, increasing ABA content regulates a developmental arrest during early maturation in wild-type seeds (between 11 DAP to 16 DAP in our experimental setup), as well as stimulates the synthesis of proteins involved in DT (Karssen *et al.*, 1983; Kermode, 2005; Koornneef *et al.*, 1989; Ooms *et al.*, 1993a). During late maturation a decline in ABA content is observed, but it is not sufficient to break the increasing dormancy imposition (from 18

DAP onwards in our experimental setup), revealing a genetic component in dormancy regulation (Karssen *et al.*, 1983; Koornneef *et al.*, 1989; Ooms *et al.*, 1993a). At 11 DAP there was no difference in the percentage of germination of all *AtLEA* mutants compared to Col-0 (Figure 1a). At 14 DAP only the mutant line *atlea19-1* displayed a significantly lower germination percentage compared to wild type seeds; however, the germination ability of this T-DNA line became comparable to Col-0 during further maturation. This observation indicates that *AtLEA19* may play a more specific role at an early stage of seed maturation. From 16 DAP onwards seeds of *atlea30-1* displayed lower percentage of germination when compared to Col-0, indicating that this gene may start to play a specific role in seed germinability around 16 DAP that affects ongoing maturation, or, alternatively, it may play this role until the end of the maturation period.

When analysing the dry storability potential during maturation we found that wild-type and *atlea* seeds were able to withstand, to a certain extent, the imposed dry storage, while the *abi3-6* mutant displayed lower and transient survival (Figure 1b). Interestingly, the mutants of *atlea6-1* and *atlea30-1* showed significantly lower ability to withstand dry storage as compared to Col-0 seeds only between 14 DAP and 16 DAP. The fact that *atlea30-1* showed a significantly lower ability to withstand dry storage and lower germinability as compared to Col-0 (Figure 1a), lead us to hypothesize that this gene may be also involved in seed dormancy together with the acquisition of germinability and DT in seeds, whereas *AtLEA6* may be specifically involved in the acquisition of DT during seed development. In fact, dormancy phenotypes in other LEA mutants suggested a correlation between acquisition of DT, dormancy and germinability (Costa *et al.*, 2016). It is also likely that a large functional redundancy and compensatory effect between specific subsets of *LEA* genes assures their correct functioning during seed maturation. As an example, the *atem6* LEA mutant displays defective seed maturation drying, and its function can be compensated by an earlier expression of its homolog *AtEM1* (Manfre *et al.*, 2009).

AtLEA30 is involved in post-embryonic processes

We also investigated the germination capacity of the *atlea* mutants after the completion of maturation using primary dormancy breaking treatments (Chahtane *et al.*, 2017) (Figure 2). Two-week-old seeds of Col-0 showed about 70% of germination without dormancy breaking treatments, indicating that the seeds are still partially dormant. On the other hand, the mutant lines *atlea19-1* and *atlea30-1* displayed a significantly lower percentage of germination when compared to Col-0. After dormancy breaking treatments by cold-stratification, GA_3, or KNO_3, both of these lines germinated to about 100%. *atlea30-1* also displayed higher t_{50} (time necessary to reach

50% of germination) when compared to Col-0 (Figure 2b). These results lead us to hypothesize that the role of *AtLEA19* and *AtLEA30* in the acquisition of germinability during seed maturation may also affect the imposition of primary dormancy. Mechanisms of dormancy are activated during late seed development (Ooms *et al.*, 1993a), and the t_{50} is a more sensitive parameter to discriminate differences in seed dormancy as compared to germination percentage (Hilhorst, 2011). Our findings on the effects of *AtLEA30* knock-out on the germinability of seeds in later stages of seed maturation (Figure 1a), and its lower germination percentage and longer t_{50} in seeds shortly after harvest (Figure 2a-b) suggest a hierarchy of the LEA proteins expressed during seed maturation, and that the *LEA30* gene may function at a higher order of regulation, commonly involved with dormancy, germinability and DT.

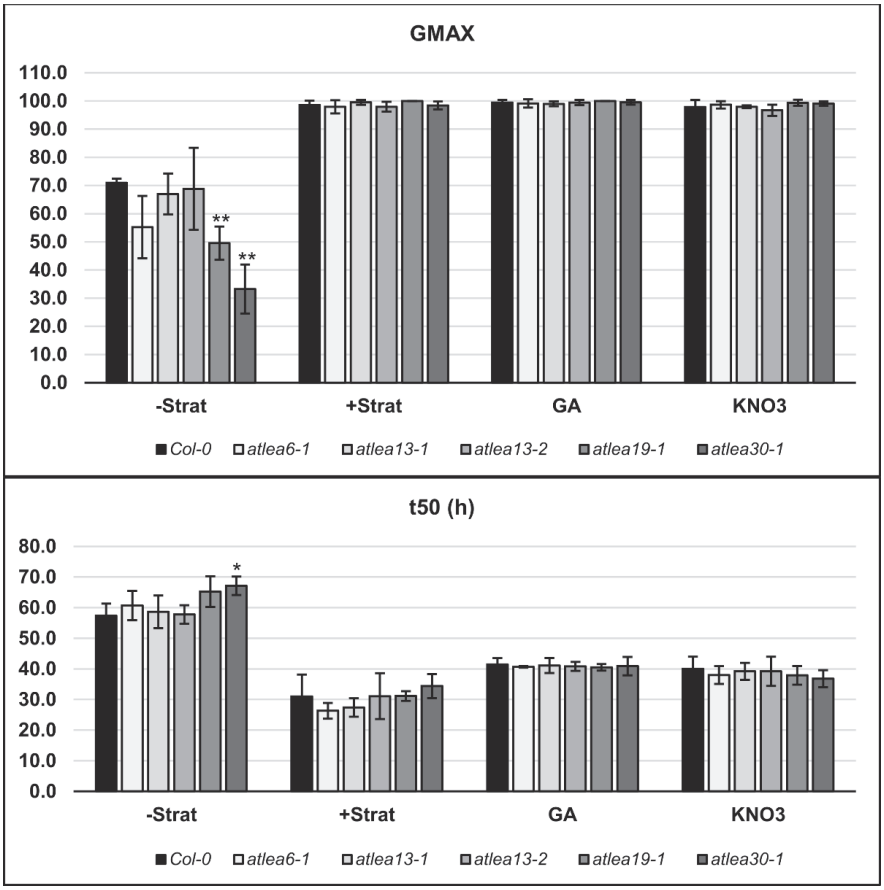


Figure 2 – Dormancy levels of *atlea* seeds. Seeds of *atleas* and Col-0 were submitted to dormancy breaking treatments by cold stratification (4°C for 48 hours), or germination in the presence of 10mM of KNO_3 or 10µM of GA_3. The experiment was performed with three biological replicates. Significant differences between the mutant lines and Col-0 within

treatments were analysed by Student's *t*-test (* $p<0.05$, ** $p<0.01$). The bars represent SD of the means.

AtLEA30 affects seed longevity and stress tolerance

Since there is a (negative) correlation between dormancy and longevity (Nguyen *et al.*, 2012), and some LEAs have been shown to be important for seed longevity (Rajjou and Debeaujon, 2008; Sugliani *et al.*, 2009), we decided to investigate longevity and accelerated aging of *atlea30-1* seeds, as this line presented a more pronounced dormancy phenotype. Longevity is the viability of seeds after dry storage (storability) and indicates seed lifespan (Rajjou and Debeaujon, 2008). Seed aging is a natural process of deterioration that can occur even under optimal storage conditions (Rajjou *et al.*, 2008). To assess *atlea30-1* longevity we conducted an Elevated Partial Pressure of Oxygen (EPPO) storage experiment (Groot *et al.*, 2012), and we performed a Controlled Deterioration Test (CDT) as a proxy for seed aging (Delouche and Baskin, 2016; Powell and Matthews, 1984). The CDT test induces a decrease in seed vigour by mimicking biochemical events that occur during the natural process of seed aging, such as protein oxidation (Bentsink *et al.*, 2000; Powell and Matthews, 1984; Rajjou *et al.*, 2008). However, some important events in seed storage, such as a decrease in tocopherol levels cannot be assessed by CDT (Groot et al., 2012). In this sense, EPPO is a more efficient and reliable method, and is also able to release primary dormancy (Buijs *et al.*, 2018).

The germination percentage of *atlea30-1* seeds under control conditions with or without stratification showed no significant differences when compared to wild-type seeds (Figure 3a). The t_{50} of seeds without stratification did not differ between *atlea30-1* and wild-type, however, after stratification the t_{50} of *atlea30-1* was significantly higher than the one of wild-type seeds (Figure 3b). It is possible that *atlea30-1* seeds are more sensitive to the low temperature treatment imposed during stratification, since several *LEA* genes have been shown to be regulated by cold (Hundertmark and Hincha, 2008). After EPPO storage, the final percentage of germination did not differ between *atlea30-1* and Col-0 seeds (Figure 3a), however, *atlea30-1* seeds showed slower germination (longer t_{50}) when compared to Col-0 for almost all time intervals (Figure 3b). Using CDT, *atlea30-1* seeds displayed a significantly lower percentage of germination after 9 days of treatment (Figure 3c), but the t_{50} did not show significant differences in any of the time intervals analysed (Figure 3d). Together, our results suggest that *AtLEA30* plays a role in seed aging and longevity.

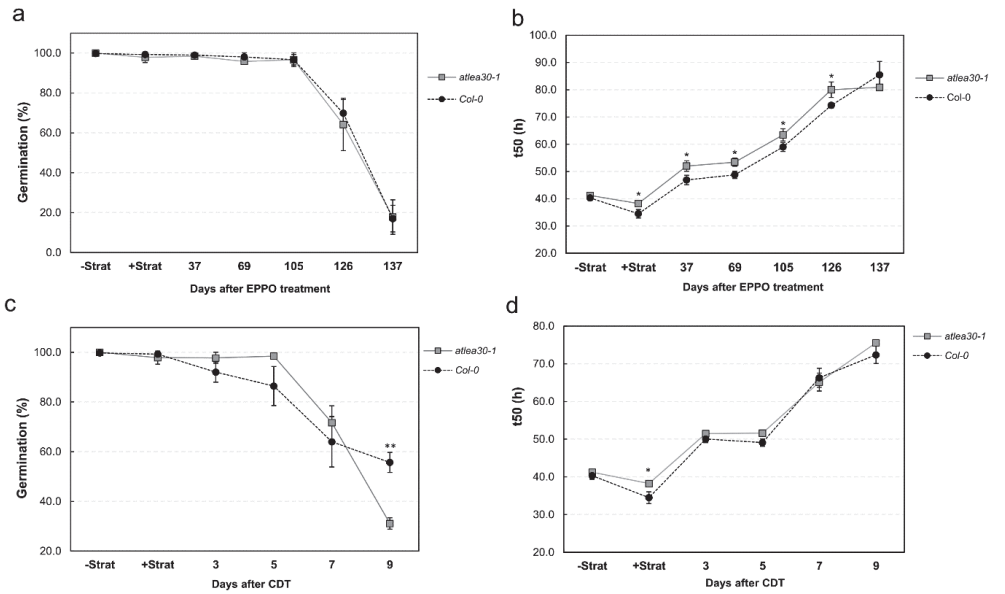


Figure 3 – Assessment of seed longevity in *atlea30-1* seeds. 16-weeks-old seeds of *atlea30-1* and Col-0 were submitted to germination with (+) or without (-) cold stratification (4°C for 48 hours), or after incubation at Elevated Partial Pressure of Oxygen (EPPO) (a,b) (Groot et al., 2012), as well as a Controlled Deterioration Test (CDT) (c,d) (Delouche and Baskin, 2016; Powell and Matthews, 1984) for different periods of time. The experiment was performed with three biological replicates and significant differences between the mutant line and Col-0 were analysed by Student's *t*-test (* $p<0.05$, ** $p<0.01$). The bars represent SD of the means.

The accumulation of LEA proteins at the end of maturation may be correlated with tolerance against environmental stresses that may occur at the beginning of seed germination (Rajjou *et al.*, 2008). Thus, we investigated the germination ability of *atlea30-1* seeds in stressful conditions, such as heat shock, salt, and ABA treatments (Figure 4).

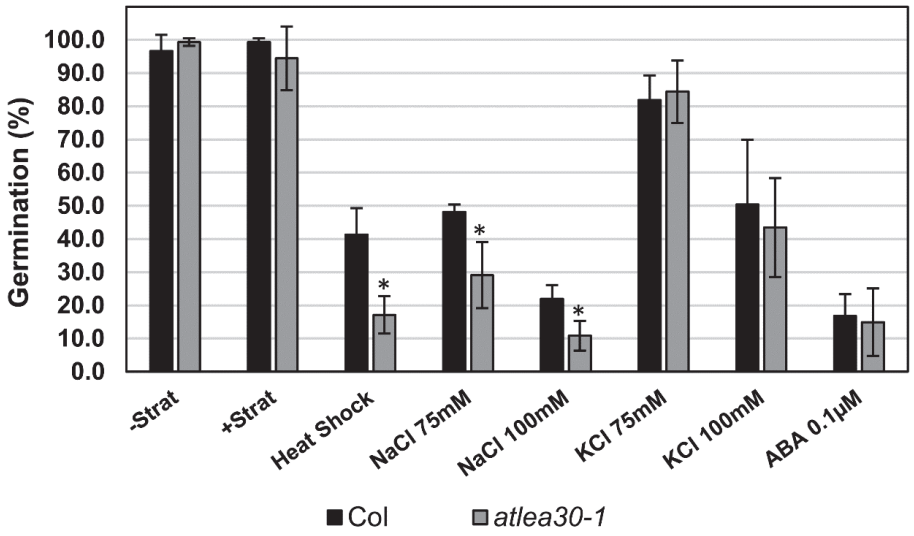


Figure 4 – Germination of *atlea30-1* seeds under stressful conditions. Heat shock was performed by incubating the seeds at 35°C for 24 hours before transferring them to 22°C. Salt stress was achieved by germinating the seeds in KCl or NaCl (75mM and 100mM). ABA treatment consisted of 0.1µM of ABA in 10mM MES, pH 5.8. Three biological replicates were used and significant differences between the mutant lines and Col-0 within treatments were analysed by Student's *t*-test (* $p<0.05$, ** $p<0.01$). The bars represent SD of the means.

Under heat shock, *atlea30-1* displayed a sensitive phenotype, with 17% germination whereas Col-0 seeds germinated for 41%. In both concentrations of NaCl (75mM and 100mM) *atlea30-1* showed a significantly lower percentage of germination, indicating that *atlea30-1* may contribute to tolerance to NaCl. In the presence of different concentrations of KCl and ABA no significant differences were observed between *atlea30-1* and Col-0. We also investigated the re-induction of DT in germinated *atlea30-1* seeds with PEG (Supplementary figure 3), but no significant differences could be observed. Taken together, these results suggest that AtLEA30 may contribute to stress tolerance in seeds during germination.

Discussion

The acquisition of longevity, dormancy, and DT occurs during seed development, with the involvement of the hormone ABA (Ooms *et al.*, 1993b). The *ABI3* gene is a master regulator of seed maturation and its mutation (*abi3*) affects the whole maturation process, including acquisition of DT, dormancy and longevity (Ooms *et al.*, 1993a). The fact that DT and longevity are acquired at different moments during seed development

implies their sequential and/or differential control (Leprince *et al.*, 2017). The interconnection between these processes indicates that the misregulation in one of the co-expression networks may have a negative impact on seed storability, longevity, and vigour (Dekkers *et al.*, 2015). The correct functioning of regulatory genes during embryogenesis and maturation phases is essential for embryo survival in the dry state for long periods of time. Some of these genes, such as LEAs, have been shown to be also expressed during germination and re-induction of DT seeds, indicating that part of the maturation-related gene network is also important for seed germination and are stress-inducible (Bai *et al.*, 2017; Maia *et al.*, 2011).

In *abi3* mutants, LEA protein accumulation has been shown to be reduced (Delahaie *et al.*, 2013). However, the unchanged accumulation of a number of LEA polypeptides raises the question if LEAs are uniquely involved in DT, and which processes would the non-changing polypeptides be involved with. The accumulation of LEA proteins concomitant with the acquisition of longevity and germinability suggests that LEAs may also be important for these processes (Leprince *et al.*, 2017). The variable expression and accumulation of LEAs in *abi3* mutants indicate that there is an intricate regulatory network of LEA proteins to assure their function (Bies-Etheve *et al.*, 2008; Delahaie *et al.*, 2013).

Together with a strong co-expression network, a high copy number of transcriptionally active *LEA* genes in the genome of ancestral plant species, as well as in resurrection plants, indicates an evolutionary constraint on functional redundancy of LEAs (Costa *et al.*, 2017a). Despite that, we hypothesized that some *LEA* genes may act upstream in the LEA network, or may have a particular role which distinguish them from other LEAs. This seems to be the case for *AtLEA30*, which mutation leads to reduced DT during maturation, as well as effects on dormancy, longevity and stress tolerance (Figures 1-4). Despite the fact that this gene is upregulated in germinated seeds treated with PEG (Maia *et al.*, 2011), it seems that the absence of *AtLEA30* expression does not affect re-induction of DT (Supplementary figure 3). *AtLEA30* has been reported to be involved in salt stress adaptation in *A. thaliana* seedlings (Huang *et al.*, 2018). Other LEAs belonging to the same family (LEA_4 family, or group 3) are known to perform multiple roles including abiotic stress tolerance such as tolerance towards salinity, freezing, and water-related stresses (reviewed by Battaglia *et al.*, (2008), Hundertmark and Hincha (2008) and Zhao *et al.* (2011)).

Only a few studies have confirmed the developmental role of LEAs using transgenic approaches (Hundertmark *et al.*, 2011; Manfre *et al.*, 2009). Our work contributes to the understanding of LEA function in seed maturation-related processes, and highlights their intrinsic functional redundancy. The precise nature and origin of the genetic and molecular mechanisms that control the expression of *LEAs* during seed

maturation and their relationships with the different developmental processes occurring in seeds remain to be elucidated. To dissect the specific contribution of LEAs in various seed traits, multiple knockout and overexpression of *LEA* genes may be suitable approaches.

Conclusion

Our study provides a better understanding of the role of LEAs during seed development and post embryonic life. We show in a small scale that there is an stablished tight regulatory network of *LEA* genes and that there is a hierarchy amongst *LEA* genes regulated during seed maturation. *AtLEA30* may be present in a higher level of LEAs hierarchy, being directly involved with the processes of maturation drying, dormancy, germination, longevity and stress tolerance.

Material and Methods

Analysis of microarray data

Genes were selected for phenotypic characterization based on the differential expression during embryo development and re-induction of desiccation tolerance. Gene expression data of *A. thaliana* ecotype Columbia during seed development were obtained from the Arabidopsis e-Northern option from BAR (Toufighi *et al.*, 2005), which includes publicly available microarray datasets such as AtGenExpress (AtGenExpress_Plus – Extended Tissue Series) (Schmid *et al.*, 2005). The absolute expression values of *LEAs* in *A. thaliana* seeds submitted to desiccation and re-induction of DT using polyethylene glycol (PEG) were obtained from GEO Series accession number GSE64227 (Maia *et al.*, 2011). A gene co-expression network was built with GENEMANIA (http://genemania.org/). Gene Onthology (GO) enrichment analysis was performed on the website of the Gene Onthology Consortium (http://geneontology.org/page/go-enrichment-analysis).

Plant material and growth conditions

Arabidopsis thaliana (L.) Heynh. accession Columbia (Col-0, N60000) was used in this study. All T-DNA insertional lines (Supplementary table 3) were obtained from the Nottingham Arabidopsis Stock Centre (NASC). The lines were named according to the identification number used by (Hundertmark and Hincha, 2008). Seeds were stratified in Petri dishes on two layers of water saturated filter paper in a cold room (4°C) for 48

hours, and subsequently transferred to 22°C and continuous light. After 24 hours, germinated seeds were transferred to Rockwool blocks (Grodan, the Netherlands) in Hyponex solution under greenhouse conditions (16h light/8h dark).

PCR genotyping and RT-qPCR expression analysis

Genomic DNA was extracted from leaves of 1 to 2-week-old plants and genotyping was performed by standard PCR using primers obtained from the T-DNA primer design of the Salk Institute Genomic Analysis Laboratory website (http://signal.salk.edu/tdnaprimers.2.html) (Supplementary table 4). Homozygous plants were double confirmed by re-extracting DNA and performing an additional PCR using the same sets of primers. RT-qPCR expression analysis was performed in dry seeds of one homozygous plant per T-DNA line. RNA was extracted using a hot borate-based protocol (Maia *et al.*, 2011). Sample preparation, RNA quality control and testing of primers were performed according to Dekkers *et al.* (2012). Approximately 700ng of RNA was reverse transcribed using the iScript cDNA synthesis kit according to the manufacturer's instructions (Bio-Rad Laboratories B.V., The Netherlands). The RT-qPCR reactions were run on a CFX machine (Bio-Rad). Three technical replicates were used per sample. The reference genes used for data normalization were At2g28390, At4g12590, and At4g34270 (Dekkers *et al.*, 2012). All primers used for RT-qPCR and the Ct (cycle threshold) values are listed in Supplementary table 5 and Supplementary table 6, respectively. After confirmation of downregulated gene expression, mature seeds from one homozygous plant per independent T-DNA line, were germinated and grown under the same conditions as described above, in a randomized complete block design with three biological replicates of at least four plants.

Assessment of desiccation tolerance during seed maturation

To obtain developing seeds at different days after pollination (DAP), at least 10 individual flowers were tagged on the day of anthesis. After 11, 14, 16, 18 and 21 DAP siliques were harvested and seeds were carefully dissected using a stereomicroscope. Part of the fresh harvested seeds were immediately submitted to germination at standard conditions (22°C and continuous light on two layers of water saturated blue filter paper (Anchor paper company, USA)). The other part of the seeds was submitted to artificial desiccation at 30% relative humidity (RH) and 25°C for 48 hours. After desiccation, seeds were germinated at standard conditions. Germination was scored daily and the results were analysed using the Germinator program to obtain the final germination percentage and time to reach 50% of the maximum germination (t_{50}) (Joosen *et al.*, 2010).

Germination assays

To assess seed dormancy, 2-week-old seeds were submitted to germination at standard conditions. Dormancy breaking treatments consisted of germination in the presence of 10mM of KNO_3, 10µM of GA_3, or stratification performed by placing the seeds at 4°C for 48 hours. Germination under stress was performed with after-ripened seeds (8-weeks-old) in the presence of 0.1µM of ABA (diluted in 10mM of MES, pH 5.8), 75mM and 100mM of NaCl or KCl, or heat shock by placing the trays with seeds at 35°C for 24 hours, and then transferring to standard conditions (22°C and continuous light).

Assessment of seed aging and longevity

To assess longevity of the T-DNA mutant seeds, after-ripened seeds (16 weeks after harvest) were submitted to artificial ageing (controlled deterioration test – CDT) (Delouche and Baskin, 2016; Powell and Matthews, 1984). Approximately 100 seeds were placed in an opened 1.5ml tube and stored above a saturated KCl solution in a closed ventilated tank at 80–85% relative humidity (RH) and 40°C for 3, 5, 7, and 9 days. RH was monitored by a Lascar data logger (Lascar Electronics). To study the effect of Elevated Partial Pressure of Oxygen (EPPO) on seed storage and longevity, about 100 after-ripened seeds were place in 2ml tubes perforated with a hole of approximately 1mm diameter closed with an oxygen-permeable polyethylene membrane (Groot *et al.*, 2012). The tubes containing the seeds were placed in 1.5L steel tanks, filled with air at a rate of 0.4MPa per minute from buffer tanks until the tank pressure reached approximately 20MPa (Buijs *et al.*, 2018). The tanks were stored at 22°C and 55% RH, and the seeds were collected after various periods. After CDT and EPPO treatments, seed germination and data analysis was performed as described above according to Joosen *et al.* (2010).

Re-induction of DT in germinated seeds

Seeds at radicle protrusion after about 28 hours of germination (stage II according to Maia *et al.* (2011)) were selected using a stereomicroscope. Seeds were submitted to a mild osmotic stress treatment achieved with -2.5MPa polyethylene glycol (PEG8000) at 22°C for 72 hours. After treatment, residual PEG was washed away with demineralised water and the seeds were immediately dried for 72 hours at 20°C and 30% RH. After drying, the seeds were transferred to 100% RH for 24 hours at 22°C in an incubator under constant light, and then rehydrated at 22°C on a Copenhagen table. Seeds that continued growth and developed into viable seedlings after 6 days of rehydration were considered desiccation tolerant. The experiment was repeated twice with three biological replicates with 20 seeds each.

Author contributions

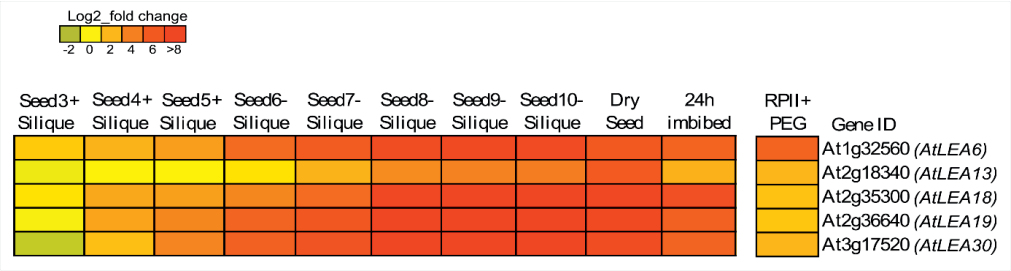
H.W.M.H. and W.L. participated in the design of the study and supervised the project. M.A.S.A and F.O. carried out the seed maturation experiments and data analysis. M.A.S.A carried out the EPPO and seed germination under stress experiments. M.A.S.A designed the experiments and H.D. carried out the re-induction of DT experiments and data analysis. M.A.S.A drafted and wrote the manuscript with significant contributions of H.W.M.H and W.L..

Acknowledgements

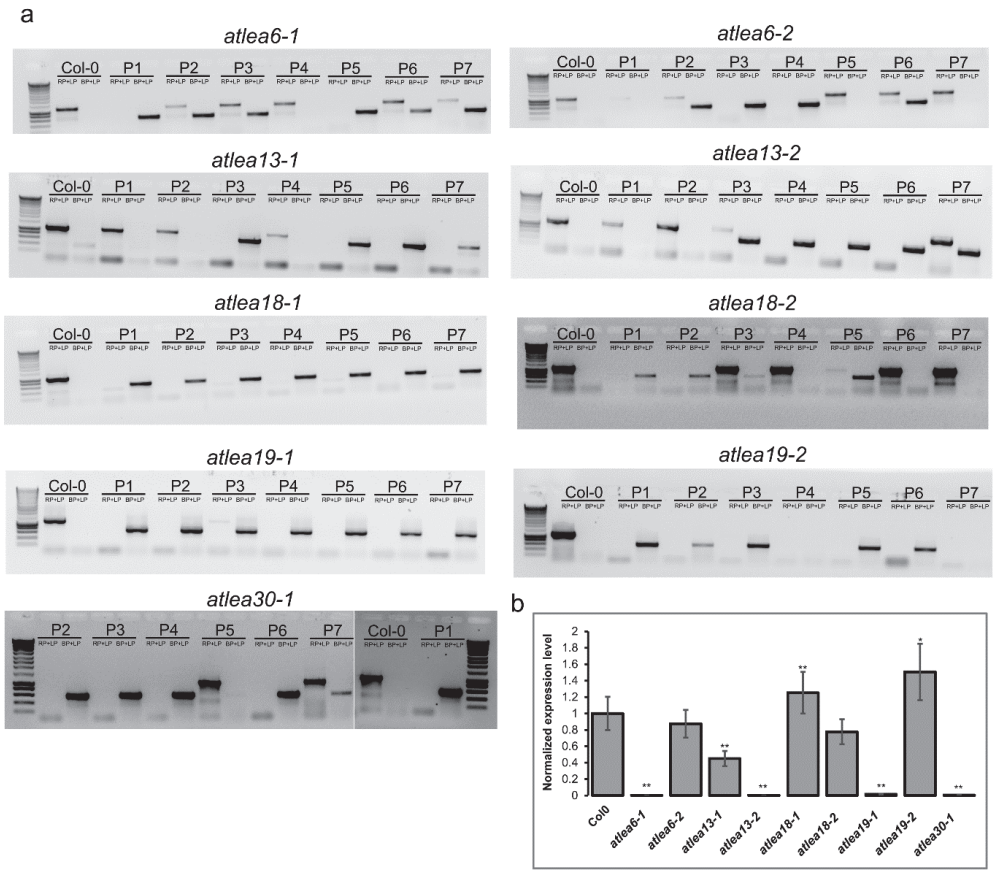
M.A.S.A received financial support from CAPES – Brazilian Federal Agency for Support and Evaluation of Graduate Education (BEX 0857/14-9). The authors thank Gonda Buijs for technical assistance with the EPPO experiments.

Supplementary Information

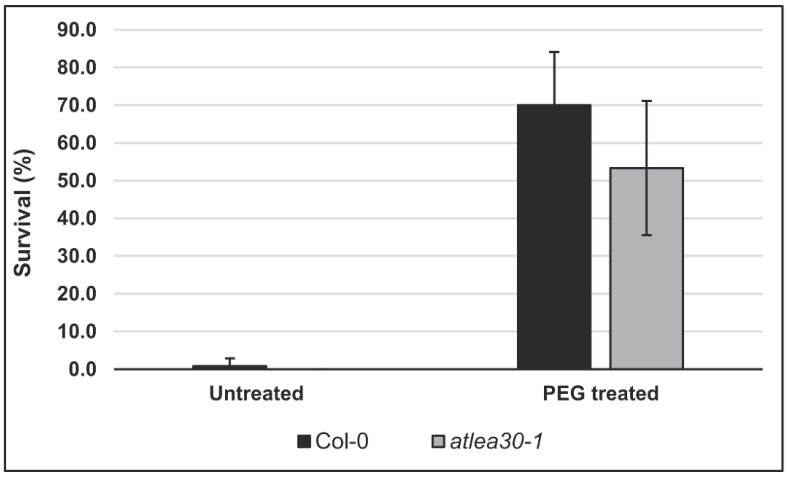
Supplementary figures



Supplementary figure 1 – Expression of *AtLEAs* in *A. thaliana*. The relative expression data of seeds from 3 to 10 days after pollination with (+) or without (-) the silique was retrieved from e-Northern (http://bar.utoronto.ca/affydb/cgi-bin/affy_db_exprss_browser_in.cgi). The expression values of *AtLEAs* in *A. thaliana* Col-0 at the stage of radicle protrusion (RPII) submitted to PEG treatment (+) compared to non-treated seeds at the same stage was retrieved from Maia et al. (2011). Log_2 fold changes in gene expression are displayed in the heat map.



Supplementary figure 2 - PCR genotyping and RT-qPCR gene expression analysis of *AtLEA* T-DNA lines. a - PCR amplification of genomic DNA of WT Col-0 and *AtLEA* T-DNA plants (seven plants per mutant: P1-7). PCR reactions were performed combining left and right, and border and right primers (LP+RP and BP+RP, respectively) (Supplementary table 4). b – mRNA transcript abundance quantified by RT-qPCR in dry seeds of Col-0 (N60000) and homozygous mutant plants. The expression levels of the T-DNA lines were determined relative to Col-0. Primers used for this analysis as well as the Ct (cycle threshold) values are shown in Supplementary tables 5 and 6, respectively. Bars are means ± SE of three technical replicates (n=3). Significant differences between the Ct values of gene expression in Col-0 and the T-DNA lines were determined by Student's *t*-test (* $p<0.05$, ** $p<0.01$).



Supplementary figure 3 - Re-induction of DT in germinated seeds of *atlea30-1*. Germinated seeds were treated with -2.5MPa PEG and posteriorly pre-humidified and rehydrated (Maia et al., 2011). Normal seedling growth 6 days after rehydration was scored. The bars represent means ± SD of three biological replicates (n=20). Student's *t*-test was used to analyse significance of differences between the lines within each treatment.

Supplementary Tables

The supplementary material of this chapter can be downloaded from http://www.wageningenseedlab.nl/thesis/masilvaartur/SI/

Chapter

6 General Discussion

On the unfolding of plant desiccation tolerance?

Water stress adaptations were crucial for the establishment of the current global plant biodiversity. The appearance of desiccation tolerance (DT) mechanisms was one of the major factors that enabled basal plant species to move from aquatic to terrestrial life and to colonize environments with lower water availability. As plant body structure and -physiology evolved and became more complex, DT mechanisms were confined to reproductive structures such as spores, (orthodox) seeds, and pollen grains. The co-option hypothesis about the evolution of DT suggests that the group of angiosperm resurrection plants were able to reactivate several of the DT mechanisms that were common to basal plants and reproductive structures in their vegetative body, which enabled their colonization of extremely harsh environments with scarce or intermittent water availability. DT involves signalling pathways for sensing the drying environment and activating responsive mechanisms. Some of these mechanisms involve protective and recovery components that are found to operate in a concerted manner at the genomic, morpho-anatomical, physiological, molecular, and biochemical levels. This network of events reveals the complex regulation of this trait.

In this thesis, I investigated genomic, physiological, molecular and biochemical aspects of DT, and focused on a major player among the protective component. In **Chapter 2** I analysed the genomic features and transcriptomic responses to desiccation and rehydration of the resurrection monocot species *Xerophyta viscosa*, and found clues that support the hypothesis of co-option of DT mechanisms from seeds by resurrection angiosperm plants. One of the clearest examples was the finding of combined gene family expansion and high expression of the so-called Late Embryogenesis Abundant protein (LEA) genes.m

LEAs were first discovered in orthodox seeds where they accumulate during the maturation-drying phase. Later, LEAs were found outside of the plant kingdom, commonly accumulating during dehydration in desiccation tolerant species of, among others, insects, nematodes, tardigrades and crustaceans. Also, *LEA* genes were found to be highly responsive to desiccation in distinct species of basal and angiosperm resurrection plants, revealing that these genes are important and universal players in DT.

Some LEAs present intriguing biochemical characteristics such as small size, high hydrophilicity, and the presence of intrinsically disordered regions (IDRs) that can acquire environment-dependent secondary structure. Due to these properties, LEAs are considered plastic and structurally dynamic proteins that play multiple roles in protecting the subcellular milieu against water stress damage. However, this protective function which depends on structural dynamics can only be assessed (to date) under

artificial conditions, by combining *in silico* and *in vitro* structural analysis of individual proteins together with *in vitro* and ectopic expression *in vivo* (e.g. using bacterial systems or heterologous/overexpression *in planta*). Thus, the functions assigned to LEAs as a family (or families) of proteins are still putative.

The major focus of three chapters of this thesis was to comprehend the evolution of LEAs in the plant lineage (**Chapter 3**), their structural dynamics *in vitro* and their function *in vivo* (**Chapter 4**), as well as their contribution to orthodox seed traits other than DT (**Chapter 5**).

In the following paragraphs, I give a brief introduction of the interesting group of resurrection plants, and then synthesize the main findings and discuss some of the insights of my thesis. First, I briefly discuss how the genome and transcriptome of the resurrection monocot *X. viscosa* can help us to understand the mechanisms of DT, aiding to previous physiological information. Second, I discuss several insights of the evolution of LEA families in plants, and provide an hypothesis of the evolutionary meaning of their distribution across multiple plant genomes. Third, I discuss about LEA protein structure and function, and how *in silico* and *in vitro* analysis can aid to functional characterization of these proteins. Finally, I point some obstacles in improving our understanding of the contribution of LEAs for plant DT, and provide what I think will be important to focus on future research in the field of DT and LEA proteins, and suggest applications of the knowledge herein generated.

Long live the resurrection plants!

Resurrection plants include a little over 130 species of ferns, fern allies, and angiosperm species which are distributed across all continents, but are restricted to environments with low water availability (xeric) were desiccation sensitive species rarely occur (Oliver *et al.*, 2000; Porembski and Barthlott, 2000). A higher proportion of ferns is able to 'resurrect' when compared to angiosperms, although angiosperm resurrection plants can tolerate more severe desiccation, likely due to the habitat where they evolved (Gaff, 1977). Despite that, the least tolerant resurrection fern is still more tolerant than most agricultural crop species (Gaff, 1977).

Resurrection plants may remain viable when desiccated from months up to at least a couple of years (Farrant and Kruger, 2001; Tuba *et al.*, 1993). Orthodox seeds produced by angiosperm species undergo maturation drying as a developmental phase, and may be stored in the dry state for up to thousands of years, producing viable plants after rehydration (Bewley *et al.*, 2012; Sallon *et al.*, 2008; Shen-Miller, 2002). Acquisition of DT and the ability to survive long periods in the dry state

(longevity) require complementary molecular players and play important roles in seed quality and storability (Leprince *et al.*, 2017; Ooms *et al.*, 1993b; Verdier *et al.*, 2013).

In fact, the most accepted hypothesis is that the DT displayed in vegetative tissues of resurrection plants is a result of co-option of DT mechanisms present in the orthodox seeds of angiosperms which, in turn, had evolved from an ancestral form of vegetative DT (Farrant and Moore, 2011; Oliver *et al.*, 2000). It is unlikely that a single DT strategy has evolved in the different resurrection species, but rather many variations of protection and repair mechanisms have been selected (Bewley, 1995). These mechanisms can be partially reactivated in desiccation sensitive seeds and seedlings via application of ABA or osmoticum (PEG) (Buitink *et al.*, 2003; Costa *et al.*, 2017a; Maia *et al.*, 2011). Comprehending how DT was established during plant evolution, and how its components allow cells to remain alive in the dry state for long periods has large potential for agronomical purposes.

How can *X. viscosa* help us to comprehend DT?

X. viscosa (Baker) is a monocotyledonous species that belongs to the Velloziaceae family, which is predominantly distributed across rocky terrains and inselbergs in exposed grasslands of tropical Africa and Madagascar (Mello-Silva *et al.*, 2011; Porembski and Barthlott, 2000). *X. viscosa* is a desiccation tolerant vascular species. Such species can tolerate loss of more than 95% of their total cellular water and are able to reassume full metabolic activity within 24 hours after rehydration (Farrant *et al.*, 2015). *X. viscosa* plants display a poikilochlorophyllous strategy during desiccation, which comprises degradation of chlorophyll and the disassembly of chloroplasts in order to avoid chlorophyll excitation and extensive formation of reactive oxygen species (ROS) during desiccation (Alpert, 2000).

Farrant *et al.* (2015) characterized three major phases in which physiological changes occur during desiccation of *X. viscosa* plants: the first phase is termed early response to drying (ERD) and consists of the decline of relative water content (RWC) down to 55%, when major photosynthetic responses are shut down. Morphologically, this phase is characterized by degreening of the leaves. The second phase consists of late responses to drying (LRD), when the RWC drops further to 10%, accompanied by a subtle leave curling and accumulation of anthocyanins in the adaxial side of the leaves. In the last phase RWC drops to below 10%, accompanied by cessation of respiration and metabolic arrest and tissues reach an air dried state (ADS). In the different stages, important mechanisms to avoid irreversible mechanical and molecular

damage, such as the ones caused by oxidative stress, have been evolutionarily selected for (Smirnoff, 1993).

At the transcriptional level, the major changes in *X. viscosa* adult plants were observed between 60% to 40% RWC (**Chapter 2, Figure 2a**). At this point, a strong repression of the expression of genes involved in energy metabolism, and increased accumulation of genes associated with chlorophyll degradation, protein folding, protection and translational control was observed. In young desiccated seedlings treated with ABA, the shutdown of energy metabolism and activation of protective mechanisms was likewise observed (**Chapter 2, Supplementary figure 3**).

The physiological and transcriptional responses of *X. viscosa* to desiccation indicate that the water content-driven coordinated deactivation of metabolism and activation of cellular protection is likely the major evolutionary characteristic that allowed the adaptation of this species to water-limiting environments.

Building blocks for successful DT

At the genome level, polyploidy events gave rise to the *X. viscosa* octoploid genome, which consists of 48 small chromosomes. Whole genome duplications (WGD) influence not only the genome size, but also the gene content, and it is one of the major sources of evolutionary novelties such as adaptation to environmental stresses (Conant and Wolfe, 2008; Panchy *et al.*, 2016). *X. viscosa* displays several genomic characteristics that correlate with its desiccation tolerant phenotype (**Chapter 2**).

Gene family analysis demonstrated that *X. viscosa* displays a very low percentage of orphan genes (which are genes not included in gene families by the lack of functional information) suggesting that DT in *X. viscosa* is a result of pre-existing genes and gene families, which gives strong support for the co-option hypothesis. Furthermore, expansion of gene families necessary for responses to light, and osmotic- and thermo- stress may have contributed to the appearance of a stress tolerant phenotype in *X. viscosa*.

Clusters of differentially expressed genes under desiccation (CoDAGs), were firstly described in an anhydrobiotic midge, and were thought to be evolutionary units that enabled the appearance of DT in these organisms (Gusev *et al.*, 2014). Based on similarity searches, we identified CoDAGs in the genome of *X. viscosa* (**Chapter 2, Figure 1**), however the majority of the genes in these CoDAGs were downregulated during desiccation, suggesting that in this species, the coordination of CoDAGs is predominantly associated with metabolic arrest.

Another remarkable genomic characteristic of *X. viscosa* is the expansion of LEA protein families as compared to other sequenced monocot genomes. In general, LEAs are highly active during desiccation in several resurrection plants (VanBuren *et al.*, 2017;

Xiao *et al.*, 2015; Xu *et al.*, 2018). LEA proteins accumulate in orthodox seeds mainly during late maturation (reviewed by Leprince *et al.* (2017)), and were found to be important for completion of the maturation program, as well as for acquisition of seed traits such as DT and longevity (Chatelain *et al.*, 2012; Hundertmark *et al.*, 2011; Manfre *et al.*, 2006; Manfre and Simon, 2008, **Chapter 5**). Thus, the coordination of upregulation of *LEAs* in *X. viscosa* and other resurrection species during desiccation also supports the hypothesis of co-option of DT mechanisms from orthodox seeds by resurrection plants (**Chapter 2**, **Figure 3**) (Farrant and Moore, 2011; Oliver *et al.*, 2000).

From an evolutionary genomic perspective, one of the most important implications of the co-option hypothesis refers to the impact of duplications on gene neofunctionalization, subfunctionalization, and dosage variation, as well as on protein biochemistry, which can change the biology of an organism (reviewed by Conant and Wolfe (2008)). In the case of duplicated copies of LEAs, we hypothesize that their contribution to the DT phenotype in *X. viscosa* may include altered protein biochemistry, likely by different abilities to undergo desiccation-induced protein folding and binding, which may enable an orchestrated protection of the cytosol and its components during the decrease in RWC.

Another important outcome of gene duplication is differential mutational rates in cis-regulatory elements (Arsovski *et al.*, 2015), which may lead to differential regulation of gene expression (temporal and spatial) between the duplicated copies. LEA proteins in seeds display variable expression profiles as well as protein accumulation during the maturation phase, indicating transcriptional and post-transcriptional regulation (Chatelain *et al.*, 2012; Espelund *et al.*, 1992; Galau *et al.*, 1987; Hughes and Galau, 1991; Verdier *et al.*, 2013). In *X. viscosa* plants under desiccation, about 71% of the *LEA* genes change in expression, and the majority of them (~43%) is upregulated between 60% to 40% RWC (1.5-1.0 $gH_2O\ g^{-1}$ dry weight) (**Chapter 2**). Despite that, genes belonging to the same LEA family display variable temporal activation and accumulation, and some of these genes are even downregulated upon desiccation (**Chapter 2, Figure 3a-b**). *In silico* analysis revealed that the different LEA families vary greatly in their upstream genomic regions, where distinct sets of transcription factors (TFs) can initiate gene expression (**Chapter 2, Supplementary material online**). It is likely that during the genomic evolution of *X. viscosa*, changes in cis-regulatory regions, including promoter binding sites, contributed to temporal variations of *LEA* gene expression, favouring the retention of duplicated LEA copies and their contribution to DT, although more empirical analysis is still needed in order to confirm this *in silico* observation.

LEAs here, LEAs there, LEA proteins everywhere!

Due to their importance for the acquisition of DT in seeds, understanding the evolution and distribution of *LEA* gene families in plant genomes can potentially aid to the selection of candidates for improving DT and other traits in seeds.

A gene family is defined as a group of genes that descended from a common ancestor gene, or homologs, and that maintained similar sequence and often similar functions during evolution (Dayhoff, 1976). Gene families include corresponding genes in different species, and genes duplicated within the same genome, also named orthologs and paralogs, respectively (Gabaldon and Koonin, 2013; Koonin, 2005). The emergence, expansion, contraction and disappearance of gene families, as well as copy number variations and per nucleotide polymorphisms, play significant roles in organismal adaptation (Demuth and Hahn, 2009). With the fast development of sequencing techniques, and availability of a number of whole-genome sequences, comparative genomic analysis has the power to provide insights into gene family evolution and its contribution to adaptation and diversity.

During the past years, genomic investigation of several plant species has shown that members of the different LEA families are widely found across evolutionary distant plant species, from monocots to eudicots (Altunoglu *et al.*, 2017; Battaglia and Covarrubias, 2013; Bies-Etheve *et al.*, 2008; Cao and Li, 2015; Charfeddine *et al.*, 2015; Du *et al.*, 2013; Hundertmark and Hincha, 2008; Lan *et al.*, 2013; Li and Cao, 2016; Magwanga *et al.*, 2018; Pedrosa *et al.*, 2015; Wang *et al.*, 2007, **Chapter 3**). It seems intuitive to speculate that, if LEA proteins are important for DT in orthodox seeds and in resurrection plants, they should be, somehow, overrepresented in the genomes of these species, and absent or in lower number in the genomes of desiccation sensitive species or species with a water-dependent lifestyle. The first assumption could be considered partially true if we analyse the distribution of LEAs in basal species (**Chapter 3, Figure 1**). Only two LEA families (LEA_5 and SMP) are present in aquatic green algae genomes (the ones investigated in this thesis), while the appearance of new LEA families and gene copy-number expansion is readily observed in the bryophyte and lycophyte clades, which may indicate that LEA family evolution was, among other factors, involved in the adaptation to land life (Rensing *et al.*, 2008). Also, the expansion of LEAs in *X. viscosa* reported in **Chapter 2**, and contraction and even disappearance of some families in aquatic (*Spirodela polyrhyza*) and marine (*Zostera marina*) (Olsen *et al.*, 2016) monocot species, could be an indication that these specific LEA families are, to a certain extent, related to DT.

However, if we analyse the genomes of species that do not present aquatic or marine lifestyles, but produce desiccation sensitive seeds (recalcitrant seeds), as well as intermediate seed phenotypes (as in intermediate seeds) (Ellis *et al.*, 1991), we find that these speculations are no longer valid. Members of all LEA families can be found in the genomes of species that produce recalcitrant or intermediate seeds such as *Theobroma cacao*, *Citrus sinensis*, *Carica papaya,* and *Elaeis guineensis*, and no apparent LEA family copy-number contraction is clearly evident or has been reported previously (**Chapter 3, Figure 1**).

Interestingly, LEAs were found, in general, in larger numbers in the genomes of the resurrection plants *S. lepidophylla*, *S. tamarisciana*, and *X. viscosa* compared with closely-related species (Costa *et al.*, 2017a; VanBuren *et al.*, 2018; Xu *et al.*, 2018), and to be highly expressed under desiccation conditions in the majority of the resurrection species for which genomes and transcriptomes are analysed (**Chapter 1, Table 1**). This indicates that the composition and functioning of distinct sets of LEAs may be one of the major protective features that were evolutionary selected in resurrection plants.

At this point, it is important to highlight that most analyses of gene families in different genomes involve clustering methods based on sequence similarity, which may vary greatly depending on the tools used and the thresholds applied. Consequently, comparative analysis may be robust within a study, but caution should be taken when comparing different studies (Demuth and Hahn, 2009). Despite these limitations, sequence similarity implies structural and likely functional similarities, thus, comparative analyses are powerful to demonstrate the dynamics of gene families, as well as functional annotation of unknown genes (Eisen, 1998).

But how could LEA gene families be a footprint of DT in resurrection plants, if they are found everywhere, from the beginning of the plant lineage in aquatic green algae, to recalcitrant angiosperm species? More likely than copy-number variation and presence or absence of entire gene families, the duplication and diversification of specific sets of genes, including *LEAs*, played an important role in the evolution of DT.

The diversification of genes within a gene family may occur via a number of events such as whole-genome duplications (WGD), segmental duplications, tandem duplications and transpositions (Freeling, 2009). All these mechanisms accelerate gene family evolution and provide material for evolutionary selection of variable duplicated copies. Synteny analysis allows the detection of the major evolutionary mechanisms that shaped the evolution of a gene family by identifying homologous genes in corresponding chromosomes of different species (Zhao *et al.*, 2017; Zhao and Schranz, 2017). The incorporation of network approaches in synteny analysis may facilitate the understanding of the evolution of duplicated gene copies in their genomic context, and ultimately, gene family evolution (Zhao and Schranz, 2017).

Large-scale synteny network analysis in 60 genomes has demonstrated that LEA families have distinct evolutionary histories, although, in general, they are largely conserved in angiosperm genomes (**Chapter 3, Table 1**). Important evolutionary mechanisms such as ancestral genomic duplications and recurrent tandem duplications contributed to the acquisition of new genomic contexts by *LEA* duplicates, resulting in gene family expansion and diversification. This implies that neofunctionalization or subfunctionalization within LEA families may have occurred, likely involving changes in protein sequences and/or gene expression profiles between the duplicated copies (Arsovski *et al.*, 2015; Conant and Wolfe, 2008). As an example, the Dehydrin family has diversified into two major syntenic groups before the origin of monocots and eudicots (**Chapter 3**). As a consequence, these two groups display distinct protein sequences with different biochemical features and motif composition. The finding of Dehydrin proteins performing distinct roles in *P. patens* indicates that functional diversification may have occurred before the origin of angiosperms (Agarwal *et al.*, 2017; Ruibal *et al.*, 2012).

Diversification between LEA proteins belonging to the same family was also observed in their subcellular localization (Avelange-Macherel *et al.*, 2018; Candat *et al.*, 2014). Protein relocalization is another mechanism by which the duplicated copies may diverge, and it is often accompanied by protein sequence and gene expression pattern alterations (Liu *et al.*, 2014). Our data revealed that the two distinct Dehydrin groups found presented diversified expression profiles in *A. thaliana*, with one group predominantly expressed during seed development and water stress, and the other expressed upon water stress only (**Chapter 3, Figure 3**). Future comparisons between our findings and experimental information of subcellular localization of LEAs, such as the work performed by Candat *et al.* (2014), may shed light onto the extent of diversification of LEA proteins located in distinct syntenic genomic regions.

Thus, bearing in mind the highly distinct evolutionary origin, biochemistry, genomic context, expression patterns, functional motifs, and subcellular localization between and within LEA families, it is likely that the combination of these factors giving rise to multiple responsiveness and functionalities of LEAs is the actual footprint of DT in which these genes take part.

LEAs stability and disorder: two sides of the same coin?

As mentioned before, it is likely that several features of LEA protein families were evolutionarily selected and contributed to the establishment of DT in plant species. One of the most remarkable characteristics of these proteins is their secondary structure or, in fact, the lack of it. Despite the lack of sequence similarities between the different LEA families, in general, LEA proteins present an overriding hydrophilicity,

high Glycine content, and low content or lack of Cysteine and Tryptophan residues (Dure *et al.*, 1989a; Garay-Arroyo *et al.*, 2000). These characteristics led to the suggestion that LEAs were mainly randomly coiled in solution, being regarded as intrinsically disordered proteins (IDPs) (Cuevas-Velazquez *et al.*, 2017; Dure *et al.*, 1989a; Goyal *et al.*, 2005c; Shih *et al.*, 2008; Soulages *et al.*, 2002; Tompa, 2002; Uversky *et al.*, 2000).

LEAs are considered as the plant protein group with the largest number of IDPs known (Sun *et al.*, 2013), since several *in silico* and *in vitro* studies have identified their lack of secondary structure in solution, confirming the disordered characteristic (Cuevas-Velazquez *et al.*, 2016; Hundertmark *et al.*, 2012; Mouillon *et al.*, 2006; Popova *et al.*, 2011; Shih *et al.*, 2012; Shih *et al.*, 2010). IDP-encoding genes were also recently found in multiple desiccation tolerant tardigrade species, induced during dehydration (Boothby *et al.*, 2017).

The fact that the IDPs accumulating in tardigrades during desiccation were able to enhance DT when expressed in heterologous systems (Boothby *et al.*, 2017) led us to investigate if a similar response was also found in *X. viscosa* leaves under desiccation. To test this, we analysed the structure of six distinct LEA proteins encoded by genes upregulated in *X. viscosa* leaves during desiccation (between 60% and 40% RWC) (**Chapter 4**). All six proteins displayed predominantly disordered structures in solution, although some level of local secondary structure was also identified. This finding suggests that LEA proteins presenting IDP characteristics may also be mediators of the DT response in *X. viscosa* leaves.

Next, we tested if the isolated IDPs of *X. viscosa* were able to undergo conformational changes in a hydrophobic environment. A characteristic of several IDPs is their propensity to undergo disorder-to-order transition *in vitro* under drying and solute perturbations, which correlates with their ability to bind and protect other molecules and membranes (Bremer *et al.*, 2017; Cuevas-Velazquez *et al.*, 2016; Furuki *et al.*, 2011; Hundertmark *et al.*, 2012; Mouillon *et al.*, 2006; Popova *et al.*, 2011; Rivera-Najera *et al.*, 2014; Shih *et al.*, 2010; Soulages *et al.*, 2002). We found that five of the six proteins studied *in vitro* displayed a substantial ability to undergo conformational changes and acquire higher levels of secondary structure upon solute perturbation (**Chapter 4**).

Interestingly, the protein XvLEA1-8 showed a highly disordered character and outstanding conformational plasticity, adopting 100% alpha-helical structure when added to a denaturing solution of acetonitrile (ACN). Concurrently, this protein displayed a highly protective function with respect to the activity of the enzyme lactate dehydrogenase (LDH) against desiccation-, heat-, and oxidation damage *in vitro*, as well as enhanced heat, salt and osmotic stress survival when heterologous expressed

in *Escherichia coli* cells and in *Arabidopsis thaliana* plants. Contrarily, XvLEA4-12 revealed a higher degree of secondary structure in water, and displayed an unsubstantial change in conformation when incubated in ACN. Furthermore, this protein did not contribute substantially to the protection of LDH upon stress, and made only a small contribution to the survival and growth of *E. coli* cells under stressful conditions.

It is not surprising that some LEA proteins display fewer IDRs or lower conformational plasticity under stressful conditions. In fact, an equilibrium between disordered and ordered forms can be found under different stressful conditions (Cuevas-Velazquez *et al.*, 2016; Kovacs *et al.*, 2008; Mouillon *et al.*, 2008; Soulages *et al.*, 2003). Also, a group of LEAs remains highly stable in their disordered state under conditions of macromolecular crowding and still are able to perform catalytic and chaperone-like activities suggesting that maintenance of the disorder may be required for their functioning (Chakrabortee *et al.*, 2010; Kovacs *et al.*, 2008; Mouillon *et al.*, 2008; Mouillon *et al.*, 2006; Soulages *et al.*, 2003). The impact of this structural versatility is reflected in their ability to perform multiple functions (Agarwal *et al.*, 2017; Alsheikh *et al.*, 2003; French-Pacheco *et al.*, 2018; Furuki and Sakurai, 2016; Goyal *et al.*, 2005a; Halder *et al.*, 2017; Hara *et al.*, 2005; Hoekstra *et al.*, 2001; Kovacs *et al.*, 2008; Liu *et al.*, 2016b; Nakayama *et al.*, 2007; Tolleter *et al.*, 2010; Tolleter *et al.*, 2007; Tompa *et al.*, 2006).

Taken together, it appears that desiccation commonly stimulates the expression of genes encoding a multitude of LEAs with distinct stability, disorder, and folding abilities which, in turn, may be able to perform variable protective functions contributing to survival in the dry state.

LEAs as an IDP (-sub)class?

The high variability in biochemistry, structure, function, and expression patterns, as well as distinct evolutionary trajectories, makes it a difficult task to define and classify LEAs. Most of their classification is based on sequence similarity, protein domains, -motif and –composition (Battaglia *et al.*, 2008; Bies-Etheve *et al.*, 2008; Dure, 1993; Dure *et al.*, 1989b; Hughes and Galau, 1989; Hunault and Jaspard, 2010; Hundertmark and Hincha, 2008; Tunnacliffe and Wise, 2007; Wise, 2003; Wise and Tunnacliffe, 2004). All of the proposed classifications make sense in the context in which they were created and applied; however, it is a great challenge to combine all existent LEA classes and groups in a consensus arrangement because they were defined based on distinct assumptions and considering distinct protein features.

Furthermore, if we consider that LEAs are IDPs or have IDRs, and that IDPs and IDRs may evolve more rapidly than proteins with defined structures due to the lack of

structural constraints (Brown *et al.*, 2011), the classification of LEA proteins based uniquely on amino acid sequence becomes less reliable. Also, most of the sequence clustering and annotation methods do not consider a large pool of segments with low sequence similarity, such as IDRs, which complicates the transfer of knowledge between homologues (van der Lee *et al.*, 2014). Thus, it would be reasonable to define a new classification of the LEA proteins which are IDPs or contain IDRs.

A systematic classification that incorporates biochemical, structural, and disorder information as classification steps could lead to a better definition of classes of IDP-LEAs. Also, the inclusion of evolutionary and synteny information could aid to the understanding of gene and protein structure as well as expression diversification which, in turn, could also aid to a better functional characterization of genes across evolutionarily distant species, and assignment of function for uncharacterized proteins, and facilitating the selection of candidate genes for future research and applications.

Despite the potential of such an approach, a hierarchical method for systematic classification that can implement so many criteria is still not fully developed (Koonin *et al.*, 2002; Orengo *et al.*, 1997; van der Lee *et al.*, 2014). While awaiting for that, LEAs will continue to be a group of 'odd ones out' in the protein classification system.

Multilevel approaches for knowledge transfer into applications

In this thesis a multilevel approach combining different research topics was employed in order to tackle curiosity-driven fundamental questions about the mechanisms governing DT in resurrection plants and seeds.

The knowledge herein generated may also contribute to the exchange between fundamental research and practical applications. In the next sections, examples of applications of this knowledge are proposed and discussed.

Comparative genomic approaches to investigate water stress adaptation in plants

With the fast development of sequencing techniques in the last decades, several projects have been developed and executed in order to sequence and assemble hundreds of plant genomes and transcriptomes. For example, the 1KP project launched in 2012 focused on providing sequence information of a thousand species, including crops and model plants (https://sites.google.com/a/ualberta.ca/onekp/). Between 2017 and 2022, a new project named 10KP (https://db.cngb.org/10kp/) will build on the 1KP project, and will include the sequencing of genomes of microbes and plants, comprising a broad spectrum of phylogenetic diversity (Van de Peer and Pires, 2018). Also, several other smaller initiatives have been developed, such as the 101 African orphan crops project, that aims to improve the nutritional value of 101 traditional

African food crops by assembling and annotating their genomes (http://africanorphancrops.org/), which provides essential tools to plant breeders. Thus, the development of new approaches to keep track of the large number of genomic and transcriptomic data becoming available may facilitate the understanding of phenotypic variation and improvement of plant performance.

In this scenario, comparative genomics (CG) has great potential in facilitating a more thorough study of the evolution, the source of genetic diversity, and the origin of phenotypical innovations in plant genomes. Typically used for gene and protein functional annotation, CG approaches have also potential for conservation of molecular biodiversity, for example by providing information about interaction between species and their threatening processes (biotic and abiotic) (Grueber, 2015; Pellegrini *et al.*, 1999). The application of CG also allows the determination and investigation of genes and genomic regions under environmental selection, providing important hypotheses for the origin of phenotypic variations.

For instance, the major molecular adaptations for plant colonization on land were inferred by comparing the genomes of aquatic and terrestrial species (Rensing *et al.*, 2008). Furthermore, CG also extends through large scale gene family analysis to understand functional diversification. For instance, in the study presented in **Chapter 3**, a large scale comparative genome analysis of LEA protein families in 60 genomes revealed important evolutionary patterns leading to synteny and protein diversification responsible for functional innovations related to water stress adaptation.

Extremophiles constitute the most suitable models in order to discover physiological and molecular mechanisms underlying the evolution of stress adaptation in plants to extreme environments using comparative investigations (Kramer, 2018). In the case of DT, several resurrection plants have been under focus of comparative analysis of transcriptomes, metabolomes and proteomes, which provided important clues about the major mechanisms underlying DT (Cushman and Oliver, 2011; Giarola *et al.*, 2017). Also, CG allowed us to investigate which are the major players of DT in the angiosperm resurrection species *X. viscosa* and to provide support to the hypothesis that these mechanisms were co-opted from orthodox seeds (**Chapter 2**).

To our knowledge, currently there are only six DT genomes fully sequenced of phylogenetically distant species: one bryophyte, two lycopodiophyta, two monocots, and one eudicot (**Table 1, Chapter 1**). The large evolutionary distance between several DT species is explained by environmental constraints, since many of the resurrection plants colonize dry rock outcrops, which favoured the appearance of DT mechanisms in non-related species (Porembski and Barthlott, 2000). Thus, caution should be taken when investigating the evolution of DT in resurrection plants because comparison between phylogenetically distant species could lead to loss of genomic information

(Zhao *et al.*, 2017), raising the urgency for more resurrection plant genomes become available.

Despite the need of more sequenced resurrection genomes in order to investigate the evolution of DT by CG analysis, some issues such as their scaffolded phylogenetic distribution (Oliver *et al.*, 2000), slow growth and unpredictable flowering times (Alpert, 2005), and the lack of molecular techniques for transformation and cultivation *in vitro* makes it problematic to test several of the hypotheses provided by CG. One way to solve these issues is by developing more non-seed model species (Rensing, 2017) such as some ferns and mosses that display DT. This approach would be extremely beneficial to comprehend which of the mechanisms acting in angiosperm resurrection plants are ancestral and derived, and how they diversified through the course of plant evolution and adaptation to climate variations.

Another potential approach consists of the comparison between species belonging to the same clade but displaying distinct stress tolerance phenotypes, the so-called sister lineage approach. Such an approach has revealed that desiccation-related genes in the anhydrobiotic insect *Polypedilum vanderplanky* have undergone massive expansions within gene clusters (Gusev *et al.*, 2014), and that horizontal gene transfer may play a role in DT of bdelloid rotifers (Nowell *et al.*, 2018).

In resurrection plants the sister lineage approach is commonly applied in order to compare the physiology of tolerant and sensitive species. For example, Ginbot and Farrant (2011) have pointed at several physiological adjustments in the resurrection plant *Eragrostis nindensis* which were misregulated in the related crop variety *Eragrostis tef*. At the genomic level, this approach has demonstrated several species-specific adaptations responsible for the distinct tolerant and sensitive phenotypes of two closely-related moss species (Xu *et al.*, 2018).

The use of CG by sister lineage approaches using extremophile species, such as resurrection plants, can help to unravel the genetic basis of several stress tolerance traits. This approach may also facilitate the discovery of new genes and pathways to be prioritized in genetic programs aiming to improve crops towards tolerance and productivity under stress.

Stabilization of matters: why does it matter?

The traditional concept that protein function critically depends on a well-defined 3D structure has changed with the finding that for many proteins and protein domains the functional state is in an unstructured state, resembling those of ordered proteins when denatured (Dunker *et al.*, 2001; Tompa, 2005; Uversky, 2002).

Advances in the field of biophysical and structural biology have favoured the development of several techniques, such as X-ray crystallography, Nuclear Magnetic

Resonance (NMR), and Circular-Dichroism (CD) analysis, which provide site-specific and global information about structural distinction between proteins and protein states (Piovesan *et al.*, 2017; Uversky, 2002). Nowadays, more than 800 entries of structures of IDPs (also termed intrinsically unstructured protein, IUPs) and IDRs can be found in the major database collection DisProt (Piovesan *et al.*, 2017).

A large number of studies support that IDPs also lack structure *in vivo*, and that they display structurally adaptive functions (Tompa, 2005). These functions are invariably dependent on their structural disorder, and the major functional modes can be classified into five major categories: entropic chains, effectors, scavengers, assemblers and display sites (Tompa, 2002). These functions are commonly found in proteins able to bind DNA, RNA, other proteins and smaller ligands (Tompa, 2002; Tompa, 2005). As mentioned before, several studies have demonstrated that LEA proteins are IDPs or contain IDRs that can undergo disorder-to-order transitions in the cell. Functions associated with this property have been demonstrated *in vitro*, and include ion binding, glass formation, and stabilization and protection of membranes and enzymes against degradation (described in **Chapter 1**).

Engineering protective proteins such as LEAs with IDP properties able to stabilize biological and non-biological materials has a tremendous socio-economic potential. For example, the storage conditions of protein and peptide pharmaceuticals is critical for their stability. The common causes of instability include oxidation, denaturation and aggregation caused by intense exposure to high and low temperatures, ultraviolet light, metal ions, free radicals, shaking, and loss of the water shell (Swain *et al.*, 2013). Often, glycerol and polyethylene glycol are employed to stabilize pharmaceuticals under cryo-storage; however, PEG does not stabilize biological materials during drying (Swain *et al.*, 2013).

Another interesting use of LEA proteins would be as stabilizers of nanoparticles, since there is a need of 'greener' strategies that use biocompatible and biodegradable stabilizing agents for metal nanoparticles for applications in biomedicine, material science, electronics, biosensors, cosmetics, food industry, and environmental remediation (Varma, 2012; Virkutyte and Varma, 2011). Interestingly, LEA proteins have been shown to be able to perform metal binding *in vitro* (French-Pacheco *et al.*, 2018; Hara *et al.*, 2005; Kruger *et al.*, 2002), which gives even more potential for their industrial application.

Taking XvLEA1-8 as an example, the heat stability, highly disordered nature, and ability to acquire complete alpha-helical structure in hydrophobic solution correlates with the preservation of enzyme activity upon desiccation, and heat-, and oxidative stress (**Chapter 4**). Engineered proteins displaying these biochemical properties could be applied for maintaining the integrity of biological structures such as DNA, RNA,

peptides, and pharmaceuticals that should be stored and delivered in a dry or frozen state. These properties might also be useful for stabilizing sensitive and harmful non-biological materials such as metal ions and chemical compounds, although further experimentation is still needed.

Taking this socio-economical potential into account, it is not surprising that LEA proteins and LEA functions have already been patented. A patent of group 1 LEA proteins (PFAM LEA_5) was deposited due to its potential for pharmaceutic-, cosmetic- and food stabilization, as well as medical uses (Cushman and Walters, 2004). Also, a patent on a method to freeze and or/dry mammalian cells for storage by expressing plant LEA proteins has been deposited (Conrad and Allen-Hoffmann, 2005).

In summary, LEA proteins have a great potential as stabilizers of biological and non-biological matters. The application of LEAs could offer a new sustainable strategy in the fields of, among others, pharmaceuticals and nanotechnology.

DT and resurrection plants : Preventing the predictable in crops

Climate change is leaving profound consequences on our planet, rising of global temperatures leading to increased sea levels, warming of the ocean, and longer heat and drought periods threatening fresh water supply and crop productivity which, consequently, impacts food security (Lindsey and Dahlman, 2018; Ma *et al.*, 2018). In the same scenario, by the year 2030, the world population is projected to grow to 8.3 billion people and the demand for food will also increase (FAO, 2017). Thus, prioritization for investments in making food systems more efficient and inclusive are currently under discussion (FAO, 2017).

One of the major objectives of breeding crops under the current climatic conditions is to develop plants of which growth and development are only lowly affected by stresses. Despite the development of traditional crop breeding techniques, some of the practices may result in intensification of land use by opening new arable lands which, in turn, has an enormous environmental cost including loss of biodiversity and increased diversification (Fischer and Connor, 2018; Ma *et al.*, 2018). Also, the cost of adapting crops to local conditions is threatening food security, especially in emerging economies (Ma *et al.*, 2018).

The improvement of biotechnological tools is promising in changing the biology of plants towards better performance under stress, offering new opportunities to improve crop performance under drought conditions (Nuccio *et al.*, 2018). However, some breeding technologies such as genetic modification via incorporation of DNA elements in a host genome to introduce new traits and induce desirable phenotypes have often generated a negative reaction from the general public (Ma *et al.*, 2018).

In this scenario, innovations in genome editing technologies can offer a powerful and elegant way to improve and accelerate the breeding of improved crops. These tools include sequence-specific nucleases (SSNs), transcription activator-like effector nucleases (TALENs), and the clustered regularly interspaced palindromic repeats (CRISPR)-associated system (Cas), which were found in nature (Jaganathan *et al.*, 2018; Ma *et al.*, 2018). CRISPR/Cas9-based genome editing has been utilized in important crop species such as rice, wheat, maize, cotton, soybean, tomato, potato, grape and citrus, in order to increase disease resistance and improve tolerance to abiotic stresses such as drought and salinity (Jaganathan *et al.*, 2018).

Knowledge of the major modulators of stress tolerance in extremophiles, for example DT in resurrection plants, could be a stepping stone for the selection of candidate genes for genome editing technologies towards improved stress tolerance in crops. In this context, LEA proteins would be suited as good candidates, since several studies have shown that the overexpression of single *LEA* genes may enhance tolerance to stresses in crop species (Babu, 2004; Chen *et al.*, 2015; Fu *et al.*, 2007; Hu *et al.*, 2016; Xiao *et al.*, 2007; Yu *et al.*, 2016). However, we have shown that single *LEA* genes have low impact in promoting seed DT and longevity (**Chapter 5**), and that the expression of several genes encoding LEA proteins with distinct biochemical characteristics are key for DT in *X. viscosa* (**Chapter 2**, **Chapter 4**). Thus, I propose that the overexpression of multiple *LEA* genes, rather than single genes, may be more efficient in promoting not only plant stress tolerance, but also enhancing seed DT in recalcitrant and intermediate species, as well as seed longevity in crops.

In this way, it would make sense to develop plants overexpressing a common regulator of multiple DT-related *LEA* genes. Previous studies have demonstrated the role of Abscisic Acid-Insensitive 5 (ABI5) as a regulator of seed maturation, DT and longevity in legumes (Verdier *et al.*, 2013; Zinsmeister *et al.*, 2016), and in altering the expression of *LEA* genes and accumulation of LEA proteins (Carles *et al.*, 2002; Zinsmeister *et al.*, 2016). Also, we have shown that the ABI5 binding motif is enriched in the promoter of several *LEAs* in *X. viscosa* (**Chapter 2**) and ABI5 may thus be, a good candidate for targeted overexpression of several *LEA* genes. However, as ABI5 also regulates other aspects of plant development mediated by hormonal crosstalk (Skubacz *et al.*, 2016), the manipulation of its expression timing, tissue localisation and expression level should also be considered in order to avoid constraints in plant growth and development.

Having in hand all the information provided in this thesis about the major processes and genes modulating DT in *X. viscosa*, one could think of several other more suitable candidates for enhancing crop tolerance by genome editing. The most important point is that nowadays we have accessible technology to study genomes

and their evolution and responses to the changing environment, and also the technology to convert this knowledge into more sustainable practices to bring back food security and fight hunger worldwide.

However, a recent determination by the Court of Justice of the European Union (ECJ) established that crops created using gene editing techniques, such as CRISPR-Cas9, will be subjected to the same regulations as those of genetically modified organisms (GMOs) (Callaway, 2018). This measure hinders advances in plant biotechnology and investments in research towards genome-edited crops that can help to assure food security at a faster rate in the current scenario of global population growth and climate change. In my opinion, it is time for us, researchers, to speak up loud about the importance, reliability and safety of such developments, and try to shift the focus of public opinion towards the long term benefits rather than negligible risks.

Where to from here?

With the above discussion in mind I provide below some outstanding questions still remaining to be addressed, which could be the focus of future research.

Outstanding questions	
⇒ What is the role of downregulated *LEAs* in *X. viscosa* leaves during desiccation?	⇒ Do LEAs form an ensemble or complex *in vivo*?
⇒ What is the role of upregulated *XvLEAs* in *X. viscosa* leaves during rehydration?	⇒ Are there biological functions of LEAs specifically dependent on intrinsic disorder or folding level?
⇒ Are CoDAGs present in the genome of other resurrection plants and do they display similar expression patterns there?	⇒ Is the cellular water potential the major driver of LEA protein folding?
⇒ Do distinct resurrection plants recruit similar sets of genes at specific water contents?	⇒ Are LEAs disordered under biological conditions?
⇒ Are there conserved elements in the promoter region of *LEAs* which are syntenically conserved across distantly-related species?	⇒ Is the overexpression of *AtLEA30* sufficient to improve longevity in seeds and rescue DT in germinated seeds of wild-type plants?
⇒ Is there a functional coordination between the genes conserved in a syntenic region?	⇒ Would double or triple knockouts of DT-related *AtLEAs* influence longevity and dormancy?

Thesis summary

The acquisition of desiccation tolerance (DT) was critical for plant colonization and diversification in environments with limited water availability. Common mechanisms underlying DT are shared amongst resurrection plants and orthodox seeds. One example is the high gene expression and accumulation of Late Embryogenesis Abundant (LEA) proteins during water loss from the cells. LEAs display variable structure and are able to perform multiple functions, which facilitates the protection of subcellular components against irreversible damage. The knowledge of how DT mechanisms evolved and work in a concerted manner during water stress has a tremendous potential for agricultural applications.

In this thesis a multilevel approach has been applied in order to investigate the genomic and molecular basis of DT in a resurrection species and in orthodox seeds. In **Chapter 2** the genomic structure and transcriptomic responses to desiccation and rehydration were investigated in the monocot resurrection species *Xerophyta viscosa*. Several genes and gene families were shown to coordinate the activation of protection and anti-senescence mechanisms. The results obtained in this chapter give strong support to the hypothesis of reactivation of orthodox seed mechanisms of DT by resurrection plants. One example of that was the finding of gene family expansions and high expression of *LEA* genes during *X. viscosa* leaf desiccation.

In **Chapter 3** we explored the evolution of *LEA* genes in 60 genomes including green algae, bryophytes, lycophytes, and angiosperms, in order to understand the evolutionary sources of their diversification and responsiveness to water stress. We found that a few LEA families are present in the green algae genomes, and the expansion of *LEA* genes correlates with the transition of plants from water to land. We propose that the differential synteny and protein evolution of *LEAs* in angiosperms were the main sources of functional diversification within and between LEA families, and were the major factors contributing to their role in water stress adaptation in plants.

In **Chapter 4** we conducted an extensive characterization of the structure and function of six LEA proteins expressed in *X. viscosa* leaves upon desiccation. *In silico* and *in vitro* analysis revealed that the selected candidates displayed characteristics of intrinsically disordered proteins (IDPs), and that defined secondary structures could be induced by solute perturbation. The structural plasticity of the proteins partially correlates with their ability to protect enzyme activity *in vitro* as well as bacterial growth against stresses. One of the candidates was also shown to enhance stress survival in

seedlings and plants of *A. thaliana* in an heterologous expression system, constituting a good candidate for future investigations.

In **Chapter 5** we investigated the physiological effects of mutations in four individual *LEA* genes expressed during seed maturation in *A. thaliana* seeds. The down regulation of three of the candidates did not strongly affect seed DT. However, one of the candidate mutants, *atlea30*, displayed lowered DT and longevity, and higher dormancy levels than wild-type seeds. The seedlings of *atlea30* were also more sensitive to germination under heat and salt stress. We propose that the network involving LEA proteins is strongly regulated, and that AtLEA30 is downstream of the common regulatory network of DT, longevity, and dormancy.

In **Chapter 6** the main findings of this thesis are integrated and discussed in the context of previous work. Also, we show how the fundamental knowledge of resurrection plant physiology and LEA protein function can contribute to applications in material science, agriculture, and food security.

References

Agarwal, T., Upadhyaya, G., Halder, T., Mukherjee, A., Majumder, A.L. and Ray, S. (2017) Different dehydrins perform separate functions in Physcomitrella patens. *Planta* **245**, 101-118.

Alpert, P. (2000) The discovery, scope, and puzzle of desiccation tolerance in plants. *Plant Ecol* **151**, 5-17.

Alpert, P. (2005) The limits and frontiers of desiccation-tolerant life. *Integr Comp Biol* **45**, 685-695.

Alpert, P. (2006) Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? *Journal of Experimental Biology* **209**, 1575-1584.

Alsheikh, M.K., Heyen, B.J. and Randall, S.K. (2003) Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. *J Biol Chem* **278**, 40882-40889.

Alsheikh, M.K., Svensson, J.T. and Randall, S.K. (2005) Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. *Plant Cell and Environment* **28**, 1114-1122.

Altunoglu, Y.C., Baloglu, M.C., Baloglu, P., Yer, E.N. and Kara, S. (2017) Genome-wide identification and comparative expression analysis of LEA genes in watermelon and melon genomes. *Physiol Mol Biol Pla* **23**, 5-21.

Amara, I., Zaidi, I., Masmoudi, K., Ludevid, M.D., Pagès, M., Goday, A. and Brini, F. (2014) Insights into late embryogenesis abundant (LEA) proteins in plants: from structure to the functions. . *American journal of plant sciences* **5**, 3440-3445.

Anderson, P.W. (1972) More Is Different - Broken Symmetry and Nature of Hierarchical Structure of Science. *Science* **177**, 393-&.

Angelovici, R., Galili, G., Fernie, A.R. and Fait, A. (2010) Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci* **15**, 211-218.

Arsovski, A.A., Pradinuk, J., Guo, X.Q., Wang, S. and Adams, K.L. (2015) Evolution of cis-regulatory elements and regulatory networks in duplicated genes of Arabidopsis. *Plant Physiol* **169**, 2982-2991.

Avelange-Macherel, M.H., Candat, A., Neveu, M., Tolleter, D. and Macherel, D. (2018) Decoding the Divergent Subcellular Location of Two Highly Similar Paralogous LEA Proteins. *International journal of molecular sciences* **19**.

Babu, R.C., Zhang, J., Blum, A., Ho, T.H.D., Wu, R., Nguyen, H.T. (2004) HVA1, a LEA gene from barley confers dehydration tolerance in transgenic rice (Oryza sativa L.) via cell membrane protection. *Plant Sci* **166**, 855-862.

Bai, B., Peviani, A., van der Horst, S., Gamm, M., Snel, B., Bentsink, L. and Hanson, J. (2017) Extensive translational regulation during seed germination revealed by polysomal profiling. *New Phytologist* **214**, 233-244.

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

Banerjee, A. and Roychoudhury, A. (2016) Group II late embryogenesis abundant (LEA) proteins: structural and functional aspects in plant abiotic stress. *Plant Growth Regul* **79**, 1-17.

Bastian, M., Heymann, S., Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media* **8**, 361-362.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L.L. (2002) The Pfam Protein Families Database. *Nucleic Acids Res* **30**, 276-280.

Battaglia, M. and Covarrubias, A.A. (2013) Late Embryogenesis Abundant (LEA) proteins in legumes. *Front Plant Sci* **4**.

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F. and Covarrubias, A.A. (2008) The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* **148**, 6-24.

Becker, B. and Marin, B. (2009) Streptophyte algae and the origin of embryophytes. *Ann Bot-London* **103**, 999-1004.

Bedi, S., Sengupta, S., Ray, A. and Chaudhuri, R.N. (2016) ABI3 mediates dehydration stress recovery response in Arabidopsis thaliana by regulating expression of downstream genes. *Plant Sci* **250**, 125-140.

Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S.P.C. and Koornneef, M. (2000) Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. *Plant Physiol* **124**, 1595-1604.

Bentsink, L., Jowett, J., Hanhart, C.J. and Koornneef, M. (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17042-17047.

Bewley, J.D. (1979) Physiological-Aspects of Desiccation Tolerance. *Annu Rev Plant Phys* **30**, 195-238.

Bewley, J.D. (1995) Physiological-Aspects of Desiccation Tolerance - a Retrospect. *Int J Plant Sci* **156**, 393-403.

Bewley, J.D., Bradford, K. and Hilhorst, H. (2012) *Seeds: physiology of development, germination and dormancy*. Springer Science & Business Media.

Bies-Etheve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Cooke, R. and 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**, 107-124.

Black, M. and Pritchard, H.W. (2002) *Desiccation and survival in plants: drying without dying*: Cabi.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**, 365-370.

Boetzer, M. and Pirovano, W. (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *Bmc Bioinformatics* **15**.

Boothby, T.C., Tapia, H., Brozena, A.H., Piszkiewicz, S., Smith, A.E., Giovannini, I., Rebecchi, L., Pielak, G.J., Koshland, D. and Goldstein, B. (2017) Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Molecular cell* **65**, 975-+.

Boudet, J., Buitink, J., Hoekstra, F.A., Rogniaux, H., Larre, C., Satour, P. and Leprince, O. (2006) Comparative analysis of the heat stable proteome of radicles of Medicago truncatula seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. *Plant Physiol* **140**, 1418-1436.

Bremer, A., Wolff, M., Thalhammer, A. and Hincha, D.K. (2017) Folding of intrinsically disordered plant LEA proteins is driven by glycerol-induced crowding and the presence of membranes. *The FEBS journal*.

Briand, L., Marcion, G., Kriznik, A., Heydel, J.-M., Artur, Y., Garrido, C., Seigneuric, R. and Neiers, F. (2016) A self-inducible heterologous protein expression system in Escherichia coli. *Scientific reports* **6**, 33037.

Brini, F., Hanin, M., Lumbreras, V., Irar, S., Pages, M. and Masmoudi, K. (2007) Functional characterization of DHN-5, a dehydrin showing a differential phosphorylation pattern in two Tunisian durum wheat (Triticum durum Desf.) varieties with marked differences in salt and drought tolerance. *Plant Sci* **172**, 20-28.

Brown, C.J., Johnson, A.K., Dunker, A.K. and Daughdrill, G.W. (2011) Evolution and disorder. *Current opinion in structural biology* **21**, 441-446.

Browne, J., Tunnacliffe, A. and Burnell, A. (2002) Anhydrobiosis - Plant desiccation gene found in a nematode. *Nature* **416**, 38-38.

Buijs, G., Kodde, J., Groot, S.P.C. and Bentsink, L. (2018) Seed dormancy release accelerated by elevated partial pressure of oxygen is associated with DOG loci. *Journal of experimental botany* **69**, 3601-3608.

Buitink, J. and Leprince, O. (2004) Glass formation in plant anhydrobiotes: survival in the dry state. *Cryobiology* **48**, 215-228.

Buitink, J., Vu, B.L., Satour, P. and Leprince, O. (2003) The re-establishment of desiccation tolerance in germinated radicles of Medicago truncatula Gaertn. seeds. *Seed Sci Res* **13**, 273-286.

Callaway (2018)

Candat, A., Paszkiewicz, G., Neveu, M., Gautier, R., Logan, D.C., Avelange-Macherel, M.H. and Macherel, D. (2014) The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in Arabidopsis offers tailored protection against abiotic stress. *The Plant cell* **26**, 3148-3166.

Cannarozzi, G., Plaza-Wuthrich, S., Esfeld, K., Larti, S., Wilson, Y.S., Girma, D., de Castro, E., Chanyalew, S., Blosch, R., Farinelli, L., Lyons, E., Schneider, M., Falquet, L., Kuhlemeier, C., Assefa, K. and Tadele, Z. (2014) Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (Eragrostis tef). *Bmc Genomics* **15**.

Cannon, S.B., Mitra, A., Baumgarten, A., Young, N.D. and May, G. (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *Bmc Plant Biol* **4**, 10.

Cao, J. and Li, X. (2015) Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (Solanum lycopersicum). *Planta* **241**, 757-772.

Carbonero, P., Iglesias-Fernandez, R. and Vicente-Carbajosa, J. (2017) The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *Journal of experimental botany* **68**, 871-880.

Carles, C., Bies-Etheve, N., Aspart, L., Leon-Kloosterziel, K.M., Koornneef, M., Echeverria, M. and Delseny, M. (2002) Regulation of Arabidopsis thaliana Em genes: role of ABI5. *Plant Journal* **30**, 373-383.

Chahtane, H., Kim, W. and Lopez-Molina, L. (2017) Primary seed dormancy: a temporally multilayered riddle waiting to be unlocked. *Journal of experimental botany* **68**, 857-869.

Chakrabortee, S., Boschetti, C., Walton, L.J., Sarkar, S., Rubinsztein, D.C. and Tunnacliffe, A. (2007) Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18073-18078.

Chakrabortee, S., Meersman, F., Schierle, G.S.K., Bertoncini, C.W., McGee, B., Kaminski, C.F. and

Tunnacliffe, A. (2010) Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16084-16089.

Chakrabortee, S., Tripathi, R., Watson, M., Schierle, G.S.K., Kurniawan, D.P., Kaminski, C.F., Wise, M.J. and Tunnacliffe, A. (2012) Intrinsically disordered proteins as molecular shields. *Mol Biosyst* **8**, 210-219.

Charfeddine, S., Saidi, M.N., Charfeddine, M. and Gargouri-Bouzid, R. (2015) Genome-wide identification and expression profiling of the late embryogenesis abundant genes in potato with emphasis on dehydrins. *Molecular biology reports* **42**, 1163-1174.

Chatelain, E., Hundertmark, M., Leprince, O., Le Gall, S., Satour, P., Deligny-Penninck, S., Rogniaux, H. and Buitink, J. (2012) Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell and Environment* **35**, 1440-1455.

Chen, Y.S., Lo, S.F., Sun, P.K., Lu, C.A., Ho, T.H. and Yu, S.M. (2015) A late embryogenesis abundant protein HVA1 regulated by an inducible promoter enhances root growth and abiotic stress tolerance in rice without yield penalty. *Plant biotechnology journal* **13**, 105-116.

Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C.Y., Ecker, J.R., Cantu, D., Rank, D.R. and Schatz, M.C. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13**, 1050-+.

Ciccarelli, F.D. and Bork, P. (2005) The WHy domain mediates the response to desiccation in plants and bacteria. *Bioinformatics* **21**, 1304-1307.

Close, T.J. (1996) Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiol Plantarum* **97**, 795-803.

Conant, G.C. and Wolfe, K.H. (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* **9**, 938-950.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.

Conrad, P. and Allen-Hoffmann, L. (2005)

Costa, M.-C.D., Artur, M.A.S., Maia, J., Jonkheer, E., Derks, M.F.L., Nijveen, H., Williams, B., Mundree, S.G., Jiménez-Gómez, J.M., Hesselink, T., Schijlen, E.G.W.M., Ligterink, W., Oliver, M.J., Farrant, J.M. and Hilhorst, H.W.M. (2017a) A footprint of desiccation tolerance in the genome of Xerophyta viscosa. **3**, 17038.

Costa, M.C.D., Artur, M.A.S., Maia, J., Jonkheer, E., Derks, M.F.L., Nijveen, H., Williams, B., Mundree, S.G., Jimenez-Gomez, J.M., Hesselink, T., Schijlen, E.G.W.M., Ligterink, W., Oliver, M.J., Farrant, J.M. and Hilhorst, H.W.M. (2017b) A footprint of desiccation tolerance in the genome of Xerophyiita viscosa. *Nature Plants* **3**.

Costa, M.C.D., Farrant, J.M., Oliver, M.J., Ligterink, W., Buitink, J. and Hilhorst, H.M.W. (2016) Key genes involved in desiccation tolerance and dormancy across life forms. *Plant Sci* **251**, 162-168.

Covarrubias, A.A., Cuevas-Velazquez, C.L., Romero-Perez, P.S., Rendon-Luna, D.F. and Chater, C.C.C. (2017) Structural disorder in plant proteins: where plasticity meets sessility. *Cellular and Molecular Life Sciences* **74**, 3119-3147.

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

Csintalan, Z., Tuba, Z., Lichtenthaler, H.K. and Grace, J. (1996) Reconstitution of photosynthesis upon rehydration in the desiccated leaves of the poikilochlorophyllous shrub Xerophyta scabrida at elevated CO2. *Journal of plant physiology* **148**, 345-350.

Csűös, M. (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910-1912.

Cuevas-Velazquez, C.L., Reyes, J.L. and Covarrubias, A.A. (2017) Group 4 late embryogenesis abundant proteins as a model to study intrinsically disordered proteins in plants. *Plant signaling & behavior* **12**, e1343777.

Cuevas-Velazquez, C.L., Saab-Rincon, G., Reyes, J.L. and Covarrubias, A.A. (2016) The Unstructured N-terminal Region of Arabidopsis Group 4 Late Embryogenesis Abundant (LEA) Proteins Is Required for Folding and for Chaperone-like Activity under Water Deficit. *J Biol Chem* **291**, 10893-10903.

Cuming, A.C., Cho, S.H., Kamisugi, Y., Graham, H. and 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.

Cushman, J. and Walters, C. (2004)

Cushman, J.C. and Oliver, M.J. (2011) Understanding vegetative desiccation tolerance using integrated functional genomics approaches within a comparative evolutionary framework. In *Plant desiccation tolerance*: Springer, pp. 307-338.

Dai, A.G. (2013) Increasing drought under global warming in observations and models (vol 3, pg 52, 2013). *Nat Clim Change* **3**, 171-171.

Das, R.K., Ruff, K.M. and Pappu, R.V. (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current opinion in structural biology* **32**, 102-112.

Dayhoff, M.O. (1976) Origin and Evolution of Protein Superfamilies. *Fed Proc* **35**, 2132-2138.

de Vries, J. and Archibald, J.M. (2018) Plant evolution: landmarks on the path to terrestrial life. *New Phytologist* **217**, 1428-1434.

Dekkers, B.J., He, H., Hanson, J., Willems, L.A., Jamar, D.C., Cueff, G., Rajjou, L., Hilhorst, H.W. and Bentsink, L. (2016) The Arabidopsis DELAY OF GERMINATION 1 gene affects ABSCISIC ACID INSENSITIVE 5 (ABI5) expression and genetically interacts with ABI3 during Arabidopsis seed development. *The Plant journal : for cell and molecular biology* **85**, 451-465.

Dekkers, B.J.W., Costa, M.C.D., Maia, J., Bentsink, L., Ligterink, W. and Hilhorst, H.W.M. (2015) Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. *Planta* **241**, 563-577.

Dekkers, B.J.W., Willems, L., Bassel, G.W., van Bolderen-Veldkamp, R.P., Ligterink, W., Hilhorst, H.W.M. and Bentsink, L. (2012) Identification of Reference Genes for RT-qPCR Expression Analysis in Arabidopsis and Tomato Seeds. *Plant and Cell Physiology* **53**, 28-37.

Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H. and Buitink, J. (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. *Journal of experimental botany* **64**, 4559-4573.

Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J.G.B.,

McCourt, P. and Samuel, M.A. (2013) ABI3 controls embryo degreening through Mendel's I locus. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E3888-E3894.

Delouche, J.C. and Baskin, C.C. (2016) Accelerated aging techniques for predicting the relative storability of seed lots.

Delseny, M., Bies-Etheve, N., Carles, C., Hull, G., Vicient, C., Raynal, M., Grellet, F. and Aspart, L. (2001) Late Embryogenesis Abundant (LEA) protein gene regulation during Arabidopsis seed maturation. *Journal of plant physiology* **158**, 419-427.

deMelo, N.F., Guerra, M., Benkoiseppon, A.M. and deMenezes, N.L. (1997) Cytogenetics and cytotaxonomy of Velloziaceae. *Plant Syst Evol* **204**, 257-273.

Demuth, J.P. and Hahn, M.W. (2009) The life and death of gene families. *BioEssays : news and reviews in molecular, cellular and developmental biology* **31**, 29-39.

Devic, M. and Roscoe, T. (2016) Seed maturation: Simplification of control networks in plants. *Plant Sci* **252**, 335-346.

Dewey, C.N. (2011) Positional orthology: putting genomic evolutionary relationships into context. *Brief Bioinform* **12**, 401-412.

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

Disfani, F.M., Hsu, W.L., Mizianty, M.J., Oldfield, C.J., Xue, B., Dunker, A.K., Uversky, V.N. and Kurgan, L. (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, I75-I83.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930.

Dosztányi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434.

Drira, M., Saibi, W., Amara, I., Masmoudi, K., Hanin, M. and Brini, F. (2015) Wheat Dehydrin K-Segments Ensure Bacterial Stress Tolerance, Antiaggregation and Antimicrobial Effects. *Appl Biochem Biotech* **175**, 3310-3321.

Du, D.L., Zhang, Q.X., Cheng, T.R., Pan, H.T., Yang, W.R. and Sun, L.D. (2013) Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in Prunus mume. *Molecular biology reports* **40**, 1937-1946.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.R., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C.H., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, M., Garner, E.C. and Obradovic, Z. (2001) Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59.

Dure, L. (1993) A Repeating 11-Mer Amino-Acid Motif and Plant Desiccation. *Plant Journal* **3**, 363-369.

Dure, L., 3rd, Crouch, M., Harada, J., Ho, T.H., Mundy, J., Quatrano, R., Thomas, T. and Sung, Z.R. (1989a) Common amino acid sequence domains among the LEA proteins of higher plants. *Plant molecular biology* **12**, 475-486.

Dure, L., Crouch, M., Harada, J., Ho, T.H.D., Mundy, J., Quatrano, R., Thomas, T. and Sung, Z.R. (1989b) Common Amino-Acid Sequence Domains among the Lea Proteins of Higher-Plants. *Plant molecular biology* **12**, 475-486.

Dure, L., Greenway, S.C. and Galau, G.A. (1981a) Developmental Biochemistry of Cottonseed Embryogenesis and Germination - Changing Messenger Ribonucleic-Acid Populations as Shown by Invitro and Invivo Protein-Synthesis .14. *Biochemistry-Us* **20**, 4162-4168.

Dure, L.S., Galau, G.A. and Greenway, S. (1981b) Changing Protein-Patterns during Cotton Cotyledon Embryogenesis and Germination as Shown by Invivo and Invitro Synthesis. *Israel J Bot* **29**, 293-306.

Dyson, H.J. and Wright, P.E. (2002) Coupling of folding and binding for unstructured proteins. *Current opinion in structural biology* **12**, 54-60.

Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Bio* **6**, 197-208.

Eisen, J.A. (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* **8**, 163-167.

Ellis, R.H., Hong, T.D. and Roberts, E.H. (1991) An Intermediate Category of Seed Storage Behavior .2. Effects of Provenance, Immaturity, and Imbibition on Desiccation-Tolerance in Coffee. *Journal of experimental botany* **42**, 653-657.

Emms, D.M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* **16**.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J.X., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C. and Gibbs, R.A. (2012) Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PloS one* **7**.

Escobar, C., De Meutter, J., Aristizabal, F.A., Sanz-Alferez, S., del Campo, F.F., Barthels, N., Van der Eycken, W., Seurinck, J., van Montagu, M., Gheysen, G. and Fenoll, C. (1999) Isolation of the LEMMI9 gene and promoter analysis during a compatible plant-nematode interaction. *Mol Plant Microbe In* **12**, 440-449.

Espelund, M., Debedout, J.A., Outlaw, W.H. and Jakobsen, K.S. (1995) Environmental and Hormonal-Regulation of Barley Late-Embryogenesis-Abundant (Lea) Messenger-Rnas Is Via Different Signal-Transduction Pathways. *Plant Cell and Environment* **18**, 943-949.

Espelund, M., Saeboelarssen, S., Hughes, D.W., Galau, G.A., Larsen, F. and Jakobsen, K.S. (1992) Late Embryogenesis-Abundant Genes Encoding Proteins with Different Numbers of Hydrophilic Repeats Are Regulated Differentially by Abscisic-Acid and Osmotic-Stress. *Plant Journal* **2**, 241-252.

FAO (2017)

Farrant, J.M., Cooper, K., Dace, H.J.W., Bentely, J. and Hilgart, A. (2016) Desiccation tolerance. In *Plant Stress Physiology* (Shabala, S., ed): CAB International, pp. 217–252.

Farrant, J.M., Cooper, K., Hilgart, A., Abdalla, K.O., Bentley, J., Thomson, J.A., Dace, H.J.W., Peton, N., Mundree, S.G. and Rafudeen, M.S. (2015) A molecular physiological review of vegetative desiccation tolerance in the resurrection plant Xerophyta viscosa (Baker). *Planta* **242**, 407-426.

Farrant, J.M., Cooper, K. and Nell, H. (2012) Desiccation tolerance. *Plant stress physiology*, 238-265.

Farrant, J.M. and Kruger, L.A. (2001) Longevity of dry Myrothamnus flabellifolius in simulated field conditions. *Plant Growth Regul* **35**, 109-120.

Farrant, J.M. and Moore, J.P. (2011) Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Curr Opin Plant Biol* **14**, 340-345.

Finkelstein, R.R. (1993) Abscisic Acid-Insensitive Mutations Provide Evidence for Stage-Specific Signal Pathways Regulating Expression of an Arabidopsis Late Embryogenesis-Abundant (Lea) Gene. *Mol Gen Genet* **238**, 401-408.

Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29-W37.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-D285.

Fischer, R.A. and Connor, D.J. (2018) Issues for cropping and agricultural science in the next 20 years. *Field Crop Res* **222**, 121-142.

Flagel, L.E. and Wendel, J.F. (2009) Gene duplication and evolutionary novelty in plants. *New Phytologist* **183**, 557-564.

Franklin, R.E. and Gosling, R.G. (1953) The Structure of Sodium Thymonucleate Fibres .1. The Influence of Water Content. *Acta Crystallogr* **6**, 673-677.

Freeling, M. (2009) Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. *Annu Rev Plant Biol* **60**, 433-453.

Freeman, T.C., Goldovsky, L., Brosch, M., Van Dongen, S., Maziere, P., Grocock, R.J., Freilich, S., Thornton, J. and Enright, A.J. (2007) Construction, visualisation, and clustering of transcription networks from Microarray expression data. *Plos Comput Biol* **3**, 2032-2042.

French-Pacheco, L., Cuevas-Velazquez, C.L., Rivillas-Acevedo, L., Covarrubias, A.A. and Amero, C. (2018) Metal-binding polymorphism in late embryogenesis abundant protein AtLEA4-5, an intrinsically disordered protein. *PeerJ* **6**.

Fu, D., Huang, B., Xiao, Y., Muthukrishnan, S. and Liang, G.H. (2007) Overexpression of barley hva1 gene in creeping bentgrass for improving drought tolerance. *Plant cell reports* **26**, 467-477.

Furuki, T. and Sakurai, M. (2016) Group 3 LEA protein model peptides protect enzymes against desiccation stress. *Bba-Proteins Proteom* **1864**, 1237-1243.

Furuki, T., Shimizu, T., Kikawada, T., Okuda, T. and Sakurai, M. (2011) Salt Effects on the Structural and Thermodynamic Properties of a Group 3 LEA Protein Model Peptide. *Biochemistry-Us* **50**, 7093-7103.

Gabaldon, T. and Koonin, E.V. (2013) Functional and evolutionary implications of gene orthology. *Nature reviews. Genetics* **14**, 360-366.

Gaff, D. and Loveys, B. (1993) Abscisic acid levels in drying plants of a resurrection grass. *Transactions of the Malaysian Society of Plant Physiology* **3**, 286-287.

Gaff, D.F. (1971) Desiccation-Tolerant Flowering Plants in Southern Africa. *Science* **174**, 1033-&.

Gaff, D.F. (1977) Desiccation Tolerant Vascular Plants of Southern-Africa. *Oecologia* **31**, 95-109.

Gaff, D.F. and Oliver, M. (2013) The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. *Funct Plant Biol* **40**, 315-328.

Galau, G.A., Bijaisoradat, N. and Hughes, D.W. (1987) Accumulation Kinetics of Cotton Late Embryogenesis-Abundant Messenger-Rnas and Storage Protein Messenger-Rnas - Coordinate Regulation during Embryogenesis and the Role of Abscisic-Acid. *Dev Biol* **123**, 198-212.

Galau, G.A., Hughes, D.W. and Dure, L. (1986) Abscisic-Acid Induction of Cloned Cotton Late Embryogenesis-Abundant (Lea) Messenger-Rnas. *Plant molecular biology* **7**, 155-170.

Gao, J. and Lan, T. (2016) Functional characterization of the late embryogenesis abundant (LEA) protein gene family from Pinus tabuliformis (Pinaceae) in Escherichia coli. *Scientific reports* **6**, 19467.

Garay-Arroyo, A., Colmenero-Flores, J.M., Garciarrubio, A. and Covarrubias, A.A. (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J Biol Chem* **275**, 5668-5674.

Garrison, E. and Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Gechev, T.S., Benina, M., Obata, T., Tohge, T., Sujeeth, N., Minkov, I., Hille, J., Temanni, M.R., Marriott, A.S., Bergstrom, E., Thomas-Oates, J., Antonio, C., Mueller-Roeber, B., Schippers, J.H., Fernie, A.R. and Toneva, V. (2013a) Molecular mechanisms of desiccation tolerance in the resurrection glacial relic Haberlea rhodopensis. *Cellular and molecular life sciences : CMLS* **70**, 689-709.

Gechev, T.S., Benina, M., Obata, T., Tohge, T., Sujeeth, N., Minkov, I., Hille, J., Temanni, M.R., Marriott, A.S., Bergstrom, E., Thomas-Oates, J., Antonio, C., Mueller-Roeber, B., Schippers, J.H.M., Fernie, A.R. and Toneva, V. (2013b) Molecular mechanisms of desiccation tolerance in the resurrection glacial relic Haberlea rhodopensis. *Cellular and Molecular Life Sciences* **70**, 689-709.

Gechev, T.S., Dinakar, C., Benina, M., Toneva, V. and Bartels, D. (2012) Molecular mechanisms of desiccation tolerance in resurrection plants. *Cellular and Molecular Life Sciences* **69**, 3175-3186.

Gekko, K., Ohmae, E., Kameyama, K. and Takagi, T. (1998) Acetonitrile-protein interactions: amino acid solubility and preferential solvation. *Bba-Protein Struct M* **1387**, 195-205.

Giarola, V., Hou, Q.C. and Bartels, D. (2017) Angiosperm Plant Desiccation Tolerance: Hints from Transcriptomics and Genome Sequencing. *Trends Plant Sci* **22**, 705-717.

Ginbot, Z.G. and Farrant, J.M. (2011) Physiological response of selected eragrostis species to water-deficit stress. *Afr J Biotechnol* **10**, 10405-10417.

Goldberg, R.B., Depaiva, G. and Yadegari, R. (1994) Plant Embryogenesis - Zygote to Seed. *Science* **266**, 605-614.

Goodstein, D.M., Shu, S.Q., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**, D1178-D1186.

Goyal, K., Pinelli, C., Maslen, S.L., Rastogi, R.K., Stephens, E. and Tunnacliffe, A. (2005a) Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. *Febs Lett* **579**, 4093-4098.

Goyal, K., Walton, L.J., Browne, J.A., Burnell, A.M. and Tunnacliffe, A. (2005b) Molecular anhydrobiology: Identifying molecules implicated in invertebrate anhydrobiosis. *Integr Comp Biol* **45**, 702-709.

Goyal, K., Walton, L.J. and Tunnacliffe, A. (2005c) LEA proteins prevent protein aggregation due to water stress. *Biochem J* **388**, 151-157.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., Chen, Z.H., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-U130.

Greenfield, N.J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890.

Griffiths, C.A., Gaff, D.F. and Neale, A.D. (2014) Drying without senescence in resurrection plants. *Front Plant Sci* **5**.

Groot, S.P.C., Surki, A.A., de Vos, R.C.H. and Kodde, J. (2012) Seed storage at elevated partial pressure of oxygen, a fast method for analysing seed ageing under dry conditions. *Ann Bot-London* **110**, 1149-1159.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R. and Osburne, M.S. (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* **209**, 95-103.

Grueber, C.E. (2015) Comparative genomics for biodiversity conservation. *Comput Struct Biotec* **13**, 370-375.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075.

Gusev, O., Suetsugu, Y., Cornette, R., Kawashima, T., Logacheva, M.D., Kondrashov, A.S., Penin, A.A., Hatanaka, R., Kikuta, S., Shimura, S., Kanamori, H., Katayose, Y., Matsumoto, T., Shagimardanova, E., Alexeev, D., Govorun, V., Wisecaver, J., Mikheyev, A., Koyanagi, R., Fujie, M., Nishiyama, T., Shigenobu, S., Shibata, T.F., Golygina, V., Hasebe, M., Okuda, T., Satoh, N. and Kikawada, T. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat Commun* **5**.

Halder, T., Upadhyaya, G. and Ray, S. (2017) YSK2 Type Dehydrin (SbDhn1) from Sorghum bicolor Showed Improved Protection under High Temperature and Osmotic Stress Condition. *Front Plant Sci* **8**.

Han, J.D., Li, X., Jiang, C.K., Wong, G.K.S., Rothfels, C.J. and Rao, G.Y. (2017) Evolutionary Analysis of the LAFL Genes Involved in the Land Plant Seed Maturation Program. *Front Plant Sci* **8**.

Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K. and Shiu, S.H. (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol* **148**, 993-1003.

Hara, M., Fujinaga, M. and Kuboi, T. (2005) Metal binding by citrus dehydrin with histidine-rich domains. *Journal of experimental botany* **56**, 2695-2703.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R. and Consortium, G.O. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-D261.

He, S., Tan, L.L., Hu, Z.L., Chen, G.P., Wang, G.X. and Hu, T.Z. (2012) Molecular characterization and functional analysis by heterologous expression in E. coli under diverse abiotic stresses for OsLEA5, the atypical hydrophobic LEA protein from Oryza sativa L. *Mol Genet Genomics* **287**, 39-54.

Hengherr, S., Heyer, A.G., Kohler, H.R. and Schill, R.O. (2008) Trehalose and anhydrobiosis in tardigrades - evidence for divergence in responses to dehydration. *Febs Journal* **275**, 281-288.

Hernandez-Sanchez, I.E., Maruri-Lopez, I., Graether, S.P. and Jimenez-Bremont, J.F. (2017) In vivo

evidence for homo- and heterodimeric interactions of Arabidopsis thaliana dehydrins AtCOR47, AtERD10, and AtRAB18. *Scientific reports* **7**.

Heyen, B.J., Alsheikh, M.K., Smith, E.A., Torvik, C.F., Seals, D.F. and Randall, S.K. (2002) The calcium-binding activity of a vacuole-associated, dehydrin-like protein is regulated by phosphorylation. *Plant Physiol* **130**, 675-687.

Hilbricht, T., Varotto, S., Sgaramella, V., Bartels, D., Salamini, F. and Furini, A. (2008) Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant Craterostigma plantagineum. *New Phytologist* **179**, 877-887.

Hilhorst, H.W. (2011) Standardizing seed dormancy research. In *Seed Dormancy*. Springer, pp. 43-52.

Hilhorst, H.W.M., Costa, M.C.D. and Farrant, J.M. (2018) A Footprint of Plant Desiccation Tolerance. Does It Exist? *Mol Plant* **11**, 1003-1005.

Hincha, D.K. and Thalhammer, A. (2012) LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochemical Society transactions* **40**, 1000-1003.

Hoekstra, F.A., Golovina, E.A. and Buitink, J. (2001) Mechanisms of plant desiccation tolerance. *Trends Plant Sci* **6**, 431-438.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. (2016) BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769.

Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O. and Pappu, R.V. (2017) CIDER: Resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys J* **112**, 16-21.

Hsiao, T.C. (1973) Plant Responses to Water Stress. *Annu Rev Plant Phys* **24**, 519-570.

Hu, T.Z., Zhou, N., Fu, M.L., Qin, J. and Huang, X.Y. (2016) Characterization of OsLEA1a and its inhibitory effect on the resistance of E. coli to diverse abiotic stresses. *Int J Biol Macromol* **91**, 1010-1017.

Huang, K.C., Lin, W.C. and Cheng, W.H. (2018) Salt hypersensitive mutant 9, a nucleolar APUM23 protein, is essential for salt sensitivity in association with the ABA signaling pathway in Arabidopsis. *Bmc Plant Biol* **18**.

Hughes, D.W. and Galau, G.A. (1987) Translation Efficiency of Lea Messenger-Rnas in Cotton Embryos - Minor Changes during Embryogenesis and Germination. *Plant molecular biology* **9**, 301-313.

Hughes, D.W. and Galau, G.A. (1989) Temporally Modular Gene-Expression during Cotyledon Development. *Gene Dev* **3**, 358-369.

Hughes, D.W. and Galau, G.A. (1991) Developmental and Environmental Induction of Lea and Leaa Messenger-Rnas and the Postabscission Program during Embryo Culture. *The Plant cell* **3**, 605-618.

Hunault, G. and Jaspard, E. (2010) LEAPdb: a database for the late embryogenesis abundant proteins. *Bmc Genomics* **11**.

Hundertmark, M., Buitink, J., Leprince, O. and Hincha, D.K. (2011) The reduction of seed-specific dehydrins reduces seed longevity in Arabidopsis thaliana. *Seed Sci Res* **21**, 165-173.

Hundertmark, M. and Hincha, D.K. (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. *Bmc Genomics* **9**, 118.

Hundertmark, M., Popova, A.V., Rausch, S., Seckler, R. and Hincha, D.K. (2012) Influence of drying on the secondary structure of intrinsically disordered and globular proteins. *Biochemical and biophysical research communications* **417**, 122-128.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z. and Dunker, A.K. (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037-1049.

Illing, N., Denby, K.J., Collett, H., Shen, A. and Farrant, J.M. (2005) The signature of seeds in resurrection plants: A molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues. *Integr Comp Biol* **45**, 771-787.

Ingram, J. and Bartels, D. (1996) The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Phys* **47**, 377-403.

Innan, H. and Kondrashov, F. (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nature reviews. Genetics* **11**, 97-108.

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S. and Venkataraman, G. (2018) CRISPR for Crop Improvement: An Update Review. *Front Plant Sci* **9**.

Jaspard, E. and Hunault, G. (2014) Comparison of Amino Acids Physico-Chemical Properties and Usage of Late Embryogenesis Abundant Proteins, Hydrophilins and WHy Domain. *PloS one* **9**, e109570.

Jia, F.J., Qi, S.D., Li, H., Liu, P., Li, P.C., Wu, C.G., Zheng, C.C. and Huang, J.G. (2014) Overexpression of Late Embryogenesis Abundant 14 enhances Arabidopsis salt stress tolerance. *Biochemical and biophysical research communications* **454**, 505-511.

Jiang, S.J., Wang, J., Liu, X.L., Liu, Y.Y., Guo, C., Zhang, L.W., Han, J.H., Wu, X.L., Xue, D., Gomaa, A.E., Feng, S., Zhang, H., Chen, Y., Ping, S.Z., Chen, M., Zhang, W., Li, L., Zhou, Z.F., Zuo, K.J., Li, X.F., Yang, Y. and Lin, M. (2017) DrwH, a novel WHy domain-containing hydrophobic LEA5C protein from Deinococcus radiodurans, protects enzymatic activity under oxidative stress. *Scientific reports* **7**, 9281.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J. and dePamphilis, C.W. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97-100.

Joh, T., Honjoh, K., Yoshimoto, M., Funabashi, J., Miyamoto, T. and Hatano, S. (1995) Molecular-Cloning and Expression of Hardening-Induced Genes in Chlorella-Vulgaris C-27 - the Most Abundant Clone Encodes a Late Embryogenesis Abundant Protein. *Plant and Cell Physiology* **36**, 85-93.

Joosen, R.V.L., Kodde, J., Willems, L.A.J., Ligterink, W., van der Plas, L.H.W. and Hilhorst, H.W.M. (2010) GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. *Plant Journal* **62**, 148-159.

Karimi, M., Inzé, D. and Depicker, A. (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci* **7**, 193-195.

Karssen, C.M., Brinkhorstvanderswan, D.L.C., Breekland, A.E. and Koornneef, M. (1983) Induction of Dormancy during Seed Development by Endogenous Abscisic-Acid - Studies on Abscisic-Acid Deficient Genotypes of Arabidopsis-Thaliana (L) Heynh. *Planta* **157**, 158-165.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066.

Kenrick, P. and Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature* **389**, 33-39.

Kermode, A.R. (1997) Approaches to elucidate the basis of desiccation-tolerance in seeds. *Seed Sci Res* **7**, 75-95.

Kermode, A.R. (2005) Role of abscisic acid in seed dormancy. *J Plant Growth Regul* **24**, 319-344.

Khandelwal, A., Cho, S.H., Marella, H., Sakata, Y., Perroud, P.F., Pan, A. and Quatrano, R.S. (2010) Role of ABA and ABI3 in Desiccation Tolerance. *Science* **327**, 546-546.

Khraiwesh, B., Qudeimat, E., Thimma, M., Chaiboonchoe, A., Jijakli, K., Alzahmi, A., Arnoux, M. and Salehi-Ashtiani, K. (2015) Genome-wide expression analysis offers new insights into the origin and evolution of Physcomitrella patens stress response. *Scientific reports* **5**.

Kikawada, T., Nakahara, Y., Kanamori, Y., Iwata, K.I., Watanabe, M., Mcgee, B., Tunnacliffe, A. and Okuda, T. (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochemical and biophysical research communications* **348**, 56-61.

Kondrashov, F.A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *P Roy Soc B-Biol Sci* **279**, 5048-5057.

Koonin, E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annual review of genetics* **39**, 309-338.

Koonin, E.V., Wolf, Y.I. and Karev, G.P. (2002) The structure of the protein universe and genome evolution. *Nature* **420**, 218-223.

Koornneef, M., Hanhart, C.J., Hilhorst, H.W. and Karssen, C.M. (1989) In Vivo Inhibition of Seed Development and Reserve Protein Accumulation in Recombinants of Abscisic Acid Biosynthesis and Responsiveness Mutants in Arabidopsis thaliana. *Plant Physiol* **90**, 463-469.

Koornneef, M., Jorna, M.L., Derswan, D.L.C.B. and Karssen, C.M. (1982) The Isolation of Abscisic-Acid (Aba) Deficient Mutants by Selection of Induced Revertants in Non-Germinating Gibberellin Sensitive Lines of Arabidopsis-Thaliana (L) Heynh. *Theor Appl Genet* **61**, 385-393.

Korf, I. (2004) Gene finding in novel genomes. *Bmc Bioinformatics* **5**.

Koster, K.L. (1991) Glass-Formation and Desiccation Tolerance in Seeds. *Plant Physiol* **96**, 302-304.

Kovacs, D., Kalmar, E., Torok, Z. and Tompa, P. (2008) Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiol* **147**, 381-390.

Kozlowski, L.P. (2016) IPC – Isoelectric Point Calculator. *Biology direct* **11**, 55.

Kramer, U. (2018) Conceptualizing plant systems evolution. *Curr Opin Plant Biol* **42**, 66-75.

Kruger, C., Berkowitz, O., Stephan, U.W. and Hell, R. (2002) A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of Ricinus communis L. *J Biol Chem* **277**, 25062-25069.

Lan, T., Gao, J. and Zeng, Q.Y. (2013) Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in Populus trichocarpa. *Tree Genet Genomes* **9**, 253-264.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-U354.

Lapinski, J. and Tunnacliffe, A. (2003) Anhydrobiosis without trehalose in bdelloid rotifers. *Febs Lett* **553**, 387-390.

Leprince, O. and Buitink, J. (2010) Desiccation tolerance: From genomics to the field. *Plant Sci* **179**, 554-564.

Leprince, O., Hendry, G. and McKersie, B. (1993) The mechanisms of desiccation tolerance in developing seeds. *Seed Sci Res* **3**, 231-246.

Leprince, O., Pellizzaro, A., Berriri, S. and Buitink, J. (2017) Late seed maturation: drying without dying. *Journal of experimental botany* **68**, 827-841.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**, W242-W245.

Li, S., Chakraborty, N., Borcar, A., Menze, M.A., Toner, M. and Hand, S.C. (2012a) Late embryogenesis abundant proteins protect human hepatoma cells during acute desiccation. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20859-20864.

Li, X. and Cao, J. (2016) Late Embryogenesis Abundant (LEA) Gene Family in Maize: Identification, Evolution, and Expression Profiles. *Plant Mol Biol Rep* **34**, 15-28.

Li, Z.H., Peng, J.Y., Wen, X. and Guo, H.W. (2012b) Gene Network Analysis and Functional Studies of Senescence-associated Genes Reveal Novel Regulators of Arabidopsis Leaf Senescence. *Journal of integrative plant biology* **54**, 526-539.

Liang, Y., Xiong, Z., Zheng, J., Xu, D., Zhu, Z., Xiang, J., Gan, J., Raboanatahiry, N., Yin, Y. and Li, M. (2016) Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in Brassica napus. *Scientific reports* **6**, 24265.

Lindsey, R. and Dahlman, L. (2018)

Ling, H., Zeng, X. and Guo, S. (2016) Functional insights into the late embryogenesis abundant (LEA) protein family from Dendrobium officinale (Orchidaceae) using an Escherichia coli system. *Scientific reports* **6**, 39693.

Liu, S.L., Pan, A.Q. and Adams, K.L. (2014) Protein Subcellular Relocalization of Duplicated Genes in Arabidopsis. *Genome Biol Evol* **6**, 2501-2515.

Liu, Y., Liang, J., Sun, L., Yang, X. and Li, D. (2016a) Group 3 LEA Protein, ZmLEA3, Is Involved in Protection from Low Temperature Stress. *Front Plant Sci* **7**, 1011.

Liu, Y., Liang, J.A., Sun, L.P., Yang, X.H. and Li, D.Q. (2016b) Group 3 LEA Protein, ZmLEA3, Is Involved in Protection from Low Temperature Stress. *Front Plant Sci* **7**.

Liu, Y., Song, Q.P., Li, D.X., Yang, X.H. and Li, D.Q. (2017a) Multifunctional Roles of Plant Dehydrins in Response to Environmental Stresses. *Front Plant Sci* **8**, 1018.

Liu, Y., Yang, M.Y., Cheng, H., Sun, N., Liu, S.M., Li, S.M., Wang, Y., Zheng, Y.Z. and Uversky, V.N. (2017b) Z The effect of phosphorylation on the salt-tolerance-related functions of the soybean protein PM18, a member of the group-3 LEA protein family. *Bba-Proteins Proteom* **1865**, 1291-1303.

Lopes, J.L., Miles, A.J., Whitmore, L. and Wallace, B.A. (2014) Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. *Protein Sci* **23**, 1765-1772.

Lopez-Molina, L., Mongrand, B., McLachlin, D.T., Chait, B.T. and Chua, N.H. (2002) ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant Journal* **32**, 317-328.

Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964.

Ma, X.L., Mau, M. and Sharbel, T.F. (2018) Genome Editing for Global Food Security. *Trends Biotechnol* **36**, 123-127.

Magwanga, R.O., Lu, P., Kirungu, J.N., Lu, H., Wang, X., Cai, X., Zhou, Z., Zhang, Z., Salih, H., Wang, K. and Liu, F. (2018) Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC genetics* **19**, 6.

Maia, J., Dekkers, B.J.W., Provart, N.J., Ligterink, W. and Hilhorst, H.W.M. (2011) The Re-Establishment of Desiccation Tolerance in Germinated Arabidopsis thaliana Seeds and Its Associated Transcriptome. *PloS one* **6**.

Malik, A.A., Veltri, M., Boddington, K.F., Singh, K.K. and Graether, S.P. (2017) Genome Analysis of Conserved Dehydrin Motifs in Vascular Plants. *Front Plant Sci* **8**, 709.

Manfre, A.J., LaHatte, G.A., Climer, C.R. and Marcotte, W.R., Jr. (2009) Seed dehydration and the establishment of desiccation tolerance during seed maturation is altered in the Arabidopsis thaliana mutant atem6-1. *Plant & cell physiology* **50**, 243-253.

Manfre, A.J., Lanni, L.M. and Marcotte, W.R. (2006) The Arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiol* **140**, 140-149.

Manfre, A.J. and Simon, A.E. (2008) Importance of coat protein and RNA silencing in satellite RNA/virus interactions. *Virology* **379**, 161-167.

Marques, A., Buijs, G., Ligterink, W. and Hilhorst, H. (2018) Evolutionary ecophysiology of seed desiccation sensitivity. *Funct Plant Biol*, -.

Mello-Silva, R., Santos, D.Y.A.C., Salatino, M.L.F., Motta, L.B., Cattai, M.B., Sasaki, D., Lovo, J., Pita, P.B., Rocini, C., Rodrigues, C.D.N., Zarrei, M. and Chase, M.W. (2011) Five vicarious genera from Gondwana: the Velloziaceae as shown by molecules and morphology. *Ann Bot-London* **108**, 87-102.

Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.Y., Bateman, A., Punta, M., Attwood, T.K., Sigrist, C.J.A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H.Z., Natale, D.A., Wu, C.H., Orengo, C., Sillitoe, I., Mi, H.Y., Thomas, P.D. and Finn, R.D. (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**, D213-D221.

Mohan, A., Oldfield, C.J., Radivojac, P., Vacic, V., Cortese, M.S., Dunker, A.K. and Uversky, V.N. (2006) Analysis of molecular recognition features (MoRFs). *Journal of molecular biology* **362**, 1043-1059.

Monke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hahnel, U., Junker, A., Weisshaar, B., Conrad, U., Baumlein, H. and Altschmied, L. (2012) Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic Acids Res* **40**, 8240-8254.

Moore, J.P., Lindsey, G.G., Farrant, J.M. and Brandt, W.F. (2007) An overview of the biology of the desiccation-tolerant resurrection plant Myrothamnus flabellifolia. *Ann Bot-London* **99**, 211-217.

Mouillon, J.M., Eriksson, S.K. and Harryson, P. (2008) Mimicking the Plant Cell Interior under Water Stress by Macromolecular Crowding: Disordered Dehydrin Proteins Are Highly Resistant to Structural Collapse. *Plant Physiol* **148**, 1925-1937.

Mouillon, J.M., Gustafsson, P. and Harryson, P. (2006) Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiol* **141**, 638-650.

Mundree, S.G., Baker, B., Mowla, S., Peters, S., Marais, S., Vander Willigen, C., Govender, K., Maredza,

A., Muyanga, S., Farrant, J.M., Thomson, J.A. (2002) Physiological and molecular insights into drought tolerance. *Afr J Biotechnol* **1**, 28-38.

Mundree, S.G. and Farrant, J.M. (2000) Some physiological and molecular insights into the mechanisms of desiccation tolerance in the resurrection plant Xerophyta viscosa Baker. In *Plant tolerance to abiotic stresses in agriculture: role of genetic engineering*: Springer, pp. 201-222.

Nakamura, S., Mano, S., Tanaka, Y., Ohnishi, M., Nakamori, C., Araki, M., Niwa, T., Nishimura, M., Kaminaka, H., Nakagawa, T., Sato, Y. and Ishiguro, S. (2010) Gateway Binary Vectors with the Bialaphos Resistance Gene, bar, as a Selection Marker for Plant Transformation. *Biosci Biotech Bioch* **74**, 1315-1319.

Nakayama, K., Okawa, K., Kakizaki, T., Honma, T., Itoh, H. and Inaba, T. (2007) Arabidopsis Cor15am is a chloroplast stromal protein that has cryoprotective activity and forms oligomers. *Plant Physiol* **144**, 513-523.

Nambara, E., Kawaide, H., Kamiya, Y. and Naito, S. (1998) Characterization of an Arabidopsis thaliana mutant that has a defect in ABA accumulation: ABA-dependent and ABA-independent accumulation of free amino acids during dehydration. *Plant and Cell Physiology* **39**, 853-858.

Nambara, E., Keith, K., Mccourt, P. and Naito, S. (1994) Isolation of an Internal Deletion Mutant of the Arabidopsis-Thaliana Ab13 Gene. *Plant and Cell Physiology* **35**, 509-513.

Nambara, E., Keith, K., Mccourt, P. and Naito, S. (1995) A Regulatory Role for the Abi3 Gene in the Establishment of Embryo Maturation in Arabidopsis-Thaliana. *Development* **121**, 629-636.

Nambara, E., Naito, S. and Mccourt, P. (1992) A Mutant of Arabidopsis Which Is Defective in Seed Development and Storage Protein Accumulation Is a New Abi3 Allele. *Plant Journal* **2**, 435-441.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J. and Finn, R.D. (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* **43**, D130-D137.

Nawrocki, E.P. and Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935.

Nelson, J.C., Saven, J.G., Moore, J.S. and Wolynes, P.G. (1997) Solvophobically driven folding of nonbiological oligomers. *Science* **277**, 1793-1796.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**, 268-274.

Nguyen, T.P., Keizer, P., van Eeuwijk, F., Smeekens, S. and Bentsink, L. (2012) Natural Variation for Seed Longevity and Seed Dormancy Are Negatively Correlated in Arabidopsis. *Plant Physiol* **160**, 2083-2092.

Nowell, R.W., Almeida, P., Wilson, C.G., Smith, T.P., Fontaneto, D., Crisp, A., Micklem, G., Tunnacliffe, A., Boschetti, C. and Barraclough, T.G. (2018) Comparative genomics of bdelloid rotifers: Insights from desiccating and nondesiccating species. *Plos Biol* **16**.

Nuccio, M.L., Paul, M., Bate, N.J., Cohn, J. and Cutler, S.R. (2018) Where are the drought tolerant crops? An assessment of more than two decades of plant biotechnology effort in crop improvement. *Plant Sci* **273**, 110-119.

Oldfield, C.J., Cheng, Y.G., Cortese, M.S., Romero, P., Uversky, V.N. and Dunker, A.K. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry-Us* **44**, 12454-12470.

Oliver, M.J., Guo, L.N., Alexander, D.C., Ryals, J.A., Wone, B.W.M. and Cushman, J.C. (2011) A Sister

Group Contrast Using Untargeted Global Metabolomic Analysis Delineates the Biochemical Regulation Underlying Desiccation Tolerance in Sporobolus stapfianus. *The Plant cell* **23**, 1231-1248.

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

Oliver, M.J., Velten, J. and Mishler, B.D. (2005) Desiccation tolerance in bryophytes: A reflection of the primitive strategy for plant survival in dehydrating habitats? *Integr Comp Biol* **45**, 788-799.

Olsen, J.L., Rouze, P., Verhelst, B., Lin, Y.C., Bayer, T., Collen, J., Dattolo, E., De Paoli, E., Dittami, S., Maumus, F., Michel, G., Kersting, A., Lauritano, C., Lohaus, R., Topel, M., Tonon, T., Vanneste, K., Amirebrahimi, M., Brakel, J., Bostrom, C., Chovatia, M., Grimwood, J., Jenkins, J.W., Jueterbock, A., Mraz, A., Stam, W.T., Tice, H., Bornberg-Bauer, E., Green, P.J., Pearson, G.A., Procaccini, G., Duarte, C.M., Schmutz, J., Reusch, T.B.H. and Van de Peer, Y. (2016) The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature* **530**, 331-+.

Olvera-Carrillo, Y., Campos, F., Reyes, J.L., Garciarrubio, A. and 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 Physiol* **154**, 373-390.

Olvera-Carrillo, Y., Luis Reyes, J. and Covarrubias, A.A. (2011) Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments. *Plant signaling & behavior* **6**, 586-589.

Ooms, J., Leon-Kloosterziel, K.M., Bartels, D., Koornneef, M. and Karssen, C.M. (1993a) Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants). *Plant Physiol* **102**, 1185-1191.

Ooms, J.J.J., Leonkloosterziel, K.M., Bartels, D., Koornneef, M. and Karssen, C.M. (1993b) Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis-Thaliana - a Comparative-Study Using Abscisic Acid-Insensitive Abi3 Mutants. *Plant Physiol* **102**, 1185-1191.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997) CATH - a hierarchic classification of protein domain structures. *Structure* **5**, 1093-1108.

Osakabe, Y., Osakabe, K., Shinozaki, K. and Tran, L.S.P. (2014) Response of plants to water stress. *Front Plant Sci* **5**.

Panchy, N., Lehti-Shiu, M. and Shiu, S.H. (2016) Evolution of Gene Duplication in Plants. *Plant Physiol* **171**, 2294-2316.

Pedrosa, A.M., Martins Cde, P., Goncalves, L.P. and Costa, M.G. (2015) Late Embryogenesis Abundant (LEA) Constitutes a Large and Diverse Family of Proteins Involved in Development and Abiotic Stress Responses in Sweet Orange (Citrus sinensis L. Osb.). *PloS one* **10**, e0145785.

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. and Yeates, T.O. (1999) Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4285-4288.

Piovesan, D., Tabaro, F., Micetic, I., Necci, M., Quaglia, F., Oldfield, C.J., Aspromonte, M.C., Davey, N.E., Davidovic, R., Dosztanyi, Z., Elofsson, A., Gasparini, A., Hatos, A., Kajava, A.V., Kalmar, L., Leonardi, E., Lazar, T., Macedo-Ribeiro, S., Macossay-Castillo, M., Meszaros, A., Minervini, G., Murvai, N., Pujols, J., Roche, D.B., Salladini, E., Schad, E., Schramm, A., Szabo, B., Tantos, A., Tonello, F., Tsirigos, K.D., Veljkovic, N., Ventura, S., Vranken, W., Warholm, P., Uversky, V.N., Dunker, A.K., Longhi, S., Tompa, P. and Tosatto, S.C.E. (2017) DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res* **45**, D219-D227.

Polanyi, M. (1968) Lifes Irreducible Structure. *Science* **160**, 1308-&.

Popova, A.V., Hundertmark, M., Seckler, R. and Hincha, D.K. (2011) Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the presence of membranes. *Bba-Biomembranes* **1808**, 1879-1887.

Popova, A.V., Rausch, S., Hundertmark, M., Gibon, Y. and Hincha, D.K. (2015) The intrinsically disordered protein LEA7 from Arabidopsis thaliana protects the isolated enzyme lactate dehydrogenase and enzymes in a soluble leaf proteome during freezing and drying. *Bba-Proteins Proteom* **1854**, 1517-1525.

Porembski, S. (2011) Evolution, diversity, and habitats of poikilohydrous vascular plants. In *Plant desiccation tolerance*: Springer, pp. 139-156.

Porembski, S. and Barthlott, W. (2000) Granitic and gneissic outcrops (inselbergs) as centers of diversity for desiccation-tolerant vascular plants. *Plant Ecol* **151**, 19-28.

Powell, A.A. and Matthews, S. (1984) Application of the Controlled Deterioration Vigor Test to Detect Seed Lots of Brussels-Sprouts with Low Potential for Storage under Commercial Conditions. *Seed Sci Technol* **12**, 649-657.

Rajjou, L. and Debeaujon, I. (2008) Seed longevity: Survival and maintenance of high germination ability of dry seeds. *Cr Biol* **331**, 796-805.

Rajjou, L., Lovigny, Y., Groot, S.P.C., Belghaz, M., Job, C. and Job, D. (2008) Proteome-wide characterization of seed aging in Arabidopsis: A comparison between artificial and natural aging protocols. *Plant Physiol* **148**, 620-641.

Rakhra, G., Kaur, T., Vyas, D., Sharma, A.D., Singh, J. and Ram, G. (2017) Molecular cloning, characterization, heterologous expression and in-silico analysis of disordered boiling soluble stress-responsive wBsSRP protein from drought tolerant wheat cv.PBW 175. *Plant Physiol Bioch* **112**, 29-44.

Receveur-Brechot, V., Bourhis, J.M., Uversky, V.N., Canard, B. and Longhi, S. (2006) Assessing protein disorder and induced folding. *Proteins* **62**, 24-45.

Reed, R.G., Feldhoff, R.C., Clute, O.L. and Peters, T. (1975) Fragments of Bovine Serum-Albumin Produced by Limited Proteolysis - Conformation and Ligand-Binding. *Biochemistry-Us* **14**, 4578-4583.

Reis, P.A.A., Rosado, G.L., Silva, L.A.C., Oliveira, L.C., Oliveira, L.B., Costa, M.D.L., Alvim, F.C. and Fontes, E.P.B. (2011) The Binding Protein BiP Attenuates Stress-Induced Cell Death in Soybean via Modulation of the N-Rich Protein-Mediated Signaling Pathway. *Plant Physiol* **157**, 1853-1865.

Rensing, S.A. (2017) Why we need more non-seed plant models. *New Phytologist* **216**, 355-360.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L.X., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S. and Boore, J.L. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64-69.

Reyes, J.L., Campos, F., Wei, H., Arora, R., Yang, Y.I., Karlson, D.T. and Covarrubias, A.A. (2008) Functional dissection of Hydrophilins during in vitro freeze protection. *Plant Cell and Environment* **31**, 1781-1790.

Riera, M., Figueras, M., Lopez, C., Goday, A. and Pages, M. (2004) Protein kinase CK2 modulates developmental functions of the abscisic acid responsive protein Rab17 from maize. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9879-9884.

Rivera-Najera, L.Y., Saab-Rincon, G., Battaglia, M., Amero, C., Pulido, N.O., Garcia-Hernandez, E., Solorzano, R.M., Reyes, J.L. and Covarrubias, A.A. (2014) A Group 6 Late Embryogenesis Abundant Protein from Common Bean Is a Disordered Protein with Extended Helical Structure and Oligomer-forming Properties. *J Biol Chem* **289**, 31995-32009.

Rodriguez, M.C.S., Edsgard, D., Hussain, S.S., Alquezar, D., Rasmussen, M., Gilbert, T., Nielsen, B.H., Bartels, D. and Mundy, J. (2010) Transcriptomes of the desiccation-tolerant resurrection plant Craterostigma plantagineum. *Plant Journal* **63**, 212-228.

Rosvall, M. and Bergstrom, C.T. (2008) Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1118-1123.

Ruibal, C., Salamo, I.P., Carballo, V., Castro, A., Bentancor, M., Borsani, O., Szabados, L. and Vidal, S. (2012) Differential contribution of individual dehydrin genes from Physcomitrella patens to salt and osmotic stress tolerance. *Plant Sci* **190**, 89-102.

Sallon, S., Solowey, E., Cohen, Y., Korchinsky, R., Egli, M., Woodhatch, I., Simchoni, O. and Kislev, M. (2008) Germination, genetics, and growth of an ancient date seed. *Science* **320**, 1464-1464.

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M. and Lepiniec, L. (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant Journal* **54**, 608-620.

Saucedo, A.L., Hernandez-Dominguez, E.E., de Luna-Valdez, L.A., Guevara-Garcia, A.A., Escobedo-Moratilla, A., Bojorquez-Velazquez, E., Del Rio-Portilla, F., Fernandez-Velasco, D.A. and Barba de la Rosa, A.P. (2017) Insights on Structure and Function of a Late Embryogenesis Abundant Protein from Amaranthus cruentus: An Intrinsically Disordered Protein Involved in Protection against Desiccation, Oxidant Conditions, and Osmotic Stress. *Front Plant Sci* **8**, 497.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. and Lohmann, J.U. (2005) A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37**, 501.

Scott, P. (2000) Resurrection plants and the secrets of eternal leaf. *Ann Bot-London* **85**, 159-166.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504.

Sharma, A., Kumar, D., Kumar, S., Rampuria, S., Reddy, A.R. and Kirti, P.B. (2016) Ectopic Expression of an Atypical Hydrophobic Group 5 LEA Protein from Wild Peanut, Arachis diogoi Confers Abiotic Stress Tolerance in Tobacco. *PloS one* **11**.

Shen-Miller, J. (2002) Sacred lotus, the long-living fruits of China Antique. *Seed Sci Res* **12**, 131-143.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A. and Waterston, R.H. (2017) DNA sequencing at 40: past, present and future. *Nature* **550**, 345-353.

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

Shi, J., Liu, M.Q., Chen, Y.Z., Wang, J.Y. and Lu, C.F. (2016) Heterologous expression of the dehydrin-

like protein gene AmCIP from Ammopiptanthus mongolicus enhances viability of Escherichia coli and tobacco under cold stress. *Plant Growth Regul* **79**, 71-80.

Shi, Z.S., Olson, C.A., Rose, G.D., Baldwin, R.L. and Kallenbach, N.R. (2002) Polyproline II structure in a sequence of seven alanine residues. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9190-9195.

Shih, M.D., Hoekstra, F.A. and Hsing, Y.I.C. (2008) Late Embryogenesis Abundant Proteins. *Advances in Botanical Research, Vol 48* **48**, 211-255.

Shih, M.D., Hsieh, T.Y., Jian, W.T., Wu, M.T., Yang, S.J., Hoekstra, F.A. and Hsing, Y.I.C. (2012) Functional studies of soybean (Glycine max L.) seed LEA proteins GmPM6, GmPM11, and GmPM30 by CD and FTIR spectroscopy. *Plant Sci* **196**, 152-159.

Shih, M.D., Hsieh, T.Y., Lin, T.P., Hsing, Y.I.C. and Hoekstra, F.A. (2010) Characterization of Two Soybean (Glycine max L.) LEA IV Proteins by Circular Dichroism and Fourier Transform Infrared Spectrometry. *Plant and Cell Physiology* **51**, 395-407.

Shinde, S., Nurul Islam, M. and Ng, C.K. (2012) Dehydration stress-induced oscillations in LEA protein transcripts involves abscisic acid in the moss, Physcomitrella patens. *The New phytologist* **195**, 321-328.

Simon, L.M., Kotorman, M., Garab, G. and Laczko, I. (2001) Structure and activity of alpha-chymotrypsin and trypsin in aqueous organic media. *Biochemical and biophysical research communications* **280**, 1367-1371.

Singh, S., Cornilescu, C.C., Tyler, R.C., Cornilescu, G., Tonelli, M., Lee, M.S. and Markley, J.L. (2005) Solution structure of a late embryogenesis abundant protein (LEA14) from Arabidopsis thaliana, a cellular stress-related protein. *Protein Sci* **14**, 2601-2609.

Skubacz, A., Daszkowska-Golec, A. and Szarejko, I. (2016) The Role and Regulation of ABI5 (ABA-Insensitive 5) in Plant Development, Abiotic Stress Responses and Phytohormone Crosstalk. *Front Plant Sci* **7**.

Smarda, P., Bures, P., Horova, L., Leitch, I.J., Mucina, L., Pacini, E., Tichy, L., Grulich, V. and Rotreklova, O. (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4096-E4102.

Smirnoff, N. (1993) Tansley Review .52. The Role of Active Oxygen in the Response of Plants to Water-Deficit and Desiccation. *New Phytologist* **125**, 27-58.

Smit, A.F.A., Hubley, R. and Green, P. (2008)

Smith, S.A. and Dunn, C.W. (2008) Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**, 715-716.

Song, L., Florea, L. and Langmead, B. (2014) Lighter: fast and memory-efficient sequencing error correction without counting. *Genome biology* **15**.

Soulages, J.L., Kim, K., Arrese, E.L., Walters, C. and Cushman, J.C. (2003) Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly (L-proline)-type II structure. *Plant Physiol* **131**, 963-975.

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

Stevenson, S.R., Kamisugi, Y., Trinh, C.H., Schmutz, J., Jenkins, J.W., Grimwood, J., Muchero, W.,

Tuskan, G.A., Rensing, S.A., Lang, D., Reski, R., Melkonian, M., Rothfels, C.J., Li, F.W., Larsson, A., Wong, G.K.S., Edwards, T.A. and Cuming, A.C. (2016) Genetic Analysis of Physcomitrella patens Identifies ABSCISIC ACID NON-RESPONSIVE, a Regulator of ABA Responses Unique to Basal Land Plants and Required for Desiccation Tolerance. *The Plant cell* **28**, 1310-1327.

Strimbeck, G.R. (2017) Hiding in plain sight: the F segment and other conserved features of seed plant SKn dehydrins. *Planta* **245**, 1061-1066.

Sugliani, M., Rajjou, L., Clerkx, E.J.M., Koornneef, M. and Soppe, W.J.J. (2009) Natural modifiers of seed longevity in the Arabidopsis mutants abscisic acid insensitive3-5 (abi3-5) and leafy cotyledon1-3 (lec1-3). *New Phytologist* **184**, 898-908.

Sun, X.L., Rikkerink, E.H.A., Jones, W.T. and Uversky, V.N. (2013) Multifarious Roles of Intrinsic Disorder in Proteins Illustrate Its Broad Impact on Plant Biology. *The Plant cell* **25**, 38-55.

Swain, S., Mondal, D., Beg, S., Niranjan Patra, C., Chandra Dinda, S., Sruti, J. and Eswara Bhanoji Rao, M. (2013) Stabilization and delivery approaches for protein and peptide pharmaceuticals: an extensive review of patents. *Recent patents on biotechnology* **7**, 28-46.

Szinay, D., Chang, S.B., Khrustaleva, L., Peters, S., Schijlen, E., Bai, Y.L., Stiekema, W.J., van Ham, R.C.H.J., de Jong, H. and Lankhorst, R.M.K. (2008) High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. *Plant Journal* **56**, 627-637.

Todaka, D., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2015) Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front Plant Sci* **6**.

Tolleter, D., Hincha, D.K. and Macherel, D. (2010) A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *Bba-Biomembranes* **1798**, 1926-1933.

Tolleter, D., Jaquinod, M., Mangavel, C., Passirani, C., Saulnier, P., Manon, S., Teyssier, E., Payet, N., Avelange-Macherel, M.H. and Macherel, D. (2007) Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *The Plant cell* **19**, 1580-1589.

Tompa, P. (2002) Intrinsically unstructured proteins. *Trends in biochemical sciences* **27**, 527-533.

Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *Febs Lett* **579**, 3346-3354.

Tompa, P. (2012) Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences* **37**, 509-516.

Tompa, P., Banki, P., Bokor, M., Kamasa, P., Kovacs, D., Lasanda, G. and Tompa, K. (2006) Protein-water and protein-buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects. *Biophys J* **91**, 2243-2249.

Toufighi, K., Brady, S.M., Austin, R., Ly, E. and Provart, N.J. (2005) The Botany Array Resource: e-Northerns, Expression Angling, and Promoter analyses. *Plant Journal* **43**, 153-163.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578.

Tuba, Z., Lichtenthaler, H.K., Maroti, I. and Csintalan, Z. (1993) Resynthesis of Thylakoids and Functional Chloroplasts in the Desiccated Leaves of the Poikilochlorophyllous Plant Xerophyta-Scabrida Upon Rehydration. *Journal of plant physiology* **142**, 742-748.

Tuba, Z., Proctor, M.C.F. and Csintalan, Z. (1998) Ecophysiological responses of homoiochlorophyllous and poikilochlorophyllous desiccation tolerant plants: a comparison and an ecological perspective (vol 24, pg 211, 1998). *Plant Growth Regul* **26**, 71-71.

Tunnacliffe, A., Hincha, D.K., Leprince, O. and Macherel, D. (2010) LEA proteins: versatility of form and function. In *Dormancy and resistance in harsh environments*: Springer, pp. 91-108.

Tunnacliffe, A., Lapinski, J. and McGee, B. (2005) A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. *Hydrobiologia* **546**, 315-321.

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

Uversky, V.N. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci* **11**, 739-756.

Uversky, V.N. (2009) Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. *Protein J* **28**, 305-325.

Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics* **41**, 415-427.

Van de Peer, Y. and Pires, J.C. (2018) Editorial overview: Genome studies and molecular genetics: Treasure troves of evolution. *Curr Opin Plant Biol* **42**, Iii-V.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E. and Babu, M.M. (2014) Classification of Intrinsically Disordered Regions and Proteins. *Chem Rev* **114**, 6589-6631.

VanBuren, R., Bryant, D., Edger, P.P., Tang, H.B., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E., Freeling, M., Bartels, D., Ten Hallers, B., Hastie, A., Michael, T.P. and Mockler, T.C. (2015) Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. *Nature* **527**, 508-U209.

VanBuren, R., Wai, C.M., Ou, S.J., Pardo, J., Bryant, D., Jiang, N., Mockler, T.C., Edger, P. and Michael, T.P. (2018) Extreme haplotype variation in the desiccation-tolerant clubmoss Selaginella lepidophylla. *Nat Commun* **9**.

VanBuren, R., Wai, C.M., Zhang, Q.W., Song, X.M., Edger, P.P., Bryant, D., Michael, T.P., Mockler, T.C. and Bartels, D. (2017) Seed desiccation mechanisms co-opted for vegetative desiccation in the resurrection grass Oropetium thomaeum. *Plant Cell and Environment* **40**, 2292-2306.

VanderEycken, W., Engler, J.D., Inze, D., VanMontagu, M. and Gheysen, G. (1996) A molecular study of root-knot nematode-induced feeding sites. *Plant Journal* **9**, 45-54.

Varma, R.S. (2012) Greener approach to nanomaterials and their sustainable applications. *Curr Opin Chem Eng* **1**, 123-128.

Venyaminov, S.Y., Baikalov, I.A., Shen, Z.M., Wu, C.S.C. and Yang, J.T. (1993) Circular Dichroic Analysis of Denatured Proteins - Inclusion of Denatured Proteins in the Reference Set. *Anal Biochem* **214**, 17-24.

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., Leprince, O., Chatelain, E., Vu, B.L., Gouzy, J., Gamas, P., Udvardi, M.K. and Buitink, J. (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. *Plant Physiol* **163**, 757-774.

Vicente-Carbajosa, J. and Carbonero, P. (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol* **49**, 645-651.

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**, 1329-1340.

Virkutyte, J. and Varma, R.S. (2011) Green synthesis of metal nanoparticles: Biodegradable polymers and enzymes in stabilization and surface functionalization. *Chem Sci* **2**, 837-846.

Vitt, H.T., Crandall-Stotler, B. and Wood, A. (2014) Bryophytes: Survival in a dry world through tolerance and avoidance. In *Plant ecology and evolution in harsh environments*. New York, NY, USA: Nova Science, pp. 276-295.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.D., Wortman, J., Young, S.K. and Earl, A.M. (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PloS one* **9**.

Wan, C.Y. and Wilkins, T.A. (1994) A Modified Hot Borate Method Significantly Enhances the Yield of High-Quality Rna from Cotton (Gossypium-Hirsutum L). *Anal Biochem* **223**, 7-12.

Wang, X.S., Zhu, H.B., Jin, G.L., Liu, H.L., Wu, W.R. and Zhu, J. (2007) Genome-scale identification and analysis of LEA genes in rice (Oryza sativa L.). *Plant Sci* **172**, 414-420.

Wang, X.Y., Zhang, L.S., Zhang, Y.N., Bai, Z.Q., Liu, H. and Zhang, D.P. (2017) Triticum aestivum WRAB18 functions in plastids and confers abiotic stress tolerance when overexpressed in Escherichia coli and Nicotiania benthamiana. *PloS one* **12**.

Wang, Y.P., Tang, H.B., DeBarry, J.D., Tan, X., Li, J.P., Wang, X.Y., Lee, T.H., Jin, H.Z., Marler, B., Guo, H., Kissinger, J.C. and Paterson, A.H. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49-e49.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.

Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738.

Williams, B., Njaci, I., Moghaddam, L., Long, H., Dickman, M.B., Zhang, X.R. and Mundree, S. (2015) Trehalose Accumulation Triggers Autophagy during Plant Desiccation. *Plos Genet* **11**.

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

Wise, M.J. and Tunnacliffe, A. (2004) POPP the question: what do LEA proteins do? *Trends Plant Sci* **9**, 13-17.

Wodniok, S., Brinkmann, H., Glockner, G., Heidel, A.J., Philippe, H., Melkonian, M. and Becker, B. (2011) Origin of land plants: Do conjugating green algae hold the key? *BMC evolutionary biology* **11**, 104.

Xiao, B., Huang, Y., Tang, N. and Xiong, L. (2007) Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *Theor Appl Genet* **115**, 35-46.

Xiao, L.H., Yang, G., Zhang, L.C., Yang, X.H., Zhao, S., Ji, Z.Z., Zhou, Q., Hu, M., Wang, Y., Chen, M., Xu, Y., Jin, H.J., Xiao, X., Hu, G.P., Bao, F., Hu, Y., Wan, P., Li, L.G., Deng, X., Kuang, T.Y., Xiang, C.B., Zhu, J.K., Oliver, M.J. and He, Y.K. (2015) The resurrection genome of Boea hygrometrica: A blueprint for survival of dehydration. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 5833-5837.

Xu, Z.C., Xin, T.Y., Bartels, D., Li, Y., Gu, W., Yao, H., Liu, S., Yu, H.Y., Pu, X.D., Zhou, J.G., Xu, J., Xi, C.C., Lei, H.T., Song, J.Y. and Chen, S.L. (2018) Genome Analysis of the Ancient Tracheophyte Selaginella tamariscina Reveals Evolutionary Features Relevant to the Acquisition of Desiccation Tolerance. *Mol Plant* **11**, 983-994.

Yasui, Y., Hirakawa, H., Oikawa, T., Toyoshima, M., Matsuzaki, C., Ueno, M., Mizuno, N., Nagatoshi, Y., Imamura, T., Miyago, M., Tanaka, K., Mise, K., Tanaka, T., Mizukoshi, H., Mori, M. and Fujita, Y. (2016) Draft genome sequence of an inbred line of Chenopodium quinoa, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Res* **23**, 535-546.

Ye, C.X., Hill, C.M., Wu, S.G., Ruan, J. and Ma, Z.S. (2016) DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Scientific reports* **6**.

Ye, C.X., Ma, Z.S.S., Cannon, C.H., Pop, M. and Yu, D.W. (2012) Exploiting sparseness in de novo genome assembly. *Bmc Bioinformatics* **13**.

Yee, A.A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., Evdokimova, E., Liu, C.S., Semesi, A. and Guido, V. (2005) NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *Journal of the American Chemical Society* **127**, 16512-16517.

Yoshida, K., Schuenemann, V.J., Cano, L.M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F.N., Kamoun, S., Krause, J., Thines, M., Weigel, D. and Burbano, H.A. (2013) The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine (vol 2, e00731, 2013). *Elife* **2**.

Yotsui, I., Serada, S., Naka, T., Saruhashi, M., Taji, T., Hayashi, T., Quatrano, R.S. and Sakata, Y. (2016) Large-scale proteome analysis of abscisic acid and ABSCISIC ACID INSENSITIVE3-dependent proteins related to desiccation tolerance in Physcomitrella patens. *Biochemical and biophysical research communications* **471**, 589-595.

Yu, J., Lai, Y.M., Wu, X., Wu, G. and Guo, C.K. (2016) Overexpression of OsEm1 encoding a group I LEA protein confers enhanced drought tolerance in rice. *Biochemical and biophysical research communications* **478**, 703-709.

Zhang, X., Lu, S.C., Jiang, C.H., Wang, Y.F., Lv, B., Shen, J.B. and Ming, F. (2014) RcLEA, a late embryogenesis abundant protein gene isolated from Rosa chinensis, confers tolerance to Escherichia coli and Arabidopsis thaliana and stabilizes enzyme activity under diverse stresses. *Plant molecular biology* **85**, 333-347.

Zhang, Z., Kuipers, G., Niemiec, L., Baumgarten, T., Slotboom, D.J., de Gier, J.W. and Hjelm, A. (2015) High-level production of membrane proteins in E-coli BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact* **14**.

Zhao, P.S., Liu, F., Zheng, G.C. and Liu, H. (2011) Group 3 late embryogenesis abundant protein in Arabidopsis: structure, regulation, and function. *Acta Physiol Plant* **33**, 1063-1073.

Zhao, T., Holmer, R., de Bruijn, S., Angenent, G.C., van den Burg, H.A. and Schranz, M.E. (2017) Phylogenomic Synteny Network Analysis of MADS-Box Transcription Factor Genes Reveals Lineage-Specific Transpositions, Ancient Tandem Duplications, and Deep Positional Conservation. *The Plant cell* **29**, 1278-1292.

Zhao, T. and Schranz, M.E. (2017) Network approaches for plant phylogenomic synteny analysis. *Curr Opin Plant Biol* **36**, 129-134.

Zhao, Y.A., Tang, H.X. and Ye, Y.Z. (2012) RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28**, 125-126.

Zhou, Y., He, P., Xu, Y.P., Liu, Q., Yang, Y.G. and Liu, S.Q. (2017) Overexpression of CsLEA11, a Y3SK2-type dehydrin gene from cucumber (Cucumis sativus), enhances tolerance to heat and cold in Escherichia coli. *Amb Express* **7**.

Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Vu, B.L., Dubois-Laurent, C., Geoffriau, E., Le Signor, A., Dalmais, M., Gutbrod, K., Dormann, P., Gallardo, K., Bendahmane, A., Buitink, J. and Leprince, O. (2016) ABI5 Is a Regulator of Seed Maturation and Longevity in Legumes. *The Plant cell* **28**, 2735-2754.

Zou, C., Lehti-Shiu, M.D., Thomashow, M. and Shiu, S.H. (2009) Evolution of Stress-Regulated Gene Expression in Duplicate Genes of Arabidopsis thaliana. *Plos Genet* **5**, e1000581.

Acknowledgements

I believe that 'gratitude brings beauty into one's life'. I would like to thank many people that were responsible for this very important accomplishment in my personal and professional life.

I thank my dear supers Henk and Wilco. Hilhorstinho, you have always been so supportive and friendly, your open mind about science helped me to develop some characteristics that I am extremely proud of today (such as creativity and passion for science). You let me be creative and innovative, and was always there to support me in the moments of frustration and weakness. Wilcoooo(!!), your joy was always inspiring me, your knowledge and humbleness made me feel confident and comfortable to go into your room and 'bother' you every time it would come into my mind that I needed your help, from technical assistance to the most complicated scientific design and insights. I am very thankful to you both for trusting me and for opening the doors of your research group and your offices for my thousands doubts and life inquiries. It means a lot to me to have met you both, to work with you and to have got your friendship, and will always bring what I have learned with you both in my personal life and professional career. I wish you all the best in your lives and hope to maintain our friendship and professional relationship still alive and shining!

I also want to thank José Márcio for feeding my imagination about how great it would be to live and to do a PhD in The Netherlands. Indeed Zé, that was the best choice that I could have made by that time in my life. You are one of the responsible ones for all this that has happened to me! I thank you very much for encouraging me and also for trusting me.

I would like to thank all members of the PPH group for the amazing time together, for the nice work discussions, for the company during conferences and symposia, and for the fun moments during our lab outings. I would like to thank Rina and Margaret for all the help and for making sure that all would flow nicely in PPH to all of us. I also would like to express my special acknowledgments to the members of the Seed Lab. I enjoyed our Friday meetings a lot (even though I used to be pretty sleepy almost always) and I have learned so much from all of you. Thank you Leo and Bas for being so patient with all my questions in the lab and for all the support, I learned a lot from you both and wish you all the best. Leónie, you have not only been a great colleague but also an inspiring woman. I am extremely happy to have had the opportunity to work with you, and also to count on your support when I was looking for a job. Your advices on thinking about my career were very valuable for me. I also would like to thank Mahdere and Nafiseh for being so friendly, for listening to my

billion complaints and for the warm advices after hours of my personal drama. Elisinha and Gondinha, dear fijne dinners, I am very thankful for your care for me and your support. I am very happy that we became more than colleagues, and we strengthen our friendship with our regular fijne dinning meetings once in a while. We will definitely keep in touch, and I wish you all the best!

Talking about woman power, I thank you so much Jill for opening the doors of your lab to me and for inspiring me to love science. Your love for what you do and for your country inspires me, and I am so happy that I had the opportunity to work with you. I wish you all the best with all my heart. I extend my thanks to the lab members that were so helpful and careful with me during my time in South Africa and afterwards, especially Tim, Astrid, Jeanne, Keren and Hal.

I also would like to thank the Brazilians that helped me to settle down in Wageningen and that shared a bit of our culture, even though we were so far from home. I would like to specially thank Deborah, Anderson, Paulo, Alexandre, Anabele, Anna Patrícya, Joana, Julio and Milene. I could count on you to improve my working skills as well as my social skills when I arrived in Wageningen, thank you so much for everything. I hope we'll keep in touch and wish you all the best.

To my dear paranymphs Renake and Juriaan, I couldn't chose differently. Renatinha, I am very thankful for all your support and your loyalty. It means a lot to me to have your friendship. Your kindness and strength have always inspired me. Thank you for listening to all my dramas and still giving me comfort and the wisest advises ever! I hope you can count on me as much as I could count on you during all these years. I wish you all the best and hope we still can share a lot of special moments in our lives! Juliaaan!!! You have been more than a colleague, collaborator, teacher, life-guru, you became a super friend! I am so happy that I got to know the great person you are. I will never forget our lab discussions and how we used to imagine the craziest sci-fi situations planning just a simple experiment. I have learned a lot from you, not only that nerd people can also be cool (haha!) but that there is always something cool to learn from my own mistakes (even the most silly cloning ones!). I wish you a lot of success in your career, I am sure you will do great! Thank you both for your friendship, I will bring that with me forever!

I also would like to thank Eric Schranz and Tao Zhao. Eric, thank you very much for opening yourself for collaborating with us, even though I had no clue about what synteny was. Even so busy you would manage to open your doors for meetings with me. Tao, I am so happy that our collaboration became a friendship. I am very thankful for all I have learned from you. Our brainstorms of sometimes 2 hours helped me to develop a lot of my ideas. I am very happy that even with some delay we managed to

combine our research fields and describe a nice story on LEAs evolution based on synteny network analysis. You guys are wonderful scientists! I wish you a lot of success!

I thank all my dear Pomonians and 'aggregates' for making my 4-year-PhD life feel like I was at home. I specially thank LooWee, Cris, Lieke, Heleen, Marlise, Siew Ling, Sacha, Arely, Diana, Jane and Simon. I am so happy that I met you all, learned about your culture and shared a bit of mine with you. I learned a lot about being friendly and social, and I will never forget our funny memories together.

Eu também agradeço aos meus queridos pretins amigos Leandro, Thiago, Jéssica, Caio, Bárbara e Léo por mantermos nossa amizade mesmo à distância, isso é muito valioso para mim! O tempo, a distância e a saudade jamais diminuíram o meu carinho e consideração por vocês!!! Obrigada por estarem sempre comigo meus queridos. Eu desejo tudo o de melhor na vida de vocês, e espero estarmos sempre presentes na vida uns dos outros! Às queridas Joana, Ana Luísa, e Tininha, obrigada pelo carinho e amizade desde nosso tempo de UFLA, sou muito grata à amizade e por poder contar com vocês sempre.

My English vocabulary is not large enough to find the words to express my love and gratitude to you Jordi, but I will try! You have been an essential person for me during my PhD. I don't like to say that I am lucky (as you know) but the truth is that I feel so lucky to have you by my side, to know that I can count on you for everything. More than my boyfriend you became my best friend. I really love and appreciate your loyalty and honesty. Knowing you were there to support me made me feel secure and strong to stand many difficult moments during these 4 years. I wish I can be there for you as much as you have been there for me, I really love who you are simply for being who you are. Thank you so much! Even though you were not in my initial PhD plan, I dedicate this thesis also to you, you were essential for it! I also extend my thanks to your family for welcoming me at your home and for the nice celebrations together.

Eu agradeço imensamente à minha mãe (Dodora) e ao meu pai (Zé Cuti). Eu sei que não foi fácil para vocês compreenderem o por que eu estava saindo de casa, pela primeira vez para tão longe dos seus cuidados. Eu espero que hoje vocês possam compreender e se sentir orgulhosos de mim assim como eu me sinto por ter vocês como meus pais. Apesar da distância vocês sempre estiveram perto, e me deram suporte e confortaram o meu coração nos momentos de saudade e angústia. Muito obrigada por me apoiarem desde sempre com os meus estudos, por me estimularem e por puxarem minha orelha na hora de parar e descansar mentalmente. Eu dedico tudo isso o que eu alcancei aqui nesses quatro anos à vida e ao trabalho de vocês. O meu carinho e dedicação também aos meus irmãos Natália e Leandro (e suas famílias) pela torcida e apoio. Naty, você é mais que uma irmã, você é uma das minhas melhores amigas. Muito obrigada por seu meu apoio em tantos momentos, eu sou muito grata

por poder contar com você nos momentos difíceis mesmo de longe. Você é uma super mulher e uma super mamãe, e você me inspira a ser forte todos os dias! Eu também dedico essa tese aos meus três sobrinhos amados Renan, Francesco e Ravi; eu espero que um dia vocês possam se orgulhar da sua titia, e possam contar comigo pro que precisar! Amo muito vocês, minha família!!!

Eu também agradeço aos meus amigos da 'Vila' e da UFLA, aos meus familiares, ao meu Padrinho Clésio e minha Madrinha Sandra, e tantos outros cujos nomes que eu não mencionei. Muito obrigada pelo carinho e pela torcida!

Eu agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brasil) e ao Governo Brasileiro pelo auxílio financeiro para a condução e conclusão do trabalho apresentado nessa tese. I also would like to thank Wageningen University and Research for the opportunity to do my PhD, and the thesis committee members, opponents and reviewers that contributed directly and indirectly to my professional development.

When I came to The Netherlands my one time goal was to accomplish a PhD. Today I look back and I see that I have accomplished many other things which are as important and valuable as the PhD diploma. The friends that I have made and the memories that I have shared with each of you will always fill my heart with joy and love!

Thank you!
Muito obrigada!

Com carinho, Mari

About the author

Mariana Aline Silva Artur was born in Lavras, Minas Gerais, Brazil on the 20th of July 1989. Daughter of a bricklayer and a housewife, Mariana had studied in public schools until the year of 2005, when she obtained a sport-scholarship for playing volleyball to do her 3-years college education at the private college school Gammon Presbyterian Institute in Lavras. In 2008 Mariana was approved in the Brazilian selection system for public universities and started her Bachelor in Biology at the Federal University of Lavras (UFLA). Her first internship was in Plant Bacteriology and she has also worked as a monitor of the course of Plant Anatomy. Mariana got fascinated with the subjects of Plant Physiology and Plant Anatomy, and did a long term internship in the Laboratories of Plant Growth and Development and Plant Tissue Culture, working with different Brazilian native plant species. Right after her bachelor's graduation in February of 2012 she achieved the 1st position for a masters in Agronomy/Plant Physiology at the same University.

During her masters Mariana worked with Prof. Amauri Alves de Alvarenga and Prof. José Márcio Rocha Faria on seed physiology and loss of desiccation tolerance in *Erythrina falcata* seeds, a Brazilian native species. During her masters, Mariana met Dr. Henk Hilhorst who was a long-term collaborator of Prof. José Márcio Rocha Faria, and felt interested on applying for a grant to fund her PhD in the Netherlands in Henk's research team. In February of 2014 Mariana concluded her Masters in Agronomy/Plant Physiology and was granted with a scholarship of the Coordination for the Improvement of Higher Education Personnel (CAPES/Brasil) for a 4-years PhD.

In August of 2014 Mariana joined the Seed Lab and the Plant Physiology Group of Wageningen University in the Netherlands and started her PhD under the supervision of Dr. Henk Hilhorst and Dr. Wilco Ligterink. Her PhD thesis was focused on understanding plant desiccation tolerance at a genomic, physiological and molecular level in the resurrection plant species, *Xerophyta viscosa*. Her PhD research was mainly focused on investigating the evolution structure of function of stress protective proteins named Late Embryogenesis Abundant (LEA proteins). During her PhD Mariana collaborated with several researchers in different fields including Prof. Eric Schranz and Dr. Tao Zhao of the Pant Biosystematics group of Wageningen University, and Prof. Jill Farrant of the University of Cape Town in South Africa, where she worked for 3 months as a guest researcher. Mariana always enjoyed communicating science

and she participated and presented posters and lectures in several regional and international conferences. In 2018 Mariana was selected to give a talk and present a poster at the great prestige Gordon Research Seminar and Conference in Salt and Water Stress in Plants, respectively.

In September 2018 (one month before submitting her PhD thesis) Mariana joined the Plant Ecophysiology Group at Utrecht University in the Netherlands to work as a postdoc under the supervision of Dr. Kaisa Kajala to investigate the environmental tuning of plant development, specifically drought induced suberization in tomato roots.

Her passion for plant science grew faster since when she first saw a plant cell under a microscope and the coordinated chloroplast 'dance' with the cytoplasmic movement (cyclosis) in her first Plant Anatomy class in 2008 at UFLA. Since then, Mariana has been fascinated with how plants respond to their environment and has contributed with the communication of plant science research by publishing scientific manuscripts including the work performed in this thesis.

Publications

ARTUR M.A.S.*, ZHAO T.*, SCHRANZ M. E., LIGTERINK W., HILHORST H. Dissecting the genomic diversification of LATE EMBRYOGENESIS ABUNDANT (LEA) protein gene families in plants. Genome Biology and Evolution, 2018.

COSTA M.C.D.*, ARTUR M.A.S.*, MAIA J., JONKHEER E., DERKS M.F.L., NIJVEEN H., WILLIAMS B., MUNDREE S.G., JIMÉNEZ-GÓMEZ J.M., HESSELINK T., SCHIJLEN E.G., LIGTERINK W., OLIVER M.J., FARRANT J.M., HILHORST H.W.. A footprint of desiccation tolerance in the genome of Xerophyta viscosa. Nature Plants, v.3,17038, 2017.

CAMPOS, N.A.; PAIVA, R.; ARTUR, M.A.S.; SILVA, L.C. . Explant type and constitution of the capsules for gabirobeira (*Campomanesia pubescens*) synthetic seed production and storage. Plant Cell Culture & Micropropagation, v. 10, p. 13-19, 2014.

CHAVES, I. DE S.; ALVARENGA, A. A.; DOUSSEAU, S.; SOUZA, E. DOS S.; ARTUR, M. A. S.; LARA, T. S.. Morphological characterization of fruits, diaspores and germination of *Miconia ligustroides* (DC.) Naundim (Melastomataceae). Acta Scientiarum. Biological Sciences, v. 35, p. 93-98, 2013.

REIS, M.V.; PAIVA, P.D.O.; PORTO, J.M.P.; PAIVA, R.; ARTUR, M.A.S.. Cytochemical and ultrastructural analysis of pro-embryogenic masses in *Tabebuia roseo-alba*. Acta Horticulturae, v. 988, p. 179-186, 2013.

CHAVES, I. de S. ; ALVARENGA, A.A. ; DOUSSEAU, S.; SOARES, G.C.M. ; SOUZA, E. dos S. ; ARTUR, M.A.S.. Germination of *Miconia ligustroides* (DC.) Naudim (Melastomataceae) diaspores submitted to different treatments for dormancy overcoming. Brazilian Journal of Botany, v. 34, p. 335-341, 2011.

**These authors contributed equally to this work.*

Education Statement

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: **Mariana Aline Silva Artur**
Date: **11 January 2019**
Group: **Laboratory of Plant Physiology**
University: **Wageningen University & Research**

1) Start-Up Phase	*date*	*cp*
► **First presentation of your project**		
The Occurrence of Desiccation Tolerance in Higher Plants: Evolution and Phylogenetic Relationships of Late Embryogenesis Abundant (LEA) Proteins	08 Dec 2014	1.5
► **Writing or rewriting a project proposal**		
The Occurrence of Desiccation Tolerance in Higher Plants: Evolution and Phylogenetic Relationships of Late Embryogenesis Abundant (LEA) Proteins	2014-2015	6.0
► **Writing a review or book chapter**		
► **MSc courses**		
	Subtotal Start-Up Phase	7.5

2) Scientific Exposure	*date*	*cp*
► **EPS PhD student days**		
The Annual EPS PhD Students Day 'Get2Gether 2016, Soest, The Netherlands	28-29 Jan 2016	0.6
The Annual EPS PhD Students Day 'Get2Gether 2018, Soest, The Netherlands	15-16 Feb 2018	0.6
► **EPS theme symposia**		
EPS Theme 4 Symposium 'Genome Biology', Wageningen, The Netherlands	03 Dec 2014	0.3
EPS Theme 3 Symposium 'Metabolism and Adaptation', Wageningen, The Netherlands	10 Feb 2015	0.3
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen, The Netherlands	21 Jan 2016	0.3
EPS Theme 4 Symposium 'Genome Biology', Wageningen, The Netherlands	16 Dec 2016	0.3
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen, The Netherlands	30 Jan 2018	0.3
EPS Theme 3 Symposium 'Metabolism and Adaptation', Wageningen, The Netherlands	13 Mar 2018	0.3
► **National meetings (e.g. Lunteren days) and other National Platforms**		
Annual meeting NWO - ALW Lunteren, The Netherlands	13-14 Apr 2015	0.6
Annual meeting NWO - ALW Lunteren, The Netherlands	11-12 Apr 2016	0.6
Annual meeting NWO - ALW Lunteren, The Netherlands	10-11 Apr 2017	0.6
Annual meeting NWO - ALW Lunteren, The Netherlands	09-10 Apr 2018	0.6
► **Seminars (series), workshops and symposia**		
Symposium: Omics Advances for Academia and Industry, Wageningen, The Netherlands	11 Dec 2014	0.3
Symposium: 3rd Dutch Seed Symposium, Wageningen, The Netherlands	07 Oct 2014	0.3
Workshop: Seed Longevity Workshop in Wernigerode, Germany	05-08 Jul 2015	1.2
Symposium: 4th Dutch Seed Symposium, Wageningen, The Netherlands	06 Oct 2015	0.3
Meeting: From Big data to Biological Solutions, Wageningen, The Netherlands	18 Jun 2015	0.3
Seminar: Prof.dr. Yves van de Peer - 'The evolutionary significance of gene and genome duplications'	03 Feb 2015	0.1

Seminar: Analysis of qPCR data. The use and usefulness of amplification curve analysis, Wageningen, The Netherlands	14 Mar 2016	0.1
Symposium: 5th Dutch Seed Symposium, Wageningen, The Netherlands	04-05 Oct 2016	0.6
Symposium: 6th Dutch Seed Symposium, Wageningen, The Netherlands	03 Oct 2017	0.3
► **Seminar plus**		
► **International symposia and congresses**		
Plant Genome Evolution Conference in Amsterdam, The Netherlands	6-8 Sep 2015	0.9
7th International Workshop on Desiccation Tolerance Accross Life Forms, Cape Town, South Africa	11-15 Jan 2016	1.5
Gordon Research Seminar, Waterville Valley, NH, USA	02-03 Jun 2018	0.6
Gordon Research Seminar, Waterville Valley, NH, USA	03-08 Jun 2018	1.8
European Plant Science Retreat (EPSR), Barcelona, Spain	20-23 Jun 2016	1.2
► **Presentations**		
Poster: Annual meeting NWO - ALW Lunteren, The Netherlands	13-14 Apr 2015	1.0
Talk: 1st PPH trip, Germany, 'The Occurrence of Desiccation Tolerance in Higher Plants'	22 Apr-01 May 2015	1.0
Poster: 7th International Workshop on Desiccation Tolerance Accross Life Forms, Cape Town, South Africa	11-15 Jan 2016	1.0
Talk: Annual meeting NWO - ALW Lunteren, The Netherlands 'Genome-wide analysis of LEAs in Xerophyta viscosa'	12 Apr 2016	1.0
Poster: 8th European Plant Science Retreat (EPSR), Barcelona, Spain	20-23 Jun 2016	1.0
Talk: EPS Theme 4 Symposium, Wageningen, The Netherlands 'LEA protein families in Xerophyta viscosa'	16 Dec 2016	1.0
Poster: Annual meeting NWO - ALW Lunteren, The Netherlands	10-11 Apr 2017	1.0
Talk: NGS Seminar, Wageningen, The Netherlands, 'A footprint of desiccation tolerance in the genome of Xerophyta viscosa'	14 Sep 2017	1.0
Talk: 6th Dutch Seed Symposium, Wageningen, The Netherlands, 'Desiccation tolerance in seeds and resurrection plants'	03 Oct 2017	1.0
Talk: Netherlands Annual Ecology Meeting - Lunteren, The Netherlands, 'Genomic and molecular characteristics of a resurrection plant'	14 Feb 2018	1.0
Talk: Gordon Research Seminar (GRS), NH, USA 'Comparative genomics to unveil the evolution of water stress adaptation in plants'	02-03 Jun 2018	1.0
Poster: Gordon Research Conference (GRC), NH, USA	03-08 Jun 2018	1.0
► **IAB interview**		
► **Excursions**		
Plant Physiology PhD trip, Germany	22 Apr-01 May 2015	1.8
Subtotal Scientific Exposure		28.7

3) In-Depth Studies	*date*	*cp*
► **EPS courses or other PhD courses**		
PhD Course 'Systems Biology: Statistical Analysis of ~Omics Data', Wageningen, The Netherlands	15-19 Dec 2014	1.5
Better Safe Than Sorry, Wageningen, The Netherlands	19 Nov 2014	0.3
Course-Bioinformatics: A User's Approach, Wageningen, The Netherlands	24-28 Aug 2015	1.5
The Power of RNA-seq, Wageningen, The Netherlands	10-12 Feb 2016	0.8
Transcription Factors and Transcriptional Regulation, Wageningen, The Netherlands	12-14 Dec 2016	1.0
► **Journal club**		
Literature Discussions in Plant Physiology, Wageningen, The Netherlands	2014-2018	3.0
► **Individual research training**		
Laboratory of Molecular and Cell Biology, University of Cape Town, South Africa	Jan-Apr 2017	3.0
Subtotal In-Depth Studies		11.1

4) Personal Development	*date*	*cp*
► **Skill training courses**		
The Essentials of Scientific Writing and Presenting (ESWP), Wageningen, The Netherlands	24 Nov-02 Dec 2014	1.2
EPS Introduction Course, Wageningen, The Netherlands	20 Jan 2015	0.2
Project and Time Management (P&TM), Wageningen, The Netherlands	30 Mar-11 May 2016	1.5
Brain Training, Wageningen, The Netherlands	2 Nov 2016	0.3
Scientific Writing , Wageningen, The Netherlands	24 Oct 2017-16 Jan 2018	1.8
► **Organisation of PhD students day, course or conference**		
► **Membership of Board, Committee or PhD council**		
	Subtotal Personal Development	5.0

TOTAL NUMBER OF CREDIT POINTS*	**52.3**
Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits.	
** A credit represents a normative study load of 28 hours of study.*	

This work was performed at the Laboratory of Plant Physiology, Wageningen University, and financially supported by the Coordination for the Improvement of Higher Education Personnel ('Cordenação de Aperfeiçoamento de Pessoal de Nivel Superior'-CAPES, Brazil). Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by: Mariana Aline Silva Artur
Cover photo: *Xerophyta viscosa* flower
Thesis Layout by: Mariana Aline Silva Artur
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress