Regulation of desiccation tolerance in *Xerophyta* seedlings and leaves.

RAFE LYALL

Thesis presented for the degree

DOCTOR OF PHILOSOPHY

In the Department of Molecular and Cell Biology

University of Cape Town

February 2016

Acknowledgements

I would like to thank my Supervisors, Prof. Nicola Illing and Dr. Rob A. Ingle, for their substantial support and guidance throughout the work involved in this thesis.

Thanks must also be given to Faezah Davids, whose help and advice was an invaluable asset throughout my studies.

I would like to thank all current and past members of Lab 425 (the EvoDevo Lab) for advice, troubleshooting and assistance on the work presented here, and for making the process fun. Special thanks to Stephen Schlebusch, for general assistance in the realm of Bioinformatics, and to Evan Milborrow, who tirelessly performed the total RNA extractions required for sequencing.

I am grateful to Nicola Mulder, Suresh Maslamoney, Ayton Meintjes and Gerrit Botha, from the UCT Computational Biology Division, for access to their hardware for the assemblies and the guidance provided.

Finally, I would like to thank my family, who supported me and put up with me throughout this process.

Table of Contents

Chapter 1: Vegetative Desiccation Tolerance in Plants

Chapter 2: Post-germinative seedling DT and the "point of no return" in Xerophyta

Chapter 3: De novo transcriptome assembly and analysis of desiccating leaves of *X. humilis*

Chapter 1: Vegetative Desiccation Tolerance in Plants

1.1 Introduction

Plants encounter numerous stresses, both biotic (e.g. parasites, pathogens, herbivory) and abiotic (e.g. climate, light, temperature), during their lifetimes. As sessile organisms, plants are generally unable to escape from unfavourable conditions, and instead have evolved an array of molecular mechanisms to tolerate stress. A stress factor common to nearly all land plants is dehydration due to the lack of available water. This could be through the prolonged absence of water completely during drought conditions, or environmental factors that restrict water potential, such as highly saline soils or low/freezing temperatures (Hasegawa et al., 2000). Drought is one of the leading causes of reduced agricultural crop production and yield, a situation that is likely to worsen as global temperatures rise while the demand for plant crops increases (Bruinsma, 2009). The most commonly cultivated crops, such as wheat, rice, corn and barley, are generally not tolerant to water stress brought on by drought, although there is modest natural intra-species variation between ecotypes. Nonetheless, many other plant species display varying levels of tolerance to drought or desiccation. It is hoped that gaining insight into the strategies tolerant plants use to protect themselves against water stress conditions could lead to improvements in agricultural crop resistance to this stressor.

1.1.1 Plant adaptations to water stress

Plant adaptations to drought can be loosely classified into three types: drought escape, drought avoidance/tolerance or desiccation tolerance (Alpert, 2006; Verslues and Juenger,

2011). Plants that use the drought escape strategy attempt to time their life cycle such that it can be completed during time when drought will not be encountered. Timing germination and/or flowering to coincide with the most amenable seasons for growth, and increasing growth rate during these periods by limiting the expression of stress response pathways are examples of drought escape adaptations. Importantly, plants that utilise this strategy are generally not able to withstand even mild loss of water from their tissues. In contrast, drought avoidance (or drought tolerance) is a strategy whereby a plant limits the loss of its own cellular water during periods of low exogenous water availability, thus maintaining metabolic activity under water stress conditions. Usually, this is achieved by regulating stomatal closure to reduce transpiration, lengthening root systems, storage of water or altering leaf area or external features. While drought tolerant plants are also able to withstand a moderate loss of cellular water compared to sensitive plants, extended or extreme water loss is still lethal (Alpert, 2005). Mechanisms of drought avoidance are widespread in plants, and different species and cultivars can show a range of phenotypic variation (Juenger, 2013). At the extreme end of the scale of land plant adaptations to water stress is vegetative desiccation tolerance (VDT). VDT is defined as the ability of an organism to lose nearly all cellular water (>95%) and recover unharmed upon rehydration (Alpert, 2005). This feat is usually achieved by entering a quiescent state with cessation of all metabolic activity. Instead, the cellular components and cells of the desiccating tissues accumulate a large amount of osmoprotectant agents, chiefly sugars, dehydrin proteins and antioxidants, as well as stored mRNA and/or proteins that are vital for successful recovery (Alpert, 2006). This form of desiccation tolerance has a high metabolic cost and is usually associated with slow-growing organisms in harsh environments where the trade-off

provides an adaptive advantage (Alpert, 2006; Oliver et al., 2000). Consequently, although VDT has evolved multiple times in higher plant species, it is still rare.

1.1.2 Phylogeny of VDT in plants

It is believed that the acquisition of VDT was a vital step in the colonisation of the land by aquatic ancestral plants. The occurrence of VDT in the plant kingdom appears to be consistent with this origin (Figure 1.1; Oliver et al. 2000; Fisher 2008). VDT is very common in basal plants, such as mosses, liverworts and hornworts, indicating that this may be the ancestral state. VDT in non-vascular plants is generally constitutive and focused on recovery from desiccation-induced damage during rehydration as opposed to induced protection of tissues prior to desiccation. Consequently, plants with this form of VDT (constitutive DT, or CDT) can usually survive very rapid desiccation (Oliver et al., 1998). With the evolution of vascular plants, however, VDT became increasingly rarer. The vascular system allowed plants to more easily control their internal water contents and thus be less influenced by external conditions; this had implications for growth rate and plant size, but also eventually gave rise to the more energy-efficient drought escapism and drought avoidance adaptations and the loss of energetically costly constitutive VDT from the majority of the lineage. Consequently, VDT is found in comparatively few angiosperm species and is completely absent in the gymnosperms.

Within the angiosperms, VDT is found in only 135 taxonomically diverse species across 44 genera - commonly known as “resurrection plants” - where it has evolved independently at least thirteen times (Gaff and Oliver, 2013). Resurrection plant species appear in both monocot and eudicot plant families, and are widespread geographically across central and

southern Africa, Europe, Asia, South America and Australia (Behnke et al., 2013; Qureshi et al., 2013; Rakić et al., 2014; Xiao et al., 2015). Generally, they grow in areas with rocky outcrops and shallow soils, where water retention is poor and periods of prolonged drought are common. Some are found in regions that are not commonly associated with drought stress, however, including tropical rainforest (Phillips et al., 2008). Resurrection plants tend to be small and slow growing, most likely due to the high energetic costs of inducing VDT (Oliver et al., 2000). The process of desiccation itself also imposes significant constraints on plant structure that makes the trait more efficient in smaller plants. For example, xylem refilling during rehydration is more easily achieved at small sizes and is likely impossible in plants greater than 3m in height (Alpert, 2005). This could explain the lack of VDT in large plants, and particularly its complete absence in mature gymnosperms.

In contrast to non-vascular plants, resurrection plants can control the rate at which water is lost from their tissues and VDT in these species involves the induction of cellular protection mechanisms prior to the onset of desiccation, rather than repairing damage after rehydration. As a consequence, however, resurrection plants are generally less able to withstand rapid dehydration (Oliver et al., 2000, 1998; Toldi et al., 2009). The major difference in VDT strategies between these two classes of plants has been highlighted as evidence that VDT in resurrection plants is not directly derived from the more primitive mechanisms seen in lower plants (Oliver and Bewley, 1997).

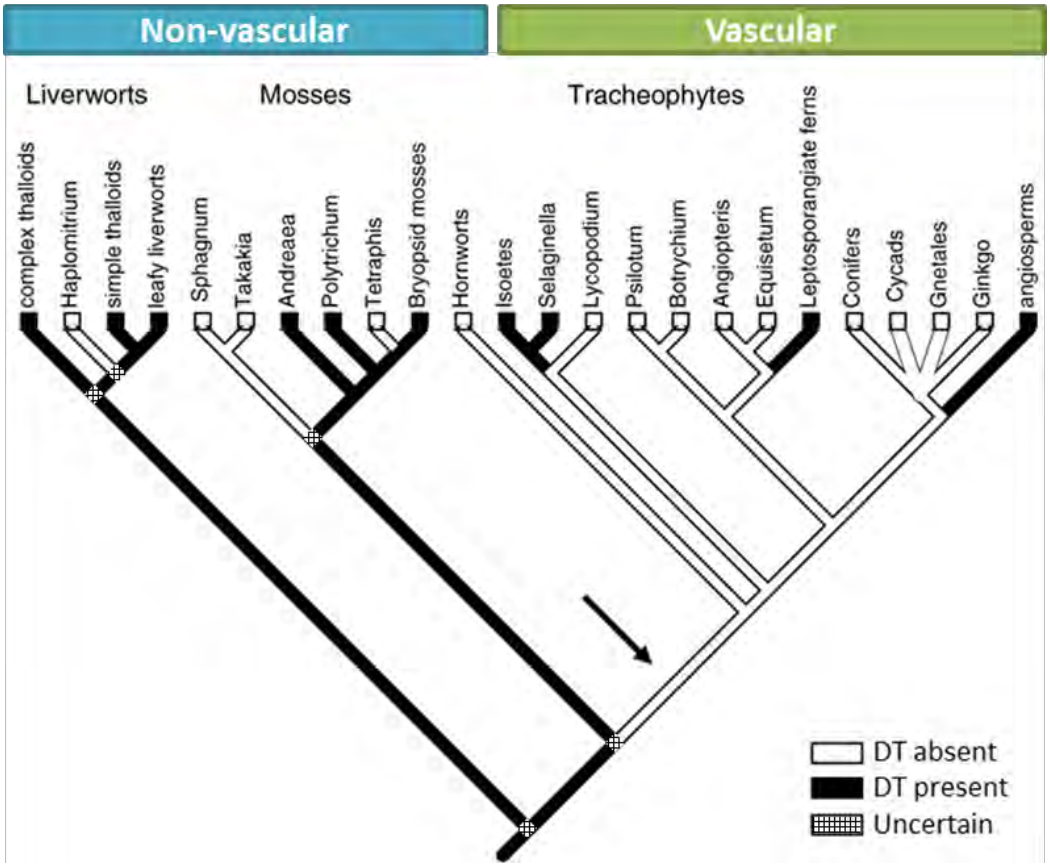


Figure 1.1: Occurrence of VDT in land plants. A simplified phylogeny of the land plants, showing the absence (white lines) or presence (black lines) of vegetative desiccation tolerance (VDT) in various land plant lineages. Ancestral states which cannot equivocally be reconstructed as either desiccation sensitive (DS) or desiccation tolerance (DT) are represented by cross-hatching. The arrow denotes where the ancestor of the vascular plants can be unequivocally reconstructed to be DS. Modified from Oliver et al. (2005).

1.1.3 Seeds: a toolbox of desiccation tolerance spare parts?

Although DT was progressively lost from the vegetative tissues of vascular plants with increasing complexity, the processes and networks involved in primitive VDT were not abandoned completely. Instead, evidence suggests that these ancestral VDT mechanisms were co-opted by early tracheophytes for abiotic stress response and the protection of their reproductive propagules, such as seeds and pollen (Fisher, 2008). Over 95% of angiosperm

and gymnosperm seeds contain DT embryos ("orthodox" seeds), whereas only a small percentage are desiccation sensitive (DS; "recalcitrant" seeds) (Gaff and Oliver, 2013). In orthodox seeds, the onset of embryo desiccation tolerance occurs during the late maturation stage of embryogenesis as a pre-programmed developmental transition (Angelovici et al., 2010). It has been proposed that resurrection plants have co-opted this developmental process to achieve a similar effect in their vegetative tissues, albeit in response to stress (Illing et al., 2005; Oliver et al., 2000). Thus, although both processes (embryo DT and VDT) could share much in common with primitive VDT mechanisms, resurrection plant VDT should be more similar to the developmentally regulated accumulation of DT protection seen in developing seeds (Illing et al., 2005; Shen, 2014).

1.2 The molecular basis of desiccation tolerance

1.2.1 Stresses invoked during desiccation

Cellular desiccation induces massive mechanical and chemical stresses within an organism. Water is essential for maintaining cellular shape and turgor, as well as the structure of macromolecules and membranes. Water is also vital for its function as a solvent and transport medium for metabolism, and is required for photosynthesis (Morgan, 1984). The loss of water thus impairs all of these functions in living cells, with potentially lethal consequences.

Plant cells undergo many structural changes during desiccation. Cells contract as water is lost, which can lead to mechanical stress and damage. Differing rates of contraction between plant cell walls and the protoplasm can exacerbate this effect, particularly at plasmalemma-cell wall junctions where the strain is the highest (Bewley and Krochko, 1982; Murai and

Yoshida, 1998). Resurrection plants often induce cell wall folding to reduce the mechanical stresses associated with cell contraction, and prevent disassociation of the cell wall and membrane (Vicré et al., 2004; Webb and Arnott, 1982). Mechanical stress and damage through the effects of cell shrinkage are reduced in cells that are not highly vacuolated and contain a high proportion of dry matter, which often accumulate in the cells of resurrection plants and plant embryos. However, the decrease of protoplasm volume also causes an increased concentration of intracellular macromolecules; coupled with a reduced masking of chemical moieties by water molecules, this can result in potentially harmful intermolecular interactions that would not normally occur in a hydrated cell. Not only cell walls, but also the maintenance of both cell membrane structure and surface area are essential for surviving dehydration and rehydration. The lipid bilayers that form the cellular membrane require water molecules to maintain their structured organization, and reduced cellular water content can lead to a loss of membrane integrity. This can have dire consequences for the cell, as leakage of hydrolases or harmful metabolic by-products into the cytoplasm could cause irreparable damage (Walters et al., 2002).

Desiccation has profound effects on nuclei and genome structure during desiccation. Maintenance of genomic integrity is obviously of vital importance for successfully surviving desiccation, but transcriptional activity must remain functional at potentially harmful water contents to coordinate a suitable cellular response. Indirectly, protection of the genome is achieved by limiting cell division and arresting the cell cycle at the G1 phase, before chromosomes have duplicated (Dekkers et al., 2015). During early seed maturation in *A. thaliana*, the nuclei shrink and chromatin becomes highly condensed. A similar response is seen in the desiccating leaves of the resurrection plant *C. plantagineum*, suggesting that this

may be a general adaptation to desiccation (van Zanten et al., 2011). Interestingly, the chromatin compaction does not appear to affect global transcription levels, as total transcript abundance and the proportion of all expressed genes in mature seeds, leaves and seedlings are similar (van Zanten et al., 2012). In mature seeds, nuclei shrinkage and chromatin compaction are both developmentally controlled by maturation genes, and can be uncoupled. Similarly, chromatin expansion during imbibition (initial intake of water by a seed, prior to germination) is developmentally regulated (van Zanten et al., 2012).

Protein misfolding induced by a lack of hydrophilic interactions, increased macromolecule density within the cell and potential membrane leakage can all severely disrupt metabolic pathways, leading to an accumulation of potentially harmful intermediate products (Farrant, 2000; Leprince et al., 2000). In particular, production of reactive oxygen species (ROS) by uncontrolled photosynthetic and respiratory machinery can lead to irreparable damage to DNA, proteins and structural macromolecules (Kranner and Birtic, 2005; Kranner et al., 2002). Down-regulation of both of these processes – respiration and photosynthesis – during early dehydration increases the chance of survival after desiccation. Loss of mitochondrial cristae is associated with mature orthodox embryos and pollen, and has also been observed in the cells of resurrection plants (Bewley and Krochko, 1982; Farrant et al., 1997; Hoekstra and van Roekel, 1983). Most resurrection plants are homochlorophyllous: their photosynthetic activity is limited during desiccation by down-regulation of photosynthetic processes or via light avoidance, such as leaf folding or accumulation of pigment molecules (Koonjul et al., 2000; Sherwin and Farrant, 1998). Poikilochlorophyllous resurrection plants, on the other hand, prevent photosynthesis by directly degrading chlorophyll and dismantling the thylakoid membranes of their chloroplasts (Tuba et al.,

1996). Poikilochlorophylly is a comparatively recent evolutionary DT adaptation, known only in some monocot resurrection plant families. Usually it is observed in species that remain desiccated for prolonged periods (6 months or longer; Oliver et al. 2000; Tuba et al. 1998).

1.2.2 Protective molecules invoked during desiccation

Desiccation tolerant plants and orthodox seed embryos induce several common response factors that aid in reducing the damage from the stresses discussed above.

1.2.2.1 Sugars

Sugars, particularly sucrose and the raffinose family oligosaccharides, have been implicated in desiccation tolerance in plants and other organisms (Berjak, 2006; Ingram and Bartels, 1996). Sugars have been predicted to play multiple roles during desiccation. Most commonly, sucrose and other sugars are believed to hydrogen bond with and stabilise membranes and macromolecules when cellular water levels drop too low to maintain a hydration shell, effectively replacing the lost water molecules (Hoekstra et al., 2001). This prevents binding to other intracellular solutes that could cause irreversible protein denaturalisation or plasma membrane phase transition. Additionally, at very low water contents (below 0.1 g H_2O/g tissue), sugars are hypothesised to facilitate the transition of the cellular cytoplasm into a vitrified state, or "glass" (Hoekstra et al., 2001). Cellular glasses are a highly viscous super-saturated liquid state that resembles a brittle, solid material. The viscous vitreous phase reduces molecular mobility, "locking" membranes, proteins and cellular structures in place and preventing their denaturisation, destabilisation or crystallisation. Experimental evidence suggests that the cytoplasmic glasses found in

desiccation tolerant tissues do not act similarly to glasses formed of the same sugars in an experimental setting, suggesting that cellular glass found in biological tissue consists of a mixture of sugars (predominantly sucrose), proteins and other solutes (Farrant and Walters, 1998; Hoekstra et al., 2001). Small amounts of oligosaccharides, particularly stachyose and raffinose, have also been shown to prevent sucrose crystallisation during glass formation, so the presence of these sugars in many DT organisms may play a similar role (Smythe, 1967). Interestingly, it has been shown that recalcitrant seeds from multiple species accumulate sugars to levels comparable to those in orthodox seeds, implying that sugars alone are not capable of imparting complete osmoprotection (Berjak and Pammenter, 2008; Farrant et al., 1993; Steadman et al., 1996). It has been argued that recalcitrant seeds fail to induce the necessary protection mechanisms during the earlier stages of dehydration, so that the protection afforded by sugars at lower water contents is insufficient to maintain cell viability (Berjak and Pammenter, 2008).

Lastly, sugars and carbohydrate signalling play important roles during seed development in angiosperms (Rolland et al., 2002). For example, sucrose is believed to regulate differentiation and storage processes associated with seed maturation, whereas glucose promotes growth and cell division in the surrounding seed coat (Weber et al., 1997; Wobus and Weber, 1999a). Exogenous sucrose can delay germination of the seeds of a number of species, including *A. thaliana* and *Brassica napus*, although the exact mechanism by which it achieves this is unclear (Gibson, 2005; Xu et al., 2010). Delay of germination by sucrose may involve accumulation of abscisic acid (ABA), as is seen in seedlings treated with exogenous glucose (Dekkers et al., 2004). It has been hypothesised that there may also be a role for

sugar signalling during desiccation in resurrection plants, possibly as a negative regulator of energy metabolism and photosynthesis (Griffiths et al., 2014).

1.2.2.2 LEA proteins

Late embryogenesis abundant (LEA) proteins are a class of proteins that, as the name suggests, were first discovered at very high levels in desiccating and dry angiosperm seeds (Dure et al., 1981), where they were predicted to be involved in the acquisition of desiccation tolerance. LEAs belong to a class of proteins known as hydrophylins, which are defined by their highly hydrophyllic nature (hydrophilicity > 1) and large proportion of glycine residues (>6%) (Battaglia et al., 2008). Hydrophylins, including LEAs and LEA-like proteins, have been found to be present in many forms of life, and play a role in desiccation tolerance (as well as other abiotic stresses) in at least plants, animals (Hand et al., 2011; Kikawada et al., 2006), fungi (Garay-Arroyo, 2000) and bacteria (Stacy and Aalen, 1998). In plants, LEA proteins are induced during seed maturation, as well as during desiccation stress in resurrection plants (Collett et al., 2004a; Illing et al., 2005; Piatkowski et al., 1990), and various abiotic stresses in desiccation sensitive plants (Shinozaki and Yamaguchi-Shinozaki, 2006). Manipulation of LEA-producing pathways has been proposed as a mechanism to improve drought and abiotic stress tolerance in crop plants (Bhatnagar-Mathur et al., 2008).

Like sugars, LEA proteins are hypothesised to function as water replacement molecules (Hoekstra et al., 2001). Some have also been shown to act via ion sequestration, and protein chaperone/transport (Grelet et al., 2005; Kovacs et al., 2014; Wise and Tunnacliffe, 2004). There is increasing evidence that LEA proteins act synergistically with sugars during cellular glass formation (reviewed by Berjak 2006). Incubation of a LEA protein isolated

from the DT pollen of *Typha latifolia* with sucrose altered its conformation during desiccation compared to drying in the absence of sucrose. Additionally, the strength of hydrogen bonding and stability of a glass state consisting of a combination of the LEA protein and sucrose exceeded that of a glass containing only sucrose (Goyal et al., 2003; Wolkers et al., 2001). Further argument has been made that LEA proteins could make up the dominant portion of the intracellular glass; sucrose and other sugars serve to increase the density of the resultant glassy state, and can be deposited in extracellular or intermembrane spaces too small for LEA proteins (Buitink and Leprince, 2004).

There are a multitude of different LEAs proteins expressed during stress in plants, and they can be divided into separate classes based on physical properties. LEA proteins have also been observed to display spatial and organ specificity (Martinelli, 2008). Of the 51 LEA proteins identified in *A. thaliana*, 36 are localised to the cytosol, most of which can also enter the nucleus; three are exclusively located in either plastids or mitochondria, and a further two are targeted to both structures; three are found specifically in the endoplasmic reticulum (ER), two in the vacuoles and one in pexophagosomes (organelles involved in selective peroxisome degradation); the remaining two LEAs are secreted (Candat et al., 2014). Non-uniform targeting of LEA proteins throughout the cellular protoplasm is consistent with local differences in cellular glass properties during desiccation, as has been suggested to occur for specialised regions and organelles (Berjak, 2006). Low levels of transcription, translation and enzyme activity have been shown to occur in the living cells of otherwise dry seeds, indicative of a non-homogenous cellular glass matrix and the presence of concentrated pockets of water within desiccated cells (Elder et al., 1987; Leubner-Metzger, 2005; Rinne et al., 1999). LEAs and LEA-like proteins known as rehydrins are also induced

during rehydration in desiccated mosses. Although it is not clear what role they play during rehydration, it has been suggested that they may stabilise membranes at low water contents as the cell rehydrates, or help control water intake (Oliver et al., 2005).

LEA proteins were originally classified based on their high abundance in embryos and seeds during embryogenesis and maturation. However, they are nonetheless a diverse set of proteins which have since been classified into multiple families that share some common features and motifs (Table 1.1; Dure et al. 1989; Bray 1993; Wise & Tunnacliffe 2004; Cuming 1999; Illing et al. 2005; Battaglia et al. 2008). For the sake of this thesis, the nomenclature determined by Illing et al. (2005) will be used. The most formally described of the LEA protein classes are the first three (Battaglia et al., 2008). Group 1 LEAs contain one to four copies of a conserved 20 amino acid repeat sequence (GGQTRREQLGEEGYSQMGRK), and are often rich in Glu, Gly and Gln residues (Cuming, 1999). Group 2 LEAs, commonly known as dehydrins, contain a signature 15-amino acid lysine-rich repeat (K-segment), EKKGIMDKIKEKLPG, and often serial repeats of serine (S) residues (Bartels et al., 2007; Close, 1996). Group 3 LEAs contain a series of partially conserved, tandem 11-amino acid repeats, TAQAAKEKAGE, which may promote the formation of an amphipathic alpha-helical protein conformation. The other LEA groups are less obviously conserved, though they can still be grouped on the basis of incompletely conserved motifs and chemical properties (Battaglia et al., 2008). Although many LEA proteins are induced during both seed development and vegetative stress in plants, the specific LEAs expressed during these two processes generally do not overlap (Hundertmark and Hincha, 2008). LEA-1 and LEA-6 LEA proteins, for example, are only expressed during seed maturation in *A. thaliana*, and are not induced by stress in vegetative tissue (Hundertmark and Hincha, 2008; Illing et al.,

2005). Bayesian reconstruction of the ancestral expression patterns of LEA genes induced in resurrection plants suggests that they are nearly all derived from propagule-specific developmental processes, that have since been co-opted into vegetative stress response pathways (Fisher, 2008). This is classed as further evidence that angiosperm VDT is likely derived from seed maturation processes, rather than gene networks related to stress in vegetative tissue.

PFAM name/code	Interpro	Dure, 1993	Bray, 1994	Wise, 2003	Illing, 2005
LEA_5; PF00477	IPR000389	D19	Group 1	Class I	LEA-1
Dehydrin; PF00257	IPR000167	D11	Group 2	Class II	LEA-2
LEA_4; PF02987	IPR004238	D7,D29	Group 3, 5	Class III	LEA-3
LEA_1; PF03760	IPR005513	D113	Group 4	Class II, III	LEA-4
SMP; PF04927	IPR007011	D34	Group 6	Class IV	LEA-6
LEA_3; PF03242	IPR004926	D73	—	Class V	LEA-7
LEA_2; PF03168	IPR004864	D95	—	Class VI	LEA-8
AWPM-19-like; PF05512	IPR008390	—	—	—	LEA-10
LEA_6; PF10741	IPR018930	—	—	—	—

Table 1.1. Different nomenclatures used to describe LEA protein groups by various authors, together with the associated Interpro and Pfam description and identifiers. The nomenclature determined by Illing et al (2005) will be used within this thesis. Table modified from Tunnacliffe and Wise (2007) and Fisher (2008).

1.2.2.3 Antioxidants

Damage caused by reactive oxygen species (ROS) can be incredibly detrimental for dehydrating or desiccating cells, thus plants exhibit multiple adaptations that lead to ROS detoxification during these abiotic stresses. These ROS-detoxifying processes occur in two ways: via molecular antioxidants, such as ascorbate, glutathione or phenolics; or by ROS-scavenging enzymes, such as peroxidases and catalases (Ahmad et al., 2010; Apel and Hirt,

2004; Bartels and Sunkar, 2005; Mittler, 2002). ROS detoxification is important during both dehydration (as discussed previously) and early rehydration (Kranner and Birtic, 2005). Interestingly, antioxidant activity may also be essential for the maintenance of plant viability in the desiccated state. The tissues of the resurrection plant *Myrothamnus flabellifolia* become less able to successfully rehydrate the longer the plant remains desiccated. The loss of viability correlates with the chemical reduction of free-radical scavenging molecular antioxidants, such as glutathione and ascorbate, which accumulated during dehydration (Kranner et al., 2002). The vitrified cellular matrix that forms in desiccated cells of DT organisms is permeable to ROS, thus the accumulation of sufficient molecular antioxidants to counteract their effects while the plant is desiccated is critical for DT (Berjak, 2006). This is also consistent with the observation that poikilochlorophylly is more common in resurrection plant species that remain desiccated for an extended period of time, as unprotected chlorophyll pigments will produce ROS when exposed to light that could accumulate in desiccated tissues over time (Oliver et al., 2000).

1.2.2.4 Heat shock proteins (HSP)

HSPs are induced during abiotic stress in many organisms, from bacteria to mammals (Gupta et al., 2010; Parsell et al., 1993; Vierling, 1991). In plants, these stresses include heat/cold, drought, high salt, oxidative stress, high light levels, wounding and toxins (Swindell et al., 2007). HSPs in plants are classified according to their weight: the 100 kDa HSPs (Hsp100), HSP90, HSP70, HSP60 and small HSP (sHSP) families (Gupta et al., 2010; Kotak et al., 2007). The sHSP gene family is particularly diverse in plants, and the genomes of different plant species can contain upwards of 40 *sHSP* genes (Vierling, 1991). Small HSPs generally occur as a complex of multiple subunits with molecular weights between 200 and

800 kDa, and are not expressed in vegetative plant tissues during normal growth conditions, but are only induced in response to stress or during seed maturation (Waters, 2013; Wehmeyer, 1996). However, the resurrection plant, *C. plantagineum*, expresses sHSPs at detectable levels even when hydrated, which accumulate to higher levels during desiccation (Alamillo et al., 1995). The exact function of HSPs is not known, but the majority of them are believed to act as molecular chaperones that directly interact with other proteins. In this manner they can selectively activate target proteins, prevent protein misfolding or aggregation, target misfolded proteins for degradation, or facilitate protein oligomer formation. They may also be involved in protein relocalisation during stress conditions (Feder and Hofmann, 1999; Gupta et al., 2010; Hu et al., 2009; Panaretou and Zhai, 2008).

1.2.2.5 Oleosins

Oil bodies are structures found in some tissues in vascular plants that store triacylglycerol (TAG) as an energy reserve. However, they are most often associated with seeds, where they accumulate alongside seed storage proteins (SSPs) and carbohydrates during seed maturation and are used to fuel the seedling during germination. Oil bodies consist of an inner matrix of TAG molecules surrounded by a phospholipid membrane. In the tissues of DT organs, such as seeds and pollen, the oil body membrane is further stabilised by a class of proteins known as oleosins (Napier et al., 1996). Oleosins have not been identified in the oil bodies of non-DT tissues, such as leaves and fruits, and have never been detected in recalcitrant seeds (Pammenter and Berjak, 1999). Thus, they likely play an integral role in preserving oil body integrity during the desiccation process. As further evidence of the link between seed maturation and VDT, an oleosin gene has previously been found to be up-regulated in the leaves of *X. humilis* during desiccation (Collett et al., 2004a; Shen, 2014).

1.3 Seed development and embryo DT

Seed development in orthodox-seed-bearing plants is a highly regulated process that assures the correct combination of embryo growth/differentiation, induction of protective factors, and accumulation of the necessary components and resources required for eventual germination (Fig. 1.2). The life-cycle of an angiosperm seed is a complicated one, and can be divided into five distinct stages, although they may differ slightly across species: fertilisation, whereby a maternally-derived egg cell is fertilised by paternally-derived pollen; embryogenesis, which involves the early development of the embryonic plant within the ovule; maturation, during which seed-specific gene pathways are activated and seed-specific traits such as reserve accumulation and desiccation tolerance arise; dormancy, where the seed enters a quiescent state; and germination, which occurs when the seed leaves dormancy due to the detection of favourable conditions for growth and the seedling begins its adult life-cycle (Park and Harada, 2008; Raz et al., 2001; Santos-Mendoza et al., 2008).

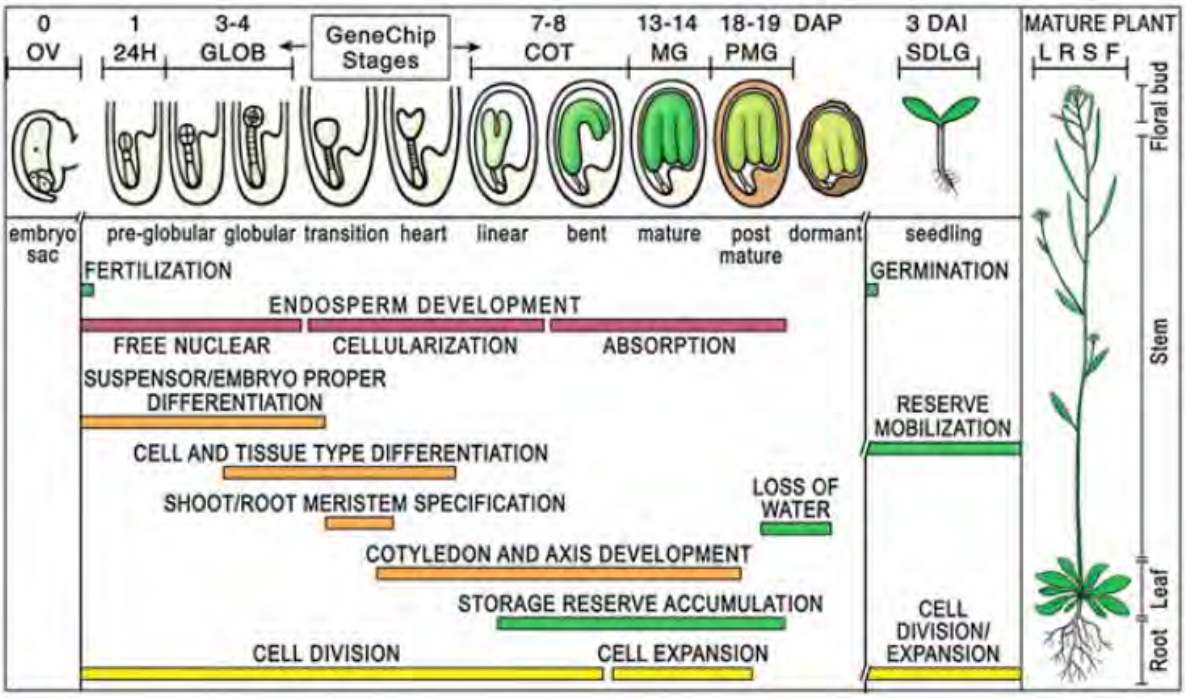


Figure 1.2. The seed life cycle in *A. thaliana*. Reproduced from Le et al. (2010), wherein global analysis of gene activity in developing seeds was analysed. Numbers along the top represent days after pollination (DAP) during seed development or days after imbibition during germination (DAI). OV, unfertilized ovule; 24H, 24-h post-pollination seed; GLOB, globular-stage seed; COT, cotyledon-stage seed; MG, mature-green-stage seed; PMG, postmature-green-stage seed; SDLG, seedling; L, leaf; R, root; S, stem; F, floral buds.

Seed development is controlled by plant hormone interactions, particularly interplay between abscisic acid (ABA) and gibberellin (GA). In *A. thaliana*, maternally-derived ABA accumulates during embryogenesis, and is important for controlling the transition from embryo morphogenesis to seed maturation (Karssen et al., 1983). ABA regulates the expression of maturation-specific genes, particularly those involved in growth arrest and subsequent seed filling by lipid and protein storage molecules (Raz et al., 2001). ABA also promotes the induction of seed desiccation tolerance, by controlling expression of osmoprotectant molecules and proteins (Santos-Mendoza et al., 2008). A second peak in ABA concentration, this time embryo-derived, promotes seed quiescence and entry into

dormancy once maturation is complete (Finkelstein et al., 2002; Karssen et al., 1983). During seed maturation and dormancy, ABA directly antagonises GA signalling and biosynthesis to prevent growth and cell division. Germination is only initiated once ABA-mediated GA inhibition is lifted via the appropriate environmental cues (Debeaujon and Koornneef, 2000).

1.3.1 Gene networks involved in seed maturation

During maturation, the embryo of orthodox seeds acquires tolerance to a wide spread of stresses, including desiccation. This process is controlled by ABA, via a network of seed-maturation-specific genes and transcription factors. In *A. thaliana*, the most prominent of these is the *LEC1-ABI3-FUS3-LEC2* (LAFL or AFL) gene network, named for the genes that act as the master regulators of seed maturation (Fig. 1.3; Santos-Mendoza et al. 2008; Jia et al. 2014). The AFL network consists of three B3-domain containing transcription factors, *ABISCISIC ACID INSENSITIVE 3* (*ABI3*), *FUSCA 3* (*FUS3*) and *LEAFY COTYLEDON 2* (*LEC2*); another *LEAFY COTYLEDON* gene, *LEC1*, acts upstream of the AFL network and encodes a HAP3 subunit, which is associated with many of the plant NFY complexes. Knock-out mutants of all of these genes show varying levels of abnormal seed development, including failure to enter into dormancy, decreased sensitivity to ABA and premature germination (reviewed in Harada 2001). Additionally, over-expression of these genes can induce somatic embryogenesis, a process whereby seed-specific factors are induced in otherwise vegetative tissue and lead to the development of an ectopic embryo (Gazzarrini et al., 2004; Kagaya et al., 2005a; Parcy et al., 1994; Stone et al., 2001).

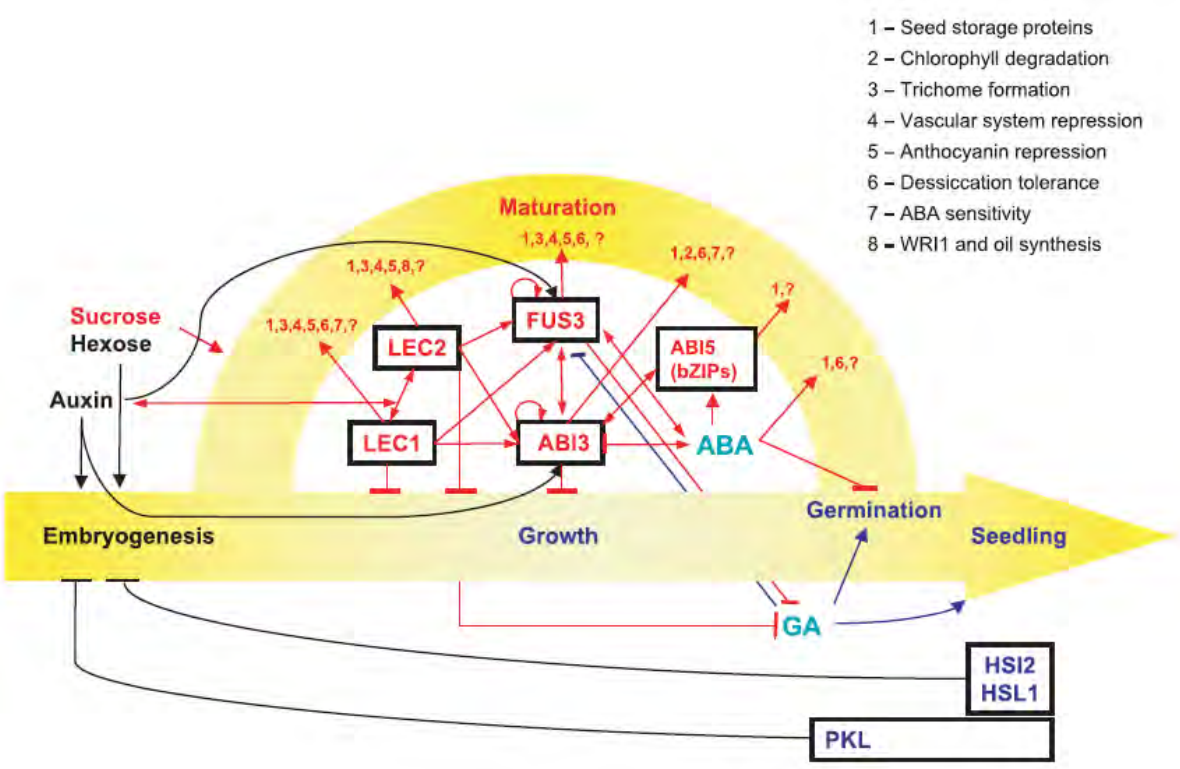


Figure 1.3. A complex network of regulatory factors regulates seed development and maturation in *A. thaliana*. The LAFL genes (*LEC1*, *ABI3*, *LEC2*, *FUS3*; centre) act as master regulators of seed development and co-ordinately control different aspects maturation (see key) and their own expression. Arrows and T-bars represent positive and negative regulation, respectively. Factors associated with seed maturation, or involved in the induction or maintenance of seed maturation, are shown in red. Factors involved in growth and differentiation are shown in blue. The hormones abscisic acid (ABA) and gibberellic acid (GA) are shown in light blue. Reproduced from Santos-Mendoza et al. (2008).

1.3.1.1 The AFL network

The individual component genes of the AFL pathway interact via a complicated network of regulation and autoregulation, which can differ across embryonic tissue or maturation stage (Figure 1.3; To et al. 2006; Santos-Mendoza et al. 2008; Gutierrez et al. 2007). Single mutants of each gene show differing levels of phenotypic severity that may hint at redundant and

synergistic functions between the AFL genes. Mutant plants lacking *ABI3* are most highly resistant to ABA treatment, and have decreased accumulation of seed storage proteins (SSPs) and seed lipids compared to either *fus3* or *lec2* mutants. *ABI3* is also the only one of these genes consistently expressed throughout the whole embryo (Santos-Mendoza et al., 2008). Of all the mutant plants, those of *lec2* have the least severe phenotype in terms of reserve accumulation, and are much more tolerant of desiccation (Stone et al., 2001). *LEC2* is a paralogue of *FUS3* that arose from a dicot-specific duplication event (Li et al., 2010), and an analysis of seed development programs in monocots and dicots suggests that only the function of the *LEC1*/*ABI3*/*FUS3* network, and not *LEC2*, is conserved in monocots (Sreenivasulu and Wobus, 2013). A central role for *FUS3* appears to be regulation of ABA biosynthesis and inhibition of germination-related GA networks in an auxin-dependent manner in the embryo epidermis. Thus, although *fus3* mutants display altered SSP, DT and seed maturation phenotypes throughout the embryo, it is likely that many of these effects occur indirectly via downstream ABA and *ABI3*-mediated gene regulation (Gazzarrini et al., 2004). However, all three AFL proteins have also been shown to be transcriptional activators that directly regulate a subset of distinct and overlapping maturation genes, as well as each other, in the tissues in which they occur (Figure 1.3) (Braybrook et al., 2006; Mönke et al., 2012; Wang and Perry, 2013; Yamamoto et al., 2010). For example, ectopic over-expression of *LEC1* can activate expression of *ABI3* and *FUS3*, and the transcript levels of these genes are reduced in a *lec1* mutant (Kagaya et al., 2005b). Similarly, ectopic expression of *LEC2* results in up-regulation of the *LEC1* and *FUS3* genes (Stone et al., 2008). *LEC1*, *LEC2* and *FUS3* have all been shown to be important for somatic embryogenesis – the production of embryonic tissue from vegetative tissue – and the subsequent up-regulation of many seed-related genes in somatic embryos that are usually downstream in the AFL network, such as bZIP

transcription factors (Figure 1.3) (Braybrook et al., 2006; Elhiti et al., 2013; Gaj et al., 2005). *ABI3* and *FUS3* have been shown to positively regulate one another in a reciprocal fashion (To et al., 2006); the ABI3 protein also interacts directly with the *FUS3* promoter (Mönke et al., 2012), and the FUS3 protein interacts with regulatory elements in the *LEC1*, *FUS3* and *ABI3* genes (Wang and Perry, 2013).

The three AFL proteins can bind the B3-domain binding site – CATGCA(TG), or the RY motif - *in vitro* (Kroj et al., 2003; Mönke et al., 2004; Reidt et al., 2000; Suzuki et al., 1997). The RY repeat is common in the promoters of seed-specific genes, and is consequently overrepresented in the gene promoters of the direct targets of the AFL genes (Baumlein et al., 1992; Bobb, 1997; Dickinson et al., 1988; Lelievre et al., 1992). Such direct targets include seed storage protein genes, LEA genes, lipid biosynthesis and seed oilbody biogenesis genes, including oleosins. As master regulators of seed maturation, downstream targets also include other transcription factors involved in gene regulation (Mönke et al., 2012; Santos-Mendoza et al., 2008; Wang and Perry, 2013).

1.3.1.2 LEC1 and the NF-Y complex

LEC1 encodes a subunit of the family of plant NF-Y complexes (Kagaya et al., 2005a). *LEC1* mutant seeds, much like those of the AFL genes, display phenotypes associated with a lack of embryonic traits, such as decreased desiccation tolerance, abnormal embryo development and failure to properly accumulate SSPs, lipids and protein. *LEC1* is active predominantly during early embryogenesis and acts upstream of the AFL network, as evidenced by the observation that ectopic expression of *LEC1* can induce *ABI3* and *FUS3* expression in *A. thaliana* (Lotan et al., 1998; Meinke, 1994; Santos-Mendoza et al., 2008).

The NF-Y protein family function as a heterotrimeric complex of the three types of subunits - NFYA, NFYB and NFYC - and can act as either transcriptional activators or repressors of genes involved in plant development and stress response (Hackenberg et al., 2012; Petroni et al., 2012). Unlike animals and fungi, which have only a single gene for each of the three subunits, the gene family has undergone an extensive duplication in plants. Different NF-Y subunit combinations are believed to form complexes that have distinct spatial-temporal expression patterns that allow fine control of developmental processes in plants (Hackenberg et al., 2012; Siefers et al., 2009). *LEC1*, an NF-YB9 subunit, is one such example, as it is expressed almost exclusively during seed embryogenesis (Lee et al., 2003). A related protein, the NF-YB6 subunit *LEC1-like* (L1L) is similarly expressed predominantly during seed development. Orthologues of *LEC1* and *L1L* have been identified in both angiosperms (monocots and dicots) and gymnosperms, and have been implicated in the switch from embryonic to vegetative growth in all species studied so far (reviewed in Petroni et al. 2012).

1.3.1.3 The bZIP family: ABI5 and the ABF genes

Analysis of the enhancer and promoter regions of seed maturation genes indicates that not only RY motifs are responsible for the intricate pattern of AFL target gene expression, but other motifs are necessary for finer control of maturation genes (Sakata et al., 1997; Vicente-Carbajosa and Carbonero, 2005; Wobus and Weber, 1999b). For example, ABA-dependent transactivation by *ABI3* requires the presence of the abscisic acid response element (ABRE) within the gene promoter as well the conserved B2 activation domain of the ABI3 protein. It is likely that ABI3 interacts with ABRE-binding protein complexes through its B2 domain, rather than with the ABRE directly (Ezcurra et al., 2000). Similarly, both *LEC1* and *LEC1-like*

proteins can regulate a subset of target genes in response to ABA via the ABRE cis-element, through direct interaction with bZIP transcription factors (Yamamoto et al., 2009).

The bZIP transcription factor (TF) family has been shown to be important for seed embryogenesis and maturation (Becker et al., 2014; Jakoby et al., 2002). The group-A bZIP gene *ABI5* is the most highly characterised bZIP involved in seed maturation. It can regulate seed gene expression in a complex with *ABI3*, and directly interacts with the ABI3 B2-domain in yeast two hybrid experiments (Nakamura et al., 2001). It can also bind the ABRE motif present in its target gene promoters *in vitro*, and has been detected as binding these promoters *in vivo* (Nambara and Marion-poll, 2003) . Interestingly, the seed-specific RY-motif recognised by ABI3 is dispensable for *ABI3*/*ABI5*/ABRE-mediated LEA gene expression, suggesting that the ABI3-ABI5 protein interaction alone is sufficient for this action, and ABI3 need not be anchored to the gene promoter (Sakata et al., 2010). The expression of *ABI5* is positively regulated by *ABI3* and ABI5 has recently been found to bind to its own promoter, so likely regulates its own expression as well (Lopez-Molina et al., 2002; Xu et al., 2014). *A. thaliana abi5* mutant seeds are similar to those of *abi3*, and have reduced sensitivity to ABA and decreased expression of a subset of LEA genes. However, they do not show reduced desiccation tolerance, and *abi3* mutants show a more severe reduction in LEA expression over a broader range of LEA genes (Finkelstein and Lynch, 2000; Finkelstein et al., 2002). Thus, it has been hypothesised that *ABI5* functions in a partially redundant fashion with other *ABI5*-like bZIP TFs in regulating the ABRE-mediated downstream effects of *ABI3* during seed maturation (Bensmihen, 2002). Interestingly, the ABF TFs can function as both positive and negative regulators of gene transcription, and can act in an antagonistic fashion. In *A. thaliana*, *ABI5* and one of its homologs, *EEL*, show opposite effects on

expression of two *ABI5*-regulated LEA genes, *AtEm1* and *AtEm6*, probably through competition for the same binding site (Bensmihen, 2002). The evolution of distinct roles for different *ABI5* -like ABFs may exist to fine-tune gene expression during seed maturation and assure the correct spatial and temporal expression of their target genes during the seed maturation process. Additionally, four group-A bZIP TFs that are not expressed in seeds, *ABF1*, *ABF2*/*AREB1*, *ABF3* and *ABF4*/*AREB2*, are considered master regulators of ABA-mediated water stress response in vegetative tissue (Uno et al., 2000; Yoshida et al., 2015, 2010). However, mutant *A. thaliana* plant lines over-expressing *ABF2*, *ABF3* and *ABF4* all resulted in ABA-hypersensitivity during germination and seedling growth, suggesting at least some overlap in gene activation activity with the seed-specific bZIP proteins (Furihata et al., 2006; Kang, 2002), and overexpression of an *ABF2* orthologue from peanut (*Arachis hypogaea*) can rescue some phenotypes of an *abi5* mutant in *A. thaliana* (Li et al. 2013). Group C bZIP proteins have also been implicated in seed maturation processes. The monocot bZIP protein *OPAQUE2* controls the expression of seed storage protein (SSP) genes in wheat and rice, but not those expressed later during maturation (Holdsworth et al., 1995; Wu et al., 1998). Several group C bZIP TFs have also independently been implicated in an as yet unidentified complex that controls expression of SSP genes in *A. thaliana* via interaction with *ABI3* (Alonso et al., 2009; Lara et al., 2003).

1.3.1.4 ABA signal transduction

ABA signalling plays an essential role in abiotic stress response and development in plants, including seed maturation; *LEC1*, *ABI3* and *ABI5* can all control a subset of seed maturation genes in an ABA dependent manner via the ABRE cis-element. However, the exact

mechanisms by which the ABA signal is recognised and acted upon within plant cells has only recently begun to be uncovered.

In *A. thaliana*, two *ABSCISIC ACID INSENSITIVE* genes, *ABI1* and *ABI2*, were originally shown to be group-A *PROTEIN PHOSPHATASE 2C* (PP2CA) proteins that negatively regulate ABA signalling in seeds and vegetative tissue in a coordinated fashion, indicating that protein phosphorylation plays an essential role in ABA signal transduction (Leung et al., 1997, 1994; Merlot et al., 2001; Meyer et al., 1994). This led to the identification of a plant specific group of protein kinases, type-2 *SUCROSE NON-FERMENTING 1*-related (SnRK2s), which activate various ABA-response proteins and TFs, including ABFs, by phosphorylation (Fujii et al., 2007; Melcher et al., 2009; Miyazono et al., 2009; Nishimura et al., 2009). In the absence of PP2CA proteins, SnRK2 protein kinases are themselves activated by phosphorylation, possibly by autophosphorylation (Fujii and Zhu, 2009); this activity is inhibited in the presence of PP2CAs, which dephosphorylate multiple Ser/Thr residues required for catalytic activity within the activation loop of the SnRK2 kinases (Umezawa et al., 2009; Vlad et al., 2009). Completing the picture, a family of cytosolic ABA receptors, termed PYR/PYL/RCAR proteins were identified. Upon perception of cytosolic ABA, the PYR/PYL/RCAR receptors bind to the catalytic site of group-A *PROTEIN PHOSPHATASE 2C* (PP2CA) proteins and prevent their interaction with and inhibition of SnRK2 kinases (Ma et al., 2009; Pandey et al., 2009; Park et al., 2009; Santiago et al., 2009).

Information on the genes that control ABA signal transduction in plants has so far mostly been derived from work in *A. thaliana*. Mutant plants lacking many of the PYL/PYR/RCAR receptor genes showed reduced expression of ABA-response genes, including SnRK2s , and severe ABA insensitivity (Gonzalez-Guzman et al., 2012). PYR/PYL protein homologs have

also been identified in rice (He et al., 2014), and have been implicated in seed development, germination and early seedling growth in this species as well (H. Kim et al., 2012). The *A. thaliana* genome contains nine group-A PP2C genes: *ABI1* and *ABI2*, which are broadly involved in ABA signalling throughout plant development; *AHG1*, *AHG3*, *HAB1* and *HAB2*, which are involved in seed development and germination; and *HAI1*, *HAI2* and *HAI3* that are induced by drought stress (Bhaskara et al., 2012). It is thought that the diversity of PP2CA protein family allows for finer control of ABA signalling during different stress conditions. The SnRK protein kinase family is diverse in *A. thaliana*, which contains 38 SnRK genes, 10 of which are SnRK2s (Hrabak et al., 2003). SnRK2s have been further divided into three classes based on phylogenetic analyses: group 1 consists of SnRK2 proteins that are not activated by ABA, group 2 contains those only weakly activated by ABA (although degree of activation varies across species), and group 3 contains those strongly activated by ABA (reviewed in Kulik et al. 2011). As general regulators of abiotic stress response, the majority of SnRK2 proteins are also activated by osmotic stress, although they are not responsive to cold (Boudsocq et al., 2004). The *A. thaliana* SnRK2s activated by ABA include the group 2 proteins SRK2C/ SnRK2.8 and SRK2F/ SnRK2.7, and the group 3 proteins SRK2D/ SnRK2.2, SRK2E/SnRK2.6, and SRK2I/ SnRK2.3 (Boudsocq et al., 2007, 2004). All five of these proteins have been linked to *ABF2* and *ABF4* phosphorylation in an *A. thaliana* cell line (Furihata et al., 2006). Those from group 3 in particular (*SRK2D*, *SRK2E* and *SRK2I*) have been shown to be essential for seed development, dormancy, germination and drought tolerance, and can induce phosphorylation of *ABI5*, *EEL* and *DPBF3* (a seed-specific ABF) proteins (Nakashima et al., 2009). The functionally redundant root-specific *SRK2C* and shoot-specific *SRK2F* may also be involved in seed development and ABF protein phosphorylation, as *srk2cf* mutant lines show a reduced induction of *ABI5* and *EEL* in response to drought, and SRK2C

interacts with and phosphorylates ABI5 protein directly based on a yeast two-hybrid assay (Mizoguchi et al., 2010; Umezawa et al., 2004). This effect is partially redundant with *SRK2D, SRK2E* and *SRK2I*, however, as a *srk2dei* mutant line eliminated induction entirely (Mizoguchi et al., 2010).

Post-translational modification by SnRK2 phosphorylation appears to be vital for ABF protein function in particular. In *A. thaliana* seeds, ABI5 is phosphorylated in response to ABA treatment, and phosphorylation is required to maintain ABI5 stability and function (Lopez-Molina et al., 2003, 2001). Additionally, a protein kinase inhibitor prevented activity of two other non-seed ABF proteins, *ABF2* and *ABF4* (Uno et al., 2000), and both of these proteins as well as *ABF3* have been shown to interact with an SnRK2 protein kinase directly (Yoshida et al., 2010). A similar pattern of SnRK2-mediated *ABI5*/ABF protein activation is seen in rice and wheat, and is probably evolutionarily conserved across most land plants (Johnson et al., 2002; Kobayashi et al., 2004a; Komatsu et al., 2013).

1.3.1.5 Long-lived mRNAs

Although a large proportion of the maturation process is involved in preparing the embryo for desiccation and dormancy, the seed also accumulates mRNA transcripts, proteins and other molecules that may not function during maturation itself but are indispensable for successful germination.

After imbibition, protein translation and mRNA transcription occur very rapidly. A transcriptomic study in germinating *A. thaliana* seeds determined that the mRNA profile of imbibed seeds was distinguishable from the profile of dry seeds within 6 hours, implying that the transcriptional machinery was functional within at least this time (Nakabayashi et

al., 2005). Of particular importance to seed germination, however, is *de novo* protein synthesis. *A. thaliana* seeds can successfully germinate even in the presence of α-amanitin, a transcriptional inhibitor, whereas those treated with cyclohexamide, which inhibits translation, cannot (Rajjou et al., 2004). This suggests that during germination the majority of *de novo* translation occurs from transcripts, known as long-lived mRNAs, which are already present in the imbibed seed. Transcriptomic analysis of long-lived mRNAs that accumulate during maturation suggests that many are themselves necessary for protein synthesis and turnover, such as gene transcripts encoding ribosomal proteins and translational initiation factors (Nakabayashi et al., 2005). Additionally, the presence of multiple transcripts associated with protein degradation, including both protease/peptidase enzymes and those related to the ubiquitin-proteasome system, support a model whereby proteins are rapidly degraded upon germination (Galland et al., 2014; Nakabayashi et al., 2005). This includes recycling of proteins as a nitrogen source, such as degradation of SSPs, and specific degradation of germination repressors (Fujiwara et al., 2002; Galland et al., 2014). For example, ABI5 is specifically ubiquitinated and targeted for 26S-proteasome-mediated degradation in the absence of ABA, which reverses its inhibition of germination (Lopez-Molina et al., 2003, 2002, 2001).

Long-lived mRNAs also appear to be important for rehydration of desiccated resurrection plants. *De novo* transcription is not required for rehydration of dried *X. humilis* leaves, for example, although *de novo* protein synthesis from stored mRNAs is essential (Dace et al., 1998).

1.3.2 Gene networks involved in seed germination

The ABA-induced seed maturation gene network not only prepares the seed for desiccation, but also inhibits growth and germination. Germination is induced in non-dormant imbibed seeds by gibberellin-regulated gene networks that reduce ABA concentrations in the seed and silence maturation-specific genes, thereby reversing the maturation-induced embryo quiescence (Finch-savage and Leubner-metzger, 2006). The end-point of germination is generally defined as the stage at which the embryo radicle induces endosperm rupture and emerges from the seed (Kucera et al., 2005). However, the period from initial imbibition until radicle protrusion is under strict hormonal control, and will only continue to completion if external environmental cues indicate conditions are suitable for growth.

The growth-inhibiting stress hormone ABA and the growth hormone GA have an antagonistic effect on seed germination, and each other. ABA induces the switch from embryogenesis to seed maturation, and increases in concentration throughout this process until it peaks in the dry seed. ABA-induced maturation genes repress GA biosynthetic processes during maturation, such that GA concentration is low in the dry seed. Thereafter, the ratio between these two hormones controls the disposition of the seed to germinate. ABA-mediated inhibition of germination is controlled by germination-repressing factors that accumulate during seed maturation and are present in high levels in the dry seed (Piskurewicz et al., 2008). Thus, high ABA and low GA concentrations promote embryo quiescence and inhibition of germination under non-ideal conditions, whereas high GA and low ABA concentrations promote growth and germination under ideal conditions by antagonising the germination-inhibitory factors (Nambara and Marion-Poll, 2005; Olszewski et al., 2002). During the initial stages of imbibition, water uptake is driven by the low water

potential of the desiccated seed and is not under control of the embryo (Bewley and Black, 1994). Under ideal germination conditions, *de novo t*ranscription and translation then occur rapidly, and are concomitant with upregulation of genes involved in GA biosynthesis and ABA catabolism, and the degradation or down-regulation of gene products or genes involved in repressing germination (Debeaujon and Koornneef, 2000; Lee et al., 2002a; Nakabayashi et al., 2005). An important class of proteins that accumulate during seed maturation and that are involved in negative regulation of GA are the DELLA subgroup of the GRAS transcription factor family. DELLA proteins are master growth regulators that inhibit GA signalling in multiple physiological and developmental processes, including flowering, germination and stress (Sun and Gubler, 2004; Thomas and Sun, 2004; Ueguchi-Tanaka et al., 2007). *A. thaliana* contains genes encoding five DELLA proteins: *GIBERRELIC ACID INSENSITIVE* (*GAI*), *REPRESSOR OF GA1-3* (*RGA*) and three *RGA-like* (RGL) genes, *RGL1*, *RGL2* and *RGL3* (Tyler et al., 2004). Although all of these genes are expressed during seed germination, only *RGL2* has been shown to be indispensable for the repression of germination in *A. thaliana* (Lee et al., 2002b; Tyler et al., 2004). A common theme of DELLA protein action is that they are deactivated by polyubiquitin-mediated protein degradation. During germination in *A. thaliana*, GA binds to the cytosolic GA-receptor *GIBBERELIN-INSENSITIVE DWARF1-like* (*AtGID1*), which targets *RGL2* to an SCF E3-ligase complex and ultimately ubiquitin-mediated proteasomal degradation (Feng et al., 2008; Sun and Gubler, 2004).

However, for a transient period during and shortly after germination, ABA levels can be re-induced by stressful conditions, leading to growth arrest and the reacquisition of a tolerant, quiescent state. Prior to germination, Piskurewicz et al. (2008) showed that the DELLA

protein *RGL2*, possibly acting in concert with an E3-ubiquitin ligase, *XERICO*, enabled an expression feedback loop that results in increased ABA biosynthesis in the seed. The increased ABA, in turn, causes an increase in expression of *RGL2*, and ultimately prevention of GA-mediated germination. The authors identified *ABI5* as the final repressor of germination in this process, specifically the SnRK2-phosphorylated form of the protein (Piskurewicz et al., 2008). Accumulation of *ABI5* mRNA and protein has also been implicated in a window of stress-induced growth arrest in developing seedlings, which are able to re-acquire DT and enter a quiescent state in response to acute stress (Finkelstein and Lynch, 2000; Lopez-Molina et al., 2001). ABI5 protein levels are further regulated in the germinating seed and seedling by ABI5 binding proteins (AFPs), which form a nuclear complex with ABI5 and target it for 26S-proteasome-mediated degradation (Lopez-Molina et al., 2003). In *A*. thaliana, *AFP2* expression is most similar to that of *ABI5*, as it is abundant in dry seeds and stressed seedlings. *AFP1, AFP3* and *AFP4* are expressed in older plants, and may interact with the vegetative ABF proteins (Garcia et al., 2008).

1.3.3 The point of no return

Embryo growth arrest can be induced during germination and for a brief period thereafter by acute stress. Not only does the embryo re-enter a quiescent state, but it re-acquires a stress and DT phenotype much like that of the mature embryo. In orthodox seeds of desiccation sensitive (DS) angiosperm species, seedling DT is generally progressively lost during germination, until the seedlings reach a so-called "point of no return" past which they can no longer reacquire DT and thus fail to survive dehydration (Evenari et al., 1971). For many species this point is reached early during the imbibition stage, prior to visible protrusion of the radicle through the testa (Maia et al., 2011; Negbi and Evenari, 1961). In *A.*

thaliana, the window is very narrow – lasting only approximately 48 hours from imbibition – and requires the action of both *ABI3* and phosphorylated ABI5 protein (Lopez-Molina et al., 2002, 2001). In other species, however, the point of no return can be extended until well after radicle emergence. This phenomenon is often seen in desert plant species, where the ability for a seedling to survive sudden and unexpected drought conditions unharmed would confer a significant selective advantage (Friedman et al., 1981; Gutterman, 2000; Huang et al., 2004). "Priming" of seeds by exposure to mild osmotic stress, such as incubation in high molecular weight polyethylene glycol (PEG), can extend the developmental window during which DT can be re-established in germinating seedlings of many species that would not be able to do so under normal conditions (Bruggink and van der Toorn, 1995), albeit narrowly in many cases (Buitink et al., 2003; Maia et al., 2011; Vieira et al., 2010). This suggests that the genetic capability to re-acquire DT may be a widespread trait in angiosperm seedlings, albeit one that is often not observed under standard experimental conditions. However, in all researched non-resurrection plant angiosperm species, the ability to re-induce DT in seedling tissues is lost shortly after germination as the plant switches to vegetative growth.

Analysis of gene expression in *A. thaliana* seedlings revealed that *ABI5*, as well as genes encoding seed-specific LEAs, seed storage and dormancy related proteins, were strongly up-regulated following PEG treatment, while those involved in energy metabolism and cell-wall modification were repressed (Fig. 1.4; Maia et al. 2011). Similarly, transcriptome analysis of *M. trunculata* seedlings undergoing re-establishment of DT following PEG treatment revealed a substantial overlap with changes in gene expression observed during the establishment of DT during seed maturation, including upregulation of *ABI3*, *ABI5*, LEA, SSP and oleosin genes (Terrasson et al., 2013). Together these data suggest that re-

establishment of DT in seedlings involves a return to a quiescent state similar to that occurring in mature orthodox seeds. However, not all genes shown to be differentially expressed during the re-induction of seedling DT in *M. trunculata* are differentially expressed during seed maturation, suggesting the possibility that these processes may occur via separate, but overlapping pathways (Buitink et al., 2006; Terrasson et al., 2013). Work by Maia et al. (2014) in the germinating seedlings of *A. thaliana* noted differences in the induction of DT by ABA-response mutant seeds and seedlings. Plants with weak mutations in *ABA2* (an ABA biosynthesis gene), *ABI3*, *ABI4* (an AP2/ERF TF) and *ABI5* could still produce DT mature seeds, and embryos could still be prompted to re-induce DT in response to mild osmotic stress at the very earliest stages of germination (testa rupture). However, all tested mutants displayed impaired ability to re-acquire DT at later stages of germination (post-testa rupture) compared to wild type seeds, implying that these two processes may occur via separate pathways. However, another likely scenario is that many of the redundant factors present during seed development (e.g. other components of the LAFL network, *ABI4*) that could mask the effect of these individual alleles on seed DT during maturation are absent after germination (Maia, 2014). Seeds with a strong *abi3* mutant allele failed to induce DT during maturation or germination, thus *ABI3* is essential in both situations. The other *ABI* genes also play important roles during seedling DT, because mutations in these genes reduced the efficacy of DT induction. A similar trend is seen in *M. truncatula*: mutant *Mt-abi3* seeds are DS even after maturation and cannot induce seedling DT; *Mt-abi5* mutant seeds are DT after maturation, but seedlings fail to re-acquire DT after PEG treatment (Terrasson et al., 2013). Interestingly, the re-induction of DT in *A. thaliana* seedlings does not require elevated ABA content, as is necessary for *RGL2*-mediated inhibition of germination (Piskurewicz et al., 2008). Despite the increased expression of ABA

biosynthesis genes, levels of ABA and its catabolic products remained the same in DS and DT seeds (Maia, 2014). Baseline levels of ABA are still essential, however, considering *aba2-1* mutants seedlings that lacked ABA could not re-induce DT. Seeds instead appear to become more sensitive to ABA, possibly by increased expression of some of the PYL-family ABA receptors, such as *PYL7* and *PYL9* (Maia 2014).

An overview of the processes that are believed to occur during reacquisition of DT within the window of seedling DT are shown in Figure 1.4. Orthodox seeds are DT until germination, after which DT is progressively lost. Prior to the point of no return the application of stress (e.g. exogenous ABA/PEG) results in increased ABA sensitivity (via increased expression of PYL proteins) but not necessarily an increase in ABA concentration, and ultimately *ABI5*/*ABF*-mediated growth arrest. LEA genes, seed maturation proteins (oleosins, SSPs) and the antioxidant system are re-activated in the arrested seedling, and it re-enters a quiescent, DT state. Despite these data, it is unclear whether embryo DT and seedling DT are in fact controlled via distinct, separate gene networks. Maturation occurs over a period of days via an integrated developmental pathway aimed at preparing the embryo for quiescence and DT, which often involves sets of partially redundant factors with overlapping functions (e.g. the LAFL network genes) (Bewley et al., 2013). On the other hand, *ABI3*- and *ABI5*-mediated osmotic priming by PEG treatment can occur over a period of hours, and is likely induced alongside general vegetative stress-response pathways and the antagonistic GA-driven germination pathways, neither of which would be present in the developing seed. This may explain the difference in genes differentially expressed between embryo maturation and DT seedling stages in *M. trunculata* (Buitink et al., 2006).

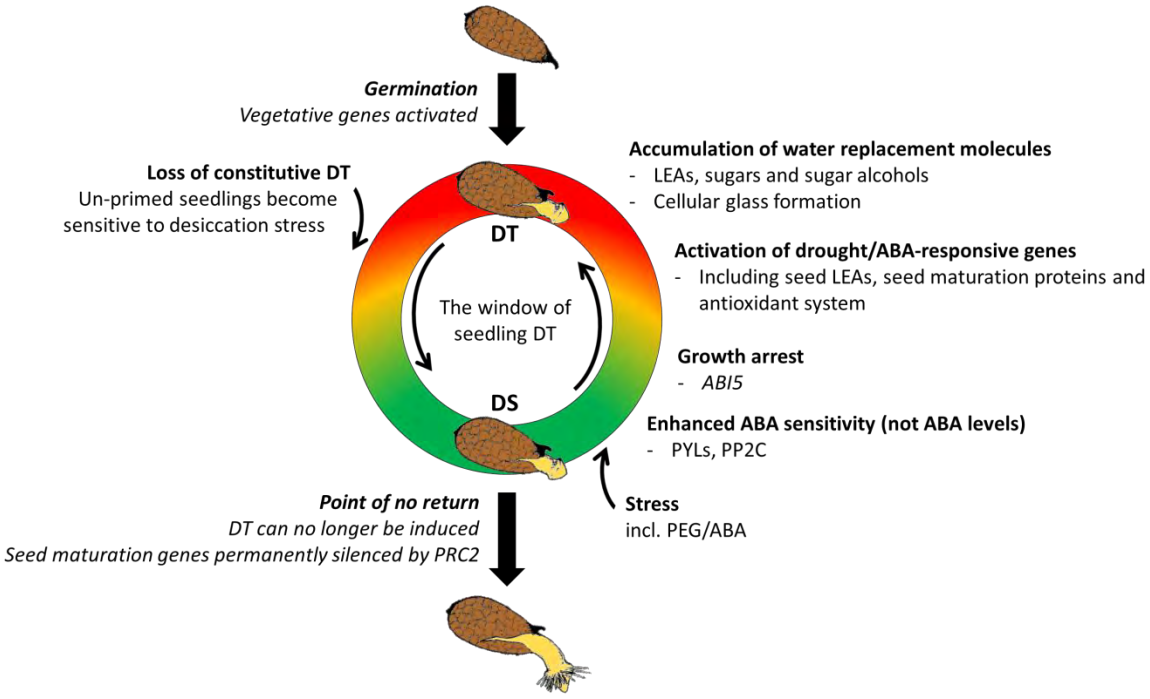


Figure 1.4. Theoretical model of the main events controlling the re-establishment of DT in the seeds of DS plants. Stress (PEG) or exogenous ABA increases seed sensitivity to ABA, possibly via PYL-type ABA receptors, resulting in growth arrest and activation of seed gene networks by *ABI5*. Activated targets include LEA genes, antioxidants and sugars, similar to those involved in seed maturation drying. Once the seedling has developed past the point of no return, DT can no longer be acquired. Modified from Maia (2014).

1.3.4 Epigenetic networks involved in seed DT

Shortly after germination in most angiosperm plant species the developing seedling loses the capability of acquiring DT in response to abiotic stress. In *A. thaliana*, this is associated with an inability of the seedling to induce *ABI5* expression, even after PEG priming or treatment with exogenous ABA (Lopez-Molina et al., 2002, 2001; Maia, 2014). From an evolutionary stand-point, this transition represents the committal of the developing seedling to vegetative growth, likely driven by a trade-off between the needs of the plant to survive early stress and also out-compete its neighbours by not restricting its own growth

unnecessarily. There is a large body of evidence showing that the switch from embryonic to vegetative growth in seed-bearing plants is controlled by epigenetic gene regulation; specifically, permanent silencing of the cohort of seed maturation genes, such as *ABI5*, by the deposition of the repressive histone 3 lysine 27 trimethylation (H3K27me3) epigenetic mark by the Polycomb Repressor Complex 2 (PRC2) (Bouyer et al., 2011).

1.3.3.1 Epigenetic gene regulation: an overview

Epigenetic regulation of gene expression is one of the key methods for regulating transitions between developmental states, which generally involves altering the expression of a large number of genes simultaneously. In plants this includes the switch from embryonic to vegetative growth, or the induction of reproductive processes such as flowering and fertilisation. The two principal modes of epigenetic gene regulation include direct DNA modification, generally cytosine methylation, or post-translational modification of the histone proteins that make up chromatin (Bird, 2002; Goldberg et al., 2007; Jenuwein and Allis, 2001). Whereas DNA methylation is generally associated with long-term gene repression, such as silencing of transposable elements or sex chromosomes, covalent histone modifications can lead to dynamic changes in both gene activation and repression.

In plants and other eukaryotes, the genome is packaged in the cell in the form of chromatin: a DNA/protein complex consisting of short loops of genomic DNA (approximately 147 bp) wrapped around nucleosomes – octameric protein complexes made up of histone proteins. Nucleosome octamers are assembled from two pairs each of histones 2A, 2B, 3 and 4. A fifth histone protein, linker histone 1, associates with the approximately 20-90 bp of genomic DNA that connects a nucleosome to the ones next to it (Margueron et al., 2005; Strahl and

Allis, 2000). Though all of the histone proteins can be post-translationally modified, it is most prominent on the N-terminal tails of H3 and H4 histones. Covalent modifications of H1 are by far the least characterised, although they do exist (Happel and Doenecke, 2009; Strahl and Allis, 2000). Variants of the canonical histone proteins have been identified in multiple species. Although they are structurally almost identical to the canonical histones, histone variants can have specifically altered residues that affect the repertoire of epigenetic marks with which they can be modified or how they interact with chromatin remodelling complexes (Happel and Doenecke, 2009; Szenker et al., 2011). For example, the mammalian H3.3 variant is preferentially found in nucleosomes of genes that are actively transcribed. The amino acid sequence of H3.3 differs in only 4 amino acids from the H3.1 and H3.2 histones, which allows it to interact with histone chaperone proteins that target it for replacement into active chromatin (Szenker et al., 2011).

The N-terminal tails of the four histone subunits that make up nucleosomes can be modified by methylation, acetylation, phosphorylation, ubiquitination and a variety of other chemical molecules (Zentner and Henikoff, 2013). It is believed that combinations of histone modifications act as an epigenetic "code", that dictates the expression status of the surrounding DNA (Jenuwein and Allis, 2001). This can be either directly, by structurally modulating DNA or nucleosome accessibility to transcription factors; or indirectly, by enabling or inhibiting the recruitment of protein complexes that lead to transcription (Bannister and Kouzarides, 2011; Bell et al., 2011; Patel and Wang, 2013; Petesch and Lis, 2008). Of the various histone modifications, histone lysine acetylation and methylation are the best characterized in various species, including plants. Histone acetylation is generally correlated with active gene expression, whereas histone lysine methylation is often

associated with gene repression (Martin and Zhang, 2005; Pikaard and Scheid, 2013). However, this is not a concrete rule for either family of modifications and depends on the degree of modification (e.g. mono-, di- or tri-methylation), the specific lysine residue modified, and the presence of other epigenetic marks (Margueron et al., 2005; K. Zhang et al., 2007). Both histone acetylation and methylation can be controlled enzymatically through histone acetylases/methylases (HAT and HMT proteins), that deposit the post-translational modification, and histone deacetylases/demethylases (HDAC and HDM proteins), that remove the corresponding modifications. The epigenetic status of chromatin regions can also be controlled indirectly by histone replacement in nucleosomes, thus resetting the combination of epigenetic marks to those on the incoming histones, as in the example of targeted replacement of H3.3 into active chromatin regions in mammals. Histone-modifying enzymes are often encoded by large gene families in plants and different components can display developmental or tissue specificity (Berr et al., 2011; Deal and Henikoff, 2011; Lauria and Rossi, 2011).

1.3.3.2 PRC2 and PRC1: regulators of developmental gene repression

One of the most well-known of the histone-modifying enzyme complexes is the Polycomb Group (PcG) family complex *PRC2*. PRC2 is a histone methyltransferase complex that regulates the deposition of the repressive H3K27me3 epigenetic mark, resulting in silencing of its target genes. In both *Drosophila* and mammals, PRC2 controls the expression of a large complement of genes during early development, preferentially silencing genes related to cell growth and division (Lund and van Lohuizen, 2004; Schwartz and Pirrotta, 2008). Similarly, approximately 15% of all genes in *A. thaliana* are marked by repressive H3K27me3 marks, and this PRC2-mediated repression is usually associated with genes that show specific

spatial or temporal expression patterns, particularly relating to developmental processes (Engelhorn and Turck, 2010; X. Zhang et al., 2007).

The *Drosophila* PRC2 complex is made up of four proteins: E(z), Su(z), NURF-55 and ESC (Lund and van Lohuizen, 2004). This tetrameric structure is maintained in plants, but many of the individual subunits have undergone gene expansion. *A. thaliana*, for example, has three E(z) homologs (*SWN, MEA* and *CLF*), three Su(z) homologs (*FIS2, EMF2* and *VRN2*), two known NURF-55 homologs (*MSI1* and *MSI4*/*FVE*) and a single ESC homolog, *FIE* (Fig. 1.5; Hennig & Derkacheva 2009). Alternative combinations of these subunits can affect the specificity of the resulting PRC2 complex, and the expression of certain subunits can be restricted by tissue type or developmental stage (Guitton et al., 2004; Pien and Grossniklaus, 2007). The existence of a PRC1 complex in plants has only recently been confirmed, although much of the information regarding its structure and function are still unknown (Hennig and Derkacheva, 2009). In *A. thaliana*, the PRC1 complex is believed to consist of two subfamilies of ubiquitin ligases (RING1 subfamily *RING1a* and *RING1b*; BMI subfamily *BMIa, BMIb* and *BMIc*), and a single H3K27me3-recognising protein, *LHP1* (Fig. 1.5; Molitor & Shen 2013). However, the *A. thaliana* ubiquitin ligase gene sub-families (*RING1a*/*b* and *BMIa*/*b*/*c*) correspond with a eudicot-specific gene expansion, so are not directly applicable to monocots (Berke and Snel, 2015). Other PcG proteins have been hypothesised to have PRC1-like functions in plants, but have not yet been shown to form part of any PRC1 complex (Calonje, 2014). One example is *EMF1*, which interacts with H3K27me3 enriched sites in the *A. thaliana* genome based on tile-based ChIP assays, and is likely involved in PRC2-mediated silencing of multiple developmental processes in plants. Some of its target genes include

ABI3, *FUS3* and *LEC2*, as well as LEA, oleosin and SSP genes, indicating it may play a role in silencing the seed maturation pathway (S. Y. Kim et al., 2012).

In animals, PRC2 is targeted to genes to be silenced by recognition of nucleosome-depleted polycomb response elements (PREs) within chromatin, where the PRC2 methyltransferase component, E(z), catalyses the tri-methylation of H3K27. A second PcG complex, PRC1, recognises H3K27me3 and promotes its bi-directional spreading to adjacent nucleosomes via H2AK119 ubiquitination, resulting in very large, multi-kilobase domains of H3K27me3-modified chromatin across the genome (Müller and Kassis, 2006; X. Zhang et al., 2007). The mechanism by which PRC2 is directed to its target genes in plants is less well understood. A few plant genes appear to contain PRE-like sequences that may influence PRC2 binding, but it is not clear if this is a common trait in plants (Berger et al., 2011; Guo et al., 2008; Lodha et al., 2013). Unlike animals, these putative PRE regions are not depleted in nucleosome content (X. Zhang et al., 2007). There is also evidence that long non-coding RNAs (lncRNA) may play a vital role in targeting plant PRC2 complexes (Heo and Sung, 2011). The interaction between PRC1 and PRC2 also appears to be very different in plants. Zones of H3K27me3-modified chromatin are small in plants, and generally associated with single target genes as opposed to local chromatin domains spanning hundreds of kilobases as in animals (K. Zhang et al., 2007). H3K27me3 and H2AK119Ub1 epigenetic modification also commonly occur independently in plant genomes (Yang et al., 2013). In fact, H2AK119Ub1 appears to be required for the deposition of H3K27me3 in a subset of PRC2-silenced genes, including those involved in seed maturation (Calonje, 2014; Molitor et al., 2014). Thus, PRC1 appears to sometimes function upstream of PRC2 in plants; the opposite of that which occurs in animals.

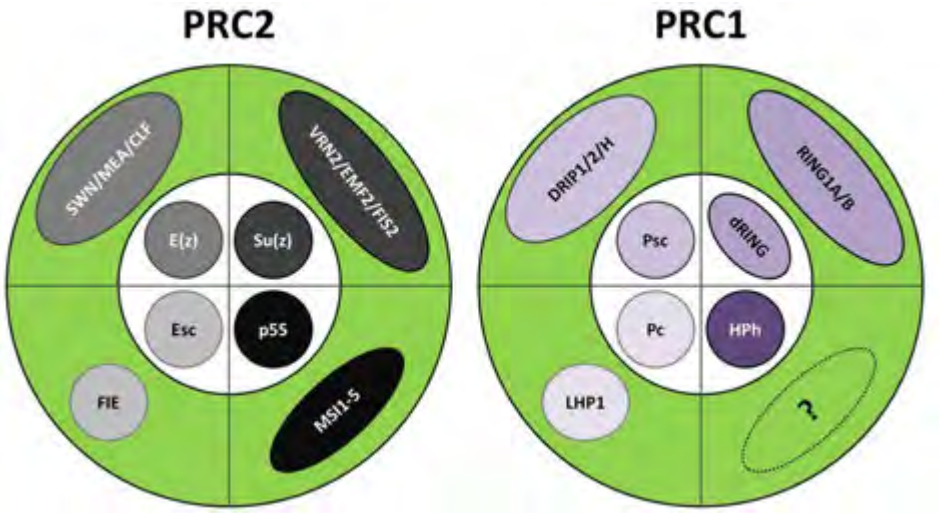


Figure 1.5. PRC2 and PRC1 protein subunits in *D. melanogaster* and *A. thaliana*. Canonical polycomb complex subunits for *D. melanogaster* PRC2 and PRC1 are shown in the centre circle (white). Known *A. thaliana* orthologue(s) of each subunit are shown in the outer circle (green). A plant orthologue of the PRC1 component Hph has so far not been identified.

1.3.3.3 Epigenetic control of the embryonic to vegetative transition

A functional PRC2 complex is essential for the transition from embryonic to vegetative growth during germination (Bouyer et al., 2011; Müller et al., 2012). Similarly, PRC1, which acts upstream of PRC2 within at least seed maturation gene promoters, is also vital for the repression of seed traits in germinating seedlings (D. Chen et al., 2010; Molitor et al., 2014; Yang et al., 2013). *Arabidopsis* mutant double knock-out lines lacking combinations of *PRC2* components exhibit de-repression of seed genes in vegetative tissue and a phenotype similar to that seen during somatic embryogenesis (Chanvivattana et al., 2004; Schubert et al., 2005). The repressive H3K27me3 modification in plants is generally mutually exclusive with the active H3K4me3 epigenetic mark, and it has been observed that the up-regulation of maturation genes in PRC2 mutant seeds is correlated with an increase in H3K4me3 in the promoters of those genes. However, loss of H3K27me3 did not correlate with a global

increase in H3K4 tri-methylation in most other PRC2-targetted genes, suggesting that an absence of H3K27me3 alone is not sufficient to induce gene activation in all target genes (Bouyer et al., 2011). The CHD3 chromatin remodelling enzyme, *PICKLE* (*PKL*) has also been implicated in the transcriptional repression of genes related to developmental identity in *A. thaliana*, including seed maturation genes (Dean Rider et al., 2003; Eshed et al., 1999; Ogas et al., 1999; Zhang et al., 2008). Roots of *pkl* mutant seedlings are visibly green and swollen (hence "pickle"), and display various embryonic traits due to a failure to repress seed maturation genes during germination (Dean Rider et al., 2003; Li et al., 2005; Zhang et al., 2008). It was observed that *pkl* mutants had decreased global H3K27me3 levels, including at the promoters of seed maturation master regulators like the *LEC* genes; *LEC1* and *LEC2* were likewise upregulated in the absence of *PKL*. It was thus assumed that *PKL* acted alongside the PRC2 complex in initiating deposition of H3K27me3 (Zhang et al., 2008). A later study by Aichinger et al. (2009) supported an indirect role for *PKL* during this process. In the roots of 5 day old *pkl* seedlings, various PRC2 components were down-regulated, including *EMF2*, *SWN* and *MEA*. ChIP analysis verified that the *PKL* protein bound to the promoters of these genes and other seed genes, but not to the promoters of *LEC1*, *LEC2* and *ABI3*. Additionally, the down-regulation of *PRC2* genes was associated with increased H3K27me3 levels at their own promoters. Interestingly, analysis of the other promoters bound by PKL indicated that *PKL* target genes were down-regulated in *pkl* mutants; thus, *PKL* functioned as a transcriptional activator with a role opposite to that of PRC2, rather than as a repressor as previously believed. The authors concluded that *PKL* protein activated PRC2 gene expression, leading to downstream gene silencing through H3K27me3 deposition. Thus, in *pkl* mutants, PRC2 remained repressed and subsequently the seed genes remained active (Aichinger et al., 2009). However, a second study by Zhang et al. (2012) contradicted these

results. In germinating seedlings – rather than in 5 day old seedling root, as in the previous study – PKL binding was indeed enriched at H3K27me3 loci, but PKL was not found to interact with PRC2 component gene promoters. PKL was however found to be associated with the *LEC1*, *LEC2* and *FUS3* genes, consistent with their upregulation. In 14 day old seedlings, these interactions were absent. However, during germination the *PKL* protein was additionally found to be associated with a number of genes that were constitutively expressed independent of *PKL* expression or the presence of H3K27me3, such as *ACT7* or *UBQ10*. Zhang and colleagues conclude three things from these results. Firstly, *PKL* does indeed function directly at the site of genes that are normally repressed during germination, including *LEC1*, *LEC2* and *FUS3*. Secondly, *PKL* becomes dispensable for the repression of seed genes during vegetative growth after germination has completed, possibly explaining the lack of association in the Aichinger et al. (2009) results. Thirdly, *PKL* appears to have a function independent of the presence of H3K27me3, and may indeed perform a general chromatin remodelling function throughout the genome (Zhang et al., 2012). Nonetheless, the precise role *PICKLE* plays in the repression of seed maturation genes during germination is still under scrutiny.

Not only histone methylation, but also histone acetylation plays a role during the germination-induced repression of seed maturation genes. The histone deacetylases *HDA19* and *HDA6* have been shown to be essential for the repression of embryonic traits in germinating seedlings. An *HDAC6-RNAi* line treated with trichostatin-A, an HDAC inhibitor, was arrested at germination and had increased levels of *LEC1*, *FUS3* and *ABI3* transcripts. This phenotype could be rescued by using *lec1*, *fus3* or *abi3* mutant plant backgrounds. *HDA6*/*HDA19* double *RNAi* lines displayed a similar phenotype without the

use of trichostatin-A, and could only be rescued in a *lec1* background (Tanaka et al., 2008). This strongly suggests that *HDA6*/*HDA19* act in a partially redundant fashion to suppress seed maturation genes that inhibit germination by removal of active histone acetylation marks from their promoters.

1.4 Gene networks involved in VDT and drought tolerance in adult tissues

The genetic networks that control plant drought and desiccation tolerance are complex, and relate to nearly all aspects of plant metabolism and physiology. Nonetheless, our current knowledge of these processes suggests a certain degree of commonality across the plant kingdom.

1.4.1 Desiccation tolerance mechanisms in lower plants

Current evidence points towards a seed-based origin for VDT in resurrection plants, rather than a re-emergence of the mechanisms seen in lower plants. However, many of proteins and mechanisms are nonetheless similar between the two VDT systems.

The physiological aspects of constitutive VDT in lower plants have been studied in multiple species from various locales, particularly those of the bryophytes (for example: Oliver et al. 2000; Robinson et al. 2000; Proctor et al. 2007; Makinde & Fajuke 2009). In general, bryophytes can survive rapid desiccation due to the constitutive presence of protective factors that maintain cellular and organelle integrity during drying. Consistent with this, bryophyte cells do not show signs of visible damage during desiccation, and transcription and translation processes are usually not induced (Oliver et al., 2005). Gene expression changes are only noticeable during early rehydration rather than dehydration, and *de novo*

protein synthesis also occurs rapidly during this period. Early transcription and translation is focused on the production of repair mechanisms, particularly a set of proteins known as rehydrins (Scott and Oliver, 1994; Wood and Oliver, 1999). The majority of rehydrins have no homology to known proteins at this time; subsequently, most of those that could be identified matched known desiccation- or light-inducible proteins, particularly LEAs, or ribosomal proteins that are presumably important for protein synthesis (Oliver et al., 2005). Unlike in DS plants, LEA proteins, particularly those of the dehydrin/LEA-2 group, are constitutively present at high levels in bryophyte cells (Bewley et al., 1993). Thus, although LEA gene expression is not induced during desiccation, they are still an important mechanism of osmoprotection in bryophytes. Similarly, soluble sugar concentrations do not increase during desiccation, but are instead present at high levels constitutively (Smirnoff, 1992). Although most bryophytes do not exhibit induced gene expression during desiccation like resurrection plants do, some have been shown to display increased DT after priming with exogenous ABA, concomitant with increased accumulation of protective proteins and sugars (Beckett, 1999; Bopp and Werner, 1993; Mayaba, 2001; Werner et al., 1991). In other species, such as *Tortula ruralis*, ABA has no effect on the induction of dehydrins and is not detectable in dry or drying tissues (Oliver and Bewley, 1997). This could suggest that an ABA-inducible VDT pathway has evolved in some bryophytes, challenging the previously-held belief that lower plants exhibit only constitutive VDT (Oliver et al., 2005; Stark et al., 2013).

The genetic underpinnings of the VDT trait in bryophytes have only recently begun to be unravelled. The first studies relied on expressed sequence tags (EST) from the desiccation tolerant moss, *T. ruralis* (Wood and Oliver, 1999). More recently, the model bryophyte

Physcomitrella patens has been of great importance, as its genome has been sequenced. Unlike *T. ruralis*, *P. patens* is an induced DT bryophyte that is DS under most conditions. However, after treatment with exogenous ABA, it exhibits VDT (Koster et al., 2010). Genomic and functional analysis of the genes involved in inducible-VDT in *P. patens* reveals a remarkable degree of overlap with the mechanisms seen in embryogenesis in seed-bearing plants, particularly in relation to ABA signalling (Marella et al., 2006; Sakata et al., 2010; Yotsui et al., 2013). Similarly, the transcriptomic desiccation response of the green alga *K. crenulatum* was determined to be similar to that seen in seeds in higher plants, suggesting that the basic principles of plant DT are inherited from the earliest ancestors of land plants (Holtzinger et al. 2014). Thus, although VDT as seen in resurrection plants is likely derived from the networks involved in seed maturation drying, this process is itself inherited from the DT mechanisms seen in early land plants.

ABI3 and ABREs

There is some evidence that a gene similar to *ABI3* is the founding member of the B3 gene family, which includes genes from the AFL network (Romanel et al., 2009). *ABI3*/*VP1* homologs have been identified in the algal species *Chlamydomonas reinhardtii* and *Volvox carteri*, and in various species of bryophytes (Romanel et al., 2009). The other AFL genes, such as *LEC2* and *FUS3*, do not appear to have orthologues in non-seed plants, and so likely represent a seed-plant-specific expansion of the B3 family that evolved to regulate seed development.

The *P. patens* genome contains three orthologues of *ABI3*/*VP1* (*PpABI3A-C*), all of which can activate *PpLEA1* gene expression in response to ABA via the native *PpLEA1* promoter and a wheat *Em*/*LEA* gene promoter, in both *P. patens* and barley tissues (Marella et al., 2006).

PpABI3 can also partially rescue the *abi3-6 A. thaliana* mutant seed phenotype, including loss of ABA insensitivity and reduced expression of seed storage and maturation genes. However, *PpABI3* cannot restore the expression of all LEA genes in the *abi3-6* background (Marella et al., 2006). There does not appear to be an *ABI5* orthologue in mosses; there are, however, multiple bZIP genes related to the rice group-C bZIP TF *OPAQUE2* (*O2*), which controls SSP expression during seed maturation by interacting with ABI3 and the ABRE (Alonso et al., 2009; Corrêa et al., 2008; Lara et al., 2003). Because *PpABI3* can rescue expression of *O2*-regulated maturation genes in *A. thaliana abi3-6* mutants (such as *CRC* and *At2S3*), but not *ABI5*-regulated genes (such as many LEAs) it is tempting to hypothesise that *O2*-related bZIP transcription factors regulated ABRE-mediated gene expression alongside ABI3 in early bryophytes. *ABI5* and the *ABI5*-related genes may represent an expansion of the bZIP transcription family in the descendants of mosses that have diversified in function (Sakata et al., 2010).

Conservation of ABA signalling pathway in plants

Recent studies have implicated group A PP2C protein phosphatases as major negative regulators of ABA signalling during development and stress in plants. *A. thaliana* mutants lacking functional *abi1* or *abi2* exhibit increased ABA sensitivity during germination and vegetative growth, leading to germination inhibition, increased dormancy and defects in drought response.

P. patens has two genes encoding group A PP2C phosphatases, designated as *PpABI1a* and *PpABI1b*. Interestingly, *P. patens* mutant plants lacking both *PpABI1* genes are fully desiccation tolerant even without ABA pre-treatment, indicating that they perform an essential role in negatively regulating the VDT response in this species (Komatsu et al.,

2013). However, unlike group-A PP2Cs in seed-bearing plants, those of *P. patens* do not interact directly with SnRK2 kinases to regulate the ABA response, but instead with unknown factors downstream of SnRK2 signalling - ultimately resulting in *ABI3*-mediated gene regulation (Fig. 1.6). As the bZIP TF(s) with which *PpABI3* interacts to regulate ABRE gene expression during VDT in *P. patens* is currently unknown, the role SnRK2 and protein phosphorylation plays in this process is not clear. However, it is evident that the ABA/PYL/PP2C/SRK2 network that plays a central role in the maturation of angiosperm seeds and the acquisition of embryonic DT, as well as in vegetative stress response, can be directly traced to a similar network that regulates VDT in bryophytes.

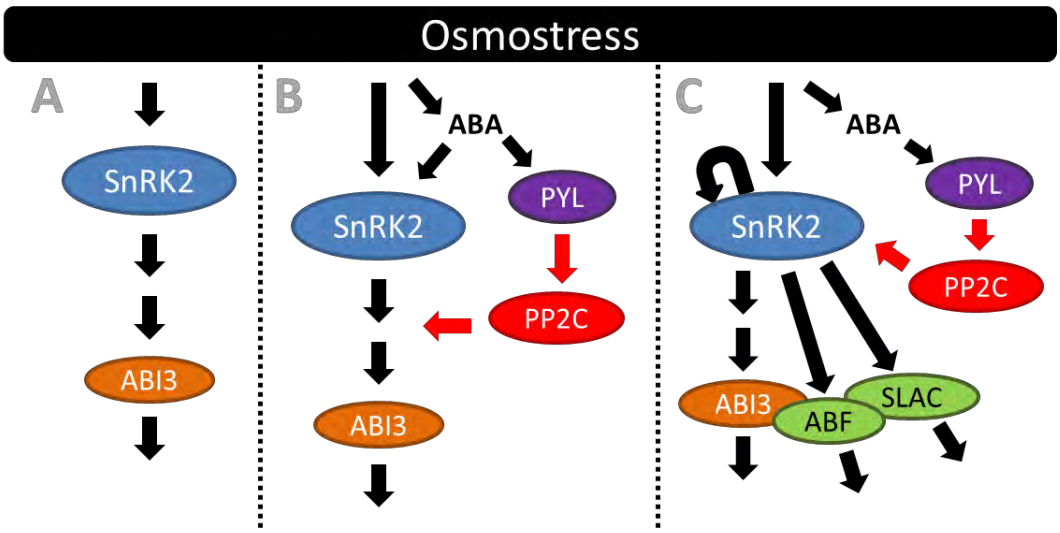


Figure 1.6. Evolution of the ABA-signalling pathway in land plants. A) *ABI3* played a central role in controlling VDT downstream of osmostress-activated SnRK2 kinases in the earliest ancestor of land plants. B) The ancestor of the bryophytes first recruited PP2C phosphatases into an ABA-PYL signalling network, although they did not interact directly with SnRK2 kinases at this stage. C) In the ancestor of angiosperms, SnRK2 is possibly activated by autophosphorylation. PP2Cs directly inhibit SnRK2 activation rather than interacting with later targets, allowing for integrated regulation of an increased number of pathways downstream of SnRK2. Black arrows = positive regulation, red arrows = negative regulation. Modified from Komatsu et al. (2013).

LEC1 and the evolution of land plants

LEC1, unlike *ABI3*, does not have a direct orthologue in bryophytes. Evidence suggests that it evolved early during the evolution of tracheophytes by duplication of an ancestral NF-YB gene, and first arose in the non-seed-bearing vascular plants (Xie et al., 2008). *LEC1* orthologues have been identified in multiple species of lycophytes and fern allies (Xie et al., 2008). The *Selaginella sinensus LEC1-like* gene, *SsLEC1*, can complement the *lec1-1* mutation in *A. thaliana*; thus, the *LEC1* orthologue seen in seedless plants is functionally equivalent to the gene expressed in embryogenesis in seed-bearing plants. An analysis of the expression pattern of the *Selaginella moellendorffii* orthologue, *SmLEC1*, showed that *SmLEC1* was preferentially expressed in both sexual and vegetative reproductive structures (strobili and bulbils, respectively), and may regulate lipid storage and the induction of DT in these structures (Kirkbride et al., 2013). *LEC1* transcripts also accumulate in the leaves of the VDT lycophyte, *S. sinensus*, and the DS fern, *A. capillus-veneris*, during the early stages of dehydration and rehydration, implicating the gene in drought stress response (Xie et al., 2008). The propagule and stress-based functions of the ancestral *LEC1*-like gene appears to have been co-opted by the seed-bearing plants to regulate similar processes specifically in embryogenesis and maturation of their seeds.

1.4.2 Drought tolerance mechanisms in DS plants

Drought tolerance in DS plants is a highly variable phenotype, but the underlying processes of how plants perceive and respond to water loss are similar across many plant species. As opposed to VDT plants, which "shut down" and enter a quiescent state, drought tolerance requires stressed plants to attempt to maintain long term growth and metabolism under

water limiting conditions. The response to water limiting conditions occurs in two phases in DS plants, known as the "acute response" and the "adaptive response" (Skirycz and Inzé, 2010). During the acute response, the plant downregulates unnecessary growth and metabolism and induces the expression of protective mechanisms, such as LEAs, as seen in VDT plants. The cessation of growth during the acute phase occurs independently of the rate of photosynthesis or carbon availability, indicating that it is an induced response rather than a secondary effect (Granier, 1999; Skirycz et al., 2010). During the adaptive response, the growth rate and metabolism of the stressed plant increase and stabilise as cellular systems adapt to acting under low water conditions. At both of these stages inhibition of growth is mediated by DELLA repression of GA signalling, similar to the process observed in seed maturation and germination. As well as directly antagonising GA biosynthesis and signalling, DELLA proteins indirectly activate stress tolerance genes by interacting with TFs, and may also play a role in protection from ROS (Achard et al., 2008, 2006; Navarro et al., 2008). During the adaptive response phase, when growth and metabolism have restarted, DELLA-GA antagonism is carefully maintained by complex cross-talk between multiple hormonal mechanisms, including ABA, ethylene and jasmonate, to assure controlled growth under stressful conditions (Skirycz and Inzé, 2010). Interestingly, DELLA-mediated growth inhibition appears to be an evolutionary conserved mechanism in angiosperms that is induced in response to multiple stresses (Achard et al., 2006).

The perception of low water conditions and subsequent genetic response occurs through a complex network of stress signalling pathways. However, these networks can be broadly divided into those that are induced in response to the stress hormone ABA, or those that are ABA-independent (Lata et al., 2011; Shinozaki and Yamaguchi-Shinozaki, 2006).

ABA dependent

ABA-dependent drought stress signalling in DS plants bears a remarkable similarity to that seen in lower plants and during seed maturation (Fig. 1.7; reviewed by Nakashima & Yamaguchi-Shinozaki 2013). The common PYL/PP2C/SnRK2 signalling cascade is the major regulator of ABA response during abiotic stress, and *A. thaliana* plants lacking SnRK2 genes, particularly *SRK2D*, *SRK2E* and *SRK2I*, show a cooperatively enhanced sensitivity to drought as well as ABA insensitivity. A similar phenotype is seen in mutants lacking *PYL* genes, which are severely insensitive to ABA during both germination and vegetative stress (Fujii and Zhu, 2009; Fujita et al., 2009; Gonzalez-Guzman et al., 2012). A bZIP family of *ABI5*-related transcription factors (ABF/AREB/ABL) act in a similarly cooperative fashion as regulators of drought tolerance in vegetative tissues, just as *ABI5* and its homologs function in seeds and seedlings. Four ABF genes in particular have been identified as "master" regulators of vegetative drought stress response: *ABF1*, *ABF2*, *ABF3* and *ABF4* (Yoshida et al., 2015, 2010). Mutant plants lacking functional copies of these genes have increased sensitivity to drought and decreased sensitivity to ABA, and impaired expression of classical drought response genes such as LEAs (Fujii and Zhu, 2009; Yoshida et al., 2015). *A .thaliana* plants over-expressing these genes exhibit enhanced drought tolerance (Fujita et al., 2005; Kang, 2002; Kim et al., 2004); over-expression of *ABF2*/*AREB1* and *ABF3* orthologues in rice and soybean have a similar effect (Barbosa et al., 2012; Oh et al., 2005). Like *ABI5*, the ABF genes responsible for drought stress response are activated by SnRK2-mediated phosphorylation (Fujita et al., 2009; Nakashima et al., 2009; Umezawa et al., 2013).

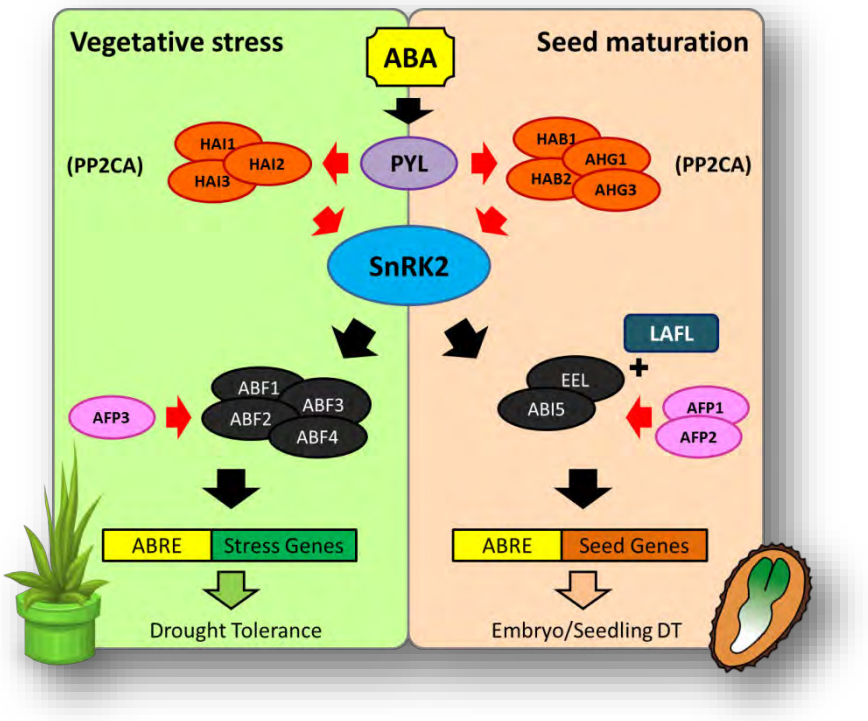


Figure 1.7. Model of the ABA-signalling pathways that control seed maturation and post-germinative stress in *A. thaliana*. Perception of ABA via cytosolic PYL receptors prevents the PP2CA-mediated inhibition of phosphorylation of ABF transcription factors via SnRK2 kinases, which activate downstream gene expression through the ABRE. ABF protein stability is also negatively regulated by AFP proteins. Despite tissue specificity of different group A PP2C genes, no such separation has yet been observed for PYL receptors or SnRK2 protein kinases. ABF proteins also interact cooperatively with other TFs that enhance specificity, such as those from the LAFL network during seed maturation. Black arrows = positive regulation, red arrows = negative regulation. Image modified from Nakashima & Yamaguchi-Shinozaki (2013).

ABA independent

ABA-independent gene expression during conditions of drought is primarily regulated by the dehydration responsive element (DRE). Despite the name, DRE elements are found in the promoters of genes associated with various abiotic stresses, including dehydration, heat, salt and cold (Yamaguchi-Shinozaki and Shinozaki, 1994). Expression of DRE-containing genes is regulated by *DEHYDRATION-RESPONSIVE ELEMENT BINDING (DREB)* proteins,

which belong to the AP2/ERF family of transcription factors. In *A. thaliana* the DREB protein family is further divided into two groups, designated *DREB1* and *DREB2* (Liu, 1998). The group 1 DREB genes studied so far (*DREB1A, DREB1B* and *DREB1C*) are induced only in response to cold, whereas the class 2 genes (*DREB2A*, *DREB2B* and *DREB2C*) are induced predominantly by drought, salt or heat stress (Liu, 1998; Nakashima et al., 2000). *DREB1C* has also shown to be induced by wounding (Nakashima et al., 2000), and probably acts as a negative regulator of the other *DREB1* genes to fine-tune their expression (Novillo et al., 2004). Early research into DREB protein activity focused primarily on the *DREB1* group; only recently have the functions of the *DREB2* group begun to be analysed (Lata and Prasad, 2011). Interestingly, alternative splicing of *DREB2* transcripts is believed to play a critical role in DRE-mediated stress response; however, the exact manner in which this is achieved is currently unknown (Egawa et al., 2006; Qin et al., 2007; Xue and Loveridge, 2004). Microarray analysis of DREB2-overexpressing transgenic plants indicates that they regulate the expression of multiple dehydration-responsive genes, including some LEAs; however, more research is needed to more clearly identify direct targets of the DREB2 proteins (Sakuma et al., 2006a).

Recently it has been shown that heat stress transcription factors (HSFs) may function downstream of the DREB2 transcriptional cascade. The HSF *HSFA3* is activated by *DREB2A* and *DREB2C* in response to drought in both maize and *A. thaliana*, where it regulates other HSF and heat shock protein (HSP) gene expression (H. Chen et al., 2010; Sakuma et al., 2006a; Schramm et al., 2008). HSF proteins are involved in abiotic stress response in plants, including responses to heat, drought, salt and osmotic stress. They generally act as transcriptional activators and are primarily involved in the induction of HSP genes, but may

have other targets as well (Banti et al., 2010; Nishizawa et al., 2006; Schramm et al., 2006). The HSF family in plants has expanded massively due to repeated duplication events, and can be divided into three groups: HSFA, HSFB and HSFC (von Koskull-Döring et al., 2007). These groups are broadly consistent between eudicots and monocots, but many of the HSF group C proteins represent a monocot specific expansion of the HSF family (Scharf et al., 2012). Similarly, the seed- specific *HSFA9* subgroup and the *HSFB3* and *HSFB5* subgroups appear to be specific to eudicots (Almoguera et al., 2002; Díaz-Martín et al., 2005; Kotak et al., 2007). Unfortunately, knowledge of monocot HSF families is currently too limited to deduce whether any have seed-specific functions. Most data suggest that the various HSF proteins work in a non-redundant fashion and, with the exception of TFs in the same subgroup, generally cannot replace each other in knock-out mutant lines (Scharf et al., 2012). HSF master regulators, that act upstream of other HSFs with more specific functions, have been identified in various species, including *HSFA1a* in tomato (Mishra et al., 2002), the entire *HSFA1* subgroup in *Arabidopsis* (Liu et al., 2011), and *HSF2Cb* in barley (Reddy et al., 2014).

Crosstalk between ABA-dependent and ABA-independent pathways

There is increasing evidence of cross-talk between the ABA-dependent and ABA-independent signalling pathways. The promoter of the abiotic stress induced gene *RD29A* contains both an ABRE and DRE, both of which are required for ABA-dependent gene expression (Narusaka et al., 2003). The *DREB1A*, *DREB2A* and *DREB2C* proteins also directly interact with ABF2, and overexpression of *DREB2C* affects ABA sensitivity (H. Chen et al., 2010). *DREB2A* gene expression is activated in an ABA-dependent fashion by binding of *ABF2*, *ABF3* or *ABF4* to an ABRE present within its promoter (Kim et al., 2011). Recently,

DREB2C has been shown to directly regulate ABA levels in *A. thaliana* via a complete DRE present in the promoter of the ABA biosynthesis gene *NCED9*. In the same study, *DREB2C* expression was induced by the application of exogenous ABA in germinating seedlings, suggesting that *DREB2C* is induced by both abiotic stress and ABA and, furthermore, positively regulates ABA signalling (Je et al., 2014). SRK2E kinase activity has been revealed to be regulated in an ABA-dependent or stress-dependent manner through different C-terminal domains of the SRK2E protein, and an exhaustive analysis of *A. thaliana* mutants lacking multiple SnRK2 genes identified multiple dysregulated target genes involved in ABA-independent pathways (Fujii et al., 2011; Yoshida et al., 2010). These results suggest that SnRK2 kinases may function as a central hub of both ABA-dependent and ABA-independent signalling during periods of stress (Nakashima and Yamaguchi-Shinozaki, 2013).

Of course, many other response factors and cis-regulatory elements interact with the ABRE and DRE stress response pathways. Known transcription factors include those from the MYB/MYC (Abe et al., 1997; Narusaka et al., 2003) , NAC (Fujita et al., 2004; Xu et al., 2013) and WRKY families (Chen et al., 2012), as well as others.

1.4.3 Desiccation tolerance mechanisms in resurrection plants

As discussed in the preceding sections, VDT has some marked differences to the mechanism of drought tolerance which is often seen in DS plant species. This includes, but is not limited to, the expression of many seed-specific genes such as LEAs, SSPs and oleosins; and entry into a dry, quiescent state during which they are metabolically inactive. While many of the genes or processes that are involved in VDT are known, the underlying transcription factors

and gene expression networks that regulate VDT in resurrection plants have not been well studied. Transcriptomic studies have identified TFs from multiple families that are highly differentially expressed during desiccation (including GRAS, HSF, MADS box, NAC, NF-YA and WRKY families), but further characterisation of individual TFs and their target genes has so far not been reported (Gechev et al., 2013). Genomic information on the commonly researched resurrection plants is also scarce; only two resurrection plant genomes have been published in the last year (*Boea hygrometrica* and *Oropethium thomaeum*), although this number will likely increase as NGS sequencing technologies become more available and affordable (Van Buren et al., 2015; Xiao et al., 2015). A lack of NGS genome data has hindered large-scale studies into conserved or unique cis-regulatory elements that are involved in VDT. Unsurprisingly, characterisation of the regulatory motifs of individual VDT-induced genes in resurrection plants has not currently resulted in the identification of novel elements, although it has confirmed the involvement of the ABRE in many cases (Bartels and Hussain, 2011; Ditzer and Bartels, 2006). ABA contents increase during desiccation in most tested resurrection plant species, and ABA-dependent signalling via the ABRE is likely essential for VDT. This does not rule out a role for ABA-independent signalling, however (Vicré et al., 2004). The involvement of non-coding RNAs in angiosperm VDT has also been reported. Studies in the resurrection plant *Craterostigma plantagineum* have identified two retrotransposon-like non-coding transcripts, *CDT-1* and *CDT-2* that, when overexpressed, can induce VDT even in the absence of exogenous ABA. Evidence indicates that *CDT-1* directs the synthesis of a 21-bp short interfering RNA (siRNA) that can induce the VDT response in *C. plantagineum* callus, including expression of desiccation- and ABA-induced genes. The second lncRNA, *CDT-2,* likely functions in a similar fashion

(Hilbricht et al., 2008). Thus, various non-coding RNA transcripts may play essential roles during VDT in other resurrection plant species.

1.5 Aims of the current study

Two related hypotheses were tested in this thesis: 1) VDT in resurrection plants evolved via the activation of the seed TF regulatory cascade in vegetative tissues in response to water loss, and 2) the predicted activation of seed TFs in vegetative tissue is due to the indefinite extension of the window in which germinating seedlings can reacquire DT in response to stress. An overview of these two models is shown in Figure 1.8.

Evidence for the first hypothesis includes the induction of a large number of canonically seed-specific genes in the vegetative tissues of desiccating *X. humilis* and other resurrection plants, and the similarity of the accumulated osmoprotectant molecules between resurrection plants tissues and orthodox seeds, in contrast to that seen in the general drought response of DS plants (Illing et al., 2005). One mechanism by which this could have evolved would be via drought and ABA-induced expression of the top-level "master regulators" of the seed maturation programme, such as the LAFL gene network or seed-specific ABFs, which are usually silenced in vegetative tissue. De-repression of these genes could then initiate the pre-established transcriptional cascade that results in quiescence and DT in orthodox seeds, but would instead result in DT in vegetative rather than embryonic tissues. The de-repression of seed maturation genes in vegetative tissue would require inhibition of the epigenetic complexes that usually prevent their expression, such as PRC2. A previous microarray study showed that the PRC2 methyltransferase, *SWINGER*, and a positive regulator of PRC2, *PICKLE*, were downregulated during VDT in *X. humilis* (Shen,

2014). This could signify that PRC2-mediated gene repression is reduced during desiccation, consistent with up-regulation of LAFL- and *ABI5*-regulated seed-specific genes (Fig. 1.8A). However, further investigation into the expression of *X. humilis* LAFL and PRC2 subunit genes during desiccation is required to substantiate this hypothesis.

In angiosperms, the seed maturation programme is usually only activated during late embryogenesis in orthodox seeds, somatic embryogenesis in vegetative tissue, and during a small window of stress-induced post-germination arrest/DT found in many angiosperm species. This window of seedling DT involves the *ABI3*-mediated re-activation of *ABI5* in response to stress, and subsequent re-accumulation of seed maturation gene products (such as seed-specific LEAs, SSPs and oleosins) in seedling tissues, "resetting" the seedling to a quiescent, embryo-like state. In DS plant species, the genes that control this process are permanently repressed by PRC2 once the seedling reaches a certain stage of development (the point of no return), and the seedling thus loses the ability to acquire DT. If these mechanisms are not permanently silenced in *X. humilis* but are instead re-activated by drought stress throughout the lifespan of the plant, it would provide an ideal mechanism for the controlled re-activation of seed genes in vegetative tissues during VDT. If this is the case, we would predict that *X. humilis* would not display a point of no return, but would instead retain the ability to induce DT indefinitely post germination (Fig. 1.8B).

Thus, the work presented in this thesis aimed to investigate the "point of no return" phenomenon in a resurrection plant, and to assess the potential role of the known seed maturation genes during VDT in this class of plants. In Chapter 2, I examine the response of developing *X. humilis* and *X. viscosa* seedlings to desiccation with or without priming by moderate osmotic stress or exogenous sucrose, in comparison to seedlings of the DS plant *A.*

thaliana. The similarities between the seedling and adult response are discussed. Chapter 3 describes the sequencing and *de novo* assembly of a *X. humilis* leaf desiccation transcriptome, as well as gene expression analysis across five stages of dehydration (100%, 80%, 60%, 40% and 5% relative water content). RNA-Seq studies were used to characterize the transcriptional response to desiccation, and to assess whether the desiccation transcriptome was activated by seed master regulators and/or by regulation of the PRC2 complex subunits and epigenetic modification, similar to the process that occurs in germinating seedlings.

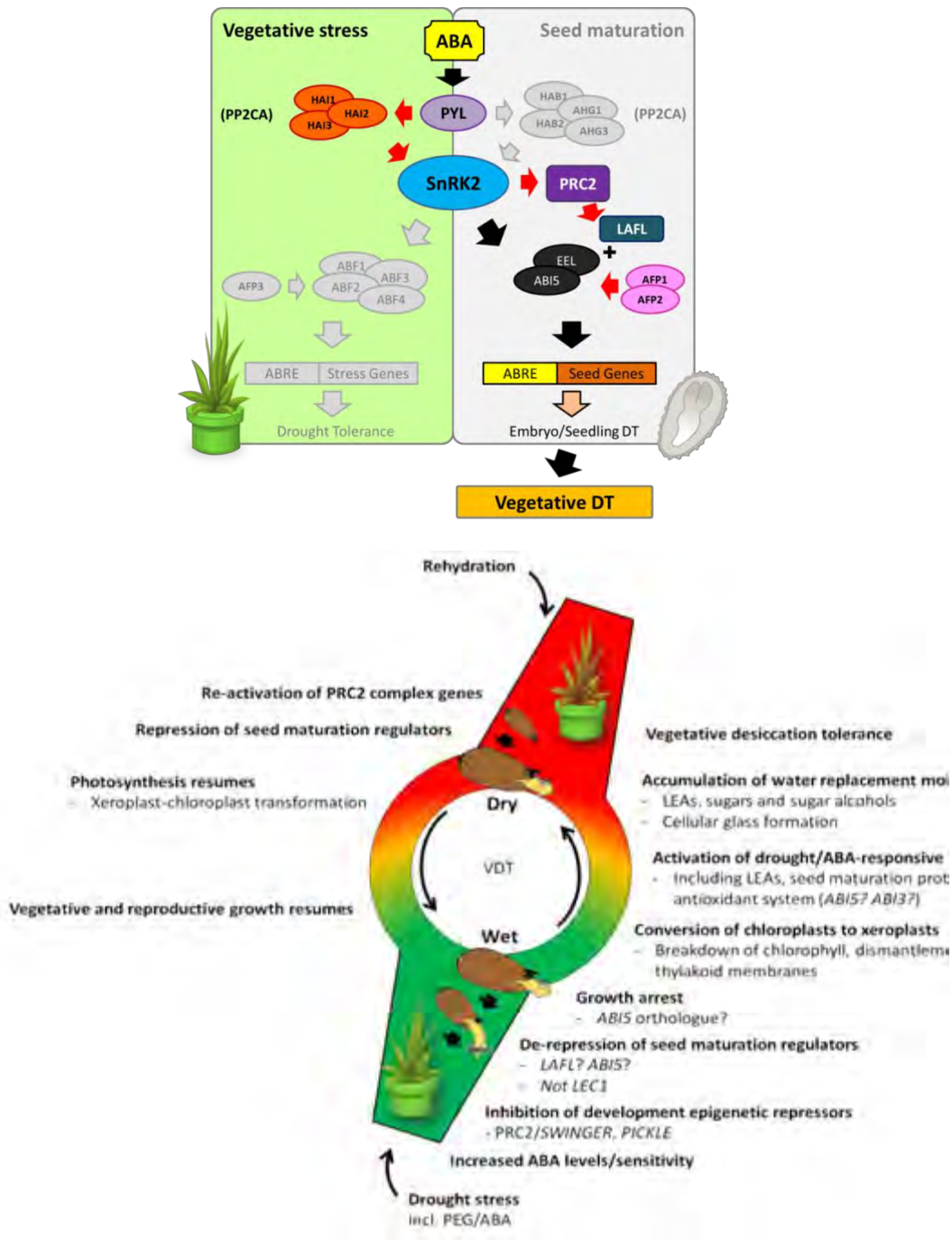


Figure 1.8. Predicted regulatory mechanisms involved in VDT acquisition in *X. humilis* based on the hypothesis that VDT evolved from co-option of the seed maturation network. A) Predicted activation of the seed maturation ABF gene network by vegetative stress in *X. humilis* compared to that seen in DS plants (see Fig. 1.7). Down-regulation of PRC2 component genes occurs as a consequence of ABA-induced stress response in this model, although it may not be due to direct action of SnRK2 as implied by this figure. B) In adult *X. humilis* plants, the window of opportunity in which seedlings can re-induce DT remains open indefinitely (compare to Fig. 1.4). Master regulators of seed maturation (e.g. the LAFL network, *ABI5*) are not irreversibly repressed in adults, but can be reactivated by down-regulation of the genes required to maintain H3K27me3 silencing over these genomic regions (e.g. *SWINGER*, *PICKLE*) resulting in the expression of seed genes in adult tissues.

Chapter 2: Post-germinative seedling DT and the "point of no return" in Xerophyta

A subset of the research in this chapter has been published independently, and can be found in full in the appendix.

Lyall R, Ingle RA, Illing N (2014) *The Window of Desiccation Tolerance Shown by Early-Stage Germinating Seedlings Remains Open in the Resurrection Plant, Xerophyta viscosa.* PLoS ONE 9(3): e93093. doi:10.1371/journal.pone.0093093

2.1 Introduction

Though VDT is rare in vegetative tissue, it is ubiquitous in the reproductive tissues and propagules of angiosperms, with the overwhelming majority (>95%) of species surveyed to date producing DT (orthodox) seed embryos and pollen (Gaff and Oliver, 2013). The seedlings of many DS species, such as *A. thaliana* and *M. truncatula*, can also re-induce DT during early development in response to stress, although this ability is quickly lost – the so-called point of no return (Buitink et al., 2003; Maia et al., 2011). The origin and mechanisms of VDT in angiosperms is a research topic of great interest, with relevance to both fundamental plant research and commercial interests in improving drought tolerance and seed viability of crop plants. In order to investigate the hypothesis that angiosperm VDT as seen in resurrection plants is derived from the same processes that control the window of DT in stressed seedlings, we analysed the response of the germinating seed of two resurrection plant species, *X. humilis* and *X. viscosa*, to desiccation stress.

Xerophyta is a genus of monocotyledonous plants in the family Velloziaceae found in Africa, Madagascar and the Arabian Peninsula (Behnke et al., 2013). Most are adapted for arid conditions and all tested species, including *X. humilis* and *X. viscosa* which are found

throughout Southern Africa, are classified as resurrection plants due to their extreme VDT (Behnke et al., 2013; Gaff and Oliver, 2013). *X. humilis* and *X. viscosa* and many other species in *Xerophyta* are poikilochlorophyllous; that is, they degrade their chlorophyll and dismantle their thylakoid membranes during desiccation in order to reduce the damage from reactive oxygen species (ROS) under water-limiting conditions (Tuba et al., 1998). Other species, such as the closely-related *Talbotia (Xerophyta) elegans*, are homochlorophyllous and do not lose chlorophyll upon desiccation. While the physiological response to desiccation of the mature plants has been studied extensively, little to nothing is known of the response of their seeds and seedlings to desiccation. If VDT has indeed evolved in resurrection plants via co-option of the genetic network responsible for the acquisition of DT in seeds, one possible mechanism by which this may have occurred is through extension of the developmental window during which seedling re-establishment of DT is possible. If this is so, we would predict that the seedlings of these resurrection plants would not display a point of no return, and would instead be able to re-establish DT throughout germination and beyond.

The aim of this section of work was to determine whether the germinating seedlings of the resurrection plants, *X. humilis* and *X. viscosa*, displayed a point of no return similar to that seen in other angiosperm species, such as *A. thaliana*, or if their seedlings retain DT throughout early development. Germinating seedlings of *A. thaliana*, *X. humilis* and *X. viscosa* were dehydrated rapidly at various stages of development, rehydrated after 48 hours, and their survival tracked over the next 5 days. Seedlings of these species were also pre-treated with either -2.5 MPa PEG-8000 for 48 hours, to simulate an extended period of mild osmotic stress, or grown on 3% sucrose (in case the ability to re-establish DT was limited by embryonic resources) in order to identify whether these treatments could increase

survival rate. Additionally, *X. humilis* and *X. viscosa* seedlings at specific developmental stages were dehydrated slowly over a number of days to compare the effects of a rapid and a reduced dehydration rate on the seedling desiccation response, or pre-treated for between 0 and 24 hours in PEG to further quantify seedling response to PEG priming. The seedlings in each experiment were tracked on an individual basis, including a photographic record of each step of the process to more accurately determine seedling size and physical description, and to track seedling survival.

2.2 Materials and Methods

2.2.1 Plant material

X. viscosa (Baker) seeds were obtained from Silverhill seeds (http://www.silverhillseeds.co.za), from wild plants located in Witsieshoek (Free State province, South Africa) and stored as packaged, at room temperature in the dark. Seeds of *X. humilis* (Baker) were obtained through pollination of plants collected from Borakalalo Game Reserve (Northwest Province, South Africa), maintained at the UCT Botany Greenhouse. Flowering plants were transferred to a plant growth chamber (Percival), where they were grown under a long-day photoperiod (16 hours light, 8 hours dark) at 22°C and cool white fluorescent light at 250 μmol $m^{-2}s^{-1}$. Pollen was transferred between flowers with gentle brushing, and fertilised flowers were allowed to ripen and dry on the parent plant. Mature seeds were excised from the ripened seed pods, separated from unfertilised seeds and organic debris, and stored in the dark at room temperature. Seeds of *A. thaliana,* Columbia (Col-0) ecotype were obtained from the European Arabidopsis Stock Centre (http://arabidopsis.info) and stored at 4°C in the dark.

2.2.2 Germination and growth conditions

In all cases, seeds were surface sterilised with 75% (v/v) ethanol for 2 minutes and air-dried in a laminar flow hood on sterile filter paper. Seeds were then germinated on half-strength Murashige and Skoog (MS) medium containing 0.7% (w/v) bacteriological agar, unless specified otherwise. Seeds of *X. humilis* and *X. viscosa* were germinated immediately while *A. thaliana* seeds were cold-stratified at 4°C for 48 h in the dark prior to germination. Plants

were grown under a long-day photoperiod (16 h light, 8 h dark) at 22°C, and cool white fluorescent light of 250 μmol $m^{-2}s^{-1}$.

2.2.3 Photography and seedling measurements

Seed and seedling photographs were obtained using a digital colour camera (JVC) mounted on a dissecting stereomicroscope (Olympus). Seedlings were photographed on a single piece of damp filter paper in order to prevent premature dehydration of their tissues due to strong lighting. Cotyledon, root and seed measurements were determined for each individual seedling using the ImageJ imaging software (http://rsbweb.nih.gov/ij/). High resolution photographs of seeds and seedlings were obtained using a Nikon Stereoscope Zoom Microscope (SMZ1500) and NIS-Elements (Nikon) digital 3D imaging software.

2.2.4 Response of X. viscosa seeds to various dormancy-breaking treatments

Batches of 25-30 *X. viscosa* seeds were treated with one of a number of dormancy breaking treatments previously described for various plant species (Emery, 1988): 4°C or 37°C incubation in the dark for 72 h on half strength MS plates; 72 h soak in the dark at 4°C or 37°C in either a 1% (w/v) thio-urea solution or 0.25% (w/v) potassium nitrate solution; 72 h stored at -20°C; acid scarification with a 50% (v/v) H_2SO_4 solution for 1 minute; 60°C soak in water for either 1 min or 1 h; twelve weeks 4°C moist chilling in constant dark on half-strength MS plates. The treated seeds were subsequently transferred to fresh half-strength MS medium plates and germinated under standard growth conditions.

2.2.5 *The effect of developmental stage on seedling survival after rapid dehydration*

A. thaliana seedlings were developmentally staged according to the system previously described by Maia et al. (2011). The developmental stages of *X. humilis* and *X. viscosa* germination have not previously been described, and so this process is described in the results section. For the purposes of the dehydration experiments, germinated seeds at specific developmental stage were transferred to damp filter paper in an open petri dish and dehydrated under constant air flow in a laminar flow hood. After 48 hours, seedlings were rehydrated on the filter paper with 2.5 mL sterile water and the plates wrapped in Parafilm® to prevent additional dehydration. Wrapped plates were transferred back to standard growth conditions. Seedlings were tracked daily over a 5 day period by inspection under an Olympus dissecting stereo light microscope, and considered DT if they resumed normal development.

2.2.6 *The impact of PEG and sucrose treatment on seedling revival after dehydration*

For the PEG-treatment, germinated seeds at appropriate developmental stages were moved to filter paper saturated with 1.2 mL of a PEG-8000 solution calculated to have an osmotic potential of -2.5 MPa at 22°. The plates were wrapped in Parafilm® to prevent evaporation of water altering the osmotic potential of the PEG solution, and were incubated under standard growth conditions for the specified time. Following incubation, the seedlings were rinsed carefully in sterile water to remove excess PEG and transferred to sterile filter paper dampened with water. Dehydration and rehydration proceeded as described above. For the sucrose treatment experiments, seeds were germinated on half-strength MS plates in the presence or absence of 3% (w/v) sucrose under standard growth conditions. Germinated

seeds at each developmental stage were then transferred to damp filter paper and subjected to dehydration as described above.

2.2.7 The effect of a reduced dehydration rate on seedling survival

Reduced dehydration rate experiments were performed on seedlings in a narrow range of developmental stages, rather than all experimental stages. Seedlings of *X. humilis* were germinated as described above. Plates of seedlings were grown for a pre-determined period of time calculated to give an average cotyledon size of approximately 0.5 mm and 2 mm. Additionally, *X. viscosa* seedlings were germinated and grown until 3.5 mm in length. Germinated seedlings of both species of the appropriate size were transferred to damp filter paper in a petri dish. The plates were closed but left unsealed so as not to prevent evaporation of water. Seedlings were allowed to dehydrate for a period of one week under standard growth conditions, and were then rehydrated with 2 ml distilled H_2O and the plates sealed and returned to standard growth conditions. Survival was tracked for five days after rehydration.

2.2.8 Statistical analysis

The seeds of both *Xerophyta* species, particularly *X. viscosa*, germinated in a slow and asynchronous manner markedly different to that of *A. thaliana* (germination defined as the piercing of the testa by the root tip). This, coupled with the difficulty in acquiring seeds of these species, meant that the experiments described occurred over the course of many months using seed batches of several hundreds to thousands of seeds. Rather than count these as biological repeats, data from multiple independent experiments was pooled and the variation within the entire pool of samples was determined via bootstrapping (Mooney et

al., 1993). The seedlings in each pooled experiment were resampled with replacement and placed into bins of pre-defined cotyledon lengths. This was repeated 10000 times, and the mean survival rate for the resampled seeds in each bin was calculated each iteration. The standard error of the mean (SEM) of the sample pools was calculated from the standard deviation of the bootstrapped mean survival values. The same procedure was applied to the data generated for the *Arabidopsis* germination experiments to facilitate comparison. Bin sizes for each experiment are given in Figure S2.2.

2.3 Results

2.3.1 Germination and growth of X. viscosa, X. humilis seedlings

The germination and subsequent development of *X. viscosa* and *X. humilis* seedlings has not been previously been reported in any controlled manner. We attempted to optimize germination conditions and developed a staging system to describe the growth of seedlings to aid further study on these plants.

The seeds of *X. viscosa* are generally elliptical in shape, approximately 1-1.5 mm in length, and surrounded by a triangular, dry, transparent husk that extends from one side of the seed (Fig 2.1A). When imbibed the seed swells with absorbed water and the embryo, located opposite the origin of the husk, is clearly visible (white arrow) (Fig 2.1B). The vast majority of the seed volume consists of nutrient endosperm, in which the embryo is embedded, and is surrounded by a thin, brown seed coat. Germination occurs over a period of 4 to 14 days thereafter. Germination rate and synchronicity were only marginally improved by various common dormancy breaking treatments, such as moist chilling (Fig. S2.1A-C). The germination of *X. viscosa* seedlings was found to resemble that of the onion, *Allium cepa*, albeit on a much smaller scale (Gray, 1887). Initially, the cotyledon lengthens within the seed and forces the developing radicle through the testa (Fig 2.1C). The embryo is reliant on nutrients from the endosperm for the first few days of germination and embryos excised during this stage are not viable. As germination progresses, the cotyledon continues to lengthen and rapidly turns green (Fig 2.1D). The radicle develops into a single primary root which develops root hairs once the cotyledons are between 1-3 mm in length (Fig 2.1E). At this stage, the embryo can be divided into four different zones: the cotyledon tip (a pale

translucent ball of tissue that absorbs food stores from the endosperm), photosynthetic cotyledon tissue, the meristem above the root, and a pointed root tip (Fig 2.1F). A central protoxylem is visible in cross-section of the germinating seedling, connecting the meristem, above the root zone, to the endosperm (white arrow) (Fig 2.1G). The cotyledon undergoes a period of rapid extension (to 4-5 mm in length), before a split in the leading end of the cotyledon (Fig 2.1Hi&ii) heralds the appearance of the first true leaf (Fig 2.1I). As the seedling develops further, additional leaves emerge from the meristem below the first leaf, and the cotyledon eventually senesces.

The seeds of *X. humilis* are very similar to *X. viscosa*, and their seedling development follows a near-identical pattern. Individual seeds are smaller than those of *X.* viscosa, at approximately half the length. The seed is oval in shape, and similarly surrounded by a dry, spongy husk. The embryo is located at one end of the seed, with the remainder of the seed volume taken up by a pale nutrient endosperm. Visible embryo growth and chlorophyll accumulation generally occurs rapidly after imbibition within the seed (within 24 hours), but germination – piercing of the radicle through the seed coat – takes between 1 and 6 days. As in *X. viscosa*, initial growth is categorised by a rapid extension of the photosynthetic cotyledon, while the cotyledon tip remains embedded in the endosperm. The primary root lengthens and develops a mass of root hairs. After the cotyledon reaches its maximal length of approximately 4 mm, the first leaf emerges through a visible slit in the meristem above the root.

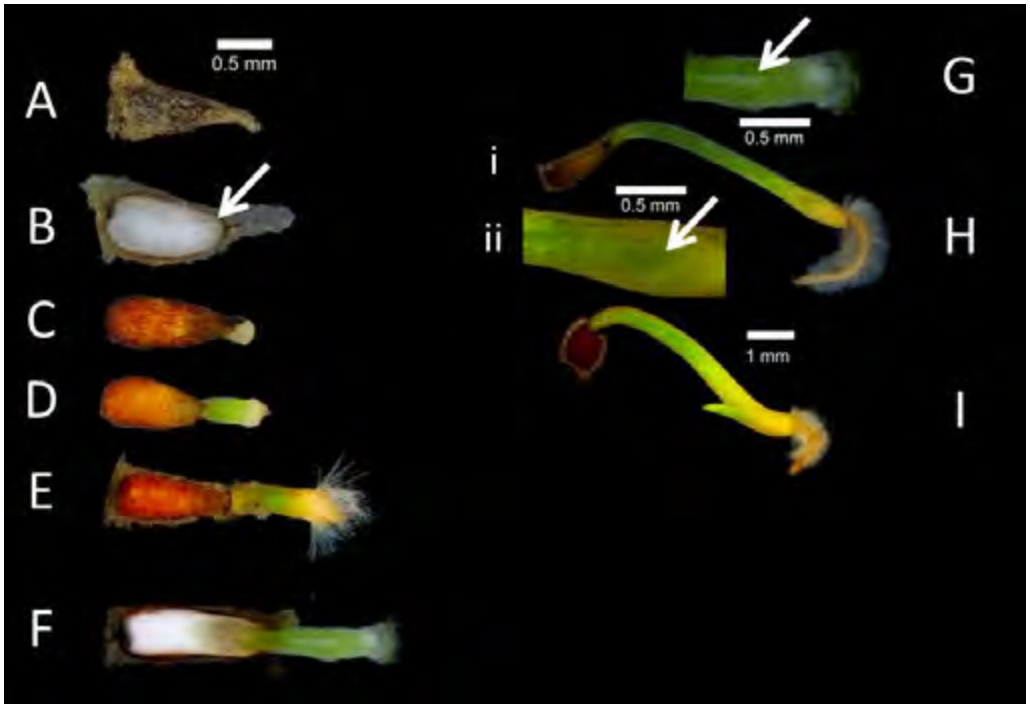


Figure 2.1. Germination and development of *X. viscosa* seeds. A) the dry seed, surrounded by husk, B) imbibed seed divided in half, showing the endosperm and embryo, C-E) the initial stages of germination (radicle protrusion, cotyledon lengthening and greening, root hair formation), F) section of a developing seedling embedded in endosperm, G) protoxylem connects the endosperm to the meristem (white arrow), H) the cotyledon lengthens dramatically (i), and eventually splits to reveal a developing leaf (white arrow) (ii), I) a representative seedling at the oldest stage of post-germinative development, showing the first true leaf. New leaves develop from below the first leaf, and eventually the cotyledon senesces. The seeds of *X. humilis* show a near identical developmental process, although the seeds are smaller.

2.3.2 The effect of developmental stage on seedling survival after rapid dehydration

As has been previously reported by Maia et al (2011), we observed that seedlings of the desiccation sensitive *A. thaliana* lost DT rapidly once the seed coat had ruptured. Slightly more than 10% of seedlings survived 48 h of dehydration once the radicle had visibly protruded through the seed coat, with the survival rate dropping to 0% at all subsequent developmental stages (Fig. 2.2A). As in *A. thaliana*, untreated seedling survival in response to rapid dehydration was also associated with seedling size at the time of drying in both *Xerophyta* species (Figure 2.2B & 3C).

The association between seedling size and seedling survival was most apparent in the seedlings of *X. viscosa*. Seedlings at an early stage of germination (cotyledon <0.4 mm) showed a near-100% survival rate (Figure 2.2B). Survival rate decreased steadily as seedlings increased in size, dropping to only 6% in seedlings with cotyledons 1.6-2 mm in length. However, in contrast to *A. thaliana*, survival rate began to increase again in larger seedlings (cotyledons > 2 mm), rising to over 40% in the oldest tested seedlings (cotyledon > 4.4 mm, around the stage at which the primary leaf emerges). In contrast, seedlings of *X. humilis* exhibited a low survival rate soon after germination (cotyledon <0.2 mm, 30%), and survival dropped dramatically soon afterwards to less than 5% in seedlings with cotyledons 0.2-0.6 mm in length (Figure 2.2C). Survival increased again, however, in larger seedlings, rising to almost 60% mean survival in seedlings with cotyledons greater than 3 mm in length.

Interestingly, survival of *X. viscosa* seedlings greater than 2 mm in length was strongly associated with the loss of chlorophyll either wholly or partially from the cotyledon (Figure 2.2D; Fig. 2.3A). Of seedlings with cotyledons > 2 mm, a total of 30% lost chlorophyll, of which 76% survived dehydration; whereas only 5% of seedlings that did not lose chlorophyll at these stages survived. This was not observed in smaller *X. viscosa* seedlings, or in *X. humilis* seedlings of any size. In a number of cases, tissue survival was restricted to regions that had degraded chlorophyll, while surrounding tissues failed to revive (Fig. 2.3B). The primary root of surviving seedlings usually also survived the desiccation process unharmed; in cases where it did not, a secondary root(s) would erupt next to or through the dead root tissue (Fig. 2.3C). Anthocyanin accumulation was also commonly observed in seedling tissues during dehydration and/or rehydration (Fig. 2.3C).

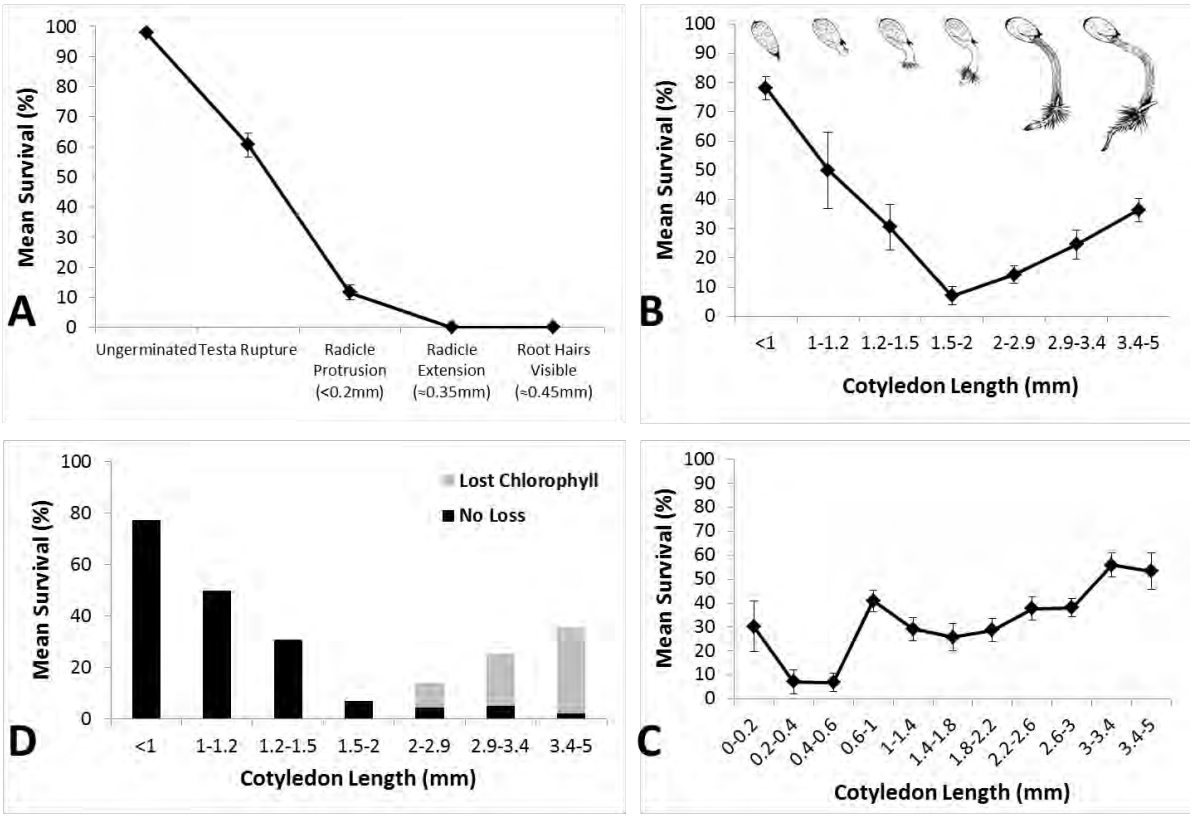


Figure 2.2. The effect of developmental stage on desiccation tolerance in untreated seedlings of *Arabidopsis* or *Xerophyta*. A) Seedlings of *A. thaliana* rapidly lose DT during germination, B) *X. viscosa* germinating seeds show a dramatic transient decline in DT, but recover at later stages, C) *X. humilis* seedlings exhibit a rapid decline in DT during early germination, but stabilise quickly thereafter, D) *X. viscosa* seedlings with cotyledons greater than 2 mm in length often lost chlorophyll from their tissues during desiccation (grey bars), a trait that correlated with seedling survival. This was not seen in any *X. humilis* seedlings of any size. Error bars on line graphs represent SEM.

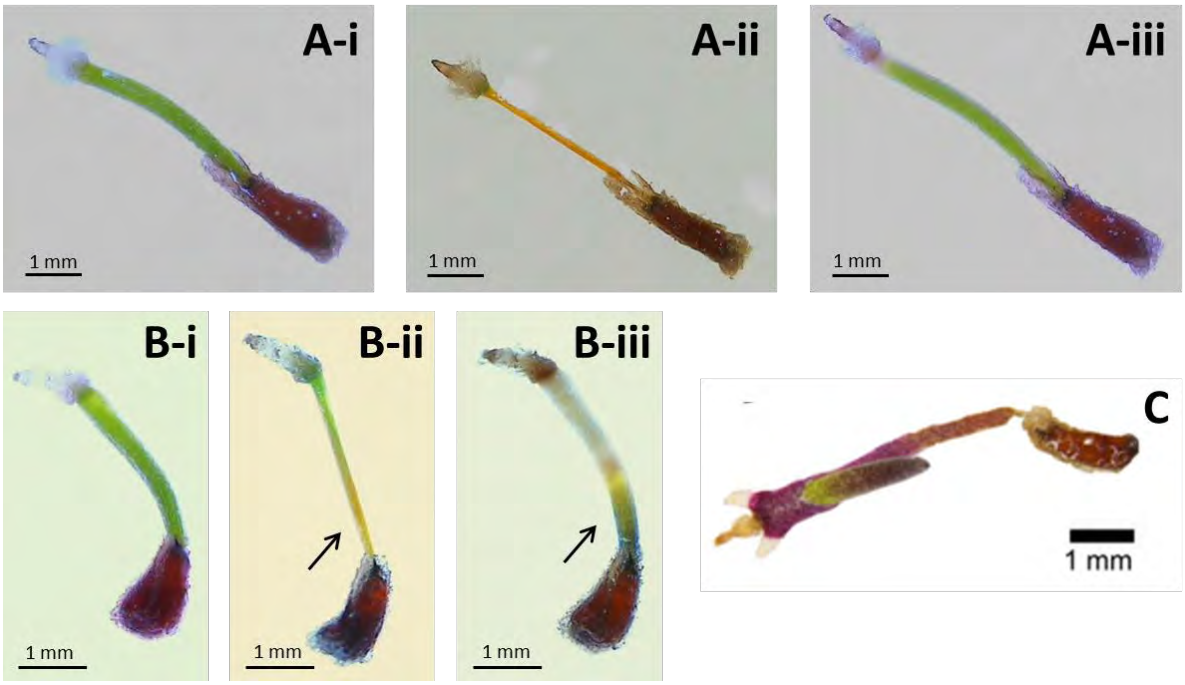


Figure 2.3. Physiological responses to desiccation in *X. viscosa* seedlings. A) A late-stage seedling when hydrated (i), after desiccation (ii), and 48 hours after rehydration (iii). Clear loss of chlorophyll can be seen in the desiccated seedling. B) A desiccated seedling showing partial chlorophyll loss from the cotyledon (ii; black arrow), and tissue survival restricted to only this region upon rehydration (iii; black arrow). In this instance, survival was restricted to one end of the cotyledon. Although the cotyledon could survive for weeks in this state, seedling development was arrested by the death of the meristematic tissue. C) A rehydrated seedling displaying strong anthocyanin accumulation in both cotyledon and first-leaf tissues. Additionally, new roots can be seen emerging alongside the former root tip that failed to survive desiccation.

2.3.3 The impact of PEG or sucrose pre-treatment on seedling survival after dehydration

Although *Xerophyta* seedlings never became 100% DS, there was a transient decline in the ability of *X. viscosa* and *X. humilis* seedlings to survive dehydration during early germination (Fig. 2.2B&C). This decline may represent a developmental stage at which DT generally cannot be re-established, such as a period of desiccation sensitivity during a switch from embryonic DT to a vegetative DT state if they are separate processes. Alternatively, young seedlings of both species may be particularly vulnerable to the effects of severe osmotic

stress as they do not have sufficient resources to mount a successful response against, or recover from, rapid dehydration. For example, depletion of the endosperm-derived nutrients upon which the germinating seedlings are dependent during initial growth, or a failure to protect the protoxylem connecting the endosperm to the meristem, might be lethal. The point of no return can be delayed in many species by priming germinating seedlings with moderate osmotic stress prior to dehydration, such as incubation in a solution of high molecular weight PEG. Sucrose has been implicated in the survival of resurrection plants in response to desiccation, and the presence of additional sucrose might also alleviate resource depletion in germinating seedlings that have not yet become self-sufficient. We thus used PEG and sucrose pre-treatment to investigate whether priming of *X. humilis, X. viscosa* or *A. thaliana* seedlings improved their ability to survive dehydration.

In agreement with previous reports, incubation in a -2.5MPa PEG-8000 solution for 48 h prior to dehydration improved the survival rate of *A. thaliana* seedlings (Fig. 2.4A), most noticeably at the stages of radicle protrusion (radicle 0.2 mm; 9% survived untreated vs. 74% PEG-treated) and radicle extension (radicle 0.35 mm; 1% vs. 20%). Nonetheless, mean survival rate dropped to less than 5% by the time the root hairs were first visible (radicle 0.45mm). Growth on 3% sucrose resulted in delayed germination in *A. thaliana* and led to the accumulation of anthocyanins, as has been reported previously for this species (Gibson, 2005; Teng et al., 2005). However, the presence of exogenous sucrose also improved seedling survival after dehydration. Whereas only 9% and 1% of untreated seedlings survived at the stages of radicle protrusion and extension stages respectively, supplementary sucrose improved survival to 59% and 42% at these stages (Fig. 2.4A). As with PEG-treated seedlings, mean survival dropped to less than 5% in seedlings with a radicle longer than 0.45

mm, suggesting that it is not possible to extend the window for re-establishment of DT beyond this point in *A. thaliana*.

PEG treatment also increased seedling survival rate after dehydration in *X. viscosa* (Figure 2.4B). Total seedling survival averaged 65%, and the transient decline in survival rate observed in untreated plants was not apparent in PEG-treated seedlings. Interestingly, the majority (83%) of seedlings of all sizes showed complete loss of chlorophyll after incubation in PEG, reminiscent of the poikilochlorophylly response seen in adult *X. viscosa* plants. Chlorophyll loss correlated strongly with seedling survival and none of the 17% of PEG-treated *X. viscosa* seedlings that failed to degrade chlorophyll survived dehydration (Fig. 2.4D). Seedling growth was halted upon initiation of PEG treatment. In contrast to its effects on delaying germination in *A. thaliana*, 3% sucrose did not have any observable effect on *X. viscosa* germination rate, nor did it induce anthocyanin accumulation in seedlings prior to dehydration. However, sucrose pre-treatment drastically improved seedling survival irrespective of cotyledon length: average survival of sucrose-treated seedlings was 91% across all developmental stages, with the lowest survival rate of 79% occurring in seedlings 2-2.9 mm in size (Figure 2.4B). As was seen in untreated *X. viscosa* seedlings, chlorophyll loss was only observed in larger seedlings, with incidence increasing with cotyledon length (Fig. 2.4E).

PEG treatment dramatically improved seedling survival rates in *X. humilis* (Figure 2.4C). The survival rate of young seedlings (cotyledon 0-1.4 mm), including the stages found to be most at risk in untreated seedlings (0.2-0.6 mm), showed a near-100% survival rate. PEG-treated seedlings did show a transient decline in survival rate in seedlings between 1.4-3 mm in length, dropping to a minimum survival rate of 60%, but this did not coincide with

previously observed sensitive period in untreated seedlings (Fig. 2.4F). Nearly all (>95%) PEG-treated *X. humilis* seedlings lost chlorophyll from their photosynthetic tissues, including many that did not survive desiccation, despite the absence of this trait in untreated seedlings. Growth on 3% sucrose caused a delay in the germination and development of *X. humilis* seedlings, as well as slightly reduced total germination. Unlike *A. thaliana,* anthocyanin accumulation was not evident and development was otherwise unaffected. However, pre-treatment with 3% sucrose did have a marked effect on seedling survival rate in response to rapid dehydration (Fig. 2.4C). The survival profile of sucrose-treated seed mirrored that of untreated seedlings, but was improved by about 35% at all early developmental stages (0-1 mm in length). Survival rose to approximately 100% at all subsequent stages. Unlike in PEG pre-treatment, chlorophyll loss was not evident in any of the sucrose-treated *X. humilis* seedlings.

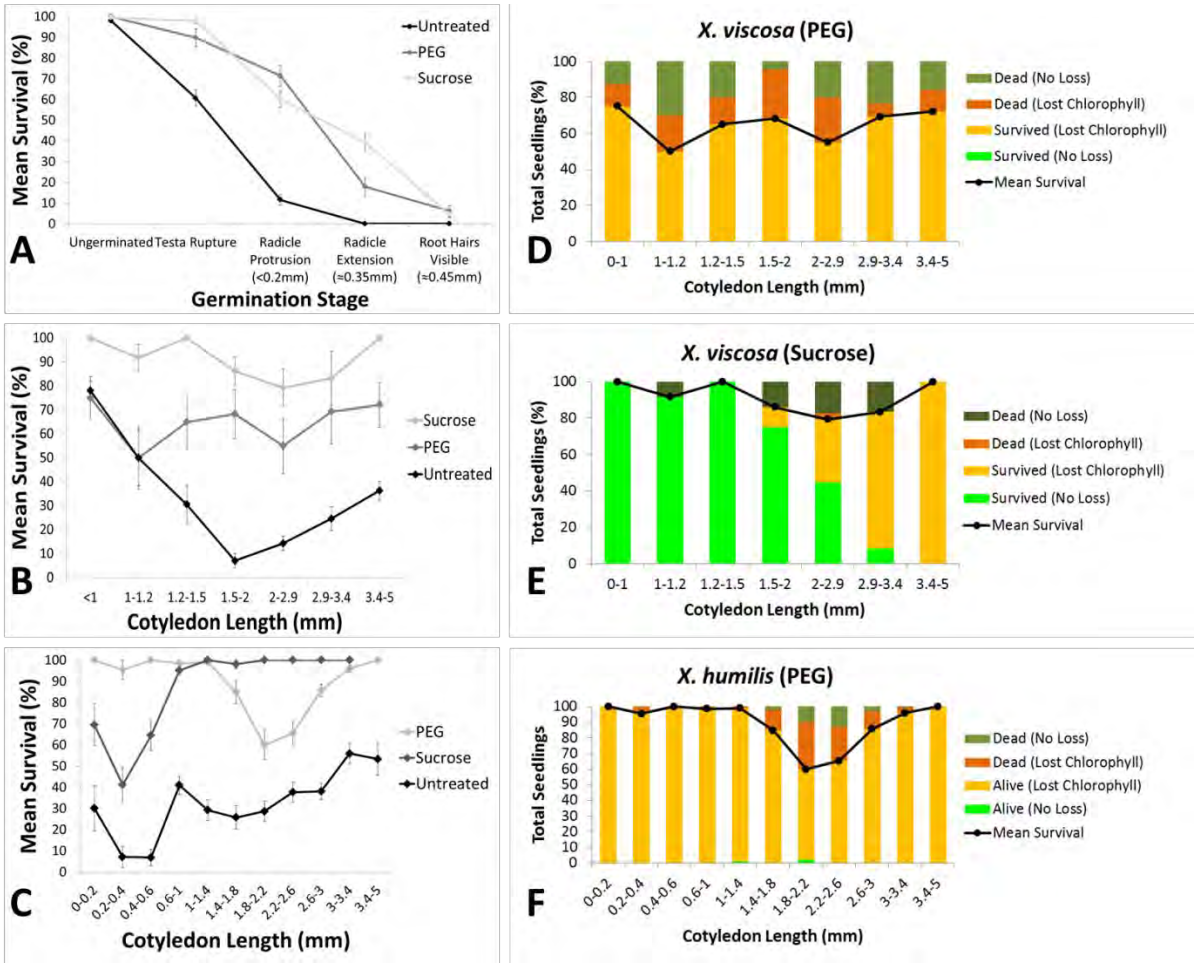


Figure 2.4. PEG and sucrose can improve seedling DT. A-C) The effect of PEG (light grey line) and sucrose (dark grey line) on seedlings of *A. thaliana* (A), *X. viscosa* (B) and *X. humilis* (C), compared to untreated seedlings of these species (black line). Error bars are SEM. Chlorophyll loss was observed in the majority of PEG treated seedlings of both species (D&F) and in large (> 1.5 mm) *X. viscosa* seedlings (E).

2.3.4 The effect of PEG priming and reduced dehydration rate on seedling survival

The exact mechanism by priming seedlings by stress treatment (e.g. incubation in PEG) induces DT in seedlings is unknown. PEG priming likely induces a transcriptional cascade that induces stress response factors in the stressed seedlings, leading to re-establishment of DT. It is unclear, however, why a presumably similar response is not observed in untreated seedlings during dehydration. It may be that PEG treatment delays lethal water loss long

enough for the DT tolerance mechanisms to be fully activated in PEG-treated seedlings, whereas desiccation in untreated seedlings occurs too rapidly for these mechanisms to be properly activated. Alternatively, PEG treatment may elicit a stronger stress response than desiccation under untreated conditions, in which case length of PEG incubation should have less of an effect on the re-establishment of DT. To test both of these possibilities, *X. humilis* seedlings of sizes 1.8-2.4 mm were treated with PEG for varying lengths of time (ranging from 0 to 24 hours) followed by rapid desiccation, and seedlings of both *X. humilis* and *X. viscosa* were left untreated and dehydrated slowly over the course of a week under standard growth conditions.

Even short periods of PEG treatment had a dramatic effect on seedling desiccation tolerance. Surprisingly, incubating seedlings for an hour or less in PEG appeared to have a negative impact on seedling survival: compared to untreated seedlings, with a 20% survival rate, PEG treated seedlings had a 5% and 0% survival rate after 30 minutes and 1 hour respectively (Fig. 2.5A). However, after 2 hours, PEG treated seedling survival rate increased 3-fold to above 60%, and rose to a maximum of over 90% in seedlings treated for 24 hours. This maximum value is higher than that observed for similarly-sized seedlings in the original 48 h *X. humilis* PEG experiments, which had only a 60% survival rate (Fig. 2.4C). As noted previously, PEG treatment alone was sufficient to induce chlorophyll loss from seedling tissues, and was clearly visible in around a third of seedlings after just 4 hours of incubation (Fig. 2.5A). Longer periods of incubation resulted in a larger percentage of seedlings that lacked chlorophyll.

Seedlings that were subjected to a reduced desiccation rate (air-dried under standard growth conditions) fared substantially better than untreated seedlings dehydrated rapidly in

a laminar flow hood at both tested stages in *X. humilis* (Fig. 2.5B). Survival rate of young (0.5 mm) and older (2 mm) seedlings was above 90% in both cases, with a 100% incidence of chlorophyll loss in both. A similar but less dramatic trend was seen in *X. viscosa* seedlings, where only older seedlings (≈3 mm and ≈4 mm) were tested (Fig. 2.5C). This indicates that many of the results observed in previous sections (failure to survive at small sizes, lack of chlorophyll loss in untreated *X. humilis* seedlings) may be related to differences in desiccation rate at different stages of development.

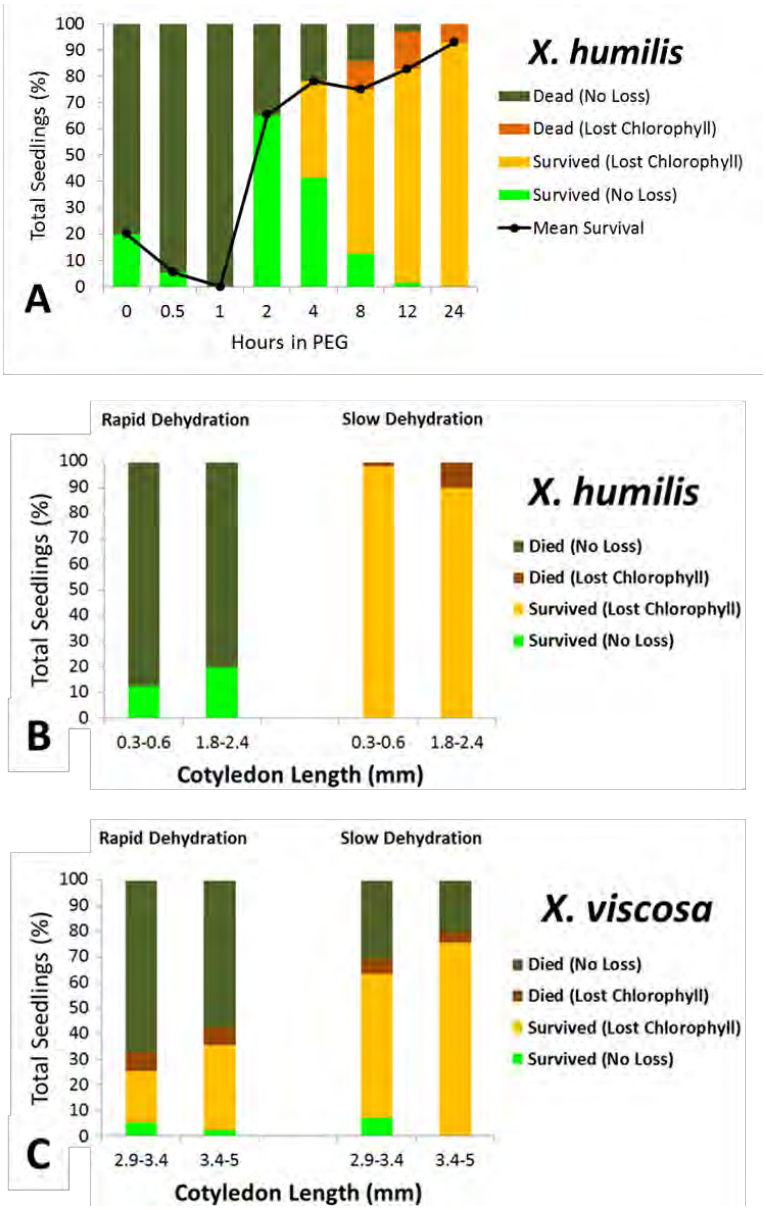


Figure 2.5. A longer period of priming improves *Xerophyta* seedling survival after desiccation. A) *X. humilis* seedlings of cotyledon length 1.8-2.4 mm were incubated in -2.5 MPa PEG-8000 for between 0 and 24 hours. Seedling survival and incidence of chlorophyll loss was tracked for each seedling. Increased survival was observed after only 2 hours, and chlorophyll loss was observed after 4 hours. B&C) Seedlings of *X. humilis* (either 0.3-0.6 mm or 1.8-2.4 mm) or *X. viscosa* (2.9-3.4 mm or 3.4-5 mm) were dried in a laminar flow hood (rapid dehydration) or for one week under standard growth conditions (slow dehydration). Survival rate and chlorophyll loss were measured for each seedling under each condition.

2.4 Discussion

The seedlings of many orthodox seed-bearing angiosperms are capable of inducing DT during the early stages of germination, even in otherwise DS species (Buitink et al., 2003; Maia et al., 2011). Current evidence suggests that this process involves resetting the seedling to a quiescent state that resembles that of mature, dry orthodox seeds – suggesting that, under the correct conditions, the embryo can halt development and revert to its former quiescent state while it still has the resources to continue germination should conditions improve (Maia et al., 2011). The point of no return that is seen in the orthodox seeds of many angiosperm species is associated with the epigenetic silencing of most seed- and maturation-associated gene networks (Bouyer et al., 2011). This could be an evolutionary trade-off between out-surviving competing seedlings during a particularly harsh, extended drought when young, and being able to outgrow competitors during periods of brief osmotic stress once the energetic costs involved in successfully desiccating completely become restrictive.

The acquisition of vegetative DT in angiosperms may have evolved through the co-option and expression of the networks regulating angiosperm seed maturation genes in adult tissues. A possible means by which this could have occurred would be through extending the temporary window during which angiosperm seedlings are able to re-establish DT indefinitely. If this is the case, we would predict that the seedlings of resurrection plants would not display a point of no return, but would instead be desiccation tolerant throughout germination. Our aim was to determine whether this was true in *X. humilis* and *X. viscosa*, two monocotyledonous, poikilochlorophyllous resurrection plants found in Southern Africa. We investigated the DT of the developing seedlings of both of these

species, as well as the desiccation sensitive *A. thaliana*, when untreated or either incubated in PEG or supplemented with exogenous sucrose.

2.4.1 Xerophyta seedlings show a transient decline in DT, which can be rescued by PEG or sucrose

In agreement with previous reports, we found that the seedlings of *A. thaliana* rapidly lost DT once the seed coat had ruptured, but chance of seedling survival could be increased by pre-treatment in PEG or, as we found, sucrose. In either case, neither treatment could substantially extend the window beyond the point at which radicle length was greater than 0.45 mm – i.e., this stage probably represents the "point of no return" of this species (or ecotype), after which the genes required for DT re-establishment are permanently silenced.

On the other hand, seedlings of the resurrection plants *X. humilis* and *X. viscosa* showed only a transient decline in survival rate in response to desiccation. The method of dehydration used in this study (desiccation under constant air flow in a laminar flow hood) is likely also far more extreme than that which would occur under natural conditions, and seedlings in the wild would probably display a much higher survival rate in response to reduced water availability. The period at which *Xerophyta* seedlings were most vulnerable to rapid desiccation occurred at different stages of development in the two species. *X. humilis* seedlings were not fully DT immediately after germination, with only a 30% survival rate, and survival dropped further in seedlings 0-1 mm in length. Seedlings older than this maintained a survival rate between 30-55%, improving slightly with increasing size. *X. viscosa* seedlings showed a 75% survival rate immediately after germination, which declined steadily down to 5% in seedlings between 1.5-2 mm in length, before rising sharply in older

seedlings. It is unclear why the lowest survival rate for seedlings of the two species occurs at different cotyledon lengths, and whether this corresponds to development or physical difference between them. For example, the seeds of *X. viscosa* are larger than those of *X. humilis* and have a greater complement of endosperm, which may indicate that *X. viscosa* seedlings have more available energy reserves during both desiccation and rehydration at early stages of development. Alternatively, a larger seed relative to the seedling is likely to correspond to a reduced comparative desiccation rate for the seed/seedling as a whole (as the larger seed would retain more liquid), which may also explain why the seedlings of *X. humilis* are less tolerant to desiccation than *X. viscosa* even immediately after radicle protrusion.

PEG and sucrose pre-treatment drastically improved DT in seedlings of both species (Fig. 2.4). PEG treatment had the most dramatic effect on *X. humilis* seedlings, where it raised seedling survival at early stages to approximately 100% - an increase of over 70% points. Oddly, although it eliminated the transient decline of DT observed in young seedlings, seedlings between 1.4 – 3 mm in length were less DT when treated for 48 hours in PEG (60% minimum survival rate) compared to other stages (Fig. 2.4C&F). However, seedlings this size had a 90% survival rate when treated with PEG for only 24 hours in a separate experiment, suggesting either a batch effect in one of these experiments or a negative effect of extended PEG incubation in these seedlings (Fig. 2.5A). In *X. viscosa*, PEG treatment completely eliminated the transient DT decline, and all seedlings had an approximately 60% survival rate irrespective of size. This is much less than the observed increase in *X. humilis*, but still significantly higher than untreated seedlings. PEG treatment was effective within

just two hours of treatment in *X. humilis* seedlings, but appeared to have a negative effect on seedling survival when seedlings were treated for 30 minutes to an hour.

Sucrose pre-treatment had a similar effect on seedling survival to that of PEG treatment, and in many cases was even more effective. Sucrose-treated *X. humilis* seedlings had improved DT at each tested developmental stage. Seedlings larger than 1 mm were completely DT (100% survival), whereas those below 1 mm in size showed a comparatively reduced rate of DT, surviving only 40-70% of the time, but still substantially higher than untreated seedlings. Interestingly, the period of reduced DT in PEG treated *X. humilis* seedlings (1.4 – 3 mm) was not apparent in sucrose-treated seedlings suggesting that, if the secondary transient decline in DT in PEG-treated seedlings is a real phenomenon, sucrose and PEG could improve DT in slightly different ways. In *X. viscosa*, meanwhile, sucrose pre-treatment also improved DT across the board; minimum survival rate was just 80% (in seedlings 2 – 2.9mm in length) and substantially higher than this at all other stages.

Although the mechanism of action is unclear, PEG treatment has previously been shown to transiently induce seedling DT in various angiosperm species. However, the effects of treating with exogenous sucrose on the DT of developing seedlings have so far not been reported. Sucrose has been shown to induce DT and freezing tolerance in axillary buds and shoot tips of various species for the purposes of cryopreservation, possibly by activation of similar ABA signalling pathways as those involved in seedling DT (Suzuki et al., 2006, 2005, 1998). Unfortunately, the exact role exogenous sucrose plays in this process cannot be determined from the results in this study. It seems unlikely that the relatively minor osmotic effects of sucrose alone (-0.22 MPa) (Sarkar, 2009), would be sufficient to prime the DT response, as osmotically stressed PEG-treated seedlings (-2.5 MPa) had a differing survival

profile (*X. humilis*) or did not recover to a remotely similar degree (*X. viscosa*). Alternatively, exogenous sucrose may have acted as an energy source that was independent of seedling stage, reducing the total number of seedlings that failed to survive purely due to a lack of sufficient resources to mount a successful desiccation response. However, the ability of PEG to induce DT in >70% of *X. humilis* seedlings within 8 hours (Fig. 2.5A) would seem to refute a general lack of available seedling resources in the majority seedlings, although it may not be in a readily-metabolisable form prior to PEG priming. Sucrose may also play a protective role during desiccation. Functionally, sugars – mainly sucrose, but also other classes of carbohydrates – have been predicted to act as osmoprotectants that, together with LEA proteins, form the cellular glass that protects tissues from desiccation (Buitink and Leprince, 2008). High background levels of intra- or extracellular sucrose could have reduced the time or carbohydrate metabolism investments required to induce glass formation during the desiccation process, and subsequently led to better protection of the seedling tissues during dehydration. A final explanation is that exogenous sucrose might prime an ABA signalling response in *Xerophyta* seedlings. Exogenous sucrose can activate ABA signalling pathways in the ripening fruits of grape and strawberry (Gambetta et al., 2010; Haifeng Jia et al., 2013), and has also been shown to have a synergistic effect with ABA in inducing DT during early seed maturation and during somatic embryo formation (Meurs et al., 1992; Saranga et al., 1990; Senaranta et al., 1989). If the window of seedling DT does involve an ABA-mediated reactivation of maturation genes, high background sucrose levels may have had a similar synergistic signalling role with ABA and its downstream processes, effectively accelerating the desiccation response.

2.4.2 Xerophyta seedling survival is influenced by desiccation rate; Poikilochlorophylly can be induced at all seed stages, but is likely uncoupled from DT protection mechanisms

Nearly all surviving PEG treated *Xerophyta* seedlings degraded chlorophyll irrespective of size or stage, suggesting that chlorophyll loss is technically inducible at all stages of seedling development in both these species. A similar effect was not observed in *A. thaliana*, and preliminary experiments show that this does not occur in the homochlorophyllous sister species, *X. elegans*. Thus, it is unlikely that this trait is a secondary effect of PEG treatment, but is rather related to the adult trait of poikilochlorophylly. Our data suggests that both the incidence of poikilochlorophylly and survival of a seedling are dependent on the rate of seedling desiccation. In both untreated and sucrose-treated *X. viscosa* seedlings, chlorophyll loss was only observed in larger seedlings (> 2 mm) whereas it was completely absent in *X. humilis* seedlings under the same treatments. However, a slow rate of dehydration under standard growth conditions over one week was sufficient to not only improve seedling survival compared to untreated seedlings in both species, but also to induce chlorophyll loss in nearly all surviving seedlings. In *X. humilis* seedlings treated with PEG for varying lengths of time from 0 to 24 hours, chlorophyll loss was first observed only 4 hours after incubation (Fig. 2.5A). Thus, it can be assumed that if the laminar-flow-dried (untreated and sucrose-treated) seedlings were desiccating at a rate faster than this, we would not expect to see an absence of chlorophyll from their tissues. It is likely then that the occurrence of chlorophyll degradation only in larger *X. viscosa* seedlings is indicative of a slower desiccation rate than seedlings at other, smaller stages. Similarly, the absence of chlorophyll degradation entirely from untreated or sucrose-treated *X. humilis* suggests that seedlings of this species desiccate more rapidly than *X. viscosa* seedlings of the same length.

Although the occurrence of chlorophyll loss in a seedling was strongly correlated with survival, multiple lines of evidence suggest that it is uncoupled from the actual cellular desiccation protection mechanisms. Whereas incubation in PEG for just 2 hours was sufficient to improve seedling survival rate from 20% to 65%, chlorophyll degradation was only visible after 4 hours (Fig. 2.5A). Additionally, although nearly all surviving PEG-treated and slow-dried seedlings lost chlorophyll, a number of seedlings across multiple experiments failed to revive despite losing chlorophyll (Fig. 2.4D&F). Thus, although chlorophyll loss can be a good predictor of the probability of seedling survival during desiccation in these species, it is neither sufficient nor required to ensure survival.

In summary, our data suggest that *Xerophyta* seedlings do not inherently lose DT during germination, even under rapid drying conditions, as other angiosperm species do; i.e. there is no developmentally-regulated point of no return in these species. One explanation for these results is that *X. humilis* and *X. viscosa* never close the window of ABA responsiveness that allows germinating angiosperm seedlings to revert to a quiescent, mature-dry state, but has instead evolved the ability to activate this response throughout its life, even in adult tissues. A particularly striking result is the potential for poikilochlorophylly at all seedling stages, further suggesting that there is a direct relationship between the seedling and VDT response in *Xerophyta.* Current evidence suggests that the desiccation response of resurrection plants is related to the seed maturation pathway, particularly the genes involved in late maturation and maturation drying. Though we show strong commonalities between the seedling and adult DT response in *Xerophyta* seedlings, further genetic and functional assays will need to be performed in order to validate the relationship between these two responses and embryonic DT found in seeds. In Chapter 3 the desiccation

transcriptome of adult *X. humilis* leaves is examined using RNA-Seq, in order to further investigate the possible expression of seed maturation genes in vegetative tissues and the relationship between VDT and the angiosperm window of DT re-establishment.

2.5 Supplementary Data

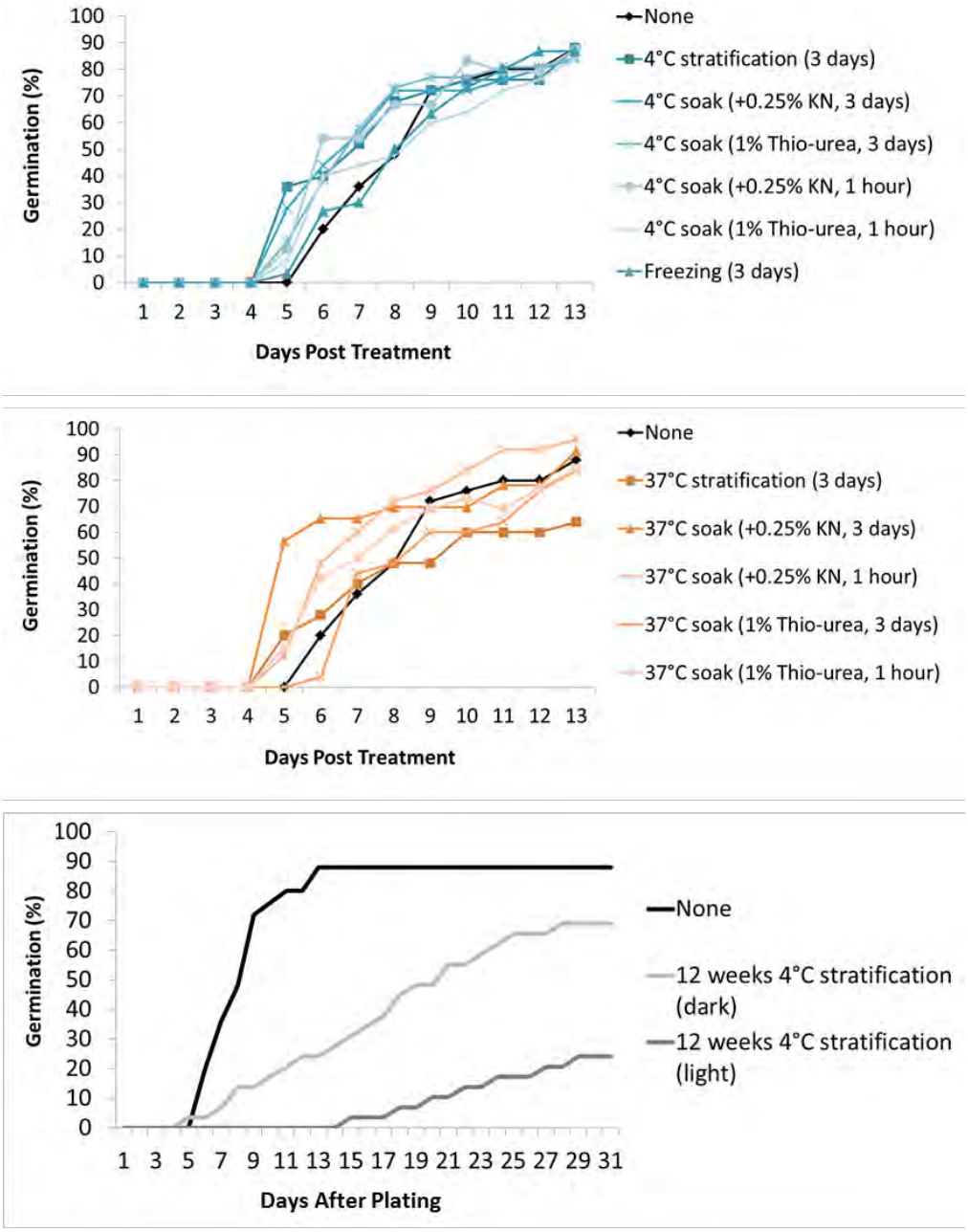


Figure S2.1. Response of *X. viscosa* seedlings to several dormancy-breaking treatments. Seeds of *X. viscosa* germinated asynchronously over the course of a week or more, so various treatments were tested to see if they could improve germination rate. None of the tested treatments could substantially increase rate or synchronicity of germination on batches of 20-30 seeds.

Figure 2.4A (*A. thaliana*)					
Developmental Stage	**Dry Seed**	**Testa Rupture**	**Radicle Protrusion**	**Radicle Extension**	**Root Hairs**
Seedling Count (PEG)	80	49	77	78	81
Seedling Count (Sucrose)	206	143	151	155	152
Seedling Count (Untreated)	191	160	164	151	135

Figure 2.4B (*X. viscosa*)							
Cotyledon Length (mm)	**<1**	**1-1.2**	**1.2-1.5**	**1.5-2**	**2-2.9**	**2.9-3.4**	**3.4-5**
Seedling Count (Untreated)	114	16	39	72	135	79	157
Seedling Count (PEG)	24	20	20	22	20	13	25
Seedling Count (Sucrose)	40	24	26	36	29	12	10

Figure 2.4B (*X. humilis*)											
Cotyledon Length (mm)	**0-0.2**	**0.2-0.4**	**0.4-0.6**	**0.6-1**	**1-1.4**	**1.4-1.8**	**1.8-2.2**	**2.2-2.6**	**2.6-3**	**3-3.4**	**3.4-5**
Seedling Count (Untreated)	20	28	44	122	89	62	84	101	155	102	45
Seedling Count (PEG)	10	22	51	65	93	46	50	78	147	147	53
Seedling Count (Sucrose)	23	34	45	123	69	51	71	52	39	10	-

Figure 2.5A (*X. humilis*)								
Hours in PEG	**0.5**	**1**	**2**	**4**	**8**	**12**	**24**	**48**
Seedling Count	85	37	39	52	41	56	64	84

Figure 2.5B (*X. humilis*)				
Drying Rate	**Rapid**		**Slow**	
Cotyledon Length (mm)	**0.3-0.6**	**1.8-2.4**	**0.3-0.6**	**1.8-2.4**
Seedling Count	89	85	69	70

Figure 2.5C (*X. viscosa*)				
Drying Rate	**Rapid**		**Slow**	
Cotyledon Length (mm)	**2.9-3.4**	**3.4-5**	**2.9-3.4**	**3.4-5**
Seedling Count	79	157	87	37

Figure S2.2. Seedling bin sizes for each analysis. Numbers given here refer only to viable, successfully germinated seedlings used in the analyses, not total seeds plated per experiment. Seedlings from multiple replicate plates were pooled, as discussed in Methods.

Chapter 3: De novo transcriptome assembly and analysis of desiccating leaves of *X. humilis*

3.1 Introduction

Resurrection plants are a diverse group of angiosperm plant species that are desiccation tolerant (DT), a rare trait in the vegetative tissues of complex organisms. Although much has been achieved regarding the origins and method of action of this trait in some resurrection plant species over the past few decades, progress has been limited by the lack of genomic and other "omics" data for these non-model plants (Dinakar and Bartels, 2013). Only recently has this begun to change, as the increased availability and decreased cost of next generation sequencing (NGS) technologies continues to revolutionise the biological sciences. The transcriptomes of two resurrection plant species (*Craterostigma plantagineum* and *Haberlea rhodopensis*) have been sequenced in the past few years, while the first genome of a resurrection plant (*Boea hygrometrica*) was released just last year (Gechev et al., 2013; Rodriguez et al., 2010; Xiao et al., 2015).

NGS is the catch-all term used to describe a range of high-throughput sequencing technologies that are considered the "second generation" to those that involve Sanger sequencing (reviewed by e.g. von Bubnoff 2008; Pettersson et al. 2009; Voelkerding et al. 2009; Metzker 2010; Su et al. 2011; Schlebusch & Illing 2012). Unlike Sanger sequencing, where increased throughput is limited by the requirement to run multiple gel electrophoresis or capillary-based sequencers in parallel, NGS techniques are massively parallel in that they can perform millions of sequencing reactions at once. The approaches of each NGS sequencing platform are slightly different, but all involve the spatial separation of

each target molecule to be sequenced (for example, by suspending the targets in water-oil emulsion droplets or attaching them to the surface of a bead or flow-cell), which allows for massively parallel sequencing to occur in the same reaction. The major NGS platforms also each use slightly different sequencing chemistries, such as pyrosequencing, reversible-terminator or sequencing-by-ligation. All of the described chemistries share the same limitation in that they produce only short runs of sequence information (termed a "read"), generally 30-400 bp in length depending on platform, and are more error-prone compared to Sanger sequencing. As a consequence, increased depth (the number of times a target locus is sequenced) is required to reliably determine the nucleotide sequence(s) of a DNA sample. Because NGS sequencing reads are generally far shorter than the DNA in the sequenced sample from which they are derived (e.g. a genome), they need to be assembled or mapped to a known sequence to be informative.

NGS technologies and associated high throughput solutions have allowed researchers to analyse the genomes and transcriptomes of a variety of organisms (or tissues) at an unprecedented scale, speed and low cost, and have understandably replaced many of the previously used techniques, such as microarrays. This has made it an ideal tool to study non-model organisms, which heretofore have had to rely on limited, low-throughput analyses. In this chapter, I use NGS RNA sequencing to assemble and analyse the transcriptome of the monocot resurrection plant, *X. humilis*.

3.1.1 RNA Sequencing

RNA sequencing relates to the analysis and quantification of RNA transcripts within an organism. This not only includes protein-coding gene transcripts (mRNA), but also the

various types of functional non-coding RNAs, such as ribosomal RNA (rRNA), micro RNAs (miRNA), long non-coding RNA (lncRNA) etc., and non-functional RNAs, such as those occurring due to transcriptional read-through (Kapranov et al., 2007a, 2007b). As most sequencing techniques do not yet have the resolution to interrogate the contents of a single cell, the term transcriptome most commonly refers to the total RNA transcripts present within a tissue or organism at a specific developmental stage or physiological status.

The most widely used methods to quantify the transcriptome are those that rely on hybridisation (such as microarray or Affymetrix) or direct sequencing (Sanger sequencing, RNA-Seq) (Mutz et al., 2013; Wang et al., 2009; Zhao et al., 2014). Hybridisation techniques can effectively quantify the transcriptome, or more often a custom or pre-chosen subset thereof, and are relatively inexpensive. However, they give no sequence information of the detected transcripts, and cannot be used to identify or classify novel transcripts or splice variants without prior knowledge about the sequence. The earliest sequencing-based techniques relied on Sanger sequencing of cDNA or EST libraries. While effective and comparatively error-free, Sanger sequencing is low throughput and expensive compared to other technologies. Updated techniques, such as serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and massively parallel signature sequencing (MPSS), were developed to bypass some of the limitations. Though successful, they were still expensive and could only sequence small portions of individual transcripts, limiting their use in the identification of transcript variants (Wang et al., 2009).

With the advent of high-throughput NGS and deep-sequencing technologies, a novel RNA sequencing method, RNA-Seq, was developed. The first step of an RNA-Seq sequencing protocol involves the conversion of an RNA sample (total RNA, or one enriched for poly-A

mRNA or non-rRNA transcripts) into cDNA. Usually, the RNA sample has been randomly fragmented and size-selected so that all sequences are the same length prior to cDNA synthesis. As with conventional NGS methods sequencing adaptors (DNA tags containing a known, unique sequence) are ligated to either one or both ends of the resultant cDNA sequences. Low concentration RNA/cDNA samples may require amplification prior to sequencing, which can effect downstream transcript quantification or error rate. Each tagged cDNA molecule finally undergoes high-throughput sequencing from either one end (single-end sequencing) or both ends (paired-end sequencing), producing a sequencing read (or reads), which can vary in size between 30-400 bp based on the sequencing technology used and the original fragment size. Unlike conventional Sanger sequencing, in which a portion of a single transcript is sequenced to high accuracy, an RNA-Seq experiment produces a staggering number of sequencing reads per run (millions to billions). The comparatively low accuracy of high-throughput sequencing methods is compensated for through sheer volume, i.e. for each read containing an error at a particular nucleotide position, there will be dozens to hundreds of reads that contain the correct base in that position.

The power of RNA-Seq lies in a number of factors. For one, as opposed to many Sanger sequencing methods, RNA-Seq is quantitative: the proportion of sequencing reads derived from a given transcript in an RNA-Seq experiment correlates directly with the abundance of that transcript within the original RNA sample (Wang et al., 2009). This makes RNA-Seq exceptionally useful for measuring relative gene expression and, by including RNA spike-in controls of known concentration, can even rival qPCR in absolute expression accuracy (Mortazavi et al., 2008). The high depth of coverage also allows for a greater dynamic range of measurable expression levels compared to hybridisation techniques. Transcript sequence

mapping by RNA-Seq is simplified in organisms where an annotated reference genome is available. By aligning the short RNA-Seq sequence reads to the genome, it is possible to accurately determine the transcript sequence and the sequences of any possible transcript variants of a particular gene or genomic locus of interest. A reference genome also makes it comparatively easy (though not trivial) to identify transcripts that correspond to unannotated genes or those derived from gene fusions (Mutz et al., 2013). The issue is more complicated when there is no genome (or a low-quality genome) to which the reads can be mapped, as is currently the case with most non-model organisms (Benjamin et al., 2014). It is then necessary to assemble the millions of contiguous sequencing reads directly, a task that is both fundamentally and computationally challenging (Miller et al., 2010).

3.1.2 De novo transcriptome assembly

Without a reference genome, it is necessary to assembly transcript sequences *de novo*, using only the information contained within just the sequencing reads (Grabherr et al., 2011). In its simplest implementation, *de novo* assembly of high-throughput sequencing reads requires loading the individual reads into memory, finding significant overlaps between two or more reads, merging them to form larger contiguous sequences ("contigs"), and repeating this in an iterative fashion until only contigs remain. In an ideal situation, these contigs should represent the original sequences (cDNA, genomic DNA) in the original sample from which the reads were derived. In actuality, this process is complicated by a number of factors. For one, NGS sequencing technologies produce a lot of sequence data – far more than is able to be processed within the memory of most current desktop computers or many computing clusters – making it necessary to manipulate or compress the sequence information and data structures before it is usable. The algorithm described above is also impractical time-wise, as

determining the network of shared edges between millions of sequences in a pairwise fashion is spectacularly time consuming and inefficient. Furthermore, *de novo* assembly is complicated by the realities of naturally occurring DNA sequences: genomes often contain stretches of low complexity sequence (mono- or –dinucleotide repeats, for example), and much of the genome is not unique due to the presence of repeated sequences (duplicated genes, transposons, LINEs, SINEs, etc.). This introduces ambiguity into the assembly process, as a single read could be contiguous with reads from unrelated genomic loci. Lastly, no sequencing technology is error-free, and the existence of errors or misclassified bases within a read can further complicate the assembly process.

Many of the mathematical and computational techniques that were created to solve or sidestep these problems were originally pioneered for the task of *de novo* genome assembly (Luo et al., 2012; Simpson et al., 2009; Zerbino and Birney, 2008). Most of these rely on the computer science abstraction known as "graphs", where the relationship between individual sequences (e.g. a read) is modelled as a network of nodes (unique sequences) and directed edges that connect them (Compeau et al., 2011; Li et al., 2012; Miller et al., 2010; Schatz et al., 2010; Zerbino and Birney, 2008). By tracing the path between a start and end node, it is possible to recreate the sequence from which those nodes were derived. In an ideal, but unrealistic, dataset all nodes would be visited only once while tracing this path, i.e. there is no ambiguity between nodes or edges, and the path does not intersect with itself. In real sequence this is unlikely, but by tracing all possible paths in a graph it is possible to determine all of the contig sequences from which the graph was derived. The most commonly used form of graph-based sequence assembly relies on a model known as a De Bruijn graph. In De Bruijn graph-based sequence assembly, the nodes do not represent

entire reads but rather "k-mers" – all possible fixed-length subsequences of length *k* that are present within the read dataset. Directed edges connect each node to all other nodes that share a *k*-1 sequence overlap. The optimal size of *k* can vary across datasets, but must be large enough that k-mer sequences can be expected to be unique within the source sequence (e.g. a genome), and is usually non-even to prevent the occurrence of palindromic k-mers. The division of reads into fixed-length k-mers, though perhaps counter-intuitive, provides many benefits for assembly. It reduces the total size of dataset by merging all repeated k-mer sequences into a single node, thus requiring less memory. It is also exceptionally useful for error-correction of genomic sequencing data, based on the fact that a sequencing error is far more likely to produce a unique k-mer than a k-mer that already exists in the dataset. Thus, very low abundance k-mers can either be corrected (converted to the most likely k-mer from which they are derived), or removed entirely, reducing the effect of errors on assembly.

The earliest attempts at RNA-Seq transcriptome assembly simply used the *de novo* genome assembly software available at the time (Birol et al., 2009; Collins et al., 2008; Jackson et al., 2009) . This had some success, but it quickly became evident that many of the assumptions that guided *de novo* genome assembly did not hold true for RNA-Seq (Schulz et al., 2012). In genome assembly, it is assumed that the k-mer coverage is more-or-less uniform (i.e. each genomic locus has been sequenced the same number of times). This simplifies error correction (as mentioned above) and makes repeat regions easier to detect. Additionally, the expected outputs of a genome assembly, namely one or more long, non-overlapping contigs or scaffolds, are vastly different to those expected of a transcriptome assembly. In contrast, transcriptomes do not have remotely uniform coverage, as reads from highly-expressed

transcripts (e.g. rRNA in total RNA) will greatly outnumber those from lowly-expressed transcripts. The presence of multiple transcript isoforms means that there will be many partially overlapping resultant contigs, and k-mer coverage can vary across even the same transcript, complicating the graph-tracing process of De Bruijn assemblers. Thus, it was necessary to implement new strategies and software for RNA-Seq assembly.

The most commonly used *de novo* transcriptome assemblers include TransAbyss, Trinity and Oases (Grabherr et al., 2011; Robertson et al., 2010; Schulz et al., 2012). Both Oases and TransAbyss are built upon software previously designed for genome assembly (Velvet and Abyss, respectively), whereas Trinity was the first assembler designed from the ground up specifically for RNA-Seq. Other assemblers also exist that utilise a variety of novel or improved algorithms or strategies, such SOAPdenovo-Trans or the recently released Bridger assembler (Chang et al., 2015; Xie et al., 2014). Previous studies that examined the performance of the top assemblers on both actual and simulated RNA-Seq data determined that there were considerable differences between the resultant assemblies derived from each, even when assembling the same sequencing data (Schulz et al., 2012). For example, TransAbyss was better at assembling lowly-expressed transcripts, whereas Trinity was best at assembling highly-expressed transcripts. Oases excelled at assembling transcriptomes with a wide range of transcript expression levels and was more sensitive than the other two assemblers, but was also less accurate overall. Thus, the choice of assembler can have a major impact on the results of a *de novo* RNA-Seq assembly. The optimal assembler or assembly strategy for a sequencing experiment can depend on the available computational resources for the study, and more importantly may vary from species to species or even based on the sequencing platform used. Recently, it is becoming increasingly common to

combine the results from multiple assemblies to produce a better representation of the "true" transcriptome (Rismani-Yazdi et al., 2012; Zhao et al., 2011). This can mean combining the transcript sequences produced by different assemblers, or the results from the same assembler using different assembly parameters, typically k-mer size. Unlike genome assembly, where there is generally an "optimal" k-mer size with which to perform sequence assembly that is dependent on the total coverage and architecture of the specific genome, it has been shown that k-mer size can severely impact the ability of an RNA-Seq assembler to correctly assemble given transcripts within a transcriptome. Specifically, smaller k-mer sizes are much better at assembling lowly-expressed transcripts, whereas larger k-mer sizes are better at assembling highly-expressed transcripts (Schulz et al., 2012). The rate of transcript misassembly is also correlated to k-mer size, with larger k-mer sizes producing less erroneous transcripts at the cost of reduced sensitivity. Effective *de novo* transcriptome assembly thus requires a careful balance of sensitivity and accuracy that is unlikely to be achieved by a single assembler.

A few methods have been used to perform post-assembly processing of merged transcriptomic data, with the goal of decreasing contig redundancy and maximising "true" transcript diversity. Oases and TransAbyss come with built-in tools to merge their output assemblies, and can even be run as a single assembly/post-processing pipeline (Robertson et al., 2010; Schulz et al., 2012). These tools simply use sequence homology to remove redundant transcripts that share high identity (user adjustable) with other transcripts in the assembly, retaining only the longest contig. Similar strategies have been used by other researchers, for example using the assembler-independent tools TGI and CAP3. TGI clusters sequences based on shared similarity; subsequently, CAP3 is used to assemble the clustered

contigs to produce a single “consensus” transcript (Rismani-Yazdi et al., 2012; Zhao et al., 2011). An alternative method instead clusters sequences based on their coding potential, for example Evidential Gene or IFRAT (Gilbert, 2013; Mbandi et al., 2015). In the Evidential Gene pipeline, the best predicted coding sequence (CDS) is first determined for each contig. Contigs are then removed based on percentage shared CDS (user adjustable), with only the longest remaining. Subsequently, contigs are clustered based on total sequence identity, as above. From the remaining sequences, alignment data and CDS are used to determine whether a sequence can be considered "primary" (putative full length) or "secondary" (putative smaller isoform/splice variant). Both of these merging strategies have their benefits and pitfalls. Techniques that merge contigs based on nucleotide similarity and length alone will often enrich for contig misassemblies or artefacts (chimeric transcripts of paralogous genes, or artificially long contigs derived from genomic contamination or intronic read-through) while not removing as many redundant contigs as coding-potential based techniques. Clustering contigs based on shared CDS leads to assemblies with worse mapping and length statistics (due to the above), but decreased redundancy and lower levels of contig misassembly, thus improving homology-based metrics (Nakasugi et al., 2014). However, by retaining only contigs with coding potential, this technique by definition will remove all non-coding transcripts, limiting their use if analysis of such RNAs is desired.

The aim of the work presented in this chapter was to assemble a reference transcriptome for *X. humilis* and use it to analyse gene expression during dehydration. While transcriptomes for two other resurrection plants, *C. plantagineum* and *H. rhodopensis*, have been published to date, *X. humilis* would represent the first available monocot resurrection plant transcriptome, and so one more closely related to the majority of major crop species

(including wheat, rice, maize and barley), as well as the first from a poikilochlorophyllous resurrection plant. Coupled with the transcriptomic data would be RNA-Seq gene expression data for a desiccation trial of adult *X. humilis* leaves at five relative water contents: 100%, 80%, 60%, 40% and 5%. This information would give an overview of the changing RNA landscape of a desiccating resurrection plant, facilitating the identification of co-regulated gene clusters. From there, enrichment for specific cellular processes in a desiccating plant through the various stages of desiccation could be identified. Particularly, it should give valuable insight into the regulatory mechanisms that control VDT and their similarity to those that control seed maturation in angiosperms, an important step in confirming the hypothesis that resurrection plant VDT is a seed-derived trait. Lastly, because the RNA-Seq data consists of stranded Illumina paired-end reads, it can form the basis for a later investigation into the role of long non-coding RNA (lncRNA) in vegetative desiccation tolerance.

3.2 Materials and methods

3.2.1 Plant material

X. humilis plants were collected from Borakalalo Nature Reserve (Northwest Province, South Africa) and transported in a desiccated state to the University of Cape Town (North West Provincial Government Permit 062 NW-12; Cape Nature Permit AAA007-01733). The dried plant mats were transferred to a total of three growth trays containing soil from the collection site and maintained in the UCT Botany greenhouse after rehydration. Three weeks prior to sampling the plant trays were transferred to a climate-controlled plant growth chamber (Conviron Adaptis A350) to acclimatise to the experimental conditions: 16h long-day, temperature setting of 22°C, an average luminosity of 250 μmol $m^{-2}s^{-1}$ and watering three times a week.

3.2.2 Leaf collection and RWC calculation

X. humilis generally grows as a clonally-produced collection of individual plants connected by a mat of interwoven roots, thus for the sake of leaf collection it was assumed that all trays contained genetically identical plants (Myers et al., 2010). For sample collection, undamaged leaves were sampled from clusters of plantlets across all trays at the same time of day (11:30 am) over a period of approximately two weeks after the cessation of watering. Because there are few non-destructive methods of calculating RWC of a single leaf without disrupting the mRNA expression profile of that leaf (particularly one depending on water content, such as desiccation status), each sampled leaf was split down the mid-vein. One half was flash frozen in liquid nitrogen and stored at -80C until RNA extraction could be performed. The wet weight of the remaining half was measured using a microbalance, and was

subsequently wrapped in labelled aluminium foil and left to dry in a drying oven at 60°C. After three days, the dehydrated leaves were removed and weighed again to determine the dry weight. RWC was then calculated as:

$$RWC = (W_{wet} - W_{dry})/(W_{dry} \times AWC)$$

Where W_{wet} is the wet weight of the leaf tissue, W_{dry} is the dry weight of the same tissue, and AWC is the absolute water constant – a measure of the maximum turgidity of the leaf.

An approximate value for AWC was calculated by averaging the absolute water constant across a sample of leaves from all trays. Five healthy leaves were selected from each of three trays, giving a total of 15 leaves. The leaves were partially immersed in sterile, distilled H_2O in a petri dish, which was sealed with Parafilm and incubated at 4°C for 12 hours. After incubation, the leaves were removed from dish, thoroughly dried on a piece of paper towel, and weighed to determine the turgid weight (W_{turgid}). Finally, each leaf was wrapped separately in aluminium foil and left to dry at 60°C. After 3 days, the dried leaves were weighed once more to determine the dry weight. The AWC for each leaf was calculated using the following formula, and the average value for the 15 leaves determined:

$$AWC = (W_{turgid} - W_{dry})/W_{dry}$$

3.2.3 Experimental design, RNA extraction and quality assessment

The desired leaf RWCs for the RNA-Seq experiments were chosen to be 100% (fully hydrated), 80%, 60%, 40% and 5% (fully desiccated). RNA was extracted from leaves that had been calculated to have a RWC within ±6% of these values. Total RNA was extracted from each leaf using Qiazol reagent (Qiagen) and a modified version of the manufacturer's

instructions. Two-millilitre curved-bottom microfuge tubes were prepared before the RNA extraction containing three 3 mm stainless steel ball bearings and 400 μl Qiazol reagent (Qiagen). Frozen leaves chosen for RNA extraction were quickly transferred to each tube from -80°C storage, and the leaf tissue was disrupted by rapid agitation in a MM400 Oscillating Mill (Retsch) for 15 minutes at 30 Hz, or until homogenisation could be visibly confirmed. After disruption, an additional 600 μl Qiazol reagent and 200 μl chloroform were added to the sample, which was vortexed for 30 seconds and incubated on ice for 5 minutes. Chilled samples were centrifuged at 12,000 x g for 15 minutes at 4°C to pellet debris and separate the supernatant. The upper aqueous layer was transferred to a new tube, diluted in an equal volume of 70% ethanol made up in DEPC-treated sterile H_2O, and total RNA purified using an RNeasy Mini Kit (Qiagen) following the manufacturer's instructions, including an on-column DNase I digestion. The purified RNA was eluted in 25 μl RNase-free water. The samples were immediately placed in storage at -80°C, except for a small volume kept aside to be used for RNA quantification and quality assessment. The RNA concentration for each sample was measured using a Nanodrop ND-1000 spectrophotometer and was calculated by averaging the readings from three individual 1ul aliquots. The purity and extent of degradation of the RNA samples was visually analysed on a denaturing agarose gel. One microliter of each sample was denatured at 60°C for 5 minutes in 2 x volume of RNA sample loading buffer and electrophoresed for 30 minutes on a 1.2% denaturing formaldehyde agarose gel. RNA extracts determined to be of high quality at each RWC were pooled to form three independent biological samples for each RWC. The biological pools were comprised of total RNA from 1-5 leaves, and no RNA was shared between pools. Each pool consisted of an equal amount of RNA (in μg) from its constituent leaf RNA extracts, and was reassessed for both concentration and quality via NanoDrop and

gel electrophoresis after pooling. For transport purposes, a 5 µg aliquot of each RWC replicate pool was treated with 20 µl of RNAstable LD (Biomatrica), following the manufacturer's instructions. Each sample was then vacuum dried using a SpeediVac (Savant), wrapped in Parafilm and sealed in a protective heat-sealed moisture-resistant bag with a separate sachet of silica-based desiccant. In parallel, replicate samples were also treated and desiccated with RNAStable LD, stored for 7 days in an air-tight container, and then rehydrated in sterile, DEPC-treated water as per the manufacturer's instructions. These re-solubilised samples were sent for analysis on a Bioanalyser (Agilent) at the Centre for Proteomic and Genomic Research (CPGR) in Cape Town, South Africa, to confirm RNA stability.

3.2.4 RNA Sequencing

The RNAstable-protected samples were sent to the Beijing Genomic Institute (BGI) for sequencing. Sequencing libraries were prepared using the Illumina TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina), and 150 bp fragment size. Sequencing was performed over 8 lanes of an Illumina flow cell using a HiSeq2000 sequencing instrument and a 90bp paired-end amplification protocol. The resulting raw reads were pre-processed by BGI, and the cleaned data was made available for download from their FTP server. The unprocessed raw reads were also uploaded by request.

3.2.5 Read quality checking and pre-processing

Read quality was assessed using the FastQC software package (v0.10.1; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For each library, Illumina adaptors and low-quality nucleotides were removed with Trimmomatic (v0.32; Bolger et al.

2014), using a sliding window size of 4, a PHRED quality score cut-off of 20 and a minimum read length of 25. Orphaned reads that lost their read pair were also discarded from the dataset. Error correction of the remaining paired reads was performed using SEECER (v1.0.3; Le et al. 2013) using a kmer size of 25, a cluster log-likelihood threshold of -1 and maximum entropy cut-off of 0.6. For *de novo* transcriptome assembly, the SEECER-corrected (assigned and unassigned) read pairs from each library were combined and read levels were normalised using the Trinity *in silico* normalisation pipeline in paired-end mode, with a kmer size of 25, a maximum coverage of 50 and a PCTS (deviation in kmer coverage divided by average kmer coverage, as a percentage) of 200. The resulting normalised paired-end reads were then merged using BBMerge v6.2, part of the BBMap software package (http://sourceforge.net/projects/bbmap/) in ungapped (overlap) mode, with a 12 bp minimum overlap and a maximum of 3 mismatches. The BBMerge unmerged and merged reads were used in the Trinity and Transabyss assemblies (which allow the input of both paired and unpaired reads simultaneously), whereas Bridger assemblies were done only on non-BBMerge processed normalised read pairs.

3.2.6 De novo transcriptome assembly

Multiple *de novo* transcriptome assemblies were performed using the assemblers Trinity v2014-04-12 and Trinity v2.0.6 (Grabherr et al., 2011), TransABYSS v1.5.2 (Robertson et al., 2010) and Bridger v2014-12-1 (Chang et al., 2015). All assemblies were created using the latest version of the assembly software (with the exception of Trinity), with default settings and the read strand-specificity set to match the Illumina dUTP stranded sequencing protocol (forward-reverse, FR). Multiple kmer sizes were also used, where possible. However, Trinity uses a recommended pre-set kmer size of 25, and so was not adjusted. Bridger assemblies

were constructed using kmer size of 19, 21, 25 and 31. Trans-ABYSS assemblies were constructed using a kmer size of 25, 31, 35, 41, 45, 51, 55, 61, 65, 71 and 81. Ultimately, eight of these assemblies (Trinity k=25, Bridger k=21, 25, 31, Transabyss k=25, 41, 61, 81) were merged using the Evidential Gene pipeline (http://arthropods.eugenes.org) (Gilbert, 2013), yielding a primary and secondary transcript set (putative coding transcripts and potential isoforms). *De novo a*ssembly and assembly post-processing were performed on a 1TB supercomputing cluster (Dakon) at the UCT Computational Biology Division.

3.2.7 Read count and FPKM calculation

The SEECER-corrected paired-end reads for each of the 15 samples (five RWC with three biological repeats) were mapped to the combined primary and secondary transcript sets using Bowtie2 (v2.2.4; Langmead & Salzberg 2012). The parameters "--fr --nofw --no-mixed --no-discordant" were set, in order to account for the strand-specific libraries and paired-end reads, and the total number of mappings for each read was counted by including the parameter "--all". The resulting BAM alignment files were used as an input to Corset (v1.03; Davidson & Oshlack 2014), which clustered the *de novo* transcripts based on both sequence similarity and expression levels, and generated a raw read count for each gene cluster (group of similar transcripts). The default Corset settings were used. The transcripts within each gene cluster produced by Corset were then assembled (if possible) using the CAP3 assembly software, to produce a consensus sequence (latest release as of March 2015 for Linux x86_64; Huang & Madan 1999). FPKM (Fragments Per Kilobase of transcript per Million mapped reads) was calculated for each gene cluster using the raw counts produced by Corset and the length of the gene cluster consensus sequence or, if a consensus sequence

was not available, the length of the longest transcript within a cluster as a representation of the gene length.

3.2.8 Transcriptome evaluation and annotation

The *X. humilis* transcriptome was evaluated against a draft version of the *X. humilis* genome (Schlebusch & Illing, unpublished data). For transcript annotation, the assembled transcripts were compared to known protein sequences with blastx against local installations of the Swissprot, Uniref90 and nr databases. Parameters included "-max_target_seqs 20" and "-evalue 1e-5". Initial alignments were performed using the settings "-word_size 6 -threshold 25" for the sake of speed; sequences with no match after the first round of alignments were run with "-word_size 4 -threshold 16", and lastly using the default word size/threshold settings. Output was formatted as XML ("-outfmt 5") for compatibility with the import function of Blast2GO. Transcripts that generated no subject hit in any of these databases were also compared to the NCBI non-redundant nucleotide database (nt) using blastn. Transcripts within the same Corset gene cluster that did not have a similar top BLAST hit were flagged as "ambiguous" in terms of protein-coding identity. Interpro terms were extracted from the consensus sequence or sequence of the longest transcript within each gene cluster using Blast2Go Pro (v3.0.10; Conesa et al. 2005), and GO ontology terms were assigned to the transcripts where possible using imported local blast results from the Swissprot database and the recommended mapping/annotation settings.

3.2.9 Differential expression analysis

Differential gene expression analysis at each RWC for each of the transcript clusters was determined using DESEq2 (Love et al, 2014) and the raw counts determined by Corset for

each gene cluster. Differential expression was performed using a log-likelihood ratio test (LRT), and default DESeq2 parameters. Genes were called as differentially expressed if the false discovery rate (FDR) was less than 0.01. The alpha value for DESeq2 independent filtering was set to match this FDR value. All DESeq2 analyses were performed using the x64 Windows GUI version of R (v3.2.0; R Development Core Team 2015) and the latest version of DESEq2 (v1.8.1; Love et al. 2014), with all required packages.

3.2.10 Gene expression validation and clustering

Gene expression values in this study were compared to that of a previous microarray study that measured that changes in gene expression in *X. humilis* leaves during desiccation (Shen, 2014). The microarray cDNA sequences were compared by reciprocal BLASTN with default settings to the assembled *X. humilis* transcriptome to confidently identify homologous transcripts. The normalised microarray expression values were exponentiated to convert them from a logarithmic scale, and the expression of genes from both datasets was normalised to the maximum expression of each gene. Least-squares regression analysis was used to obtain an estimate of the correlation between the two datasets at each specific RWC, and between each RWC. A null correlation value was calculated by determining the average correlation for each gene to all other genes at that RWC by resampling.

Expression vectors for differentially expressed genes were predicted using Multi-Experiment Viewer (MeV; www.tm4.org). Only gene clusters with a maximum FPKM > 1 and fold-change > 2 were considered for analysis so as to reduce the size of the dataset. Regularised log expression values for all genes were calculated using DESeq2 after averaging replicate counts, using the "collapseReplicates" and "rlog" (blind = FALSE)

DESeq2 functions. These values were used as input for MeV. The expression levels of all genes were centred and normalised within MeV, and expression profiles were clustered using K-means clustering (k=16, 1000 iterations, distance metric: Euclidian).

3.2.11 GO enrichment analysis

GO enrichment for each expression cluster was calculated and visualised using the BINGO software package (v3.03; Maere et al. 2005), part of the CytoScape network analysis tool (v3.2.1; Shannon et al. 2003). Only GO terms with an adjusted p-value of less than 0.01 were considered significant. A custom annotation file containing the *X. humilis* transcript IDs and associated GO terms was used as input, and the full set of successfully-annotated *X. humilis* gene clusters was used as a reference set.

3.3 Results

3.3.1 Plant material

It was important to determine the AWC prior to calculating RWCs of each leaf sample. The AWC across leaves and trays was fairly variable, as would be expected in a biological system, but averaged at around 3.4 (±0.36 SD) (Table 3.1). Leaves of dehydrating *X. humilis* plants were collected at 11:30am daily over a period of twelve days during late July and early August. In total, 105 leaves were collected across all three trays. A dehydration curve (RWC versus day sampled) can be seen in Figure 3.1. Leaves were grouped into bins of 100%, 80%, 60%, 40% and 5% RWC (±6% RWC) for further analysis (Supplementary Table S3.1).

	Leaf 1	Leaf 2	Leaf 3	Leaf 4	Leaf 5	Average
Tray A	3.64	3.36	3.55	2.24	2.29	3.02
Tray B	4.26	4.13	3.72	3.40	3.11	3.73
Tray C	3.68	4.18	3.81	3.24	2.37	3.45
					AWC	3.4 ± 0.36

Table 3.1. Individual AWC of *X. humilis* leaves from each plant tray. AWC of the 15 leaves used to determine global AWC (± SD) for the experiment.

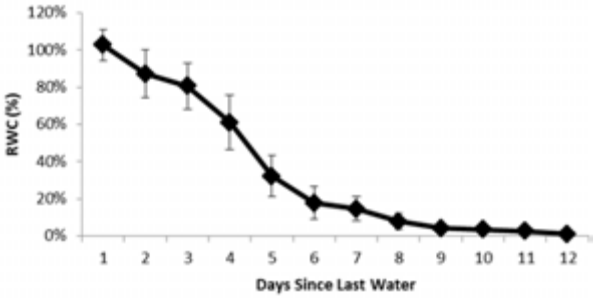


Figure 3.1. Average RWC (±SEM) for *X. humilis* leaves collected on each day after the cessation of watering.

	100% RWC			80% RWC			60% RWC			40% RWC			5% RWC		
Pool	Leaf	RWC	(ng/µl)	Leaf	RWC	(ng/µl)	Leaf	RWC	(ng/µl)	Leaf	RWC	(ng/µl)	Leaf	RWC	(ng/µl)
A	(A)003	107.5%	352.6	(B)034	82.0%	212	(A)010	60.5%	324.2	(A)027	40.4%	94.9	(A)091	5.1%	435.3
	(A)009	101.3%	165.4	(B)035	79.0%	291.3	(A)005	55.8%	229.9	(C)039	40.5%	793.5	(A)102	4.9%	119.5
	(A)002	100.7%	293.3							(B)013	37.3%	98.4	(A)045	4.8%	346.1
	(A)008	97.5%	398.9							(B)053	36.8%	108.5	(A)099	4.6%	193.9
													(A)0054	4.5%	291.5
	RIN: 6.0			RIN: 4.6			RIN: 5.5			RIN: 5.6			RIN: 6.1		
B	(B)019	112.6%	160.9	(B)032	77.3%	264.9	(C)003	64.3%	328	(A)031	40.0%	422.9	(B)066	6.7%	437.2
	(B)002	105.0%	429.2	(C)001	85.2%	391.8				(C)032	37.3%	951.3	(B)067	5.7%	434.8
	(B)012	102.4%	287.7							(B)052	40.9%	250	(B)070	5.2%	644.6
	(B)001	99.5%	463.5							(B)057	38.2%	648	(B)083	5.1%	553.4
	(B)018	97.5%	283.5										(B)092	4.4%	362.7
	RIN: 5.7			RIN: 4.7			RIN: 5.2			RIN: 5.6			RIN: 6.4		
C	(C)009	120.7%	804.2	(B)030	75.0%	575.1	(A)006	59.4%	594.5	(A)020	36.0%	84.6	(C)097	5.4%	628.8
	(C)002	108.7%	707.3	(C)011	82.6%	734.4	(C)018	63.7%	791.5	(C)033	36.2%	210.5	(C)074	5.1%	785.4
	(C)007	108.4%	418.4							(B)061	43.7%	326.5	(C)066	5.0%	444.5
	(C)005	100.8%	567							(A)033	35.1%	184.7	(C)049	4.6%	522.5
	(C)012	100.7%	629.6										(C)047	4.6%	665.8
	RIN: 5.1			RIN: 3.5			RIN: 5.1			RIN: 5.9			RIN: 6.0		

Table 3.2. Leaf samples in each biological pool. RNA extractions corresponding to each leaf in the biological sample pools. The letter in brackets next to each leaf corresponds to which of the three plant trays the leaf was collected. Also included is the RIN value determined by BGI for each pooled leaf sample.

RNA extractions were performed on each leaf individually. After quality assessment of each RNA sample via gel electrophoresis, three independent biological pools were created from leaves at each of the five selected water contents (Table 3.2). The RIN (RNA integrity) scores of the pooled RNA samples were relatively low, especially those of the 80% RWC samples. However, variation in RIN scores can differ across tissue type, species and RNA extraction techniques, and often does not reflect the true degradation status of the measured sample (Johnson et al., 2012). The differences in terms of gene expression, secondary metabolite concentration and other physiological attributes between a hydrated, drying and desiccated *X. humilis* leaf could arguably be greater than between two unrelated tissues in a non-VDT plant, which may account for the variation in RIN values observed here. Additionally, BGI routinely use a mammalian RNA standard during RIN calculation of plant samples, further skewing the meaningfulness of the result.

3.3.2 Preprocessing of reads

After pre-filtering, clean reads consisting of over 350 million paired-end sequences, spread across 15 experimental samples, were downloaded from the BGI servers. The unfiltered ("raw") reads were also requested for diagnostic purposes, but the quality of the clean reads was considered sufficient for downstream assembly and differential expression analysis. Pre-processing of the sequencing reads occurred in two steps: quality assessment, trimming, and error correction, which were performed on each sample individually to account for library-specific biases; followed by *in silico* normalisation and merging of paired-end reads, which was performed on all remaining combined read pairs for the purpose of *de novo* transcriptome assembly. The pre-processing and assembly pipeline is shown in Figure 3.2, and an overview of the scripts and commands used can be found at the end of the supplementary data (Figure S3.12).

3.3.2.1 Quality assessment

There are a number of measures used to quantify the suitability of short sequence reads for downstream processes. These include: per-base and per-read sequencing quality (confidence that the base call for a particular nucleotide is correct), nucleotide content (the distribution of ATGC and N across the library and across reads) and GC content. Additional measures include the average length of the reads in the library, the level of duplication of sequences, and the presence of over-represented sequences or kmers (sequences of a specific length *k*). The expected observations for each of these measures will be different depending on sequencing strategy used (Illumina, 454, PacBio), sample type (metagenomics, single cell culture or tissue sample), library preparation (sonication vs. enzymatic digest) and target

molecule (genome sequencing vs RNA-Seq). The quality of our pre-filtered sequencing data was consistent with general RNA-Seq data (Fig. S3.1A-D). Per-base sequence quality (the distribution of PHRED quality scores across each read) matched the expected distribution for Illumina sequencing reads (Fig. S3.1A). The average PHRED score for all bases across each read was generally well above 35 (Fig. S3.1B) and the reads were exactly 90bp in length with 0% N content, likely due to pre-filtering by BGI. Per-base sequence, GC and kmer content analyses were marked as "failed" by the FastQC software due to a non-random distribution of bases in approximately the first 10bp of each read (Fig. S3.1C&D), but this is to be expected for RNA-Seq data. As a consequence, the per-sequence GC% content was also marked as irregular in a number of libraries. Duplication levels varied between 66.53% and 77.18% across the libraries, and did not seem to correlate with the RIN value of the underlying samples (Fig. 3.3). Duplication levels tended to be slightly higher in the reverse reads by about 1-5%, and was higher in leaves with low relative water contents. Despite high duplication levels there were no overrepresented sequences in the libraries in general.

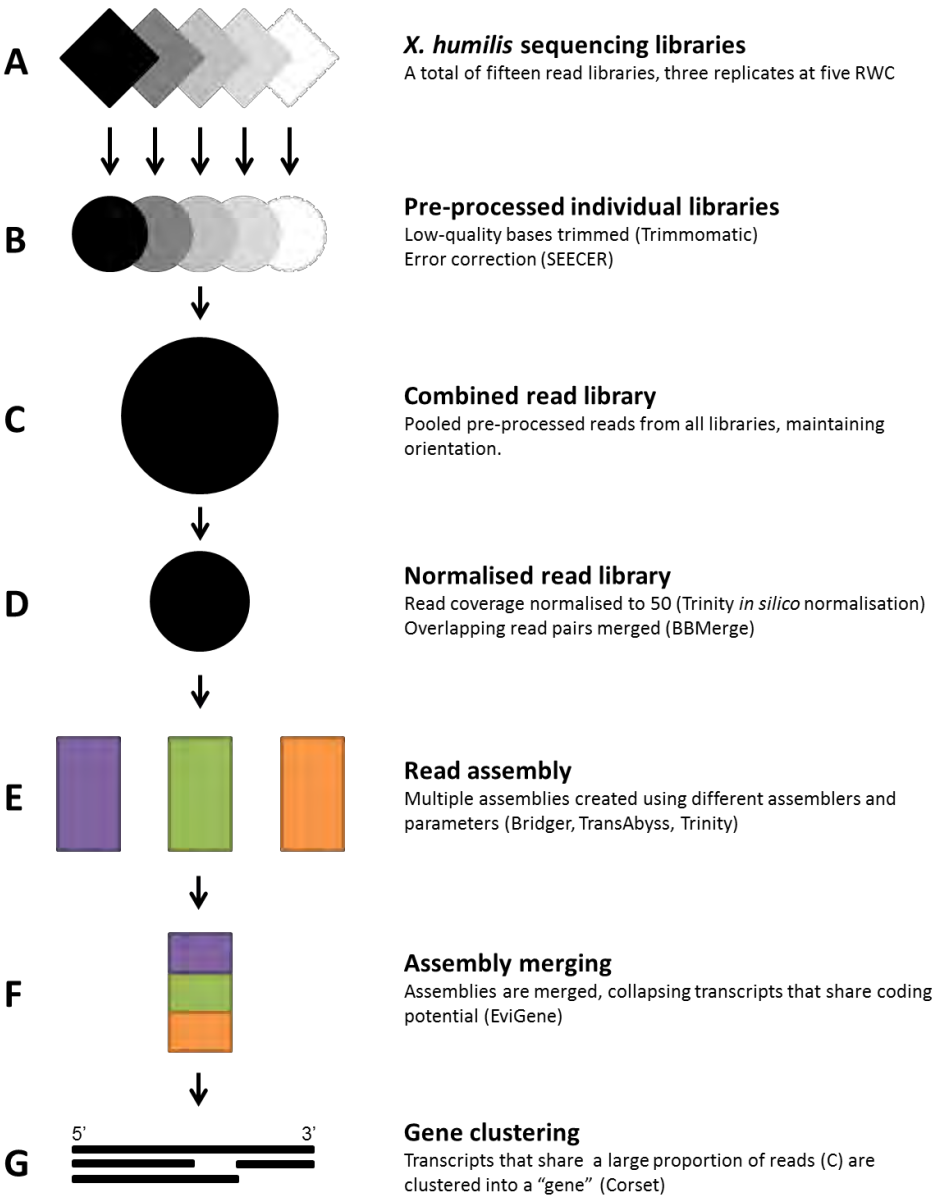


Figure 3.2. Read processing and assembly pipeline. A-D) Read pre-processing, including trimming and error correction (library-specific), and *in silico* normalisation and read-pair merging (performed on all reads). E-G) Assembly, merging of assemblies and transcript clustering.

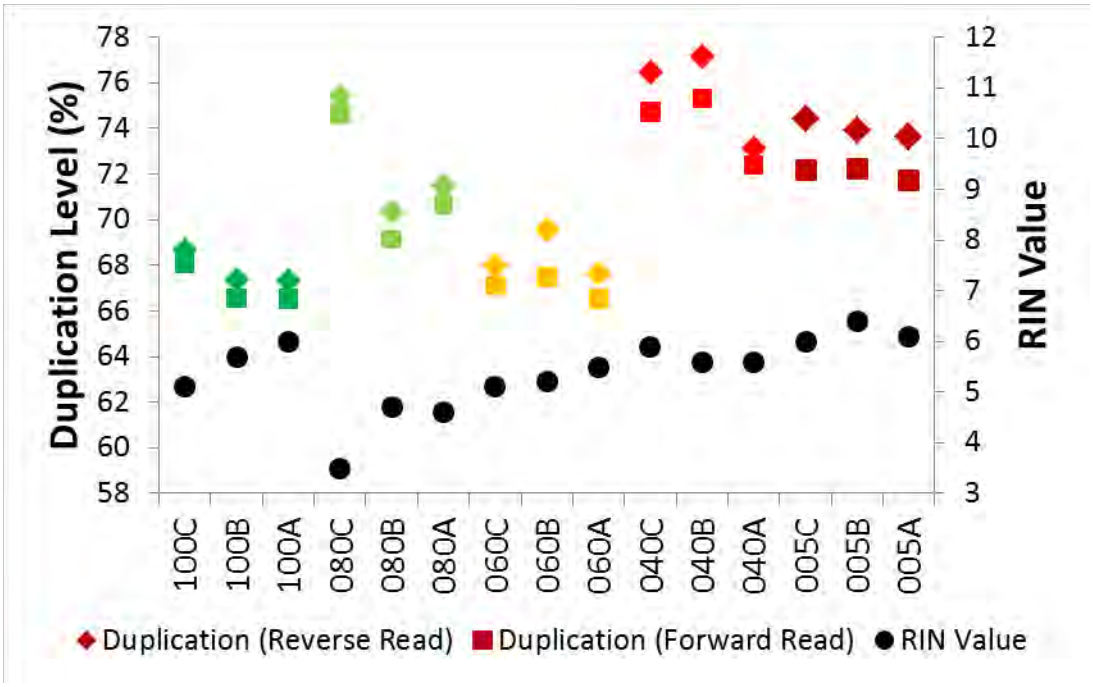


Figure 3.3. Read quality scores for *X. humilis* RNA-Seq Illumina paired-end reads. Per-library duplication rates for forward (square) and reverse (diamond) reads, and RNA quality (RIN) for the pooled RNA sample corresponding to each library.

3.3.2.2 Read trimming, deduplication and error correction

Trimmomatic was used to remove low-quality nucleotides from the read ends prior to assembly and/or mapping for differential expression analysis. Trimming had minimal effect on the dataset, as it had already been pre-processed by BGI (Fig. 3.4A). After trimming, read errors were corrected using SEECER. In brief: because the kmer profiles of RNA-Seq data differ drastically from those of genomic NGS data, SEECER first pre-assembles reads into clusters based on overlapping sequences. Then, reads that are assigned to a particular cluster that pass a threshold suggesting that they originated from that cluster are corrected based on the consensus sequence of that cluster. On average, approximately 85% of the

reads in each library could be assigned to a SEECER cluster, and 28% of these reads had at least one error corrected by SEECER (Fig. 3.4A).

3.3.2.3 In silico normalisation and paired-end merging

For *de novo* transcriptome assembly, the reads from each library were merged into a combined paired-end dataset. To reduce both assembly errors due to excess coverage of high-abundance reads, as well as the computational power required to run the assemblies, the combined reads were normalised to a maximum kmer coverage level of 50 (with k = 25) using the Trinity *in silico* normalisation pipeline. This succeeded in reducing the number of reads from 350 million pairs to just 27.1 million, a reduction of well over 90% (Fig. 3.4B). Most of the reads were discarded due to being over the kmer coverage limit, but a small number (0.25%, 1.75 million reads), were discarded due to aberrant kmer coverage across the read indicative of sequencing errors. The normalisation step took approximately 15 hours, using 8 parallel threads of a 1TB supercomputing cluster. The total number of reads was further reduced by merging any remaining read pairs that had an overlap due to the reported average insert size (150 bp) being smaller than double the read length (90 bp*2 = 180 bp). Two merging tools were tested: FLASH and BBmerge. Both succeeded in merging multiple pairs into single reads. FLASH merged a larger number of sequences (56.1%) with default settings compared to the more conservative BBMerge, which merged 51.25% of reads (Fig. 3.4C). Ultimately, the BBMerge merged reads were used for *de novo* transcriptome assembly, so as to reduce the risk of introducing assembly errors that could arise due to overly aggressive merging of read pairs. The mode of the histogram of merged reads was approximately 140bp, consistent with the desired average insert size (150bp) of our TruSeq libraries and the minimum BBMerge overlap requirement (Fig. 3.4D).

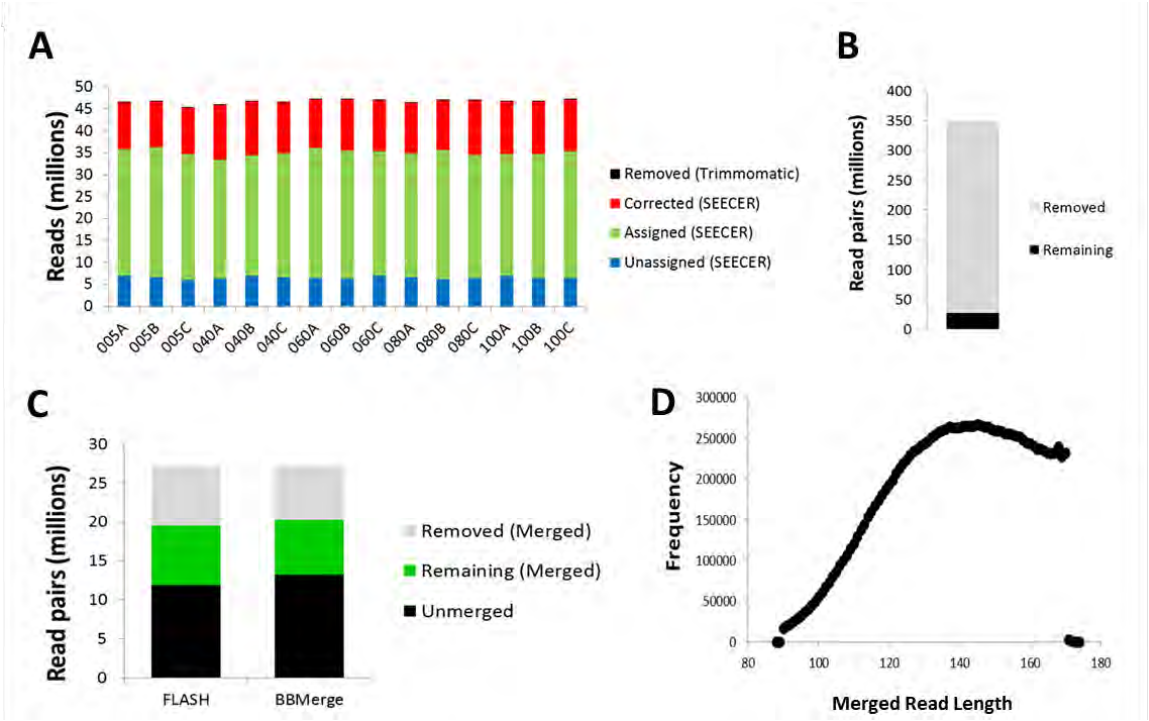


Figure 3.4. Pre-processing of *X. humilis* reads. A) The effect of pre-processing on read count in each library: removed by trimming (black), assigned to a SEECER HMM (green and red), corrected by SEECER (red), not assigned to a SEECER HMM (blue), B) the number of reads removed by Trinity after *in silico* normalisation, C) the number of paired reads merged into a single read by BBMerge and FLASH, D) histogram showing the frequency of read merging by BBMerge versus combined length of the merged reads. This value should tend towards the average insert size for the RNA-Seq libraries.

3.3.3 Transcriptome assembly

Multiple *de novo* transcriptome assemblies were created, using a number of assemblers, assembly parameters and kmer sizes. The three assemblers used were Trinity (versions "r20140413p1" and the latest development version of release 2.0.6 as of March 2015), Bridger (release 2014-12-01) and Trans-ABYSS (v1.5.2). The latest stable release of Trinity (2.0.6) failed to complete the assembly pipeline on the computing cluster used for these assemblies, possibly due to a known file organisation bug in that version of the software. A

development version (beta version of the next release) was used instead, which fixed many of the errors of the previous release.

3.3.3.1 Read normalisation

The effect of *in silico* read normalisation was investigated for each assembler. For the sake of comparison, these assemblies were all performed with a kmer size of 25 on 8 parallel CPU threads and no memory limit on a 1TB RAM machine. Only the latest version of the Trinity (development branch), Bridger and TransAbyss assemblers were compared, and only contigs with a length greater than 200 were reported.

Read normalisation had a big impact on assembler run time (Fig. 3.5A). Trinity took by far the longest of all three assemblers to run, topping out at nearly 140 hours when using all reads and default settings. Normalisation dropped this by half, reducing runtime by over 72 hours, as well as decreasing the assembly disc space and memory requirements. The Bridger assembler took nearly 48 hours to assemble the entire read set, but runtime could be reduced by over 75% when using normalised reads, reducing assembly time by over 35 hours. Transabyss took under a day to assemble even the un-normalised read set, and the effect of read normalisation on runtime was significantly less than with the other assemblers – assembly time was reduced by 25%, which amounted to less than a 3 hour difference for a single assembly. In Trinity, read normalisation lead to an increase in the total number of putative transcripts assembled, as well as an increase in the total size of the assembly (Fig. 3.5B&C). This trend was reversed in Transabyss, which assembled fewer bases and contigs, and the effect was negligible on the Bridger assembly. However, read normalisation did lead to an increase in contig N50 across all assemblers (Fig. 3.5D). Read normalisation had only a

minor effect on the mapping rate of reads against the resultant assemblies. The total number of non-normalised reads that mapped concordantly to each assembly with Bowtie2 was reduced in those created from the normalised reads with Bridger (by 8%), but increased very slightly in the Transabyss and the Trinity assemblies (by 1.5% and 0.35% respectively) (Fig. 3.5E).

Read normalisation thus improved many of the reported assembly statistics of *X. humilis* transcriptome assembly, while also reducing runtime and storage space, and so was used for all final assemblies. Although the time and space benefit of read normalisation is minimal for a single assembly run on most assemblers, it is extremely worthwhile when multiple assemblies are to be run consecutively when testing assembly parameters.

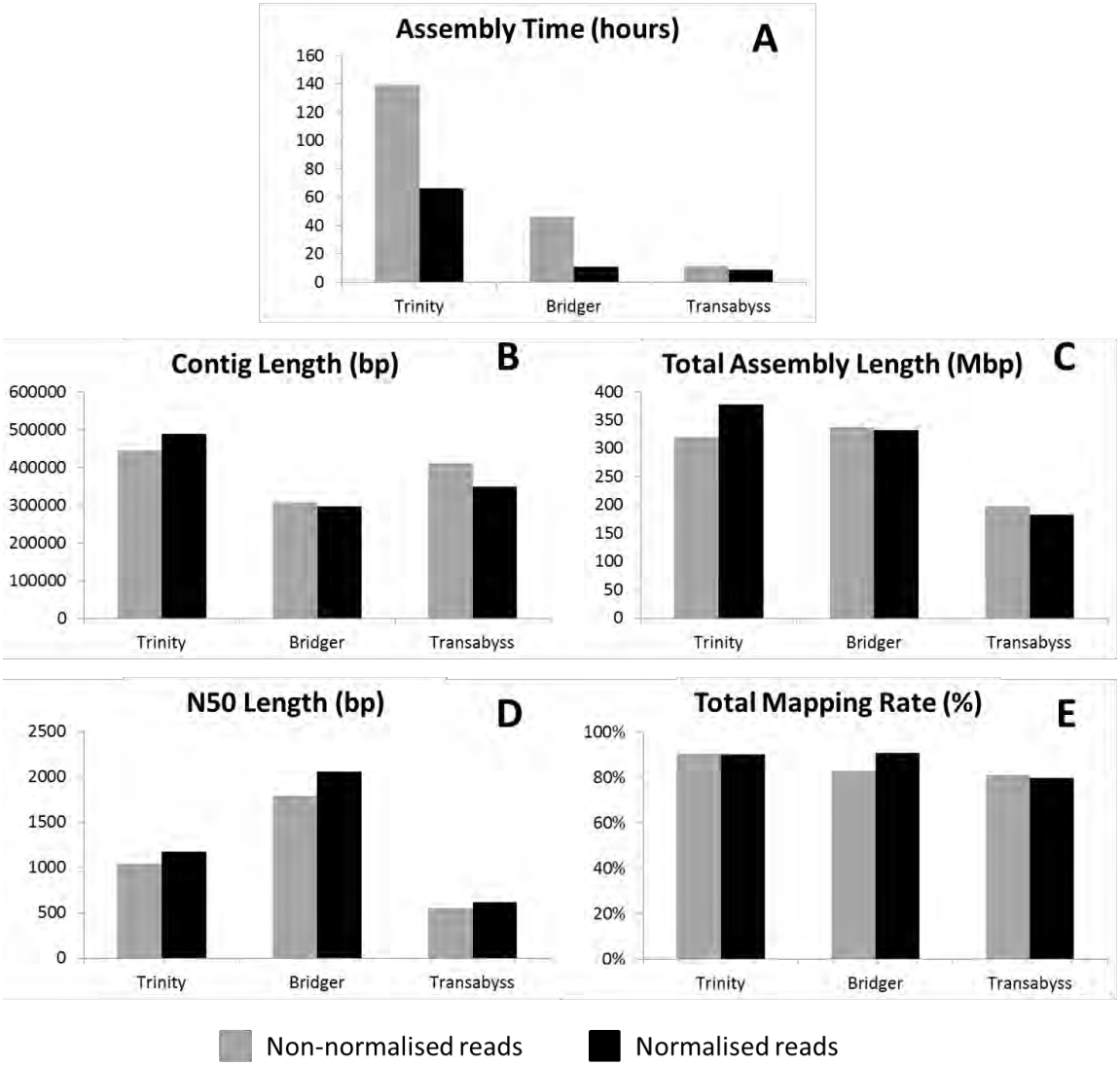


Figure 3.5. The effect of *in silico* normalisation on assembly statistics. Normalisation had varying degrees of effect on the assembly process and output of the three tested assemblers, including time taken to assemble (A), number of contigs (B), total length of the assembly (C), N50 contig length (D) and mapping rate (E).

3.3.3.2 De novo assembly results

In total, 17 assemblies were created using a range of kmer sizes across the three assemblers.

A comparison of the main assembly statistics can be found in Figure 3.6.

Two versions of Trinity were compared, using the recommended kmer size of 25. The latest version of Trinity (2.0.6, development branch) produced both the longest total assembly (as measured by the total assembled bases), as well as the greatest number of individual contigs (defined here as assembled sequences >200bp in length) of all the assemblers (Fig. 3.6A). This far exceeded those assembled by the previous version of Trinity, although the N50 contig length was subsequently lower: 1177bp using the latest version, compared to 1427bp using the older version. Bridger, a relatively new assembler, can handle kmer sizes of up to 32bp in length. *De novo* assemblies were performed using a kmer size of 19, 21, 25 and 31. On average, Bridger created large assemblies with fewer, longer putative transcripts, as can be observed by their comparatively high N50 contig lengths compared to the other assemblers. Total length of assembled bases and number of assembled contigs increased with kmer size between k = 19 to 25, but dropped slightly at k = 31. In contrast, the latest version of Transabyss produced the shortest, most fragmented assemblies. Although it produced fewer total contigs than Trinity, the total length of assembled bases averaged at half that of Trinity, and was also substantially lower than that of Bridger. Consequently, the N50 contig lengths of the Transabyss assemblies were universally lower than the other assemblers. The largest number of putative transcripts were produced at lower kmer sizes (k < 41), and dropped as kmer size increased. The length of the total assembled bases increased from k = 25 to 45, but dropped as kmer size approached read length. Mapping rates of reads against eight selected assemblies, determined using Bowtie 2, can be seen in Figure 3.6B. Only concordant alignments (both pairs align within a given maximum insert size from each other) in the correct orientation (Forward-Reverse (FR) for read pairs 1 and 2, respectively, due to the use of the dUTP Illumina strandedness protocol) were counted (see Methods). Total mapping rate as well as the proportion of multi-mapped reads (read pairs that aligned more than once

within the dataset) differed across assemblies (Fig. 3.6B). Average mapping rate was approximately 80%. Attempting to map reads in the incorrect orientation (e.g. RF vs FR) resulted in negligible read alignments (< 1%), suggesting stranded information was correctly accounted for by the assemblers used (data not shown).

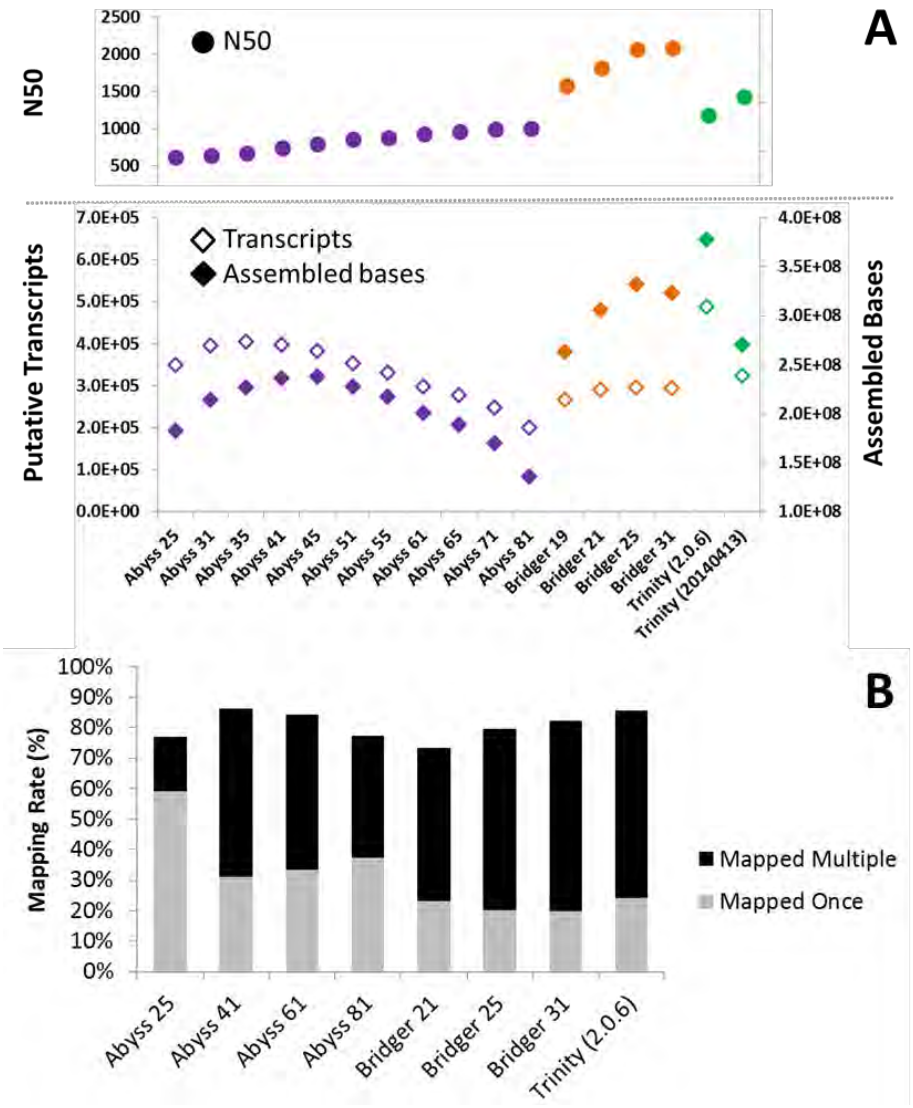


Figure 3.6. Statistics of *de novo* assembly of *X. humilis* reads with varying assemblers and parameters. A) Chart showing the N50 value (circles), number of transcripts (open diamond) and total assembled bases (solid diamond) for each assembly with Abyss (purple), Bridger (orange) and Trinity (green), B) concordant read pair mapping rate for each of eight selected assemblies. The proportions of reads that map uniquely are shown in light grey, whereas those that map multiple times within the assembly are shown in black.

3.3.3.3 Merging assemblies

Different assemblers and assembly conditions, such as kmer size, can assemble "true" transcript sequences to varying degrees of accuracy. To account for this, multiple assemblies are often merged and then filtered to remove redundant contigs - in theory outputting only the best-assembled version of each unique transcript. One approach to do this is embodied by the Abyss-merge pipeline, which merges multiple assemblies produced using different assembly parameters, and then uses a modified BLAT script to remove redundant contigs based on percent sequence similarity. This method was tested on the *X. humilis* Transabyss *de novo* assemblies using either all or a subset of the individual kmer assemblies. The average number of contigs produced by Transabyss was approximately 330,000, although this number varied across kmer sizes (Fig. 3.7A). Merging the first four assemblies (k = 25, 31 35 and 41) using a 95% sequence identity threshold and an indel tolerance of 1, resulted in a total output of nearly 650,000 contigs (Fig. 3.7A). The total number of contigs produced via Abyss-merge increased with the number of input assemblies, albeit in a nonlinear fashion. Merging all the Transabyss assemblies resulted in a "non-redundant" merged output of nearly 850,000 contigs, far in excess of what would be realistic for any organism, and indicative that this approach is unsuited to this *X. humilis* dataset. An alternative strategy for producing a biologically relevant reference transcriptome is to cluster contigs based on shared coding potential, rather than entire sequence identity. The Evidential Gene (EviGene) pipeline determines whether contigs are potentially protein-coding based on predicted coding sequence. Transcripts with a partial ORF (e.g. a low coverage transcript) or multiple ORFs (unspliced transcripts/intronic contamination) are penalised if a similar transcript codes for the same protein sequence and is of a better quality. This method is preferred over

percent-based merging techniques in many cases, as assembly errors are much less likely to produce viable coding sequence. Additionally, misassembled chimeric contigs are generally longer than their component sequences, thus merging multiple transcripts and selecting the longest can actually enrich for assembly artefacts. The Evidential Gene (EviGene) pipeline was used to merge a subset of the assemblies in this fashion: specifically, Trinity (v2.0.6 development branch), Bridger (k = 21, 25 and 31) and Transabyss (k = 25, 41, 61 and 81). The EviGene pipeline scripts were modified such that ORF prediction was restricted to only the forward strand of each contig, to account for the stranded RNA-Seq protocol used. This would prevent the potential merging of contigs produced from opposite strands, and the false output of antisense sequences. Prior to merging, the pool of contigs consisted of slightly over 2.6 million redundant sequences. After merging, EviGene designated 72,893 putative transcripts as "primary" (most complete predicted ORF-containing sequences), and a further 93,478 sequences as "secondary" (sequences with >50% coding sequence similarity to primary sequences – putative isoforms or transcript variants). The remaining contigs were classified by EviGene as "dropped", which implies that they had no coding potential or were redundant to the sequences in the primary or secondary sets. The proportion of the primary and secondary set contigs that were derived from each assembler largely correlated with the number of assemblies used as input (Fig. 3.7B&C). The majority of contigs could be assembled using smaller kmer sizes (k <= 25). However, the use of larger kmer sizes was clearly beneficial in the case of Abyss and Bridger when compared to Trinity, as they produced more unique transcripts (Fig. 3.7B&C).

I used the Corset *de novo* transcript clustering software to further combine the primary and secondary set of contigs into a single set of "gene clusters" – putative transcripts derived

from a single gene. Rather than compare contigs based on sequence similarity, Corset groups contigs based on shared read pairs and read-pair coverage: transcripts derived from the same genomic locus should have most reads in common, and differentially expressed isoforms should have an altered read coverage compared to similar transcripts across the dataset or time series. The non-normalised reads of all 15 samples (five RWC, three repeats each) were independently mapped to a combined dataset including both the primary and secondary sequences using Bowtie 2. In this instance all alignments for each read pair were counted, as recommended by Corset (Bowtie 2 parameter "--all"). Mapping rates were fairly consistent across all libraries (77±0.02%, Fig. 3.7D). This is lower than the average mapping rate for individual assemblies, as would be expected for a filtered dataset. Of the original 166,371 primary/secondary sequences, Corset found 9,715 to have read counts too low to be considered genuine transcripts (total count < 10 across all 15 replicates). The remaining 156,656 contigs were grouped into 76,868 putative gene clusters. In addition to grouping information, Corset also includes raw read counts for each gene cluster with its output, which was used for downstream differential expression analysis.

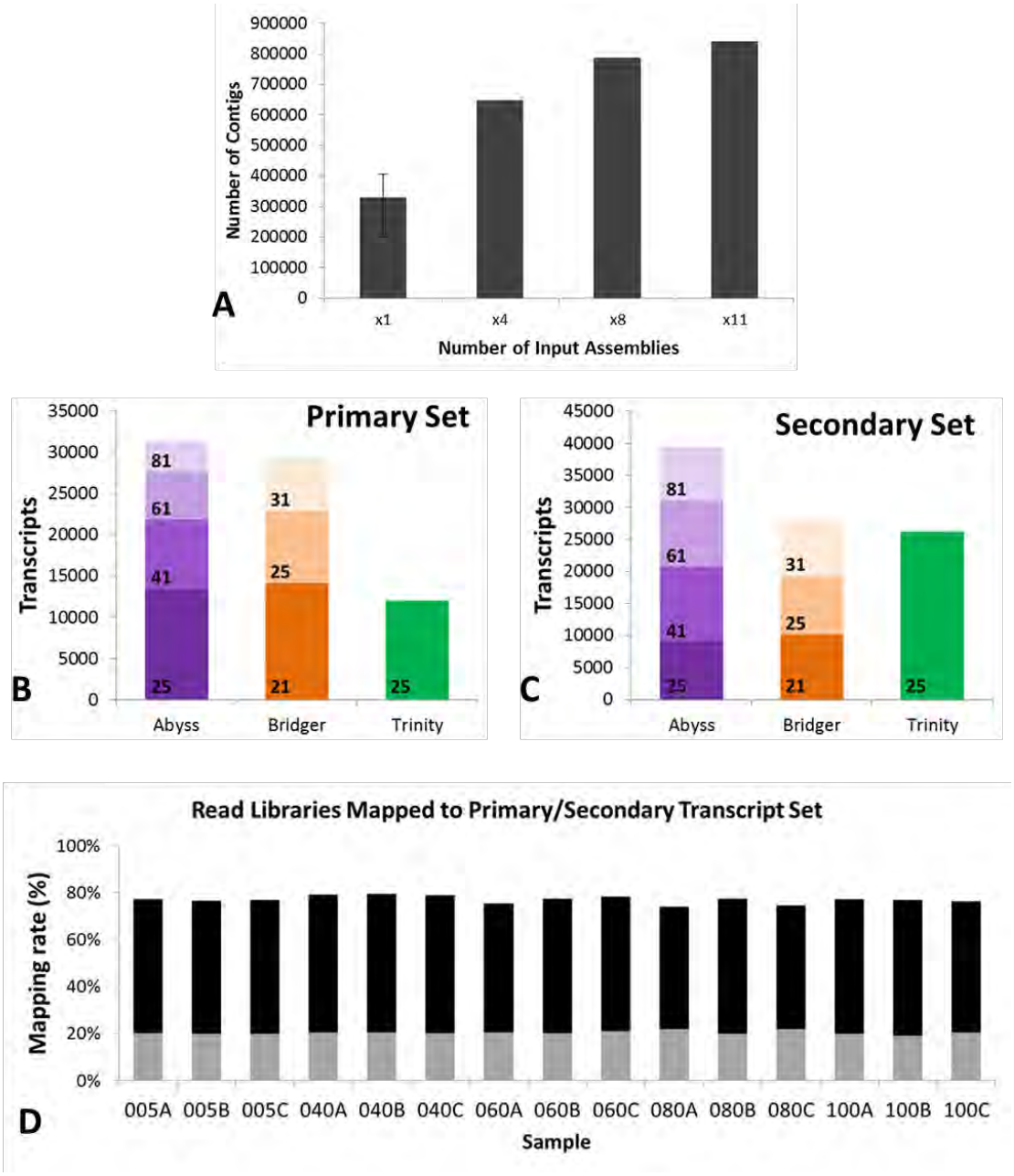


Figure 3.7. Merging *de novo* transcriptome assemblies. A) The results of merging varying numbers of Transabyss assemblies created with different kmer parameters using the TransAbyss-merge pipeline. The first column (x1) is the average number transcripts produced by all 11 TransAbyss assemblies and includes the SD, B) results of merging using the EviGene pipeline, C&D) the proportion of transcripts derived from each assembler and kmer size (shaded columns, the values represent kmer length) for the primary and secondary EviGene transcript sets, E) percentage of reads in each library that map to the combined Primary/Secondary EviGene set either uniquely (light grey) or multiple times (black).

3.3.3.4 Transcriptome annotation

Preliminary annotation of the final set of 156,656 contigs (independent of Corset clustering) was done using a local installation of NCBI Blast and Blast2Go Pro. The full set of sequences was aligned to the Swissprot, UniRef90 and nr/nt databases using blastx/blastn (see Methods). For large databases, altering the default word size/threshold score (to either 6/25 or 4/16 depending on the size of the database) of the blastx search was found to have a large impact on alignment speed (as much as a 30-fold increase), without negatively impacting the search result. The majority of the contigs (127,151 – 81.2% of the total) were assigned a BLAST hit with an e-value less than 1e-5 (Fig. 3.8A&B). The probability of matching a known protein, as well as the significance of the sequence alignment, scaled with contig size (Fig. 3.8A). In total, the contigs matched to a set of approximately 44000 unique protein ids. The vast majority of alignments were to proteins in the Swissprot or UniRef90 databases; the nr and nt databases combined supplied less than 100 protein hits that were not present in other databases (Fig. 3.8B). Information regarding the taxonomy of each sequence was predicted based on the Uniref90 alignments. The majority of sequences matched known plant proteins; however, a large proportion aligned to fungal proteins, or had no alignment (Fig. 3.8C). Sequences with matches to animal and bacterial proteins, or subjects with no available taxonomic group, made up a much smaller proportion of the dataset (Fig. 3.8C). Plant matches were mainly comprised of hits to proteins within *Phoenix dactylifera* (45%) and *Musa acuminata* (18%), but included contributions by *Vitis vinifera*, *Nelumbo nucifera*, *Oryza sativa* and *Theobroma cacao* (Fig. 3.8D). Fungal matches were largely spread across the *Pleosporineae* (15%), *Phaeosphaeria* (15%), *Leptosphaeria* (10%) and *Pleosporaceae* (10%) genera (Fig. 3.8E).

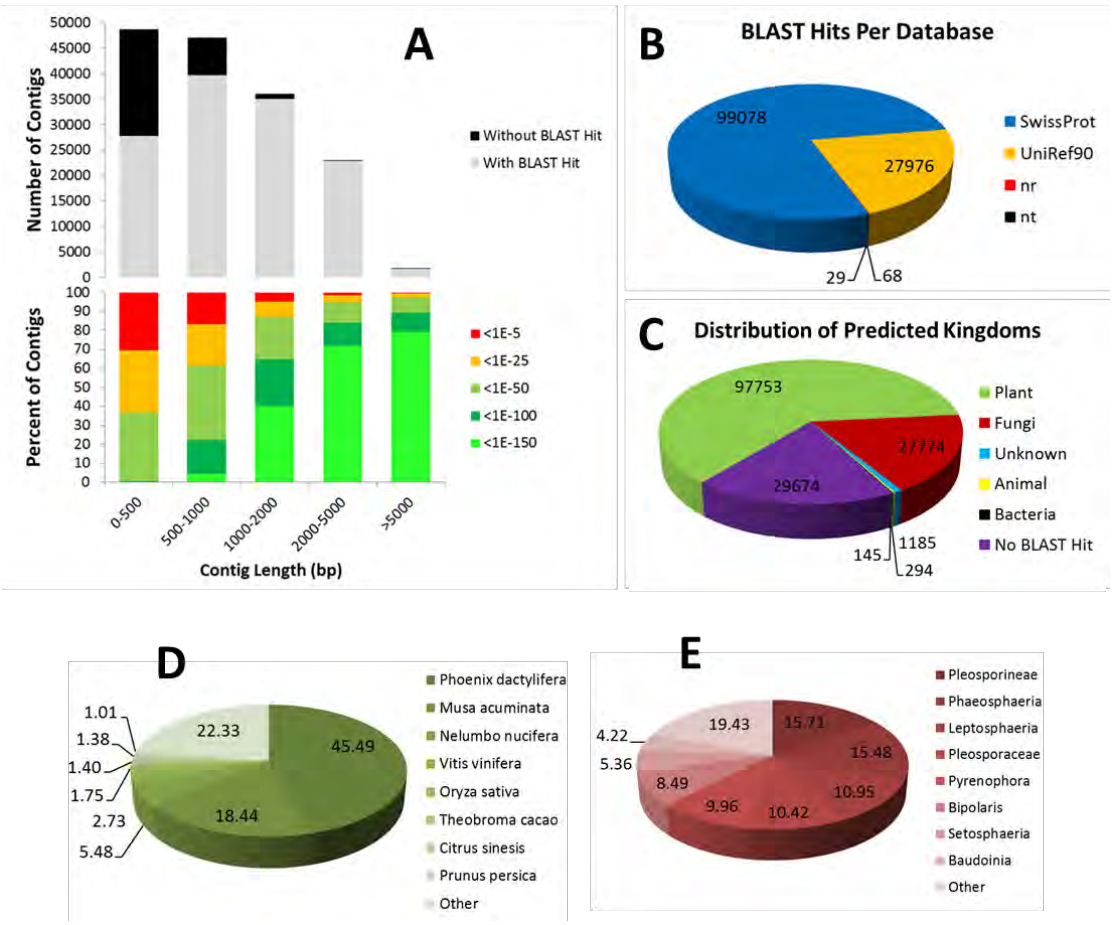


Figure 3.8. Annotation of the EviGene *X. humilis* transcriptome. A) Top: number of *X. humilis* contigs with a hit in an NCBI Blast database (hits, light grey; no hits, black) grouped by contig size; Bottom: the distribution of e-values for hits in each size group as a percentage, B) number of hits per database when searched by database priority (Swissprot, Uniref90, nr, nt); C) taxonomic distribution of *X. humilis* contig blastx top hits, D) distribution of plant species amongst aligned protein hits, E) distribution of fungal genera amongst the aligned protein hits.

Despite the disproportionate number contigs with fungal protein matches, these sequences represented less than 2% of all RNA-Seq reads across all biological samples (Fig. 3.9A). To verify the presence of possible fungal contaminants, the *de novo* assembled contigs were aligned to a rough draft of the *X. humilis* genome using blastn. Despite the fragmented nature of the current genome assembly (Stephen Schlebusch, personal communication), the

rate of alignment was impressive. Nearly 100% of contigs predicted to match known plant proteins aligned to the genome, as did approximately 50% of sequences with "Animal", "Bacterial" and "Unknown" predictions (Fig. 3.9B). Similarly, 91% of contigs with no match to known proteins nonetheless appeared to derive from *Xerophyta* genomic sequence. In contrast, only 2.7% of contigs that were classified as originating from fungal transcripts aligned to the *X. humilis* genome (Fig. 3.9B).

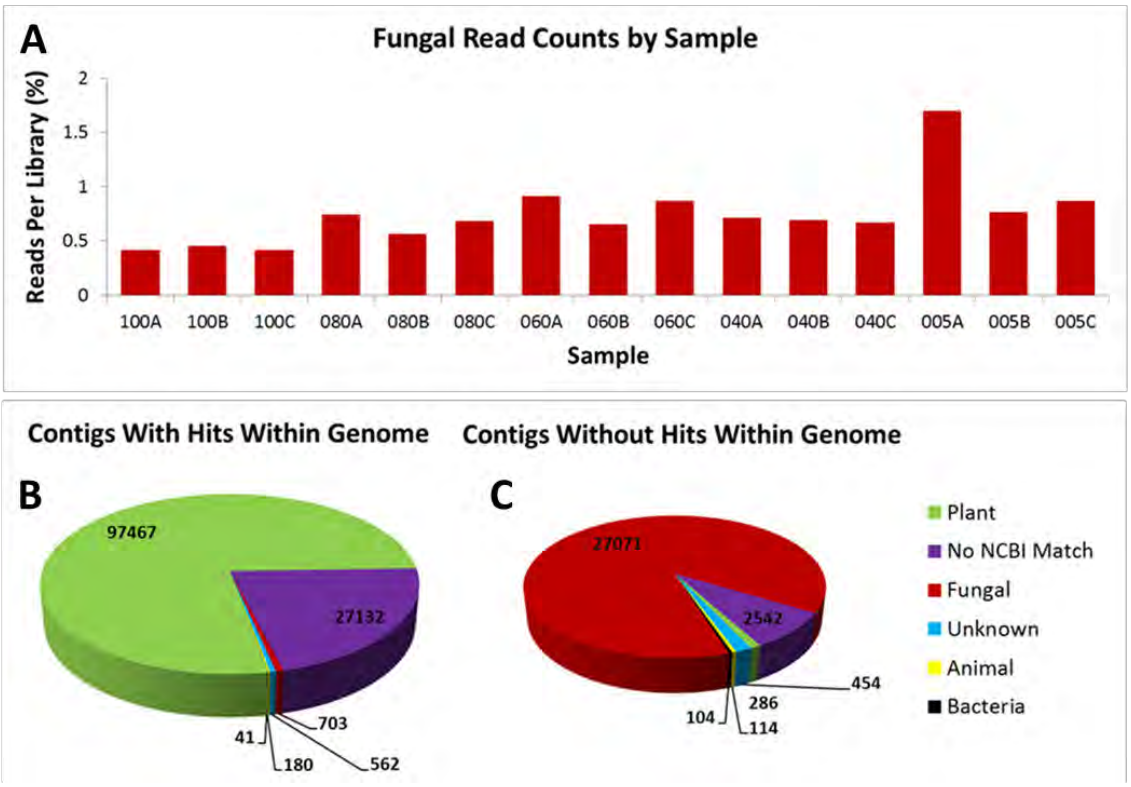


Figure 3.9. Verification of *X. humilis* transcript origin. A) The percentage of reads within each library that mapped to contigs predicted to be fungal based on blast alignment, A&B) number of sequences that aligned (A) or did not align (B) to the draft *X. humilis* genome and their predicted taxonomic origin.

As an independent validation of the Corset clustering method, the protein identity and taxonomic classification of each contig was compared to the other contigs within its Corset gene cluster. Of the 76,686 total clusters, 30,560 (39%) contained more than a single contig. Of these, only 147 gene clusters contained contigs with conflicting taxonomic origin of the best protein hit (i.e. hits to an animal and plant sequence), and 2257 gene clusters differed in the top protein hit between contigs. However, the majority of these contigs still appeared to match similar classes of proteins. The local blastx results for each contig were imported into Blast2GO for gene ontology mapping and annotation (see Methods). The GO annotations for contigs within the same Corset gene cluster were merged where necessary, resulting in 43,934 annotated clusters (57% of the total). The contigs within each of these gene clusters were assembled using CAP3 if possible to provide a representative consensus sequence; otherwise the longest sequence was chosen as representative of the cluster. Interpro scans could thus be performed for each gene cluster via Blast2GO, which succeeded in adding protein domain information for 16,600 clusters, and additional GO information for a further 8,950 gene clusters.

3.3.3.5 Differential expression testing

Differential expression of the 76,868 gene clusters was tested using the R-based DESeq2 package for Bioconductor. As input, DESeq2 requires raw read counts (not pre-normalised values, or size corrected ones such as RPKM/FPKM) for each gene sample, which were provided via the Corset step of our *de novo* assembly pipeline. As a first step, regularised-log transformed (rlog) sample data was compared using Principal Components Analysis (PCA). Generally, biological replicates for each RWC clustered together across the principal components. The major exception to this was the 80% RWC samples, which were spread out

across both axes (Fig. 3.10). None of the post-sequencing analyses indicated anything erroneous with these libraries. The 80% RWC samples had the lowest RIN values of all the RNA samples (Fig. 3.3), although RNA quality looked similar to other samples when visualised on a denaturing RNA gel. Additionally, the 80% RWC samples had some of the fewest biological samples per RNA pool (average 2 leaves) compared to other RWC; however, the 60% RWC samples did not display such variation despite containing even fewer leaves per pool (1.6; Table 3.2). This points toward the variation possibly being a biological phenomenon that was not masked due to the low number of leaves sampled. As no sample was an obvious outlier and there was no scientific basis to remove them even if they were, the 80% RWC samples were kept intact. With the exception of the 80% samples, the majority of the variance (first principal component, 81%) seemed to quite clearly separate the "wet" samples (100% and 60% RWC) from the "dry" samples (40% and 5% RWC; Fig. 3.10).

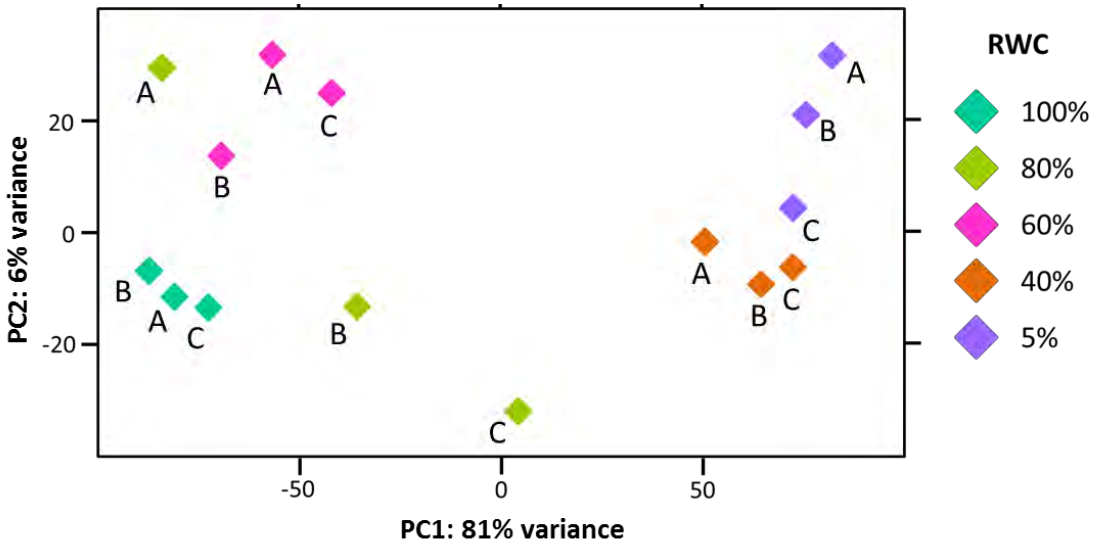


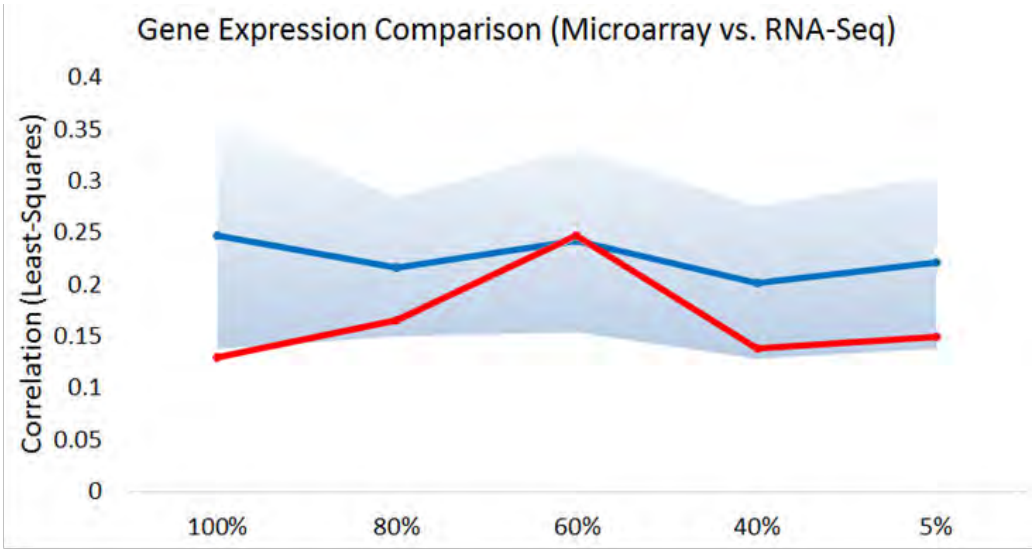
Figure 3.10: Principal Component Analysis of the regularised log read counts per gene for each library. Raw read counts were converted to regularised log (rlog) values using DESeq2 and subsequently used to create a diagnostic PCA plot.

Differential expression was calculated using the log-likelihood ratio test (LTR) function of DESeq2, which tests for genes that are differentially expressed at any of the five time points (or RWC) tested. A significance cut-off was selected at a false discovery rate (FDR) less than 0.01. Of the 76,868 total gene clusters: 19,179 (25%) were removed via the DESeq2 independent filtering step due to having read counts too low to reliably determine significance; 856 (1%) genes were removed entirely as outliers (FDR = NA); and a total of 30,478 (40%) were found to be differentially expressed (DE) at an FDR <= 0.01 during desiccation. Given the high number of DE genes identified by this analysis, a secondary filter based on absolute expression levels (genes expressed at a robust level) and fold change (genes that show a large change in expression during desiccation) was also used. Because raw counts are not comparable across genes of differing lengths, expression levels were first normalised by gene length by calculating the FPKM of each gene cluster. An FPKM of 1 was chosen as an expression cut-off, and genes with maximum expression below this value were removed. A fold-change cut-off of 2 was also implemented, so that all remaining genes showed an expression change of at least two-fold between their minimum and maximum expression levels. This reduced our dataset to 18,737 gene clusters of interest – still a sizeable pool of genes to investigate.

3.3.3.6 Validation of RNA-Seq gene expression profiles

The expression of a subset of *X. humilis* genes in response to desiccation was previously studied via microarray (Collett et al., 2004b; Shen, 2014). In the study performed by Shen, the change in abundance of 1,680 *X. humilis* transcripts was investigated in leaves at six RWCs: 100%, 80%, 60%, 40%, 20% and 5%. Of these genes, 1,361 (81%) were differentially expressed in desiccating *X. humilis* leaves.

A



B

Microarray RWC (%)

RNA-Seq RWC (%)	100	80	60	40	20	5
100	**0.130**	0.165	0.262	0.301	0.283	0.296
80	0.179	**0.166**	0.198	0.282	0.282	0.297
60	**0.165**	0.174	0.248	0.307	0.297	0.310
40	0.332	0.227	0.150	**0.138**	0.153	0.162
5	0.316	0.264	0.239	0.173	**0.119**	0.150

Figure 3.11. Correlation of changes in transcript abundance of a subset *X. humilis* transcripts across a previous microarray and this RNA-Seq study. A) Graph of correlation values for a set of 1,337 *X. humilis* transcripts at each RWC between the Shen, 2014 microarray study and this RNA-Seq study (red line), compared to a null correlation determined by resampling of the expression data for each gene compared to every other gene at each RWC (blue line). The blue shaded portion corresponds to the standard deviation of the null correlation values at each RWC, and thus represents the variance in expression values of the underlying dataset. B) Correlation of gene expression at each RWC compared to all other RWCs between the microarray and RNA-Seq studies. Values which correspond to the highest correlation between the two techniques at a specific RWC are bolded and underlined. Gene expression at 60% RWC was not positively correlated between the two samples.

The 1860 microarray cDNA sequences were compared by reciprocal BLASTN to the assembled *X. humilis* transcriptome. This resulted in the confident identification of 1,337 sequences that could be assumed to correspond to the same transcript in both datasets. In

order to compare expression levels between the two datasets, the microarray expression values were converted from a logarithmic scale and both datasets were normalised to the maximum expression of each transcript. The expression of all transcripts was compared at each RWC across the two normalised datasets by least-squares regression analysis, resulting in the calculation of correlation coefficients corresponding to each RWC (Fig. 3.11A, red line). Additionally, a value for a null correlation was determined by resampling analysis of the data for each RWC (Fig. 3.11A, blue line). A value for inter-RWC expression correlation was also calculated (Fig. 3.11B).

Gene expression at four of the five measured RWCs was positively correlated between the two datasets (100%, 80%, 40% and 5%; Fig. 3.11A). Surprisingly, there was no correlation between gene expression at 60% RWC in the microarray and RNA-Seq experiments (Fig. 3.11A). Gene expression at 60% RWC as measured by RNA-Seq was most closely correlated with that at 100% RWC (0.165) and 80% RWC (0.174) in the Shen microarray study (Fig. 3.11B). In contrast, gene expression at 60% RWC in the microarray study was most closely correlated with the expression at 40% RWC in this study (Fig. 3.11B). Interestingly, gene expression at 5% RWC as measured by RNA-Seq was most similar to gene expression at 20% RWC in the Shen microarray (Fig. 3.11B).

To investigate the discrepancies between the measured expression values by the two quantification technologies, the 1,337 transcripts were grouped based on a simple representation of their expression profiles. Transcripts were considered to have "high" expression when their normalised expression was greater than 0.5, and "low" expression when it was less than or equal to 0.5. Transcripts were then grouped into sets based on their expression at 100% and 5% RWC (four possible expression patterns: High-High, High-Low,

Low-High and Low-Low) (Table 3.3A), or by their expression at 100%, 60% and 5% RWC (eight possible expression patterns) (Table 3.3B) in both the microarray and RNA-Seq datasets. The proportion of all 1,337 transcripts that fell into each expression profile bin was calculated, in an attempt to reveal distinct patterns of transcript expression between the two datasets.

The largest proportion of transcripts in both datasets had an expression profile of "HL" - representing genes that were higher at 100% RWC than at 5% RWC (Table 3.3A). Transcripts with the opposite pattern of expression ("LH") were similarly high in both datasets (27.7% and 23.9%). However, the number of transcripts that showed consistent expression at both 100% and 5% RWCs ("HH" and "LL", respectively) differed between the two technologies. Over 20% of transcripts from the microarray dataset had "low" expression at both 100% and 5% RWC, compared to only 7.4% in the RNA-Seq study. In contrast, 27% of transcripts had "high" expression at 100% and 5% in RNA-Seq study, compared to only 10% in the Shen microarray study (Table 3.3B). Genes that show robust changes in transcript abundance between hydrated and desiccated tissue could thus make up a large part of the observed correlation between these datasets (Fig. 3.11A), as opposed to genes that show changes in expression between 100% and 5% RWC.

To investigate this further, a similar analysis was performed comparing transcript expression between 100%, 60% and 5%, consisting of eight possible expression paths (Table 3.3B). The most obvious differences between these two datasets consists of transcripts in the "HHH", "HHL" and "HLL" expression bins. A large number of transcripts were measured to be in the "HHH" bin in the RNA-Seq study (25.4%), compared to only 2.8% in the microarray study (Table 3.3B). A similar trend is seen in the "HHL" bin, which has 34% of

all transcripts in the RNA-Seq study, and only 11.1% in the microarray study. The opposite is seen in the "HLL" bin, containing 34.5% of transcripts as measured by microarray, and only 4% in the RNA-Seq study (Table 3.3B). A related trend is observed between the "LLH" and "LHH" bins in the two studies. Although a large number of genes share the "LHH" bin - suggesting that both studies agreed on the expression of genes that were activated during later desiccation - many of the genes with "LLH" expression were measured as "LHH" in the Shen microarray study (Table 3.3B).

Because the "HHL" vs. "HLL" and "LLH" vs. "LHH" bins alone account for such a large proportion of the difference between the two datasets (Table 3.3B), despite the fact that the "HL" and "LH" bins are the most similar between the two datasets (Table 3.3A), it is consistent with a scenario whereby the gene expression as measured at 60% RWC by the two studies in fact corresponds to expression at two different RWCs (Fig. 3.12A&B). Specifically, if the measured RWC at 60% RWC in this study is at a higher RWC than that of the Shen microarray study, it could result in the observed differences in measured expression of genes with these specific expression profiles (Fig. 3.12A&B).

A

RNA-Seq \ Microarray	HH	HL	LH	LL	
HH	58	185	63	55	27.0
HL	44	326	44	93	37.9
LH	34	71	188	77	27.7
LL	7	22	25	45	7.4
	10.7	45.2	23.9	20.2	

B

RNA-Seq \ Microarray	HHH	HLH	HHL	HLL	LLL	LHL	LLH	LHH	
HHH	15	38	52	123	10	40	32	29	25.4
HLH	1	4	3	7	1	4	1	1	1.6
HHL	12	25	75	213	9	79	17	24	34.0
HLL	2	5	4	34	0	5	1	2	4.0
LLL	1	3	4	6	1	31	10	13	5.2
LHL	0	3	2	10	0	13	0	2	2.2
LLH	6	24	6	51	17	47	99	69	23.9
LHH	1	3	2	12	1	12	9	11	3.8
	2.8	7.9	11.1	34.1	2.9	17.3	12.6	11.3	

Table 3.3. Comparison of the number of genes following specific simple expression patterns across microarray and RNA-Seq datasets. The 1,337 shared transcripts were grouped into bins depending on whether their expression followed one of a number of simple expression patterns. A) Transcripts were grouped based on whether their normalised expression was "high" ("H"; > 0.5) or "low" ("L"; <= 0.5) at 100% and 5% RWC, resulting in four possible expression patterns ("HH", "HL", "LH", "LL"). Values on the end of each row and column correspond to what proportion of all transcripts fell within that expression profile for the RNA-Seq and microarray study, respectively. B) Transcripts were grouped based on whether their normalised expression was high or low at 100%, 60% and 5% RWC, resulting in eight possible expression profiles. Values on the end of each row and column correspond to what proportion of all transcripts fell within that expression profile for the RNA-Seq and microarray study, respectively.

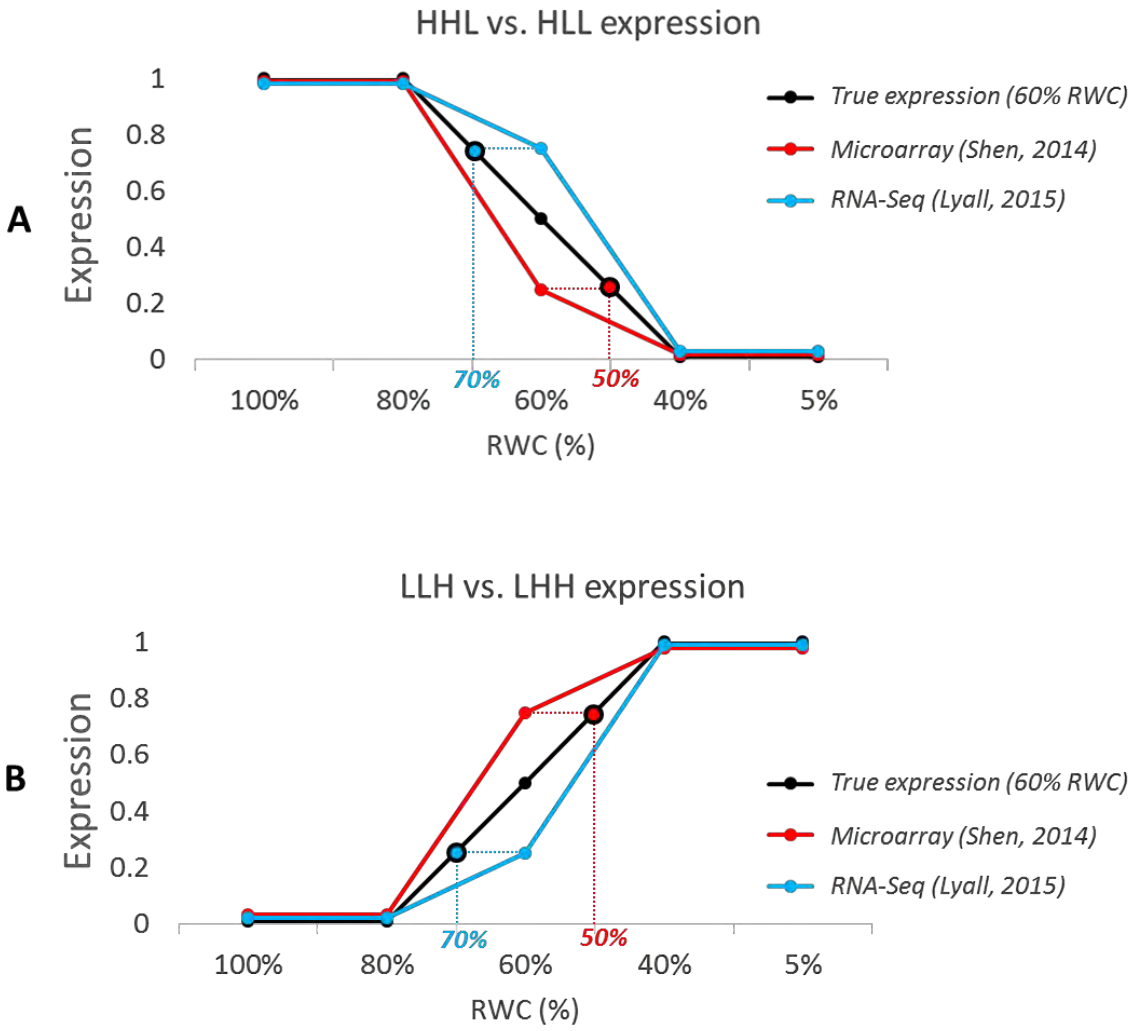


Figure 3.12. Hypothetical scenario in which the 60% RWC samples between the RNA-Seq and microarray studies corresponded to higher and lower actual RWC, respectively. A) The measured expression of down-regulated genes ("HL") is elevated in this RNA-Seq study if measured earlier during dehydration (e.g. 70% RWC, blue dashed line) compared to later dehydration (e.g. 50% RWC, red dashed line) in the microarray study by Shen, resulting in skewed measurements of "HHL" and "HLL" gene expression between the two studies. B) The reverse is true for up-regulated genes ("LH"), where gene expression measured as "LLH" in this study (blue dashed line) could correspond to "LHH" expression measured by microarray at a later RWC (red dashed line).

3.3.3.7 Expression path clustering and GO analysis

For the purposes of gene expression profiling, the 18,737 DE gene clusters were separated based on whether their origin was predicted to be plant (i.e. *X. humilis* derived; 18,534), or fungal (203). The transcript sets were independently grouped into expression paths using k-means clustering within Multi-Experiment Viewer (MeV). *X. humilis* transcripts were grouped into a set of 16 expression paths across the five stages of dehydration (Fig. 3.13).

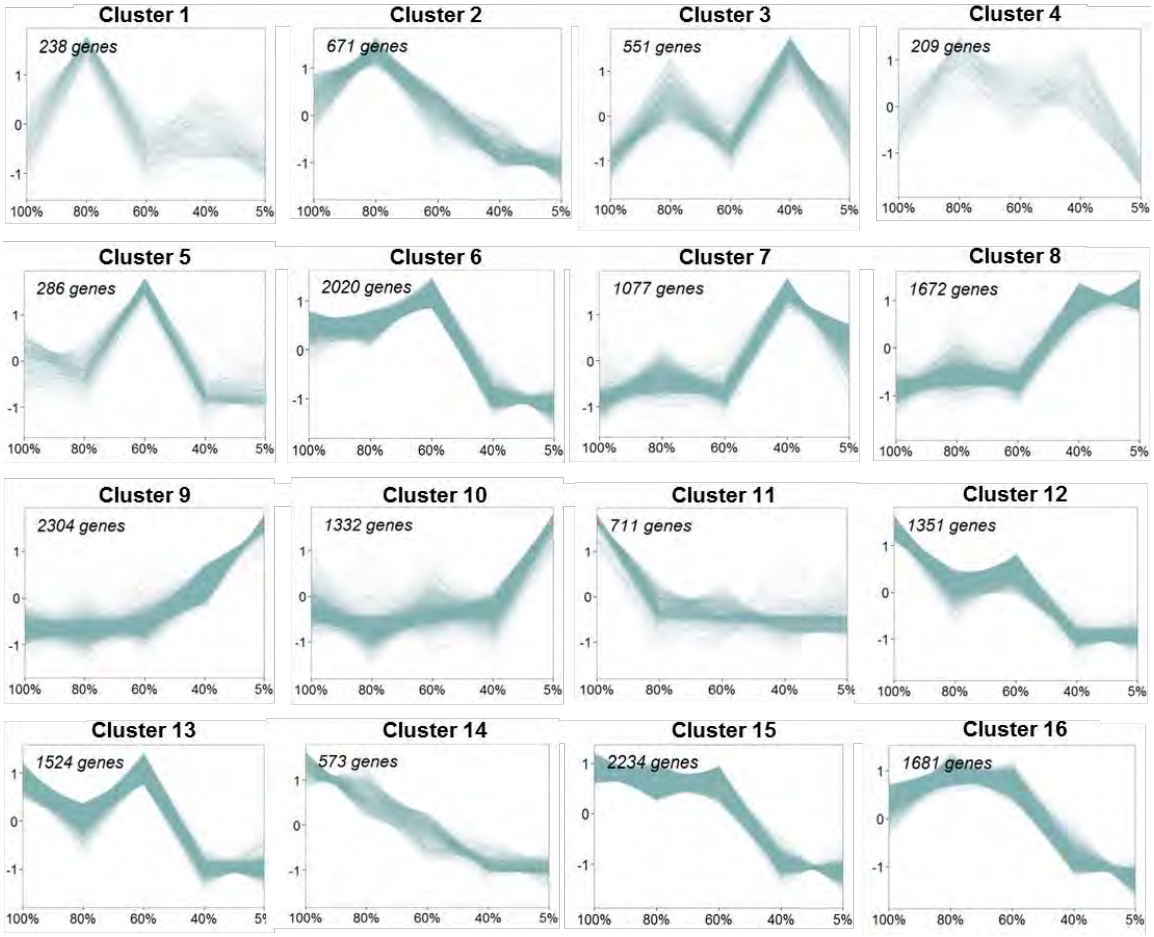


Figure 3.13: *X. humilis* gene expression clusters. K-means clustering of all 18,534 DE *X. humilis* transcripts (FDR < 0.01; FPKM > 1, FC > 2) using normalised rlog expression values.

Furthermore, GO enrichment analysis for Biological Processes of the selected expression clusters was performed using the Cytoscape plugin BINGO, which produces a network of related GO terms (nodes). Nodes with an adjusted p-value less than 0.01 were removed, resulting in the retention of only well-supported sub-networks of GO terms. Due to the large changes in gene expression that occurred during desiccation, many of these networks were far too vast to include in this chapter, and thus only selected sub-networks of interest are given here. Where possible, all nodes within a network are shown; however, superfluous nodes or edges may have been trimmed for the sake of visual clarity. In cases where the network of GO terms was too large to display, the lowest-level GO terms of the network were isolated and grouped based on their position within their local sub-network.

Up-regulated genes

The first two clusters contained genes that were transiently up-regulated at 80% RWC. Both contained few enriched GO networks. Cluster 1 had only one, relating to cell wall biogenesis, whereas Cluster 2 had networks related to ABA catabolism, flowering, skotomorphogenesis (seedling growth without light) and seed development (Fig. 3.14).

In comparison to the first two clusters, Clusters 3 and 4 both consisted of genes that were highly expressed at 80% and also 40% RWC. Despite their similar expression profiles, the GO networks enriched in each cluster were quite different. Cluster 3 was enriched for numerous terms related to abiotic stress response, including response to light, heat, cold, hyperosmolarity, water deprivation and hydrogen peroxide, as well as response to signalling molecules abscisic acid, jasmonic acid and carbohydrate stimulus. A smaller network related to the release of seed dormancy (Fig. 3.14). Cluster 4, which contained fewer genes, had only a single enriched network related to flavonol synthesis (Fig. 3.15).

Cluster 5 contained genes that were transiently up-regulated at 60% RWC, before returning to the levels seen in hydrated tissue. Although this cluster consisted of a comparatively small number of genes (286), it contained a large and diverse assortment of enriched GO networks. The two largest related to response to various biotic, abiotic and signalling factors: response to triglyceride, ABA, jasmonic and salicylic acids, ethylene and auxin; response to wounding, herbicide, cold, salt and water deprivation/desiccation, and various defence responses, including response to bacteria, nematodes, insects and fungi. The second large GO network related to growth and development, particularly of reproductive structures: pollen, seed and fruit development, as well as root growth. A third network consisted of GO terms to do with cell wall biogenesis and thickening, and synthesis of cell wall-related polysaccharides, such as xylans. Additionally, a small but significantly enriched network was present that related to nucleosome assembly (Fig. 3.16).

Cluster 6 consisted of genes that were also slightly up-regulated at 60% RWC, but heavily down-regulated thereafter. The vast majority of these enriched GO terms were concerned with regulation of gene expression, such as histone lysine methylation, post-transcriptional gene regulation (mRNA splicing, ta-siRNA and miRNA interference) and translation (rRNA and tRNA maturation). GO terms related to DNA replication and DNA repair were also enriched (Fig. 3.17).

Clusters 7 and 8 correspond to genes that are transiently or permanently up-regulated at 40% RWC respectively. Both clusters produced networks that shared a large number of GO terms, including large networks relating to stress response and hormone signalling, chlorophyll degradation and the release of seed dormancy (Fig. 3.18).

Cluster 9 consisted of genes that were progressively induced between 40% and 5% RWC, and contained the largest number of enriched GO terms of all up-regulated clusters. The largest group of GO terms pertained to chemical biosynthesis processes, including synthesis of carbohydrates, lipids, hormones and pigments (Fig. 3.19A), and the second largest contained terms related to stress response, both biotic and abiotic (Fig. 3.19B). GO terms associated with signalling pathways, particularly sugars, jasmonic acid, ethylene and abscisic acid were also enriched (Fig. 3.19C). Lastly, GO terms related to development of reproductive structures and cell division/death (Fig. 3.19D), protein ubiquitination/degradation (Fig. 3.19E) and regulation of transcription (Fig. 3.19F) were also present.

The last of the up-regulated genes were present in Cluster 10, which contained genes that were induced specifically at 5% RWC. Despite the large number of genes (1332), comparatively few GO terms were enriched in this cluster. The largest network related to lipid biosynthesis, particularly triterpenoids, galactolipids and jasmonic acid. Two other networks contained terms associated with ER to Golgi vesicle-mediated transport and the targeting of proteins to peroxisomes, as well as catabolism of misfolded proteins (Fig. 3.20).

Down-regulated genes

Clusters 11-16 consisted of genes that were down-regulated in desiccated versus hydrated tissues. Despite containing 22% fewer total genes, these clusters contained nearly 15% more enriched GO terms than the previous clusters (1-10), resulting in large, functionally diverse networks.

Clusters 11, 12, 13 and 14 contained genes down-regulated to some degree at 80% RWC. Despite the large differences in overall expression pattern between these clusters, they shared numerous terms within their respective GO networks. These enriched terms could be loosely grouped into four processes: 1) response to hormones/endogenous signals, particularly jasmonic acid, auxin, ABA, GA and brassinosteroids (Fig. 3.21), 2) response to exogenous signals and stress, such as bacteria, desiccation or high light (Fig. 3.22), 3) chemical biosynthesis processes, particularly of flavonols, chlorophyll and carbohydrates involved in cell wall biogenesis (Fig. 3.23), 4) growth and development, particularly of the roots and reproductive structures, such as seeds and flowers (Fig. 3.24).

Clusters 15 and 16 had very similar expression profiles, and both contained genes that were strongly down-regulated after 60% RWC. Unlike Clusters 11-14, GO terms related to stress response, endogenous signalling and synthesis were not enriched. A network containing GO terms associated with growth and development, predominantly those related to flowers/seeds, leaves and leaf vasculature was found in Cluster 15 (Fig. 3.24). Cluster 15 also contained a network with GO terms associated with photosynthesis and photosynthetic apparatus, many of which were shared with a similar, smaller network found in Cluster 12 (Fig. 3.25A). Lastly, networks related to the regulation of gene expression were found in Cluster 13 (H3K9 methylation, mRNA stability and splicing), Cluster 15 (RNA interference by miRNA and ta-siRNA) and Cluster 16 (nonsense-mediated decay, mRNA splicing) (Fig. 3.25B).

Seed genes

GO terms associated with seed development, particularly the processes of seed dormancy and release of dormancy, were over-represented in multiple clusters of both up- and down-

regulated genes. A summary of the top protein hits of these *X. humilis* genes associated with enriched seed-related GO terms, based on BLASTX analysis against known proteins in the Swissprot database, is shown in Supplementary Table 3.2. The presence of many seed-related genes, such as those with similarity to the antioxidant *1-CYS PERODOXIN* (*PER1*), oil-body related genes such as oleosins and caleosins (*PEROXYGENASE 1*/*ATS1* and other peroxygenases), and desiccation-induced LEAs, were to be expected based on previous literature on *X. humilis* and resurrection plants. Genes predicted to play a general role in ABA biosynthesis and signalling but that are not necessarily seed specific, such as transcripts with similarity to *9-CIS-EPOXYCARETENOID DIOXYGENASE 1*, *ABSCISIC ACID 8'-HYDROXYLASE 1* and PP2C protein phosphatases were also present across numerous clusters, as well as various other enzymes or transcription factors involved in general development or signalling.

Fungal genes

Despite the large number of predicted fungal transcripts compared to total transcripts (17%, Fig. 3.8C), only 203 predicted fungal gene clusters were determined to be DE using our cut-offs. Due to their small number they could be clustered into only four expression profiles. For the most part, fungal genes that were detected were upregulated in response to dehydration of *X. humilis*. These clusters are available in the supplementary data (Fig. S3.2).

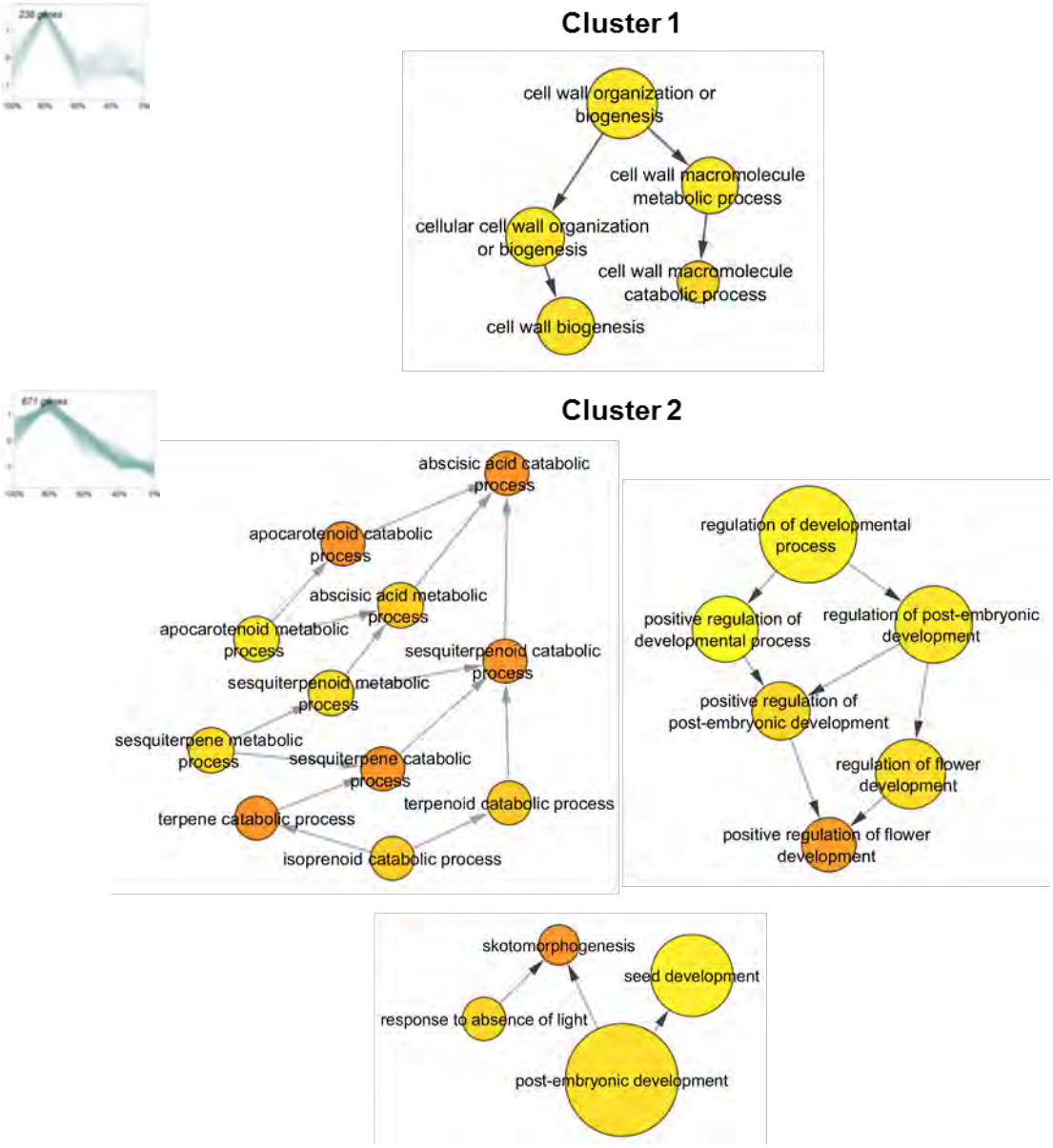
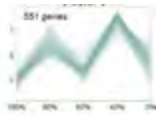
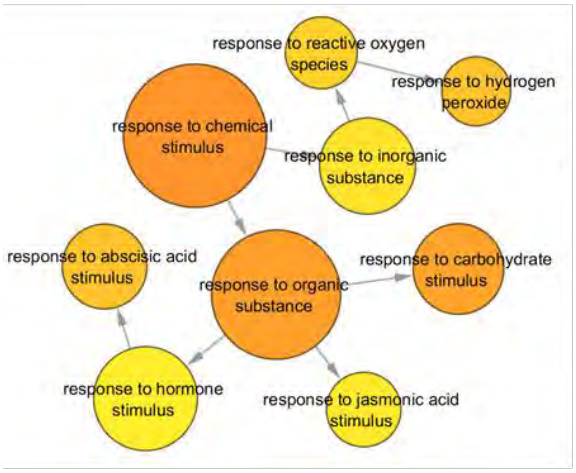


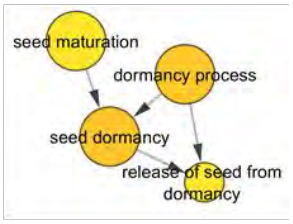
Figure 3.14. Enriched GO term networks of genes transiently up-regulated at 80% RWC. Networks of significantly enriched GO terms (FDR < 0.01) as determined using BINGO, visualised using Cytoscape. Node size correlates with the number of input genes containing that term; node colour

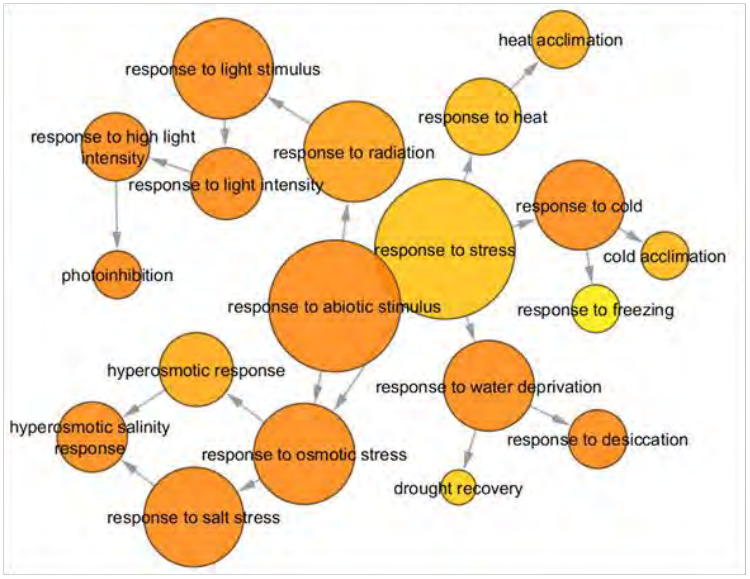
represents the level of significance (darker nodes are more significant).

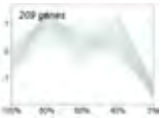


Cluster 3









Cluster 4

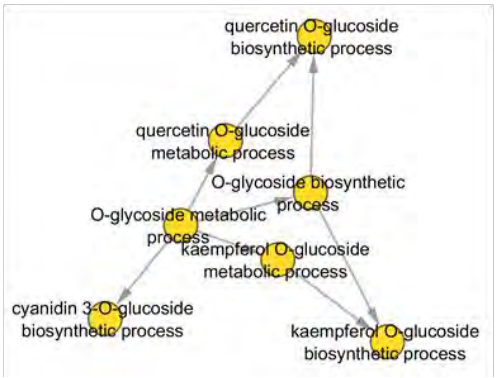


Figure 3.15. Enriched GO term networks of genes transiently up-regulated at 80% and 40% RWC. Networks of significantly enriched GO terms (FDR < 0.01) as determined using BINGO, visualised using Cytoscape. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

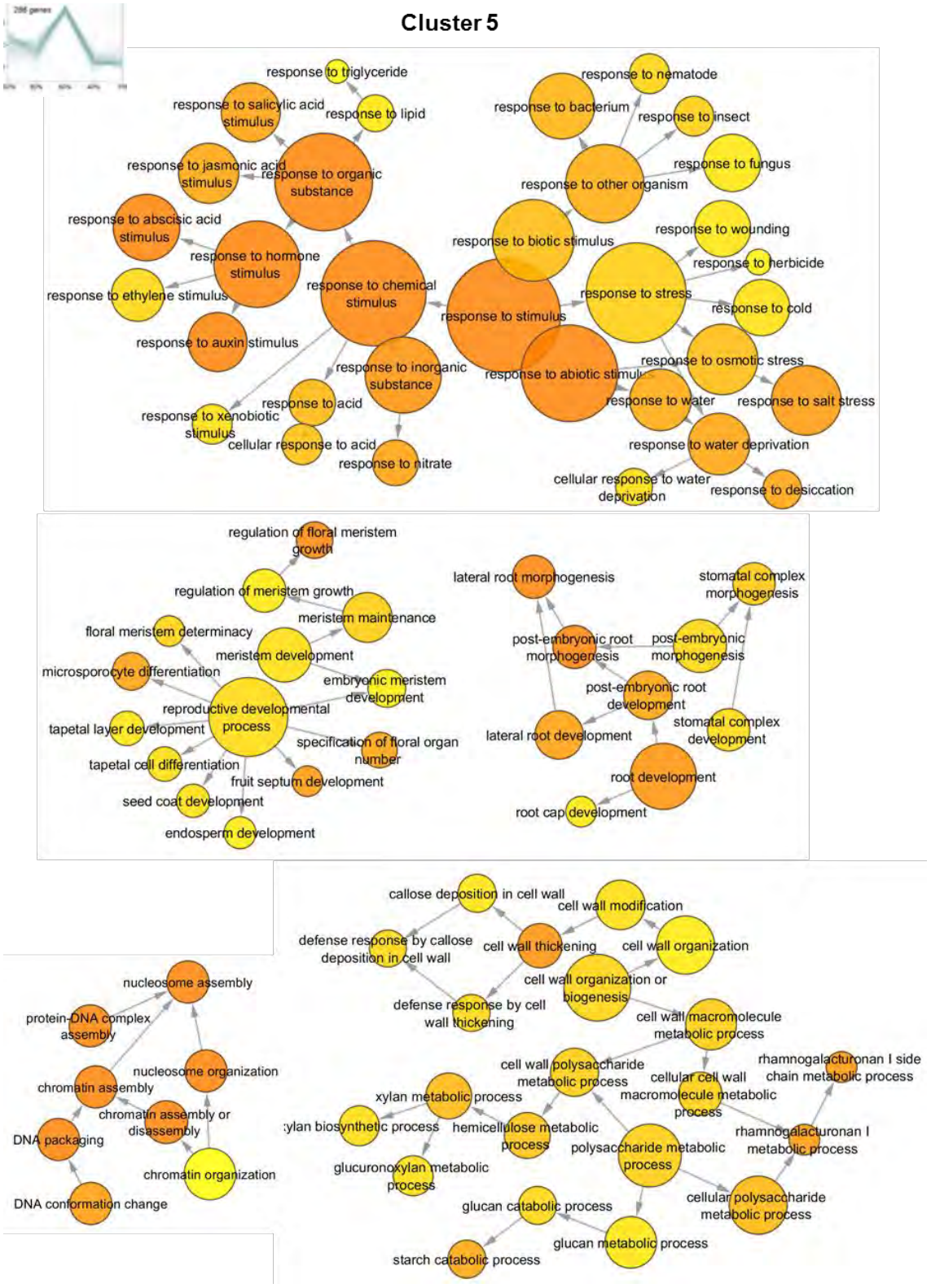


Figure 3.16. Enriched GO term networks of genes up-regulated at 60% RWC. Networks of significantly enriched GO terms (FDR < 0.01) as determined using BINGO, visualised using Cytoscape. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

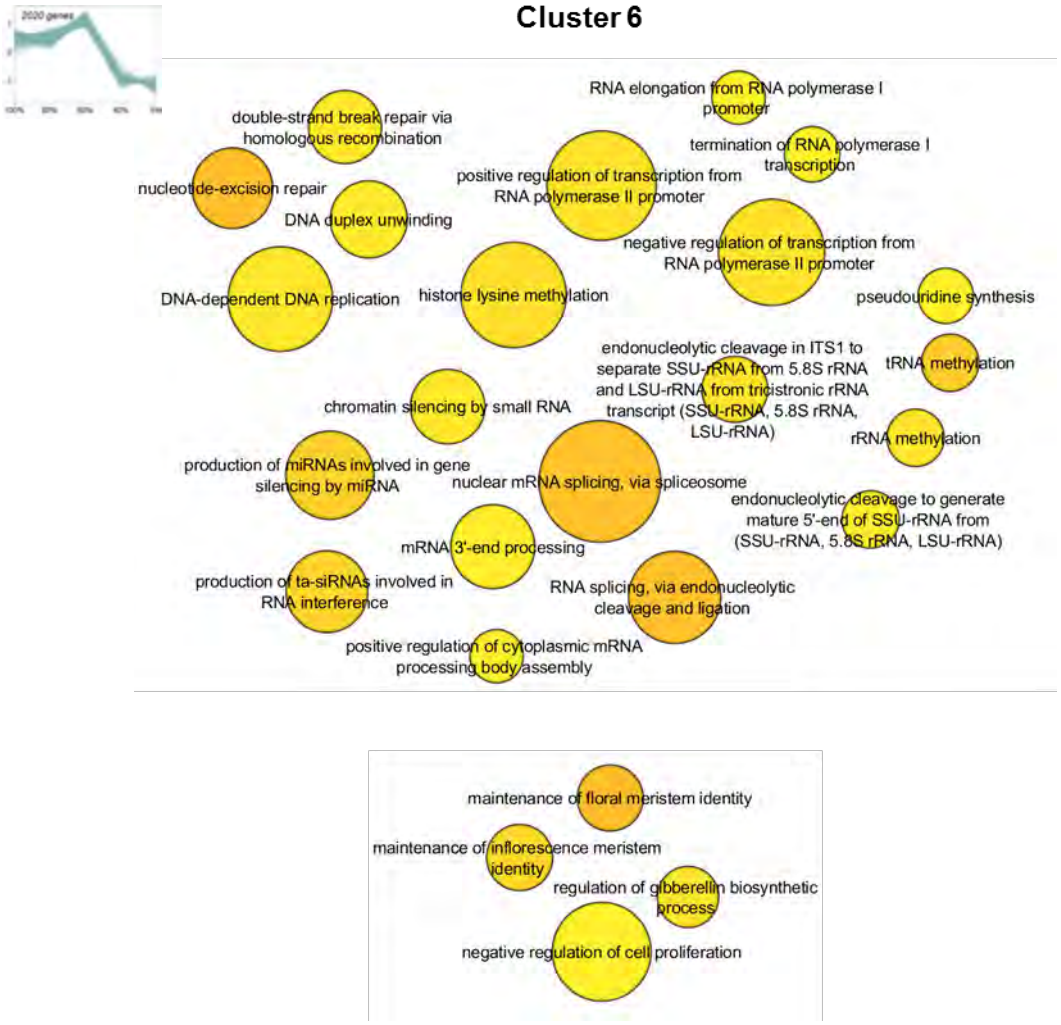
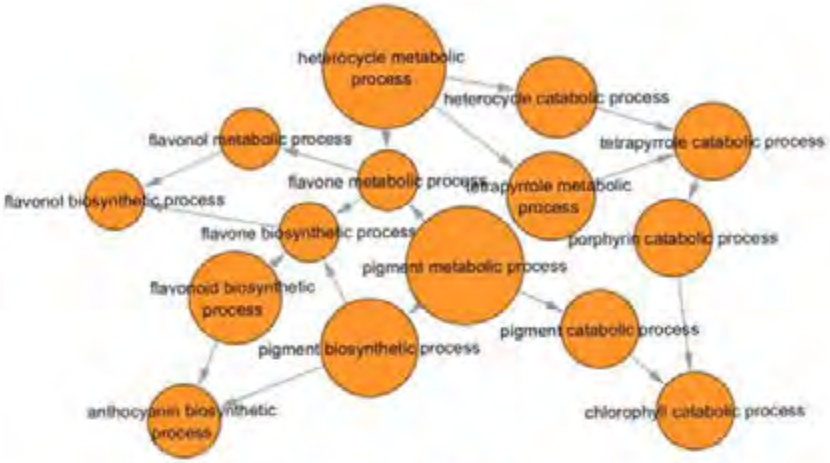
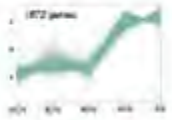


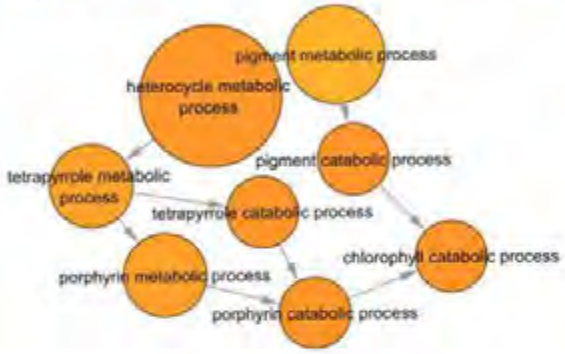
Figure 3.17. Lowest-level terms within the enriched GO networks of Cluster 6. Networks of significantly enriched GO terms (FDR < 0.01) as determined using BINGO, visualised using Cytoscape. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

Cluster 7



Cluster 8





Cluster 7

Cluster 8

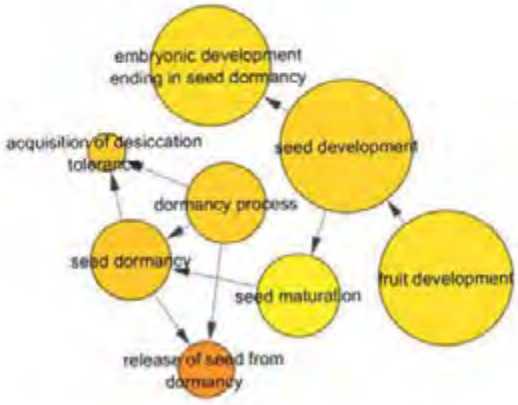
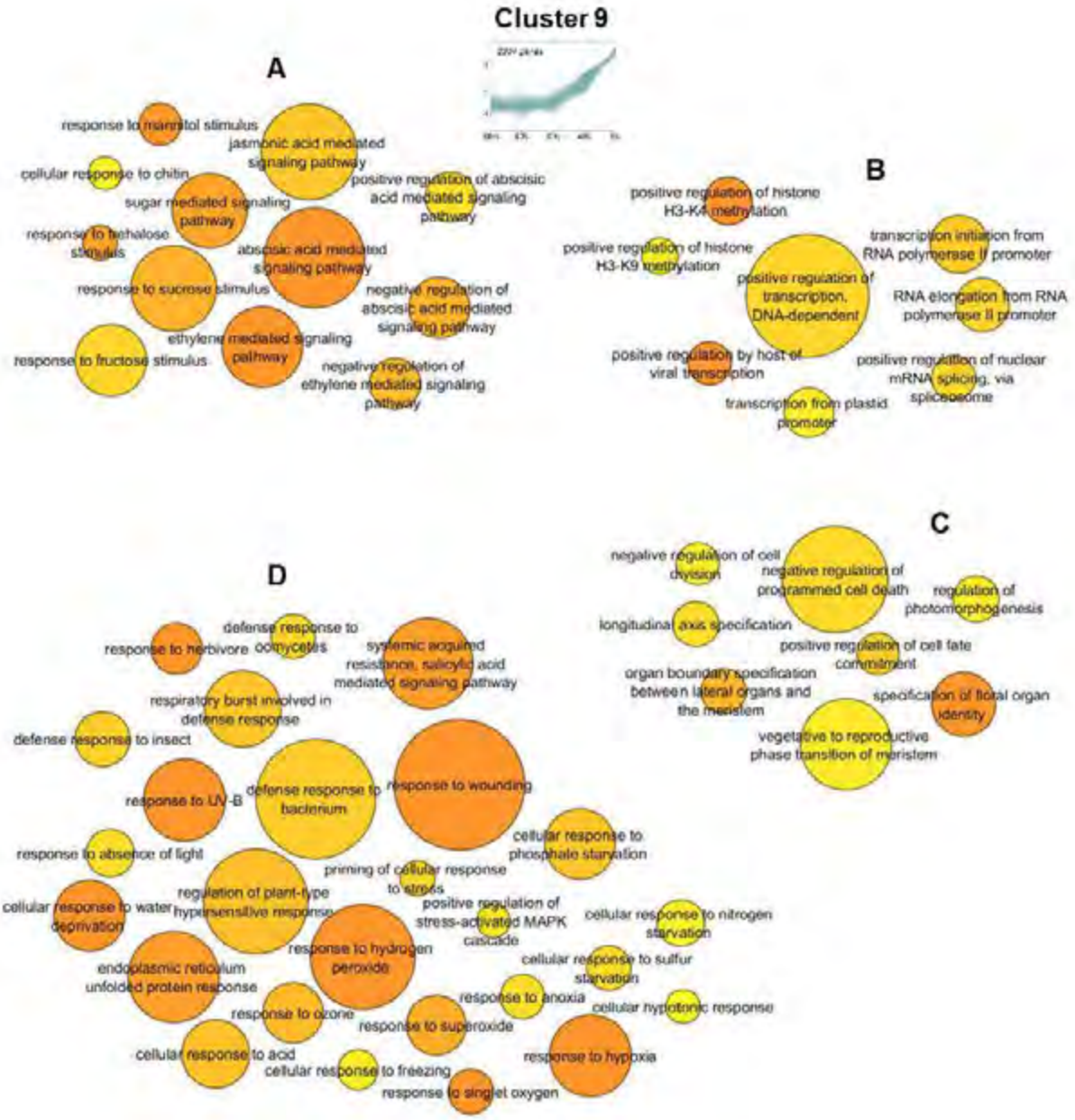


Figure 3.18. Enriched GO networks of genes up-regulated at 40% RWC. Networks of significantly enriched GO terms (FDR < 0.01) as determined using BINGO, visualised using Cytoscape. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).


Cluster 9
A
response to mannitol stimulus
jasmonic acid mediated signaling pathway
cellular response to chitin
positive regulation of abscisic acid mediated signaling pathway
sugar mediated signaling pathway
response to trehalose stimulus
abscisic acid mediated signaling pathway
response to sucrose stimulus
negative regulation of abscisic acid mediated signaling pathway
ethylene mediated signaling pathway
response to fructose stimulus
negative regulation of ethylene mediated signaling pathway
B
positive regulation of histone H3-K4 methylation
transcription initiation from RNA polymerase II promoter
positive regulation of histone H3-K9 methylation
positive regulation of transcription, DNA-dependent
RNA elongation from RNA polymerase II promoter
positive regulation by host of viral transcription
positive regulation of nuclear mRNA splicing, via spliceosome
transcription from plastid promoter
C
negative regulation of cell division
negative regulation of programmed cell death
regulation of photomorphogenesis
longitudinal axis specification
positive regulation of cell fate commitment
organ boundary specification between lateral organs and the meristem
specification of floral organ identity
vegetative to reproductive phase transition of meristem
D
defense response to oomycetes
response to herbivore
systemic acquired resistance, salicylic acid mediated signaling pathway
respiratory burst involved in defense response
defense response to insect
response to UV-B
defense response to bacterium
response to wounding
cellular response to phosphate starvation
response to absence of light
priming of cellular response to stress
regulation of plant-type hypersensitive response
cellular response to water deprivation
positive regulation of stress-activated MAPK cascade
cellular response to nitrogen starvation
response to hydrogen peroxide
endoplasmic reticulum unfolded protein response
cellular response to sulfur starvation
response to anoxia
response to ozone
cellular hypotonic response
response to superoxide
cellular response to acid
response to hypoxia
cellular response to freezing
response to singlet oxygen

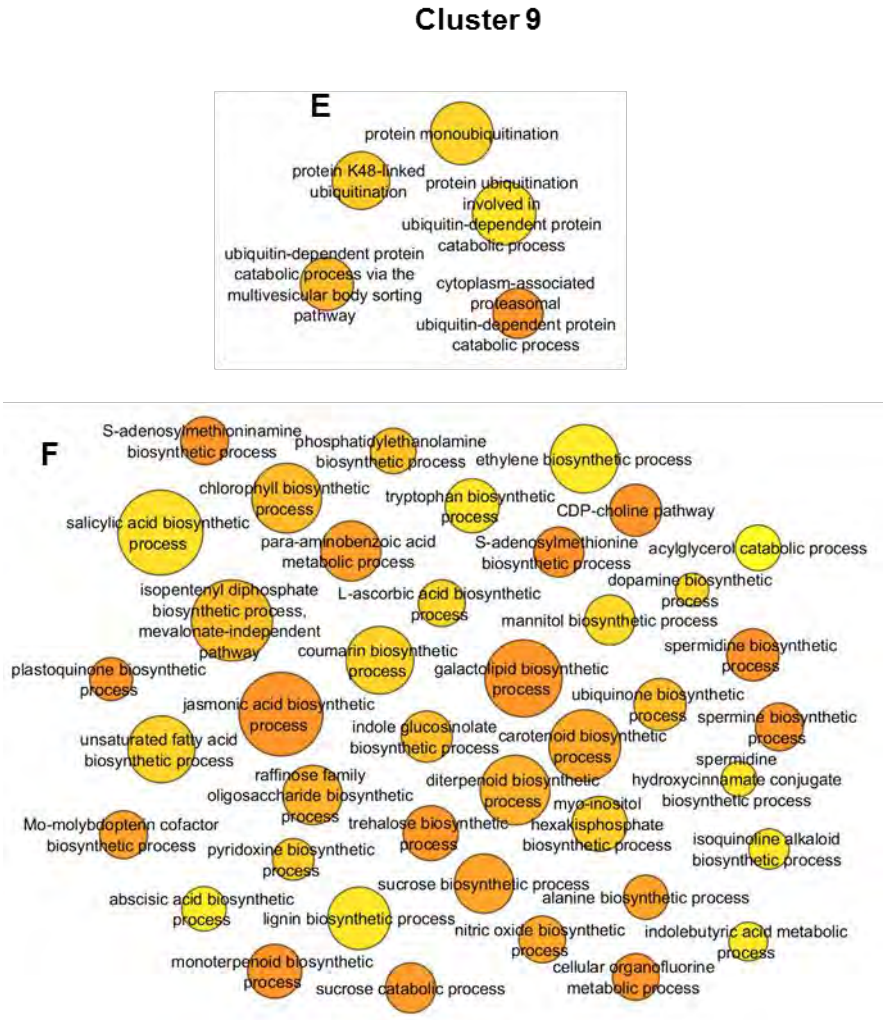


Figure 3.19. Lowest-level terms of the enriched GO network of genes up-regulated between 40% and 5% RWC. Nodes containing no outgoing edges from were isolated from sub-networks of significantly enriched GO terms (FDR < 0.01) determined using the Cytoscape plugin, BINGO. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

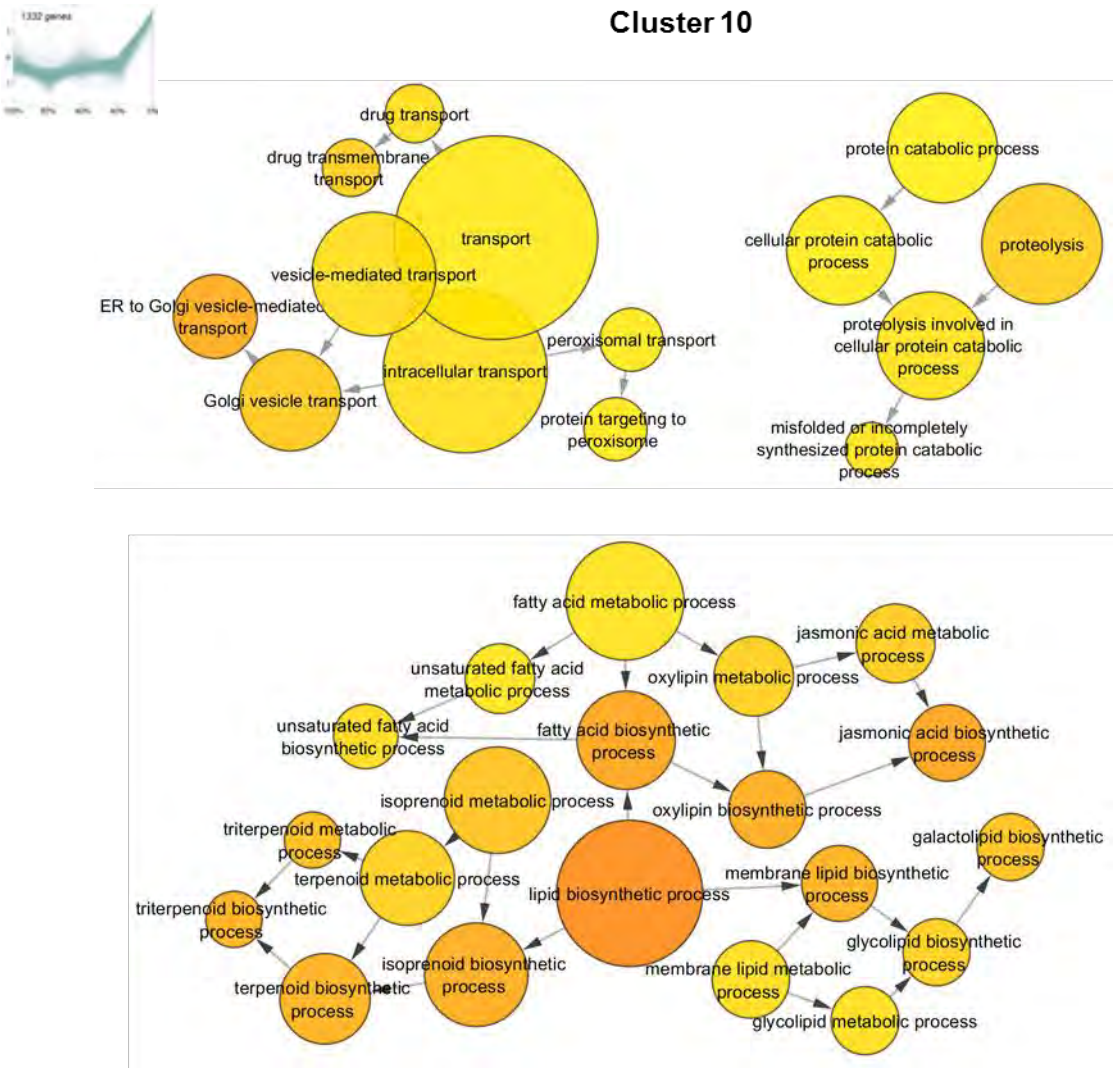


Figure 3.20. Enriched GO networks of genes up-regulated at 5% RWC. Networks of significantly enriched GO terms (FDR < 0.01) as determined using BINGO, visualised using Cytoscape. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

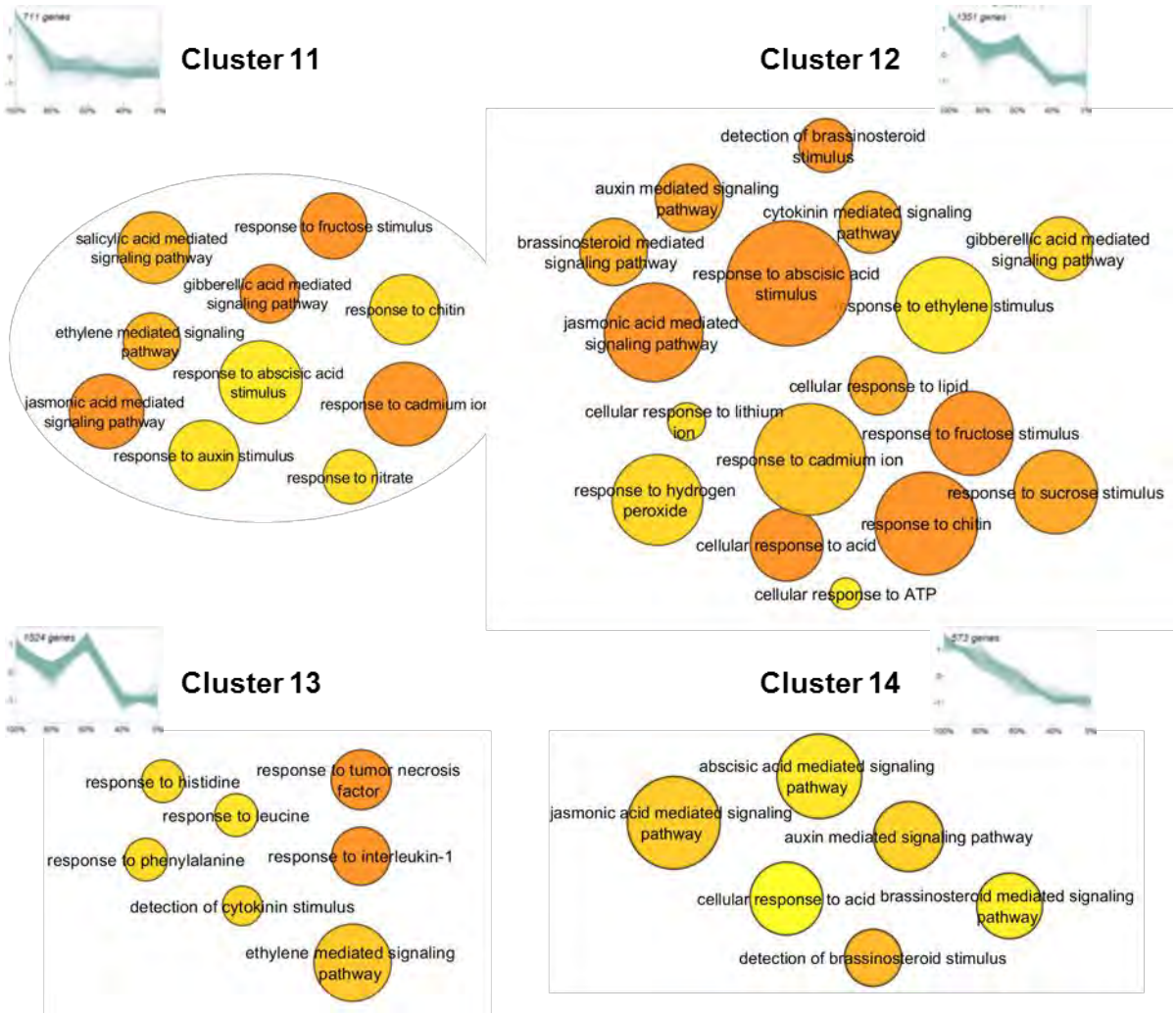


Figure 3.21. Enriched GO terms related to endogenous signalling in down-regulated gene clusters. Nodes containing no outgoing edges were isolated from sub-networks of significantly enriched GO terms (FDR < 0.01) determined using the Cytoscape plugin, BINGO. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

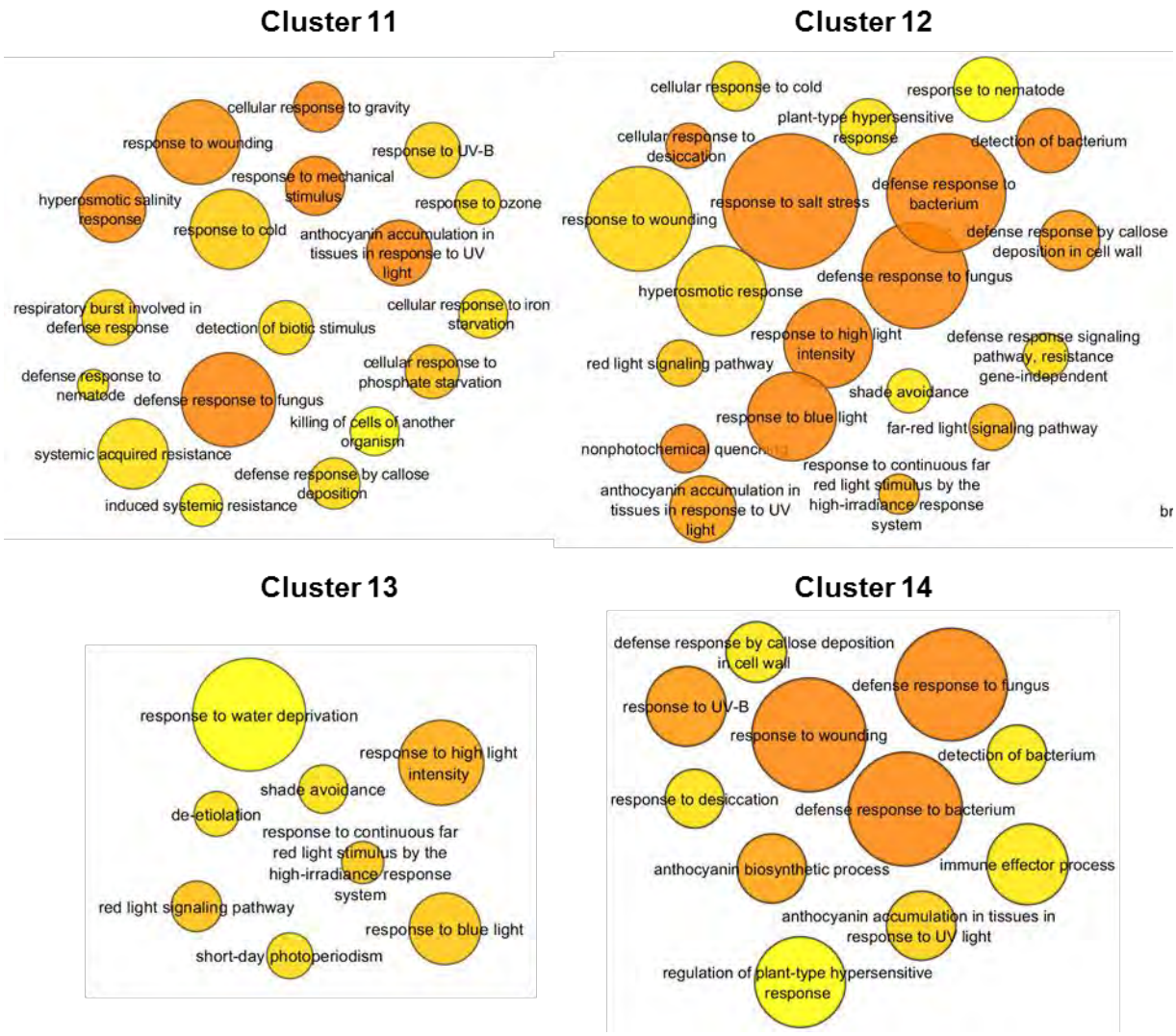


Figure 3.22. Enriched GO terms related to abiotic and biotic stress responses in down-regulated gene clusters. Nodes containing no outgoing edges were isolated from sub-networks of significantly enriched GO terms (FDR < 0.01) determined using the Cytoscape plugin, BINGO. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

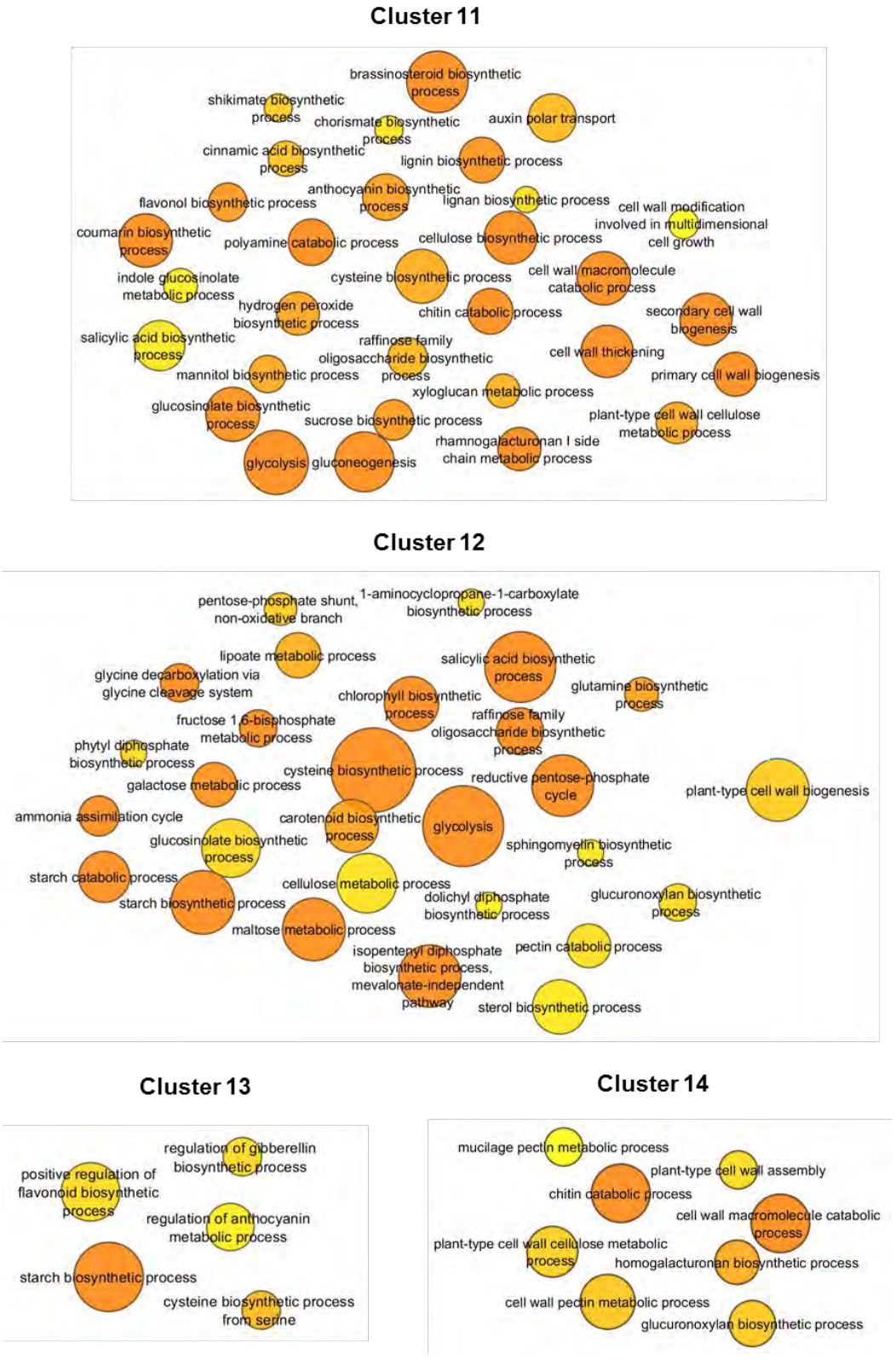


Figure 3.23. Enriched GO terms related to chemical biosynthesis in down-regulated gene clusters. Nodes containing no outgoing edges were isolated from sub-networks of significantly enriched GO terms (FDR < 0.01) determined using the Cytoscape plugin, BINGO. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

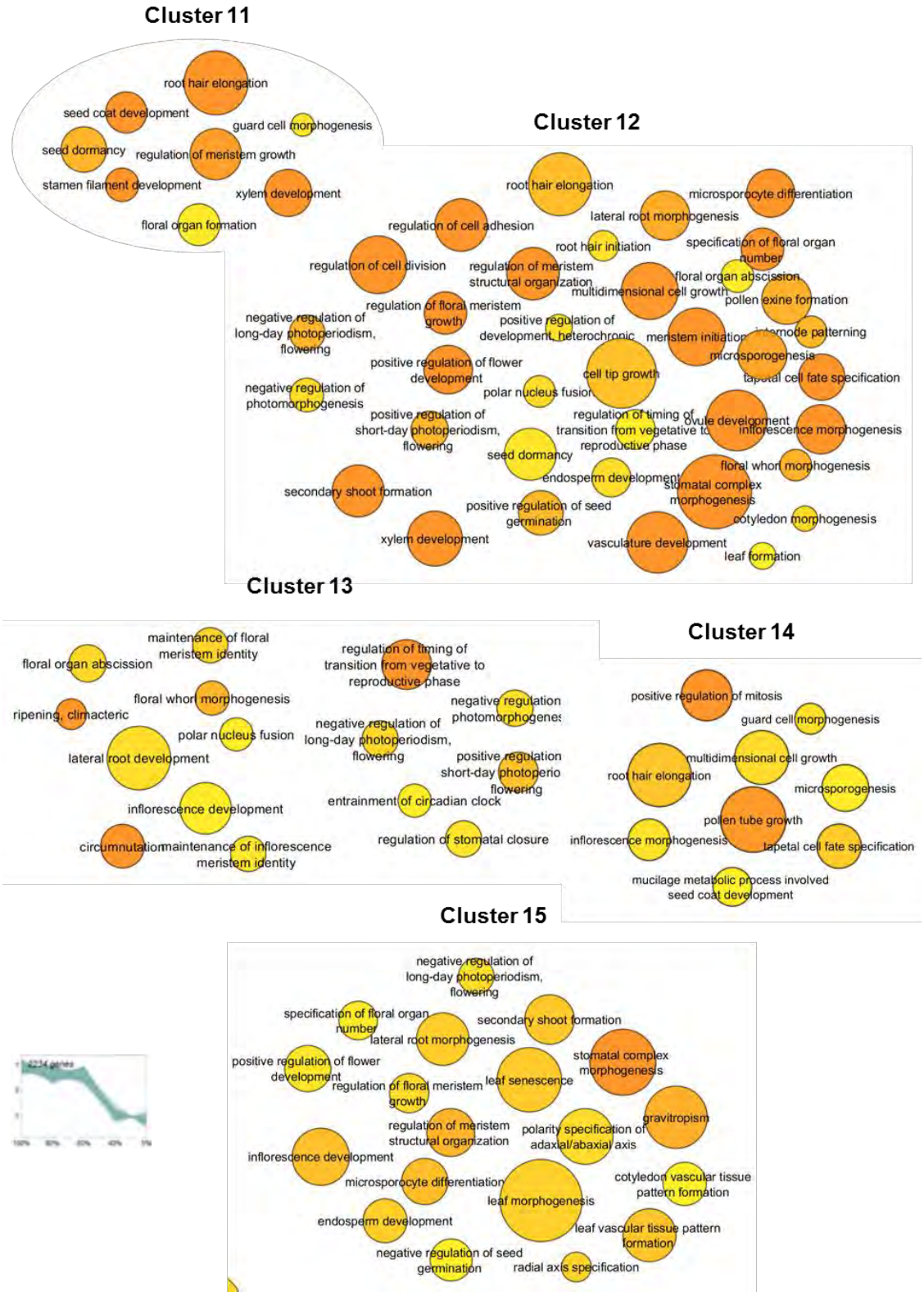


Figure 3.24. Enriched GO terms related to growth and development in down-regulated gene clusters. Nodes containing no outgoing edges were isolated from sub-networks of significantly enriched GO terms (FDR < 0.01) determined using the Cytoscape plugin, BINGO. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

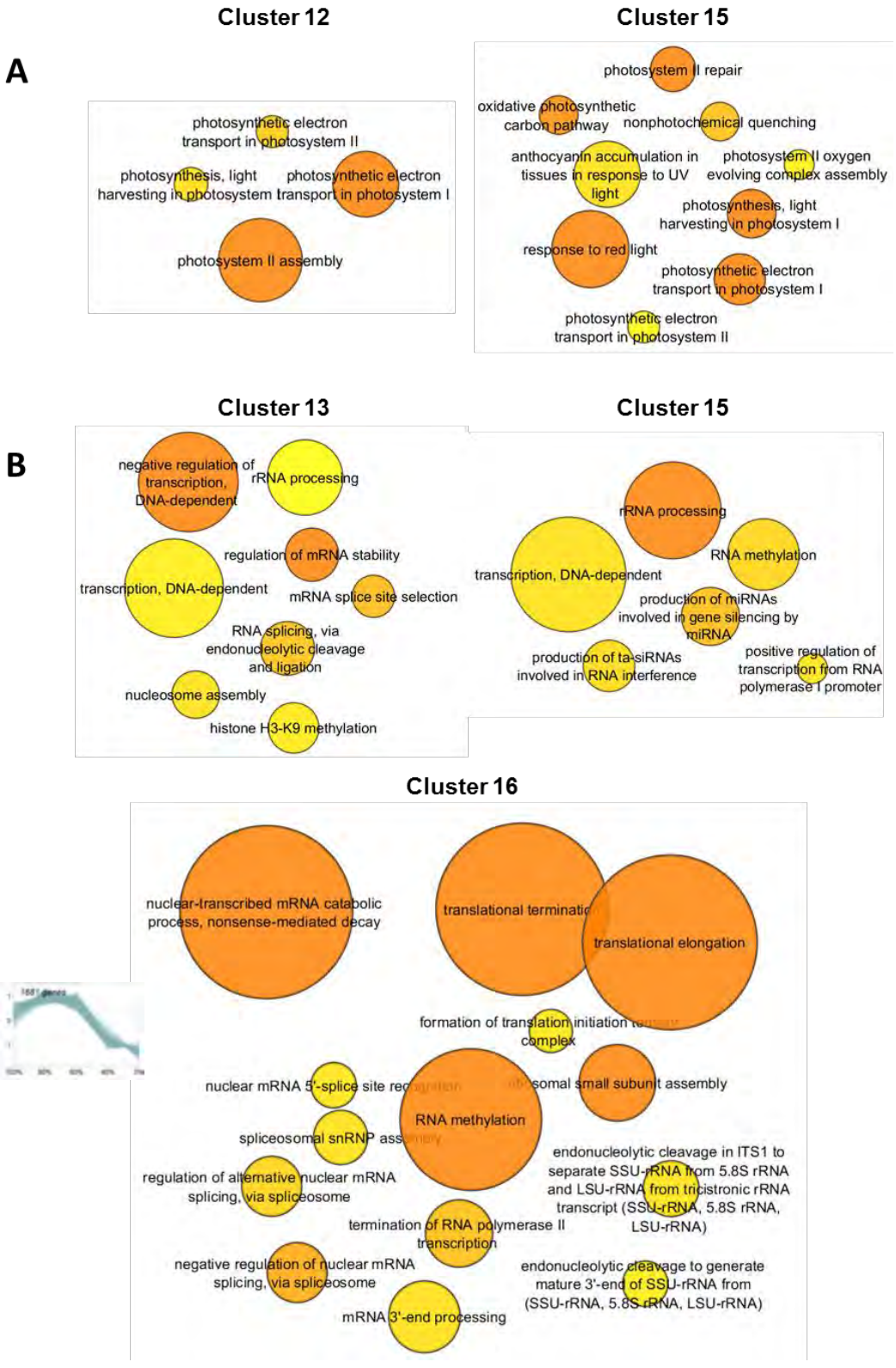


Figure 3.25. Other GO terms enriched in down-regulated gene clusters. Groups of GO terms related to photosynthesis (A) and gene regulation (B) in Clusters 11-16. Nodes containing no outgoing edges from were isolated from sub-networks of significantly enriched GO terms (FDR < 0.01) determined using the Cytoscape plugin, BINGO. Node size correlates with the number of input genes containing that term; node colour represents the level of significance (darker nodes are more significant).

3.3.4 Gene expression analysis of desiccating X. humilis leaves

Using the transcriptome annotation and differential expression data, I investigated the changes in expression of key genes or gene families in the desiccating leaves of *X. humilis*.

3.3.4.1 LEA genes are upregulated in response to dehydration in X. humilis

Putative LEA transcripts were some of the most highly upregulated genes in response to dehydration in *X. humilis*. Nearly 150 gene clusters were predicted through sequence similarity and the presence of known LEA Pfam domains to be LEA or LEA-like gene transcripts (Fig. 3.26A), and many of the top most highly-expressed genes at 5% RWC were transcripts that were annotated as LEA protein-encoding sequences (Fig. 3.26B). The total number of predicted LEA transcripts found is far in excess of the number of LEA genes from each family currently identified in the genome of *A. thaliana* and other DS plant species, let alone expressed at any one time (Fig. 3.26C). In order to ascertain whether the observed increase in LEA-encoding transcripts was due to an expansion in the number of LEA genes in *X. humilis* or due to extensive splicing of individual LEA genes, each transcript was aligned to the draft *X. humilis* genome (Stephen Schlebusch, personal correspondence) using BLASTN. Transcripts that shared sequence from the same genomic scaffold coordinates were considered to be isoforms derived from the same gene for the sake of this study. This resulted in the collapsing of the 150 LEA gene clusters into 124 putative *X. humilis* LEA genes (Fig. S3.4) a still greater number than that found in *A. thaliana* and other DS species (Fig. 3.26C).

LEA protein transcripts were identified from all known LEA protein families (Fig. 3.26A). Although the majority of identified *X. humilis* LEA protein transcripts were up-regulated in

response to desiccation, a number were either down-regulated or not differentially expressed across the selected RWCs - most of which could be found in the LEA-8 family (Fig. 3.26A; Fig. S3.3). The degree to which LEA transcripts are generally upregulated in desiccating *X. humilis* tissues is highlighted when analysing the reads that map to these contigs. In total, over 15% of sequenced reads in the 40% and 5% RWC libraries were derived from LEA transcripts, compared to just 1% at 100% RWC (Fig. 3.26D). After normalising for transcript length, this could imply that as many as 1 in 5 transcripts present during the later stages of desiccation could code for LEA proteins. One particular *X. humilis* LEA-3 transcript alone, with similarity to the *A. thaliana* LEA protein *At1g52690*/*F6D8.9*, accounted for over 2% of all sequenced reads across the 5% RWC RNA libraries (Fig. 3.26B). The increase in expression of many LEA transcripts appeared to peak (albeit at different magnitudes) in two separate stages, 80% and 40% RWC, giving their expression profiles a distinct appearance (Fig. 3.26E). This trend was not seen in the unrelated 80% and 60% RWC "B" samples, however (Fig. 3.26D). In general, LEA expression was far higher during late desiccation (after 60% RWC) than during earlier RWCs (Fig. 3.26E).

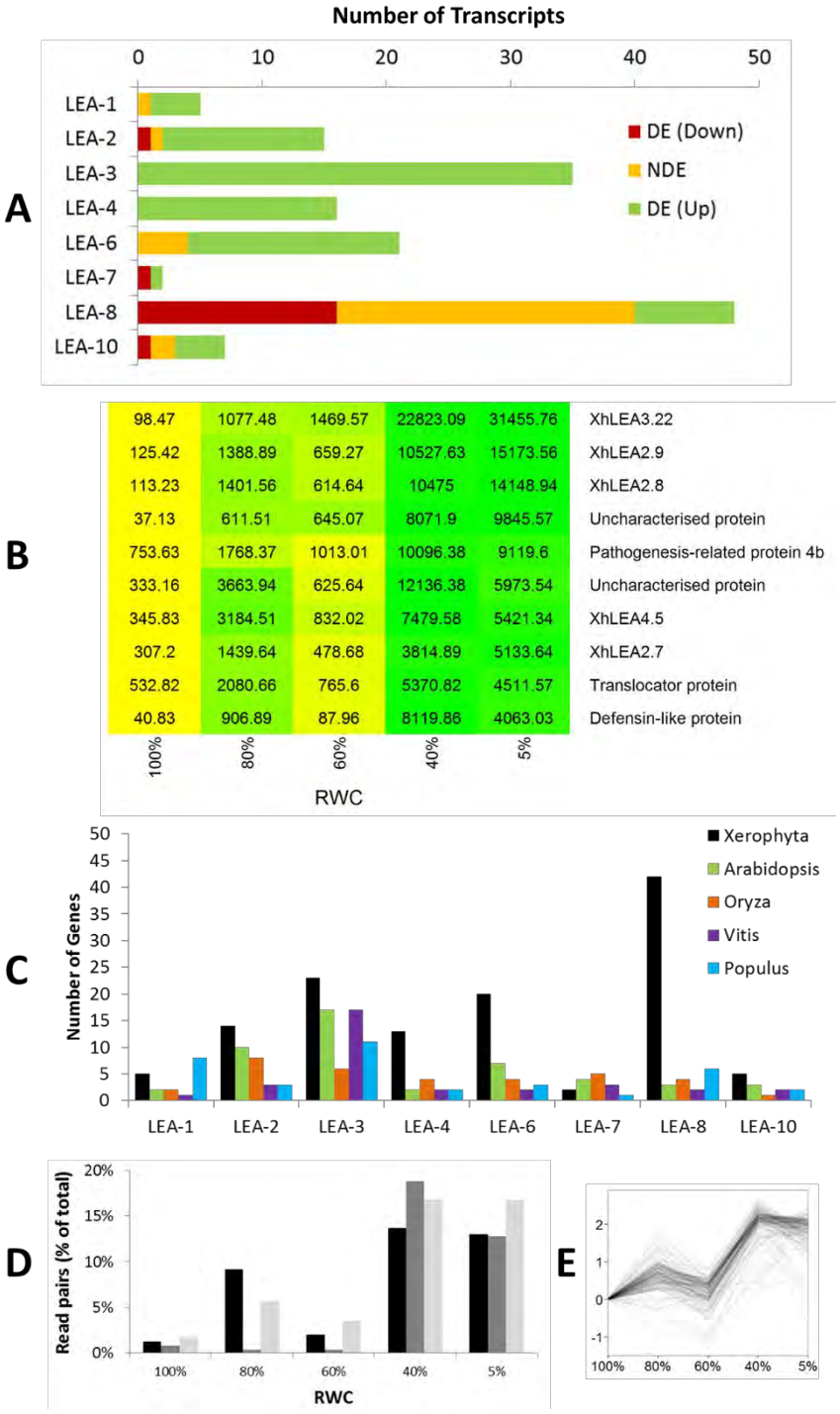


Figure 3.26: LEA gene expression in desiccating *X. humilis*. A) Transcripts from each LEA family up-regulated (green), down-regulated (red) or NDE (yellow) in the *X. humilis* transcriptome during desiccation. B) Ten of the most highly-expressed genes at 5% RWC; cell values correspond to FPKM, whereas colour denotes the distance from the expression level at 100% RWC. C) Comparison of the number of identified LEA genes in each family in *X. humilis* and four DS plant species. D) The percentage of RNA-Seq reads mapped to LEA transcripts in each sequencing library. E) Normalised gene expression profile of up-regulated LEA genes.

3.3.4.2 Seed maturation master regulators, with the exception of ABI3, are not differentially expressed during desiccation

I aimed to determine whether DT in *X. humilis* vegetative tissue evolved from the activation of the *LEC1*/*ABI3*/*FUS3*/*LEC2* (LAFL) seed maturation transcription factor cascade. Four identified *X. humilis* transcripts had similarity to key regulators of the LAFL pathway, namely *XhABI3.1*, *XhABI3.2*, *XhABI3.3* and *XhLEC1.1* (Fig. 3.27A). Only one of these transcripts, *XhABI3.1*, was differentially expressed in response to desiccation in *X. humilis* based on our selected cut-offs. By relaxing the requirements for a minimum 2-fold expression change, the *XhABI3.2* transcript was also statistically DE based on the FDR value, but total fold change and maximum expression level for both genes was still very low (FC ≈ 2, FPKM ≈ 5). The *XhABI3.3* and *XhLEC1* transcripts were not DE and, when present, were generally expressed at a level below 1 FPKM. Phylogenetic and sequence analysis of the proteins encoded by the three identified *X. humilis ABI3* transcripts revealed that all three clustered together with the *ABI3* protein sequences of other monocots. *XhABI3.1* and *XhABI3.2* appeared to be splice variants of the same gene, and so encoded proteins more similar to each other than they were to the NDE *XhABI3.3* (Fig. 3.27B). Both *XhABI3.1* and *XhABI3.2* were enriched for acidic residues at the N-terminal domain to a similar degree as the *AtABI3* protein. Interestingly, all three of the identified *Xerophyta ABI3* transcripts encoded a truncated protein that contained a stop codon shortly after the B2 domain, and thus would lack the seed motif-binding B3 domain that is the defining feature of the gene family (Fig. 3.27C).

The LAFL gene network, including *ABI3*, regulates the expression of multiple seed-specific genes, including LEAs and SSPs, during embryogenesis. Multiple LEA transcript were up-

regulated, as previously discussed, and a search for seed nutrient storage-related protein transcripts in *X. humilis* resulted in the identification of a number of such transcripts, including those related to oleosins, caleosins, globulins and bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin family genes. With the exception of one bifunctional 2S albumin, all were induced in vegetative tissue during desiccation (Fig. 3.28). Interestingly, all up-regulated oleosin, caleosin and bifunctional 2S albumin genes were induced as early as 80% RWC, whereas all predicted globulin SSP genes were only up-regulated during late desiccation.

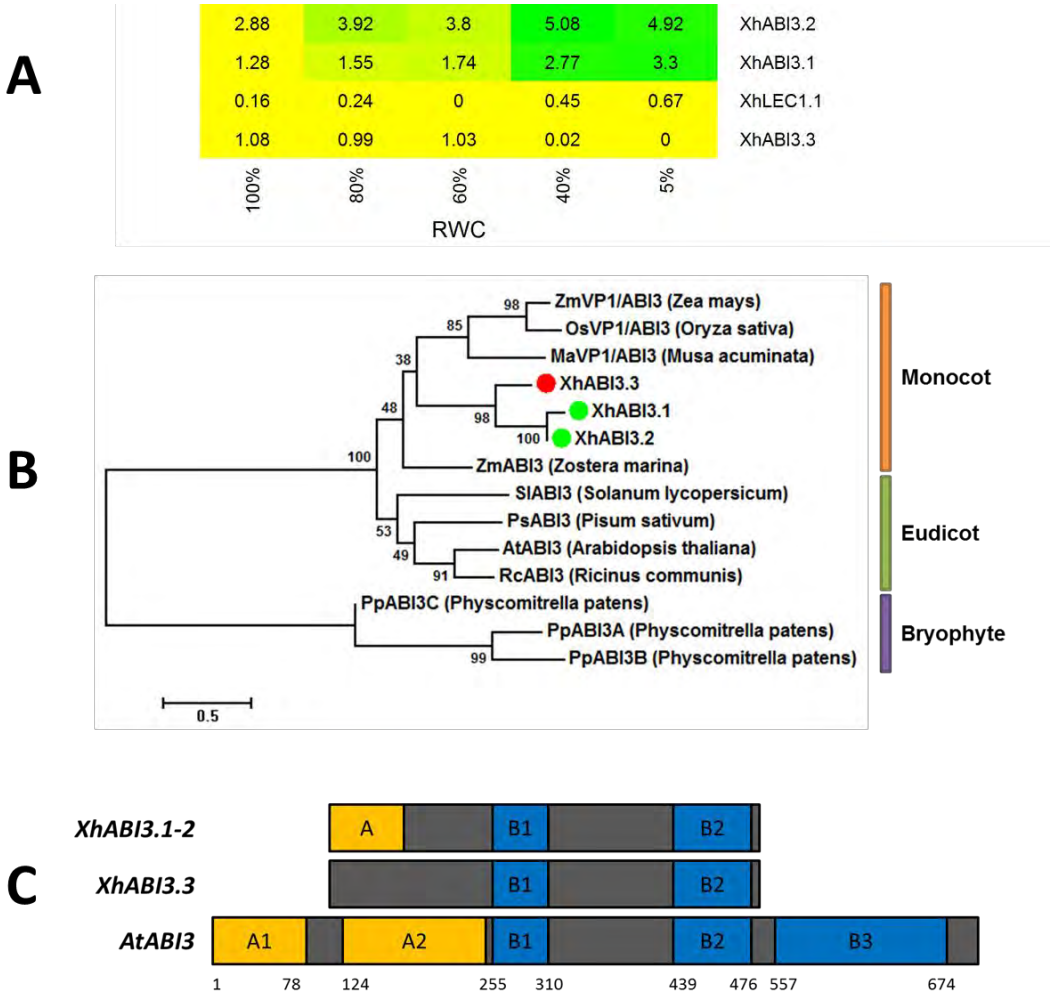


Figure 3.27. Two *ABI3* transcripts are differentially expressed in desiccating *X. humilis* leaves. A) Heatmap showing expression changes of the only identified LAFL network gene transcripts, *XhABI3.1*, *XhABI3.2*, *XhABI3.3* and *XhLEC1*. Only two, the transcript isoforms *XhABI3.1* and *XhABI3.2*, are DE during desiccation. B) Neighbour-joining phylogenetic tree of aligned *ABI3* protein sequences from multiple plant species, including those DE in *X. humilis* (green circles) and NDE (red circle). C) Conservation of B1 and B2 domains (blue) across *A. thaliana* and *X. humilis* ABI3 protein sequences. The B3 domain is absent from all *XhABI3* sequences due to a stop codon shortly after the B2 domain. A region enriched for acidic residues, but with little other similarity to A1 or A2 from *AtABI3*, is found only in *XhABI3.1* and *XhABI3.2*.

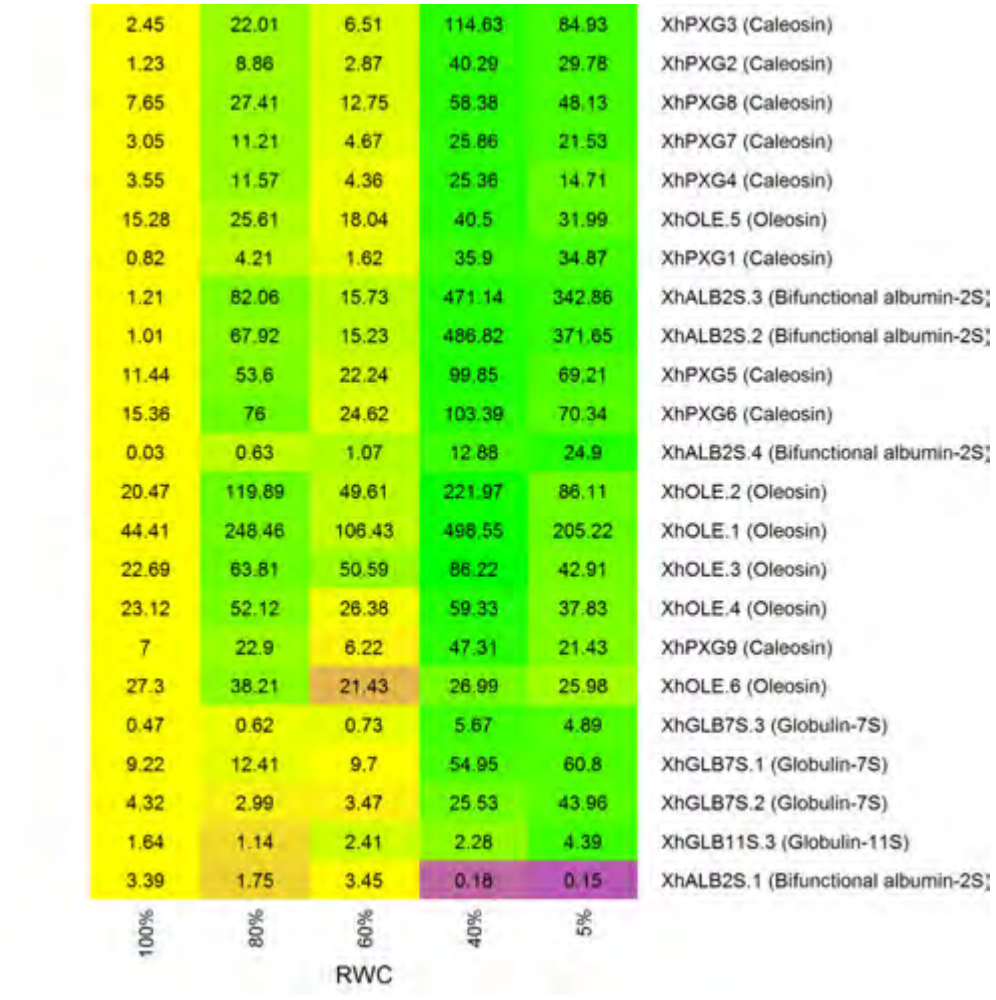


Figure 3.28. Predicted nutrient storage related genes identified in desiccating *X. humilis* leaves. Various genes associated with nutrient or lipid storage, usually found abundantly in seeds, were expressed in vegetative tissues during desiccation in *X. humilis*. Cell value is the FPKM, whereas cell colour denotes the difference in FPKM relative to 100% RWC. Entirely yellow rows denote genes that are not significantly DE.

3.3.4.3 Components of the ABA-signalling pathway are activated during desiccation

I aimed to explore the role of ABA signalling during desiccation in *X. humilis* by investigating the genes involved in ABA metabolism, ABA response and ABRE-mediated gene expression.

In plants, ABA is synthesised through a series of intermediate reactions controlled by a specific set of enzymes (Figure 3.29; Milborrow 2001; Nambara & Marion-poll 2005; Baron et al. 2012). The ABA biosynthesis pathway is controlled through both transcriptional and post-transcriptional regulation, and is proposed to be carefully regulated during development (e.g. seed maturation) and in response to stress (Milborrow, 2001; Nambara and Marion-Poll, 2005). For example, most of the transcripts of genes involved in ABA biosynthesis are up-regulated during drought or salt stress in *A. thaliana* (Xiong and Zhu, 2003). Active ABA can also be glycosylated to form ABA-glucose ester (ABA-GE), the predominant inactive form of the hormone, and can be converted back to the active form by glucose hydrolases such as *AtBG1* in *A. thaliana* (Lee et al., 2006). ABA catabolism is controlled by ABA hydroxylases, such as *AtCYP707A1-4* in *A. thaliana* (Nambara and Marion-Poll, 2005).

Orthologues of all of these genes, with the exception of *AtBG1*, could be identified in the *X. humilis* transcriptome. The biosynthesis genes *ABA1* and *ABA2* appeared to have only a single expressed homologue each in *X. humilis*, *XhABA1.1* and *XhABA2.1*, whereas there were four transcripts with high similarity to *ABA4*. Six *X. humilis* transcripts had similarity to the NCED/CCD family of carotenoid cleavage dioxigenases. However, only three of these (*XhNCED2, 5* and *6*) were most similar to the NCED genes believed to be involved in ABA

biosynthesis in *A. thaliana* or *O. sativa* (Fig. 3.30A; Tan et al. 2003; Zhu et al. 2009). Three *X. humilis* transcripts had similarity to the *A. thaliana* xanthine dehydrogenase family, only two of which (*XhAAO1* and *XhAAO2*) were in the acetaldehyde oxidase (AAO) clade that contains *AtAAO3*, the enzyme that catalyses the final step of ABA synthesis in *A. thaliana* (Fig. 3.30B; Baron et al. 2012). However, a phylogenetic comparison of the protein sequences of this family was not sufficient to determine if the identified *XhAAO* genes were necessarily functionally equivalent to *XhAAO3*. Lastly, twelve *X. humilis* transcripts encoded proteins that were highly similar to the *A. thaliana* CYP707A ABA hydroxylases based on phylogenetic analysis (Fig. 3.30C). Three (*XhCYP707A4, 10* and *12*) were orthologous to *AtCYP707A4*, while the remainder were most similar to *AtCYP707A1* and *AtCYP707A3*.

Changes in gene expression were not consistent across all *X. humilis* genes predicted to be involved in ABA metabolism. *XhABA1.1* and *XhABA2.1*, which catalyse the first and fourth steps of the ABA biosynthesis pathway, respectively, were both down-regulated during late desiccation (after 60% RWC; Fig. 3.29 and Fig. 3.30D). However, both of these genes were also expressed at a very high level compared to other genes in the pathway (>60 FPKM) during early desiccation. The three *X. humilis* NCED genes predicted to function during ABA biosynthesis (*XhNCED2, 5* and *6*) were up-regulated in leaves during both early and late desiccation (Fig. 3.30D), as were the *X. humilis ABA4* orthologues (Fig. 3.30D). *XhAAO1* and *XhAAO2*, which are similar to the ABA biosynthesis gene *AtAAO3*, were highly up-regulated only during late desiccation (40% and 5% RWC; Fig. 3.30D). *X. humilis CYP707A*-related ABA hydroxylase genes were generally down-regulated after 80% RWC, but those orthologous to *AtCYP707A4* (*XhCYP707A4, 10*, and *12*) were highly up-regulated during late desiccation.

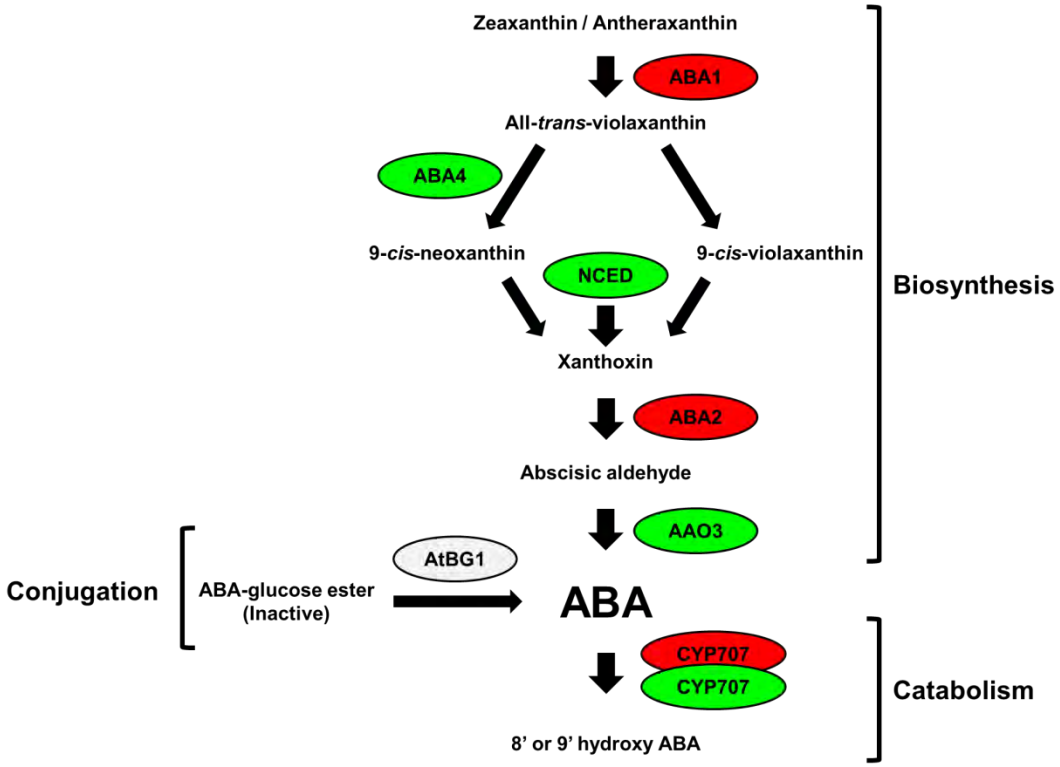


Figure 3.29. The expression changes of key ABA biosynthesis pathway genes in *X. humilis* leaves during desiccation. During ABA biosynthesis, zeaxanthin is converted to ABA via a series of intermediate reactions controlled by specific enzymes (given in coloured circles; green = up-regulated in *X. humilis* during desiccation, red = down-regulated). ABA can also accumulate through conversion of inactive glycosylated ABA to the active form by glucosyl hydrolases such as AtBG1; however, no orthologue of this gene could be found in the *X. humilis* transcriptome. ABA catabolism is achieved via ABA 8'- and 9'-hydroxylases (CYP707A1-4 in *A. thaliana*), transcripts of which were both up- and down-regulated in *X. humilis* during desiccation. Modified from Baron et al., 2012.

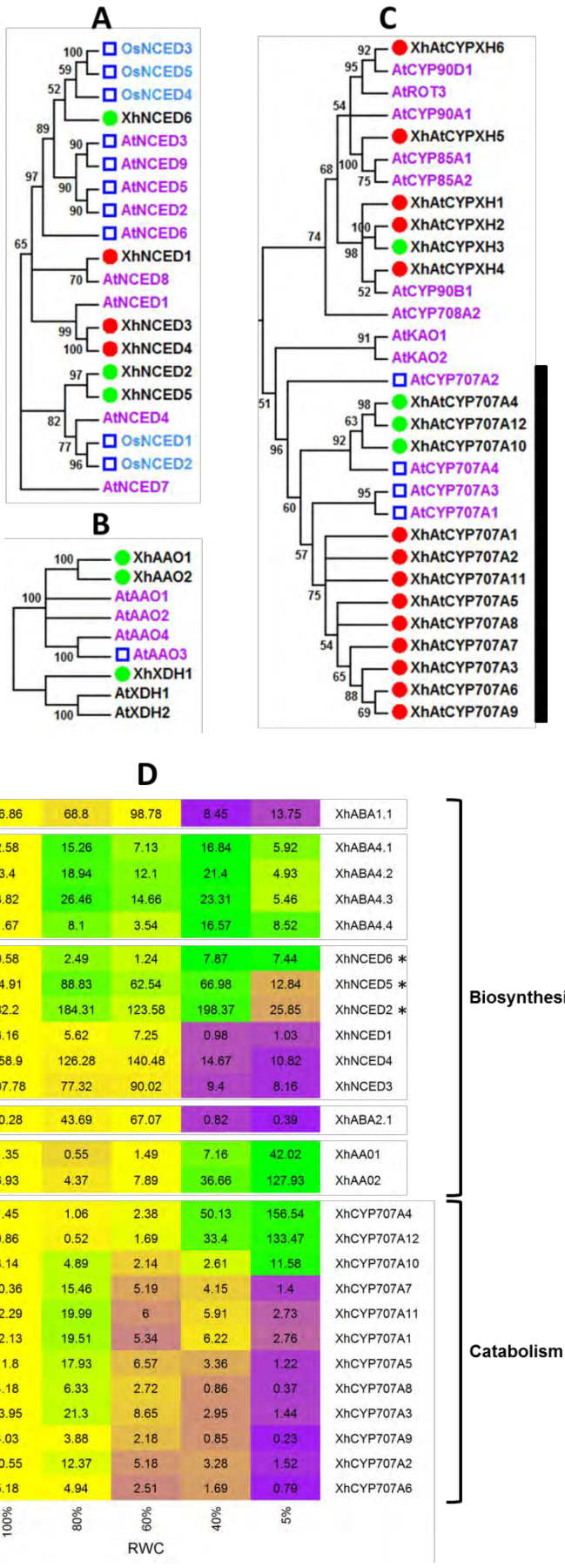

A
OsNCED3
OsNCED5
OsNCED4
XhNCED6
AtNCED3
AtNCED9
AtNCED5
AtNCED2
AtNCED6
XhNCED1
AtNCED8
AtNCED1
XhNCED3
XhNCED4
XhNCED2
XhNCED5
AtNCED4
OsNCED1
OsNCED2
AtNCED7
B
XhAAO1
XhAAO2
AtAAO1
AtAAO2
AtAAO4
AtAAO3
XhXDH1
AtXDH1
AtXDH2
C
XhAtCYPXH6
AtCYP90D1
AtROT3
AtCYP90A1
XhAtCYPXH5
AtCYP85A1
AtCYP85A2
XhAtCYPXH1
XhAtCYPXH2
XhAtCYPXH3
XhAtCYPXH4
AtCYP90B1
AtCYP708A2
AtKAO1
AtKAO2
AtCYP707A2
XhAtCYP707A4
XhAtCYP707A12
XhAtCYP707A10
AtCYP707A4
AtCYP707A3
AtCYP707A1
XhAtCYP707A1
XhAtCYP707A2
XhAtCYP707A11
XhAtCYP707A5
XhAtCYP707A8
XhAtCYP707A7
XhAtCYP707A3
XhAtCYP707A6
XhAtCYP707A9
D
96.86 68.8 98.78 8.45 13.75 XhABA1.1
2.58 15.26 7.13 16.84 5.92 XhABA4.1
3.4 18.94 12.1 21.4 4.93 XhABA4.2
3.82 26.46 14.66 23.31 5.46 XhABA4.3
1.67 8.1 3.54 16.57 8.52 XhABA4.4
0.58 2.49 1.24 7.87 7.44 XhNCED6 *
34.91 88.83 62.54 66.98 12.84 XhNCED5 *
82.2 184.31 123.58 198.37 25.85 XhNCED2 *
6.16 5.62 7.25 0.98 1.03 XhNCED1
158.9 126.28 140.48 14.67 10.82 XhNCED4
107.78 77.32 90.02 9.4 8.16 XhNCED3
60.28 43.69 67.07 0.82 0.39 XhABA2.1
1.35 0.55 1.49 7.16 42.02 XhAA01
6.93 4.37 7.89 36.66 127.93 XhAA02
Biosynthesis
1.45 1.06 2.38 50.13 156.54 XhCYP707A4
0.86 0.52 1.69 33.4 133.47 XhCYP707A12
3.14 4.89 2.14 2.61 11.58 XhCYP707A10
10.36 15.46 5.19 4.15 1.4 XhCYP707A7
12.29 19.99 6 5.91 2.73 XhCYP707A11
12.13 19.51 5.34 6.22 2.76 XhCYP707A1
11.8 17.93 6.57 3.36 1.22 XhCYP707A5
4.18 6.33 2.72 0.86 0.37 XhCYP707A8
13.95 21.3 8.65 2.95 1.44 XhCYP707A3
4.03 3.88 2.18 0.85 0.23 XhCYP707A9
10.55 12.37 5.18 3.28 1.52 XhCYP707A2
5.18 4.94 2.51 1.69 0.79 XhCYP707A6
Catabolism
100% 80% 60% 40% 5%
RWC

Figure 3.30. Differentially expressed genes involved in ABA metabolism in *X. humilis* leaves during desiccation. A-C) Unrooted Neighbour-Joining phylogenetic trees showing the evolutionary relationship between *X. humilis* NCED carotenoid cleavage dioxigenase genes (A), abscisic aldehyde oxidase (AAO) genes (B) and ABA 8'-hydroxylase genes (C) compared to *A. thaliana* (purple) and/or *O. sativa* (light blue) protein sequences. Gene homologues known to be involved in ABA biosynthesis are marked with an empty blue square. Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed. A) Only *XhNCED6* falls within the same clade as the NCED genes thought to be involved in ABA biosynthesis in *A. thaliana* (*AtNCED2,3,5,6* and *9*), although *XhNCED2* and *XhNCED5* group with the rice ABA-biosynthetic genes *OsNCED1* and *OsNCED2*. B) *XhAAO1* and *XhAAO2* group non-specifically with all four *A. thaliana* AAO gene homologues; *XhXDH1* is more similar to the related *A. thaliana* xanthine dehydrogenase (XDH) genes, and so is unlikely to be involved in ABA biosynthesis. C) Of the *X. humilis* CYP genes with significant sequence similarity to *AtCYP707A1-4*, only 12 cluster within the same clade (black bar) when compared to all *A. thaliana* CYPA protein sequences. The full tree can be found in the supplementary data (Fig. S3.5). D) Expression changes of the identified ABA metabolism genes in the *X. humilis* transcriptome during desiccation, related to ABA biosynthesis (top half) and ABA catabolism (bottom half). Values are FPKM, whereas cell colour denotes the difference in FPKM relative to 100% RWC.

The response to endogenous ABA is controlled by multiple families of ABA receptors and response factors. Group A bZIP proteins (ABRE binding factors; ABFs) are important activators of many ABA response genes downstream of ABA signalling during both seed development/germination and vegetative stress. Although no transcript encoding a protein that was directly orthologous to *AtABI5* was found in the desiccation transcriptome of *X. humilis*, four transcripts were identified with similarity to *ABI5*-like the group A bZIP TFs (*XhABF1*, *XhABF2*, *XhABF3* and *XhABF4*) all of which were differentially expressed in leaf tissue during desiccation based on our cut-offs (Fig. 3.31A). *XhABF1* was strongly induced by desiccation at nearly all stages, whereas *XhABF2* and *XhABF3* were transiently induced at low levels at 80% RWC, followed by decreased expression in dry tissue. *XhABF4* was present at very low levels in hydrated tissue, and decreased further in transcript abundance at later stages. A comparison of the predicted *X. humilis* ABF protein sequences with those of

A. thaliana, as well as the seed-specific ABI5 orthologues of *O. sativa* and *T. aestivum*, revealed that with the exception of *XhABF4* they were more similar to ABF proteins expressed during vegetative stress than those involved in seed development (Fig. 3.31B). Several ABI5 binding proteins (AFPs), which are predicted to negatively regulate ABF protein function, were also differentially expressed during desiccation (Fig. 3.31C). The majority of AFPs were present at low levels in hydrated tissue, but increased in abundance at the later stages of desiccation.

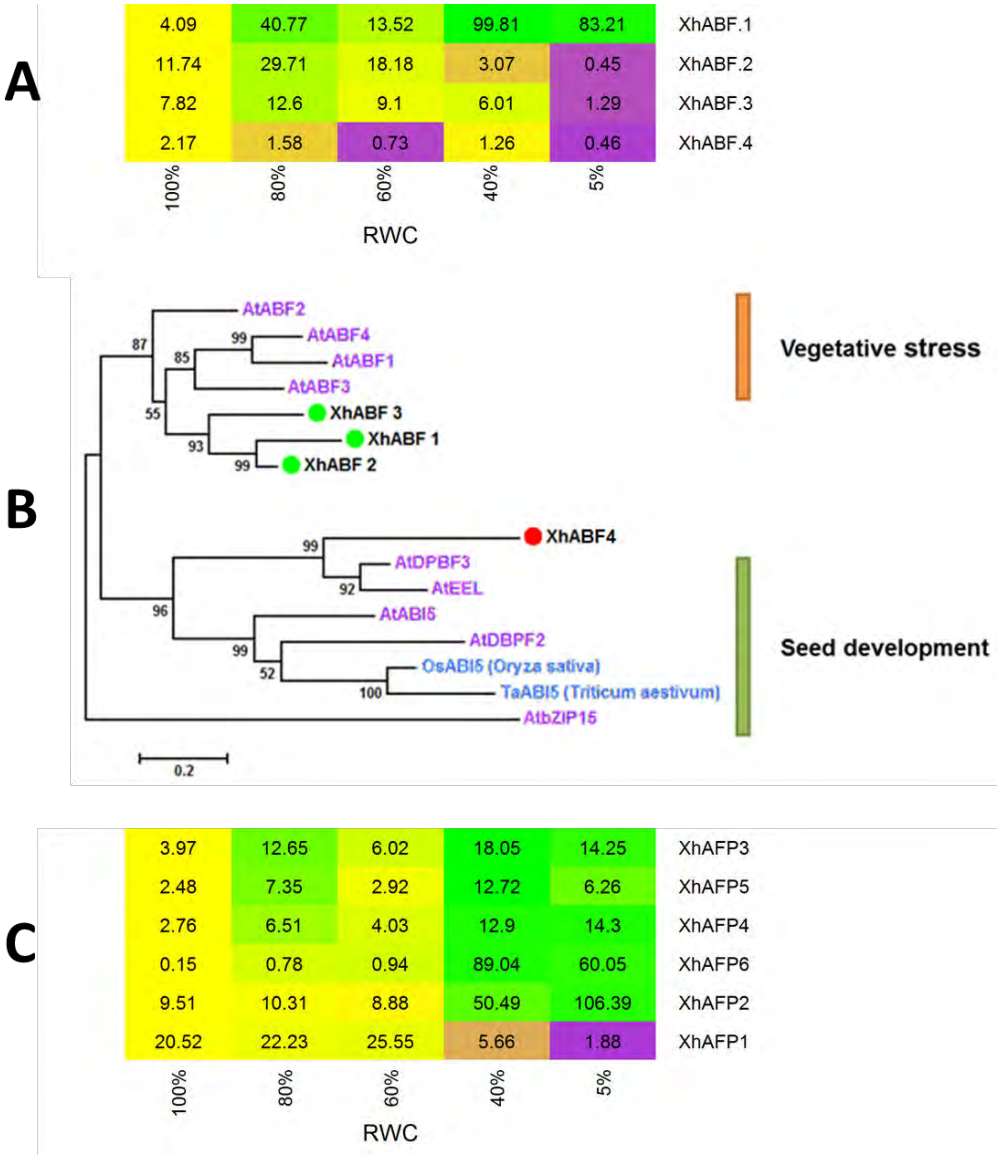
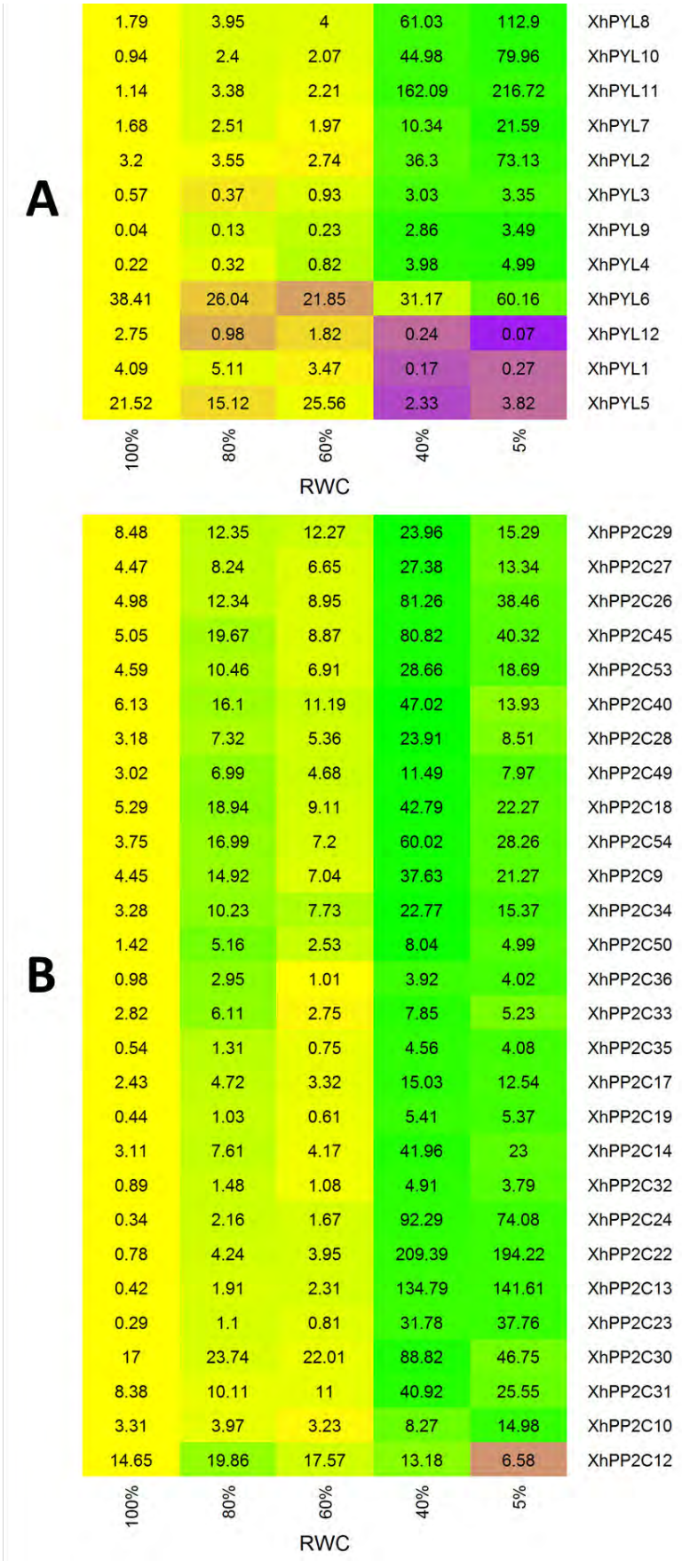


Figure 3.31. ABF proteins expressed in *X. humilis*. A) Expression change of four identified *X. humilis* ABF-encoding transcripts. B) Phylogenetic comparison of predicted *X. humilis* ABF protein sequences with those of *A. thaliana,* as well as known seed-specific ABI5 sequences of *O. sativa* and *T. aestivum*. C) Expression of predicted *X. humilis* ABF-binding AFP protein transcripts.

Upstream components of the ABA signal transduction pathway were also differentially expressed during desiccation in *X. humilis*. With the exception of *XhPYL1*, *XhPYL5* and *XhPYL12*, multiple transcripts predicted to encode PYL-class ABA receptors were strongly induced during late desiccation (Fig. 3.32A; Fig. S3.6). Additionally, sixty-one transcripts encoding PP2C protein phosphatases were identified in the *X. humilis* transcriptome, most of

which were up-regulated. Phylogenetic comparison of the predicted *X. humilis* PP2C proteins with those from *A. thaliana* showed that just under half (26/61) could be clustered with the group A PP2C proteins, which are involved in ABA signalling (Fig 3.32B; Fig. S3.7). Finally, multiple SnRK2 protein kinase genes were expressed during desiccation. Comparison of the predicted protein sequences of these genes with the known SnRK2 proteins from *A. thaliana* and *O. sativa* showed that three clustered with the Group 3 SnRK2 proteins (strongly ABA activated; *XhSnRK2.4*, *XhSnRK2.5* and *XhSnRK2.6*) and three clustered with those from Group 2 (weakly activated by ABA; *XhSnRK2.1, XhSnRK2.2* and *XhSnRK2.11*) (Fig. 3.32C; Fig. S3.8).


A
1.79 3.95 4 61.03 112.9 XhPYL8
0.94 2.4 2.07 44.98 79.96 XhPYL10
1.14 3.38 2.21 162.09 216.72 XhPYL11
1.68 2.51 1.97 10.34 21.59 XhPYL7
3.2 3.55 2.74 36.3 73.13 XhPYL2
0.57 0.37 0.93 3.03 3.35 XhPYL3
0.04 0.13 0.23 2.86 3.49 XhPYL9
0.22 0.32 0.82 3.98 4.99 XhPYL4
38.41 26.04 21.85 31.17 60.16 XhPYL6
2.75 0.98 1.82 0.24 0.07 XhPYL12
4.09 5.11 3.47 0.17 0.27 XhPYL1
21.52 15.12 25.56 2.33 3.82 XhPYL5
100% 80% 60% 40% 5%
RWC
B
8.48 12.35 12.27 23.96 15.29 XhPP2C29
4.47 8.24 6.65 27.38 13.34 XhPP2C27
4.98 12.34 8.95 81.26 38.46 XhPP2C26
5.05 19.67 8.87 80.82 40.32 XhPP2C45
4.59 10.46 6.91 28.66 18.69 XhPP2C53
6.13 16.1 11.19 47.02 13.93 XhPP2C40
3.18 7.32 5.36 23.91 8.51 XhPP2C28
3.02 6.99 4.68 11.49 7.97 XhPP2C49
5.29 18.94 9.11 42.79 22.27 XhPP2C18
3.75 16.99 7.2 60.02 28.26 XhPP2C54
4.45 14.92 7.04 37.63 21.27 XhPP2C9
3.28 10.23 7.73 22.77 15.37 XhPP2C34
1.42 5.16 2.53 8.04 4.99 XhPP2C50
0.98 2.95 1.01 3.92 4.02 XhPP2C36
2.82 6.11 2.75 7.85 5.23 XhPP2C33
0.54 1.31 0.75 4.56 4.08 XhPP2C35
2.43 4.72 3.32 15.03 12.54 XhPP2C17
0.44 1.03 0.61 5.41 5.37 XhPP2C19
3.11 7.61 4.17 41.96 23 XhPP2C14
0.89 1.48 1.08 4.91 3.79 XhPP2C32
0.34 2.16 1.67 92.29 74.08 XhPP2C24
0.78 4.24 3.95 209.39 194.22 XhPP2C22
0.42 1.91 2.31 134.79 141.61 XhPP2C13
0.29 1.1 0.81 31.78 37.76 XhPP2C23
17 23.74 22.01 88.82 46.75 XhPP2C30
8.38 10.11 11 40.92 25.55 XhPP2C31
3.31 3.97 3.23 8.27 14.98 XhPP2C10
14.65 19.86 17.57 13.18 6.58 XhPP2C12
100% 80% 60% 40% 5%
RWC

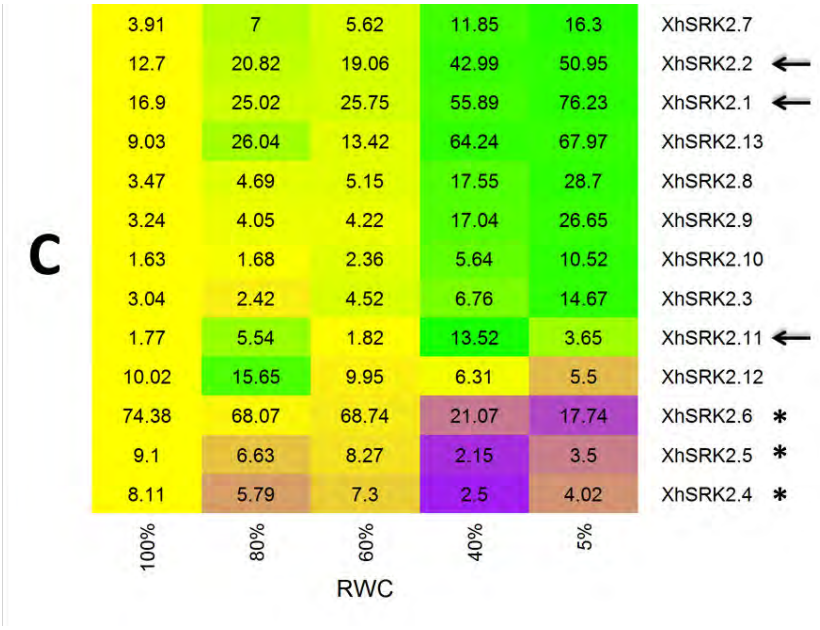


Figure 3.32. Genes involved in ABA signal transduction in desiccating *X. humilis* leaves. A) PYL-family ABA receptors. B) Predicted group A PP2C genes based on similarity to *A. thaliana* PP2C protein sequences. C) SnRK2 protein kinase genes. Those predicted to encode group 2 SnRK2 proteins based on similarity to *A. thaliana* and *O. sativa* protein sequences are marked with an arrow; predicted group 3 SnRK2s are marked with an asterisk.

3.3.4.4 Distinct transcription factors and TF families are regulated throughout desiccation

Out of the 42,958 putative *X. humilis* genes clusters (ignoring fungal transcripts), 18,534 (43%) were differentially expressed during a cycle of dehydration. These massive changes in transcript abundance imply the existence of massive transcriptional re-programming during desiccation. The potential changes in transcriptional activity was investigated via two related routes: transcription factors (TFs), and chromatin epigenetic modifications of the promoters on which they act.

Transcription factors expressed in desiccating X. humilis

Over 3,000 gene clusters were identified as putative TFs based on the presence of the GO molecular functions "sequence-specific DNA binding" (GO:0043565) and "sequence-specific

DNA transcription activity" (GO:0003700). Nearly 1,300 of these genes were also differentially expressed during desiccation in *X. humilis*, of which slightly over 1,000 could be grouped into one of the recognised plant transcription factor families based on Interpro binding domains or annotation (Table 3.3; H. Zhang et al. 2011). Those that could not be classified were dropped from further analysis to reduce errors due to possible misclassification of non-TF genes.

The majority of TFs in the desiccation transcriptome of *X. humilis* were from the *APETALA2/ethylene-responsive binding factor* (AP2/ERF), MYB, Homeobox, Zinc-finger C3H and WRKY transcription factor families, but a significant number could also be found from the basic-helix-loop-helix (bHLH), NAM/ATAF/CUC2 (NAC), zinc-finger C2H2, *AUXIN-RESPONSE FACTOR* (ARF), basic leucine-zipper (bZIP) and heat-stress factor (HSF) families (Table 3.3). TFs from all of these families have been implicated as significant players in both stress response and seed development. To further dissect the role that the detected TFs might play during desiccation, TFs from known families were clustered into expression profiles using the MeV k-means algorithm. In total six expression clusters were selected (Fig. 3.33). The identified TFs displayed consecutive waves of expression as desiccation progressed: clusters U1 and U2 were induced during early dehydration (80% and 60% RWC respectively), whereas clusters U3 and U4 predominantly had peak expression at 40% and 5% RWC. Gene down-regulation was most prominent at 80% RWC and 40% RWC (cluster D1 and D2). For the most part, the distribution of the top ten TF families within each cluster was similar to that for the dataset as a whole, with AP2/ERF, Myb and homeobox-domain TFs being the most common (Fig. 3.33).

TF Family	Interpro Term	Interpro Domain	Interpro Match	BLAST Match	Total
AP2/ERF	IPR001471	AP2/ERF domain	96	33	129
MYB	IPR001005	SANT/Myb domain	90	15	105
Homeobox domain	IPR001356	Homeobox domain	67	36	103
C3H	IPR001878	Zinc finger, CCHC-type	3	84	87
WRKY	IPR003657	WRKY domain	57	28	85
bHLH	IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain	36	27	63
NAC	IPR003441	NAC domain	40	14	54
C2H2	IPR007087	Zinc finger, C2H2	40	0	40
ARF	IPR010525	Auxin response factor	13	25	38
bZIP	IPR004827	Basic-leucine zipper domain	36	0	36
CO-like	IPR010402	CTT domain	30	0	30
GRAS	IPR005202	Transcription factor GRAS	27	0	27
RING-FYVE-PHD	IPR011011	Zinc finger, FYVE/PHD-type	9	18	27
GATA	IPR000679	Zinc finger, GATA-type	19	5	24
HSF	IPR027709	Heat shock transcription factor, plant	21	0	21
TCP	IPR005333	Transcription factor, TCP	17	0	17
NF-YB	IPR003958	Transcription factor CBF/NF-Y/archaeal histone domain	11	5	16
Trihelix	IPR027759	Trihelix transcription factor GT3	1	12	13
DOF	IPR003851	Zinc finger, Dof-type	13	0	13
FAR1	IPR004330	FAR1 DNA binding domain	11	0	11
EIL	IPR023278	Ethylene insensitive 3-like protein, DNA-binding domain	9	0	9
DBB	IPR000315	B-box-type zinc finger	9	0	9
B3	IPR003340	B3 DNA binding domain	8	0	8
NF-YA	IPR001289	CCAAT-binding transcription factor, subunit B	6	0	6
MADS	IPR002100	Transcription factor, MADS-box	4	0	4
LSD1	IPR005735	Zinc finger, LSD1-type	3	0	3
YABBY	IPR006780	YABBY protein	3	0	3
SBP	IPR004333	Transcription factor, SBP-box	2	0	2
LOB	IPR004883	Lateral organ boundaries, LOB	2	0	2
ZF-HD	IPR006455	Homeodomain, ZF-HD class	2	0	2
NF-X1	IPR000967	Zinc finger, NF-X1-type	1	0	1
CPP	IPR005172	CRC domain	1	0	1
DUF573	IPR007592	Protein of unknown function DUF573	1	0	1
DUF822	IPR008540	BES1/BZR1 plant transcription factor, N-terminal	1	0	1
WRC	IPR014977	WRC domain	1	0	1
		Total	**696**	**310**	**1006**

Table 3.4: Distribution of transcription factor (TF) gene families in the *X. humilis* transcriptome. Putative TFs were identified based on the presence of the GO terms "sequence-specific DNA binding" (GO:0043565) and "sequence-specific DNA transcription activity" (GO:0003700) . TF gene transcripts were then further categorised by the presence of common plant TF motifs or binding domains (Interpro), or by similarity to known plant TFs in the NCBI Swissprot/Uniref90 databases.

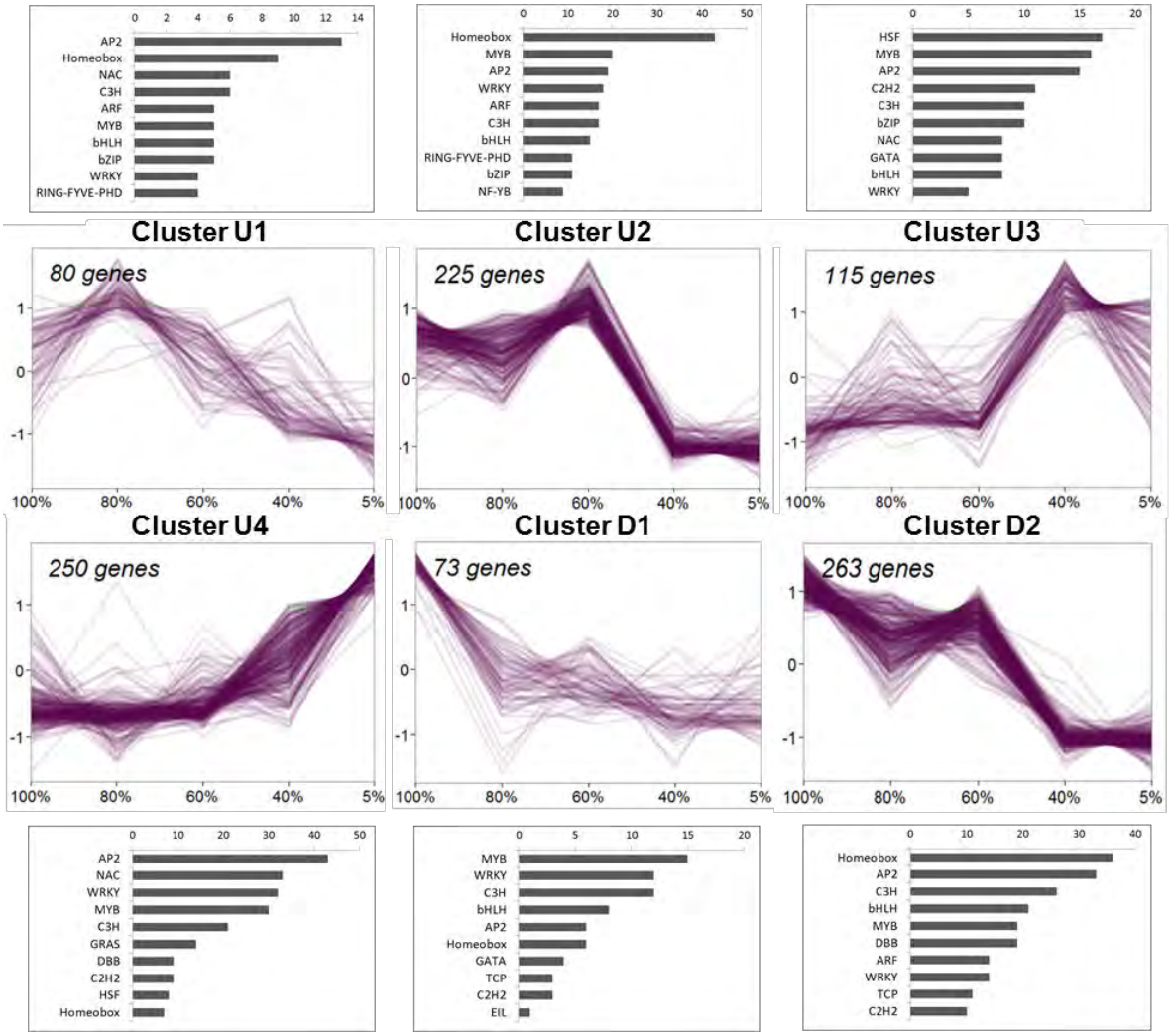


Figure 3.33: Expression profiles and distribution of TF families in the *X. humilis* desiccation transcriptome. TF gene transcripts were clustered into expression profiles by MeV k-means clustering. The numbers of genes from each plant TF family were tallied, and the top 10 families in each cluster are displayed in this figure (bar graphs; x-axis is number of transcripts).

In an attempt to characterise the diversity of TF functions during desiccation in *X. humilis*, GO enrichment analysis was performed on the predicted TF expression clusters. A very large number of GO terms were enriched across and between clusters. For example, all six clusters were enriched for GO terms associated with positive and negative regulation of DNA-templated transcription compared to the rest of the transcriptome, as would be expected for TF genes. However, numerous other terms were also shared between all

clusters, irrespective of expression profile (Fig. 3.34). These included terms related to defence response (regulation of defence response and plant-type hypersensitive response; respiratory burst, and response to bacteria/fungi), signalling (response to chitin, cyclopentane, cytokinin, auxin, ABA and far-red light; GA and ethylene signalling pathways; jasmonic and salicylic acid induced systemic resistance), and growth and development – particularly of reproductive structures (regulation of flower/seed development, regulation of seed dormancy/germination; regulation of photomorphogenesis and leaf senescence). Due to the very large number of significantly-enriched GO terms (even after accounting for the above shared GO terms) further analysis focused largely on the lowest-level GO terms that were uniquely found in each cluster (Table 3.5). However, a list of terms shared by two or more TF expression clusters is also given in Table 3.6.

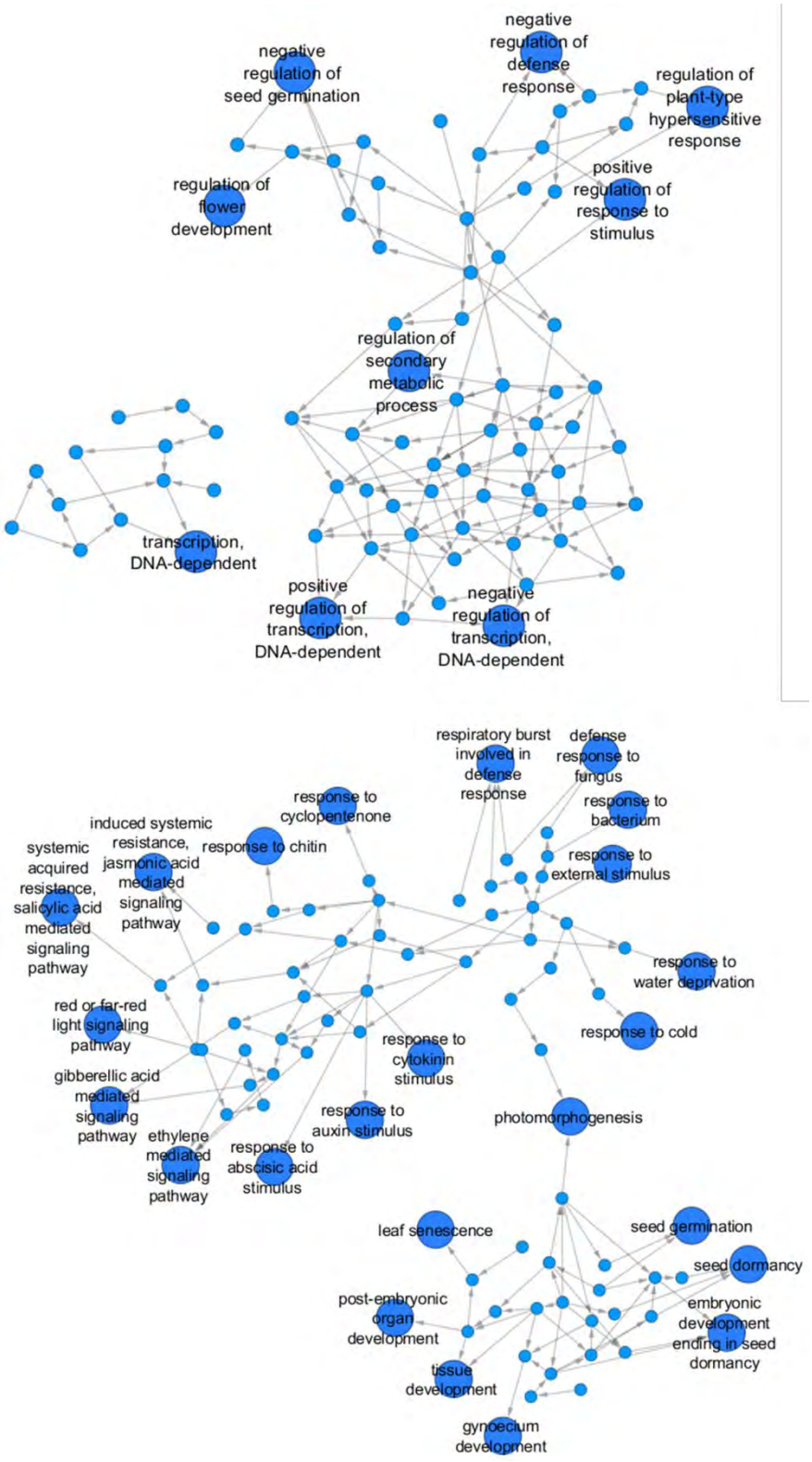


Figure 3.34. Enriched GO terms shared between all TF expression profiles. Only the lowest-level GO terms within the network are labelled (large blue nodes), although all visible nodes were significantly enriched in each respective cluster (adjusted p-value < 0.01).

GO Code	Biological Process	U1	U2	U3	U4	D1	D2
GO:0001666	response to hypoxia						
GO:0009615	response to virus						
GO:0009648	photoperiodism						
GO:0009785	blue light signaling pathway						
GO:0009880	embryonic pattern specification						
GO:0010160	formation of organ boundary						
GO:0010358	leaf shaping						
GO:0050826	response to freezing						
GO:0006333	chromatin assembly or disassembly						
GO:0006346	methylation-dependent chromatin silencing						
GO:0006366	transcription from RNA polymerase II promoter						
GO:0007155	cell adhesion						
GO:0008285	negative regulation of cell proliferation						
GO:0009410	response to xenobiotic stimulus						
GO:0009555	pollen development						
GO:0009624	response to nematode						
GO:0009691	cytokinin biosynthetic process						
GO:0009911	positive regulation of flower development						
GO:0010048	vernalization response						
GO:0010080	regulation of floral meristem growth						
GO:0010220	positive regulation of vernalization response						
GO:0010413	glucuronoxylan metabolic process						
GO:0010440	stomatal lineage progression						
GO:0010492	maintenance of shoot apical meristem identity						
GO:0010500	transmitting tissue development						
GO:0016246	RNA interference						
GO:0016573	histone acetylation						
GO:0017055	negative regulation of RNA polymerase II transcriptional preinitiation complex assembly						
GO:0019375	galactolipid biosynthetic process						
GO:0019432	triglyceride biosynthetic process						
GO:0031048	chromatin silencing by small RNA						
GO:0032107	regulation of response to nutrient levels						
GO:0042631	cellular response to water deprivation						
GO:0043457	regulation of cellular respiration						
GO:0045010	actin nucleation						
GO:0045492	xylan biosynthetic process						
GO:0045723	positive regulation of fatty acid biosynthetic process						
GO:0060772	leaf phyllotactic patterning						
GO:0060774	auxin mediated signaling pathway involved in phyllotactic patterning						
GO:0070918	production of small RNA involved in gene silencing by RNA						
GO:0071249	cellular response to nitrate						
GO:0071345	cellular response to cytokine stimulus						
GO:0006990	positive regulation of gene-specific transcription involved in unfolded protein response						
GO:0009220	pyrimidine ribonucleotide biosynthetic process						
GO:0009745	sucrose mediated signaling						
GO:0010218	response to far red light						
GO:0010286	heat acclimation						
GO:0034605	cellular response to heat						
GO:0055062	phosphate ion homeostasis						
GO:0071236	cellular response to antibiotic						
GO:0071499	cellular response to laminar fluid shear stress						
GO:0000398	nuclear mRNA splicing, via spliceosome						
GO:0001510	RNA methylation						
GO:0002237	response to molecule of bacterial origin						
GO:0009741	response to brassinosteroid stimulus						
GO:0009749	response to glucose stimulus						
GO:0009834	secondary cell wall biogenesis						
GO:0009965	leaf morphogenesis						
GO:0010047	fruit dehiscence						
GO:0010072	primary shoot apical meristem specification						
GO:0010199	organ boundary specification between lateral organs and the meristem						
GO:0010353	response to trehalose stimulus						
GO:0010449	root meristem growth						
GO:0010602	regulation of 1-aminocyclopropane-1-carboxylate metabolic process						
GO:0010618	aerenchyma formation						
GO:0010896	regulation of triglyceride catabolic process						
GO:0016567	protein ubiquitination						
GO:0019915	lipid storage						
GO:0042991	transcription factor import into nucleus						
GO:0048504	regulation of timing of organ formation						
GO:0051301	cell division						
GO:0051457	maintenance of protein location in nucleus						
GO:0051782	negative regulation of cell division						

GO Code	Biological Process	U1	U2	U3	U4	D1	D2
GO:0006979	response to oxidative stress						
GO:0009595	detection of biotic stimulus						
GO:0009697	salicylic acid biosynthetic process						
GO:0009816	defense response to bacterium, incompatible interaction						
GO:0010114	response to red light						
GO:0010193	response to ozone						
GO:0046621	negative regulation of organ growth						
GO:0048498	establishment of petal orientation						
GO:0080125	multicellular structure septum development						
GO:0006281	DNA repair						
GO:0006808	regulation of nitrogen utilization						
GO:0006826	iron ion transport						
GO:0006857	oligopeptide transport						
GO:0009585	red, far-red light phototransduction						
GO:0009590	detection of gravity						
GO:0009637	response to blue light						
GO:0009759	indole glucosinolate biosynthetic process						
GO:0009805	coumarin biosynthetic process						
GO:0009938	negative regulation of gibberellic acid mediated signaling pathway						
GO:0009944	polarity specification of adaxial/abaxial axis						
GO:0009959	negative gravitropism						
GO:0010052	guard cell differentiation						
GO:0010091	trichome branching						
GO:0010233	phloem transport						
GO:0010311	lateral root formation						
GO:0010345	suberin biosynthetic process						
GO:0015706	nitrate transport						
GO:0016226	iron-sulfur cluster assembly						
GO:0032527	protein exit from endoplasmic reticulum						
GO:0034976	response to endoplasmic reticulum stress						
GO:0035304	regulation of protein amino acid dephosphorylation						
GO:0042744	hydrogen peroxide catabolic process						
GO:0043433	negative regulation of transcription factor activity						
GO:0045036	protein targeting to chloroplast						
GO:0048446	petal morphogenesis						
GO:0048497	maintenance of floral organ identity						
GO:0048573	photoperiodism, flowering						
GO:0052545	callose localization						

Table 3.5. Lowest-level GO terms unique to each TF expression cluster GO network. The ID and description of each of the lowest-level child GO terms within the networks of enriched GO terms for each TF expression clusters were isolated, and those found uniquely in a single cluster were grouped. Columns U1 to D2 correspond to TF clusters U1 to D2.

GO Code	Biological Process	U1	U2	U3	U4	D1	D2
GO:0009736	cytokinin mediated signaling pathway						
GO:0009738	abscisic acid mediated signaling pathway						
GO:0009910	negative regulation of flower development						
GO:0048825	cotyledon development						
GO:0042538	hyperosmotic salinity response						
GO:0030968	endoplasmic reticulum unfolded protein response						
GO:0010093	specification of floral organ identity						
GO:0048527	lateral root development						
GO:0010582	floral meristem determinacy						
GO:0048510	regulation of timing of transition from vegetative to reproductive phase						
GO:0031537	regulation of anthocyanin metabolic process						
GO:0000436	positive regulation of transcription from RNA polymerase II promoter by carbon catabolites						
GO:0010262	somatic embryogenesis						
GO:0034059	response to anoxia						
GO:0010105	negative regulation of ethylene mediated signaling pathway						
GO:0015996	chlorophyll catabolic process						
GO:0009933	meristem structural organization						
GO:0010084	specification of organ axis polarity						
GO:0010182	sugar mediated signaling pathway						
GO:0010228	vegetative to reproductive phase transition of meristem						
GO:0048829	root cap development						
GO:0009630	gravitropism						
GO:0009269	response to desiccation						
GO:0009734	auxin mediated signaling pathway						
GO:0009788	negative regulation of abscisic acid mediated signaling pathway						
GO:0010167	response to nitrate						
GO:0009718	anthocyanin biosynthetic process						
GO:0009961	response to 1-aminocyclopropane-1-carboxylic acid						
GO:0052542	defense response by callose deposition						
GO:0000303	response to superoxide						
GO:0010119	regulation of stomatal movement						
GO:0048576	positive regulation of short-day photoperiodism, flowering						
GO:0080113	regulation of seed growth						
GO:0009407	toxin catabolic process						
GO:0010380	regulation of chlorophyll biosynthetic process						
GO:0080022	primary root development						
GO:0010082	regulation of root meristem growth						
GO:0009611	response to wounding						
GO:0048574	long-day photoperiodism, flowering						
GO:0010310	regulation of hydrogen peroxide metabolic process						
GO:0009641	shade avoidance						
GO:0009686	gibberellin biosynthetic process						
GO:0010018	far-red light signaling pathway						
GO:0010050	vegetative phase change						
GO:0010161	red light signaling pathway						
GO:0010223	secondary shoot formation						
GO:0048572	short-day photoperiodism						
GO:0048579	negative regulation of long-day photoperiodism, flowering						
GO:0080060	integument development						
GO:0009625	response to insect						
GO:0009744	response to sucrose stimulus						
GO:0009846	pollen germination						
GO:0009942	longitudinal axis specification						
GO:0010229	inflorescence development						
GO:0030010	establishment of cell polarity						
GO:0042762	regulation of sulfur metabolic process						
GO:0043620	regulation of transcription in response to stress						
GO:0070370	cellular heat acclimation						
GO:0009800	cinnamic acid biosynthetic process						
GO:0010371	regulation of gibberellin biosynthetic process						
GO:0048354	mucilage biosynthetic process involved in seed coat development						
GO:0048453	sepal formation						
GO:0009722	detection of cytokinin stimulus						
GO:0010055	atrichoblast differentiation						
GO:0010090	trichome morphogenesis						
GO:0016036	cellular response to phosphate starvation						
GO:0048451	petal formation						

GO Code	Biological Process	U1	U2	U3	U4	D1	D2
GO:0006417	regulation of translation						
GO:0007263	nitric oxide mediated signal transduction						
GO:0007623	circadian rhythm						
GO:0009610	response to symbiotic fungus						
GO:0009616	virus induced gene silencing						
GO:0009658	chloroplast organization						
GO:0009963	positive regulation of flavonoid biosynthetic process						
GO:0010014	meristem initiation						
GO:0010027	thylakoid membrane organization						
GO:0010031	circumnutation						
GO:0010051	xylem and phloem pattern formation						
GO:0010076	maintenance of floral meristem identity						
GO:0010077	maintenance of inflorescence meristem identity						
GO:0010089	xylem development						
GO:0010100	negative regulation of photomorphogenesis						
GO:0010158	abaxial cell fate specification						
GO:0010197	polar nucleus fusion						
GO:0010201	response to continuous far red light stimulus by the high-irradiance response system						
GO:0010227	floral organ abscission						
GO:0042659	regulation of cell fate specification						
GO:0045604	regulation of epidermal cell differentiation						
GO:0048457	floral whorl morphogenesis						
GO:0071472	cellular response to salt stress						
GO:0080127	fruit septum development						
GO:0080147	root hair cell development						
GO:0009612	response to mechanical stimulus						
GO:0009646	response to absence of light						
GO:0009693	ethylene biosynthetic process						
GO:0009644	response to high light intensity						
GO:0010600	regulation of auxin biosynthetic process						
GO:0071497	cellular response to freezing						
GO:0006406	mRNA export from nucleus						
GO:0009631	cold acclimation						
GO:0009643	photosynthetic acclimation						
GO:0009695	jasmonic acid biosynthetic process						
GO:0010099	regulation of photomorphogenesis						
GO:0010117	photoprotection						
GO:0010224	response to UV-B						
GO:0031539	positive regulation of anthocyanin metabolic process						
GO:0032922	circadian regulation of gene expression						
GO:0035264	multicellular organism growth						
GO:0042542	response to hydrogen peroxide						
GO:0045926	negative regulation of growth						
GO:0048235	pollen sperm cell differentiation						
GO:0070483	detection of hypoxia						
GO:0071456	cellular response to hypoxia						
GO:0006661	phosphatidylinositol biosynthetic process						
GO:0042753	positive regulation of circadian rhythm						
GO:0006612	protein targeting to membrane						
GO:0043069	negative regulation of programmed cell death						
GO:0043090	amino acid import						
GO:0000162	tryptophan biosynthetic process						
GO:0009861	jasmonic acid and ethylene-dependent systemic resistance						
GO:0009956	radial pattern formation						
GO:0015824	proline transport						
GO:0042793	transcription from plastid promoter						
GO:0046482	para-aminobenzoic acid metabolic process						
GO:0048571	long-day photoperiodism						
GO:0048831	regulation of shoot development						
GO:0002213	defense response to insect						
GO:0009835	ripening						
GO:0010106	cellular response to iron ion starvation						
GO:0010438	cellular response to sulfur starvation						
GO:0010439	regulation of glucosinolate biosynthetic process						
GO:0016132	brassinosteroid biosynthetic process						
GO:0045962	positive regulation of development, heterochronic						

Table 3.6. Lowest-level GO terms shared by five or less TF expression cluster GO networks. The ID and description of each of the lowest-level child GO terms within the networks of enriched GO terms for each TF expression clusters were isolated, and those shared by fewer than six clusters were grouped. Columns U1 to D2 correspond to TF clusters U1 to D2.

Transcription factors expressed during early dehydration

Two clusters contained TFs that were expressed or upregulated during early dehydration in X. *humilis* (defined for the purposes of this section as > 40% RWC) compared to hydrated tissue: namely, cluster U1, at 80% RWC, and Cluster U2, at 60% RWC (Fig. 3.33). The majority of these genes were down-regulated after 60% RWC, although a number of TFs in cluster U1 were up-regulated at 40% RWC as well. Total fold change and average transcript abundance of the TFs in these clusters were marginal compared to many of the TFs found in clusters specific to late desiccation, which often underwent very large changes in expression levels (Fig. 3.35A&B).

Cluster U1 contained the fewest GO terms, with only unique terms related to anther dehiscence, double fertilisation, blue-light signalling and leaf shaping enriched (Table 3.4). Cluster U2, in comparison, contained a large number of unique GO terms that were markedly different from U1, suggesting that the transcriptional changes at these two stages of dehydration involve very different physiological responses. Interestingly, GO terms related to negative regulation of gene expression, RNA-mediated chromatin silencing and regulation of translation were enriched in this cluster, which is consistent with massive transcriptional and epigenetic changes occurring during this stage of dehydration. Multiple terms associated with cell wall biogenesis, membrane lipids and cytoskeleton structures were also enriched (Table 3.4).

Many of the putative *X. humilis* TFs in these clusters were most similar to genes with currently little to no direct functional characterisation in other plant species, making it difficult to infer their specific role (if any) in the dehydration response. The genes that show the highest peak fold-change in their respective clusters (at 80% and 60% RWC) are shown

in Figure 3.35. Of these genes, cluster U1 contained a number of AP2/ERF-domain TFs (such as *ERF25*, *TINY/ERF40*, *APETALA2*) and an AP2/B3-like TF (Fig. 3.35A). Additionally, a predicted DELLA protein, with similarity to the maize protein *DWARF8*, was found in this cluster, as was one of the few *HsfB* family HSFs identified in the *X. humilis* transcriptome (*XhHSFB2*, see Fig. 3.36C below). Cluster U2 was dominated by TFs from the WRKY and homeobox-leucine zipper (HOX) families (Fig. 3.35B).

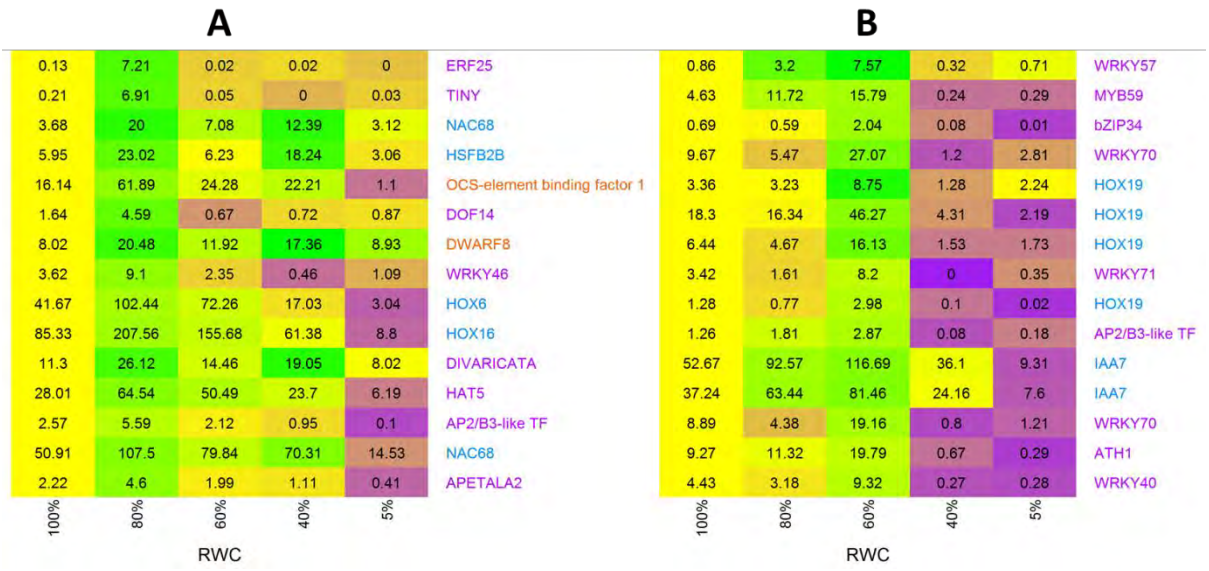


Figure 3.35: Most highly induced TF genes during early dehydration in *X. humilis*. A) Most highly up-regulated genes at 80% RWC (cluster U1), B) most highly up-regulated genes at 80% RWC (cluster U2). Gene names are the top BLASTX hit from the Swissprot database, colour coded by plant species of the top hit (purple: *A. thaliana*, blue: *O. sativa*; orange: *Z. mays*). Cell values represent FPKM, cell colour denotes the similarity to transcript abundance at 100% RWC.

Transcription factors expressed during late dehydration (40%/5% RWC)

Clusters U3 and U4 corresponded to TFs whose expression peaked at 40% RWC and 5% RWC respectively. However, a large number of TFs in Cluster U3 also showed a lesser increase in expression at 80% RWC (Fig. 3.33).

Cluster U3 contained only a few unique enriched GO terms, including phosphatifylinositol 3'-kinase signalling, sucrose-mediated signalling and transcription in response to unfolded protein. Cluster U4 was enriched for multiple terms related to mRNA processing and protein ubiquitination; response to endogenous signals, such as brassinosteroid and trehalose, and terms associated with development, specifically root meristem and aerenchyma (Table 3.4). However, both clusters also specifically shared a large number of GO terms that were not found in any other clusters, mostly related to cellular response to various stresses such cold/freezing, high light and hydrogen peroxide (Table 3.5).

An analysis of some of the individual TFs within the U3 and U4 clusters also revealed a large similarity in the TF families that were expressed. For example, both clusters contained many HSF transcription factors. However, cluster U3 contained predominantly HSFs from the *HsfC* family as defined in the literature (based on comparison to *A. thaliana* and *O. sativa* HSF protein sequences) whereas cluster U4 almost exclusively contained those from the *HsfA* family (Fig. 3.36A-C). ZAT/AZF family protein transcripts were also found spread across clusters U3 and U4 (Fig. 3.37A-C). Transcripts encoding proteins similar to *DEHYDRATION RESPONSE-ELEMENT BINDING 2* (*DREB2*) TFs were present in both clusters, although the vast majority were found in cluster U4 (Fig. 3.37D&E). Similarly, gibberellin-signalling associated *SCARECROW -like* (SCL) TFs were present in both but more

abundant in cluster U4 (Fig. 3.37D&E). A phylogenetic analysis of the *X. humilis* DREB and SCL gene families can be found in Figure S3.9.

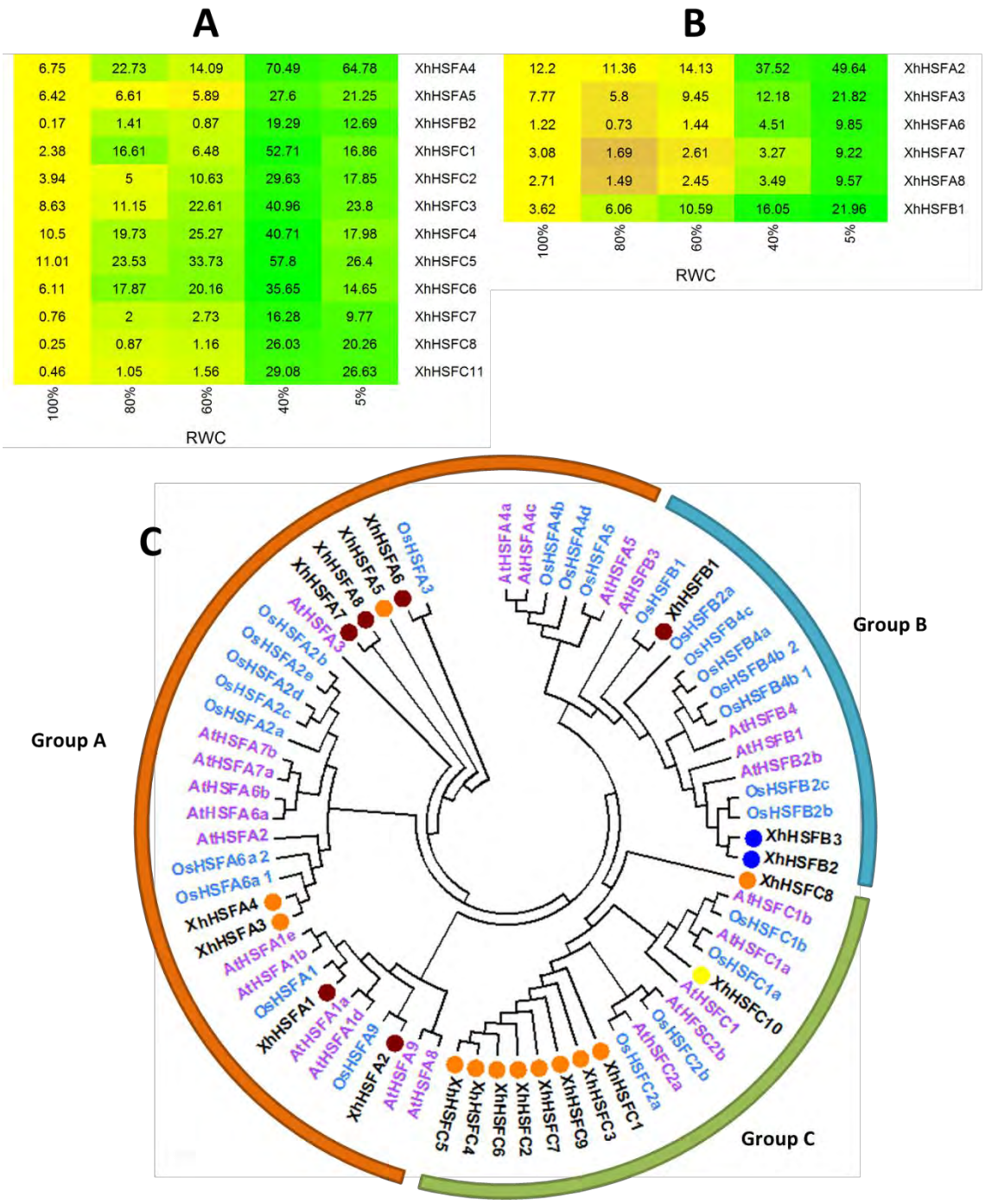


Figure 3.36. Heat shock transcription factors induced during late desiccation. Expression pattern of HSF protein coding transcripts found in cluster U3 (A; induced at 40% RWC) and cluster U4 (B; induced at 5% RWC). C) Bootstrap consensus tree derived from the alignment of predicted full length HSF protein sequences from *X. humilis* (black), *A. thaliana* (purple) and *O. sativa* (blue). HSF protein transcripts from cluster U3 (orange circles) are predominantly clustered with the C-family of HSFs, whereas those from cluster U4 (dark red circles) are from the A-family of HSFs. HSFs found in Cluster U1 and U2 are marked with a blue and yellow circle, respectively.

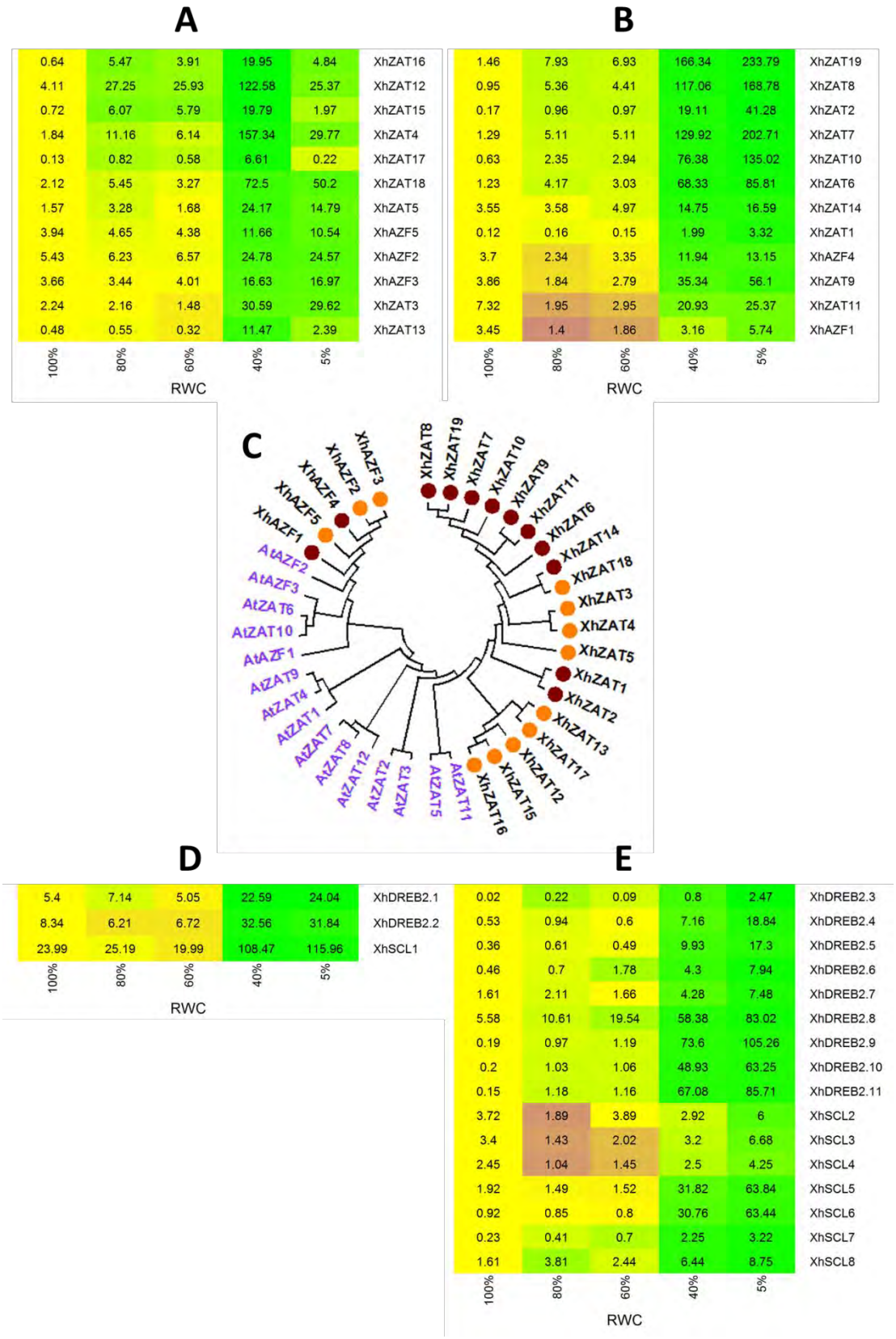

A
0.64 5.47 3.91 19.95 4.84 XhZAT16
4.11 27.25 25.93 122.58 25.37 XhZAT12
0.72 6.07 5.79 19.79 1.97 XhZAT15
1.84 11.16 6.14 157.34 29.77 XhZAT4
0.13 0.82 0.58 6.61 0.22 XhZAT17
2.12 5.45 3.27 72.5 50.2 XhZAT18
1.57 3.28 1.68 24.17 14.79 XhZAT5
3.94 4.65 4.38 11.66 10.54 XhAZF5
5.43 6.23 6.57 24.78 24.57 XhAZF2
3.66 3.44 4.01 16.63 16.97 XhAZF3
2.24 2.16 1.48 30.59 29.62 XhZAT3
0.48 0.55 0.32 11.47 2.39 XhZAT13
100% 80% 60% 40% 5%
RWC
B
1.46 7.93 6.93 166.34 233.79 XhZAT19
0.95 5.36 4.41 117.06 168.78 XhZAT8
0.17 0.96 0.97 19.11 41.28 XhZAT2
1.29 5.11 5.11 129.92 202.71 XhZAT7
0.63 2.35 2.94 76.38 135.02 XhZAT10
1.23 4.17 3.03 68.33 85.81 XhZAT6
3.55 3.58 4.97 14.75 16.59 XhZAT14
0.12 0.16 0.15 1.99 3.32 XhZAT1
3.7 2.34 3.35 11.94 13.15 XhAZF4
3.86 1.84 2.79 35.34 56.1 XhZAT9
7.32 1.95 2.95 20.93 25.37 XhZAT11
3.45 1.4 1.86 3.16 5.74 XhAZF1
100% 80% 60% 40% 5%
RWC
C
XhAZF3
XhAZF2
XhAZF4
XhAZF5
XhAZF1
AtAZF2
AtAZF3
AtZAT6
AtZAT10
AtAZF1
AtZAT9
AtZAT4
AtZAT1
AtZAT7
AtZAT8
AtZAT12
AtZAT2
AtZAT3
AtZAT5
AtZAT11
XhZAT16
XhZAT15
XhZAT12
XhZAT17
XhZAT13
XhZAT2
XhZAT1
XhZAT5
XhZAT4
XhZAT3
XhZAT18
XhZAT14
XhZAT6
XhZAT11
XhZAT9
XhZAT10
XhZAT7
XhZAT19
XhZAT8
D
5.4 7.14 5.05 22.59 24.04 XhDREB2.1
8.34 6.21 6.72 32.56 31.84 XhDREB2.2
23.99 25.19 19.99 108.47 115.96 XhSCL1
100% 80% 60% 40% 5%
RWC
E
0.02 0.22 0.09 0.8 2.47 XhDREB2.3
0.53 0.94 0.6 7.16 18.84 XhDREB2.4
0.36 0.61 0.49 9.93 17.3 XhDREB2.5
0.46 0.7 1.78 4.3 7.94 XhDREB2.6
1.61 2.11 1.66 4.28 7.48 XhDREB2.7
5.58 10.61 19.54 58.38 83.02 XhDREB2.8
0.19 0.97 1.19 73.6 105.26 XhDREB2.9
0.2 1.03 1.06 48.93 63.25 XhDREB2.10
0.15 1.18 1.16 67.08 85.71 XhDREB2.11
3.72 1.89 3.89 2.92 6 XhSCL2
3.4 1.43 2.02 3.2 6.68 XhSCL3
2.45 1.04 1.45 2.5 4.25 XhSCL4
1.92 1.49 1.52 31.82 63.84 XhSCL5
0.92 0.85 0.8 30.76 63.44 XhSCL6
0.23 0.41 0.7 2.25 3.22 XhSCL7
1.61 3.81 2.44 6.44 8.75 XhSCL8
100% 80% 60% 40% 5%
RWC

Figure 3.37. Gene families induced by desiccation stress in *X. humilis*. A&B) Expression pattern of predicted stress-related ZAT/AZF protein coding transcripts found in cluster U3 (A) and U4 (B). C) Bootstrap consensus tree derived from alignment of predicted full-length *X. humilis* and *A. thaliana* ZAT/AZF protein sequences, showing the distribution of these protein-coding transcripts between clusters U3 (orange circles) and U4 (brown circles). D&E) Various *DEHYDRATION RESPONSE ELEMENT BINDING 2* (*DREB2*) and *SCARECROW-like* (*SCL*) transcripts are found in clusters U3 (D) and U4 (E).

Transcription factors downregulated during desiccation

Slightly under half of all TF coding transcripts were downregulated in response desiccation, split between Clusters D1 and D2. TF down-regulation occurred in two consecutive waves, the first occurring in the first measured stage (80% RWC; Cluster D1). Many TFs in Cluster D2 were also transiently downregulated at 80% RWC, but transcript levels dropped to their lowest only after 60% RWC (Fig. 3.33).

Fewer GO terms were uniquely enriched in these clusters in comparison to those up-regulated during desiccation. Cluster D1 contained a small assortment of enriched GO terms, including membrane and ER membrane organisation, response to ozone and salicylic acid biosynthesis (Table 3.4). Cluster D2 had unique GO terms related to red/far-red light phototransduction, negative regulation of gibberellic acid signalling, heterochromatin formation and DNA repair, amongst others (Table 3.4).

Many of the TFs in these clusters currently have little to no functional characterisation. Nonetheless, a number of TF families seemed to occur predominantly in one or both of these clusters. Cluster D1 contained the fewest number of TFs of any cluster, and consequently fewer genes with known functions. Examples of genes with predicted functions included

various genes with similarity to *RADIALIS-like* (*RAD-like*) and *TCP* family TFs, which are often involved in leaf development and patterning (Fig. 3.38A; Corley et al. 2005). Cluster D2 meanwhile contained a large proportion of ethylene-responsive *RAP* and *CONSTANS-like* (*COL*) gene homologs (Fig. 3.38B).

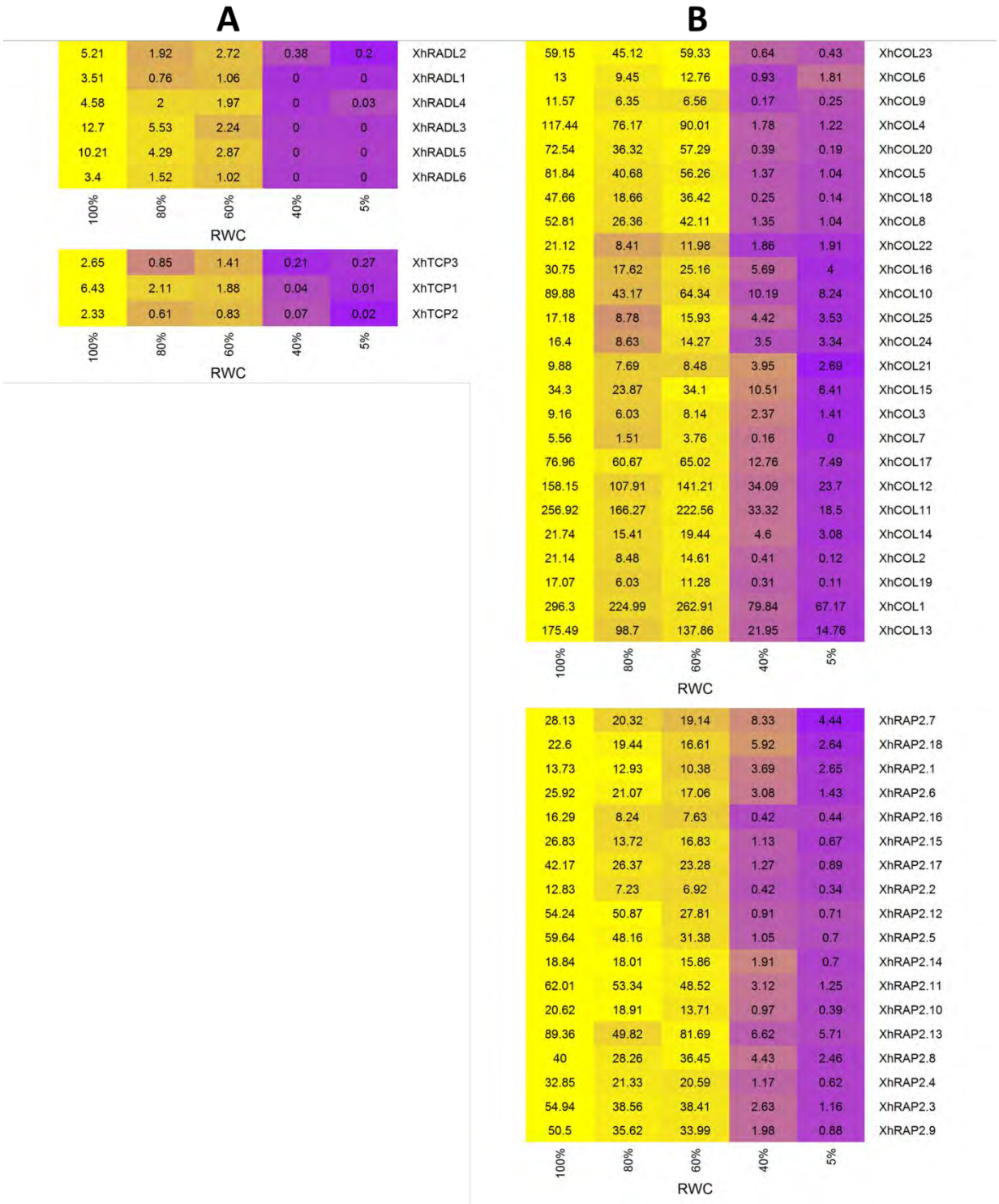


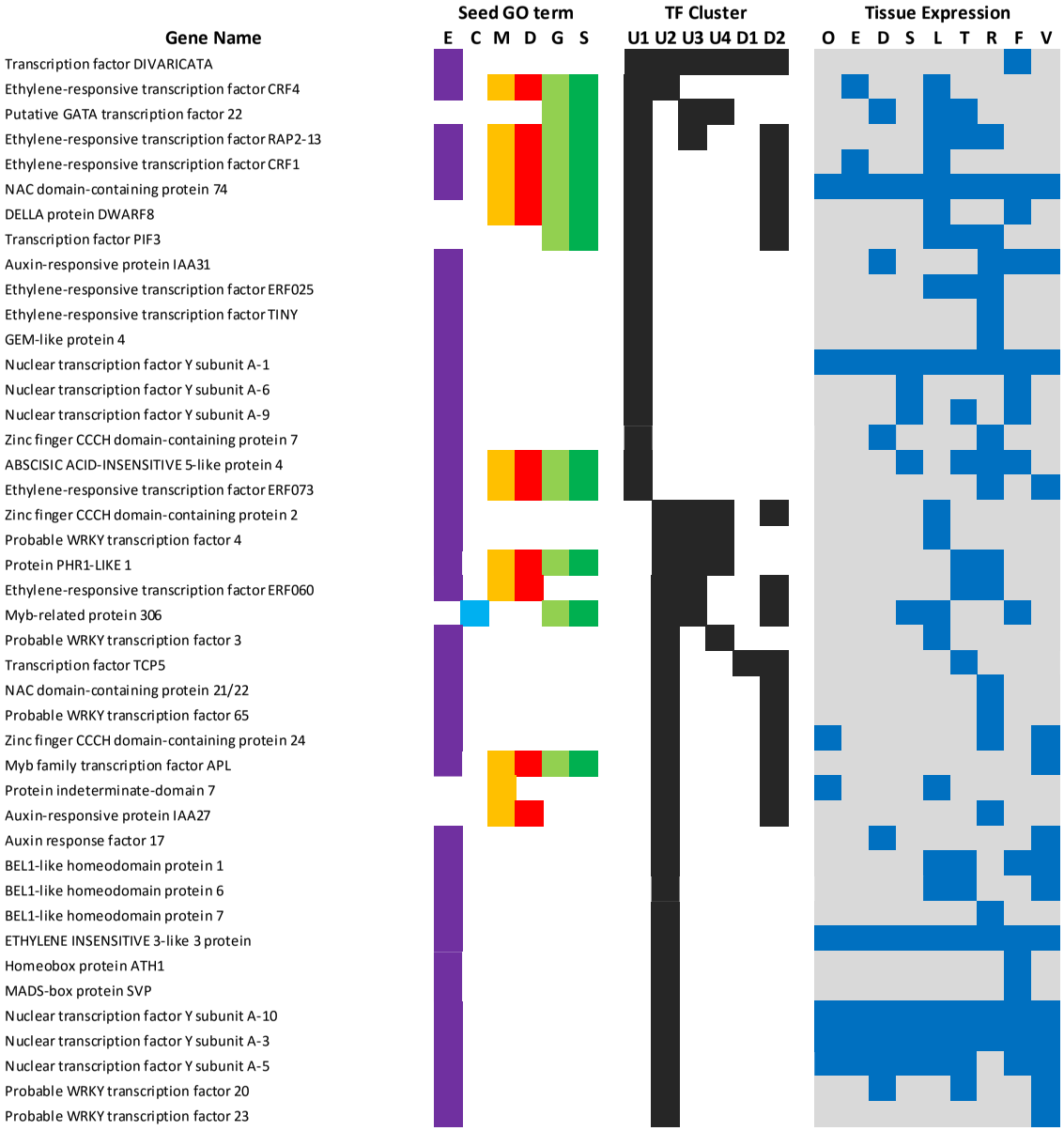
Figure 3.38: Selected transcription factor genes downregulated during dehydration in *X. humilis*. A) Genes down-regulated at 60% RWC (Cluster D1), B) genes down-regulated after 100% RWC (Cluster D2). Cell values represent FPKM, cell colour denotes the similarity to transcript abundance at 100% RWC.

Transcription factors with seed-related functions

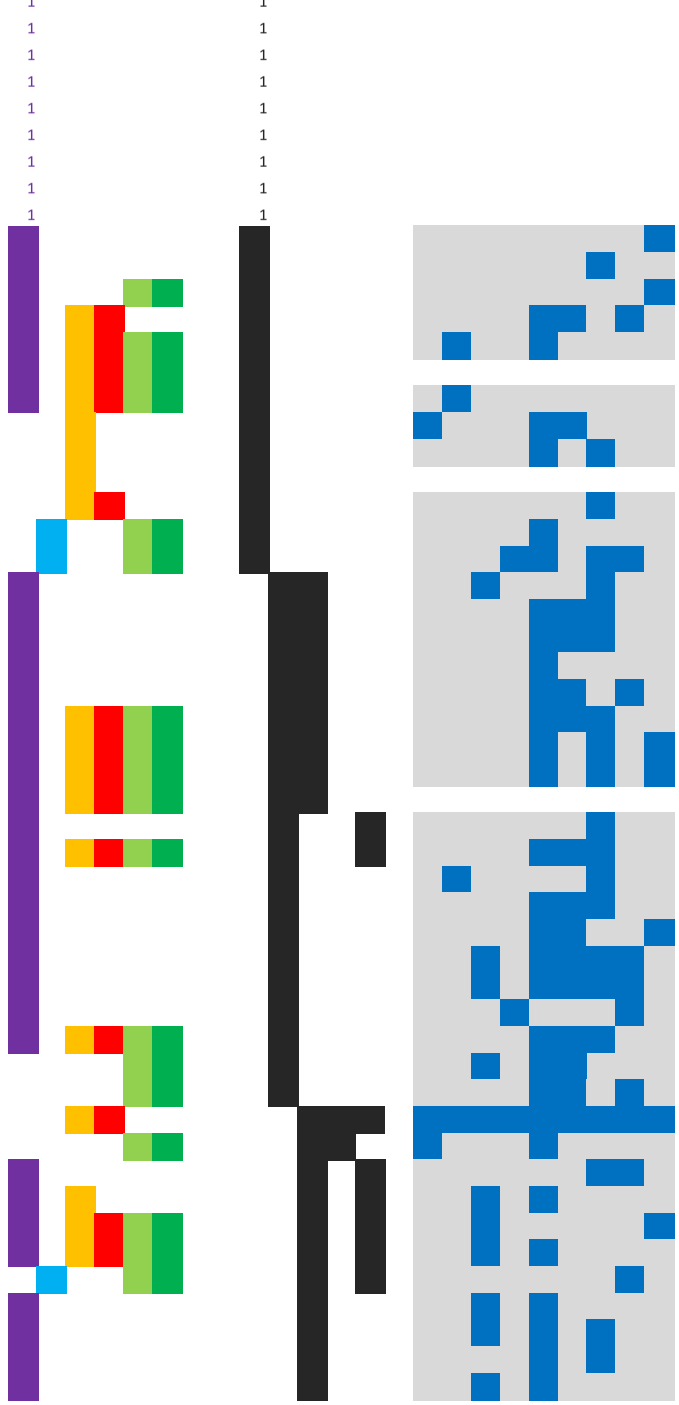
GO terms related to seed development and germination were common to all transcription factor expression clusters (Fig. 3.34). In order to identify whether these terms were associated with genes that are involved specifically in seed development (as opposed to general development or stress response) in other species, all *X. humilis* TF transcripts that contained a seed-related GO term (specifically the seed development child GO terms "embryo development ending in seed dormancy; GO:0009793", "seed coat development; GO:0010214", "seed maturation; GO:0010431", "seed dormancy; GO:0010162", "seed germination; GO:0009845" and "seedling development; GO:0090351", the only ones present in the set) were extracted and the top protein BLAST hits in the Swissprot database were identified. For each top protein hit, the absence or presence of a matching transcript in any of the six TF expression clusters was determined, which could account for possible gene isoforms expressed at different stages of dehydration (Table 3.6; references Table S3.3). Additionally, domains of expression and putative function of each predicted protein were determined from current literature, UniProt, Genevisible or GeneAtlas expression databases. Transcripts encoding proteins with functions outside of seed development would be expected to be present in tissues or developmental stages outside of the seed.

As expected, transcripts with similarity to the same top Swissprot proteins were generally found in the same expression clusters. A few, such as those similar to the TF *DIVARICATA* (Table 3.7, first row), could be found in multiple up- and down-regulated clusters, and so likely encode similar proteins with diversified functions. The majority of identified proteins with similarity to *X. humilis* TF genes containing seed-related GO terms were expressed in

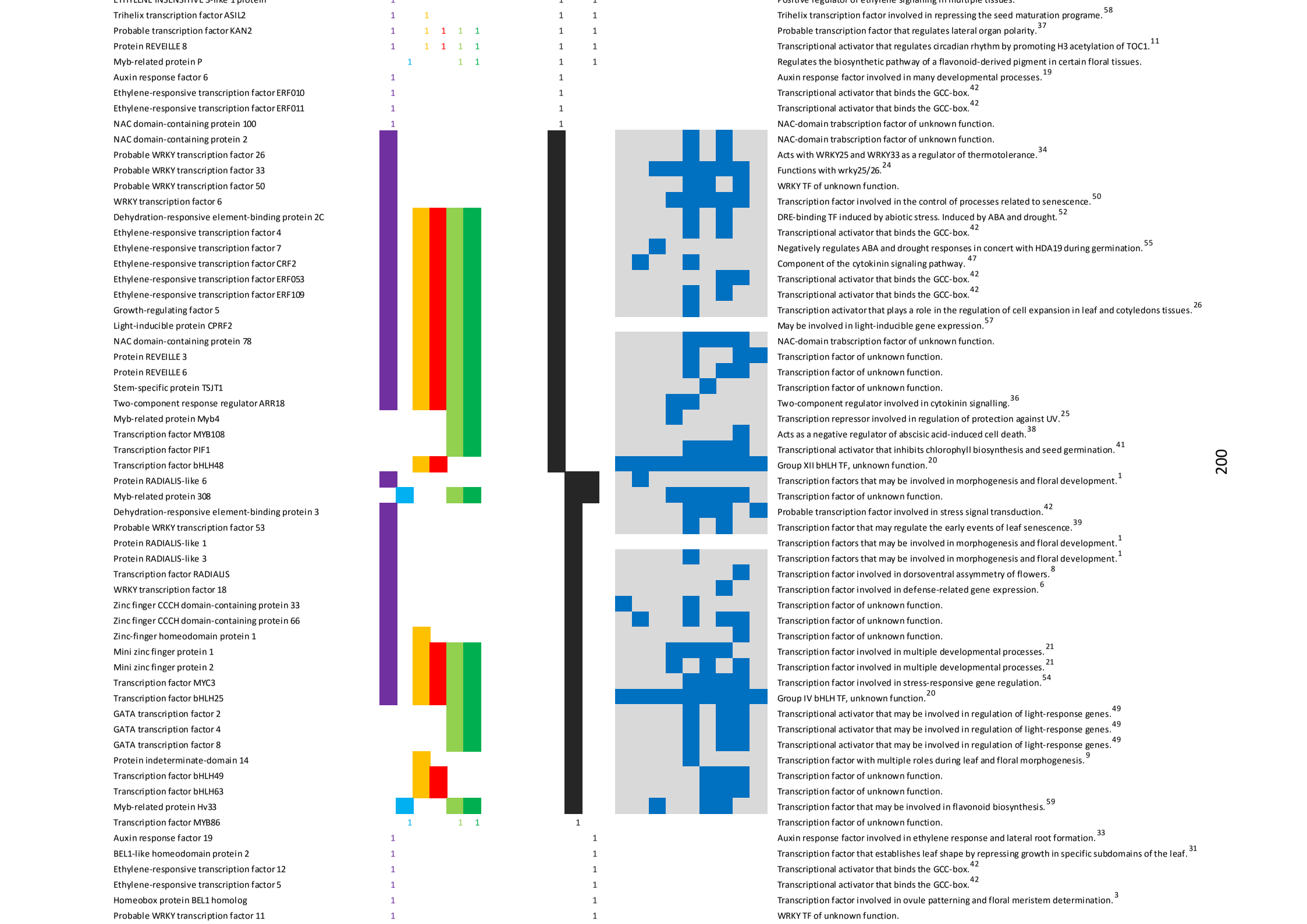
tissues outside of the seed, and had functions associated with stress response or developmental processes in seed and vegetative tissues.



Gene Name	Gene function/tissue specificity
Transcription factor DIVARICATA	Transcription factor involved in flower development.[14]
Ethylene-responsive transcription factor CRF4	Component of the cytokinin signaling pathway.[47]
Putative GATA transcription factor 22	Transcriptional activator that may be involved in regulation of light-response genes.[49]
Ethylene-responsive transcription factor RAP2-13	Transcriptional activator that binds the GCC-box.[42]
Ethylene-responsive transcription factor CRF1	Component of the cytokinin signaling pathway.[47]
NAC domain-containing protein 74	Transcription factor of unknown function.
DELLA protein DWARF8	Transcriptional regulator that acts as a repressor of the gibberellin (GA) signaling pathway.[45]
Transcription factor PIF3	Transcription factor acting positively in the phytochrome signaling pathway.[43]
Auxin-responsive protein IAA31	Short-lived transcriptional factors that function as repressors of early auxin response..[23]
Ethylene-responsive transcription factor ERF025	Transcriptional activator that binds the GCC-box.[42]
Ethylene-responsive transcription factor TINY	ERF family transcription factor of unknown function.[52]
GEM-like protein 4	Transcription factor of unknown function.
Nuclear transcription factor Y subunit A-1	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Nuclear transcription factor Y subunit A-6	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Nuclear transcription factor Y subunit A-9	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Zinc finger CCCH domain-containing protein 7	Transcription factor of unknown function.
ABSCISIC ACID-INSENSITIVE 5-like protein 4	Binds to the ABA-responsive element (ABRE).[7]
Ethylene-responsive transcription factor ERF073	Transcriptional activator that binds the GCC-box.[42]
Zinc finger CCCH domain-containing protein 2	Negative regulator of phytochrome-mediated promotion of seed germination.28
Probable WRKY transcription factor 4	Involved in biotic and abiotic stress response. Induced by infection, SA, JA, ABA and leaf senescence.[32]
Protein PHR1-LIKE 1	Transcription factor acting as central integrator of phosphate starvation responses.[4]
Ethylene-responsive transcription factor ERF060	Transcriptional activator that binds the GCC-box. Expressed in shoot and root.[42]
Myb-related protein 306	Transcription factor of unknown function.
Probable WRKY transcription factor 3	WRKY TF of unknown function. Induced by SA during leaf senescence.[50]
Transcription factor TCP5	Transcription factor involved in shoot organ morphogenesis.[30]
NAC domain-containing protein 21/22	Transcriptional activator that mediates auxin signaling to promote lateral root development.[60]
Probable WRKY transcription factor 65	WRKY TF of unknown function.
Zinc finger CCCH domain-containing protein 24	Transcription factor of unknown function.
Myb family transcription factor APL	Transcription factor required for phloem identity.[2]
Protein indeterminate-domain 7	Transcription factor of unknown function.
Auxin-responsive protein IAA27	Short-lived transcriptional factors that function as repressors of early auxin response.[23]
Auxin response factor 17	Auxin response factor that may help modulate early auxin response.[19]
BEL1-like homeodomain protein 1	Homeodomain transcription factor of unknown function.
BEL1-like homeodomain protein 6	Homeodomain transcription factor of unknown function.
BEL1-like homeodomain protein 7	Homeodomain transcription factor of unknown function.
ETHYLENE INSENSITIVE 3-like 3 protein	Positive regulator of ethylene signalling in multiple tissues.[5]
Homeobox protein ATH1	Transcription factor which may be involved in flower development.[46]
MADS-box protein SVP	Transcriptional repressor that controls the identity of the floral meristem.[16]
Nuclear transcription factor Y subunit A-10	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Nuclear transcription factor Y subunit A-3	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Nuclear transcription factor Y subunit A-5	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Probable WRKY transcription factor 20	WRKY TF of unknown function.
Probable WRKY transcription factor 23	WRKY TF of unknown function.
Probable WRKY transcription factor 40	WRKY TF of unknown function.
NAC domain-containing protein 86	Transcription factor directing sieve element enucleation and cytosol degradation.[13]
B3 domain-containing protein Os07g0563300	B3-domain transcription factor of unknown function.
Ethylene-responsive transcription factor CRF3	Component of the cytokinin signaling pathway.[47]
Pathogenesis-related genes transcriptional activator PTI6	Transcriptional activator that binds the GCC-box.[42]
WUSCHEL-related homeobox 8	Probable transcription factor, which may be involved in embryonic patterning.[18]
Protein indeterminate-domain 2	Transcription factor of unknown function.



BEL1-like homeodomain protein 6	1	1	Homeodomain transcription factor of unknown function.
BEL1-like homeodomain protein 7	1	1	Homeodomain transcription factor of unknown function.
ETHYLENE INSENSITIVE 3-like 3 protein	1	1	Positive regulator of ethylene signalling in multiple tissues.[5]
Homeobox protein ATH1	1	1	Transcription factor which may be involved in flower development.[46]
MADS-box protein SVP	1	1	Transcriptional repressor that controls the identity of the floral meristem.[16]
Nuclear transcription factor Y subunit A-10	1	1	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Nuclear transcription factor Y subunit A-3	1	1	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Nuclear transcription factor Y subunit A-5	1	1	Stimulates the transcription of various genes by recognizing and binding to a CCAAT motif in promoters.[17]
Probable WRKY transcription factor 20	1	1	WRKY TF of unknown function.
Probable WRKY transcription factor 23			WRKY TF of unknown function.
Probable WRKY transcription factor 40			WRKY TF of unknown function.
NAC domain-containing protein 86			Transcription factor directing sieve element enucleation and cytosol degradation.[13]
B3 domain-containing protein Os07g0563300			B3-domain transcription factor of unknown function.
Ethylene-responsive transcription factor CRF3			Component of the cytokinin signaling pathway.[47]
Pathogenesis-related genes transcriptional activator PTI6			Transcriptional activator that binds the GCC-box.[42]
WUSCHEL-related homeobox 8			Probable transcription factor, which may be involved in embryonic patterning.[18]
Protein indeterminate-domain 2			Transcription factor of unknown function.
Zinc finger protein STOP1 homolog			Transcription factor of unknown function.
Zinc finger protein WIP3			Transcription factor of unknown function.
Transcription factor bHLH128			Transcription factor of unknown function.
Transcription factor MYB59			Transcription factor of unknown function.
Transcription factor TRY			Transcription factor oinvolved in trichome and trichoblast development.[53]
Ethylene-responsive transcription factor RAP2-10			Transcriptional activator that binds the GCC-box.[42]
Zinc finger protein AZF1			Inhibits growth under abiotic stress by repressing gene expression.[29]
Zinc finger protein ZAT10			Transcriptional repressor involved in abiotic stress responses.[51]
Zinc finger protein ZAT11			Transcriptional repressor involved in abiotic stress responses.[35]
Zinc finger protein ZAT12			Transcriptional repressor involved in light acclimation, cold and oxidative stress responses.[10]
Dehydration-responsive element-binding protein 2A			DRE-binding TF induced by abiotic stress. Induced by ABA and drought.[52]
Ethylene-responsive transcription factor 11			Transcriptional activator that binds the GCC-box.[42]
Ethylene-responsive transcription factor 8			Transcriptional activator that binds the GCC-box.[42]
Ethylene-responsive transcription factor ERF110			Transcriptional activator that binds the GCC-box.[42]
Probable WRKY transcription factor 51			Positive regulator of salicylic acid negative regulator of jasmonic acid signaling.[15]
Ethylene-responsive transcription factor ERF071			Transcriptional activator that binds the GCC-box.[42]
AP2/ERF and B3 domain-containing protein Os05g0549800			AP2/ERF/B3 transcription factor of unknown function.
NAC domain-containing protein 68			Transcriptional activator that regulates cytokinin signalling during cell division.[27]
Zinc finger CCCH domain-containing protein 30			Transcription factor of unknown function.
Zinc finger CCCH domain-containing protein 35			Transcription factor of unknown function.
Zinc finger protein AZF2			Inhibits growth under abiotic stress by repressing gene expression.[29]
Zinc finger protein ZAT5			Transcription factor of unknown function.
Protein REVEILLE 1			Positive regulator for cold-responsive gene expression, cold tolerance and circadian rhythm.[61]
Protein REVEILLE 2			Morning-phased transcription factor integrating the circadian clock and auxin pathways.[48]
Transcription factor MYB113			Transcription factor of unknown function.
Transcription factor bHLH62			Group XII bHLH TF, unknown function.[20]
GATA transcription factor 5			Transcriptional activator that may be involved in regulation of light-response genes.
ETHYLENE INSENSITIVE 3-like 1 protein			Positive regulator of ethylene signalling in multiple tissues.[5]
Trihelix transcription factor ASIL2			Trihelix transcription factor involved in repressing the seed maturation programe.[58]
Probable transcription factor KAN2			Probable transcription factor that regulates lateral organ polarity.[37]
Protein REVEILLE 8			Transcriptional activator that regulates circadian rhythm by promoting H3 acetylation of TOC1.[11]
Myb-related protein P			Regulates the biosynthetic pathway of a flavonoid-derived pigment in certain floral tissues.
Auxin response factor 6			Auxin response factor involved in many developmental processes.[19]
Ethylene-responsive transcription factor ERF010			Transcriptional activator that binds the GCC-box.[42]
Ethylene-responsive transcription factor ERF011			Transcriptional activator that binds the GCC-box.[42]
NAC domain-containing protein 100			NAC-domain trabscription factor of unknown function.
NAC domain-containing protein 2	1	1	NAC-domain trabscription factor of unknown function.
Probable WRKY transcription factor 26	1	1	Acts with WRKY25 and WRKY33 as a regulator of thermotolerance.[34]
Probable WRKY transcription factor 33	1	1	Functions with wrky25/26.[24]
Probable WRKY transcription factor 50	1	1	WRKY TF of unknown function.
WRKY transcription factor 6	1	1	Transcription factor involved in the control of processes related to senescence.[50]
Dehydration-responsive element-binding protein 2C	1	1	DRE-binding TF induced by abiotic stress. Induced by ABA and drought.[52]
Ethylene-responsive transcription factor 4	1	1	Transcriptional activator that binds the GCC-box.[42]


Trihelix transcription factor ASIL2
Probable transcription factor KAN2
Protein REVEILLE 8
Myb-related protein P
Auxin response factor 6
Ethylene-responsive transcription factor ERF010
Ethylene-responsive transcription factor ERF011
NAC domain-containing protein 100
NAC domain-containing protein 2
Probable WRKY transcription factor 26
Probable WRKY transcription factor 33
Probable WRKY transcription factor 50
WRKY transcription factor 6
Dehydration-responsive element-binding protein 2C
Ethylene-responsive transcription factor 4
Ethylene-responsive transcription factor 7
Ethylene-responsive transcription factor CRF2
Ethylene-responsive transcription factor ERF053
Ethylene-responsive transcription factor ERF109
Growth-regulating factor 5
Light-inducible protein CPRF2
NAC domain-containing protein 78
Protein REVEILLE 3
Protein REVEILLE 6
Stem-specific protein TSJT1
Two-component response regulator ARR18
Myb-related protein Myb4
Transcription factor MYB108
Transcription factor PIF1
Transcription factor bHLH48
Protein RADIALIS-like 6
Myb-related protein 308
Dehydration-responsive element-binding protein 3
Probable WRKY transcription factor 53
Protein RADIALIS-like 1
Protein RADIALIS-like 3
Transcription factor RADIALIS
WRKY transcription factor 18
Zinc finger CCCH domain-containing protein 33
Zinc finger CCCH domain-containing protein 66
Zinc-finger homeodomain protein 1
Mini zinc finger protein 1
Mini zinc finger protein 2
Transcription factor MYC3
Transcription factor bHLH25
GATA transcription factor 2
GATA transcription factor 4
GATA transcription factor 8
Protein indeterminate-domain 14
Transcription factor bHLH49
Transcription factor bHLH63
Myb-related protein Hv33
Transcription factor MYB86
Auxin response factor 19
BEL1-like homeodomain protein 2
Ethylene-responsive transcription factor 12
Ethylene-responsive transcription factor 5
Homeobox protein BEL1 homolog
Probable WRKY transcription factor 11
Positive regulator of ethylene signaling in multiple tissues.
Trihelix transcription factor involved in repressing the seed maturation programe.[58]
Probable transcription factor that regulates lateral organ polarity.[37]
Transcriptional activator that regulates circadian rhythm by promoting H3 acetylation of TOC1.[11]
Regulates the biosynthetic pathway of a flavonoid-derived pigment in certain floral tissues.
Auxin response factor involved in many developmental processes.[19]
Transcriptional activator that binds the GCC-box.[42]
Transcriptional activator that binds the GCC-box.[42]
NAC-domain trabscription factor of unknown function.
NAC-domain trabscription factor of unknown function.
Acts with WRKY25 and WRKY33 as a regulator of thermotolerance.[34]
Functions with wrky25/26.[24]
WRKY TF of unknown function.
Transcription factor involved in the control of processes related to senescence.[50]
DRE-binding TF induced by abiotic stress. Induced by ABA and drought.[52]
Transcriptional activator that binds the GCC-box.[42]
Negatively regulates ABA and drought responses in concert with HDA19 during germination.[55]
Component of the cytokinin signaling pathway.[47]
Transcriptional activator that binds the GCC-box.[42]
Transcriptional activator that binds the GCC-box.[42]
Transcription activator that plays a role in the regulation of cell expansion in leaf and cotyledons tissues.[26]
May be involved in light-inducible gene expression.[57]
NAC-domain trabscription factor of unknown function.
Transcription factor of unknown function.
Transcription factor of unknown function.
Transcription factor of unknown function.
Two-component regulator involved in cytokinin signalling.[36]
Transcription repressor involved in regulation of protection against UV.[25]
Acts as a negative regulator of abscisic acid-induced cell death.[38]
Transcriptional activator that inhibits chlorophyll biosynthesis and seed germination.[41]
Group XII bHLH TF, unknown function.[20]
Transcription factors that may be involved in morphogenesis and floral development.[1]
Transcription factor of unknown function.
Probable transcription factor involved in stress signal transduction.[42]
Transcription factor that may regulate the early events of leaf senescence.[39]
Transcription factors that may be involved in morphogenesis and floral development.[1]
Transcription factors that may be involved in morphogenesis and floral development.[1]
Transcription factor involved in dorsoventral assymmetry of flowers.[8]
Transcription factor involved in defense-related gene expression.[6]
Transcription factor of unknown function.
Transcription factor of unknown function.
Transcription factor of unknown function.
Transcription factor involved in multiple developmental processes.[21]
Transcription factor involved in multiple developmental processes.[21]
Transcription factor involved in stress-responsive gene regulation.[54]
Group IV bHLH TF, unknown function.[20]
Transcriptional activator that may be involved in regulation of light-response genes.[49]
Transcriptional activator that may be involved in regulation of light-response genes.[49]
Transcriptional activator that may be involved in regulation of light-response genes.[49]
Transcription factor with multiple roles during leaf and floral morphogenesis.[9]
Transcription factor of unknown function.
Transcription factor of unknown function.
Transcription factor that may be involved in flavonoid biosynthesis.[59]
Transcription factor of unknown function.
Auxin response factor involved in ethylene response and lateral root formation.[33]
Transcription factor that establishes leaf shape by repressing growth in specific subdomains of the leaf.[31]
Transcriptional activator that binds the GCC-box.[42]
Transcriptional activator that binds the GCC-box.[42]
Transcription factor involved in ovule patterning and floral meristem determination.[3]
WRKY TF of unknown function.
200

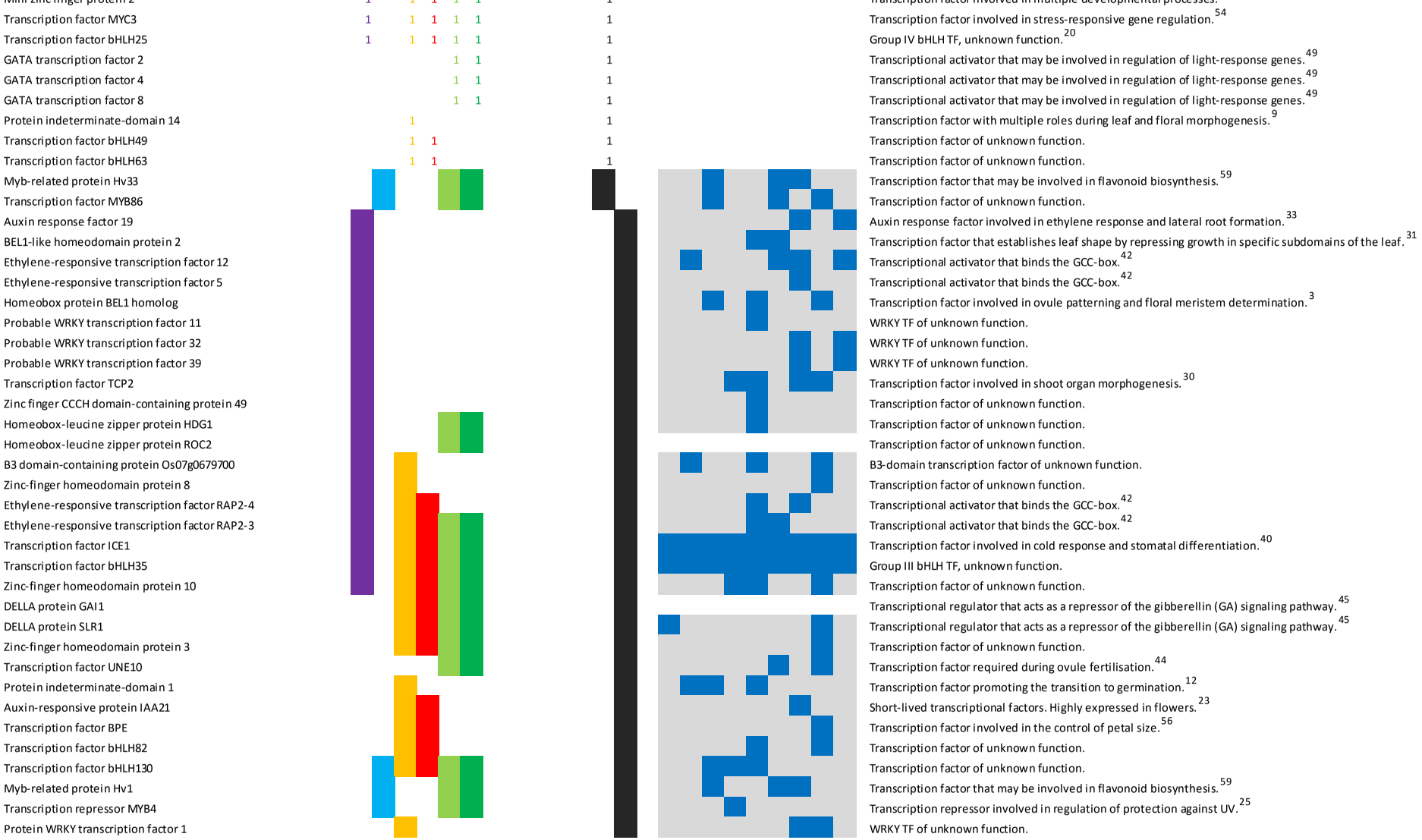


Table 3.7. Top protein hit in the Swissprot BLAST database of *X. humilis* TFs associated with seed-related GO terms. Proteins are marked with a series of boxes, which denote which GO terms (coloured boxes; columns E-S), expression clusters (black boxes; columns U1-D2) or expression pattern (blue boxes; columns O-V) with which they are associated. Columns: E = seed embryo development ending in dormancy; C = seed coat development; M = seed maturation; D = seed dormancy; G= seed germination; S = seedling development. The labels U1 to D2 correspond to Clusters U1 to D2, respectively. Columns: O = ovule; E = embryo; D = seed; S = silique/fruit; L = leaf; T = stem; R = root; F = flower; V = xylem/phloem.

3.3.4.5 Desiccation in Xerophyta is associated with broad changes in chromatin regulation

Epigenetic regulation of gene expression is one of the key methods for regulating transitions between developmental states, which generally involve transcriptional re-programming of a large number of genes simultaneously (Wollmann and Berger, 2015). GO terms associated with epigenetic gene regulation were shown to be enriched in multiple *X. humilis* expression clusters for both DE genes in general and TF-coding genes. We chose to further investigate gene transcripts that were involved in epigenetic activation or silencing of chromatin, to better understand the role this may play during the desiccation response.

Histones

Over forty transcripts annotated as histones were differentially expressed during desiccation in *X. humilis*. Based on sequence similarity and Interpro domain identification, these could be grouped into 10 H1, 8 H2A, 11 H2b, 3 H3 and 10 H4 protein-encoding transcripts. The vast majority of putative differentially expressed histones had a similar expression profile: varying degrees of induction at 60% RWC, followed by a drop in expression to generally well below that seen in hydrated tissue (Fig. 3.39A-F). Only a handful of histone transcripts had expression profiles that differed from the general trend. An H1 variant (*XhH1.9*) was the only histone that showed constant robust upregulation in response to desiccation, increasing nearly 8-fold from 100% to 5% RWC (Fig. 3.39A). A comparison of the predicted protein sequences of the *X. humilis* and *A. thaliana* H1 family revealed that *XhH1.9* was most similar to the *A. thaliana* drought-inducible histone variant, *H1.3* (Fig. 3.39F). Two H2A-coding transcripts showed slightly increased expression during late desiccation: *XhH2A.2* at 40% RWC, and *XhH2A.1* at 5% RWC; similarly, an H3-encoding transcript (*XhH3.1*) was induced

to a low level at 5% RWC (Fig. 3.39B-D). Lastly, an H4 transcript, *XhH4.1*, had increased expression at 40% RWC (Fig. 3.39E).

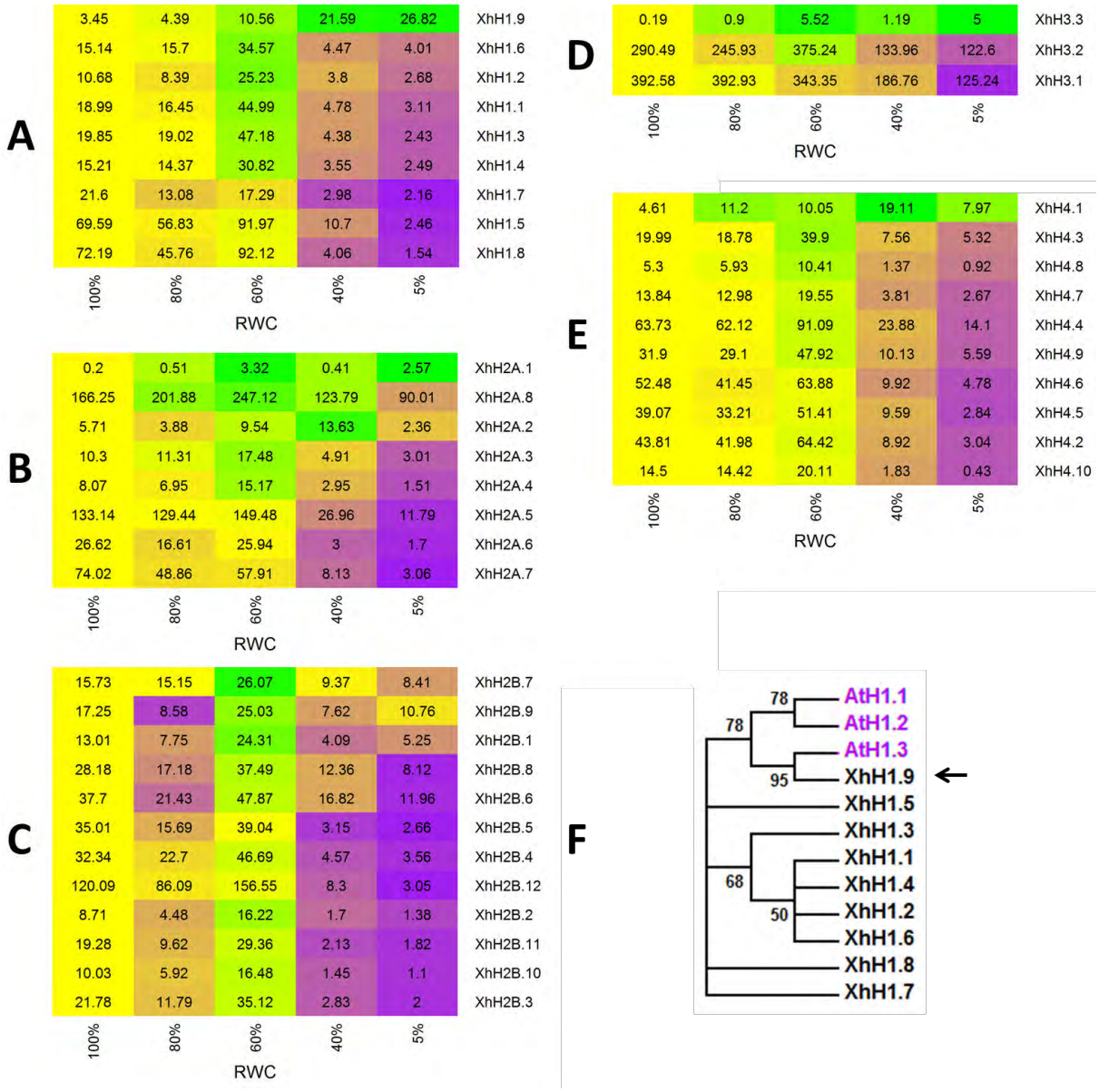


Figure 3.39: Differentially expressed histone-coding transcripts in *X. humilis*. A-E) Expression changes of histone-coding transcripts during desiccation in *X. humilis*. Cell values are FPKM, whereas cell colour denotes the difference in expression compared to 100% RWC. F) Phylogenetic tree showing the evolutionary relationship between *X. humilis* and *A. thaliana* H1 protein sequences. The clade consisting of the drought inducible *AtH1.3* and desiccation-induced *XhH1.9* is marked with an arrow.

Histone-modifying enzyme genes: HAT, HDAC, HMT and HDM

The transition from embryogenesis to seed maturation, and the later transition from quiescent embryo to vegetative growth, are both mediated by epigenetic regulation of gene expression. Of particular importance is PRC2, a PcG complex that inhibits the expression of seed maturation genes through deposition of the repressive H3K27me3 epigenetic mark with the aid of PRC1 (Müller et al., 2012; Yang et al., 2013). In order to investigate the possible roles of PRC2 and other histone modifying enzymes during desiccation, known chromatin-modifying genes were identified in the *X. humilis* transcriptome. Because histone modification factors tend to be large complexes of multiple proteins with various functions, further analysis was focused on genes associated with histone acetylation/methylation directly, with further emphasis on the *PRC2* and *PRC1* complexes themselves.

Histone modifying enzymes that were identified as differentially expressed in *X. humilis* could be grouped into four classes, with a varying number of different gene transcripts for each class: histone methyltransferases (HMT; 21), demethylases (HDM; 4), acetyltransferases (HAT; 7) and deacetylases (HDAC; 10). Interestingly, the majority of all histone modifying enzyme classes displayed a similar expression profile - a sudden decrease in transcript abundance after 60% RWC - irrespective of histone or modification specificity (Fig. 3.40 & 3.41). Only a few up-regulated genes displayed an expression profile that did not follow this trend.

Predicted HMTs were the most abundant of the identified histone-modifying enzyme genes. However, only one was slightly up-regulated during desiccation (*XhHMT20*), while the remainder were expressed in hydrated leaves and consistently down-regulated after 60% RWC (Fig. 3.40A). A phylogenetic analysis of the *X. humilis* and *A. thaliana* full-length HMT

protein sequences suggests that *XhHMT20* is most similar to the Class V SET-domain group HMTs (e.g. the SUVH group of *A. thaliana*), which mediate deposition of the H3K9 methylation epigenetic mark (Fig. S3.10A). Similarly, only one of the four identified HDM-encoding transcripts was up-regulated during desiccation (*XhHDM3*; Fig. 3.40C). *XhHDM3* was found to be similar to HDMs of the JMJD2/KDM4 family, which preferentially target H3K9 sites for demethylation (Fig. S3.10B).

Fewer HAT protein genes were identified in *X. humilis*, although a greater number were up-regulated in desiccating tissue (Fig. 3.41A). In total, four HAT gene transcripts were induced to varying degrees during late desiccation. Three of these (*XhHAT5, XhHAT6* and *XhHAT7*) were predicted to encode CPB family HAT proteins based on similarity to *A. thaliana* HAT protein sequences, whilst the fourth (*XhHAT4*) encoded a MYST family HAT protein, similar to *AtHAM1* and *AtHAM2* (Fig. 3.41B). Two of the ten identified HDAC genes were likewise up-regulated (*XhHDA3* and *XhHDA4*). Both encoded RPD3/HDA1-family HDAC proteins: *XhHDA3* had similarity to class I RPD3/HDA1 proteins, such as *AtHDA19* and *AtHDA6*, whereas *XhHDA4* had similarity to those from class II, such as *AtHDA5* and *AtHDA18* (Fig. 3.41C&D).

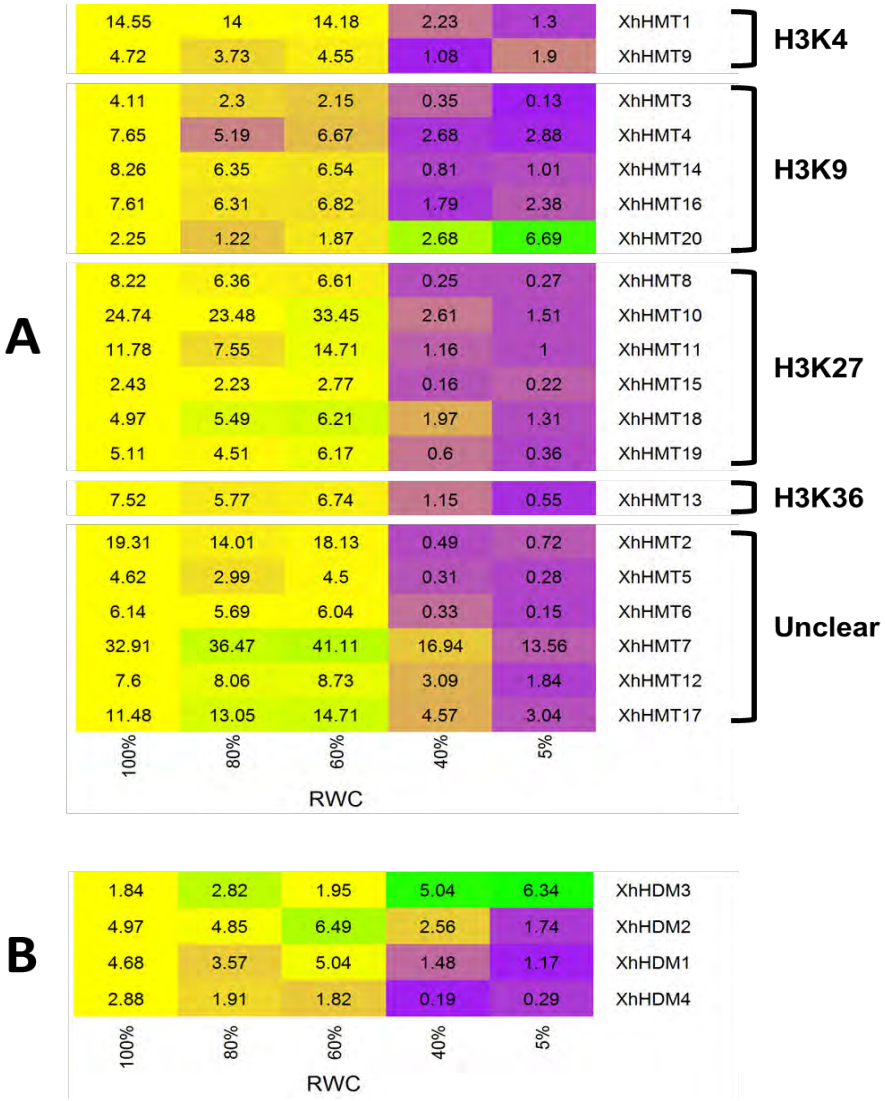


Figure 3.40. Differentially expressed histone methyltransferase and demethylase enzyme genes in *X. humilis*. A) Expression of 21 putative HMT genes, as well as their predicted specificity as based on a phylogenetic comparison to known HMTs from other species (see Supplementary Figure S3.10A). C) Expression of *X. humilis* HDM proteins.

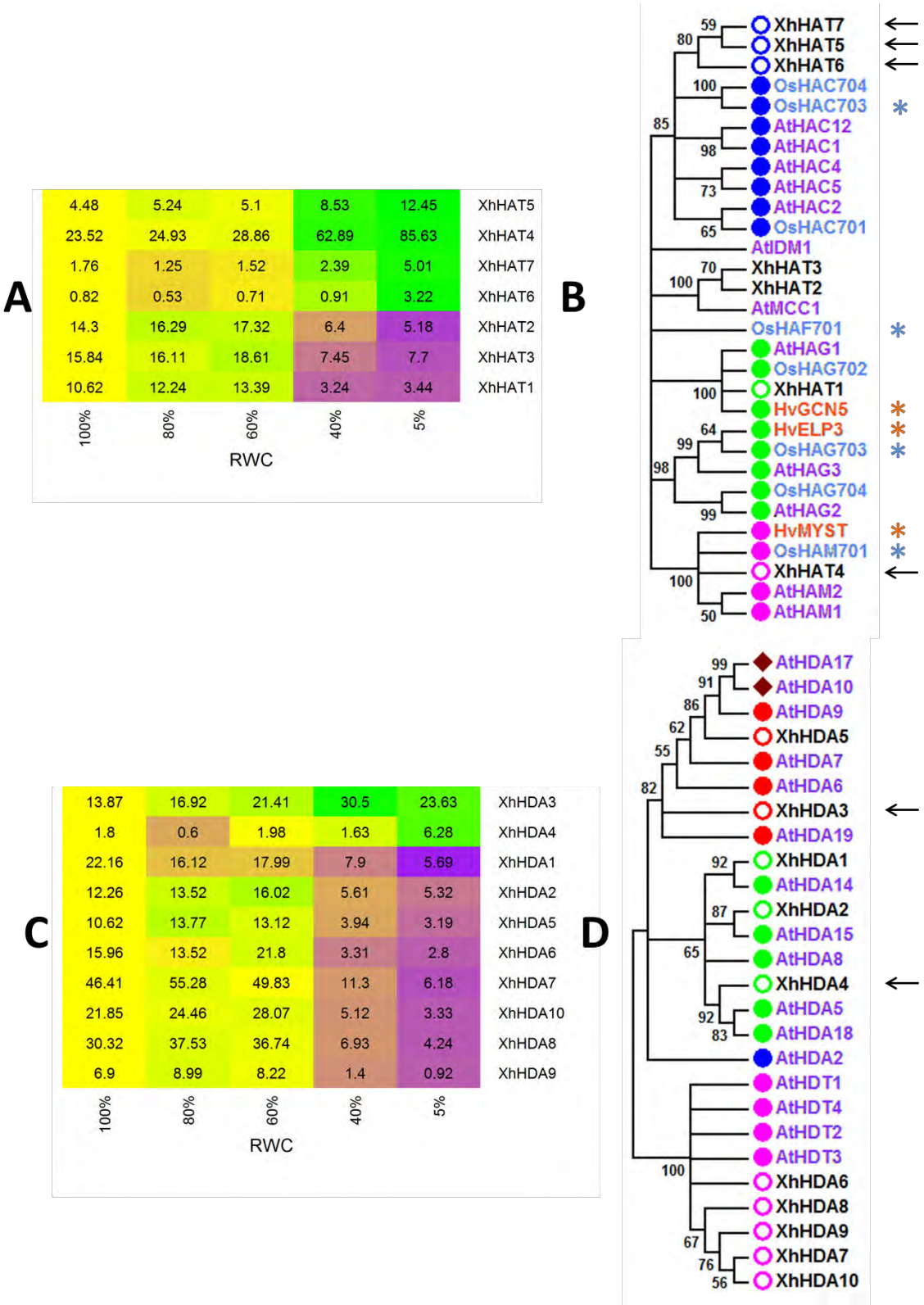

A
4.48 5.24 5.1 8.53 12.45 XhHAT5
23.52 24.93 28.86 62.89 85.63 XhHAT4
1.76 1.25 1.52 2.39 5.01 XhHAT7
0.82 0.53 0.71 0.91 3.22 XhHAT6
14.3 16.29 17.32 6.4 5.18 XhHAT2
15.84 16.11 18.61 7.45 7.7 XhHAT3
10.62 12.24 13.39 3.24 3.44 XhHAT1
100% 80% 60% 40% 5%
RWC
B
XhHAT7
XhHAT5
XhHAT6
OsHAC704
OsHAC703
AtHAC12
AtHAC1
AtHAC4
AtHAC5
AtHAC2
OsHAC701
AtIDM1
XhHAT3
XhHAT2
AtMCC1
OsHAF701
AtHAG1
OsHAG702
XhHAT1
HvGCN5
HvELP3
OsHAG703
AtHAG3
OsHAG704
AtHAG2
HvMYST
OsHAM701
XhHAT4
AtHAM2
AtHAM1
C
13.87 16.92 21.41 30.5 23.63 XhHDA3
1.8 0.6 1.98 1.63 6.28 XhHDA4
22.16 16.12 17.99 7.9 5.69 XhHDA1
12.26 13.52 16.02 5.61 5.32 XhHDA2
10.62 13.77 13.12 3.94 3.19 XhHDA5
15.96 13.52 21.8 3.31 2.8 XhHDA6
46.41 55.28 49.83 11.3 6.18 XhHDA7
21.85 24.46 28.07 5.12 3.33 XhHDA10
30.32 37.53 36.74 6.93 4.24 XhHDA8
6.9 8.99 8.22 1.4 0.92 XhHDA9
100% 80% 60% 40% 5%
RWC
D
AtHDA17
AtHDA10
AtHDA9
XhHDA5
AtHDA7
AtHDA6
XhHDA3
AtHDA19
XhHDA1
AtHDA14
XhHDA2
AtHDA15
AtHDA8
XhHDA4
AtHDA5
AtHDA18
AtHDA2
AtHDT1
AtHDT4
AtHDT2
AtHDT3
XhHDA6
XhHDA8
XhHDA9
XhHDA7
XhHDA10

Figure 3.41. Differentially expressed histone acetyltransferase and deacetylase enzyme genes in *X. humilis*. A) Predicted HAT genes, B) Phylogenetic analysis of predicted *X. humilis* and *A. thaliana* HAT proteins. The major families in *A. thaliana* are identified by fully coloured circles (blue: CPB; pink: MYST; green: GNAT). Open circles represent predicted *X. humilis* HDACs that cluster with one of these families. HAT genes up-regulated in *X. humilis* are marked with an arrow, whereas those up-regulated by stress or ABA in other species are marked with an asterisk. C) HDAC proteins, D) Phylogenetic analysis of predicted *X. humilis* and *A. thaliana* HDAC proteins. The major HDAC families in *A. thaliana* (red: RPD3/HDA1 class I; green: RPD3/HDA1 class II; blue: RPD3/HDA1 class III; pink: HD2; brown diamonds: non-functional HDACs). Open circles represent predicted *X. humilis* proteins that cluster with one of these families, and those that are further up-regulated during desiccation are marked with an arrow.

PRC1 and PRC2 gene expression

Genes corresponding to each of the major plant PRC1 and PRC2 subunits could be identified in *X. humilis* based on BLASTX and BLASTP analysis. Several transcripts encoded proteins with similarity to the known plant PRC1 components (LHP1, BMI1/DRIP and RING1; Fig. 3.42A). With the exception of *XhLHP1*, which was slightly down-regulated during desiccation, all PRC1-related transcripts were not differentially expressed (Fig. 3.42B).

I also identified multiple transcripts with similarity to PRC2 subunit genes. Su(z) and ESC were represented by only a single gene each, *XhEMF2* and *XhFIE*. Three gene clusters (*XhHMT8*, *XhHMT10* and *XhHMT11*) were predicted to code for E(z) orthologues, and seven had similarity to PRC2 p55-subunit MSI proteins (Fig. 3.42C; Fig. S3.11). With the exception of *XhFIE* and *XhMSI4*, all identified PCR2 gene transcripts were heavily downregulated during late desiccation (Fig. 3.42D).

A

Gene Cluster	BLAST hit	E-value	Gene ID
XhLHP1	LHP1	4E-25	At5g17690
	LHP1*	8E-51	Os10g0324900
XhRING1.1	RING1a	1E-80	At5g44280
	RING1b	1E-68	At1g03770
XhRING1.2	RING1a	5E-61	At5g44280
	RING1b	3E-56	At1g03770
XhDRIP1	DRIP2	1E-95	At2g30580
	DRIP1	1E-91	At1g06770
XhDRIP2	DRIP2	4E-129	At2g30580
	DRIP1	7E-120	At1g06770
XhDRIP3	DRIP2	1E-132	At2g30580
	DRIP1	1E-121	At1g06770
XhDRIP4	DRIP2	4E-128	At2g30580
	DRIP1	7E-120	At1g06770

B

Gene Cluster	BLAST hit	E-value	Gene ID
XhMSI3	MSI1	0.0	At5g58230
XhMSI4	MSI1	0.0	At5g58230
XhMSI5	MSI1	0.0	At5g58230
XhMSI6	MSI1	0.0	At5g58230
XhMSI7	MSI1	0.0	At5g58230
XhMSI1	MSI4	0.0	At2g19520
XhMSI2	MSI4	0.0	At2g19520
XhEMF2	EMF2	5E-40	At5g51240
XhHMT10	CLF/MEA/SWN	0.0	-
XhHMT11	CLF/MEA/SWN	0.0	-
XhHMT8	CLF/MEA/SWN	0.0	-
XhFIE	FIE	0.0	At3g20740

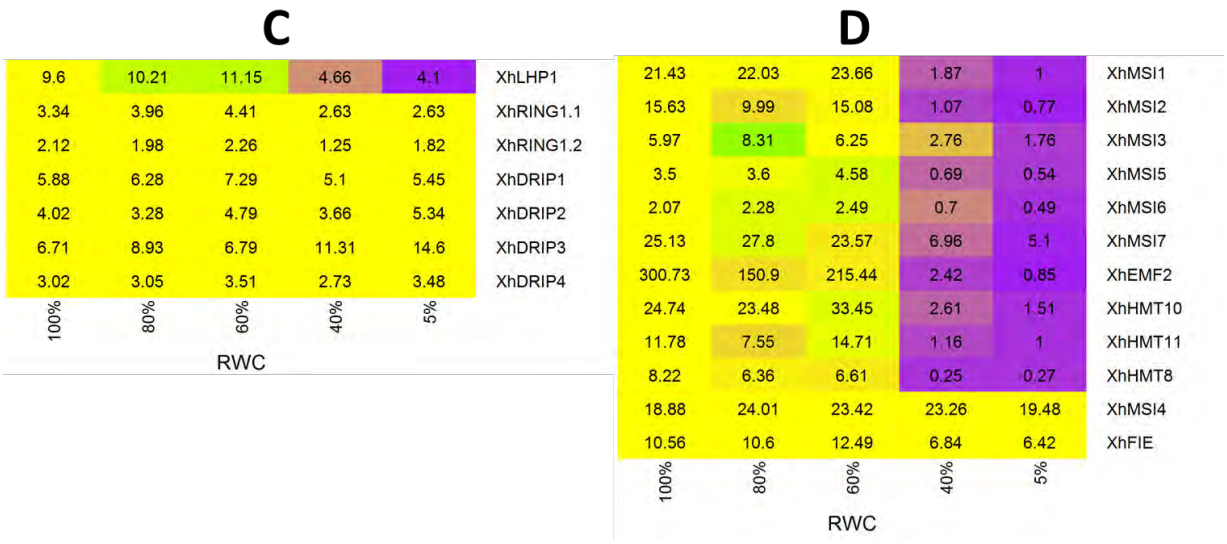


Figure 3.42. PRC1- and PRC2-related genes expressed in *X. humilis*. A&B) *X. humilis* gene clusters with similarity to *A. thaliana* PRC1 and PRC2 protein sequences in the Swissprot database. The putative PRC1 component gene *XhLHP1* also shows the comparison to the rice *LHP1* orthologue. The *X. humilis* HMT genes *XhHMT8, XhHMT10* and *XhHMT11* were equally similar to all three *A. thaliana* E(z) homologues (e-value = 0). C) Expression of *X. humilis* PRC1 subunit transcripts during desiccation; only the putative *XhLHP1* gene is DE. D) With the exception of *XhFIE* and *XhMSI4*, all PRC2 subunit genes were significantly down-regulated after 60% RWC.

3.3.4.6 Response of co-sequenced fungi to desiccation in Xerophyta humilis

Despite the large proportion of initial putative transcripts that were predicted to be fungal, only 203 were found to be DE at our selected cut-offs. Due to the low levels of fungal sequence contamination in comparison to the plant tissue (Fig. 3.10A), and the uncontrolled

nature of these sequences (sample and library size are inconsistent, and not accounted for by DESeq2 testing), these values should be considered statistically unreliable. Nonetheless, they were analysed out of interest.

Fungal genes were clustered into four expression profiles via k-means clustering, three of which would be classified as up-regulated during desiccation and contained the vast majority of transcripts (Fig. S3.2). Nearly half of the DE fungal sequences (80/203) coded for 40S- and 60S-ribosomal proteins, which were for the most part upregulated in the later stages of desiccation (Fig. 3.43A). A large portion of the remaining sequences could not be matched to known proteins (46/203). At least one of these could be classified as a putative LEA/dehydrin based on an Interpro domain search and, as would be expected, was highly upregulated at various stages of desiccation. Other upregulated transcripts included those with similarity to heat shock proteins and a trihydrophobin (Fig. 3.43A). Downregulated genes included tubulin-α and elongation factor-α, and various other ribosomal proteins (Fig. 3.43A).

There were no DE genes detected that were obviously involved in fungal pathogenesis. However, a number of *X. humilis* DE genes were associated with response to infection, fungal or otherwise. Transcripts encoding proteins similar to common plant anti-fungal proteins, such osmotin, defensin, zeamatin, various chitinases, pathogenesis-related proteins, and plant defence transcription factors - such genes similar to *WRKY6* - were upregulated in *X. humilis* during desiccation (Fig. 3.43B). However, transcripts of similar genes were often down-regulated as well (Fig. 3.43B), and many of these genes serve dual functions as general stress response factors.

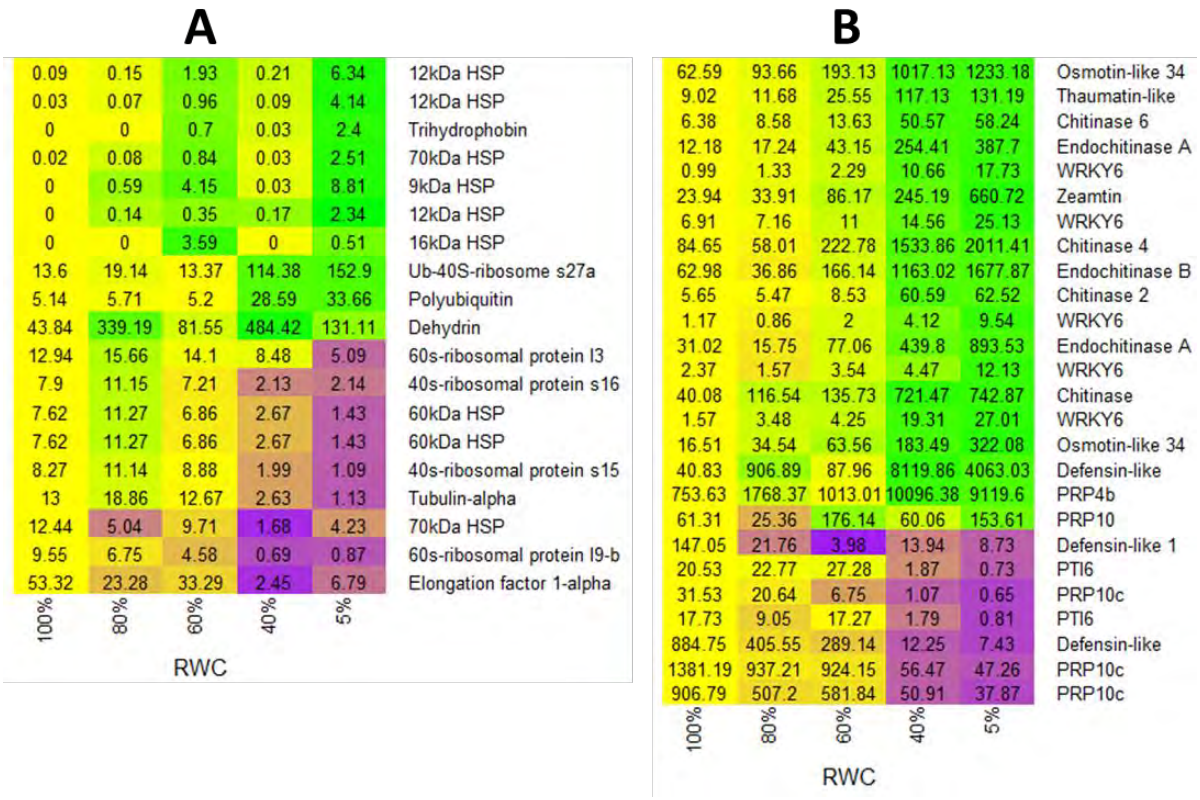


Figure 3.43: Fungal contamination of wild *X. humilis* plants. Some differentially expressed genes of fungal origin (A) and plant-derived anti-fungal genes (B), as determined by the top protein hit for these genes in the Swissprot BLAST database, in the *X. humilis* desiccation transcriptome assembled in this study. Cell values are FPKM, whereas cell colour corresponds to difference in transcript abundance relative to 100% RWC.

3.3.5 Discussion

3.3.5.1 De novo assembly of an X. humilis reference transcriptome

A reference transcriptome of the coding sequences of the poikilochlorophyllous resurrection plant, *X. humilis* was successfully assembled. The transcriptome was assembled using pooled strand-specific RNA-Seq reads sequenced from total RNA extracted at five stages of dehydration: 100%, 80%, 60%, 40% and 5%. As such, it is likely that most of the transcripts that are expressed in leaves during a desiccation cycle in this species have been identified. Using a combination of multiple assemblers and assembly parameters, the Evidential Gene pipeline, and Corset, 76,868 putative gene clusters (genomic loci that are likely to represent a single "gene") were identified. The common assembly merging approach, represented by the Transabyss-merge pipeline, resulted in a massive increase in transcript numbers when used on *X. humilis* data, rather than the expected decrease due to removal of transcripts that had assembled multiple times. This suggests a higher than expected rate of SNP or insertion/deletions in the assembly, which prevented the merging of similar transcripts. It is unclear whether this is due to accumulated assembly errors or whether it represents a biological characteristic unique to *X. humilis*, such as a higher than average mutation rate. The latter explanation is partially supported by data derived from a draft *X. humilis* genome assembly, and is possibly related to the repeated cycles of desiccation and associated DNA damage encountered by these plants over the course of their lifespan (Schlebusch & Illing, personal correspondence). This problem was largely alleviated by using the EviGene merging pipeline which focuses on merging transcripts based on shared coding sequence, indicating that the errors were likely concentrated in non-coding/chimeric regions or contig

ends. As with any *de novo* assembly, it is nonetheless important to verify that the assembled transcripts exist in a biological system, for example by PCR.

Approximately 80% of sequencing reads could be mapped back to the reference transcriptome using Bowtie2, indicating that the sequencing reads were well represented within the reference assembly. Annotation was achieved using a combination of a local installation of NCBI blast and the Swissprot/Uniref90/nr/nt databases and the Blast2Go Pro package. Approximately 60% of the transcriptome could be confidently annotated using Blast2GO, and another 21% was partially annotated based on homology to known proteins in the Swissprot or Uniref90 databases. Only 19% of the transcriptome had no match to known proteins, suggestive of a good quality dataset. *X. humilis* gene clusters that were differentially expressed during desiccation were identified using DESeq2, and those that fell within our selected cut-offs (FDR < 0.01, FPKM >= 1, FC >= 2) accounted for 24% of the transcriptome (18,534 gene clusters).

Differentially expressed genes were clustered into 16 total expression clusters using MeV. Gene expression during desiccation could be divided into four main profiles: 1) A comparatively small proportion of genes (2,380) were responsive (either down- or up-regulated) to the initial stages of desiccation, and thus may contain genes important for the early dehydration response (80% RWC; Clusters 1, 2, 3, 4 and 11); 2) A very large number of genes (8, 810; 47.5% of all DE genes) were heavily down-regulated after 60% RWC compared to their levels in hydrated leaves (Clusters 6, 12, 13, 15 and 16), suggesting that many of these genes are not important for the late-stage desiccation response, or act as repressors of the desiccation response. 3) Another large subset of genes (6, 385; 34.5% of all DE genes) were subsequently up-regulated after 60% RWC (Clusters 7, 8, 9 and 10) and may be

important for stress response at low water contents, or accumulated for rehydration. 4) Lastly, a small number of genes (760) showed a "biphasic" response in this dataset, as they were up-regulated during early (80% RWC) and late (40% and 5% RWC) desiccation, but did not show highly elevated expression levels at 60% RWC (Clusters 3 and 4). This subset contained many genes that are associated with drought and desiccation tolerance in other plant species.

The measured changes in gene expression were validated by comparison to a previous microarray study that investigated the changes in transcripts abundance during a desiccation cycle in *X. humilis* leaves at 100%, 80%, 60%, 40%, 20% and 5% RWC (Shen, 2014). This ultimately facilitated the comparison of the expression of 1,337 transcripts between the two studies. Although there was in general a positive correlation in gene expression between this study and the microarray study, as expected, the gene expression of transcripts at 60% RWC was not correlated between the two studies (Fig. 3.11A). A PCA plot does not reveal any irregularities with these samples: the RNA-Seq 60% RWC libraries cluster together, and are clearly more similar to samples from hydrated tissues than desiccated tissues (Fig. 3.10). This is similar to the PCA analysis of the microarray biological replicates, although they are arguably more equidistant from samples derived from hydrated and desiccated tissues (Shen, 2014). There is some evidence that this could be due to differences in the measured RWC of leaves at 60% RWC in each study. As described in Figure 3.12, many of the expression differences between the transcripts can be explained by reasoning that the leaves measured to be 60% RWC in this study are at a higher RWC than those measured as 60% RWC in the Shen study (Fig. 3.12A&B). This is further supported by the inter-RWC sample correlation values, which show that the measured gene expression at

60% RWC in this RNA-Seq study most closely correlates to gene expression at 100% and 80% RWC in the microarray (i.e. higher than 60% RWC), whereas gene expression at 60% RWC in the Shen microarray study most closely correlates with the gene expression at 40% RWC in this study (Fig. 3.11B). This effect would be predicted to be enhanced if many genes undergo a massive switch in gene expression near or at 60% RWC, which is borne out by the expression clustering data which shows nearly 70% of genes have a large shift (either upwards or downwards) in expression at this RWC (Clusters 5-9, 13, 15-16; Fig. 3.13).

This explanation is consistent with the results obtained but does have some weaknesses. Although a lot of the data matches a scenario in which the leaves measured at 60% RWC in this study were at a higher RWC than in the previous microarray study, there are still expression profiles that do fit this prediction (Fig. 3.11B). This could be due to general underlying differences between the datasets, or variables unrelated to desiccation. It is also difficult to compare RNA-Seq and microarray data, due to the differing methods of analysing expression changes and the expression scales. The two technologies also have different biases in terms of technical artefacts. For example, microarrays generally compare well with NGS count-based studies, especially for genes that show large, robust changes in expression. However, microarray studies show poorer correlation when transcripts are expressed at particularly low or high levels, due to masking by background noise or signal saturation at high concentrations, respectively, which could affect downstream comparisons (Richard et al., 2014). Because this present work and the Shen, 2014 study are the only two large-scale investigations into gene expression changes in *X. humilis* during desiccation, it is impossible at the current time to tell which measured RWC is closer to a putative "true" 60% RWC without external validation.

It is important to note here that this study only looked at direct changes in mRNA transcript abundance, which may not necessarily correlate with either resultant protein levels or protein activity (for example, activation by post-translational modifications). Additionally, resurrection plants often accumulate long-lived mRNA transcripts to be transcribed during rehydration, which can complicate inferences into gene function as not all up-regulated genes can be assumed to necessarily be translated during desiccation.

3.3.5.2 Multiple classes of LEA genes are upregulated during desiccation in X. humilis

A very large proportion of expression Clusters 3 and 4 (the genes that showed a biphasic expression pattern) consisted of LEA protein transcripts. Nearly 150 transcripts were annotated as LEA protein-coding genes, the majority of which (69%) were up-regulated during desiccation. As would be expected, LEA gene transcripts were some of the most abundantly expressed genes in response to desiccation in *X. humilis*. The LEA-3 *XhLEA3.22* transcript, with highest similarity to the *A. thaliana At1g52690/F6D8* gene, was by far the most highly expressed of all *Xerophyta* transcripts in the dry leaf, accounting for over 2% of all sequencing reads alone in the 5% RWC libraries. Assuming that the abundance of this transcript equates to protein levels, it could suggest that this LEA-3 protein may be particularly widespread within the cells of desiccated *X. humilis* tissue, or is an essential component of the cytoplasmic glassy state during desiccation. Many of the identified LEA genes were also expressed at comparatively high levels even in hydrated tissues (>300 FPKM). Thus, LEA proteins or the transcripts that encode them may be accumulated to some degree even in unstressed tissue.

The biphasic induction pattern observed for many LEA genes in this study has not been previously observed to our knowledge; a recently-published microarray study in *X. humilis* identified over 40 differentially expressed LEA transcripts, which were generally up-regulated in leaves at 60% and 40% RWC, as determined by that study (Shen, 2014). The observed biphasic expression pattern is seen across all identified LEAs from all families in this dataset, and is not due to the presence of LEA genes from specific groups or families. It is possible that this is an artefact of our experimental setup - for example, if some of the selected leaves were not truly at the calculated RWC; and may also be a consequence of the variability in the 80% RWC libraries, as seen in the PCA plots of both the Shen study and the data presented in this study (Fig. 3.10). This result could also represent an error in the assumptions of RNA-Seq about transcriptome uniformity across samples in this particular study. If substantially more transcripts are expressed in desiccating leaves at 60% compared to 80% and 40% RWC, the proportion of LEA transcripts will decrease irrespective of their absolute level, resulting in proportionally fewer reads during sequencing. However, this explanation is unlikely as the observed biphasic expression was found in only a minority of up-regulated genes. Alternatively, the expression profile may be the result of a biologically relevant physiological response. It has been well established that even desiccation-sensitive plant species induce LEA proteins in response to drought and other abiotic stresses, which would presumably need to occur prior to their cells reaching lethal water content levels (59-30% RWC; Höfler et al. 1941). Thus, an alternative hypothesis is that the 80% RWC LEA induction represents a stage of drought or stress responsiveness that is separate from the desiccation response seen at the later stages of desiccation in *X. humilis* (after 60% RWC), and may perhaps be analogous to the acute response seen in DS plant species during the first stages of drought. Additionally, in poikilochlorophyllous resurrection plants,

chloroplast and mitochondria can undergo structural changes during the initial stages of dehydration, particularly dismantlement of the inner membrane structures (thylakoids and cristae). In *Xerophyta scabrida*, the disappearance of internal chloroplast and mitochondrial membranes is visible between 100% and 60% RWC (Tuba, 1993a). In *X. humilis*, photosynthetic rate rapidly decreases below 80% RWC and has ceased entirely by 60% RWC (Beckett et al., 2012), consistent with chloroplast-xeroplast transformation also occurring during this stage in this species. LEA proteins may be important for membrane and macromolecule protection and stabilisation during this process, and could accumulate in these organelles during the early stages of dehydration. A similar process may occur in mitochondria, although it has yet to be investigated whether *X. humilis* mitochondrial cristae are lost during desiccation as is seen in *X. scabrida* and *X. villosa* (Tuba, 2008; Tuba et al., 1993). However, although specific LEA proteins have been found to be targeted to both chloroplasts and mitochondria in other plant species (Candat et al., 2014; Mueller et al., 2003), it is unlikely that the early, simultaneous induction of so many LEA genes in *X. humilis* relates solely to plastid or mitochondrial membrane protection. Instead, their increased expression may be as a general consequence of an ABA-induced abiotic stress response pathway.

By aligning the nearly 150 *X. humilis* LEA gene transcripts to the draft *X. humilis* genome, it was possible to determine that they were derived from 124 putative *X. humilis* "genes" – defined here as distinct, non-overlapping genomic regions – suggesting that at least some LEA gene families have been duplicated in *X. humilis* or *Xerophyta* when compared to other DS plant species (Fig. 3.26C). This analysis is not entirely confirmative, due to the fragmented nature of the current *X. humilis* draft genome assembly. Similar transcripts

derived from recently-duplicated genes, for example, may be collapsed into a single "gene" if one copy of the duplicated genes is missing from the genome assembly. In contrast, a partially-assembled LEA gene sequence split across multiple scaffolds may incorrectly result in the classification of diverse isoforms of that gene as coming from different genomic regions.

With the exception of the LEA-7 group, nearly all LEA families were expanded to some degree in *X. humilis* when compared to other plants. The functional significance of these gene expansions is not yet clear. LEA-3 is the largest LEA family in many DS plants, including *A. thaliana, O. sativa* and *V. vinifera*, and LEA-3 proteins display the largest variety of subcellular localisations (Candat et al., 2014; Hundertmark and Hincha, 2008). The LEA-3 gene family was noticeably expanded in *X. humilis*, and all LEA-3 transcripts were up-regulated in response to desiccation. A LEA-3 gene transcript, *XhLEA3.22*, was also the most highly-expressed of all transcripts in the dry leaf. Overall, this suggests that the LEA-3 protein family may be of general importance to the response to dehydration stress, and the *XhLEA3.22* protein may be highly abundant throughout the leaves of desiccated *X. humilis* plants. By far the most obviously expanded class of LEAs were those belonging to the LEA-8 family, which contained 42 members compared to the 3 found in *A. thaliana*. Interestingly, only eight LEA-8 gene transcripts were up-regulated in response to desiccation; 24 were not differentially expressed (NDE), and a further 16 were down-regulated. Both the constitutively expressed and down-regulated LEA-8 genes were present at a low level in hydrated tissue (generally well below 20 FPKM), which could suggest that the massive expansion of genes containing the LEA-8 motif is unrelated to VDT directly, or that the LEA-8 motif is not specific to genes that have been previously classified as LEAs. The LEA-8

family is atypical in that it contains the most hydrophobic of the LEA proteins (Battaglia et al., 2008). The three *A. thaliana* LEA-8 proteins have been found to be present in the cytosol and nucleus, and as a whole the class may be associated with both membrane protection (e.g. *LEA26* in *A. thaliana*) or enzyme protection and prevention of protein aggregation (e.g. *MtPM25* in *M. truncatula*; Boucher et al. 2010; Candat et al. 2014), thus LEA-8 proteins have been predicted to constitute both a protection and repair mechanism during desiccation in other plant species. The various NDE/down-regulated LEA-8 transcripts could perhaps be induced during rehydration to facilitate protein de-aggregation, similar to the proposed role of some rehydrins in lower plants (Oliver et al., 2005). Alternatively, they may function during specific stages of development or are only induced beyond low background levels in response to abiotic stresses that were not encountered during our experimental conditions, such as extreme temperatures. Also of interest is the identification of 21 LEA-6 (seed maturation protein; SMP) protein-coding transcripts in *X. humilis*, all but four of which are induced in leaves during desiccation. *A. thaliana* encodes six LEA-6 genes that are expressed specifically during seed development and not in response to stress in vegetative tissues (Hundertmark and Hincha, 2008; Illing et al., 2005). Five LEA-6 genes are also found in the genome of *O. sativa*, only a few of which appear to be weakly inducible in vegetative tissue (Wang et al., 2007). The expansion of the LEA-6 family and subsequent expression of these genes in the vegetative tissues of *X. humilis* may thus represent further evidence of the importance of seed-specific DT expression pathways during VDT.

Apart from the aforementioned NDE/down-regulated LEA-6 and LEA-8 genes, only four other LEA gene transcripts were NDE during desiccation (*XhLEA1.5*, *XhLEA2.14*, *XhLEA10.5* and *XhLEA10.6*), and a further three were down-regulated (*XhLEA2.15*, *XhLEA7.2* and

XhLEA10.7). Thus, with the exception of the LEA-8 family, nearly 95% of the identified LEA protein transcripts were up-regulated during vegetative desiccation in *X. humilis*. This is in contrast to *A. thaliana*, where the majority of LEA proteins are either expressed exclusively during vegetative stress or seed development (such as the seed-specific LEA-1 and LEA-6 families), with little overlap between the two types (Hundertmark and Hincha, 2008; Illing et al., 2005). This separation of LEAs also occurs during the reacquisition of DT in germinating *A. thaliana* seedlings, where only seed-specific LEAs are re-induced (Maia, 2014). Assuming that the large number of LEA transcripts identified in this study constitutes the majority of LEA genes within the *X. humilis* genome, this suggests that VDT in *X. humilis* may be associated with simultaneous induction of both the vegetative stress and seed development repertoire of LEA proteins, rather than the expression of only seed-specific LEAs - as would be expected if VDT is inherited from the window of seedling DT. However, this would need to be verified against a complete list of *X. humilis* LEA genes obtained from the genome, when a full draft becomes available.

3.3.5.3 Various seed specific genes are activated in desiccating leaves

As has been reported previously, desiccation of *X. humilis* leaves induced accumulation of transcripts generally associated with seed maturation-specific functions (Illing et al., 2005; Shen, 2014). These included seed-specific LEA families (e.g. LEA-1 and LEA-6), multiple transcripts associated with nutrient or energy storage in seeds, such as oleosins, caleosins and globulins, as well as seed-specific antioxidants, such as *PER1* (Fig. 3.28; Table 3.2). Orthologues of these genes are generally inactive in non-reproductive vegetative tissues in plants, and are regulated by ABA and the LAFL network during seed development (Haiyan Jia et al., 2013; Jia et al., 2014). They likely play a role in desiccated leaves that is similar to

the one they have in seeds, such as protein and membrane protection (LEAs), regulators of or direct sources of energy and amino acids during leaf rehydration (SSPs, oleosins/caleosins) and antioxidant activity (*PER1*). Their presence in the transcriptome of desiccating *X. humilis* is further confirmation that at least some aspects of vegetative DT are similar to the processes seen in seed maturation. Intriguingly, oleosin and caleosin genes showed an expression pattern similar to that of the LEA genes, and were up-regulated early during dehydration (80% RWC) and again during late desiccation (< 60% RWC). In contrast, globulin SSP genes were specifically induced only after 60% RWC. This very early expression phase is indicative that some seed-specific genes are activated rapidly during desiccation, and was also observed in *X. humilis* in a previous microarray study (Shen, 2014). It is unclear for what reason oil-body related genes are induced during early dehydration, although the similar expression profile to that of the LEAs suggests that they could be induced via the same abiotic stress pathway. Oleosin proteins have been found to be targeted to chloroplast plastoglobuli (plastid oil bodies) in seeds, tapetum cells and pollen (Parthibane et al., 2012; Wang et al., 1997). Additionally, plastoglobuli accumulate in chloroplasts during their conversion into xeroplasts in both *X. scabrida* and *X. humilis*, where they likely contain lipid storage materials necessary for reconstitution of thylakoids during rehydration, and additionally may play a role in photoprotection (Bréhélin et al., 2007; Solymosi et al., 2013; Tuba et al., 1993). Thus, oil-body related genes may play a role in the process of dismantling internal chloroplast membranes in *X. humilis*, as well as lipid storage related functions in the cell throughout desiccation. On the other hand, the globulin SSPs appear to be important only during late desiccation and likely correspond to the accumulation of storage proteins in the final stages of water loss.

Despite the induction of a number of well-characterised LAFL target genes, only transcripts with similarity to the seed master regulators *ABI3* (*XhABI3.1*, *XhABI3.2* and *XhABI3.3*) and *LEC1* (*XhLEC1.1*) were identified in the *X. humilis* transcriptome; none with significant similarity to *FUS3* or *LEC2* appeared to be expressed in hydrated or dehydrating tissues. Of these identified putative LAFL gene transcripts, only two *X. humilis ABI3* orthologue splice variants (*XhABI3.1* and *XhABI3.2*) were slightly but significantly DE during desiccation. This refutes our working hypothesis that VDT in resurrection plants involves re-activation of the central gene networks that control seed development. *ABI3* is vital for desiccation response during maturation drying in angiosperms and VDT in some lower plants (Santos-Mendoza et al., 2008; Yotsui et al., 2013). The very low induction of the two DE *ABI3* transcripts in *X. humilis* (barely 2-fold increase, maximum FPKM < 5) makes it unlikely that *ABI3* is acting as a master regulator of VDT; additionally, maximal *XhABI3* expression occurs during late desiccation, whereas many desiccation-related genes, such as LEAs, are strongly induced as early as 80% RWC. Furthermore, all three *X. humilis ABI3* transcripts encoded proteins that lacked the B3 domain, which binds to the seed-specific RY motif. The *ABI3* gene is alternatively spliced in other plant species, and has been investigated in some detail in developing pea embryos (Gagete et al., 2009). Although none of the *PsABI3* isoforms lacked the B3 domain, the authors identified an *A. thaliana* EST (AU238055) that would code for a truncated ABI3 protein that lacked the B3 domain due to a stop codon within a retained intron. The fact that all three identified *X. humilis ABI3* transcripts encoded proteins that lacked this domain suggests that this result is not an assembly artefact, and may represent an *ABI3* variant that is expressed in the vegetative tissues of *X. humilis*. Genomic scaffolds to which these sequences mapped in the current draft *X. humilis* genome likewise did not contain sequence for a B3 domain (Stephen Schlebusch, personal correspondence). Due to

the fragmented nature of the current genome assembly, it is possible that a corresponding B3 domain may be found on a genomic contig that was not completely assembled, but no evidence for its existence has been found in either the transcriptomic or genomic assemblies so far. Because the B3 domain mediates the interaction of ABI3 with the seed-specific RY motif, it appears unlikely that the function of the proteins produced by these transcripts involves gene activation via this CRE. In *A. thaliana*, ABI3 interacts with bZIP transcription factors, such as *ABI5*, to mediate expression of some embryogenesis genes, including LEAs and SSPs, via the ABRE (Lara et al., 2003; Suzuki and Mccarty, 2008). This interaction is independent of the B3 domain, and does not require ABI3 to be anchored to an RY element (Sakata et al., 2010); thus, it is tempting to hypothesise that the role of *ABI3* during late desiccation in *X. humilis* could involve cooperative interaction with other transcription factors, including ABA-responsive TFs of the bZIP family.

3.3.5.4 ABA metabolism and an ABA-signalling pathway is activated by desiccation

ABA signalling is known to play an important role in both drought and desiccation tolerance in plants. In the *X. humilis* transcriptome, there is evidence for the up-regulation of many ABA biosynthesis genes as well as activation of ABA-signalling pathway genes during desiccation.

X. humilis orthologues of the ABA biosynthesis genes *ABA4* and *NCED* were up-regulated throughout desiccation, most prominently at 80% and 40% RWC (Fig. 3.30D). Two *X. humilis* genes with similarity to *AtAAO3* (*XhAAO1* and *XhAA02*) were also up-regulated specifically during late desiccation. However, it is unclear based on a phylogenetic analysis whether these genes might perform a function equivalent to *AtAAO3* (oxidation of abscisic aldehyde

to abscisic acid), or are perhaps orthologous to one of the other closely-related *A. thaliana AAO* genes, which are not believed to be involved in ABA metabolism (Fig. 3.30B; Seo et al. 2004). Interestingly, the predicted *X. humilis* orthologues of *ABA1* and *ABA2* were not up-regulated during desiccation and were both down-regulated after 60% RWC in *X. humilis* leaves (Fig. 3.30D). *ABA1* expression does not appear to be consistently responsive to stress in leaves across plant species. *ABA1* transcripts are not responsive to drought in tobacco and tomato leaves - although they are differentially expressed in roots - but are differentially expressed in *A. thaliana* leaves in response to drought, salt and PEG (Audran et al., 1998; Thompson et al., 2000; Xiong et al., 2002). *ABA2* transcript levels, on the other hand, are not differentially expressed in response to stress in *A. thaliana* (Cheng et al., 2002; González-Guzmán et al., 2002). Both *ABA1* and *ABA2* were expressed at a very high level during early dehydration in *X. humilis* compared to other genes involved in ABA metabolism (60-90 FPKM) before they were down-regulated, and so it is possible that the baseline levels of their corresponding proteins are sufficient to prevent limitation of ABA production during stress despite their lack of up-regulation, as is presumably the case in tomato and tobacco (Xiong and Zhu, 2003). Furthermore, in *A. thaliana*, the *NCED* family of proteins are believed to be the rate-limiting enzymes in ABA production (Iuchi et al., 2001), so the up-regulation of three putative *NCED* genes suggests that ABA biosynthesis is likely increased during desiccation. Conversely, the majority of genes predicted to be involved in ABA catabolism by similarity to *A. thaliana CYP707A* ABA hydroxylases showed a very slight increase in expression at 80% RWC, but were down-regulated thereafter. However, three *XhCYP707A* genes were up-regulated during late desiccation (40% and 5% RWC; Fig. 3.30D), and as such could play a role in attenuating ABA biosynthesis at low water contents or upon rehydration. In general, these results are consistent with increased ABA levels during

desiccation through both an increase in biosynthetic gene expression levels (*ABA4*, *NCED* and *AAO* genes) and downregulation of the majority of predicted *XhCYP707A* ABA catabolism genes. Additionally, ABA levels (as predicted by ABA biosynthesis gene transcript abundance) appear to peak during both early dehydration (80% RWC) and late desiccation, a pattern similar to the expression profile of most LEAs (Fig. 3.26) and seed genes (Fig. 3.28). This profile is also comparable to the measured changes in ABA concentration in desiccating *X. humilis* leaves in a previous study, where ABA levels were maximal at ≈70% RWC, with a secondary, smaller peak in concentration at ≈50% RWC (Beckett, 2011). Measured ABA levels were much reduced after 40% RWC, which may be due to the action of ABA catabolysis enzymes, such as those encoded by the *XhCYP707A4*, *XhCYO707A10* and *XhCYP707A12* genes identified in this study.

Concomitant with a predicted increase in ABA hormone concentrations, a large number of genes involved in ABA signalling were also up-regulated in desiccating *X. humilis* leaves at various stages. These included PYL/PYR-family ABA receptors, SnRK2 and PP2C protein kinases/phosphatases, and ABRE-binding TF factors (ABFs). ABFs are a related family of group A bZIP transcription factors that act as important regulators of the ABA response in plants downstream of ABA signalling (Fujita et al., 2013). In *A. thaliana*, different ABF proteins are important for seed development (e.g. *AtABI5*, *AtEEL*) and vegetative stress response (e.g. *ABF1-4*; Fujita et al. 2013; Yoshida et al. 2015). *X. humilis* and *X. viscosa* seedlings display VDT from the earliest stages of germination (see Chapter 2), the same developmental stages at which *AtABI5* regulates seedling growth arrest and DT re-acquisition in *A. thaliana* (Lopez-Molina et al., 2002; Maia, 2014). Based on our earlier hypothesis that VDT in *Xerophyta* involves de-repression of the central seed maturation

regulatory network, an *ABI5* orthologue seemed an ideal candidate as a seed TF that could be an important regulator of the VDT response. Four genes that encoded proteins with similarity to ABFs were identified in the *X. humilis* transcriptome, three of which were up-regulated during desiccation. Surprisingly, phylogenetic analysis revealed that all three up-regulated ABF gene protein sequences were most similar to *AtABF1-4*, the "master regulators" of vegetative stress response in *A. thaliana*, whereas *XhABF4*, the only predicted ABF to belong to the seed development clade, was expressed at a very low level and down-regulated during desiccation. This suggests that components of the vegetative stress response, rather than seed maturation factors, may be more important in regulating VDT in *X. humilis*. Of particular interest is the gene *XhABF1*, which was very strongly induced during both the early (10-fold at 80% RWC) and late (over 20-fold at 40/5% RWC) desiccation response, and has an expression pattern that matches that of many desiccation-induced genes, including most LEAs (Fig. 3.26E) and seed nutrient storage genes (Fig. 3.28). Further consistent with a role for ABF proteins during VDT in *X. humilis*, several *ABI FIVE-binding PROTEIN* (*AFP*) genes were also significantly DE in the dataset. AFP proteins are induced by ABA and stress, and are predicted to fine-tune ABF function by acting as negative regulators that target them for ubiquitination and degradation (Garcia et al., 2008).

The ABA signal in plants is mediated by cytosolic ABA receptors of the PYL/PYR/RCAR family, which interact with PP2CA protein phosphatases in response to ABA and ultimately lead to ABF activation by SnRK2 protein kinases. Multiple PYL-family ABA receptor genes were up-regulated during the late stages of desiccation (40% and 5% RWC). The tissue or temporal specificity of these genes in other species, if any, has not been thoroughly investigated; however, PYL mutants display both seed and vegetative mutant phenotypes,

suggesting they are important for ABA signalling throughout the plant life cycle (Gonzalez-Guzman et al., 2012). PYL genes also show patterns of transcriptional regulation in response to abiotic stress across plant species. For example, all *SlPYL* genes are down-regulated by dehydration in tomato leaves, with the exception of *SlPYL1* (up-regulated) and *SlPYL8* (not differentially regulated) (Sun et al., 2011). Similarly, in sorghum, all *SbPYL* genes were down-regulated in leaves in response to dehydration except for *SbPYL1* and *SbPYL7* (Dalal and Inupakutika, 2014). In maize *ZmPYL4*, *ZmPYL7* and *ZmPYL8* were significantly up-regulated in response to dehydration stress in leaves and the remaining *ZmPYL* genes were downregulated (Fan et al., 2016). An investigation into the role of PYL receptors in ABA-sensitivity during stress-induced growth arrest in *A. thaliana* seedlings identified three PYL genes that were either up-regulated (*AtPYL7* and *AtPYL9*) or down-regulated (*AtPYL5*) in response to growth arrest by PEG treatment (Maia, 2014). In contrast, we see the majority of *X. humilis* PYL genes are up-regulated during vegetative desiccation (as much as 200-fold, in the case of *XhPYL11*), although a handful of PYL genes were also down-regulated, suggesting a possible non-redundant role in ABA signal transduction amongst different *X. humilis* PYLs (Fig. 3.32A). For the most part, up-regulated PYL genes were highly induced during late rather than early desiccation (40% and 5%). Thus, *X. humilis* may possibly become more sensitive to ABA as desiccation progresses, despite the apparently low levels of ABA in leaves at this stage (Beckett, 2011). The up-regulation of more PYL gene transcripts during VDT than during dehydration in DS plants may be due to underlying regulatory differences between these two stresses, or is a response associated with *X. humilis* or *Xerophyta* in particular that may be unrelated to VDT.

Group-A PP2C proteins are induced by ABA, but negatively regulate ABA-signalling in higher plants. They inhibit SnRK2 protein kinases in the absence of ABA, but are themselves inhibited by PYL proteins in the presence of ABA. The *A. thaliana* genome contains nine PP2CA genes, which often show a separation in tissue and developmental stage specificity (Bhaskara et al., 2012). In contrast, nearly all predicted *X. humilis* group-A PP2C protein genes (based on similarity to *A. thaliana* and *O. sativa* PP2CA protein sequences) were up-regulated during the late stages of desiccation (Fig. 3.32B; Fig. S3.7). Many PP2CA genes were up-regulated during both early and late desiccation, although transcript abundance was generally higher at 40% RWC than at other time points. In *X. humilis* PP2CAs likely serve to attenuate the ABA response throughout desiccation, and have been proposed to act as an "off switch" for VDT in the case that water stress is removed prior to commitment to complete desiccation.

Lastly, multiple genes predicted to encode SnRK2-family protein kinases were also identified in the *X. humilis* desiccation transcriptome. In higher plants, these proteins are classified into three groups: Group 1, those that are not activated by ABA; Group 2, those that are weakly activated by ABA; and Group 3, those that are highly activated by ABA (Kobayashi et al., 2005). A phylogenetic analysis of predicted *X. humilis* SnRK2 protein sequences could confidently assign genes to all three of these groups, with the exception of *XhSRK2.3*, *XhSRK2.9* and *XhSRK2.10* (Fig. S3.8). Interestingly, all three putative *X. humilis* group 3 SnRK2 genes (*XhSRK2.4*, *XhSRK2.5* and *XhSRK2.6*) were down-regulated in desiccating tissue, whereas those from group 2 (*XhSRK2.1*, *XhSRK2.2* and *XhSRK2.11*) were up-regulated after 60% RWC (Fig. 3.32C). There is precedence for SnRK2 gene repression in response to ABA, as two group 3 SnRK2 genes in rice, *SAPK9* and *SAPK10*, are down-

regulated specifically in the roots by ABA treatment (Kobayashi et al., 2004b). Additionally, both group 2 and 3 SnRK2 kinases have been implicated in abiotic stress response, including drought and dehydration, and can activate expression of downstream stress response genes (Fujita et al., 2013; Kulik et al., 2011; Mizoguchi et al., 2010). Proteins from both groups have also been shown to both induce expression of and directly activate ABF proteins to regulate ABA-mediated gene expression (Mizoguchi et al., 2010). Thus, our results could suggest that group 2 SnRK2-mediated signalling is important for the *X. humilis* desiccation response, as opposed to that of the strongly-ABA-activated group 3 SnRK2s. Group 1 SnRK2 proteins are not activated by ABA, but are nonetheless believed to also be involved in abiotic stress response in many cases (Kulik et al., 2011). All four identified *X. humilis* group 1 SnRK2 genes (*XhSRK2.7*, *XhSRK2.8*, *XhSRK2.12* and *XhSRK2.13*) were induced by desiccation, further implicating the ABA-independent stress-response signalling pathways in VDT response.

3.3.5.5 Dehydration in X. humilis is associated with successive waves of transcription factor activity

Well over a thousand TF-coding genes were differentially expressed during desiccation in *X. humilis*, of which approximately 1000 could be grouped into known plant TF families based on the presence of predicted protein domains or sequence homology to known TFs. Unsurprisingly, the vast transcriptional changes required to halt metabolism and enter a quiescent, desiccated state involved the up- or down-regulation of TFs from nearly all TF families. Grouping TF transcripts by shared expression profiles revealed that TF genes were often induced in successive waves, suggesting consecutive regulation of multiple gene sets as dehydration progresses. For example, Clusters U1, U2, U3 and U4 represented TFs

upregulated predominantly at 80%, 60%, 40% and 5% RWC respectively, whereas Clusters D1 and D2 represented TFs down-regulated immediately upon desiccation and predominantly after 60% RWC, respectively. Although they shared a number of GO terms, the enriched GO terms and over-represented TF families often differed across expression clusters, consistent with distinct roles for each set of expressed TFs at each RWC.

All differentially expressed TFs contain a shared set of regulated biological processes

Genes from all six expression clusters shared a large number of enriched GO terms, suggesting some commonalities between the expressed TFs. Unsurprisingly - considering the sample set was composed entirely of TF-coding sequences - this included multiple terms related to DNA-templated transcription. The GO term "response to water deprivation" was found in all clusters, as would perhaps be expected considering water deprivation was the common variable across all samples in the dataset. Interestingly, the GO term "response to cold" was also shared by all TF clusters compared to the transcriptome as a whole, suggesting a degree of overlap between the desiccation response and genes involved in response to cold stress. Many TFs involved in individual stress responses display cross-talk in other abiotic stresses (Nakashima et al., 2014); over-expression of the cold-induced *DREB1* genes in *A. thaliana*, for example, can improve drought tolerance (Liu, 1998). Similarly, there are multiple lines of evidence for cross-talk between transcriptional networks that regulate abiotic and biotic stress responses (Kissoudis et al., 2014; Walley et al., 2007). The defence response GO terms "response to bacteria", "defense response to fungus" and "respiratory burst involved in defence response" were also found in all six expression clusters, as were the terms "systemic acquired resistance, salicylic acid mediated signalling pathway" and "induced systemic resistance, jasmonic acid mediated signalling pathway" – suggesting

possible overlap between TFs differentially expressed during VDT and the biotic stress response. However, there was some evidence that the wild *X. humilis* plants used in this study were host to a fungus/fungi, based on the consistent levels of fungal sequence contamination, so it cannot be ruled out that some of the genes induced during desiccation were involved in preventing opportunistic pathogenic infection by fungi or bacteria in response to stress. Even the presence of benign endosymbiotic fungi could still trigger a transcriptomic response, for example, through response to fungal PAMPs (pathogen associated molecular patterns) such as chitin, which are not necessarily associated with pathogenesis. Additionally, resurrection plants accumulate secondary metabolites to protect against biotic factors like herbivory while in a desiccated state (Gechev et al., 2014), and so may also express genes that produce antibiotic agents that prevent infection or colonisation by microorganisms while desiccated.

Endogenous signalling, specifically by cytokinins, auxin, ABA, ethylene and GA were also found to be enriched in all six TF clusters, likely due to a coordinated response by TFs from multiple signalling pathways throughout the desiccation process. Finally, various GO terms related to growth and development were enriched across all clusters, particularly of seeds ("embryonic development ending in seed dormancy", "seed dormancy", "seed germination", "negative regulation of seed germination") and flowers ("regulation of flower development", "gynoecium development"), but also noticeably "leaf senescence". Prevention of senescence and programmed cell death is vital for successfully completing the desiccation process in resurrection plants, and is an aspect of VDT that has only recently begun to be investigated (Griffiths et al., 2014). Due to the preponderance of seed-related GO terms enriched in the TFs identified in this dataset, I attempted to analyse the putative

function of the predicted proteins of similar genes in other plant species. Although many were found to have functions during seed development or germination, the overwhelming majority were not specifically associated with seed development processes (Table 3.7); rather, they are involved in morphological or general stress responses that would be common in embryonic and vegetative tissues, suggesting that VDT is not associated with the specific up-regulation of seed-specific TFs.

TFs expressed during early desiccation (80% RWC) may regulate the initial response to general stress

Clusters U1 and D1 contained TF genes that had peak expression or were down-regulated at 80% RWC, respectively. Both contained a comparatively small number of TFs and enriched GO terms compared to other clusters. A small number of stress response GO terms were unique to these clusters (response to hypoxia, freezing, viruses; response to ozone, oxidative stress), as well terms related to growth and response to light (photoperiodism, blue and red/far-red light signalling), but there was no overarching collection of GO terms that gave insight into the regulatory mechanisms that specifically occurred during this stage of desiccation stress response.

It is possible that inhibition of growth plays a role during the early response to dehydration in *X. humilis*. Growth arrest during the acute response phase of drought and during germination is generally mediated by DELLA proteins, which inhibit GA signalling. A gene encoding a protein with similarity to the DELLA protein *DWARF8*, the only predicted DELLA that was differentially expressed in *X. humilis*, was found in cluster U1 and may play a role in an equivalent acute response in *Xerophyta*. Analysis of the few genes with

similarity to proteins with known functions in these clusters also revealed that a large number of *RADIALIS, RAD-like* and *TCP* proteins were down-regulated, gene families that are often involved in leaf patterning and development (Corley et al., 2005). Amongst proteins similar to up-regulated genes, *HAT5* is a homeodomain TF that may be involved in leaf growth and senescence (Aoyama et al., 1995). Over-expression of the DREB-like ethylene-responsive gene *TINY*, similar to a gene up-regulated in cluster U1, also results in reduced cell proliferation and growth (Wilson et al., 1996). Thus, this stage of dehydration may include a period of general stress response and growth arrest similar to the acute stress response seen in DS plants during drought.

TFs expressed at 60% RWC are associated with large changes in gene regulation and chromatin modification

Clusters U2 and D2 corresponded to TFs that were up-regulated at 60% RWC and then down-regulated, or prominently down-regulated immediately after 60% RWC. Both clusters contained a large number of genes, and a diverse range of enriched GO terms.

By far the most noticeable group of unique GO terms in cluster U2 were those related to the regulation of gene expression, both epigenetic and post-transcriptionally. This suggests that largescale changes in regulation of gene expression occur at 60% RWC, a similar result to that seen in cluster 6 of all DE genes (Fig. 3.17). This is likely associated with the induction of many of the genes necessary for VDT, and the silencing of unnecessary developmental programmes during desiccation; for example, multiple *CONSTANTS* (*CO*) and *CO-like* TFs, which control many of the aspects of reproduction and flowering in *A. thaliana*, were down-regulated at this stage of desiccation (Fig. 3.38B; Putterill et al. 1995). The prevalence of GO

terms associated with RNA-mediated epigenetic gene regulation is also of particular interest. *CDT-1* and *CDT-2*, two non-coding RNAs identified in the resurrection plant *C. plantagineum*, direct the synthesis of a 21 bp siRNA that can independently induce VDT in this species without the application of exogenous ABA (Hilbricht et al., 2008). The data observed here could indicate that siRNA-mediated gene silencing plays a similarly important function during desiccation in *X. humilis*. Cluster D2 was enriched for a diverse range of GO terms associated with multiple developmental and metabolic processes, perhaps due to the predicted epigenetic silencing of multiple unrelated gene programmes at this stage.

Additionally, cluster U2 had multiple GO terms related to biosynthesis of xylan/glucuronoxylan and triglycerides, likely related to the numerous cell wall and membrane changes that occur during desiccation in resurrection plants, such as cell wall thickening, cell wall folding and vesicle fragmentation (Griffiths et al., 2014; Hoekstra et al., 2001; Toldi et al., 2009). These processes may be already occurring at this state of dehydration, or the pathways that regulate these mechanisms may be activated at this stage.

Multiple representative genes of stress-response TF families are induced in late desiccation (40% and 5% RWC)

The last two stages of desiccation appeared to be largely involved in cellular response to stress. The GO terms "cellular response to heat" and "heat acclimation" were unique to cluster U3, and "response to high light intensity", "cellular response to freezing", "cold acclimation", "photosynthetic acclimation", "photoprotection", "response to UV-B", "response to hydrogen peroxide" and "cellular response to hypoxia" were shared

specifically between cluster U3 and U4 (Table 3.5 & 3.6). Both clusters also shared expression of multiple genes from the same abiotic stress-related TF families.

HSF-family transcription factors were found in both late-desiccation specific clusters. Cluster U3 contained the highest proportion of HSFs, the vast majority of which were from the HSFC group. In contrast, cluster U4 predominantly contained HSFs from group A. Only one HSFB was found in each cluster (Fig. 3.36). The functional characterisation of the HSF family in monocots is poor, but the expression of a different set of HSF genes at 40% and 5% RWC suggests that the HSFs in each group are functionally non-redundant, or HSFs from group C (present in U3) act upstream of group A HSFs (found in U4). HSF master regulators have been identified in *A. thaliana* (the HSFA1 family) and tomato (the *HSFA1a* gene), and there is evidence that the *HSFC2b* gene in barley regulates a number of downstream HSFs during seed development and heat stress (Reddy et al., 2014; Scharf et al., 2012). Nearly all predicted *X. humilis* HSFC proteins were most closely grouped with the *HSFC2* clade, so it is possible that these genes also play a central regulatory function in *Xerophyta.* On the other hand, it has previously been reported that the HSF and DREB2 transcriptional cascades may converge during stress response, as the *AtHSFA3* gene in *A. thaliana* is activated by both *AtDREB2a* and *AtDRBE2c* in response to drought (H. Chen et al., 2010; Sakuma et al., 2006a; Schramm et al., 2008). A variety of predicted *DREB2* genes were found that were up-regulated during desiccation, predominantly below 40% RWC (Fig. 3.37E). The *DREB* protein family are important transcription factors involved in gene activation during abiotic stress, in a mostly ABA-independent manner (Lata and Prasad, 2011). The *DREB2* subgroup, to which all up-regulated *X. humilis* DREB genes in these clusters belong, are specifically induced by heat, drought and hypersalinity; they are rapidly induced by stress, and activate

the expression of numerous downstream genes, including some LEAs (Sakuma et al., 2006b). Interestingly, *AtDREB2a* gene expression is also positively regulated by *AtABF4* in *A. thaliana*, due to the presence of an ABRE within its promoter (Kim et al., 2011); *XhABF1*, a related *Xerophyta* ABF protein identified in this study, is strongly up-regulated during these same stages of desiccation, and so may be involved in *DREB2* gene activation.

As well as HSF and DREB proteins, a large number of genes with similarity to the Zn-C2H2 *ZAT/AZF* family of TFs were spread across both clusters U3 and U4 (Fig. 3.37). The *ZAT* and *AZF* genes encode a family of related transcriptional repressors that, in *A. thaliana*, are induced by ABA and involved in repression of ABA-inhibitory and auxin-related factors, as well as the induction of growth arrest during drought and other abiotic stress (Drechsel et al., 2010; Miller et al., 2008; Sakamoto, 2004). *AtZAT11* and *AtZAT12* play a central role in the cellular response to ROS and light-inducible stresses (Davletova, 2005; X.-M. Liu et al., 2014; Qureshi et al., 2013). *AtZAT10* has also been implicated in the maintenance of seed quiescence: mRNA transcript abundance for this gene peaks at the end of maturation, and inhibition of *AtZAT10* in response to imbibition is believed to be important for the initiation of germination (Galland et al., 2014). Similarly, *AtAZF2* has been shown to be important for ABA-signalling in the seed and the onset of germination (Drechsel et al., 2010). The large number of distinct *X. humilis ZAT*-like transcripts, their desiccation-specific expression pattern and large fold-change induction (over 100-fold in many cases) suggests that these genes maybe play an important role during late VDT response in *X. humilis*, although their precise function is unclear.

Lastly, multiple *SCARECROW-LIKE* (SCL) gene transcripts were expressed during late desiccation (Fig. 3.37D&E). SCL proteins are a class of GA-responsive TFs that are expressed

in seeds and adult plants, and make up one branch of the GRAS family of plant transcription factors. Many have been shown to be involved in response to biotic or abiotic stresses, although the function of most of them is still unclear. As positive regulators of GA signalling, *SCL* proteins often oppose ABA-induced stress responses. For example, *AtSCL3* is a positive regulator of GA signalling that is repressed by GA, but acts as a mutual antagonist of ABA-induced DELLA proteins through competitive interaction at the same binding site (Yoshida and Ueguchi-Tanaka, 2014; Z.-L. Zhang et al., 2011). *AtSCL3* reaches peak mRNA abundance levels in the dry seed, but expression is reduced in response to either exogenous ABA or after imbibition as GA levels increase (Feurtado et al., 2011). The related protein *AtSCL6* has been found to mediate GA-DELLA interaction during leaf growth and chloroplast biosynthesis in light (Ma et al., 2014). In *X. humilis* these genes may play an important role in fine-tuning the ABA-GA signalling networks during late desiccation and/or rehydration, similar to their proposed roles during drought stress and germination.

3.3.5.6 VDT may be associated with large-scale chromatin remodelling

The activation of the VDT response in *X. humilis* requires a massive shift in gene expression, as evidenced by the large number of DE genes and TF families present during the desiccation process. Epigenetic gene regulation is a commonly used mechanism to control the activation and repression of multiple gene networks during developmental processes, and the transition to a desiccated state in *X. humilis* is likely no different.

Genes associated with chromatin rearrangement and epigenetic changes during VDT were centred on 60% RWC in our data. Numerous genes and TFs related to chromatin remodelling or epigenetic regulation were specifically induced at 60% RWC (DE clusters 5, 6

& 9, TF cluster U2), or specifically silenced after 60% RWC (DE clusters 13, 15 & 16, TF cluster D2). This suggests that chromatin changes occurring at and around 60% RWC may particularly important for VDT, a time-point which also coincides with the induction of expression of the majority of stress response genes, based on the widespread enrichment of stress-related GO terms in genes up-regulated between 60% and 5% RWC (DE clusters 5-9; Fig. 3.13 & Fig. 3.16-3.19). However, not all of the chromatin remodelling processes observed during VDT are necessarily related directly with specific gene regulation. Chromatin compaction occurs in the nuclei of maturing seeds of *A. thaliana*, and in the nuclei of desiccating leaves of the resurrection plant *C. plantagineum* (van Zanten et al., 2012), so it is likely that a similar process occurs in *X. humilis*.

Histones

I identified over 40 histone-coding genes that were DE during a cycle of dehydration, from each of the histone families: H1 (10 genes), H2A (8), H2B (3), H3 (3) and H4 (10). For comparison, *A. thaliana* contains 15 H3 gene variants (Okada et al., 2005), 3 H1 variants (Wierzbicki, 2005), 7 H4 variants (Piontkivska et al., 2002) and, based on proteins curated by Swissprot, 11 H2b and 13 H2a variants. The vast majority of histone protein genes showed a very similar expression profile: varying degrees of induction at 60% RWC, and a massive decrease in transcript abundance thereafter. This pattern of expression likely correlates with a major chromatin rearrangement process occurring at 60% RWC, followed by a substantial decrease or cessation of DNA replication and nucleosome assembly/reassembly at lower water contents, possibly to avoid an increased risk of DNA damage (Dekkers et al., 2015). A drought-inducible H1 variant, *XvDIH1v*, that was isolated from the related resurrection

plant *X. viscosa* was reported as exhibiting a similar expression profile (Holiday, 2007). Our data suggests that this could be a common response in the majority of histone protein genes.

Of all the histones, only a family of H1 variants have been directly associated with water stress in higher plants, such *H1.3* in *A. thaliana* and *H1-S* in tomato, and which likely function by regulating gene expression rather than overall chromatin structure (Scippa et al., 2004, 2000; Zhu et al., 2012). I identified only one H1 variant, *XhH1.9*, which was strongly induced during desiccation in *X. humilis* (Fig. 3.39A). This gene was found to be most similar to the *H1.3* histone variant in *A. thaliana* based on protein sequence alignment, suggesting that drought-inducible linker histones may accumulate in chromatin during desiccation in *X. humilis* (Fig. 3.39F). There is currently no evidence that drought-inducible H1 histones are important components of chromatin within maturing or dry seeds, although the level of canonical histone *H1.1* is elevated in dry maize seed nuclei compared to the level normally seen in plant tissues (Bewley and Black, 1994; Groot et al., 2007). Thus, up-regulation of this histone variant during desiccation in *X. humilis* may be specific to vegetative tissues.

Only a handful of other histone gene transcripts were up-regulated during late desiccation. These included two H2A genes (*XhH2A.1* and *XhH2A.2*), an H3 gene (*XhH3.3*) and an H4 gene (*XhH4.1*). As even less is known about the role of non-H1 histone variants during drought stress or seed development (Bewley and Black, 1994), it is difficult to ascertain the specific function of these histone variants during vegetative desiccation. They may facilitate DNA repair processes, result in altered epigenetic regulation of developmental or stress response genes, or be important for heterochromatin formation, chromatin compaction and chromatin protection while the cell is in a desiccated state.

Histone modifying enzymes

I focused my analysis of epigenetic gene regulation on the expression of histone modifying enzymes, specifically those associated with histone acetylation and methylation. Consistent with a general decrease in chromatin remodelling or modification during the later stages of desiccation, the majority of histone modifying enzymes – irrespective of target histone or protein residue – were down-regulated after 60% RWC (Fig. 3.40 & 3.41). The global down-regulation of chromatin modifying enzymes during late desiccation may be because a more limited range of genes are active during this time, requiring a less diverse range of histone epigenetic modifications to regulate expression. It is also likely linked to a general reduction in cell division and/or DNA replication as vegetative tissues enter a quiescent state, as it matches the down-regulation expression profile of most histone protein genes. Newly-synthesised histones undergo a number of post-translational modifications during translation and prior to import into the nucleus, thus a global decrease in histone synthesis – as suggested by the decrease in abundance of histone-coding transcripts – could also result in a general reduction in the level of required histone modification complexes (Keck and Pemberton, 2012; Rivera et al., 2015). However, a small number of *X. humilis* histone-modifying enzymes of varying specificities did show up-regulation during the desiccation process, and thus may be interesting candidates as mediators of the epigenetic regulation of VDT.

Histone acetylation may have diverse functions during VDT

HAT proteins are generally associated with gene activation, and have been implicated in the response to abiotic stress in a number of species (Fang et al., 2014; Kim et al., 2015; Papaefthimiou et al., 2010). In *X. humilis*, only MYST (*XhHAT4*) and CPB (*XhHAT5, XhHAT6*

and *XhHAT7*) family HATs were up-regulated in response to desiccation (Fig. 3.41A). Genes from these families are up-regulated by drought stress and ABA in other species, such as *OsHAC703* and *OsHAM701* in rice and *HvMYST* in barley (Fig. 3.41B). Interestingly, multiple predicted members of the GNAT family of HATs were down-regulated during desiccation stress, although members of this family have been shown to be responsive to drought or ABA in other species (*OsHAG703*, *HvELP3* and *HvGCN5*). During stress, HAT proteins are predicted to lead to induction of stress response genes by acetylation of histones in the promoters of those genes (Kim et al., 2015). However, there is currently little information regarding direct targets of HAT proteins that are involved in stress response. In maize, the up-regulation of the *ZmEXPB2* and *ZmXET1* genes, which are involved in cell wall maintenance in response to high salinity, is associated with increased H3K9 acetylation of their promoters - likely mediated by two HAT genes, *ZmHATB* and *ZmGCN5*, which show a concomitant increase in transcript abundance (Li et al., 2014). It is probable that many of the genes identified as up-regulated during desiccation in *X. humilis* are regulated epigenetically, and may contain similarly acetylated promoter histones. Only three of the seven identified HAT-encoding genes were downregulated during desiccation in *X. humilis*: *XhHAT1*, which was highly similar to *AtHAG1*/*HvGCN5*/*OsHAG702*, as well as *XhHAT2* and *XhHAT3*, which were both most similar to *AtMCC1* (Fig. 3.41B). In *A. thaliana*, *MCC1* has been shown to be involved in chromosome recombination and segregation during meiosis by histone hyperacetylation (Perrella et al., 2010). If *XhHAT2* and *XhHAT3* perform a similar function in *Xerophyta*, their observed reduction in transcript abundance likely correlates with the cessation of meiotic processes in desiccating tissues. In comparison, *AtHAG1* has been associated with multiple processes related to development and response to external stimuli, and is likely involved in further as yet unidentified systems (Servet et al., 2010).

Histone acetylation is not only important for direct activation of gene expression. Acetylation of histones, particularly at H4K16, prevents the spread of heterochromatin formation and thus could indirectly control gene expression by preventing heterochromatin-mediated gene silencing (Shahbazian and Grunstein, 2007; Soppe et al., 2002). Increased acetylation of histones is also observed in plants irradiated with UV-B (Casati et al., 2008). Knock-out mutants of the *A. thaliana* HAT gene *HAM1* and its functionally redundant homolog, *HAM2*, exhibit increased levels of DNA damage after UV treatment (Campi et al., 2012). As one of the up-regulated *X. humilis* HAT genes, *XhHAT4*, has high similarity to both *AtHAM1* and *AtHAM2*, this could suggest that it may play a role in histone acetylation-mediated DNA repair mechanisms during desiccation and/or rehydration.

As opposed to HATs, HDACs are associated with gene repression because they remove the active histone acetylation epigenetic mark. The majority of identified HDAC-coding genes were down-regulated in desiccating tissue below 60% RWC, a trend that is probably related to the global repression of histones and other classes of histone-modifying enzyme genes observed in the dataset. Only two *X. humilis* HDAC genes, *XhHDA3* and *XhHDA4*, show increased expression during desiccation. *XhHDA3* is likely a RPD3/HDA1 class I HDAC, whereas *XhHDA4* is predicted to belong to the class II RPD3/HDA1 HDACs. Genes from both of these families as well as a large number of predicted HD2 HDACs were also down-regulated in desiccating tissues. However, no genes with similarity to the third class of RPD3/HDA1 HDACs (*AtHDA2*) were identified (Pandey, 2002).

XhHDA3 was induced at approximately 60% RWC, and was most similar to the *A. thaliana* proteins *AtHDA19*, *AtHDA6* and *AtHDA7* based on full-length protein sequence alignment. Of these three, *AtHDA19* and *AtHDA6* have been the most thoroughly researched, and have

been found to be involved in numerous developmental and physiological responses (X. Liu et al., 2014). *AtHDA7*, on the other hand, is less well-characterised, although it has been shown to be important for gametophyte and embryo development (Cigliano et al., 2013). Interestingly, both *AtHDA19* and *AtHDA6* together are involved in both seed development and abiotic stress response, where they act to repress the expression of seed maturation genes during germination (such as *ABI3* and *OLEOSIN1*) and ABA-regulated stress response genes during vegetative drought stress (Chen and Wu, 2010; X. Liu et al., 2014). This raises the interesting possibility that *XhHDA3* is involved in attenuating stress-related gene expression during desiccation, or perhaps maintains repression of a subset of seed maturation genes that are not expressed during VDT. Alternatively, *XhHDA3* mRNA or protein may accumulate during desiccation and acts during rehydration to repress the ABA signalling network, similarly to what would occur during germination. The only other up-regulated HDAC, *XhHDA4*, was found to have greatest similarity to the *A. thaliana* proteins *AtHDA5* and *AtHDA18*. *AtHDA18* is believed to have arisen from the ancestral version of *AtHDA5* via a dicot-specific duplication event, thus *XhHDA4* may represent the monocot orthologue of *AtHDA5* in *X. humilis* (Pandey, 2002). Unfortunately, *AtHDA5* function has not been extensively investigated, although it has been implicated in the regulation of flowering time (Luo et al., 2015). As the transcript abundance of this gene is only elevated at water contents below 40% RWC, it may be important for the very late stages of desiccation response; alternatively, this increase may represent mRNA that has been accumulated to be translated during rehydration.

H3K9 methylation may delineate activated and silenced genome regions during desiccation

By far the largest class of histone-modifying enzyme genes identified as DE in *X. humilis* during desiccation were histone methyltransferases, with over 20 distinct transcripts present in the dataset. However, with the exception of *XhHMT20*, every single HMT gene was down-regulated after 60% RWC irrespective of predicted family or histone specificity. Additionally, although only four HDM genes were identified in the *X. humilis* transcriptome, only one – *XhHDM3* – was similarly not down-regulated. Interestingly, both of these genes are predicted to encode proteins involved in H3K9 methylation. *XhHMT20* is most closely related to the Class V Su(var) family of HMTs, which are important for H3K9 deposition, whereas *XhHDM3* is predicted to be a member of the JMJ2/KDM4 family of H3K9 demethylases, such as the rice HDM *JMJ706* (Lu et al., 2008; Naumann et al., 2005; Sun and Zhou, 2008; Yu et al., 2009). Taken together, this suggests that H3K9 methylation may be of particular importance during desiccation in *X. humilis*.

H3K9 methylation is important for regulation of gene expression, particularly by the linked processes of heterochromatin formation and DNA methylation (Yu et al., 2009). In plants, the SUVH family **o**f HMTs are essential for the formation of heterochromatin and regulation of DNA methylation (Jackson et al., 2004). In *A. thaliana*, this function is controlled in a partially redundant fashion by *SUVH1*, *SUVH4-6* and *SUVH9*; *SUVH2* is believed to have a more central role as *suvh2* is the only SUVH gene mutant to exhibit large changes in global H3K9 methylation patterns (Naumann et al., 2005). Antagonism of these heterochromatin marks by Jumonji C-domain containing (jmjC) proteins is seen in many organisms, including fission yeast (*S. pombe*), filamentous fungi (*N. crassa*) and *A. thaliana* (Tamaru, 2010). The rice JmjC HDM, *JMJ706*, is constitutively expressed in vegetative tissue and is

involved in at least flower organ development, where it co-localises with and antagonises the SUVH group HMTs in heterochromatic regions of the genome (Sun and Zhou, 2008).

Heterochromatin and DNA methylation are generally associated with silenced regions of the genome; for example, centromeric and telomeric regions, developmentally repressed genes and transposable elements (Yu et al., 2009). During seed development in *A. thaliana*, widespread heterochromatin formation and associated chromatin compaction is also believed to protect the genome from the stresses of maturation drying, as the more compact structure limits access to genomic DNA. A similar process is predicted to occur in the resurrection plant *C. plantagineum*, suggesting that this may be a common response to desiccation (van Zanten et al., 2012). The results reported here could thus support a model whereby *XhHMT20* mediates H3K9 methylation deposition during dehydration stress, perhaps as a regulator of desiccation-related heterochromatin formation and/or DNA methylation. *XhHDM3* could act to suppress this effect, possibly in distinct genomic regions associated with desiccation-specific gene expression.

Seed-specific gene de-repression during desiccation is not associated with unique patterns of PRC2 or PRC1 gene expression

In this study, I analysed transcripts coding for the various subunits of the PRC2 and PRC1 histone-modification complexes in an attempt to confirm a role for PRC2-mediated gene silencing during VDT in *X. humilis.*

Representatives of each of the PRC2 subunits could be found in the *X. humilis* transcriptome, including three E(z) orthologues (*XhHMT8, XhHMT10* and *XhHMT11*), seven P55 orthologues (*XhMSI1-7*), one Su(z) orthologue (*XhEMF2*) and one ESC orthologue (*XhFIE*).

With the exception of *XhFIE* and *XhMSI4*, which were not differentially expressed, all identified PRC2 subunit genes were heavily downregulated after 60% RWC in *X. humilis*. Multiple genes believed to be associated with the plant PRC1 complex were also identified, including an LHP1 orthologue (*XhLHP1*), two RING1 orthologues (*XhRING1.1-2*) and four DRIP/BMI orthologues (*XhDRIP1-4*). Of these genes only *LHP1* was significantly down-regulated, the remainder were not DE. These results therefore confirm and expand upon the original *X. humilis* microarray data collected by Shen et al. 2014: namely, the PRC2 methyl-transferase *SWINGER*, as well as the majority of PRC2 components, is down-regulated during desiccation in *X. humilis*. However, the increased scope of an RNA-Seq experiment has made it apparent that the expression profile of these PRC2 subunit genes during desiccation is by no means unique amongst histone modifying factor genes, and is in fact shared by the vast majority of chromatin-modifying protein gene transcripts irrespective of target histone and/or modification. A similar trend is also seen for all histone protein transcripts and, as discussed above, may suggest a general reduction in nucleosome modification and re-arrangement during the later stages of desiccation.

Although it is possible that the reduced levels of PRC2 during desiccation could still lead to a global reduction in genomic H3K27me3, a more in-depth analysis of the expression of all predicted PRC2 target genes in *X. humilis* (for comparison, 15% of genes in *A. thaliana* are marked by H3K27me3) is required to identify whether PRC2-silenced genes are significantly enriched amongst up-regulated genes during desiccation. Many seed maturation genes have been shown to be activated in desiccating *X. humilis* in this and previous studies, suggesting that H3K27me3-mediated gene silencing must be lifted from at least a subset of desiccation-induced genes (Illing et al., 2005; Shen, 2014). Our results, however, suggest that this may

not achieved by specific global antagonism of the PCR2/PCR1 complexes. Transcript abundance of the PRC1/PRC2 subunits did not show a unique profile amongst histone modification enzymes in general, and the expression profile of these genes did not correlate with the up-regulation of many of the seed maturation-associated genes identified in this study, such as many LEAs and oleosins, which were up-regulated prior to any change in PRC2 subunit gene expression. The down-regulation of PRC2 subunit genes after 60% RWC could be associated with the increased expression of seed genes at lower water contents compared to 80% RWC; however, it is equally possible that the enhanced up-regulation of these genes during late desiccation is due to increased ABA perception and/or signalling at this stage (due to the very high up-regulation of PYLs, group II SnRK2 kinases and downstream effectors such as *XhABF1*). Thus, these results do not rule out the possibility that many of the genes specifically induced during VDT are epigenetically regulated; in fact, it seems likely that histone acetylation and H3K9 methylation could play an important part of the desiccation response. However, more evidence is required to specifically link PRC2 and PRC1 gene activity, and by extension H3K27me3 methylation, to the induction or repression of desiccation-responsive genes as has been proposed previously.

3.3.6 General conclusion

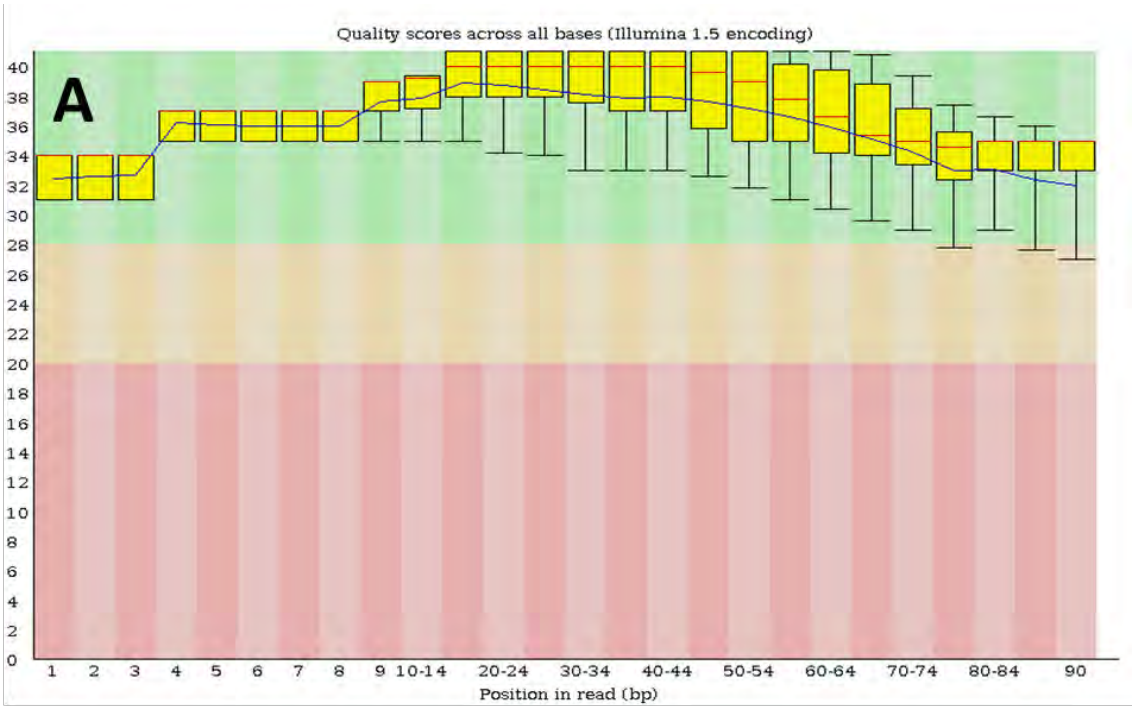
In summary, the data presented in this chapter is not consistent with a role for seed-maturation master regulatory TFs in the induction of VDT in *X. humilis*, as hypothesised at the outset of this study (see the Introduction, Fig. 1.8). Gene expression analysis of the protein-coding desiccation transcriptome of *X. humilis* has strengthened support for a role for ABA in resurrection plant VDT, as ABA biosynthetic genes and multiple components of the ABA signalling network were up-regulated in desiccating leaves. However, the only up-

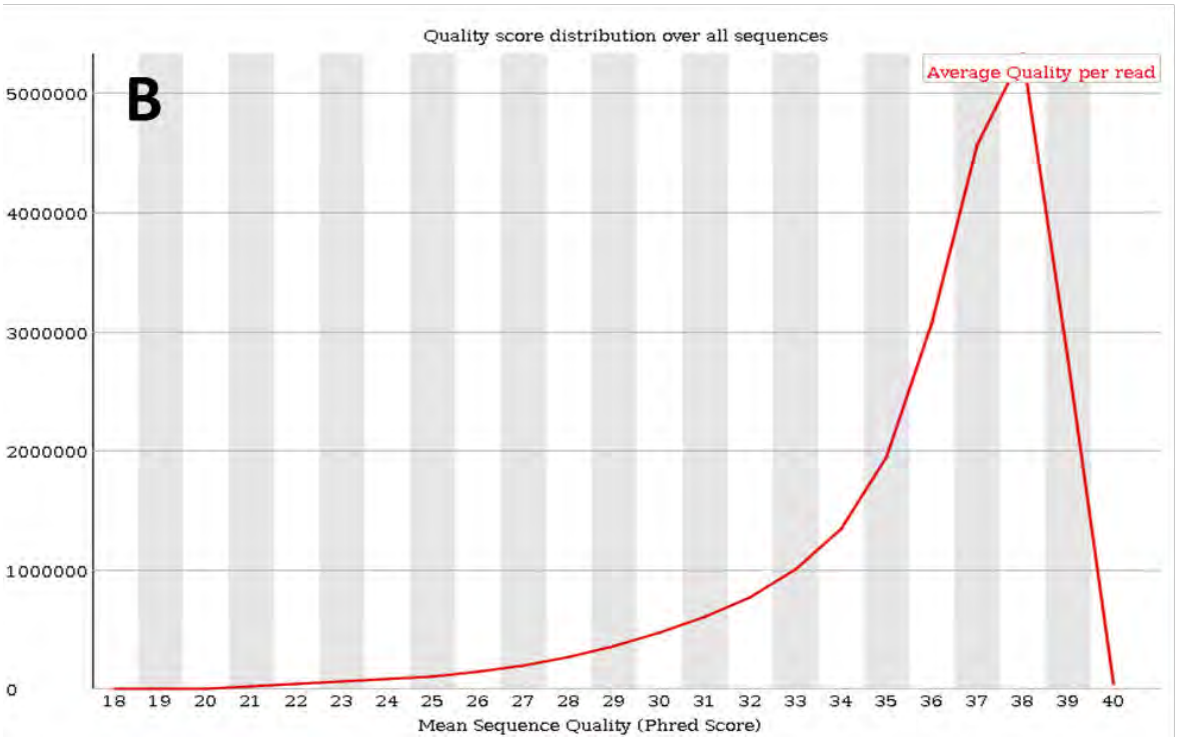
regulated bZIP ABRE-binding factor (ABF) genes were most closely related to those that control abiotic stress response in *A. thaliana* – as opposed to seed development – implying that VDT may be regulated by vegetative stress factors. VDT is however associated with the up-regulation of seed-specific genes, such as oleosins, caleosins and the antioxidant *PER1*, as has been shown in previous studies; furthermore, a very large number of LEA protein genes were upregulated in desiccating tissue, including many from seed-specific families. This suggests that some seed maturation genes are de-repressed in adult tissues, but there was no clear evidence that this correlated with the expression of PRC2 subunit genes, which silence these genes in other species by deposition of the H3K27me3 epigenetic mark. PRC2 genes, along with the majority of other histone modification enzyme and histone protein gene transcripts, were heavily down-regulated in desiccating tissue. The exception to this were a small number of histone variants, including a desiccation-responsive H1 most closely related to the drought-responsive *AtH1.3* found in *A. thaliana*, as well as several histone modifying enzymes: an H3K9 HMT and H3K9 HDM, which could be cooperatively involved in heterochromatin formation, and various HAT/HDACs. This dataset should provide valuable insight into many of the key processes and genes involved in regulating VDT in *X. humilis*.

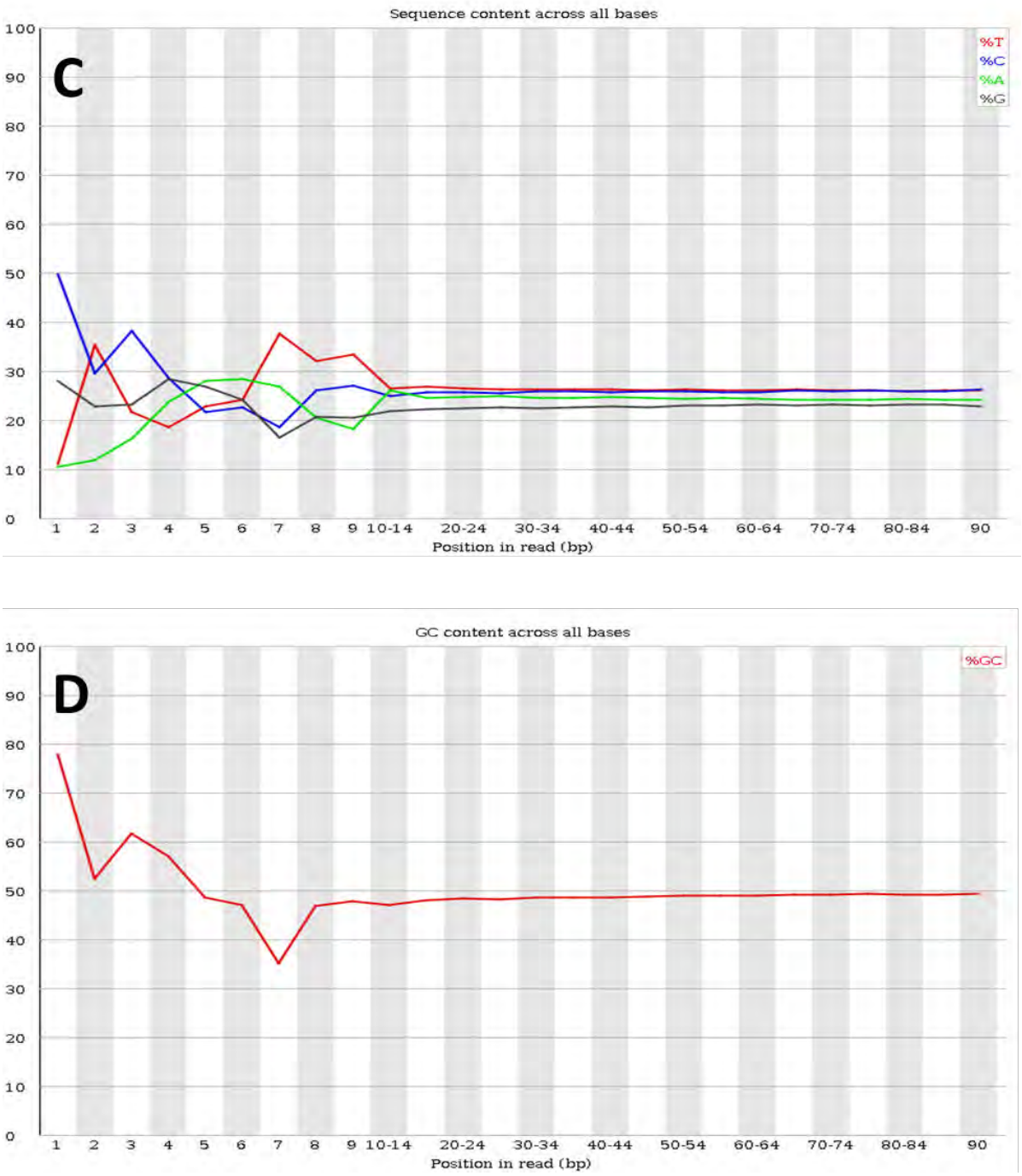
3. 4 Supplementary Data

100% ± 9% RWC		80% ± 5% RWC		60% ± 5% RWC		40% ± 6% RWC		5% ± 2% RWC	
Leaf ID	RWC	Leaf ID	RWC	Leaf ID	RWC	Leaf ID	RWC	Leaf ID	RWC
(A)003	107.5%	(B)034	80.2%	(A)016	64.8%	(A)017	45.7%	(A)091	5.1%
(A)009	101.3%	(B)035	79.0%	(A)011	64.6%	(A)027	40.4%	(A)102	4.9%
(A)002	100.7%	(B)031	78.9%	(A)010	60.5%	(A)031	40.0%	(A)045	4.8%
(A)004	99.5%	(B)032	77.3%	(A)006	59.4%	(A)020	35.9%	(A)099	4.6%
(A)008	97.5%	(B)030	75.0%	(A)005	55.8%	(A)033	35.1%	(A)054	4.5%
(B)019	112.6%	(C)001	85.2%	(B)015	57.7%	(B)052	40.9%	(B)066	6.7%
(B)002	105.0%	(C)011	82.6%	(C)003	64.3%	(B)055	40.4%	(B)067	5.7%
(B)012	102.4%			(C)018	63.6%	(B)047	38.8%	(B)070	5.2%
(B)001	99.5%					(B)057	38.2%	(B)083	5.1%
(B)018	97.5%					(B)013	37.3%	(B)092	4.4%
(C)009	120.7%					(C)039	40.5%	(C)097	5.4%
(C)002	108.7%					(C)032	37.3%	(C)074	5.1%
(C)007	108.4%					(C)033	36.2%	(C)066	5.0%
(C)005	100.8%							(C)049	4.6%
(C)012	100.7%							(C)047	4.6%

Supplementary Table S3.1. Sample IDs and RWC for sample *Xerophyta* leaves at our selected RWC. Leaves at approximately 100%, 80%, 60%, 40% or 5% RWC were chosen for downstream total RNA extraction.


Quality scores across all bases (Illumina 1.5 encoding)
A
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10-14 20-24 30-34 40-44 50-54 60-64 70-74 80-84 90
Position in read (bp)


Quality score distribution over all sequences
B
Average Quality per read
5000000
4000000
3000000
2000000
1000000
0
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Mean Sequence Quality (Phred Score)



Supplementary Figure S3.1. Representative sequencing quality statistics as determined by FASTQC. A) The PHRED quality score distribution across each nucleotide per read in the read library, B) Average PHRED quality score for each read across the library, C&D) Sequence content (AGCT) and GC percentage distribution across the length of each read in the library, showing the non-random distribution of bases in the first <10 base pairs that is distinctive of RNA-Seq library preparation. Results shown here are for 5% RWC sample A, but are comparable to all other libraries.

Gene Name (BLAST subject title)	Number of Gene Clusters	Highest E-value
Cluster 2 - (***Seed development; GO:0048316)***		
50S ribosomal protein L21, chloroplastic	2	3.0E-07
50S ribosomal protein L31, chloroplastic	1	4.0E-47
ABC transporter B family member 11	1	1.0E-58
ABSCISIC ACID-INSENSITIVE 5-like protein 4	1	1.0E-66
ABSCISIC ACID-INSENSITIVE 5-like protein 5	1	1.0E-71
AP2-like ethylene-responsive transcription factor TOE3	1	1.0E-50
Abscisic acid 8'-hydroxylase 1	9	3.0E-84
Ankyrin repeat domain-containing protein, chloroplastic	1	4.0E-119
Aspartic proteinase PCS1	1	7.0E-116
Auxin-responsive protein IAA31	1	3.0E-29
Cycloartenol-C-24-methyltransferase 1	2	0.0E+00
E3 ubiquitin-protein ligase ATL59	1	2.0E-11
E3 ubiquitin-protein ligase At3g02290	1	6.0E-18
E3 ubiquitin-protein ligase Os04g0590900	1	3.0E-68
Ethylene-responsive transcription factor CRF1	1	7.0E-28
Ethylene-responsive transcription factor CRF4	1	6.0E-36
Ethylene-responsive transcription factor ERF073	1	7.0E-16
Ethylene-responsive transcription factor RAP2-13	1	5.0E-43
Eukaryotic translation initiation factor 6-2	1	7.0E-68
Growth-regulating factor 1	1	2.0E-47
MATE efflux family protein LAL5	1	6.0E-09
NAC domain-containing protein 74	2	5.0E-83
Nuclear transcription factor Y subunit A-6	1	4.0E-31
Nuclear transcription factor Y subunit A-7	1	3.0E-08
PLASMODESMATA CALLOSE-BINDING PROTEIN 3	1	2.0E-30
Pentatricopeptide repeat-containing protein At5g15010, mitochondrial	1	0.0E+00
Phosphate transporter PHO1-2	1	0.0E+00
Probable LRR receptor-like protein kinase At1g51890	1	1.0E-84
Probable LRR receptor-like serine/threonine-protein kinase At4g08850	1	0.0E+00
Probable cellulose synthase A catalytic subunit 9 [UDP-forming]	1	0.0E+00
Probable indole-3-pyruvate monooxygenase YUCCA8	2	0.0E+00
Probable prefoldin subunit 2	1	9.0E-52
Protein ROOT INITIATION DEFECTIVE 3	1	4.0E-129
Protein kinase APK1B, chloroplastic	1	5.0E-147
Pyrophosphate-fructose 6-phosphate 1-phosphotransferase subunit beta 1	1	4.0E-39
QWRF motif-containing protein 8	1	6.0E-59
Ras-related protein Rab7	1	5.0E-96
Receptor-like protein kinase At3g21340	1	5.0E-95
Serine/threonine-protein kinase PBS1	1	0.0E+00
Splicing factor U2af small subunit B	1	1.0E-112
T-complex protein 1 subunit theta	1	2.0E-35
Transcription factor DIVARICATA	1	1.0E-87
U-box domain-containing protein 33	2	4.0E-68
Ubiquitin-conjugating enzyme E2 28	1	4.0E-103
Ubiquitin-conjugating enzyme E2-17 kDa	1	2.0E-88
WD-40 repeat-containing protein MSI1	1	0.0E+00
Wall-associated receptor kinase 3	1	2.0E-103
Cluster 3 - ***(Seed dormancy; GO:0010162)***		
1-Cys peroxiredoxin	3	1.0E-123
1-Cys peroxiredoxin PER1	1	9.0E-64
E3 ubiquitin-protein ligase listerin	1	2.0E-06
Ethylene-responsive transcription factor ERF060	1	5.0E-39
Ethylene-responsive transcription factor RAP2-13	1	9.0E-35
Iron-sulfur assembly protein IscA-like 1, mitochondrial	1	8.0E-34

LOB domain-containing protein 37	1	2.E-52
Oleosin 16 kDa	1	3.E-32
Oleosin 5	3	2.E-28
Peroxygenase	3	2.E-54
Peroxygenase 1	1	1.E-62
Probable peroxygenase 3	1	4.E-106
Probable protein phosphatase 2C 51	1	5.E-60
Probable protein phosphatase 2C 8	2	9.E-81
Probable protein phosphatase 2C 9	2	3.E-112
mRNA cap guanine-N7 methyltransferase 1	1	5.E-79

Cluster 7 - ***(Seed dormancy; GO:0010162)***

Em protein H5	1	2.E-15
Ethylene-responsive transcription factor ERF071	1	2.E-33
Ethylene-responsive transcription factor ERF110	1	4.E-24
Late embryogenesis abundant protein B19.4	1	2.E-62
Oleosin 5	1	7.E-28
Peroxygenase	5	1.E-111
Probable protein phosphatase 2C 24	2	1.E-66
Probable protein phosphatase 2C 51	1	2.E-29
Probable protein phosphatase 2C 75	1	1.E-79
Probable protein phosphatase 2C 8	8	2.E-111
Probable protein phosphatase 2C 9	7	4.E-112
Protein REVEILLE 1	2	4.E-33
Protein phosphatase 2C 3	3	2.E-93

Cluster 8 - ***(Seed dormancy; GO:0010162)***

21 kDa protein	1	3.E-23
29 kDa ribonucleoprotein B, chloroplastic	1	7.E-67
9-cis-epoxycarotenoid dioxygenase 1, chloroplastic	1	0.E+00
ACT domain-containing protein ACR3	1	8.E-96
ACT domain-containing protein ACR8	1	6.E-146
Adenine nucleotide transporter BT1, chloroplastic/mitochondrial	3	5.E-155
Autophagy-related protein 36	1	5.E-10
BTB/POZ domain-containing protein At5g66560	1	5.E-29
Dehydration-responsive element-binding protein 2A	4	9.E-37
Dehydration-responsive element-binding protein 2C	1	5.E-31
E3 ubiquitin-protein ligase ATL41	1	9.E-20
E3 ubiquitin-protein ligase ATL6	1	4.E-52
E3 ubiquitin-protein ligase RNF185	1	2.E-23
Em protein	1	3.E-13
Ethylene-responsive transcription factor 11	2	5.E-18
Ethylene-responsive transcription factor 4	2	7.E-19
Ethylene-responsive transcription factor 8	2	1.E-22
Ethylene-responsive transcription factor ERF010	1	3.E-33
Ethylene-responsive transcription factor RAP2-10	1	3.E-32
F-box protein PP2-A13	1	7.E-116
F-box protein PP2-B10	1	7.E-67
Ferritin-1, chloroplastic	1	8.E-17
Ferritin-3, chloroplastic	9	6.E-26
Glucan endo-1,3-beta-glucosidase	7	3.E-172
Glucan endo-1,3-beta-glucosidase 14	2	1.E-145
Glucan endo-1,3-beta-glucosidase 6	1	4.E-164
Glucose-6-phosphate/phosphate translocator 1, chloroplastic	1	2.E-157
Glucose-6-phosphate/phosphate translocator 2, chloroplastic	1	4.E-70
Glycine-rich protein 2	1	1.E-33
Late embryogenesis abundant protein B19.3	1	1.E-60
Late embryogenesis abundant protein B19.4	3	4.E-18
Lysine-specific demethylase JMJ706	2	0.E+00

Lysine-specific demethylase JMJ706	2	0.E+00
Mannose-6-phosphate isomerase 1	1	2.E-135
NAC domain-containing protein 2	1	2.E-59
Nuclear transcription factor Y subunit A-7	2	1.E-44
PHD and RING finger domain-containing protein 1	1	2.E-10
Pentatricopeptide repeat-containing protein At1g26900, mitochondrial	1	3.E-126
Peroxygenase	3	9.E-67
Phospholipase D beta 1	3	4.E-122
Phospholipase D delta	2	0.E+00
Probable 6-phosphogluconolactonase 4, chloroplastic	3	6.E-110
Probable LRR receptor-like serine/threonine-protein kinase At3g47570	1	1.E-74
Probable WRKY transcription factor 33	3	3.E-112
Probable WRKY transcription factor 4	2	3.E-43
Probable alpha,alpha-trehalose-phosphate synthase [UDP-forming] 9	0	3.E-79
Probable histone deacetylase 19	1	2.E-171
Probable indole-3-pyruvate monooxygenase YUCCA8	1	0.E+00
Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial	3	4.E-94
Probable protein phosphatase 2C 24	2	1.E-42
Probable protein phosphatase 2C 68	2	8.E-72
Probable protein phosphatase 2C 75	4	2.E-82
Probable protein phosphatase 2C 8	3	7.E-67
Probable protein phosphatase 2C 9	3	7.E-30
Probable receptor-like protein kinase At5g15080	1	3.E-54
Probable receptor-like protein kinase At5g47070	1	1.E-128
Protein EARLY FLOWERING 3	1	5.E-33
Protein PHR1-LIKE 1	2	2.E-46
Protein TOPLESS	1	0.E+00
Protein TRANSPARENT TESTA 12	3	3.E-108
Protein disulfide isomerase-like 2-2	1	2.E-106
Protein phosphatase 2C 3	4	1.E-75
Protein phosphatase 2C 37	1	6.E-93
RING-H2 finger protein ATL2	1	1.E-18
RNA-binding protein CP29B, chloroplastic	1	3.E-44
Serine decarboxylase 1	1	0.E+00
Serine/threonine-protein kinase D6PKL2	1	0.E+00
Serine/threonine-protein kinase OXI1	1	2.E-112
Stem-specific protein TSJT1	1	2.E-32
Transcription factor DIVARICATA	2	9.E-16
WRKY transcription factor 6	1	1.E-112
Zinc finger CCCH domain-containing protein 2	8	4.E-46
Zinc finger protein AZF1	2	5.E-35
Zinc finger protein AZF2	1	2.E-29
Zinc finger protein ZAT10	2	3.E-34
Zinc finger protein ZAT11	6	1.E-19

Cluster 11 - ***(Seed dormancy; GO:0010162)***

Cytochrome P450 90B1	1	0.E+00
Gibberellin-regulated protein 6	3	6.E-37
Glucan endo-1,3-beta-glucosidase 1	1	0.E+00
Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1	2	0.E+00
Mini zinc finger protein 1	1	8.E-20
Mini zinc finger protein 2	1	5.E-23
PLASMODESMATA CALLOSE-BINDING PROTEIN 3	1	3.E-29
Protein GAST1	1	3.E-31
Snakin-1	1	4.E-21
Transcription factor bHLH49	1	4.E-34
Transcription factor bHLH62	1	8.E-54
Transcription factor bHLH63	1	4.E-39
Tubulin beta-1 chain	7	0.E+00

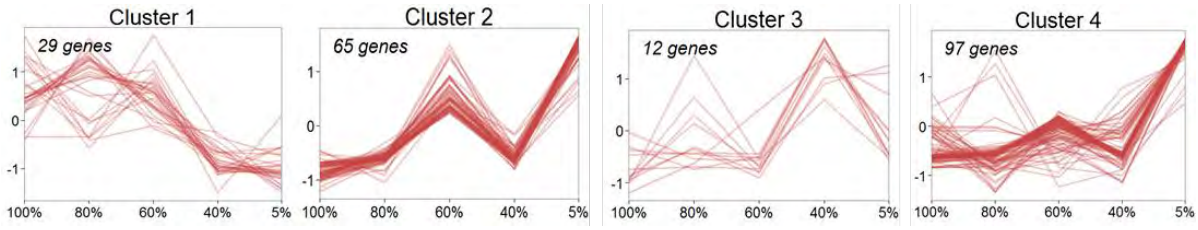
Tubulin beta-1 chain	7	0.E+00
Tubulin beta-4 chain	1	0.E+00
Tubulin beta-5 chain	2	0.E+00
Tubulin beta-6 chain	2	0.E+00
Tubulin beta-7 chain	1	2.E-152

Cluster 12 - ***(Seed dormancy; GO:0010162)***

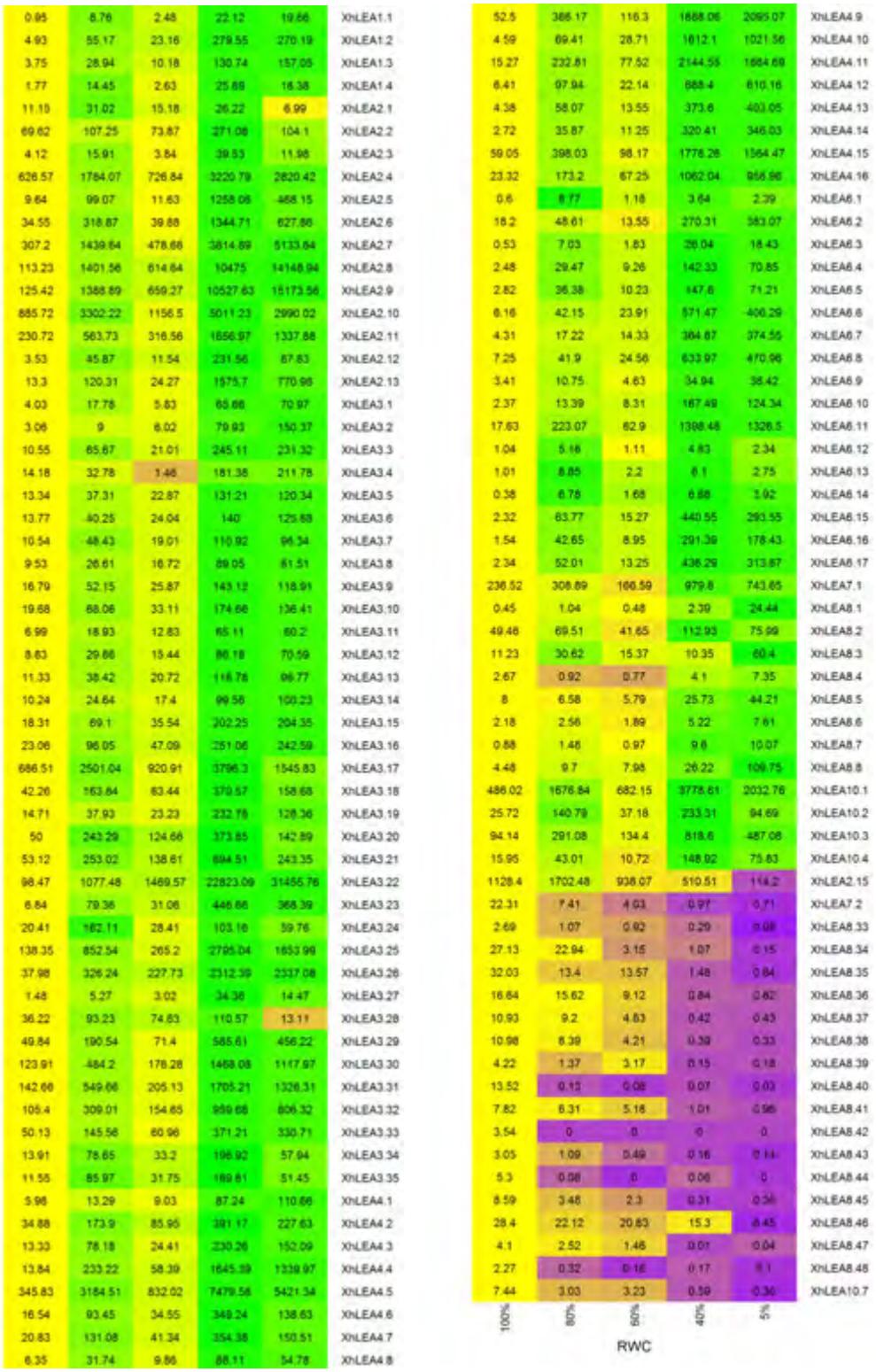
24-methylenesterol C-methyltransferase 2	1	0.E+00
ABC transporter B family member 1	2	9.E-100
ABC transporter B family member 19	1	0.E+00
ACT domain-containing protein ACR3	1	2.E-124
ARF guanine-nucleotide exchange factor GNOM	1	0.E+00
Auxin efflux carrier component 1	3	7.E-126
Auxin response factor 19	1	0.E+00
Auxin-responsive protein IAA21	1	9.E-62
Auxin-responsive protein IAA27	1	1.E-88
BEL1-like homeodomain protein 2	2	8.E-70
BTB/POZ domain-containing protein At1g67900	1	0.E+00
BTB/POZ domain-containing protein NPY1	1	0.E+00
Chloroplast stem-loop binding protein of 41 kDa b, chloroplastic	3	0.E+00
Cytochrome P450 90B1	1	0.E+00
Cytochrome P450 94A1	1	1.E-122
DELLA protein DWARF8	3	0.E+00
DNA-damage-repair/toleration protein DRT100	1	8.E-29
Disease resistance protein RFL1	1	5.E-94
E3 ubiquitin-protein ligase Os04g0590900	2	1.E-70
ETHYLENE INSENSITIVE 3-like 1 protein	1	2.E-109
Ethylene-responsive transcription factor 5	1	6.E-16
Ethylene-responsive transcription factor CRF1	1	5.E-18
Ethylene-responsive transcription factor RAP2-13	5	6.E-27
Gibberellin-regulated protein 1	1	1.E-13
Glucan endo-1,3-beta-glucosidase 13	1	0.E+00
Glucan endo-1,3-beta-glucosidase 7	1	0.E+00
Homeobox protein BEL1 homolog	1	2.E-26
LRR receptor-like serine/threonine-protein kinase EFR	1	2.E-122
LRR receptor-like serine/threonine-protein kinase ERECTA	2	5.E-62
LRR receptor-like serine/threonine-protein kinase FLS2	1	7.E-61
Leucine-rich repeat receptor-like protein kinase PXL1	1	4.E-43
Leucine-rich repeat receptor-like serine/threonine-protein kinase At1g17230	1	6.E-40
Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM1	1	0.E+00
Leucine-rich repeat receptor-like serine/threonine/tyrosine-protein kinase SOBIR1	1	8.E-117
Leucine-rich repeat receptor-like tyrosine-protein kinase At2g41820	1	1.E-44
MATE efflux family protein 9	1	6.E-129
Malate dehydrogenase, glyoxysomal	2	6.E-166
Myb family transcription factor APL	1	2.E-43
NAC domain-containing protein 21/22	1	2.E-67
Obtusifoliol 14-alpha demethylase	1	0.E+00
Pentatricopeptide repeat-containing protein At1g79490, mitochondrial	1	0.E+00
Pentatricopeptide repeat-containing protein At5g46100	1	5.E-151
Polyadenylate-binding protein 2	1	0.E+00
Probable LRR receptor-like serine/threonine-protein kinase At4g08850	9	9.E-58
Probable LRR receptor-like serine/threonine-protein kinase At5g10290	1	8.E-35
Probable LRR receptor-like serine/threonine-protein kinase MRH1	2	9.E-119
Probable WRKY transcription factor 11	1	1.E-69
Probable WRKY transcription factor 65	1	3.E-46
Probable auxin efflux carrier component 1b	1	2.E-19
Probable inactive leucine-rich repeat receptor-like protein kinase At1g66830	1	2.E-174
Probable inactive leucine-rich repeat receptor-like protein kinase At3g03770	2	0.E+00

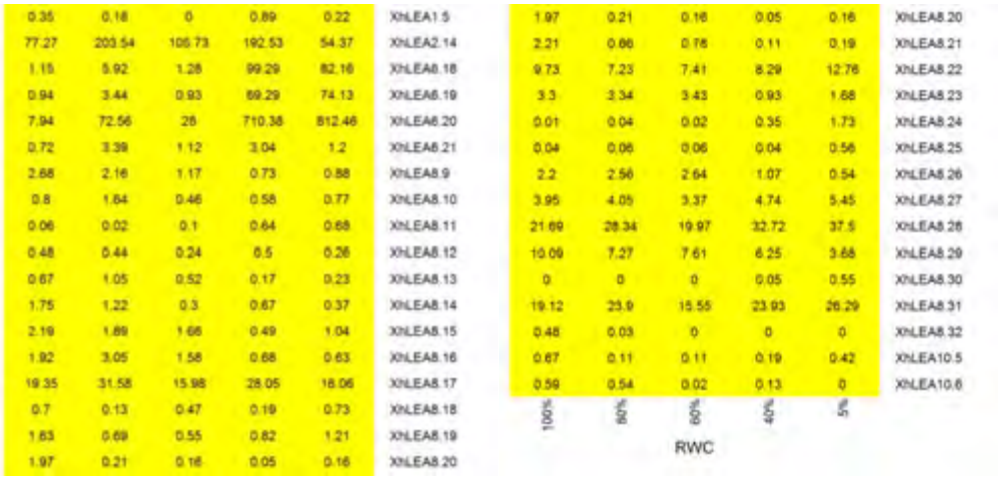
Probable leucine-rich repeat receptor-like protein kinase At1g35710	14	3.E-45
Probable leucine-rich repeat receptor-like protein kinase At5g63930	1	6.E-33
Probable receptor-like protein kinase At1g33260	1	4.E-74
Probable receptor-like protein kinase At5g24010	1	0.E+00
Probably inactive leucine-rich repeat receptor-like protein kinase At5g48380	1	0.E+00
Protein ETHYLENE INSENSITIVE 3	1	1.E-153
Protein GAMETE EXPRESSED 1	1	1.E-07
Protein NSP-INTERACTING KINASE 3	1	0.E+00
Protein RADIALIS-like 3	2	1.E-22
Protein RADIALIS-like 6	1	3.E-14
Protein STRUBBELIG-RECEPTOR FAMILY 1	1	4.E-102
Protein STRUBBELIG-RECEPTOR FAMILY 6	1	0.E+00
Protein TRANSPARENT TESTA 12	4	2.E-125
Protein argonaute 1A	1	0.E+00
Protein argonaute 4B	1	0.E+00
RING-H2 finger protein ATL13	1	7.E-63
RING-H2 finger protein ATL7	1	2.E-42
Receptor-like protein 12	1	6.E-46
Receptor-like protein kinase HSL1	5	5.E-158
Receptor-like serine/threonine-protein kinase ALE2	2	2.E-143
Root phototropism protein 2	1	7.E-92
Rop guanine nucleotide exchange factor 1	1	0.E+00
Serine/threonine-protein kinase D6PKL2	2	1.E-151
Serine/threonine-protein kinase KIPK	1	0.E+00
Systemin receptor SR160	1	0.E+00
TITAN-like protein	1	2.E-23
Transcription factor MYC3	1	5.E-128
Transcription factor TCP2	2	3.E-37
Transcription factor TCP5	3	5.E-39
Transcription factor bHLH25	1	8.E-51
Transcription factor bHLH82	1	3.E-38
Transcriptional corepressor SEUSS	1	3.E-27
Trihelix transcription factor ASIL2	2	2.E-08
Tubby-like F-box protein 5	1	0.E+00
Two-component response regulator ARR8	2	2.E-56
Very-long-chain 3-oxoacyl-CoA reductase 1	1	2.E-71
WD-40 repeat-containing protein MSI1	1	8.E-154
Wall-associated receptor kinase 3	1	3.E-131
Xylulose 5-phosphate/phosphate translocator, chloroplastic	3	7.E-120

Supplementary Table S3.2. Top protein hits of *X. humilis* genes associated with enriched seed-related GO terms. Top BLASTX protein hits within the Swissprot database of *X. humilis* transcripts annotated by Blast2GO as having possible seed-related functions. Multiple transcripts with the same top hit within each cluster have been grouped, and the highest (least significant) e-value amongst those transcripts is shown.

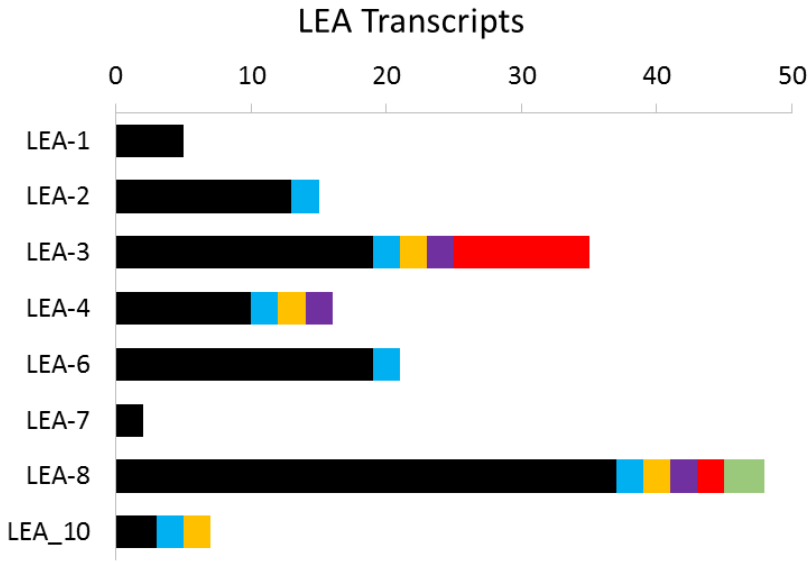


Supplementary Figure S3.2. Putative fungal gene expression clusters. K-means clustering of 203 transcripts predicted to be fungal in origin (FDR < 0.01; FC > 2) using normalised rlog expression values.

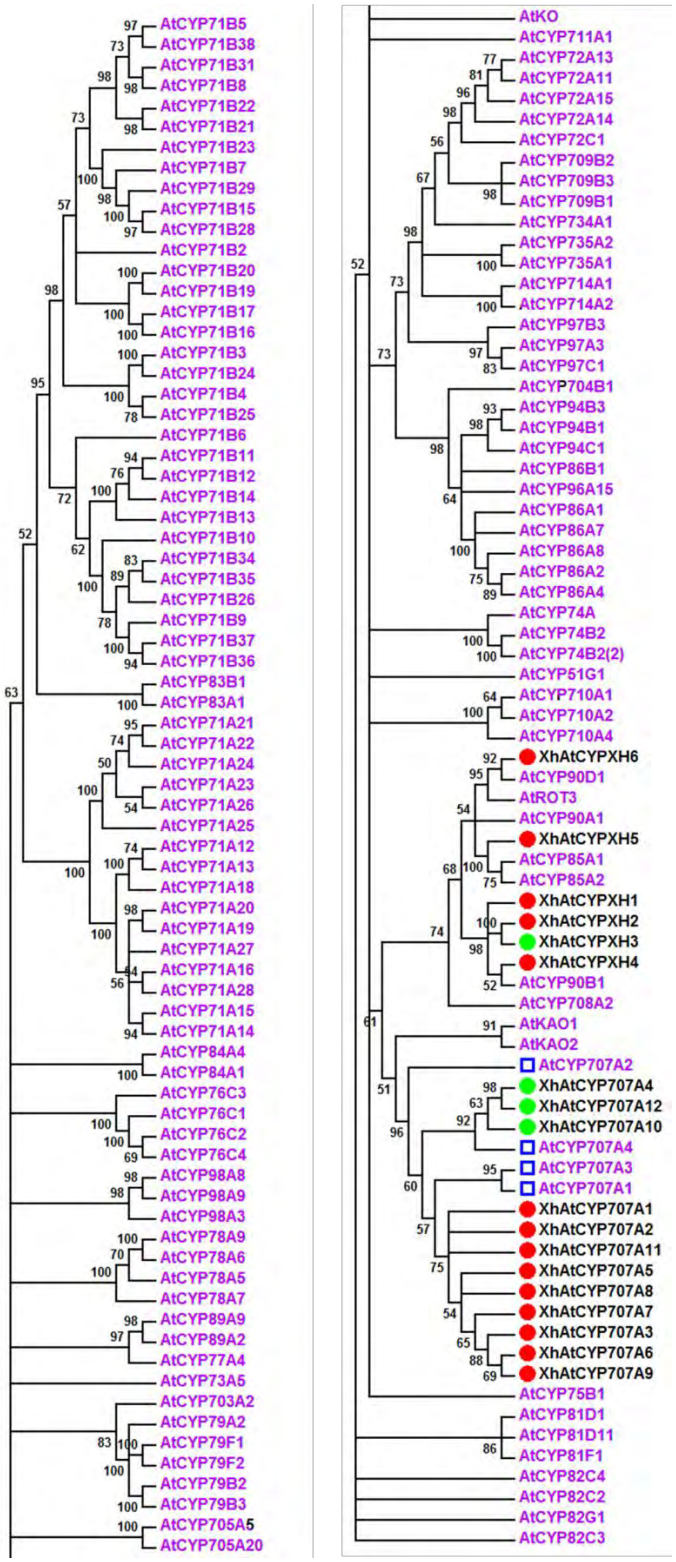

0.95 8.76 2.48 22.12 19.66 XhLEA1.1
4.93 55.17 23.16 279.55 270.19 XhLEA1.2
3.75 28.94 10.18 130.74 157.05 XhLEA1.3
1.77 14.45 2.63 25.69 16.38 XhLEA1.4
11.15 31.02 15.18 26.22 6.99 XhLEA2.1
69.62 107.25 73.87 271.06 104.1 XhLEA2.2
4.12 15.91 3.84 39.53 11.98 XhLEA2.3
626.57 1784.07 726.84 3220.79 2820.42 XhLEA2.4
9.64 99.07 11.63 1258.08 466.15 XhLEA2.5
34.55 318.87 39.88 1344.71 627.86 XhLEA2.6
307.2 1439.64 478.66 3814.89 5133.64 XhLEA2.7
113.23 1401.56 614.64 10475 14148.94 XhLEA2.8
125.42 1388.89 659.27 10527.63 15173.56 XhLEA2.9
885.72 3302.22 1156.5 5011.23 2990.02 XhLEA2.10
230.72 563.73 316.56 1656.97 1337.88 XhLEA2.11
3.53 45.87 11.54 231.56 67.83 XhLEA2.12
13.3 120.31 24.27 1575.7 770.98 XhLEA2.13
4.03 17.78 5.83 69.66 70.97 XhLEA3.1
3.06 9 6.02 79.93 150.37 XhLEA3.2
10.55 65.67 21.01 245.11 231.32 XhLEA3.3
14.18 32.78 1.46 181.38 211.78 XhLEA3.4
13.34 37.31 22.87 131.21 120.34 XhLEA3.5
13.77 40.25 24.04 140 125.68 XhLEA3.6
10.54 48.43 19.01 110.92 96.34 XhLEA3.7
9.53 26.61 16.72 89.05 61.51 XhLEA3.8
16.79 52.15 25.87 143.12 118.91 XhLEA3.9
19.68 88.06 33.11 174.66 136.41 XhLEA3.10
6.99 18.93 12.83 65.11 60.2 XhLEA3.11
8.63 29.66 15.44 86.18 70.59 XhLEA3.12
11.33 38.42 20.72 116.76 96.77 XhLEA3.13
10.24 24.64 17.4 99.56 100.23 XhLEA3.14
18.31 69.1 35.54 202.25 204.35 XhLEA3.15
23.06 96.05 47.09 251.06 242.59 XhLEA3.16
686.51 2501.04 920.91 3796.3 1545.83 XhLEA3.17
42.26 163.64 83.44 370.57 158.68 XhLEA3.18
14.71 37.93 23.23 232.78 126.36 XhLEA3.19
50 243.29 124.66 373.85 142.89 XhLEA3.20
53.12 253.02 138.61 894.51 243.35 XhLEA3.21
98.47 1077.48 1469.57 22823.09 31455.76 XhLEA3.22
6.84 79.36 31.06 446.86 368.39 XhLEA3.23
20.41 182.11 28.41 103.16 59.76 XhLEA3.24
138.35 852.54 265.2 2795.04 1693.99 XhLEA3.25
37.98 326.24 227.73 2312.39 2337.08 XhLEA3.26
1.48 5.27 3.02 34.36 14.47 XhLEA3.27
36.22 93.23 74.83 110.57 13.11 XhLEA3.28
49.84 190.54 71.4 585.61 456.22 XhLEA3.29
123.91 484.2 178.28 1468.08 1117.97 XhLEA3.30
142.66 549.06 205.13 1705.21 1326.31 XhLEA3.31
105.4 309.01 154.65 959.66 806.32 XhLEA3.32
50.13 145.56 60.96 371.21 330.71 XhLEA3.33
13.91 78.65 33.2 196.92 57.94 XhLEA3.34
11.55 85.97 31.75 169.81 51.45 XhLEA3.35
5.96 13.29 9.03 87.24 110.66 XhLEA4.1
34.88 173.9 85.95 391.17 227.63 XhLEA4.2
13.33 78.18 24.41 230.26 152.09 XhLEA4.3
13.84 233.22 58.39 1645.39 1339.97 XhLEA4.4
345.83 3184.51 832.02 7479.56 5421.34 XhLEA4.5
16.54 93.45 34.55 349.24 138.63 XhLEA4.6
20.83 131.08 41.34 354.38 150.51 XhLEA4.7
6.35 31.74 9.86 88.11 34.78 XhLEA4.8
52.5 386.17 116.3 1888.06 2095.07 XhLEA4.9
4.59 69.41 28.71 1612.1 1021.56 XhLEA4.10
15.27 232.81 77.52 2144.55 1684.68 XhLEA4.11
6.41 97.94 22.14 888.4 610.16 XhLEA4.12
4.38 58.07 13.55 373.6 403.05 XhLEA4.13
2.72 35.87 11.25 320.41 346.03 XhLEA4.14
59.05 398.03 98.17 1778.26 1364.47 XhLEA4.15
23.32 173.2 67.25 1062.04 966.96 XhLEA4.16
0.6 8.77 1.18 3.64 2.39 XhLEA6.1
16.2 48.61 13.55 270.31 383.07 XhLEA6.2
0.53 7.03 1.83 28.04 18.43 XhLEA6.3
2.48 29.47 9.26 142.33 70.85 XhLEA6.4
2.82 36.38 10.23 147.6 71.21 XhLEA6.5
6.16 42.15 23.91 571.47 466.29 XhLEA6.6
4.31 17.22 14.33 384.87 374.55 XhLEA6.7
7.25 41.9 24.56 633.97 470.96 XhLEA6.8
3.41 10.75 4.63 34.94 38.42 XhLEA6.9
2.37 13.39 8.31 167.48 124.34 XhLEA6.10
17.63 223.07 62.9 1398.48 1328.5 XhLEA6.11
1.04 5.16 1.11 4.83 2.34 XhLEA6.12
1.01 8.85 2.2 6.1 2.75 XhLEA6.13
0.38 6.76 1.68 6.88 3.92 XhLEA6.14
2.32 63.77 15.27 440.55 293.55 XhLEA6.15
1.54 42.65 8.95 291.39 178.43 XhLEA6.16
2.34 52.01 13.25 436.29 313.87 XhLEA6.17
236.52 308.89 166.59 979.8 743.65 XhLEA7.1
0.45 1.04 0.48 2.39 24.44 XhLEA8.1
49.46 69.51 41.65 112.93 75.99 XhLEA8.2
11.23 30.62 15.37 10.35 60.4 XhLEA8.3
2.67 0.92 0.77 4.1 7.35 XhLEA8.4
8 6.58 5.79 25.73 44.21 XhLEA8.5
2.18 2.56 1.89 5.22 7.61 XhLEA8.6
0.88 1.46 0.97 9.6 10.07 XhLEA8.7
4.48 9.7 7.98 26.22 109.75 XhLEA8.8
486.02 1676.84 682.15 3778.61 2032.76 XhLEA10.1
25.72 140.79 37.18 233.31 94.69 XhLEA10.2
94.14 291.08 134.4 818.6 487.08 XhLEA10.3
15.95 43.01 10.72 148.92 75.83 XhLEA10.4
1128.4 1702.48 938.07 510.51 114.2 XhLEA2.15
22.31 7.41 4.63 0.97 0.71 XhLEA7.2
2.69 1.07 0.92 0.29 0.08 XhLEA8.33
27.13 22.94 3.15 1.07 0.15 XhLEA8.34
32.03 13.4 13.57 1.48 0.84 XhLEA8.35
16.64 15.62 9.12 0.84 0.62 XhLEA8.36
10.93 9.2 4.83 0.42 0.43 XhLEA8.37
10.98 6.39 4.21 0.38 0.33 XhLEA8.38
4.22 1.37 3.17 0.15 0.18 XhLEA8.39
13.52 0.13 0.08 0.07 0.03 XhLEA8.40
7.82 6.31 5.18 1.01 0.96 XhLEA8.41
3.54 0 0 0 0 XhLEA8.42
3.05 1.09 0.49 0.16 0.14 XhLEA8.43
5.3 0.06 0 0.06 0 XhLEA8.44
8.59 3.46 2.3 0.31 0.36 XhLEA8.45
28.4 22.12 20.83 15.3 8.45 XhLEA8.46
4.1 2.52 1.46 0.01 0.04 XhLEA8.47
2.27 0.32 0.16 0.17 0.1 XhLEA8.48
7.44 3.03 3.23 0.59 0.36 XhLEA10.7
100% 80% 60% 40% 5%
RWC



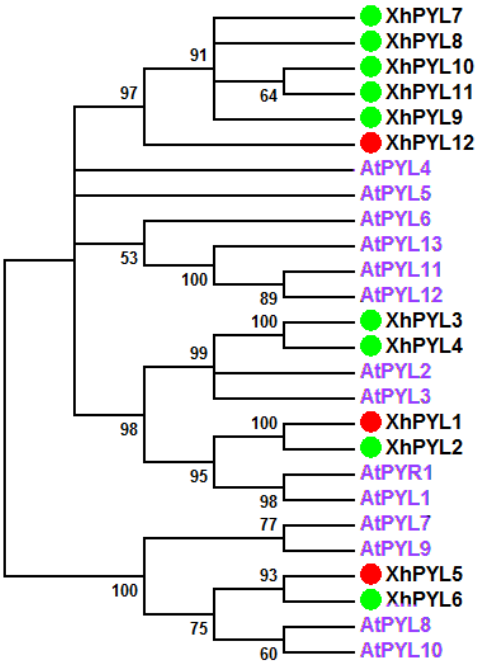
Supplementary Figure S3.3. Expression values for all identified LEA transcripts in *X. humilis*. The top half of the figure shows LEAs that were found to be DE during desiccation, whereas the bottom half contains LEAs that were NDE. LEA transcripts in each half are grouped by family. Cell values are FPKM, whereas cell colour denotes the difference to transcript abundance at 100% RWC (yellow).



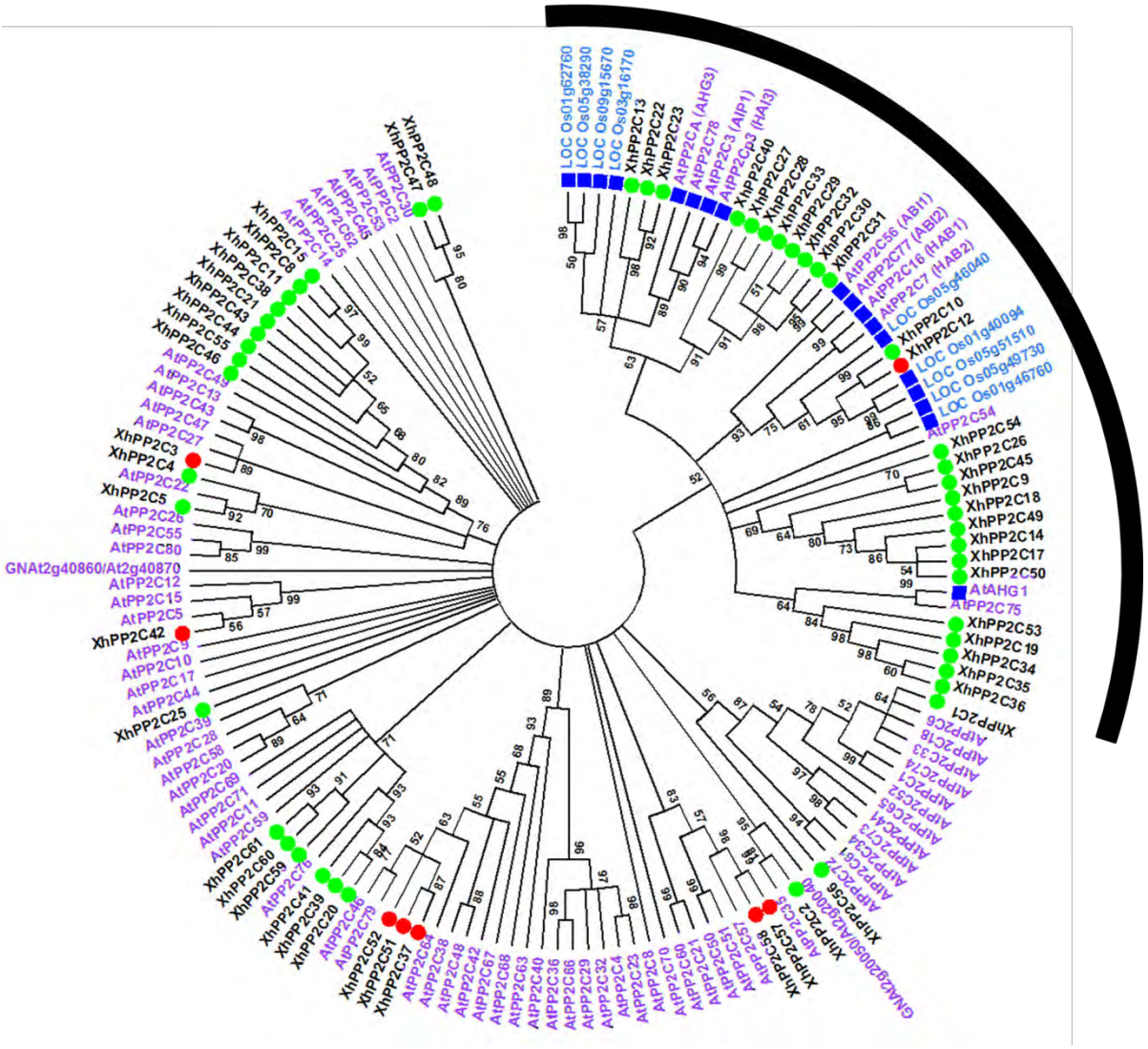
Supplementary Figure S3.4. The predicted number of LEA genes in *X. humilis* and their associated transcript isoforms. In order to assess the possible expansion of LEA genes in *X. humilis*, the LEA transcripts identified in the transcriptome were compared to scaffold sequences of the draft *X. humilis* genome assembly. Any transcripts that shared scaffold sequence (and thus contained a shared exon) were considered to be derived from a single LEA "gene", and were grouped accordingly. LEA genes with only a single isoform are coloured black, those with two or more isoforms are marked in colour (a single colour per gene per LEA family).



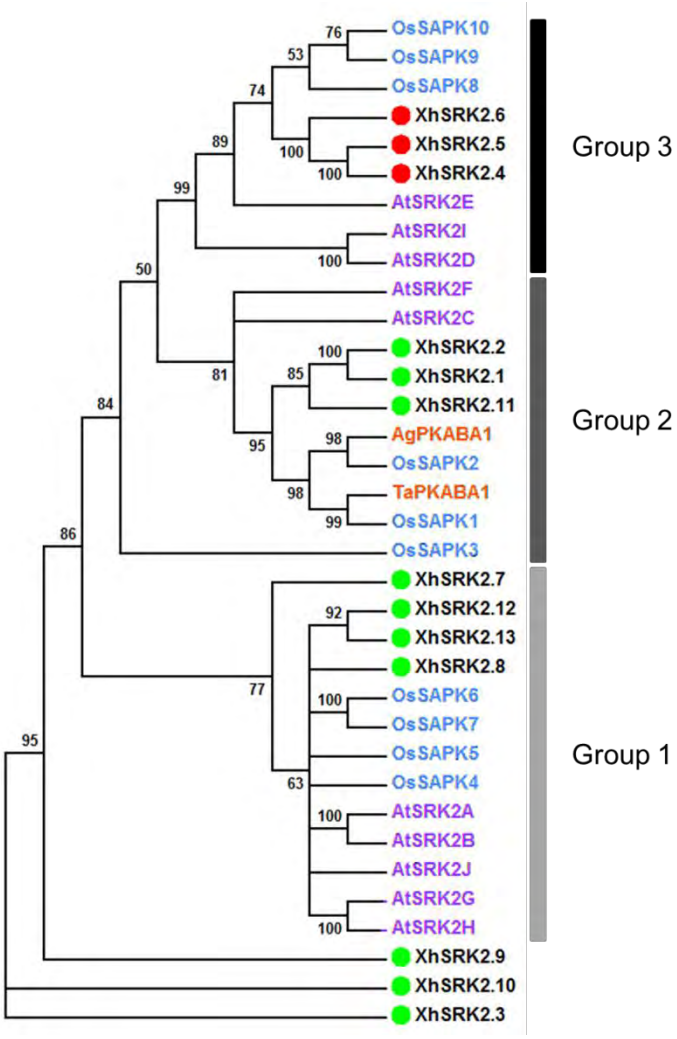
Supplementary Figure S3.5. Phylogenetic analysis of *X. humilis* cytochrome P450 (CYP) genes. An unrooted Neighbour-Joining phylogenetic tree of *X. humilis CYP707A*-like protein sequences (black) and those of *A. thaliana* (purple). Only twelve grouped with the *CYP707A* clade of *A. thaliana* (marked with open blue squares). Up-regulated genes in *X. humilis* are marked with a green circle, and those that were down-regulated are marked with a red circle. Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.



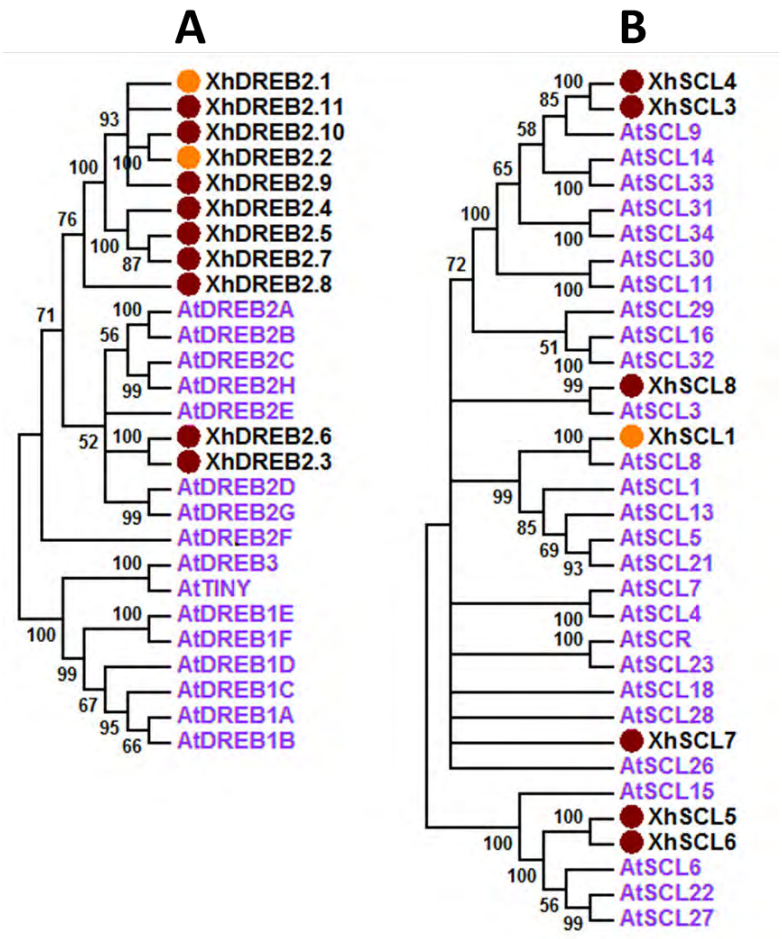
Supplementary Figure S3.6. Phylogenetic analysis of predicted *X. humilis* PYL-family ABA receptors. An unrooted Neighbour-Joining phylogenetic tree showing the evolutionary relationship between full-length predicted *X. humilis* PYL/PYR/RCAR protein sequences (black) and those of *A. thaliana* (purple). Up-regulated *X. humilis* PYL genes are marked with a green circle, and those that were down-regulated are marked with a red circle. Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.



Supplementary Figure S3.7. Phylogenetic analysis of predicted *X. humilis* PP2C proteins. An unrooted Neighbour-Joining phylogenetic tree showing the evolutionary relationship between full-length predicted *Xerophyta* PP2C protein sequences and those of *A. thaliana* (purple), as well as the ABA-interacting group A rice PP2C proteins (light blue). Up-regulated *X. humilis* PP2C genes are marked with a green circle, and those that were down-regulated are marked with a red circle. The clade that likely represents ABA-interacting PP2CA protein sequences (marked by blue squares in *A. thaliana* and rice) is indicated by the black arc. Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.



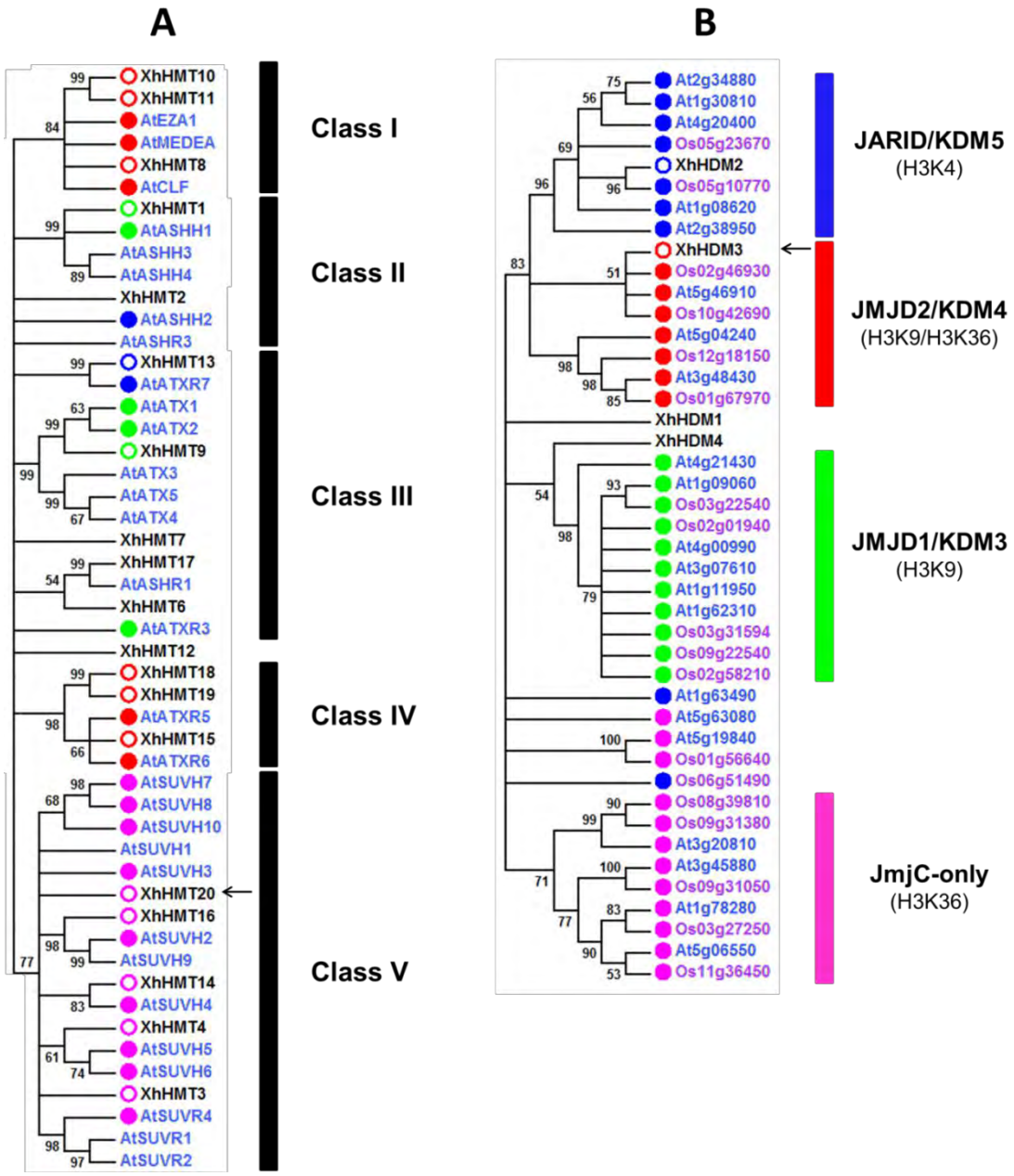
Supplementary Figure S3.8. Phylogenetic analysis of predicted *X. humilis* SnRK2 proteins. An unrooted Neighbour-Joining phylogenetic tree showing the evolutionary relationship between full-length predicted *Xerophyta* SnRK2 protein sequences and those of *A. thaliana* (purple), rice (light blue) and the known ABA-responsive barley and wheat SnRK2 proteins (PKABA1; orange). Up-regulated *X. humilis* SnRK2 genes are marked with a green circle, and those that were down-regulated are marked with a red circle. The tree clearly separates the SnRK2 protein groups as defined by the literature: Group 3 (strongly ABA-activated), Group 2 (weakly ABA activated) and group 1 (not activated by ABA). Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.



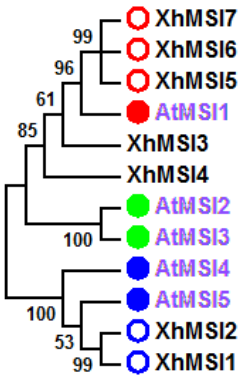
Supplementary Figure S3.9. Phylogenetic analysis of predicted *X. humilis DREB2* and *SCL* proteins. Unrooted Neighbour-Joining phylogenetic trees showing the evolutionary relationship between full-length *DREB2* (A) and *SCL* (B) protein sequences and those of *A. thaliana* (purple). Genes present in cluster U3 are marked with an orange circle, and those from cluster U4 are marked with a dark red circle. Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.

Code	Reference(s)	Code	Reference(s)
1	Baxter et al., 2007	32	Lai et al., 2008; Robatzek & Somssich, 2001
2	Bonke et al., 2003	33	Li et al., 2006
3	Brambilla et al., 2007	34	Li et al., 2011
4	Bustos et al., 2010	35	Liu et al., 2014
5	Chao et al., 1997	36	Mason et al., 2004
6	Chen & Chen, 2002	37	McAbee et al., 2006
7	Choi et al, 2000	38	Mengiste et al., 2003; Cui et al., 2013
8	Corley et al., 2005	39	Miao et al., 2010
9	Cui et al., 2013	40	Miura et al., 2007; Kanaoka et al., 2008
10	Davletova et al., 2005	41	Moon et al., 2008
11	Farinas & Mas, 2011	42	Nakano et al., 2006
12	Feurtado et al., 2011	43	Ni et al., 1999
13	Furuta et al., 2014	44	Pagnussat et al., 2005
14	Galego & Almeida, 2002	45	Peng et al, 1999
15	Gao et al., 2011	46	Proveniers et al, 2007
16	Gregis et al., 2009	47	Rashotte et al., 2006
17	Gusmaroli et al., 2001	48	Rawat et al., 2009
18	Haecker et al., 2004	49	Reyes et al., 2004
19	Hagen & Guilfoyle, 2002	50	Robatzek & Somssich, 2001
20	Heim et al., 2002	51	Sakamoto et al., 2004; Mittler et al., 2006
21	Hu and Ma, 2006	52	Sakuma et al., 2002
22	Itoh et al, 2002	53	Schellmann et al., 2002
23	Jain et al, 2006	54	Smolen et al., 2002
24	Jiang & Deyholos, 2009; Mao et al., 2011	55	Song et al., 2005
25	Jin et al., 2000	56	Szecsi et al., 2006
26	Kim et al., 2003	57	Weisshaar et al., 1991
27	Kim et al., 2006	58	Willmann et al., 2011
28	Kim et al., 2008	59	Wissenbach et al., 1993
29	Kodaira et al., 2011	60	Xie et al., 2000
30	Koyama et al., 2007	61	Zhang et al., 2007
31	Kumar et al., 2007		

Supplementary Table S3.3. Selected reference list for Table 3.7.



Supplementary Figure S3.10. Phylogenetic analysis of predicted *X. humilis* HMT and HDM proteins. Unrooted Neighbour-Joining phylogenetic trees showing the evolutionary relationship between predicted *X. humilis* histone methyltransferase (HMT) (A) and histone demethylase (HDM) (B) protein sequences and those of *A. thaliana* (blue) and rice (purple). A) Many sequences could be grouped into recognised HMT protein classes (I-V). HMT proteins with known specificity are marked by a coloured circle, and *X. humilis* HMTs that may share specificities with a closely-related *A. thaliana* orthologue are marked with an open circle of that colour (H3K4 = green, H3K9 = purple, H3K27 = red, H3K36 = blue). B) HDM protein sequences that belong to a known HDM family are marked by the coloured bars and/or circles (see labels at right). *X. humilis* HDMs that are predicted to belong to one of these families are marked with an open circle of that colour. Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.



Supplementary Figure S3.11. Phylogenetic analysis of predicted *X. humilis* MSI proteins. Unrooted Neighbour-Joining phylogenetic trees showing the evolutionary relationship between predicted *X. humilis* MSI proteins and those of *A. thaliana* (blue). Bootstrap values for each branch were calculated from 1000 replicate trees, and those with a bootstrap value less than 50 were collapsed.

Read Quality Assessment (FASTQC v0.10.1)

```
#!/bin/bash

directory=( "005A" "005B" "005C" "040A" "040B" "040C" "060A" "060B" "060C" "080A"
"080B" "080C" "100A" "100B" "100C" )

files=( "L3" "L4" "L6" "L8" )

for i in "${directory[@]}"
        do
        for j in "${files[@]}"
                do
                inputdir='../unprocessed_reads/*'$i'*/*'$j'*.fq'
                output='./'$i'_bgiclean_fastqc_output/'
                fastqc -o $output -t 8 -f fastq $inputdir
                done
        done
```

Read trimming (Trimmomatic v0.32)

```
#!/bin/sh

directory=( "005A" "005B" "005C" "040A" "040B" "040C" "060A" "060B" "060C" "080A"
"080B" "080C" "100A" "100B" "100C" )

files=( "L3" "L4" "L6" "L8" )

for i in "${directory[@]}"
       do
       mkdir  './'$i'_bgiclean_trimmed_phred20/'
       for j in "${files[@]}"
              do
              input1='../unprocessed_bgi_clean/*'$i'*/*'$j'*_1.fq'
              input2='../unprocessed_bgi_clean/*'$i'*/*'$j'*_2.fq'
              java -jar ~/software/Trimmomatic-0.32/trimmomatic-0.32.jar  PE -
threads 8 -phred33 -trimlog
'./'$i'_bgiclean_trimmed_phred20/log_'$i'_'$j'_bgiclean_trimmed_phred20.logfile'
$input1 $input2
'./'$i'_bgiclean_trimmed_phred20/'$i'_'$j'_bgiclean_phred20trim_paired_1.fastq'
'./'$i'_bgiclean_trimmed_phred20/'$i'_'$j'_bgiclean_phred20trim_unpaired_1.fastq'
'./'$i'_bgiclean_trimmed_phred20/'$i'_'$j'_bgiclean_phred20trim_paired_2.fastq'
'./'$i'_bgiclean_trimmed_phred20/'$i'_'$j'_bgiclean_phred20trim_unpaired_2.fastq'
ILLUMINACLIP:/vol1/rafe/xerophyta/rna-seq-
2014/trimmed_bgi_clean/illumina_barcodes.list:2:40:15 SLIDINGWINDOW:4:20 MINLEN:25
              done
done
```

Read error correction (SEECER v1.0.3)

```
#!/bin/sh

files=( "005A" "005B" "005C" "040A" "040B" "040C" "060A" "060B" "060C" "080A"
"080B" "080C" "100A" "100B" "100C" )

for i in "${files[@]}"
        do
              output_dir="/vol1/rafe/xerophyta/rna-seq-
2014/raw_cleaned_reads/seecer_corr_phred20/"
              pair1=$input_dir'xh_trimmed_P20_'$i'_pair1.fq'
        pair2=$input_dir'xh_trimmed_P20_'$i'_pair2.fq'
        stderrout='out_seecer_corr_phred20_'$i'.log'
        run_seecer.sh -k 25 -t $output_dir $pair1 $pair2
        done
```

Read in silico normalisation (Trinity v2.0.6)

```
#!/bin/bash

left='xh_trimmed_phred20_seecer-corrected_merged_pair1.fa'
right='xh_trimmed_phred20_seecer-corrected_merged_pair2.fa'
kmer=25
cov=50

~/software/trinityrnaseq-2.0.6/util/insilico_read_normalization.pl --seqType fa --
JM 300G --max_cov $cov --left $left --right $right --pairs_together --SS_lib_type
RF --CPU 24 --KMER_SIZE $kmer > out_normalise_reads_k25_C50_pcts200.log 2>&1
```

Merging reads (BBMerge v6.2)

```
#!/bin/bash

read1="xh_phred20_seecer_insilico_pair1.fasta"
read2="xh_phred20_seecer_insilico_pair2.fasta"
outmerged="bbmerged_singles.fq"
outpair1="bbmerged_pair1.fq"
outpair2="bbmerged_pair2.fq"

bbmerge.sh in1=$read1 in2=$read2 out=$outmerged outu1=$outpair1 outu2=$outpair2 >
out_bbmerging.log 2>&1
```

Transcriptome de novo assembly (Trinity v2.0.6)

```
#!/bin/bash

left="bbmerge_cat_pair1_and_singles.fa"
right="bbmerged_pair2.fa"
output_dir="trinity_bbmerge-postnorm_devel_default"
stderrout='out_trinity_bbmerged-postnorm_devel_default.log'

/home/rafe/software/trinityrnaseq-devel/Trinity --output $output_dir --seqType fa -
-max_memory 350G --SS_lib_type RF --left $left --right $right --CPU 8 --
bflyGCThreads 4 --bflyHeapSpaceMax 50G --bflyHeapSpaceInit 50G > $stderrout 2>&1
```

Transcriptome de novo assembly (Trinity v2014-04-12)

```
#!/bin/bash

left="bbmerge_cat_pair1_and_singles.fa"
right="bbmerged_pair2.fa"
output_dir="trinity_bbmerge-postnorm_devel_default"
stderrout='out_trinity_bbmerged-postnorm_devel_default.log'

Trinity --output $output_dir --seqType fa  --JM 350G --SS_lib_type RF --left $left
--right $right --CPU 8 --bflyGCThreads 4 > $stderrout 2>&1
```

Transcriptome de novo assembly (Bridger v2014-12-1)

```
#!/bin/bash

kmers=(21 25 31 19)
gap_len=500

left="seecer_norm_pair1.fasta"
right="seecer_norm_pair2.fasta"

for j in ${kmers[@]}
       do
       outdir=$j'mer_bridger_BAT_seecer_norm_gap'$gap_len'_v1'
```

```
	stderrout='out_bridger_BAT_seecer_norm_k'$j'_gap'$gap_len'_v1.log'
	Bridger.pl --seqType fa --left $left --right $right --CPU 8 --SS_lib_type RF
-k $j --pair_gap_length $gap_len --output $outdir > $stderrout 2>&1
	done
```

Transcriptome de novo assembly (TransABYSS v1.5.2)

```
#!/bin/bash

kmers=(25 41 61 81 19 31 35 45 51 55 65)

left="bbmerged_pair1.fa"
right="bbmerged_pair2.fa"
singles="bbmerged_singles.fa"

for j in ${kmers[@]}
	do
	out_name='transabyss_'$j'mer_bbmerge_postnorm_K25C50'
	output_dir=$j'mer_bbmerge_postnorm/'
	stderrout='out_transabyss_'$j'k_bbmerge_postnorm.log'
	transabyss --kmer $j --pe $left $right --se $left $right $singles --SS --
name $out_name --outdir $output_dir --threads 8 --length 200 > $stderrout 2>&1
	done
```

TransABYSS-merge (TransABYSS v1.5.2)

```
#!/bin/bash

assembly_dir="/vol1/rafe/xerophyta/rna-seq-
2014/final_Assemblies_2015/Abyss_assemblies/Assemblies_for_EviGene/"

fa25=$assembly_dir'abyss_1.5.2_default_k25_forEvigene.fasta'
fa31=$assembly_dir'abyss_1.5.2_default_k31_forEvigene.fasta'
fa35=$assembly_dir'abyss_1.5.2_default_k35_forEvigene.fasta'
fa41=$assembly_dir'abyss_1.5.2_default_k41_forEvigene.fasta'
fa45=$assembly_dir'abyss_1.5.2_default_k45_forEvigene.fasta'
fa51=$assembly_dir'abyss_1.5.2_default_k51_forEvigene.fasta'
fa55=$assembly_dir'abyss_1.5.2_default_k55_forEvigene.fasta'
fa61=$assembly_dir'abyss_1.5.2_default_k61_forEvigene.fasta'
fa65=$assembly_dir'abyss_1.5.2_default_k65_forEvigene.fasta'
fa71=$assembly_dir'abyss_1.5.2_default_k71_forEvigene.fasta'
fa81=$assembly_dir'abyss_1.5.2_default_k81_forEvigene.fasta'

output="transabyss_merged_k25-k81_NAT.fasta"

transabyss-merge $fa25 $fa31 $fa35 $fa41 $fa45 $fa51 $fa55 $fa61 $fa65 $fa71 $fa81
--prefixes k25_ k31_ k35_ k41_ k45_ k51_ k55_ k61_ k65_ k71_ k81_ --mink 25 --maxk
81 --SS --out $output --threads 8 --length 200
```

Evidential Gene (EVIGENE)

```
trformat.pl -input $assemblies.fa -output $assemblies.tr > out_trformat.log 2>&1
tr2aacds.pl -mrnaseq $assemblies.tr -NCPU 16 -MAXMEM 500000 -logfile >
out_tr2aacds.log
```

Read mapping (Bowtie2 v2.2.4)

```
#!/bin/bash

samples=( "005A" "005B" "005C" "040A" "040B" "040C" "060A" "060B" "060C" "080A"
"080B" "080C" "100A" "100B" "100C")

max_insert_size=500
```

```
read_type="-f"

db_index_name="combined_eviGene"

for j in ${samples[@]}
       do
       left='/vol1/rafe/xerophyta/rna-seq-
2014/raw_cleaned_reads/seecer_corr_phred20/xh_phred20_seecer-
corrected_reads/xh_trimmed_P20_'$j'_pair1.fq_corrected.fa'
       right='/vol1/rafe/xerophyta/rna-seq-
2014/raw_cleaned_reads/seecer_corr_phred20/xh_phred20_seecer-
corrected_reads/xh_trimmed_P20_'$j'_pair2.fq_corrected.fa'
       bam_file='bowtie2_combined_assembly_'$i'_reads_aligned_all_PE.bam'
       errout=$i'_evigene_align_stats.log'
       bowtie2 --fr --nofw --threads 8 --no-mixed --no-discordant --all --end-to-
end $read_type -X $max_insert_size -x $db_index_name -1 $left -2 $right 2> $errout
| samtools view -S -b -o $bam_file -
done
```

Gene cluster creation (CORSET v1.03)

```
#!/bin/bash

bam_dir="/vol1/rafe/xerophyta/rna-seq-
2014/final_Assemblies_2015/EviGene_assemblies/Corset_map_biologicals/mapping_biolog
icals/"

x005A=$bam_dir'bowtie2_combined_assembly_005A_reads_aligned_all_PE.bam'
x005B=$bam_dir'bowtie2_combined_assembly_005B_reads_aligned_all_PE.bam'
x005C=$bam_dir'bowtie2_combined_assembly_005C_reads_aligned_all_PE.bam'
x040A=$bam_dir'bowtie2_combined_assembly_040A_reads_aligned_all_PE.bam'
x040B=$bam_dir'bowtie2_combined_assembly_040B_reads_aligned_all_PE.bam'
x040C=$bam_dir'bowtie2_combined_assembly_040C_reads_aligned_all_PE.bam'
x060A=$bam_dir'bowtie2_combined_assembly_060A_reads_aligned_all_PE.bam'
x060B=$bam_dir'bowtie2_combined_assembly_060B_reads_aligned_all_PE.bam'
x060C=$bam_dir'bowtie2_combined_assembly_060C_reads_aligned_all_PE.bam'
x080A=$bam_dir'bowtie2_combined_assembly_080A_reads_aligned_all_PE.bam'
x080B=$bam_dir'bowtie2_combined_assembly_080B_reads_aligned_all_PE.bam'
x080C=$bam_dir'bowtie2_combined_assembly_080C_reads_aligned_all_PE.bam'
x100A=$bam_dir'bowtie2_combined_assembly_100A_reads_aligned_all_PE.bam'
x100B=$bam_dir'bowtie2_combined_assembly_100B_reads_aligned_all_PE.bam'
x100C=$bam_dir'bowtie2_combined_assembly_100C_reads_aligned_all_PE.bam'


corset -p corset_output -g
005,005,005,040,040,040,060,060,060,080,080,080,100,100,100 -n
005A,005B,005C,040A,040B,040C,060A,060B,060C,080A,080B,080C,100A,100B,100C "$x005A"
"$x005B" "$x005C" "$x040A" "$x040B" "$x040C" "$x060A" "$x060B" "$x060C" "$x080A"
"$x080B" "$x080C" "$x100A" "$x100B" "$x100C" > out_corset_clustering.out 2>&1
```

Differential gene expression testing (DESeq2)

```
> library(DESeq2)
> counts<-read.delim(input,header=TRUE,row.names=1)
> samples<-
data.frame(row.names=c(colnames(counts)),condition=as.factor(c(rep("RWC100",3),rep(
"RWC80",3),rep("RWC60",3),rep("RWC40",3),rep("RWC5",3))))
> cds<-DESeqDataSetFromMatrix(countData=counts,colData=samples,design=~condition)
> source("plotPCAWithSampleNames.R")
> plotPCAWithSampleNames(cds)
> rcds=rlog(cds)
> decds<-DESeq(cds,test="LRT",full=~condition,reduced=~1)
> res=results(decds,alpha=0.01)
> resSig=res[which(res$padj < 0.01),]
```

Supplementary Figure S3.12. Example scripts to run the programs used to generate the transcriptomic data in this study. An overview of the scripts and commands used during various steps of the transcriptome assembly pipeline (see Figure 3.2).

Chapter 4: Conclusion and Future Work

4.1 Conclusion

The broad aim of the research presented in this thesis was to investigate the hypothesis that vegetative desiccation tolerance (VDT) in angiosperm resurrection plants was derived from the mechanisms that control desiccation tolerance (DT) in mature orthodox seeds. Particular emphasis was given to the theory that angiosperm VDT could be related to the brief window seen during angiosperm seed germination within which seedlings are able to re-induce embryonic traits, including DT, in response to imposed stress. These questions were approached in two ways. A seed germination system was used to compare the ability of the seeds and seedlings of the desiccation sensitive (DS) model plant *Arabidopsis thaliana* and the resurrection plants *Xerophyta humilis* and *Xerophyta viscosa* to re-induce DT in response to severe osmotic stress. The point of no return (the developmental stage at which germinating seedlings lose the ability to re-establish DT) of these three species was investigated by priming seeds or seedlings with PEG, sucrose or slow drying. The sequencing, *de novo* assembly and expression analysis of the desiccation transcriptome of *X. humilis* leaves at varying stages of water loss was characterized to identify the transcription factors networks and their downstream target genes active during the desiccation programme. Expression analysis was focused on the expression of LEA genes, transcription factors (TFs) and regulators of epigenetic gene regulation, in order to test the hypothesis that VDT results from the de-repression of seed maturation genes in *X. humilis* due to relaxation of PRC2-mediated H3K27me3 gene silencing of various seed master regulatory TFs.

Analysis of the germination and window of DT in *Xerophyta* seedlings has not been reported previously. An experimental protocol was developed to efficiently track seedling growth and development under normal or stress conditions, and various optimisations were also attempted to improved germination synchronicity. The results obtained show strong support for a link between seedling DT and VDT in *Xerophyta*. Seedlings of *X. humilis* and *X. viscosa* exhibited only a transient decline in DT during early seedling development, although DT was never lost completely, and which could be rescued by PEG, sucrose or slow drying at all stages. In contrast, *A. thaliana* swiftly lost the ability to induce DT in seedlings during germination under standard growth conditions, and priming with PEG or sucrose could improve survival rate during only early development but not at later stages. PEG priming has been shown to improve DT in various DS plant species during germination, thus it is not surprising that it had a similar effect on *Xerophyta* seedlings. Improvement of seedling DT by the addition of exogenous sucrose has not so far been reported; however, the experiments performed here were insufficient to determine how sucrose has this effect. Two possibilities include the activation of a sugar- or sucrose-mediated signalling pathway that results in activation of VDT-related stress response pathways (e.g. via ABA), or that the extra carbon source provision due to high levels of exogenous or endogenous sucrose improves the seedling response to desiccation. Further analysis is required to unravel the mechanism by which applied sucrose improves DT in germinating seedlings.

Interestingly, seedlings of both *X. humilis* and *X. viscosa*, which are poikilochlorophyllous resurrection plants, exhibited the ability to induce chlorophyll degradation from the earliest stages at which chlorophyll was present in seedling tissues. This suggests that the mechanisms that control poikilochlorophylly are indelibly tied to the desiccation response

during both seedling and adult developmental stages. There is also evidence, however, that the acquisition of DT occurs much more rapidly than those that control poikilochlorophylly in seedlings, and chlorophyll loss is likely uncoupled from the DT acquisition pathways. Many seedlings could still survive desiccation despite retaining chlorophyll, while some seedlings that lost chlorophyll nonetheless failed to survive rehydration. Although the seedling data presented in this chapter supports a link between seedling and adult DT in *Xerophyta*, the genetic relationship between seedling/adult DT and embryonic DT regulatory mechanisms remains to be elucidated.

As a first step in understanding the molecular mechanisms underpinning adult VDT, the transcriptome of *X. humilis* was assembled from leaves across five RWC (100%, 80%, 60%, 40% and 5%), resulting in the identification of 76,868 gene clusters – transcripts derived from a single genomic locus that likely represent a single gene. Contamination of *X. humilis* leaf samples by a fungus or fungi was observed – accounting for 17% of all identified genes, although only 1% of sequencing reads – but was determined to not negatively impact the sequencing or expression results. Approximately 80% of the transcriptome was successfully annotated against known proteins. Gene expression analysis at varying stages of RWC revealed that 18,534 genes were differentially expressed during desiccation at a fold-change > 2 and FPKM > 1 (24% of all transcripts). Gene expression, with the exception of the 60% RWC sample, positively corresponded with the gene expression observed in a previous microarray study (Shen, 2014). These differences may be due to variation in the measured RWC of the leaves used between the two studies. We then analysed the expression of genes known or suspected to be involved in VDT in *X. humilis* and other resurrection plant species.

LEA and LEA-like proteins are important for desiccation in most known DT organisms and tissues. Consequently, many individual LEA proteins have been identified as being differentially expressed during desiccation in VDT plants. In this study, nearly 150 gene clusters were annotated as coding for *X. humilis* LEA proteins, and most of the *X. humilis* LEA gene families were massively expanded compared to those of other DS species. This included the LEA-6 family of seed-specific LEAs, but also those from LEA-4, LEA-8 and, to a lesser extent, LEA-1 and LEA-10 families. LEA gene expression during desiccation was induced at two stages: transiently at 80% RWC, and again at 40% and 5% RWC. Additionally, the simultaneous induction of nearly all identified LEA proteins in *X. humilis*, including many usually involved in seed development (e.g. LEA-1, LEA-6 families), is particularly notable. In *A. thaliana*, there is little overlap between LEA genes highly expressed in seeds (or stressed seedlings) and vegetative tissues, suggesting that they are controlled by different pathways. This appears to be different in *X. humilis* and has important implications regarding the evolution of VDT in this species, as it implies that either both signalling pathways (vegetative drought response or seed maturation) are active simultaneously during desiccation of adult tissues, or that one pathway has evolved the ability of activating both stress-specific and seed-specific LEAs in vegetative tissues.

A comprehensive examination of TFs differentially expressed during desiccation was undertaken in order to identify classes of transcriptional regulators that may be important for VDT. TFs from nearly all of the major plant TF families were identified, consistent with broad changes in gene regulation and expression. Dehydration in *X. humilis* was associated with consecutive waves of TF induction, suggesting that VDT is regulated by a transcriptional cascade that coordinates gene expression as water is lost from the cell.

Unexpectedly, the analysis of the TFs induced during desiccation in *X. humilis* found little to support the hypothesis of seed maturation-derived transcriptional regulation of VDT. Only genes with similarity to two predicted seed maturation master regulators were expressed in *X. humilis* during desiccation, *ABI3* and *LEC1*. However, the *LEC1* transcript was expressed at a very low level and was not differentially expressed during desiccation; two alternatively-spliced *XhABI3* transcripts were only slightly but significantly up-regulated as desiccation progressed, although their expression was also low. Furthermore, the predicted proteins of the *XhABI3* genes lacked the B3 domain which binds to the seed-specific RY motif, suggesting that they may function cooperatively with other TFs rather than interact with seed gene CREs. Although GO terms related to seed development, particularly maintenance of seed dormancy and germination, were enriched in a number of TF clusters, closer investigation revealed that nearly all of these TFs functioned during biotic or abiotic cellular stress response during both seed development and vegetative growth, and virtually none were seed-specific.

The initial hypothesis outlined at the start of this study predicted that de-repression of seed maturation genes would involve down-regulation of PRC2 component genes, particularly *SWINGER*, resulting in in a switch from silencing H3K27 methylation to activating epigenetic marks in VDT gene promoters. Although this study could confirm that virtually all PRC2 component genes were downregulated during desiccation, the expression profile of these genes was mirrored by the majority of genes encoding histones and histone modification enzymes, irrespective of target histone or residue. This change is likely indicative of a global reduction in histone modification during desiccation, possibly due to a massive reduction in genome replication and increase in heterochromatin formation and

chromatin compaction as water is lost. Additionally, although many seed genes were up-regulated during early dehydration (80% RWC), PRC2 subunit gene down-regulation only occurred after 60% RWC. This study has no data linking PRC2 gene levels with protein levels or PRC2 complex activity, however, and a general reduction in H3K27me3 during late desiccation is at least consistent with the greater up-regulation multiple classes of genes at this stage. Additional analyses, such as ChIP analysis to identify epigenetic modifications significantly enriched or depleted in desiccation-related genes, are required to test the role of PRC2 and other epigenetic complexes in regulating VDT.

Recently it has become clear that the ABA-signalling pathway present in land plants is highly conserved, particularly amongst higher plants. Common elements in this pathway include cytosolic ABA receptors of the PYL family, group-A PP2C phosphatases and group-2 SnRK protein kinases which activate downstream ABRE-binding bZIP TFs (ABFs) in conjunction with other proteins, ultimately resulting in ABA-induced gene expression. Early seedless vascular plants, for example ferns, likely inherited the components of this pathway from ancestral non-vascular plants (such as mosses), where it controlled VDT in concert with an ancestral variant of *ABI3*. Many of these component genes have since diversified in seed-bearing plants, resulting in two distinct variants of the original ABA-signalling pathway: one that controls seed maturation via ABI3 and the bZIP protein ABI5 and its homologs, and a separate pathway that controls ABA responsiveness during vegetative stress via a related set of bZIP transcription factors (such as ABF1-4 in *A. thaliana*) in an ABI3 independent manner.

Many genes associated with ABA metabolism, particularly biosynthesis, and genes involved in the ABA signal transduction pathway were significantly up-regulated in *X. humilis* during

desiccation, including multiple PYL genes, group-A PP2Cs, SnRK2s and bZIP transcription factors. Additionally, a group of *AFP* genes, a recently-discovered class of proteins that bind ABFs and target them for degradation, were likewise up-regulated (Garcia et al., 2008). This is thus strong evidence for a central role for ABA signalling during VDT in *X. humilis*. However, the *X. humilis* bZIP TFs up-regulated during desiccation encoded proteins most similar to the ABFs that control the response to vegetative drought stress in *A. thaliana* rather than those from the seed maturation pathway. This finding was unexpected based on the hypothetical model of VDT induction proposed at the start of this study (see Chapter 1, Figure 1.7). Coupled with the observed absence of seed-maturation master regulatory gene transcripts – with the exception of a lowly-expressed, truncated version of *ABI3* – it instead appears that VDT may be regulated by vegetative stress response factors, with *XhABF1* being the most promising candidate in relation to ABA signalling and initiating the response to water loss.

An alternative model of VDT induction in *Xerophyta* seedlings and plants is outlined in Figures 4.1 and 4.2, which highlights key regulators and their downstream targets identified in this study and provides testable hypotheses that could guide future work.

In many angiosperms, DT can be re-established in germinating seedlings during a short developmental window. This process is controlled by *ABI3* and *ABI5*, which are not present in unstressed seedlings, and requires ABI5 protein activation by SnRK2-mediated phosphorylation downstream of the ABA signalling pathway (Lopez-Molina et al., 2002; Maia, 2014). Activation of ABI5 results in growth arrest, expression of seed-specific LEAs, seed maturation proteins and antioxidant systems and, ultimately, desiccation tolerance (Fig. 4.1A, right). The existence of a similar *ABI3*/*ABI5*-mediated pathway specific to the

seedlings of *X. humilis* is not inconsistent with the results of this study (Fig. 4.1B, "window of seedling DT"), but evidence for such a pathway is outside the scope of thesis. However, it is clear that in vegetative tissues VDT is not controlled by *ABI3*/*ABI5* orthologues – which are absent or expressed at low levels in desiccating leaves – but is instead activated by a separate pathway mediated by vegetative stress response factors, such as the *XhABF1* gene identified in this study (Fig. 4.1A, left). As the seedling develops, either one or both of these pathways are activated in response to drought stress, resulting in the re-establishment of DT (Fig. 4.1B). Evidence for the early activation of a vegetative VDT pathway in seedlings is given by the observation that *Xerophyta* seedlings of all developmental stages are technically capable of inducing poikilochlorophylly, a trait not usually associated with seedling DT re-establishment in DS angiosperm species – but one that is found in adult *X. humilis* and *X. viscosa* plants. By the time the seedling has committed to vegetative growth, however, the vegetative pathway is sufficient to induce VDT (Fig. 4.1B, bottom). In this way, DT can be re-established in seedlings even after any putative *ABI3*/*ABI5*-mediated pathway has been permanently silenced (the "point of no return" in DS species, Fig. 4.1B), explaining why such a phenomenon was not observed in *Xerophyta* seedlings. The transient decline in survival rate of rapidly dried *X. humilis* and *X. viscosa* seedlings during development, and the presence of poikilochlorophylly only in *X. viscosa* seedlings of the developmental stage after this decline, could be associated with a switch from a seedling-specific *ABI3*/*ABI5*-mediated DT pathway to a vegetative stress-associated VDT pathway. However, there is currently insufficient evidence to say if this is the case. Multiple classes of seed-specific genes (many LEAs, SSPs, oleosins, caleosins; Fig. 4.1B) are activated in desiccating leaves despite the absence of seed maturation TFs/ABFs. Assuming expression of these genes is controlled via ABA and the ABRE, as during seed development, this suggests that the vegetative ABF(s)

identified in this study are capable of activating seed genes in adult tissues, either alone or in a complex with other proteins, in addition to activating classically drought-responsive genes (Fig. 4.1A). For example, even though the *ABI3* transcripts expressed in desiccating *X. humilis* would produce a protein that lacked the seed RY motif-binding B3 domain, a putative *ABI3*/*ABF* complex could still target and activate gene expression at ABREs present in seed gene promoters, which would not require the B3 domain (Bensmihen, 2002; Nakamura et al., 2001). With this interpretation, VDT thus represents the synergistic activation of both the abiotic stress response and a subset of seed maturation genes (Fig. 4.1A). The simultaneous up-regulation of nearly all identified LEA protein genes, consisting of both seed-specific and abiotic stress-related LEA families, is evidence of the shared expression pattern of genes from these two pathways.

The proposed model is thus consistent with both the physiological seedling results discussed in Chapter 2 and the gene expression analysis performed in Chapter 3. However, it should be noted that this model assumes that the separation of seed maturation and vegetative stress ABA-signalling pathways that has been found in other angiosperms (for example, *AtABI5*-related bZIPs and *AtABF1-4* during seed development and abiotic stress, respectively, in *A. thaliana*) is conserved in *X. humilis*, and that the clustering of the identified *X. humilis* ABFs (*XhABF1-3*) with the vegetative clade of *A. thaliana* ABFs implies that they are associated with abiotic stress in *Xerophyta*. However, because only adult leaves were sampled in this study, there is no direct evidence that this is the case. An alternative hypothesis is that the gene networks that control seed development and maturation in *Xerophyta* have diverged from those found in DS plants, and that *XhABF1-3* are in fact functionally equivalent to the *ABI5*-related seed maturation-specific ABFs that control seed

maturation in *A. thaliana* and other plants. If this is the case, the conclusion presented herein that VDT is not activated by a seed maturation related gene network would be erroneous. No study so far has investigated the gene networks that control seed development and maturation in *X. humilis* or any other resurrection plant; such an investigation would be a good first step in verifying or disproving the model described above.

A general overview of the proposed mechanisms and processes that occur as water is lost from desiccating *X. humilis* leaves is given in Figure 4.2, based on the gene expression analysis performed in this study.

X. humilis orthologues of the ABA biosynthetic genes *ABA4* and *NCED* are first up-regulated at 80% RWC, suggesting an increase in ABA levels at or after this stage – consistent with a previous investigation into ABA levels of desiccating *X. humilis* leaves (Fig. 4.2.1; Beckett 2011). The initial response to drought-induced ABA signalling occurs via at least three ABF proteins, *XhABF1-3*, which are activated by group II or III SnRK2 kinases. Although *XhABF1* is active at both early (80% RWC) and late (40% and 5% RWC) stages of the desiccation response, *XhABF2* and *XhABF3* are specifically induced during this early stage and so may be involved in general stress response (e.g. growth arrest, repression of GA signalling) as opposed to desiccation (Fig. 4.2.6). Various TFs are up-regulated at this stage, including a DELLA protein gene, an orthologue of the DREB-like TF gene *TINY*, and a subset of HSF genes, which may all play a role in the initial drought response (Fig. 4.2.7) The first wave of seed gene expression also occurs at this stage, including all LEAs, oleosins, caleosins and *PER1* – but not globulin SSPs (Fig. 4.2.7). As *XhABF1* is the only ABF with an expression profile that matches the identified seed genes, it is a likely candidate as central coordinator of this response. By 60% RWC photosynthesis has ceased and chloroplast-xeroplast

transformation has likely occurred or is occurring (Fig. 4.2.3; Beckett et al. 2012). The expression of histone and chromatin-remodelling gene transcripts is severely curtailed after 60% RWC, possibly associated with H3K9-mediated chromatin compaction and heterochromatin formation which serves to protect the genome while desiccated (Fig. 4.2.7). Certain histone variant transcripts are specifically up-regulated during late desiccation, and may be associated with structural or epigenetic changes to the chromatin state that are important for VDT (Fig. 4.2.6). Gene expression does not cease, however, possibly because of maintenance of actively-transcribed euchromatin regions by the action of the H3K9 demethylase, *XhHDM3*, and several CPB and MYST family histone acetyltransferases (4.2.4). A very large number of genes were specifically up-regulated after 60% RWC, and the expression of abiotic stress response genes and TFs (LEAs/seed genes/SSPs, most HSFs, *DREB2*, *ZAT/AZF*, *SCL* TFs) increases dramatically (Fig. 4.2.7). ABA levels have been found to be low in *X. humilis* leaves after 40% RWC compared to earlier RWCs (Fig. 4.2.1), however ABA sensitivity at these stages might also be heightened due to the very high induction of PYL-family ABA receptors and downstream components of the ABA signalling pathway, such as group 2 (but not group 3) SnRK2 kinases (Fig. 4.2.5). Late desiccation is probably associated with the accumulation of protective molecules (secondary metabolites, LEAs, other proteins), storage reserves (SSPs, lipids) and factors necessary for rehydration (Fig. 4.2.7). Sucrose accumulates throughout the final stages of desiccation, and is probably vital for the formation of a sugar/LEA cellular glass (Fig. 4.2.1; McDonald 2008). The late desiccation response is likely controlled by ABA (via *XhABF1*) and/or the TFs mentioned above, some of which may be directly regulated by *XhABF1* (Fig. 4.2.6). This could indicate that *X. humilis* leaves are not completely desiccation tolerant until sometime after 60% RWC,

although VDT competence/longevity may increase over time between 60% and 5% RWC rather than requiring a critical concentration of these gene products (Fig. 4.2.8).

In conclusion, the data presented herein disputes the hypothesis that desiccation tolerance in vegetative tissue evolved by the activation of the seed maturation regulatory network in leaves. Although seed-specific factors, such as many LEAs, SSPs, oleosins and antioxidants were strongly induced, none of the upstream TFs that specifically activate their expression in seeds were up-regulated in *X. humilis* during desiccation. Additionally, the simultaneous pattern of LEA gene up-regulation is consistent with the induction of genes from both the seed maturation and vegetative stress networks, rather than just one of these pathways. An alternative model is proposed, whereby many of the ABA-responsive LEAs and seed genes are activated by the vegetative stress response regulatory network.

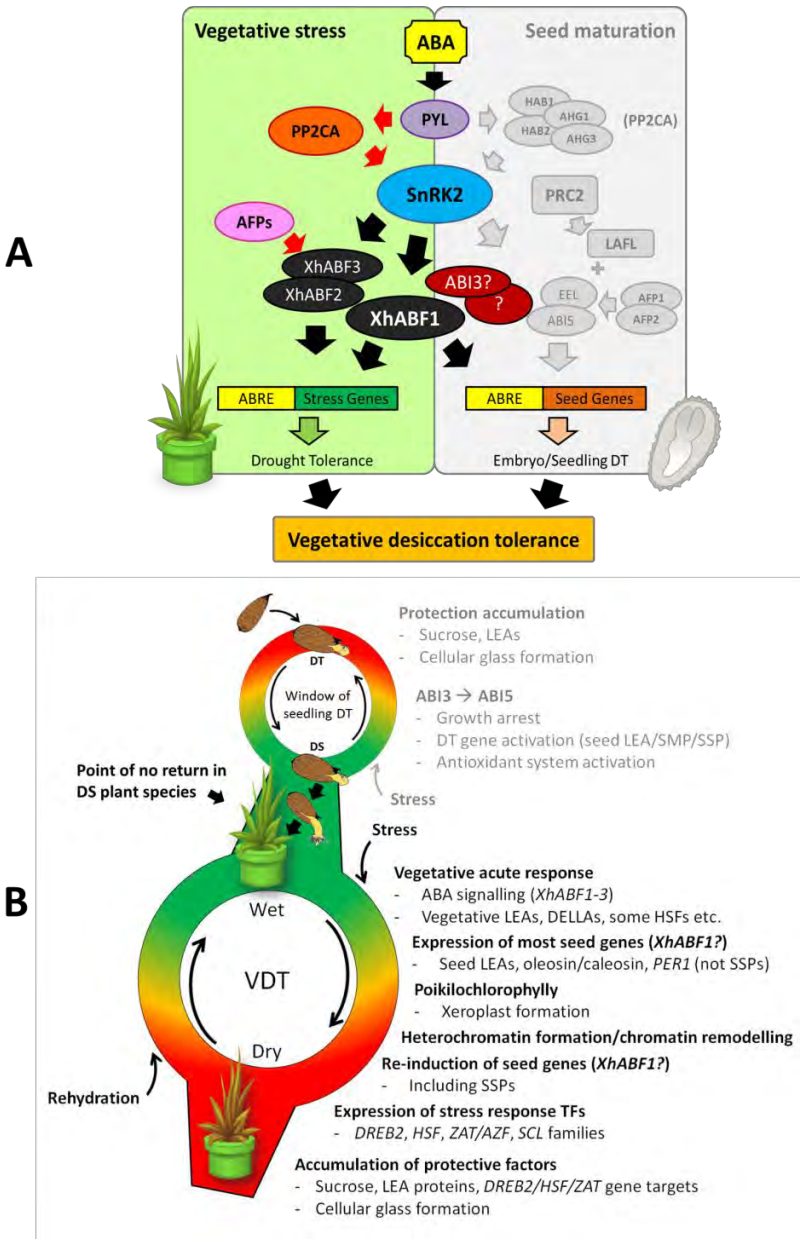


Figure 4.1. Proposed model of the processes governing seed gene expression and VDT acquisition during desiccation in *X. humilis*. A) Only three ABF genes were up-regulated during desiccation in *X. humilis*, all related to those involved in vegetative stress response in *A. thaliana*. Only *XhABF1* was expressed throughout desiccation, including all stages at which seed maturation genes were induced. The *XhABF1* protein may be capable of activating expression of ABRE-containing seed-specific genes, either alone or in cooperation with other TFs (e.g. *ABI3*) similar to what occurs during seed maturation. This would allow activation of a subset of ABA-responsive seed genes required for DT by the vegetative stress response network. B) Proposed model of VDT induction in *Xerophyta* seedlings and adult plants. In DS angiosperms, seedlings can re-establish DT in response to stress during a small window controlled by ABI3 and ABI5 ("window of seedling DT"). A similar pathway may exist in *X. humilis*, although evidence for such a pathway is out of the scope of this thesis. However, adult plants likely induce VDT via a vegetative stress response pathway, as described in the first part of this figure. This pathway may be active in *Xerophyta* seedlings as well, resulting in poikilochlorophylly and the lack of a point of no return in seedlings of all developmental stages from these species.

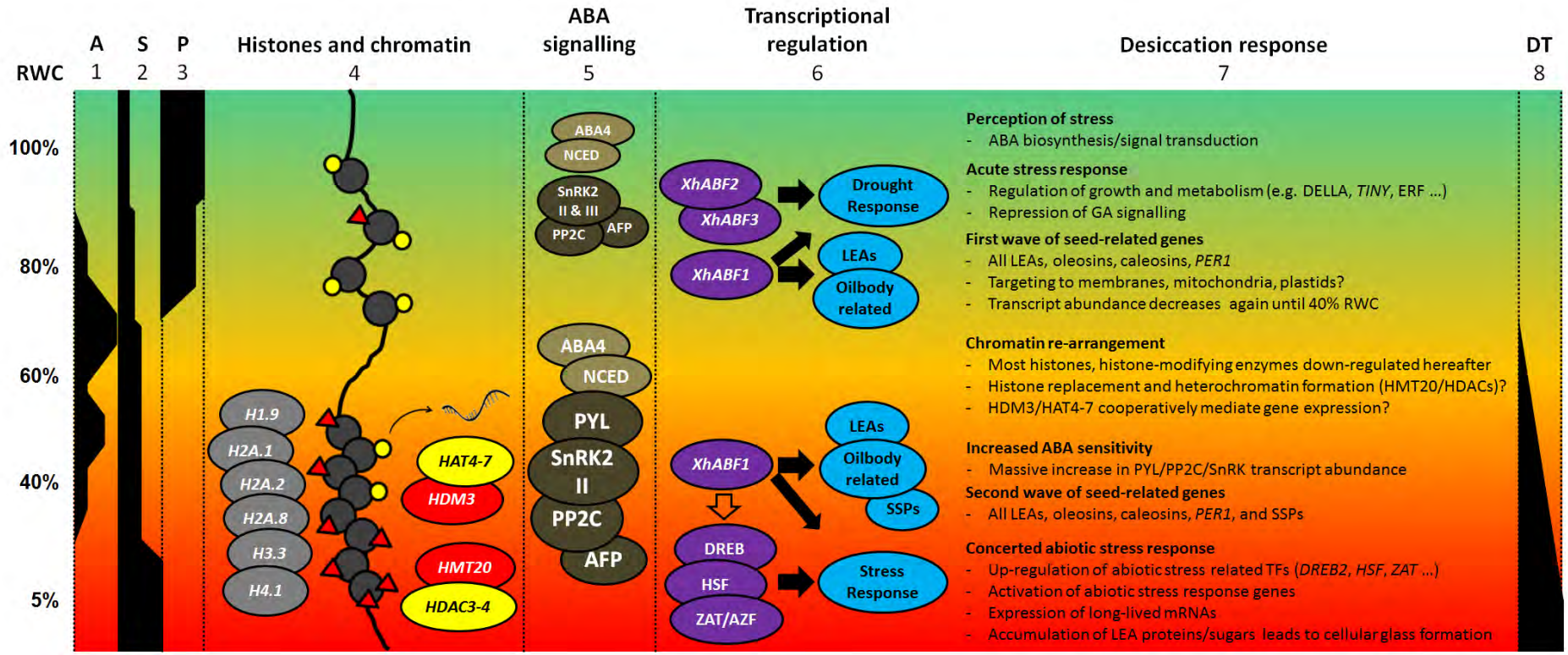


Figure 4.2. Overview of the predicted desiccation process in *X. humilis*. Various changes are induced by dehydration as water is lost from *X. humilis* tissues. 1) ABA levels start to increase shortly before 80% RWC, and peak just before 60% RWC (Beckett, 2011). 2) Sucrose concentration in the leaves increases as they desiccate, reaching the highest levels in dry tissue (MacDonald, 2008). 3) Photosynthetic rate rapidly decreases between 100% and 60% RWC (Beckett et al., 2013). 4) Desiccation is most likely associated with large-scale changes in chromatin structure, including possibly heterochromatin formation, after 60% RWC. This is likely controlled by histone methylation (red triangles), particularly of H3K9, and histone acetylation (yellow circles). Important genes in this process include a handful of histone variants, H3K9 methylases/demethylases (*XhHMT20/XhHDM3*) and histone acetylases/deactylases (*XhHAT4-7/XhHDAC3&4)*. 5) ABA is important throughout desiccation, although the abundance of signalling genes, especially PYLs, increases dramatically after 60% RWC . 6) ABA induces expression of ABF bZIP transcription factors, which are involved in general stress response (*XhABF2&3*) and possibly the expression of desiccation-associated genes, including seed genes (*XhABF1*). These genes may also regulate the expression of downstream regulators, such as *DREB2* and *HSF* transcription factors. 7) An overview of the changes that occur across each. 8) Due to the specific expression of many stress-response genes only after 60% RWC, it seems likely that *X. humilis* leaves are not fully desiccation tolerant until after 60% RWC. The probability of successful dehydration/rehydration, or the length of time for which a leaf can remain desiccated, could improve depending on how quickly the tissue reaches an air-dry state.

4.2 Future work

The results presented in this thesis, despite providing an internally consistent model of VDT induction in the resurrection plant *X. humilis*, are based only on an RNA-Seq transcriptome analysis of a single tissue type – leaf – and thus require further work to test and confirm the models proposed here.

Firstly, evidence of the *in vivo* existence of the transcripts highlighted within this work, as well as independent verification of their expression profiles, is necessary to confirm these results. For gene transcripts for which their assembled sequence deviates from the orthologous sequences in other species – for example, the lack of a region encoding a B3 domain in the identified *X. humilis* ABI3 transcripts – it may be necessary to perform 3'/5' RACE (Rapid Amplification of cDNA Ends) on RNA samples form the corresponding tissues to verify the assembled sequences (Frohman, 1993). Functional characterisation of the described genes and transcripts is also warranted, as it cannot be taken for granted that orthologous genes across two distantly-related species necessarily perform the same function. As a non-model organism, transgenic studies involving knock-out or over-expressing mutant lines of *X. humilis* are likely not possibly in the near future; however, complementation studies using *X. humilis* genes in a model plant, such as *A. thaliana*, could provide important information about their function. Studies utilising RNA interference, or the recently-developed CRISPR technologies, could also potentially be used to great effect in this regard.

Because *Xerophyta* seedlings exhibit a desiccation response remarkably similar to adult plants, a seedling model could be viably used to study VDT in this species in a more

controlled manner than vegetative tissue in many instances. The visible cue of chlorophyll loss could be leveraged to assess seedlings that have successfully or unsuccessfully acquired DT through pre-treatment, and further tease apart the genetic networks that are related specifically to VDT as opposed to general stress response. A much neglected aspect of the seed-derived model of VDT in resurrection plants is an in-depth investigation into seed maturation in these species, and precise identification of the important TFs that control this process. For example, as a preliminary experiment, it would be interesting to assess the expression of the *X.* humilis *ABI3* and *ABF* genes identified in this study in dry seeds and during seedling development. If, as in DS plants, levels of orthologues of *XhABI3* and *XhABI5* are high in seeds and during germination, but drop as development progresses – despite the retention of DT in *Xerophyta* seedlings – it would provide clear evidence for a regulatory disconnect between the angiosperm window of seedling DT and resurrection plant VDT. Alternatively, if the desiccation-induced *XhABI3* isoforms and *ABF* genes identified in this study are highly expressed in seeds, it would suggest that they are in fact associated with the seed maturation regulatory network rather than with a vegetative response.

The rise in prominence and decrease in cost of Next Generation Sequencing (NGS) technologies make them ideal tools to explore the phenomenon of VDT in resurrection plants. Although this study made use of RNA-Seq to investigate the desiccation transcriptome of adult leaves, it could similarly be used to examine the transcriptomes of rehydrating leaves, or *X. humilis* seedlings subjected to stress. A draft genome of *X. humilis* is currently under assembly, which should massively expand the available genomic sequence information of this plant. A complete draft of the *X. humilis* genome could be leveraged to

investigate the mechanism of the apparent LEA gene diversification in *Xerophyta* – for example, tandem duplication – or enable the use of bioinformatics approaches to statistically analyse cis-regulatory elements and TF binding within the promoters of selected desiccation-induced genes, which could reveal important patterns of regulatory motifs relevant to VDT. For example, a prediction of the model proposed in the previous section is that the ABRE motif should be enriched in the promoters of many genes (especially seed-specific genes) that are up-regulated during desiccation. Related to this, it would be informative to determine whether the ABF gene proteins identified in this study, particularly *XhABF1*, can bind to these promoter regions *in vivo*; for example, via yeast one-hybrid analysis or ChIP-Seq using antibodies raised to these proteins in desiccating leaves.

Lastly, the activation of VDT-related mechanisms during desiccation is likely associated with changes in the epigenetic landscape of the *X. humilis* genome, affecting both gene expression and chromatin structure. ChIP/ChiP-Seq or similar approaches could be used to characterise the epigenetic state of the genome during desiccation, and identify histone modifications enriched in down- or up-regulated gene sets that may be important in regulating VDT.

References

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., Shinozaki, K., 1997. Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell 9, 1859–68. doi:10.1105/tpc.9.10.1859

Achard, P., Cheng, H., De Grauwe, L., Decat, J., Schoutteten, H., Moritz, T., Van Der Straeten, D., Peng, J., Harberd, N.P., 2006. Integration of plant responses to environmentally activated phytohormonal signals. Science 311, 91–4. doi:10.1126/science.1118642

Achard, P., Renou, J.-P., Berthomé, R., Harberd, N.P., Genschik, P., 2008. Plant DELLAs restrain growth and promote survival of adversity by reducing the levels of reactive oxygen species. Curr. Biol. 18, 656–60. doi:10.1016/j.cub.2008.04.034

Ahmad, P., Jaleel, C.A., Salem, M.A., Nabi, G., Sharma, S., 2010. Roles of enzymatic and nonenzymatic antioxidants in plants during abiotic stress. Crit. Rev. Biotechnol.

Aichinger, E., Villar, C.B.R., Farrona, S., Reyes, J.C., Hennig, L., Köhler, C., 2009. CHD3 proteins and polycomb group proteins antagonistically determine cell identity in Arabidopsis. PLoS Genet. 5, e1000605. doi:10.1371/journal.pgen.1000605

Alamillo, J., Almoguera, C., Bartels, D., Jordano, J., 1995. Constitutive expression of small heat shock proteins in vegetative tissues of the resurrection plant Craterostigma plantagineum. Plant Mol. Biol. 29, 1093–1099. doi:10.1007/BF00014981

Almoguera, C., Rojas, A., Díaz-Martín, J., Prieto-Dapena, P., Carranco, R., Jordano, J., 2002. A seed-specific heat-shock transcription factor involved in developmental regulation during embryogenesis in sunflower. J. Biol. Chem. 277, 43866–72. doi:10.1074/jbc.M207330200

Alonso, R., Oñate-Sánchez, L., Weltmeier, F., Ehlert, A., Diaz, I., Dietrich, K., Vicente-Carbajosa, J., Dröge-Laser, W., 2009. A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. Plant Cell 21, 1747–61. doi:10.1105/tpc.108.062968

Alpert, P., 2006. Constraints of tolerance : why are desiccation-tolerant organisms so small or rare ? 1575–1584. doi:10.1242/jeb.02179

Alpert, P., 2005. The limits and frontiers of desiccation-tolerant life. Integr. Comp. Biol. 45, 685–95. doi:10.1093/icb/45.5.685

Angelovici, R., Galili, G., Fernie, A.R., Fait, A., 2010. Seed desiccation: a bridge between maturation and germination. Trends Plant Sci. 15, 211–8. doi:10.1016/j.tplants.2010.01.003

Aoyama, T., Dong, C.H., Wu, Y., Carabelli, M., Sessa, G., Ruberti, I., Morelli, G., Chua, N.H., 1995. Ectopic expression of the Arabidopsis transcriptional activator Athb-1 alters leaf cell fate in tobacco. Plant Cell 7, 1773–85.

Apel, K., Hirt, H., 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu. Rev. Plant Biol. 55, 373–99. doi:10.1146/annurev.arplant.55.031903.141701

Audran, Borel, Frey, Sotta, Meyer, Simonneau, Marion-Poll, 1998. Expression studies of the zeaxanthin epoxidase gene in nicotiana plumbaginifolia. Plant Physiol. 118, 1021–8.

Bannister, A.J., Kouzarides, T., 2011. Regulation of chromatin by histone modifications. Cell Res. 21, 381–95. doi:10.1038/cr.2011.22

Banti, V., Mafessoni, F., Loreti, E., Alpi, A., Perata, P., 2010. The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in Arabidopsis. Plant Physiol. 152, 1471–83. doi:10.1104/pp.109.149815

Barbosa, E.G.G., Leite, J.P., Marin, S.R.R., Marinho, J.P., de Fátima Corrêa Carvalho, J., Fuganti-Pagliarini, R., Farias, J.R.B., Neumaier, N., Marcelino-Guimarães, F.C., de Oliveira, M.C.N., Yamaguchi-Shinozaki, K., Nakashima, K., Maruyama, K., Kanamori, N., Fujita, Y., Yoshida, T., Nepomuceno, A.L., 2012. Overexpression of the ABA-Dependent AREB1 Transcription Factor from Arabidopsis thaliana Improves Soybean Tolerance to Water Deficit. Plant Mol. Biol. Report. 31, 719–730. doi:10.1007/s11105-012-0541-4

Baron, K.N., Schroeder, D.F., Stasolla, C., 2012. Transcriptional response of abscisic acid (ABA) metabolism and transport to cold and heat stress applied at the reproductive stage of development in Arabidopsis thaliana. Plant Sci. 188-189, 48–59. doi:10.1016/j.plantsci.2012.03.001

Bartels, D., Hussain, S.S., 2011. Resurrection plants: physiology and molecular biology, in: Plant Desiccation Tolerance. Springer, pp. 339–364.

Bartels, D., Phillips, J., Chandler, J., 2007. Desiccation tolerance: gene expression, pathways, and regulation of gene expression. Plant Desiccation Toler. 35, 115–137.

Bartels, D., Sunkar, R., 2005. Drought and Salt Tolerance in Plants. CRC. Crit. Rev. Plant Sci. 24, 23–58. doi:10.1080/07352680590910410

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., Covarrubias, A.A., 2008. The Enigmatic LEA Proteins and Other Hydrophilins. PLANT Physiol. 148, 6–24. doi:10.1104/pp.108.120725

Baumlein, H., Nagy, I., Villarroel, R., Inze, D., Wobus, U., 1992. Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. Plant J. 2, 233–239. doi:10.1046/j.1365-313X.1992.t01-45-00999.x

Becker, M.G., Hsu, S.-W., Harada, J.J., Belmonte, M.F., 2014. Genomic dissection of the seed. Front. Plant Sci. 5, 464. doi:10.3389/fpls.2014.00464

Beckett, M., 2011. Investigation into the role of volatile organic compounds, and abscisic acid in stomatal regulation, in the resurrection plant Xerophyta humilis. University of Cape Town.

Beckett, M., Loreto, F., Velikova, V., Brunetti, C., Di Ferdinando, M., Tattini, M., Calfapietra, C., Farrant, J.M., 2012. Photosynthetic limitations and volatile and non-volatile isoprenoids in the poikilochlorophyllous resurrection plant Xerophyta humilis during dehydration and rehydration. Plant. Cell Environ. 35, 2061–74. doi:10.1111/j.1365-3040.2012.02536.x

Beckett, R.P., 1999. Partial dehydration and ABA induce tolerance to desiccation-induced ion

leakage in the moss Atrichum androgynum. South African J. Bot. 65, 212–217.

Behnke, H.D., Hummel, E., Hillmer, S., Sauer-Gürth, H., Gonzalez, J., Wink, M., 2013. A revision of African Velloziaceae based on leaf anatomy characters and rbcL nucleotide sequences. Bot. J. Linn. Soc. 172, 22–94. doi:10.1111/boj.12018

Bell, O., Tiwari, V.K., Thomä, N.H., Schübeler, D., 2011. Determinants and dynamics of genome accessibility. Nat. Rev. Genet. 12, 554–64. doi:10.1038/nrg3017

Benjamin, A.M., Nichols, M., Burke, T.W., Ginsburg, G.S., Lucas, J.E., 2014. Comparing reference-based RNA-Seq mapping methods for non-human primate data. BMC Genomics 15, 570. doi:10.1186/1471-2164-15-570

Bensmihen, S., 2002. The Homologous ABI5 and EEL Transcription Factors Function Antagonistically to Fine-Tune Gene Expression during Late Embryogenesis. PLANT CELL ONLINE 14, 1391–1403. doi:10.1105/tpc.000869

Berger, N., Dubreucq, B., Roudier, F., Dubos, C., Lepiniec, L., 2011. Transcriptional Regulation of Arabidopsis LEAFY COTYLEDON2 Involves RLE, a cis-Element That Regulates Trimethylation of Histone H3 at Lysine-27. Plant Cell 23, 4065–4078. doi:10.1105/tpc.111.087866

Berjak, P., 2006. Unifying perspectives of some mechanisms basic to desiccation tolerance across life forms. Seed Sci. Res. 16, 1–15. doi:10.1079/SSR2005236

Berjak, P., Pammenter, N.W., 2008. From Avicennia to Zizania: seed recalcitrance in perspective. Ann. Bot. 101, 213–28. doi:10.1093/aob/mcm168

Berke, L., Snel, B., 2015. The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. BMC Evol. Biol. 15, 44. doi:10.1186/s12862-015-0319-z

Berr, A., Shafiq, S., Shen, W.-H., 2011. Histone modifications in transcriptional activation during plant development. Biochim. Biophys. Acta 1809, 567–76. doi:10.1016/j.bbagrm.2011.07.001

Bewley, J.D., Black, M.J., 1994. Seeds: Physiology of Development and Germination. Plenum Press, New York, NY.

Bewley, J.D., Bradford, K.J., Hilhorst, H.W.M., Nonogaki, H., 2013. Development and maturation, in: Seeds. Springer, pp. 27–83.

Bewley, J.D., Krochko, J.E., 1982. Desiccation-tolerance, in: Physiological Plant Ecology II. Springer, pp. 325–378.

Bewley, J.D., Reynolds, T.L., Oliver, M.J., 1993. Evolving strategies in the adaptation to desiccation. Curr. Top. plant Physiol.

Bhaskara, G.B., Nguyen, T.T., Verslues, P.E., 2012. Unique drought resistance functions of the highly ABA-induced clade A protein phosphatase 2Cs. Plant Physiol. 160, 379–95. doi:10.1104/pp.112.202408

Bhatnagar-Mathur, P., Vadez, V., Sharma, K.K., 2008. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. Plant Cell Rep. 27, 411–24. doi:10.1007/s00299-007-0474-9

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6–21.

doi:10.1101/gad.947102

Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A., Jones, S.J.M., 2009. De novo transcriptome assembly with ABySS. Bioinformatics 25, 2872–7. doi:10.1093/bioinformatics/btp367

Bobb, A., 1997. Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. Nucleic Acids Res. 25, 641–647. doi:10.1093/nar/25.3.641

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–20. doi:10.1093/bioinformatics/btu170

Bopp, M., Werner, O., 1993. Abscisic Acid and Desiccation Tolerance in Mosses. Bot. Acta 106, 103–106. doi:10.1111/j.1438-8677.1993.tb00344.x

Boucher, V., Buitink, J., Lin, X., Boudet, J., Hoekstra, F.A., Hundertmark, M., Renard, D., Leprince, O., 2010. MtPM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. Plant. Cell Environ. 33, 418–30. doi:10.1111/j.1365-3040.2009.02093.x

Boudsocq, M., Barbier-Brygoo, H., Laurière, C., 2004. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana. J. Biol. Chem. 279, 41758–66. doi:10.1074/jbc.M405259200

Boudsocq, M., Droillard, M.-J., Barbier-Brygoo, H., Laurière, C., 2007. Different phosphorylation mechanisms are involved in the activation of sucrose non-fermenting 1 related protein kinases 2 by osmotic stresses and abscisic acid. Plant Mol. Biol. 63, 491–503. doi:10.1007/s11103-006-9103-1

Bouyer, D., Roudier, F., Heese, M., Andersen, E.D., Gey, D., K, M., Goodrich, J., Renou, J., Grini, P.E., Colot, V., Schnittger, A., 2011. Polycomb Repressive Complex 2 Controls the Embryo-to- Seedling Phase Transition. Transition 7. doi:10.1371/journal.pgen.1002014

Bray, E.A., 1993. Molecular Responses to Water Deficit. Plant Physiol. 103, 1035–1040.

Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., Harada, J.J., 2006. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc. Natl. Acad. Sci. U. S. A. 103, 3468–73. doi:10.1073/pnas.0511331103

Bréhélin, C., Kessler, F., van Wijk, K.J., 2007. Plastoglobules: versatile lipoprotein particles in plastids. Trends Plant Sci. 12, 260–6. doi:10.1016/j.tplants.2007.04.003

Bruggink, T., van der Toorn, P., 1995. Induction of desiccation tolerance in germinated seeds. Seed Sci. Res. 5, 1–4. doi:10.1017/S096025850000252X

Bruinsma, J., 2009. The Resource Outlook to 2050: By How Much Do Land, Water Use and Crop Yields Need to Increase by 2050? Paper prepared for the Expert Meeting on How to Feed the World in 2050, Food and Agriculture Organization, Rome.

Buitink, J., Leger, J.J., Guisle, I., Vu, B.L., Wuillème, S., Lamirault, G., Le Bars, A., Le Meur, N., Becker, A., Küster, H., Leprince, O., 2006. Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-

sensitive to desiccation-tolerant stages in Medicago truncatula seeds. Plant J. 47, 735–50. doi:10.1111/j.1365-313X.2006.02822.x

Buitink, J., Leprince, O., 2008. Intracellular glasses and seed survival in the dry state. C. R. Biol. 331, 788–95. doi:10.1016/j.crvi.2008.08.002

Buitink, J., Leprince, O., 2004. Glass formation in plant anhydrobiotes: survival in the dry state. Cryobiology 48, 215–28. doi:10.1016/j.cryobiol.2004.02.011

Buitink, J., Vu, B.L., Satour, P., Leprince, O., 2003. The re-establishment of desiccation tolerance in germinated radicles of Medicago truncatula Gaertn. seeds. Seed Sci. Res. 13, 273–286. doi:10.1079/SSR2003145

Calonje, M., 2014. PRC1 marks the difference in plant PcG repression. Mol. Plant 7, 459–71. doi:10.1093/mp/sst150

Campi, M., D'Andrea, L., Emiliani, J., Casati, P., 2012. Participation of chromatin-remodeling proteins in the repair of ultraviolet-B-damaged DNA. Plant Physiol. 158, 981–95. doi:10.1104/pp.111.191452

Candat, A., Paszkiewicz, G., Neveu, M., Gautier, R., Logan, D.C., Avelange-Macherel, M.-H., Macherel, D., 2014. The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in Arabidopsis offers tailored protection against abiotic stress. Plant Cell 26, 3148–66. doi:10.1105/tpc.114.127316

Casati, P., Campi, M., Chu, F., Suzuki, N., Maltby, D., Guan, S., Burlingame, A.L., Walbot, V., 2008. Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. Plant Cell 20, 827–42. doi:10.1105/tpc.107.056457

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C.L., Huang, X., 2015. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol. 16, 30. doi:10.1186/s13059-015-0596-2

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z.R., Goodrich, J., 2004. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development 131, 5263–76. doi:10.1242/dev.01400

Chen, D., Molitor, A., Liu, C., Shen, W., 2010. The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. Nat. Publ. Gr. 20, 1332–1344. doi:10.1038/cr.2010.151

Chen, H., Hwang, J.E., Lim, C.J., Kim, D.Y., Lee, S.Y., Lim, C.O., 2010. Arabidopsis DREB2C functions as a transcriptional activator of HsfA3 during the heat stress response. Biochem. Biophys. Res. Commun. 401, 238–44. doi:10.1016/j.bbrc.2010.09.038

Chen, L., Song, Y., Li, S., Zhang, L., Zou, C., Yu, D., 2012. The role of WRKY transcription factors in plant abiotic stresses. Biochim. Biophys. Acta 1819, 120–8. doi:10.1016/j.bbagrm.2011.09.002

Chen, L.-T., Wu, K., 2010. Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. Plant Signal. Behav. 5, 1318–20. doi:10.4161/psb.5.10.13168

Cheng, W.-H., Endo, A., Zhou, L., Penney, J., Chen, H.-C., Arroyo, A., Leon, P., Nambara, E., Asami, T., Seo, M., Koshiba, T., Sheen, J., 2002. A unique short-chain dehydrogenase/reductase in Arabidopsis glucose signaling and abscisic acid

biosynthesis and functions. Plant Cell 14, 2723–43.

Cigliano, R.A., Cremona, G., Paparo, R., Termolino, P., Perrella, G., Gutzat, R., Consiglio, M.F., Conicella, C., 2013. Histone deacetylase AtHDA7 is required for female gametophyte and embryo development in Arabidopsis. Plant Physiol. 163, 431–440.

Close, T.J., 1996. Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. Physiol. Plant. 97, 795–803. doi:10.1111/j.1399-3054.1996.tb00546.x

Collett, H., Shen, A., Gardner, M., Farrant, J.M., Denby, K.J., Illing, N., 2004a. Towards transcript profiling of desiccation tolerance in Xerophyta humilis: Construction of a normalized 11 k X. humilis cDNA set and microarray expression analysis of 424 cDNAs in response to dehydration. Physiol. Plant. 122, 39–53. doi:10.1111/j.1399-3054.2004.00381.x

Collett, H., Shen, A., Gardner, M., Farrant, J.M., Denby, K.J., Illing, N., 2004b. Towards transcript profiling of desiccation tolerance in Xerophyta humilis: Construction of a normalized 11 k X. humilis cDNA set and microarray expression analysis of 424 cDNAs in response to dehydration. Physiol. Plant. 122, 39–53. doi:10.1111/j.1399-3054.2004.00381.x

Collins, L.J., Biggs, P.J., Voelckel, C., Joly, S., 2008. An approach to transcriptome analysis of non-model organisms using short-read sequences. Genome Inform. 21, 3–14.

Compeau, P.E.C., Pevzner, P.A., Tesler, G., 2011. How to apply de Bruijn graphs to genome assembly. Nat. Biotechnol. 29, 987–91. doi:10.1038/nbt.2023

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–6. doi:10.1093/bioinformatics/bti610

Corley, S.B., Carpenter, R., Copsey, L., Coen, E., 2005. Floral asymmetry involves an interplay between TCP and MYB transcription factors in Antirrhinum. Proc. Natl. Acad. Sci. U. S. A. 102, 5068–73. doi:10.1073/pnas.0501340102

Corrêa, L.G.G., Riaño-Pachón, D.M., Schrago, C.G., dos Santos, R.V., Mueller-Roeber, B., Vincentz, M., 2008. The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS One 3, e2944. doi:10.1371/journal.pone.0002944

Cuming, A.C., 1999. LEA proteins, in: Seed Proteins. Springer, pp. 753–780.

Dace, H., Sherwin, H.W., Illing, N., Farrant, J.M., 1998. Use of metabolic inhibitors to elucidate mechanisms of recovery from desiccation stress in the resurrection plant Xerophyta humilis. Plant Growth Regul. 171–177.

Dalal, M., Inupakutika, M., 2014. Transcriptional regulation of ABA core signaling component genes in sorghum (Sorghum bicolor L. Moench). Mol. Breed. 34, 1517–1525. doi:10.1007/s11032-014-0114-3

Davidson, N.M., Oshlack, A., 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol. 15, 410. doi:10.1186/s13059-014-0410-6

Davletova, S., 2005. The Zinc-Finger Protein Zat12 Plays a Central Role in Reactive Oxygen and Abiotic Stress Signaling in Arabidopsis. PLANT Physiol. 139, 847–856. doi:10.1104/pp.105.068254

Deal, R.B., Henikoff, S., 2011. Histone variants and modifications in plant gene regulation. Curr. Opin. Plant Biol. 14, 116–22. doi:10.1016/j.pbi.2010.11.005

Dean Rider, S., Henderson, J.T., Jerome, R.E., Edenberg, H.J., Romero-Severson, J., Ogas, J., 2003. Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis. Plant J. 35, 33–43. doi:10.1046/j.1365-313X.2003.01783.x

Debeaujon, I., Koornneef, M., 2000. Gibberellin Requirement for Arabidopsis Seed Germination Is Determined Both by Testa Characteristics and Embryonic Abscisic Acid. PLANT Physiol. 122, 415–424. doi:10.1104/pp.122.2.415

Dekkers, B.J.W., Costa, M.C.D., Maia, J., Bentsink, L., Ligterink, W., Hilhorst, H.W.M., 2015. Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. Planta 241, 563–77. doi:10.1007/s00425-014-2240-x

Dekkers, B.J.W., Schuurmans, J.A.M.J., Smeekens, S.C.M., 2004. Glucose delays seed germination in Arabidopsis thaliana. Planta 218, 579–88. doi:10.1007/s00425-003-1154-9

Díaz-Martín, J., Almoguera, C., Prieto-Dapena, P., Espinosa, J.M., Jordano, J., 2005. Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter. Plant Physiol. 139, 1483–94. doi:10.1104/pp.105.069963

Dickinson, C.D., Evans, R.P., Nielsen, N.C., 1988. RY repeats are conserved in the 5′-flanking regions of legume seed-protein genes. Nucleic Acids Res. 16, 371–371. doi:10.1093/nar/16.1.371

Dinakar, C., Bartels, D., 2013. Desiccation tolerance in resurrection plants: new insights from transcriptome, proteome and metabolome analysis. Front. Plant Sci. 4, 482. doi:10.3389/fpls.2013.00482

Ditzer, A., Bartels, D., 2006. Identification of a dehydration and ABA-responsive promoter regulon and isolation of corresponding DNA binding proteins for the group 4 LEA gene CpC2 from C. plantagineum. Plant Mol. Biol. 61, 643–63. doi:10.1007/s11103-006-0038-3

Drechsel, G., Raab, S., Hoth, S., 2010. Arabidopsis zinc-finger protein 2 is a negative regulator of ABA signaling during seed germination. J. Plant Physiol. 167, 1418–1421. doi:10.1016/j.jplph.2010.05.010

Dure, L., Crouch, M., Harada, J., Ho, T.H., Mundy, J., Quatrano, R., Thomas, T., Sung, Z.R., 1989. Common amino acid sequence domains among the LEA proteins of higher plants. Plant Mol. Biol. 12, 475–86. doi:10.1007/BF00036962

Dure, L., Greenway, S.C., Galau, G.A., 1981. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. Biochemistry 20, 4162–4168. doi:10.1021/bi00517a033

Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., Nakamura, C., Takumi, S., 2006. Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. Genes Genet. Syst. 81, 77–91. doi:10.1266/ggs.81.77

Elder, R.H., Dell'Aquila, A., Mezzina, M., Sarasin, A., Osborne, D.J., 1987. DNA ligase in

repair and replication in the embryos of rye, Secale cereale. Mutat. Res. Mol. Mech. Mutagen. 181, 61–71. doi:10.1016/0027-5107(87)90288-0

Elhiti, M., Stasolla, C., Wang, A., 2013. Molecular regulation of plant somatic embryogenesis. Vitr. Cell. Dev. Biol. - Plant 49, 631–642. doi:10.1007/s11627-013-9547-3

Emery, D.E., 1988. Seed propagation of native California plants. Santa Barbara Botanic Garden.

Engelhorn, J., Turck, F., 2010. Metaanalysis of ChIP-chip data. Methods Mol. Biol. 631, 185–207. doi:10.1007/978-1-60761-646-7_14

Eshed, Y., Baum, S.F., Bowman, J.L., 1999. Distinct Mechanisms Promote Polarity Establishment in Carpels of Arabidopsis. Cell 99, 199–209. doi:10.1016/S0092-8674(00)81651-7

Evenari, M., Shanan, L., Tadmor, N., 1971. The Negev: The challenge of a desert., in: The Negev. The Challenge of a Desert. Cambridge, Mass.: Harvard University Press/London: Oxford University Press.

Ezcurra, I., Wycliffe, P., Nehlin, L., Ellerstrom, M., Rask, L., 2000. Transactivation of the Brassica napus napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. Plant J. 24, 57–66. doi:10.1046/j.1365-313x.2000.00857.x

Fan, W., Zhao, M., Li, S., Bai, X., Li, J., Meng, H., Mu, Z., Peleg, Z., Blumwald, E., Cutler, S., Rodriguez, P., Finkelstein, R., Abrams, S., González-Guzmán, M., Rodríguez, L., Lorenzo-Orts, L., Pons, C., Sarrión-Perdigones, A., Fernández, M., Shen, Y., Wang, X., Wu, F., Du, S., Cao, Z., Shang, Y., Pandey, S., Nelson, D., Assmann, S., Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., Park, S., Fung, P., Nishimura, N., Jensen, D., Fujii, H., Zhao, Y., Santiago, J., Rodrigues, A., Saez, A., Rubio, S., Antoni, R., Dupeux, F., Nishimura, N., Sarkeshik, A., Nito, K., Park, S., Wang, A., Carvalho, P., Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S., Umezawa, T., Nakashima, K., Miyakawa, T., Kuromori, T., Tanokura, M., Shinozaki, K., Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., Zhao, Y., Chan, Z., Gao, J., Xing, L., Cao, M., Yu, C., Saavedra, X., Modrego, A., Rodríguez, D., González-García, M., Sanz, L., Nicolás, G., Chai, Y., Jia, H., Li, C., Dong, Q., Shen, Y., Sun, L., Wang, Y., Chen, P., Ren, J., Ji, K., Li, Q., Kim, H., Hwang, H., Hong, J., Lee, Y., Ahn, I., Yoon, I., Romero, P., Lafuente, M., Rodrigo, M., Merilo, E., Laanemets, K., Hu, H., Xue, S., Jakobson, L., Tulva, I., Klingler, J., Batelli, G., Zhu, J., Hauser, F., Waadt, R., Schroeder, J., Komatsu, K., Suzuki, N., Kuwamura, M., Nishikawa, Y., Nakatani, M., Ohtawa, H., Sharp, R., LeNoble, M., Antoni, R., Gonzalez-Guzman, M., Rodriguez, L., Peirats-Llobet, M., Pizzio, G., Fernandez, M., Gonzalez-Guzman, M., Pizzio, G., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G., Zhao, Y., Xing, L., Wang, X., Hou, Y., Gao, J., Wang, P., Yamaguchi-Shinozaki, K., Shinozaki, K., Virlouvet, L., Ding, Y., Fujii, H., Avramova, Z., Fromm, M., Ribeiro, D., Desikan, R., Bright, J., Confraria, A., Harrison, J., Hancock, J., Pantin, F., Renaud, J., Barbier, F., Vavasseur, A., Thiec, D., Rose, C., Wang, Y., Yu, H., Zhang, Y., Lai, C., She, Y., Li, W., Broz, A., Thelen, J., Muszynski, M., Miernyk, J., Randall, D., Wei, K., Pan, S., Huai, J., Wang, M., He, J., Zheng, J., Dong, Z., Lv, H., Ying, S., Zhang, D., Li, H., Liu, Y., Shi, Y.,

Song, Y., Wei, K., Wang, Y., Xie, D., Wei, K., Wang, Y., Zhong, X., Pan, S., Alexandrov, N., Brover, V., Freidin, S., Troukhan, M., Tatarinova, T., Zhang, H., Soderlund, C., Descour, A., Kudrna, D., Bomhoff, M., Boyd, L., Currie, J., Boursiac, Y., Léran, S., Corratgé-Faillie, C., Gojon, A., Krouk, G., Lacombe, B., Fan, W., Li, J., Jia, J., Wang, F., Cao, C., Hu, J., Fujii, H., Zhu, J., Fujita, Y., Nakashima, K., Yoshida, T., Katagiri, T., Kidokoro, S., Kanamori, N., Finkelstein, R., Pizzio, G., Rodriguez, L., Antoni, R., Gonzalez-Guzman, M., Yunta, C., Merilo, E., Kim, H., Lee, K., Hwang, H., Bhatnagar, N., Kim, D., Yoon, I., Wang, Y., Wu, Y., Duan, C., Chen, P., Li, Q., Dai, S., Ramegowda, V., Senthil-Kumar, M., Fujita, Y., Fujita, M., Shinozaki, K., Yamaguchi-Shinozaki, K., Szostkiewicz, I., Richter, K., Kepka, M., Demmel, S., Ma, Y., Korte, A., Xu, Z., Kim, D., Hwang, I., Yin, Y., Adachi, Y., Ye, W., Hayashi, M., Nakamura, Y., Kinoshita, T., Shi, W., Li, H., Liu, T., Polle, A., Peng, C., Luo, Z., 2016. Contrasting transcriptional responses of PYR1/PYL/RCAR ABA receptors to ABA or dehydration stress between maize seedling leaves and roots. BMC Plant Biol. 16, 99. doi:10.1186/s12870-016-0764-x

Fang, H., Liu, X., Thorn, G., Duan, J., Tian, L., 2014. Expression analysis of histone acetyltransferases in rice under drought stress. Biochem. Biophys. Res. Commun. 443, 400–5. doi:10.1016/j.bbrc.2013.11.102

Farrant, J.M., 2000. A comparison of mechanisms of desiccation tolerance among three angiosperm resurrection plant species. Plant Ecol. 29–39.

Farrant, J.M., Pammenter, N.W., Berjak, P., 1993. Seed development in relation to desiccation tolerance: A comparison between desiccation-sensitive (recalcitrant) seeds of Avicennia marina and desiccation-tolerant types. Seed Sci. Res. 3, 1–13. doi:10.1017/S0960258500001513

Farrant, J.M., Pammenter, N.W., Berjak, P., Walters, C., 1997. Subcellular organization and metabolic activity during the development of seeds that attain different levels of desiccation tolerance. Seed Sci. Res. 7, 135–144. doi:10.1017/S0960258500003470

Farrant, J.M., Walters, C., 1998. Ultrastructural and biophysical changes in developing embryos of Aesculus hippocastanum in relation to the acquisition of tolerance to drying. Physiol. Plant. 104, 513–524. doi:10.1034/j.1399-3054.1998.1040401.x

Feder, M.E., Hofmann, G.E., 1999. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. Annu. Rev. Physiol. 61, 243–82. doi:10.1146/annurev.physiol.61.1.243

Feng, S., Martinez, C., Gusmaroli, G., Wang, Y., Zhou, J., Wang, F., Chen, L., Yu, L., Iglesias-Pedraz, J.M., Kircher, S., Schäfer, E., Fu, X., Fan, L.-M., Deng, X.W., 2008. Coordinated regulation of Arabidopsis thaliana development by light and gibberellins. Nature 451, 475–9. doi:10.1038/nature06448

Feurtado, J.A., Huang, D., Wicki-Stordeur, L., Hemstock, L.E., Potentier, M.S., Tsang, E.W.T., Cutler, A.J., 2011. The Arabidopsis C2H2 zinc finger INDETERMINATE DOMAIN1/ENHYDROUS promotes the transition to germination by regulating light and hormonal signaling during seed maturation. Plant Cell 23, 1772–94. doi:10.1105/tpc.111.085134

Finch-savage, W.E., Leubner-metzger, G., 2006. Seed dormancy and the control of germination. New Phytol. 501–523.

Finkelstein, R.R., Gampala, S.S.L., Rock, C.D., 2002. Abscisic Acid Signaling in Seeds and Seedlings. Science (80-.). 15–46. doi:10.1105/tpc.010441.would

Finkelstein, R.R., Lynch, T.J., 2000. The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. Plant Cell 12, 599–609.

Fisher, K.M., 2008. Bayesian Reconstruction of Ancestral Expression of the LEA Gene Families Reveals Propagule-derived Desiccation Tolerance in Resurrection Plants. Am. J. Bot. 95, 506–515.

Friedman, J., Stein, Z., Rushkin, E., 1981. Drought tolerance of germinating seeds and young seedlings of Anastatica hierochuntica L. Oecologia 51, 400–403.

Frohman, M.A., 1993. [24] Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: Thermal race. Methods Enzymol. 218, 340–356. doi:10.1016/0076-6879(93)18026-9

Fujii, H., Verslues, P.E., Zhu, J.-K., 2011. Arabidopsis decuple mutant reveals the importance of SnRK2 kinases in osmotic stress responses in vivo. Proc. Natl. Acad. Sci. U. S. A. 108, 1717–22. doi:10.1073/pnas.1018367108

Fujii, H., Verslues, P.E., Zhu, J.-K., 2007. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in Arabidopsis. Plant Cell 19, 485–94. doi:10.1105/tpc.106.048538

Fujii, H., Zhu, J.-K., 2009. Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. Proc. Natl. Acad. Sci. U. S. A. 106, 8380–5. doi:10.1073/pnas.0903144106

Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L.-S.P., Yamaguchi-Shinozaki, K., Shinozaki, K., 2004. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J. 39, 863–76. doi:10.1111/j.1365-313X.2004.02171.x

Fujita, Y., Fujita, M., Satoh, R., Maruyama, K., Parvez, M.M., Seki, M., Hiratsu, K., Ohme-Takagi, M., Shinozaki, K., Yamaguchi-Shinozaki, K., 2005. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. Plant Cell 17, 3470–88. doi:10.1105/tpc.105.035659

Fujita, Y., Nakashima, K., Yoshida, T., Katagiri, T., Kidokoro, S., Kanamori, N., Umezawa, T., Fujita, M., Maruyama, K., Ishiyama, K., Kobayashi, M., Nakasone, S., Yamada, K., Ito, T., Shinozaki, K., Yamaguchi-Shinozaki, K., 2009. Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in Arabidopsis. Plant Cell Physiol. 50, 2123–32. doi:10.1093/pcp/pcp147

Fujita, Y., Yoshida, T., Yamaguchi-Shinozaki, K., 2013. Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. Physiol. Plant. 147, 15–27. doi:10.1111/j.1399-3054.2012.01635.x

Fujiwara, T., Nambara, E., Yamagishi, K., Goto, D.B., Naito, S., 2002. Storage proteins. Arabidopsis Book 1, e0020. doi:10.1199/tab.0020

Furihata, T., Maruyama, K., Fujita, Y., Umezawa, T., Yoshida, R., Shinozaki, K., Yamaguchi-Shinozaki, K., 2006. Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. Proc. Natl. Acad. Sci. U. S. A. 103, 1988–93.

doi:10.1073/pnas.0505667103

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

Gagete, A.P., Riera, M., Franco, L., Rodrigo, M.I., 2009. Functional analysis of the isoforms of an ABI3-like factor of Pisum sativum generated by alternative splicing. J. Exp. Bot. 60, 1703–14. doi:10.1093/jxb/erp038

Gaj, M.D., Zhang, S., Harada, J.J., Lemaux, P.G., 2005. Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. Planta 222, 977–88. doi:10.1007/s00425-005-0041-y

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., Rajjou, L., 2014. Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. Mol. Cell. Proteomics 13, 252–68. doi:10.1074/mcp.M113.032227

Gambetta, G.A., Matthews, M.A., Shaghasi, T.H., McElrone, A.J., Castellarin, S.D., 2010. Sugar and abscisic acid signaling orthologs are activated at the onset of ripening in grape. Planta 232, 219–34. doi:10.1007/s00425-010-1165-2

Garay-Arroyo, A., 2000. Highly Hydrophilic Proteins in Prokaryotes and Eukaryotes Are Common during Conditions of Water Deficit. J. Biol. Chem. 275, 5668–5674. doi:10.1074/jbc.275.8.5668

Garcia, M.E., Lynch, T., Peeters, J., Snowden, C., Finkelstein, R., 2008. A small plant-specific protein family of ABI five binding proteins (AFPs) regulates stress response in germinating Arabidopsis seeds and seedlings. Plant Mol. Biol. 67, 643–58. doi:10.1007/s11103-008-9344-2

Gazzarrini, S., Tsuchiya, Y., Lumba, S., Okamoto, M., 2004. The Transcription Factor FUSCA3 Controls Developmental Timing in Arabidopsis through the Hormones Gibberellin and Abscisic Acid. Cell 7, 373–385.

Gechev, T.S., Benina, M., Obata, T., Tohge, T., Sujeeth, N., Minkov, I., Hille, J., Temanni, M.-R., Marriott, A.S., Bergström, E., Thomas-Oates, J., Antonio, C., Mueller-Roeber, B., Schippers, J.H.M., Fernie, A.R., Toneva, V., 2013. Molecular mechanisms of desiccation tolerance in the resurrection glacial relic Haberlea rhodopensis. Cell. Mol. Life Sci. 70, 689–709. doi:10.1007/s00018-012-1155-6

Gechev, T.S., Hille, J., Woerdenbag, H.J., Benina, M., Mehterov, N., Toneva, V., Fernie, A.R., Mueller-Roeber, B., 2014. Natural products from resurrection plants: potential for medical applications. Biotechnol. Adv. 32, 1091–101. doi:10.1016/j.biotechadv.2014.03.005

Gibson, S.I., 2005. Control of plant development and gene expression by sugar signaling. Curr. Opin. Plant Biol. 8, 93–102. doi:10.1016/j.pbi.2004.11.003

Gilbert, D., 2013. Gene-omes built from mRNA seq not genome DNA., in: 7th Annual Arthropod Genomics Symposium. Notre Dame.

Goldberg, A.D., Allis, C.D., Bernstein, E., 2007. Epigenetics : A Landscape Takes Shape. Cell 635–638. doi:10.1016/j.cell.2007.02.006

González-Guzmán, M., Apostolova, N., Bellés, J.M., Barrero, J.M., Piqueras, P., Ponce, M.R.,

Micol, J.L., Serrano, R., Rodríguez, P.L., 2002. The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde. Plant Cell 14, 1833–46.

Gonzalez-Guzman, M., Pizzio, G.A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G.W., Fernández, M.A., Holdsworth, M.J., Perez-Amador, M.A., Kollist, H., Rodriguez, P.L., 2012. Arabidopsis PYR/PYL/RCAR receptors play a major role in quantitative regulation of stomatal aperture and transcriptional response to abscisic acid. Plant Cell 24, 2483–96. doi:10.1105/tpc.112.098574

Goyal, K., Tisi, L., Basran, A., Browne, J., Burnell, A., Zurdo, J., Tunnacliffe, A., 2003. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. J. Biol. Chem. 278, 12977–84. doi:10.1074/jbc.M212007200

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–52. doi:10.1038/nbt.1883

Granier, C., 1999. Water Deficit and Spatial Pattern of Leaf Development. Variability in Responses Can Be Simulated Using a Simple Model of Leaf Development. PLANT Physiol. 119, 609–620. doi:10.1104/pp.119.2.609

Gray, A., 1887. The elements of botany for beginners and for schools. American Book Company.

Grelet, J., Benamar, A., Teyssier, E., Avelange-Macherel, M.-H., Grunwald, D., Macherel, D., 2005. Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. Plant Physiol. 137, 157–67. doi:10.1104/pp.104.052480

Griffiths, C.A., Gaff, D.F., Neale, A.D., 2014. Drying without senescence in resurrection plants. Front. Plant Sci. 5, 36. doi:10.3389/fpls.2014.00036

Groot, S.P.C., Soeda, Y., van der Geest, A.H.M., Adkins, S.W., Ashmore, S., Navie, S.C., others, 2007. Gene Expression in Relation to Seed Development and Longevity. Seeds Biol. Dev. Ecol. 28.

Guitton, A., Page, D.R., Chambrier, P., Lionnet, C., Faure, J., 2004. Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Image (Rochester, N.Y.) 2971–2981. doi:10.1242/dev.01168

Guo, M., Thomas, J., Collins, G., Timmermans, M.C.P., 2008. Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of Arabidopsis. Plant Cell 20, 48–58. doi:10.1105/tpc.107.056127

Gupta, S.C., Sharma, A., Mishra, M., Mishra, R.K., Chowdhuri, D.K., 2010. Heat shock proteins in toxicology: how close and how far? Life Sci. 86, 377–84. doi:10.1016/j.lfs.2009.12.015

Gutierrez, L., Wuytswinkel, O. Van, Castelain, M., Bellini, C., 2007. Combined networks

regulating seed maturation. Trends Plant Sci. 12. doi:10.1016/j.tplants.2007.06.003

Gutterman, Y., 2000. Environmental factors and survival strategies of annual plant species in the Negev Desert, Israel. Plant Species Biol. 15, 113–125.

Hackenberg, D., Wu, Y., Voigt, A., Adams, R., Schramm, P., Grimm, B., 2012. Studies on differential nuclear translocation mechanism and assembly of the three subunits of the Arabidopsis thaliana transcription factor NF-Y. Mol. Plant 5, 876–88. doi:10.1093/mp/ssr107

Hand, S.C., Menze, M.A., Toner, M., Boswell, L., Moore, D., 2011. LEA proteins during water stress: not just for plants anymore. Annu. Rev. Physiol. 73, 115–34. doi:10.1146/annurev-physiol-012110-142203

Happel, N., Doenecke, D., 2009. Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431, 1–12. doi:10.1016/j.gene.2008.11.003

Harada, J.J., 2001. Role of Arabidopsis LEAFY COTYLEDON genes in seed development. J. Plant Physiol. 158, 405–409. doi:10.1078/0176-1617-00351

Hasegawa, P.M., Bressan, R.A., Zhu, J.-K., Bohnert, H.J., 2000. PLANT CELLULAR AND MOLECULAR RESPONSES TO HIGH SALINITY. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51, 463–499. doi:10.1146/annurev.arplant.51.1.463

He, Y., Hao, Q., Li, W., Yan, C., Yan, N., Yin, P., 2014. Identification and characterization of ABA receptors in Oryza sativa. PLoS One 9, e95246. doi:10.1371/journal.pone.0095246

Hennig, L., Derkacheva, M., 2009. Diversity of Polycomb group complexes in plants : same rules , different players ? Trends Genet. doi:10.1016/j.tig.2009.07.002

Heo, J.B., Sung, S., 2011. Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science (80-.). 331. doi:10.1126/science.1197349

Hilbricht, T., Varotto, S., Sgaramella, V., Bartels, D., Salamini, F., Furini, A., 2008. Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant Craterostigma plantagineum. New Phytol. 179, 877–87. doi:10.1111/j.1469-8137.2008.02480.x

Hoekstra, F., Golovina, E., Buitink, J., 2001. Mechanisms of plant desiccation tolerance. Trends Plant Sci. 6, 431–438.

Hoekstra, F.A., van Roekel, T., 1983. Isolation-Inflicted Injury to Mitochondria from Fresh Pollen Gradually Overcome by an Active Strengthening during Germination. PLANT Physiol. 73, 995–1001. doi:10.1104/pp.73.4.995

Höfler, K., Migsch, H., Rottenburg, W., 1941. {Ü}ber die Austrocknungresistenz landwirtschaftlicher Kulturpflanzen. Forschungsdienst 12, 50–61.

Holdsworth, M.J., Mu&#xF;oz-Blanco, J., Hammond-Kosack, M., Colot, V., Schuch, W., Bevan, M.W., 1995. The maize transcription factor Opaque-2 activates a wheat glutenin promoter in plant and yeast cells. Plant Mol. Biol. 29, 711–720. doi:10.1007/BF00041162

Holiday, J., 2007. Characterisation of the potentially drought induced H1 variant found in Xerophyta viscosa. UCT.

Hrabak, E.M., Chan, C.W.M., Gribskov, M., Harper, J.F., Choi, J.H., Halford, N., Kudla, J., Luan, S., Nimmo, H.G., Sussman, M.R., Thomas, M., Walker-Simmons, K., Zhu, J.-K.,

Harmon, A.C., 2003. The Arabidopsis CDPK-SnRK superfamily of protein kinases. Plant Physiol. 132, 666–80. doi:10.1104/pp.102.011999

Hu, W., Hu, G., Han, B., 2009. Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. Plant Sci. 176, 583–590. doi:10.1016/j.plantsci.2009.01.016

Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res. 9, 868–77.

Huang, Z., Dong, M., Gutterman, Y., 2004. Factors influencing seed dormancy and germination in sand, and seedling survival under desiccation, of Psammochloa villosa (Poaceae), inhabiting the moving sand dunes of Ordos, China. Plant Soil 259, 231–241. doi:10.1023/B:PLSO.0000020971.66784.fc

Hundertmark, M., Hincha, D.K., 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics 9, 118. doi:10.1186/1471-2164-9-118

Illing, N., Denby, K.J., Collett, H., Shen, A., Farrant, J.M., 2005. The Signature of Seeds in Resurrection Plants: A Molecular and Physiological Comparison of Desiccation Tolerance in Seeds and Vegetative Tissues. Integr. Comp. Biol. 45, 771–787. doi:10.1093/icb/45.5.771

Ingram, J., Bartels, D., 1996. the Molecular Basis of Dehydration Tolerance in Plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 377–403. doi:10.1146/annurev.arplant.47.1.377

Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., Tabata, S., Kakubari, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., 2001. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. Plant J. 27, 325–33.

Jackson, B.G., Schnable, P.S., Aluru, S., 2009. Parallel short sequence assembly of transcriptomes. BMC Bioinformatics 10 Suppl 1, S14. doi:10.1186/1471-2105-10-S1-S14

Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., Schubert, I., Jenuwein, T., Jacobsen, S.E., 2004. Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 308–315. doi:10.1007/s00412-004-0275-7

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F., 2002. bZIP transcription factors in Arabidopsis. Trends Plant Sci. 7, 106–111. doi:10.1016/S1360-1385(01)02223-3

Je, J., Chen, H., Song, C., Lim, C.O., 2014. Arabidopsis DREB2C modulates ABA biosynthesis during germination. Biochem. Biophys. Res. Commun. 452, 91–8. doi:10.1016/j.bbrc.2014.08.052

Jenuwein, T., Allis, C.D., 2001. Translating the histone code. Science 293, 1074–80. doi:10.1126/science.1063127

Jia, H., McCarty, D.R., Suzuki, M., 2013. Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. Plant Physiol. 163, 1293–305. doi:10.1104/pp.113.220988

Jia, H., Suzuki, M., McCarty, D.R., 2014. Regulation of the seed to seedling developmental

phase transition by the LAFL and VAL transcription factor networks. Wiley Interdiscip. Rev. Dev. Biol. 3, 135–45. doi:10.1002/wdev.126

Jia, H., Wang, Y., Sun, M., Li, B., Han, Y., Zhao, Y., Li, X., Ding, N., Li, C., Ji, W., Jia, W., 2013. Sucrose functions as a signal involved in the regulation of strawberry fruit development and ripening. New Phytol. 198, 453–465. doi:10.1111/nph.12176

Johnson, M.T.J., Carpenter, E.J., Tian, Z., Bruskiewich, R., Burris, J.N., Carrigan, C.T., Chase, M.W., Clarke, N.D., Covshoff, S., Depamphilis, C.W., Edger, P.P., Goh, F., Graham, S., Greiner, S., Hibberd, J.M., Jordon-Thaden, I., Kutchan, T.M., Leebens-Mack, J., Melkonian, M., Miles, N., Myburg, H., Patterson, J., Pires, J.C., Ralph, P., Rolf, M., Sage, R.F., Soltis, D., Soltis, P., Stevenson, D., Stewart, C.N., Surek, B., Thomsen, C.J.M., Villarreal, J.C., Wu, X., Zhang, Y., Deyholos, M.K., Wong, G.K.-S., 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One 7, e50226. doi:10.1371/journal.pone.0050226

Johnson, R.R., Wagner, R.L., Verhey, S.D., Walker-Simmons, M.K., 2002. The abscisic acid-responsive kinase PKABA1 interacts with a seed-specific abscisic acid response element-binding factor, TaABF, and phosphorylates TaABF peptide sequences. Plant Physiol. 130, 837–46. doi:10.1104/pp.001354

Juenger, T.E., 2013. Natural variation and genetic constraints on drought tolerance. Curr. Opin. Plant Biol. 16, 274–81. doi:10.1016/j.pbi.2013.02.001

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., 2005a. LEAFY COTYLEDON1 Controls Seed Storage Protein Genes through Its Regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell 46, 399–406. doi:10.1093/pcp/pci048

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T., 2005b. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiol. 46, 399–406. doi:10.1093/pcp/pci048

Kang, J. -y., 2002. Arabidopsis Basic Leucine Zipper Proteins That Mediate Stress-Responsive Abscisic Acid Signaling. PLANT CELL ONLINE 14, 343–357. doi:10.1105/tpc.010362

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R., 2007a. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484–8. doi:10.1126/science.1138341

Kapranov, P., Willingham, A.T., Gingeras, T.R., 2007b. Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. 8, 413–23. doi:10.1038/nrg2083

Karssen, C.M., Brinkhorst-van der Swan, D.L., Breekland, A.E., 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, 158–65. doi:10.1007/BF00393650

Keck, K.M., Pemberton, L.F., 2012. Histone chaperones link histone nuclear import and chromatin assembly. Biochim. Biophys. Acta 1819, 277–89.

doi:10.1016/j.bbagrm.2011.09.007

Kikawada, T., Nakahara, Y., Kanamori, Y., Iwata, K., Watanabe, M., McGee, B., Tunnacliffe, A., Okuda, T., 2006. Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. Biochem. Biophys. Res. Commun. 348, 56–61. doi:10.1016/j.bbrc.2006.07.003

Kim, H., Hwang, H., Hong, J.-W., Lee, Y.-N., Ahn, I.P., Yoon, I.S., Yoo, S.-D., Lee, S., Lee, S.C., Kim, B.-G., 2012. A rice orthologue of the ABA receptor, OsPYL/RCAR5, is a positive regulator of the ABA signal transduction pathway in seed germination and early seedling growth. J. Exp. Bot. 63, 1013–24. doi:10.1093/jxb/err338

Kim, J.-M., Sasaki, T., Ueda, M., Sako, K., Seki, M., 2015. Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. Front. Plant Sci. 6, 114. doi:10.3389/fpls.2015.00114

Kim, J.-S., Mizoi, J., Yoshida, T., Fujita, Y., Nakajima, J., Ohori, T., Todaka, D., Nakashima, K., Hirayama, T., Shinozaki, K., Yamaguchi-Shinozaki, K., 2011. An ABRE promoter sequence is involved in osmotic stress-responsive expression of the DREB2A gene, which encodes a transcription factor regulating drought-inducible genes in Arabidopsis. Plant Cell Physiol. 52, 2136–46. doi:10.1093/pcp/pcr143

Kim, S., Kang, J.-Y., Cho, D.-I., Park, J.H., Kim, S.Y., 2004. ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. Plant J. 40, 75–87. doi:10.1111/j.1365-313X.2004.02192.x

Kim, S.Y., Lee, J., Eshed-Williams, L., Zilberman, D., Sung, Z.R., 2012. EMF1 and PRC2 cooperate to repress key regulators of Arabidopsis development. PLoS Genet. 8, e1002512. doi:10.1371/journal.pgen.1002512

Kirkbride, R.C., Fischer, R.L., Harada, J.J., 2013. LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of Selaginella moellendorffii. PLoS One 8, e67971. doi:10.1371/journal.pone.0067971

Kissoudis, C., van de Wiel, C., Visser, R.G.F., van der Linden, G., 2014. Enhancing crop resilience to combined abiotic and biotic stress through the dissection of physiological and molecular crosstalk. Front. Plant Sci. 5, 207. doi:10.3389/fpls.2014.00207

Kobayashi, Y., Murata, M., Minami, H., Yamamoto, S., Kagaya, Y., Hobo, T., Yamamoto, A., Hattori, T., 2005. Abscisic acid-activated SNRK2 protein kinases function in the gene-regulation pathway of ABA signal transduction by phosphorylating ABA response element-binding factors. Plant J. 44, 939–49. doi:10.1111/j.1365-313X.2005.02583.x

Kobayashi, Y., Yamamoto, S., Minami, H., Kagaya, Y., Hattori, T., 2004a. Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid. Plant Cell 16, 1163–77. doi:10.1105/tpc.019943

Kobayashi, Y., Yamamoto, S., Minami, H., Kagaya, Y., Hattori, T., 2004b. Differential activation of the rice sucrose nonfermenting1-related protein kinase2 family by hyperosmotic stress and abscisic acid. Plant Cell 16, 1163–77. doi:10.1105/tpc.019943

Komatsu, K., Suzuki, N., Kuwamura, M., Nishikawa, Y., Nakatani, M., Ohtawa, H., Takezawa, D., Seki, M., Tanaka, M., Taji, T., Hayashi, T., Sakata, Y., 2013. Group A PP2Cs evolved in land plants as key regulators of intrinsic desiccation tolerance. Nat.

Commun. 4, 2219. doi:10.1038/ncomms3219

Koonjul, P.K., Brandt, W.F., Lindsey, G.G., Farrant, J.M., 2000. Isolation and Characterisation of Chloroplasts from Myrothamnus flabellifolius Welw. J. Plant Physiol. 156, 584–594. doi:10.1016/S0176-1617(00)80217-5

Koster, K.L., Balsamo, R.A., Espinoza, C., Oliver, M.J., 2010. Desiccation sensitivity and tolerance in the moss Physcomitrella patens: assessing limits and damage. Plant Growth Regul. 62, 293–302. doi:10.1007/s10725-010-9490-9

Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., Scharf, K.-D., 2007. Complexity of the heat stress response in plants. Curr. Opin. Plant Biol. 10, 310–6. doi:10.1016/j.pbi.2007.04.011

Kovacs, D., Agoston, B., Tompa, P., 2014. Disordered plant LEA proteins as molecular chaperones. Plant Signal. Behav. 3, 710–713. doi:10.4161/psb.3.9.6434

Kranner, I., Beckett, R.P., Wornik, S., Zorn, M., Pfeifhofer, H.W., 2002. Revival of a resurrection plant correlates with its antioxidant status. Plant J. 31, 13–24. doi:10.1046/j.1365-313X.2002.01329.x

Kranner, I., Birtic, S., 2005. A modulating role for antioxidants in desiccation tolerance. Integr. Comp. Biol. 45, 734–40. doi:10.1093/icb/45.5.734

Kroj, T., Savino, G., Valon, C., Giraudat, J., Parcy, F., 2003. Regulation of storage protein gene expression in Arabidopsis. Development 130, 6065–73. doi:10.1242/dev.00814

Kucera, B., Cohn, M.A., Leubner-Metzger, G., 2005. Plant hormone interactions during seed dormancy release and germination. Seed Sci. Res. 15, 281–307. doi:10.1079/SSR2005218

Kulik, A., Wawer, I., Krzywińska, E., Bucholc, M., Dobrowolska, G., 2011. SnRK2 protein kinases--key regulators of plant response to abiotic stresses. OMICS 15, 859–72. doi:10.1089/omi.2011.0091

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–9. doi:10.1038/nmeth.1923

Lara, P., Oñate-Sánchez, L., Abraham, Z., Ferrándiz, C., Díaz, I., Carbonero, P., Vicente-Carbajosa, J., 2003. Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. J. Biol. Chem. 278, 21003–11. doi:10.1074/jbc.M210538200

Lata, C., Prasad, M., 2011. Role of DREBs in regulation of abiotic stress responses in plants. J. Exp. Bot. 62, 4731–48. doi:10.1093/jxb/err210

Lata, C., Yadav, A., Prasad, M., 2011. Role of plant transcription factors in abiotic stress tolerance. Abiotic Stress response plants--Physiological, Biochem. Genet. Perspect. 10, 23172.

Lauria, M., Rossi, V., 2011. Epigenetic control of gene regulation in plants. Biochim. Biophys. Acta - Gene Regul. Mech. 1809, 369–378. doi:10.1016/j.bbagrm.2011.03.002

Le, H.-S., Schulz, M.H., McCauley, B.M., Hinman, V.F., Bar-Joseph, Z., 2013. Probabilistic error correction for RNA sequencing. Nucleic Acids Res. 41, e109. doi:10.1093/nar/gkt215

Lee, H., Fischer, R.L., Goldberg, R.B., Harada, J.J., 2003. Arabidopsis LEAFY COTYLEDON1

represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc. Natl. Acad. Sci. U. S. A. 100, 2152–6. doi:10.1073/pnas.0437909100

Lee, K.H., Piao, H.L., Kim, H.-Y., Choi, S.M., Jiang, F., Hartung, W., Hwang, I., Kwak, J.M., Lee, I.-J., Hwang, I., 2006. Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. Cell 126, 1109–20. doi:10.1016/j.cell.2006.07.034

Lee, S., Cheng, H., King, K.E., Wang, W., He, Y., Hussain, A., Lo, J., Harberd, N.P., Peng, J., 2002a. Gibberellin regulates Arabidopsis seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition. Genes Dev. 16, 646–58. doi:10.1101/gad.969002

Lee, S., Cheng, H., King, K.E., Wang, W., He, Y., Hussain, A., Lo, J., Harberd, N.P., Peng, J., 2002b. Gibberellin regulates Arabidopsis seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition. Genes Dev. 16, 646–58. doi:10.1101/gad.969002

Lelievre, J.-M., Oliveira, L.O., Nielsen, N.C., 1992. 5'CATGCAT-3' Elements Modulate the Expression of Glycinin Genes. PLANT Physiol. 98, 387–391. doi:10.1104/pp.98.1.387

Leprince, O., Hoekstra, F.A., Harren, F.J.M., 2000. Unravelling the responses of metabolism to dehydration points to a role for cytoplasmic viscosity in desiccation tolerance. Seed Biol. Adv. Appl. CABI, New York, USA 57–66.

Leubner-Metzger, G., 2005. beta-1,3-Glucanase gene expression in low-hydrated seeds as a mechanism for dormancy release during tobacco after-ripening. Plant J. 41, 133–45. doi:10.1111/j.1365-313X.2004.02284.x

Leung, J., Bouvier-Durand, M., Morris, P., Guerrier, D., Chefdor, F., Giraudat, J., 1994. Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase. Science (80-.). 264, 1448–1452. doi:10.1126/science.7910981

Leung, J., Merlot, S., Giraudat, J., 1997. The Arabidopsis ABSCISIC ACID-INSENSITIVE2 (ABI2) and ABI1 genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. Plant Cell 9, 759–71. doi:10.1105/tpc.9.5.759

Li, H., Yan, S., Zhao, L., Tan, J., Zhang, Q., Gao, F., Wang, P., Hou, H., Li, L., 2014. Histone acetylation associated up-regulation of the cell wall related genes is involved in salt stress induced maize root swelling. BMC Plant Biol. 14, 105. doi:10.1186/1471-2229-14-105

Li, H.-C., Chuang, K., Henderson, J.T., Rider, S.D., Bai, Y., Zhang, H., Fountain, M., Gerber, J., Ogas, J., 2005. PICKLE acts during germination to repress expression of embryonic traits. Plant J. 44, 1010–22. doi:10.1111/j.1365-313X.2005.02602.x

Li, X.-Y., Liu, X., Yao, Y., Li, Y.-H., Liu, S., He, C.-Y., Li, J.-M., Lin, Y.-Y., Li, L., 2013. Overexpression of Arachis hypogaea AREB1 Gene Enhances Drought Tolerance by Modulating ROS Scavenging and Maintaining Endogenous ABA Content. Int. J. Mol. Sci. 14, 12827–42. doi:10.3390/ijms140612827

Li, Y., Jin, K.E., Zhu, Z.H.U., Yang, J.I., 2010. Stepwise origin and functional diversification of the afl subfamily b3 genes during land plant evolution. J. Bioinform. Comput. Biol. 8, 33–45. doi:10.1142/S0219720010005129

Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B., Yang, B., Fan, W., 2012. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. Brief. Funct. Genomics 11, 25–37. doi:10.1093/bfgp/elr035

Liu, H.-C., Liao, H.-T., Charng, Y.-Y., 2011. The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. Plant. Cell Environ. 34, 738–51. doi:10.1111/j.1365-3040.2011.02278.x

Liu, Q., 1998. Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression, Respectively, in Arabidopsis. PLANT CELL ONLINE 10, 1391–1406. doi:10.1105/tpc.10.8.1391

Liu, X., Yang, S., Zhao, M., Luo, M., Yu, C.-W., Chen, C.-Y., Tai, R., Wu, K., 2014. Transcriptional repression by histone deacetylases in plants. Mol. Plant 7, 764–772.

Liu, X.-M., An, J., Han, H.J., Kim, S.H., Lim, C.O., Yun, D.-J., Chung, W.S., 2014. ZAT11, a zinc finger transcription factor, is a negative regulator of nickel ion tolerance in Arabidopsis. Plant Cell Rep. 33, 2015–2021. doi:10.1007/s00299-014-1675-7

Lodha, M., Marco, C.F., Timmermans, M.C.P., 2013. The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. Genes Dev. 27, 596–601. doi:10.1101/gad.211425.112

Lopez-Molina, L., Mongrand, S., Chua, N.H., 2001. A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 98, 4782–7. doi:10.1073/pnas.081594298

Lopez-Molina, L., Mongrand, S., Kinoshita, N., Chua, N.-H., 2003. AFP is a novel negative regulator of ABA signaling that promotes ABI5 protein degradation. Genes Dev. 17, 410–8. doi:10.1101/gad.1055803

Lopez-Molina, L., Mongrand, S., McLachlin, D.T., Chait, B.T., Chua, N.-H., 2002. ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. Plant J. 32, 317–28.

Lotan, T., Ohto, M., Yee, K.M., West, M.A.L., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B., Harada, J.J., 1998. Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. Cell 93, 1195–1205.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. doi:10.1186/s13059-014-0550-8

Lu, F., Li, G., Cui, X., Liu, C., Wang, X.-J., Cao, X., 2008. Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice. J. Integr. Plant Biol. 50, 886–96. doi:10.1111/j.1744-7909.2008.00692.x

Lund, A.H., van Lohuizen, M., 2004. Polycomb complexes and silencing mechanisms. Curr. Opin. Cell Biol. 16, 239–46. doi:10.1016/j.ceb.2004.03.010

Luo, M., Tai, R., Yu, C.-W., Yang, S., Chen, C.-Y., Lin, W.-D., Schmidt, W., Wu, K., 2015. Regulation of flowering time by the histone deacetylase HDA5 in Arabidopsis. Plant J. 82, 925–36. doi:10.1111/tpj.12868

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J.,

Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1, 18. doi:10.1186/2047-217X-1-18

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., Grill, E., 2009. Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science 324, 1064–8. doi:10.1126/science.1172408

Ma, Z., Hu, X., Cai, W., Huang, W., Zhou, X., Luo, Q., Yang, H., Wang, J., Huang, J., 2014. Arabidopsis miR171-Targeted Scarecrow-Like Proteins Bind to GT cis-Elements and Mediate Gibberellin-Regulated Chlorophyll Biosynthesis under Light Conditions. PLoS Genet. 10, e1004519. doi:10.1371/journal.pgen.1004519

Maere, S., Heymans, K., Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448–9. doi:10.1093/bioinformatics/bti551

Maia, J., 2014. Unravelling desiccation tolerance in germinated Arabidopsis seeds. Wageningen University.

Maia, J., Dekkers, B.J.W., Provart, N.J., Ligterink, W., Hihorst, H.W.M., 2011. The re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds and its associated transcriptome. PLoS One 6, 1–11. doi:10.1371/journal.pone.0029123

Makinde, A., Fajuke, A.A., 2009. Adaptive Strategies of Mosses to Desiccation. Not. Bot. Horti Agrobot. Cluj-Napoca. doi:10.15835/nbha3713119

Marella, H.H., Sakata, Y., Quatrano, R.S., 2006. Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from Physcomitrella patens. Plant J. 46, 1032–44. doi:10.1111/j.1365-313X.2006.02764.x

Margueron, R., Trojer, P., Reinberg, D., 2005. The key to development: interpreting the histone code? Curr. Opin. Genet. Dev. 15, 163–176. doi:10.1016/j.gde.2005.01.005

Martin, C., Zhang, Y., 2005. The diverse functions of histone lysine methylation. Nat. Rev. Mol. Cell Biol. 6, 838–49. doi:10.1038/nrm1761

Martinelli, T., 2008. In situ localization of glucose and sucrose in dehydrating leaves of Sporobolus stapfianus. J. Plant Physiol. 165, 580–7. doi:10.1016/j.jplph.2007.01.019

Mayaba, N., 2001. ABA Increases the Desiccation Tolerance of Photosynthesis in the Afromontane Understorey Moss Atrichum androgynum. Ann. Bot. 88, 1093–1100. doi:10.1006/anbo.2001.1552

Mbandi, S.K., Hesse, U., van Heusden, P., Christoffels, A., 2015. Inferring bona fide transfrags in RNA-Seq derived-transcriptome assemblies of non-model organisms. BMC Bioinformatics 16, 58. doi:10.1186/s12859-015-0492-5

McDonald, Z.E., 2008. Sucrose Phosphate Synthase activity and gene expression in relation to dehydration induced sucrose accumulation in the resurrection plan Xerophyta Humilis. University of Cape Town.

Meinke, D.W., 1994. Seed development in Arabidopsis thaliana [WWW Document]. Cold Spring Harb. Monogr. Arch. 27. URL https://dspace.fandm.edu/bitstream/handle/11016/24115/DaoThesis.pdf?sequence=1

(accessed 9.1.15).

Melcher, K., Ng, L.-M., Zhou, X.E., Soon, F.-F., Xu, Y., Suino-Powell, K.M., Park, S.-Y., Weiner, J.J., Fujii, H., Chinnusamy, V., Kovach, A., Li, J., Wang, Y., Li, J., Peterson, F.C., Jensen, D.R., Yong, E.-L., Volkman, B.F., Cutler, S.R., Zhu, J.-K., Xu, H.E., 2009. A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. Nature 462, 602–8. doi:10.1038/nature08613

Merlot, S., Gosti, F., Guerrier, D., Vavasseur, A., Giraudat, J., 2001. The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. Plant J. 25, 295–303. doi:10.1046/j.1365-313x.2001.00965.x

Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46. doi:10.1038/nrg2626

Meurs, C., Basra, A.S., Karssen, C.M., Van Loon, L.C., 1992. Role of abscisic acid in the induction of desiccation tolerance in developing seeds of Arabidopsis thaliana. Plant Physiol. 3, 1484–1493.

Meyer, K., Leube, M., Grill, E., 1994. A protein phosphatase 2C involved in ABA signal transduction in Arabidopsis thaliana. Science (80-.). 264, 1452–1455. doi:10.1126/science.8197457

Milborrow, B.V., 2001. The pathway of biosynthesis of abscisic acid in vascular plants: a review of the present state of knowledge of ABA biosynthesis. J. Exp. Bot. 52, 1145–1164. doi:10.1093/jexbot/52.359.1145

Miller, G., Shulaev, V., Mittler, R., 2008. Reactive oxygen signaling and abiotic stress. Physiol. Plant. 133, 481–9. doi:10.1111/j.1399-3054.2008.01090.x

Miller, J.R., Koren, S., Sutton, G., 2010. Assembly algorithms for next-generation sequencing data. Genomics 95, 315–27. doi:10.1016/j.ygeno.2010.03.001

Mishra, S.K., Tripp, J., Winkelhaus, S., Tschiersch, B., Theres, K., Nover, L., Scharf, K.-D., 2002. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. Genes Dev. 16, 1555–67. doi:10.1101/gad.228802

Mittler, R., 2002. Oxidative stress, antioxidants and stress tolerance. Trends Plant Sci. 7, 405–410. doi:10.1016/S1360-1385(02)02312-9

Miyazono, K.-I., Miyakawa, T., Sawano, Y., Kubota, K., Kang, H.-J., Asano, A., Miyauchi, Y., Takahashi, M., Zhi, Y., Fujita, Y., Yoshida, T., Kodaira, K.-S., Yamaguchi-Shinozaki, K., Tanokura, M., 2009. Structural basis of abscisic acid signalling. Nature 462, 609–14. doi:10.1038/nature08583

Mizoguchi, M., Umezawa, T., Nakashima, K., Kidokoro, S., Takasaki, H., Fujita, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., 2010. Two closely related subclass II SnRK2 protein kinases cooperatively regulate drought-inducible gene expression. Plant Cell Physiol. 51, 842–7. doi:10.1093/pcp/pcq041

Molitor, A., Shen, W.-H., 2013. The polycomb complex PRC1: composition and function in plants. J. Genet. Genomics 40, 231–8. doi:10.1016/j.jgg.2012.12.005

Molitor, A.M., Bu, Z., Yu, Y., Shen, W.-H., 2014. Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in

repression of seed developmental genes. PLoS Genet. 10, e1004091. doi:10.1371/journal.pgen.1004091

Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.-P., Bäumlein, H., Conrad, U., 2004. Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. Planta 219, 158–66. doi:10.1007/s00425-004-1206-9

Mönke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hähnel, U., Junker, A., Weisshaar, B., Conrad, U., Bäumlein, H., Altschmied, L., 2012. Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. Nucleic Acids Res. 40, 8240–54. doi:10.1093/nar/gks594

Mooney, C.Z., Duval, R.D., Duval, R., 1993. Bootstrapping: A nonparametric approach to statistical inference. Sage.

Morgan, J.M., 1984. Osmoregulation and Water Stress in Higher Plants. Annu. Rev. Plant Physiol. 35, 299–319. doi:10.1146/annurev.pp.35.060184.001503

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–8. doi:10.1038/nmeth.1226

Mueller, J.K., Heckathorn, S.A., Fernando, D., 2003. Identification of a Chloroplast Dehydrin in Leaves of Mature Plants. Int. J. Plant Sci. 164, 535–542. doi:10.1086/375376

Müller, J., Kassis, J.A., 2006. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr. Opin. Genet. Dev. 16, 476–84. doi:10.1016/j.gde.2006.08.005

Müller, K., Bouyer, D., Schnittger, A., Kermode, A.R., 2012. Evolutionarily Conserved Histone Methylation Dynamics during Seed Life-Cycle Transitions. PLoS One 7, e51532. doi:10.1371/journal.pone.0051532

Murai, M., Yoshida, S., 1998. Evidence for the Cell Wall Involvement in Temporal Changes in Freezing Tolerance of Jerusalem Artichoke (Helianthus tuberosus L.) Tubers during Cold Acclimation. Plant Cell Physiol. 39, 97–105. doi:10.1093/oxfordjournals.pcp.a029295

Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G., Stahl, F., 2013. Transcriptome analysis using next-generation sequencing. Curr. Opin. Biotechnol. 24, 22–30. doi:10.1016/j.copbio.2012.09.004

Myers, M.Y., Farrant, J.M., Roden, L.C., 2010. Preliminary characterization of floral response of Xerophyta humilis to desiccation, vernalisation, photoperiod and light intensity. Plant Growth Regul. 62, 213–216. doi:10.1007/s10725-010-9460-2

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., Nambara, E., 2005. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. Plant J. 41, 697–709. doi:10.1111/j.1365-313X.2005.02337.x

Nakamura, S., Lynch, T.J., Finkelstein, R.R., 2001. Physical interactions between ABA response loci of Arabidopsis. Plant J. 26, 627–635. doi:10.1046/j.1365-313x.2001.01069.x

Nakashima, K., Fujita, Y., Kanamori, N., Katagiri, T., Umezawa, T., Kidokoro, S., Maruyama, K., Yoshida, T., Ishiyama, K., Kobayashi, M., Shinozaki, K., Yamaguchi-Shinozaki, K.,

2009. Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. Plant Cell Physiol. 50, 1345–63. doi:10.1093/pcp/pcp083

Nakashima, K., Shinwari, Z.K., Sakuma, Y., Seki, M., Miura, S., Shinozaki, K., Yamaguchi-Shinozaki, K., 2000. Organization and expression of two Arabidopsis DREB2 genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. Plant Mol. Biol. 42, 657–665. doi:10.1023/A:1006321900483

Nakashima, K., Yamaguchi-Shinozaki, K., 2013. ABA signaling in stress-response and seed development. Plant Cell Rep. 32, 959–70. doi:10.1007/s00299-013-1418-1

Nakashima, K., Yamaguchi-Shinozaki, K., Shinozaki, K., 2014. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front. Plant Sci. 5, 170. doi:10.3389/fpls.2014.00170

Nakasugi, K., Crowhurst, R., Bally, J., Waterhouse, P., 2014. Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant Nicotiana benthamiana. PLoS One 9, e91776. doi:10.1371/journal.pone.0091776

Nambara, E., Marion-Poll, A., 2005. Abscisic acid biosynthesis and catabolism. Annu. Rev. Plant Biol. 56, 165–85. doi:10.1146/annurev.arplant.56.032604.144046

Nambara, E., Marion-poll, A., 2003. ABA action and interactions in seeds. Trends Plant Sci. 8, 213–217. doi:10.1016/S1360-1385(03)00060-8

Napier, J.A., Stobart, A.K., Shewry, P.R., 1996. The structure and biogenesis of plant oil bodies: the role of the ER membrane and the oleosin class of proteins. Plant Mol. Biol. 31, 945–956. doi:10.1007/BF00040714

Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K., Yamaguchi-Shinozaki, K., 2003. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses. Plant J. 34, 137–148. doi:10.1046/j.1365-313X.2003.01708.x

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.-C., Dorn, R., Jenuwein, T., Reuter, G., 2005. Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. EMBO J. 24, 1418–29. doi:10.1038/sj.emboj.7600604

Navarro, L., Bari, R., Achard, P., Lisón, P., Nemri, A., Harberd, N.P., Jones, J.D.G., 2008. DELLAs Control Plant Immune Responses by Modulating the Balance of Jasmonic Acid and Salicylic Acid Signaling. Curr. Biol. 18, 650–655. doi:10.1016/j.cub.2008.03.060

Negbi, M., Evenari, M., 1961. The means of survival of some desert summer annuals. Plant/Water Relationships Arid Semiarid Cond. 249.

Nishimura, N., Hitomi, K., Arvai, A.S., Rambo, R.P., Hitomi, C., Cutler, S.R., Schroeder, J.I., Getzoff, E.D., 2009. Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science 326, 1373–9. doi:10.1126/science.1181829

Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., Shigeoka, S., 2006. Arabidopsis heat shock transcription factor A2 as a key regulator in response to several

types of environmental stress. Plant J. 48, 535–47. doi:10.1111/j.1365-313X.2006.02889.x

Novillo, F., Alonso, J.M., Ecker, J.R., Salinas, J., 2004. CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 101, 3985–90. doi:10.1073/pnas.0303029101

Ogas, J., Kaufmann, S., Henderson, J., Somerville, C., 1999. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proc. Natl. Acad. Sci. 96, 13839–13844. doi:10.1073/pnas.96.24.13839

Oh, S.-J., Song, S.I., Kim, Y.S., Jang, H.-J., Kim, S.Y., Kim, M., Kim, Y.-K., Nahm, B.H., Kim, J.-K., 2005. Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. Plant Physiol. 138, 341–51. doi:10.1104/pp.104.059147

Okada, T., Endo, M., Singh, M.B., Bhalla, P.L., 2005. Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Analysis 557–568. doi:10.1111/j.1365-313X.2005.02554.x

Oliver, M., Tuba, Z., Mishler, B., 2000. The evolution of vegetative desiccation tolerance in land plants. Plant Ecol. 85–100.

Oliver, M.J., Bewley, D.J., 1997. Desiccation-tolerance of plant tissues: a mechanistic overview. Hortic. Rev. (Am. Soc. Hortic. Sci). 18, 171–213.

Oliver, M.J., O'Mahony, P., Wood, A.J., 1998. "To dryness and beyond" – Preparation for the dried state and rehydration in vegetative desiccation-tolerant plants. Plant Growth Regul. 24, 193–201. doi:10.1023/A:1005863015130

Oliver, M.J., Velten, J., Mishler, B.D., 2005. Desiccation tolerance in bryophytes: a reflection of the primitive strategy for plant survival in dehydrating habitats? Integr. Comp. Biol. 45, 788–99. doi:10.1093/icb/45.5.788

Olszewski, N., Sun, T., Gubler, F., 2002. Gibberellin Signaling: Biosynthesis, Catabolism, and Response Pathways. Plant Cell 14, S61–80. doi:10.1105/tpc.010476

Pammenter, N.W., Berjak, P., 1999. A review of recalcitrant seed physiology in relation to desiccation-tolerance mechanisms. Seed Sci. Res. 9, 13–37. doi:10.1017/S0960258599000033

Panaretou, B., Zhai, C., 2008. The heat shock proteins: Their roles as multi-component machines for protein folding. Fungal Biol. Rev. 22, 110–119. doi:10.1016/j.fbr.2009.04.002

Pandey, R., 2002. Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Res. 30, 5036–5055. doi:10.1093/nar/gkf660

Pandey, S., Nelson, D.C., Assmann, S.M., 2009. Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. Cell 136, 136–48. doi:10.1016/j.cell.2008.12.026

Papaefthimiou, D., Likotrafiti, E., Kapazoglou, A., Bladenopoulos, K., Tsaftaris, A., 2010. Epigenetic chromatin modifiers in barley: III. Isolation and characterization of the barley GNAT-MYST family of histone acetyltransferases and responses to exogenous

ABA. Plant Physiol. Biochem. PPB / Socieťé Fr. Physiol. vegétale 48, 98–107. doi:10.1016/j.plaphy.2010.01.002

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., Giraudat, J., 1994. Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. Plant Cell 6, 1567–82. doi:10.1105/tpc.6.11.1567

Park, S., Harada, J.J., 2008. Arabidopsis embryogenesis. Methods Mol. Biol. 427, 3–16. doi:10.1007/978-1-59745-273-1_1

Park, S.-Y., Fung, P., Nishimura, N., Jensen, D.R., Fujii, H., Zhao, Y., Lumba, S., Santiago, J., Rodrigues, A., Chow, T.-F.F., Alfred, S.E., Bonetta, D., Finkelstein, R., Provart, N.J., Desveaux, D., Rodriguez, P.L., McCourt, P., Zhu, J.-K., Schroeder, J.I., Volkman, B.F., Cutler, S.R., 2009. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science 324, 1068–71. doi:10.1126/science.1173041

Parsell, D.A., Taulien, J., Lindquist, S., 1993. The role of heat-shock proteins in thermotolerance, in: Molecular Chaperones. Springer, pp. 23–30.

Parthibane, V., Iyappan, R., Vijayakumar, A., Venkateshwari, V., Rajasekharan, R., 2012. Serine/threonine/tyrosine protein kinase phosphorylates oleosin, a regulator of lipid metabolic functions. Plant Physiol. 159, 95–104. doi:10.1104/pp.112.197194

Patel, D.J., Wang, Z., 2013. Readout of epigenetic modifications. Annu. Rev. Biochem. 82, 81–118. doi:10.1146/annurev-biochem-072711-165700

Perrella, G., Consiglio, M.F., Aiese-Cigliano, R., Cremona, G., Sanchez-Moran, E., Barra, L., Errico, A., Bressan, R.A., Franklin, F.C.H., Conicella, C., 2010. Histone hyperacetylation affects meiotic recombination and chromosome segregation in Arabidopsis. Plant J. 62, 796–806. doi:10.1111/j.1365-313X.2010.04191.x

Petesch, S.J., Lis, J.T., 2008. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. Cell 134, 74–84. doi:10.1016/j.cell.2008.05.029

Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F., Mantovani, R., 2012. The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. Plant Cell 24, 4777–92. doi:10.1105/tpc.112.105734

Pettersson, E., Lundeberg, J., Ahmadian, A., 2009. Generations of sequencing technologies. Genomics 93, 105–11. doi:10.1016/j.ygeno.2008.10.003

Phillips, J.R., Fischer, E., Baron, M., van den Dries, N., Facchinelli, F., Kutzer, M., Rahmanzadeh, R., Remus, D., Bartels, D., 2008. Lindernia brevidens: a novel desiccation-tolerant vascular plant, endemic to ancient tropical rainforests. Plant J. 54, 938–48. doi:10.1111/j.1365-313X.2008.03478.x

Piatkowski, D., Schneider, K., Salamini, F., Bartels, D., 1990. Characterization of Five Abscisic Acid-Responsive cDNA Clones Isolated from the Desiccation-Tolerant Plant Craterostigma plantagineum and Their Relationship to Other Water-Stress Genes. PLANT Physiol. 94, 1682–1688. doi:10.1104/pp.94.4.1682

Pien, S., Grossniklaus, U., 2007. Polycomb group and trithorax group proteins in Arabidopsis. Biochim. Biophys. Acta - Gene Struct. Expr. 1769, 375–382. doi:10.1016/j.bbaexp.2007.01.010

Pikaard, C.S., Scheid, O.M., 2013. Epigenetic regulation in plants. Cold Spring Harb. Perspect. Biol 6, a019315.

Piontkivska, H., Rooney, A.P., Nei, M., 2002. Purifying Selection and Birth-and-death Evolution in the Histone H4 Gene Family. Mol. Biol. Evol. 19, 689–697. doi:10.1093/oxfordjournals.molbev.a004127

Piskurewicz, U., Jikumaru, Y., Kinoshita, N., Nambara, E., Kamiya, Y., Lopez-Molina, L., 2008. The gibberellic acid signaling repressor RGL2 inhibits Arabidopsis seed germination by stimulating abscisic acid synthesis and ABI5 activity. Plant Cell 20, 2729–45. doi:10.1105/tpc.108.061515

Proctor, M.C.F., Oliver, M.J., Wood, A.J., Alpert, P., Stark, L.R., Cleavitt, N.L., Mishler, B.D., 2007. Desiccation-tolerance in bryophytes: a review. Bryologist 110, 595–621. doi:10.1639/0007-2745(2007)110[595:DIBAR]2.0.CO;2

Putterill, J., Robson, F., Lee, K., Simon, R., Coupland, G., 1995. The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80, 847–857. doi:10.1016/0092-8674(95)90288-0

Qin, F., Kakimoto, M., Sakuma, Y., Maruyama, K., Osakabe, Y., Tran, L.-S.P., Shinozaki, K., Yamaguchi-Shinozaki, K., 2007. Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in Zea mays L. Plant J. 50, 54–69. doi:10.1111/j.1365-313X.2007.03034.x

Qureshi, M.K., Sujeeth, N., Gechev, T.S., Hille, J., 2013. The zinc finger protein ZAT11 modulates paraquat-induced programmed cell death in Arabidopsis thaliana. Acta Physiol. Plant. 35, 1863–1871. doi:10.1007/s11738-013-1224-y

R Development Core Team, 2015. R: A Language and Environment for Statistical Computing.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., Job, D., 2004. The effect of alpha-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiol. 134, 1598–613. doi:10.1104/pp.103.036293

Rakić, T., Lazarević, M., Jovanović, Z.S., Radović, S., Siljak-Yakovlev, S., Stevanović, B., Stevanović, V., 2014. Resurrection plants of the genus Ramonda: prospective survival strategies - unlock further capacity of adaptation, or embark on the path of evolution? Front. Plant Sci. 4, 550. doi:10.3389/fpls.2013.00550

Raz, V., Bergervoet, J.H.W., Koornneef, M., 2001. Sequential steps for developmental arrest in Arabidopsis seeds. Development 252, 243–252.

Reddy, P.S., Kavi Kishor, P.B., Seiler, C., Kuhlmann, M., Eschen-Lippold, L., Lee, J., Reddy, M.K., Sreenivasulu, N., 2014. Unraveling regulation of the small heat shock proteins by the heat shock factor HvHsfB2c in barley: its implications in drought stress response and seed development. PLoS One 9, e89125. doi:10.1371/journal.pone.0089125

Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, andreas, Tewes, A., Ezcurra, I., Rask, L., Baumlein, H., 2000. Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. Plant J. 21, 401–408. doi:10.1046/j.1365-313x.2000.00686.x

Richard, A.C., Lyons, P.A., Peters, J.E., Biasci, D., Flint, S.M., Lee, J.C., McKinney, E.F., Siegel, R.M., Smith, K.G., Allison, D., Cui, X., Page, G., Sabripour, M., Loewe, R., Nelson, P., Hoffmann, R., Seidl, T., Dugas, M., Bolstad, B., Irizarry, R., Astrand, M., Speed, T., Cope, L., Irizarry, R., Jaffee, H., Wu, Z., Speed, T., Irizarry, R., Wu, Z., Jaffee, H., Shippy, R., Fulmer-Smentek, S., Jensen, R., Jones, W., Wolber, P., Johnson, C., Pine, P., Boysen, C., Guo, X., Chudin, E., Sun, Y., Willey, J., Thierry-Mieg, J., Thierry-Mieg, D., Setterquist, R., Wilson, M., Lucas, A., Novoradovskaya, N., Papallo, A., Turpaz, Y., Baker, S., Warrington, J., Shi, L., Herman, D., Irizarry, R., Bolstad, B., Collin, F., Cope, L., Hobbs, B., Speed, T., Leek, J., Scharpf, R., Bravo, H., Simcha, D., Langmead, B., Johnson, W., Geman, D., Baggerly, K., Irizarry, R., Lazar, C., Meganck, S., Taminau, J., Steenhoff, D., Coletta, A., Molter, C., Weiss-Solis, D., Duque, R., Bersini, H., Nowe, A., Chen, C., Grennan, K., Badner, J., Zhang, D., Gershon, E., Jin, L., Liu, C., Luo, J., Schumacher, M., Scherer, A., Sanoudou, D., Megherbi, D., Davison, T., Shi, T., Tong, W., Shi, L., Hong, H., Zhao, C., Elloumi, F., Shi, W., Thomas, R., Lin, S., Tillinghast, G., Liu, G., Zhou, Y., Herman, D., Li, Y., Deng, Y., Fang, H., Bushel, P., Woods, M., Zhang, J., Johnson, W., Li, C., Rabinovic, A., Wong, M., Medrano, J., Flagella, M., Bui, S., Zheng, Z., Nguyen, C., Zhang, A., Pastor, L., Ma, Y., Yang, W., Crawford, K., McMaster, G., Witney, F., Luo, Y., Mieczkowski, J., Tyburczy, M., Dabrowski, M., Pokarowski, P., McCall, M., Irizarry, R., Millenaar, F., Okyere, J., May, S., Zanten, M. van, Voesenek, L., Peeters, A., Jiang, N., Leach, L., Hu, X., Potokina, E., Jia, T., Druka, A., Waugh, R., Kearsey, M., Luo, Z., Seo, J., Hoffman, E., Canales, R., Luo, Y., Willey, J., Austermiller, B., Barbacioru, C., Boysen, C., Hunkapiller, K., Jensen, R., Knight, C., Lee, K., Ma, Y., Maqsodi, B., Papallo, A., Peters, E., Poulter, K., Ruppel, P., Samaha, R., Shi, L., Yang, W., Zhang, L., Goodsaid, F., Larkin, J., Frank, B., Gavras, H., Sultana, R., Quackenbush, J., Yuen, T., Wurmbach, E., Pfeffer, R., Ebersole, B., Sealfon, S., Geiss, G., Bumgarner, R., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D., Fell, H., Ferree, S., George, R., Grogan, T., James, J., Maysuria, M., Mitton, J., Oliveri, P., Osborn, J., Peng, T., Ratcliffe, A., Webster, P., Davidson, E., Hood, L., Dimitrov, K., Bustin, S., Nolan, T., Prokopec, S., Watson, J., Waggott, D., Smith, A., Wu, A., Okey, A., Pohjanvirta, R., Boutros, P., Lee, J., Lyons, P., McKinney, E., Sowerby, J., Carr, E., Bredin, F., Rickman, H., Ratlamwala, H., Hatton, A., Rayner, T., Parkes, M., Smith, K., Lyons, P., Koukoulaki, M., Hatton, A., Doggett, K., Woffendin, H., Chaudhry, A., Smith, K., McKinney, E., Lyons, P., Carr, E., Hollis, J., Jayne, D., Willcocks, L., Koukoulaki, M., Brazma, A., Jovanovic, V., Kemeny, D., Pollard, A., Macary, P., Chaudhry, A., Smith, K., Carvalho, B., Irizarry, R., Kauffmann, A., Gentleman, R., Huber, W., Reis, P., Waldron, L., Goswami, R., Xu, W., Xuan, Y., Perez-Ordonez, B., Gullane, P., Irish, J., Jurisica, I., Kamel-Reid, S., Zilliox, M., Irizarry, R., McCall, M., Uppal, K., Jaffee, H., Zilliox, M., Irizarry, R., Vandesompele, J., Preter, K. De, Pattyn, F., Poppe, B., Roy, N. Van, Paepe, A. De, Speleman, F., Bourgon, R., Gentleman, R., Huber, W., Hackstadt, A., Hess, A., 2014. Comparison of gene expression microarray data with count-based RNA measurements informs microarray interpretation. BMC Genomics 15, 649. doi:10.1186/1471-2164-15-649

Rinne, P.L.H., Kaikuranta, P.L.M., van der Plas, L.H.W., van der Schoot, C., 1999. Dehydrins in cold-acclimated apices of birch (Betula pubescens Ehrh.): production, localization and potential role in rescuing enzyme function during dehydration. Planta 209, 377–388. doi:10.1007/s004250050740

Rismani-Yazdi, H., Haznedaroglu, B.Z., Reeves, D., Peccia, J., 2012. Optimization of de novo

transcriptome assembly from high-throughput short read sequencing data improves functional annotation for non-model organisms.

Rivera, C., Saavedra, F., Alvarez, F., Díaz-Celis, C., Ugalde, V., Li, J., Forné, I., Gurard-Levin, Z.A., Almouzni, G., Imhof, A., Loyola, A., 2015. Methylation of histone H3 lysine 9 occurs during translation. Nucleic Acids Res. 43, 9097–106. doi:10.1093/nar/gkv929

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.-L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J.M., Hoodless, P.A., Birol, I., 2010. De novo assembly and analysis of RNA-seq data. Nat. Methods 7, 909–12. doi:10.1038/nmeth.1517

Robinson, S.A., Wasley, J., Popp, M., Lovelock, C.E., 2000. Desiccation tolerance of three moss species from continental Antarctica. Funct. Plant Biol. 27, 379–388. doi:10.1071/PP99133

Rodriguez, M.C.S., Edsgärd, D., Hussain, S.S., Alquezar, D., Rasmussen, M., Gilbert, T., Nielsen, B.H., Bartels, D., Mundy, J., 2010. Transcriptomes of the desiccation-tolerant resurrection plant Craterostigma plantagineum. Plant J. 63, 212–28. doi:10.1111/j.1365-313X.2010.04243.x

Rolland, F., Moore, B., Sheen, J., 2002. Sugar Sensing and Signaling in Plants. Plant Cell 14, S185–205. doi:10.1105/tpc.010455

Romanel, E.A.C., Schrago, C.G., Couñago, R.M., Russo, C.A.M., Alves-Ferreira, M., 2009. Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. PLoS One 4, e5791. doi:10.1371/journal.pone.0005791

Sakamoto, H., 2004. Arabidopsis Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. PLANT Physiol. 136, 2734–2746. doi:10.1104/pp.104.046599

Sakata, Y., Chiba, Y., Fukushima, H., Matsubara, N., Habu, Y., Naito, S., Ohno, T., 1997. The RY sequence is necessary but not sufficient for the transcription activation of a winged bean chymotrypsin inhibitor gene in developing seeds. Plant Mol. Biol. 34, 191–197. doi:10.1023/A:1005841125832

Sakata, Y., Nakamura, I., Taji, T., Tanaka, S., Quatrano, R.S., 2010. Regulation of the ABA-responsive Em promoter by ABI3 in the moss Physcomitrella patens. Plant Signal. Behav. 5, 1061–1066. doi:10.4161/psb.5.9.11774

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki, K., 2006a. Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. Plant Cell 18, 1292–309. doi:10.1105/tpc.105.035881

Sakuma, Y., Maruyama, K., Qin, F., Osakabe, Y., Shinozaki, K., Yamaguchi-Shinozaki, K., 2006b. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. Proc. Natl. Acad. Sci. U. S. A. 103, 18822–7. doi:10.1073/pnas.0605639103

Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.-Y., Jamin, M., Cutler, S.R.,

Rodriguez, P.L., Márquez, J.A., 2009. The abscisic acid receptor PYR1 in complex with abscisic acid. Nature 462, 665–8. doi:10.1038/nature08591

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., Lepiniec, L., 2008. Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. Plant J. 54, 608–20. doi:10.1111/j.1365-313X.2008.03461.x

Saranga, Y., Rhodes, D., Janick, J., 1990. Factors inducing desiccation tolerance in celery somatic embryos. HortScience 25, 1088.

Sarkar, A., 2009. Plant Stem Cells. Discovery Publishing House.

Scharf, K.-D., Berberich, T., Ebersberger, I., Nover, L., 2012. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. Biochim. Biophys. Acta 1819, 104–19. doi:10.1016/j.bbagrm.2011.10.002

Schatz, M.C., Delcher, A.L., Salzberg, S.L., 2010. Assembly of large genomes using second-generation sequencing. Genome Res. 20, 1165–73. doi:10.1101/gr.101360.109

Schlebusch, S., Illing, N., 2012. Next generation shotgun sequencing and the challenges of de novo genome assembly: review article. S. Afr. J. Sci. 108, 1–8.

Schramm, F., Ganguli, A., Kiehlmann, E., Englich, G., Walch, D., von Koskull-Döring, P., 2006. The heat stress transcription factor HsfA2 serves as a regulatory amplifier of a subset of genes in the heat stress response in Arabidopsis. Plant Mol. Biol. 60, 759–72. doi:10.1007/s11103-005-5750-x

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., von Koskull-Döring, P., 2008. A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. Plant J. 53, 264–74. doi:10.1111/j.1365-313X.2007.03334.x

Schubert, D., Clarenz, O., Goodrich, J., 2005. Epigenetic control of plant development by Polycomb-group proteins. Curr. Opin. Plant Biol. 8, 553–61. doi:10.1016/j.pbi.2005.07.005

Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–92. doi:10.1093/bioinformatics/bts094

Schwartz, Y.B., Pirrotta, V., 2008. Polycomb complexes and epigenetic states. Curr. Opin. Cell Biol. 20, 266–273. doi:10.1016/j.ceb.2008.03.002

Scippa, G.S., Di Michele, M., Onelli, E., Patrignani, G., Chiatante, D., Bray, E.A., 2004. The histone-like protein H1-S and the response of tomato leaves to water deficit. J. Exp. Bot. 55, 99–109. doi:10.1093/jxb/erh022

Scippa, G.S., Griffiths, A., Chiatante, D., Bray, E.A., 2000. The H1 histone variant of tomato, H1-S, is targeted to the nucleus and accumulates in chromatin in response to water-deficit stress. Planta 211, 173–181. doi:10.1007/s004250000278

Scott, H.B., Oliver, M.J., 1994. Accumulation and polysomal recruitment of transcripts in response to desiccation and rehydration of the moss Tortula ruralis. J. Exp. Bot. 45, 577–583. doi:10.1093/jxb/45.5.577

Senaranta, T., McKersie, B.D., Bowley, S.R., 1989. Desiccation tolerance of alfalfa (Medicago

sativa L.) somatic embroys. Influence of abscisic acid, stress pretreatments and drying rates. Plant Sci. 65, 253–259.

Seo, M., Aoki, H., Koiwai, H., Kamiya, Y., Nambara, E., Koshiba, T., 2004. Comparative studies on the Arabidopsis aldehyde oxidase (AAO) gene family revealed a major role of AAO3 in ABA biosynthesis in seeds. Plant Cell Physiol. 45, 1694–703. doi:10.1093/pcp/pch198

Servet, C., Conde e Silva, N., Zhou, D.-X., 2010. Histone acetyltransferase AtGCN5/HAG1 is a versatile regulator of developmental and inducible gene expression in Arabidopsis. Mol. Plant 3, 670–7. doi:10.1093/mp/ssq018

Shahbazian, M.D., Grunstein, M., 2007. Functions of site-specific histone acetylation and deacetylation. Annu. Rev. Biochem. 76, 75–100. doi:10.1146/annurev.biochem.76.052705.162114

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–504. doi:10.1101/gr.1239303

Shen, A.Y.-H., 2014. The transcriptome response of leaves of the resurrection plant, Xerophyta humilis to desiccation.

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

Shinozaki, K., Yamaguchi-Shinozaki, K., 2006. Gene networks involved in drought stress response and tolerance. J. Exp. Bot. 58, 221–227. doi:10.1093/jxb/erl164

Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E., Tayrose, G., Holt, B.F., 2009. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant Physiol. 149, 625–41. doi:10.1104/pp.108.130591

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117–23. doi:10.1101/gr.089532.108

Skirycz, A., De Bodt, S., Obata, T., De Clercq, I., Claeys, H., De Rycke, R., Andriankaja, M., Van Aken, O., Van Breusegem, F., Fernie, A.R., Inzé, D., 2010. Developmental stage specificity and the role of mitochondrial metabolism in the response of Arabidopsis leaves to prolonged mild osmotic stress. Plant Physiol. 152, 226–44. doi:10.1104/pp.109.148965

Skirycz, A., Inzé, D., 2010. More from less: plant growth under limited water. Curr. Opin. Biotechnol. 21, 197–203. doi:10.1016/j.copbio.2010.03.002

Smirnoff, N., 1992. The carbohydrates of bryophytes in relation to desiccation tolerance. J. Bryol. 17, 185–191. doi:10.1179/jbr.1992.17.2.185

Smythe, B., 1967. Sucrose crystal growth. II. Rate of crystal growth in the presence of impurities. Aust. J. Chem. 20, 1097. doi:10.1071/CH9671097

Solymosi, K., Tuba, Z., Böddi, B., 2013. Desiccoplast-etioplast-chloroplast transformation

under rehydration of desiccated poikilochlorophyllous Xerophyta humilis leaves in the dark and upon subsequent illumination. J. Plant Physiol. 170, 583–90. doi:10.1016/j.jplph.2012.11.022

Soppe, W.J.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., Fransz, P.F., 2002. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J. 21, 6549–6559.

Sreenivasulu, N., Wobus, U., 2013. Seed-development programs: a systems biology-based comparison between dicots and monocots. Annu. Rev. Plant Biol. 64, 189–217. doi:10.1146/annurev-arplant-050312-120215

Stacy, R.A.P., Aalen, R.B., 1998. Identification of sequence homology between the internal hydrophilic repeated motifs of Group 1 late-embryogenesis-abundant proteinsin plants and hydrophilic repeats of the general stress protein GsiB of Bacillus subtilis. Planta 206, 476–478. doi:10.1007/s004250050424

Stark, L.R., Greenwood, J.L., Brinda, J.C., Oliver, M.J., 2013. The desert moss Pterygoneurum lamellatum (Pottiaceae) exhibits an inducible ecological strategy of desiccation tolerance: effects of rate of drying on shoot damage and regeneration. Am. J. Bot. 100, 1522–31. doi:10.3732/ajb.1200648

Steadman, K.J., Pritchard, H.W., Dey, P.M., 1996. Tissue-specific Soluble Sugars in Seeds as Indicators of Storage Category. Ann. Bot. 77, 667–674. doi:10.1093/aob/77.6.667

Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.-F., Fischer, R.L., Goldberg, R.B., Harada, J.J., 2008. Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. Proc. Natl. Acad. Sci. U. S. A. 105, 3151–6. doi:10.1073/pnas.0712364105

Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L., Goldberg, R.B., Harada, J.J., 2001. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proc. Natl. Acad. Sci. U. S. A. 98, 11806–11. doi:10.1073/pnas.201413498

Strahl, B.D., Allis, C.D., 2000. The language of covalent histone modifications. Nature 403, 41–5. doi:10.1038/47412

Su, Z., Ning, B., Fang, H., Hong, H., Perkins, R., Tong, W., Shi, L., 2011. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev. Mol. Diagn. 11, 333–43. doi:10.1586/erm.11.3

Sun, L., Wang, Y.-P., Chen, P., Ren, J., Ji, K., Li, Q., Li, P., Dai, S.-J., Leng, P., 2011. Transcriptional regulation of SlPYL, SlPP2C, and SlSnRK2 gene families encoding ABA signal core components during tomato fruit development and drought stress. J. Exp. Bot. 62, 5659–69. doi:10.1093/jxb/err252

Sun, Q., Zhou, D.-X., 2008. Rice jmjC domain-containing gene JMJ706 encodes H3K9 demethylase required for floral organ development. Proc. Natl. Acad. Sci. U. S. A. 105, 13679–84. doi:10.1073/pnas.0805901105

Sun, T.-P., Gubler, F., 2004. Molecular mechanism of gibberellin signaling in plants. Annu. Rev. Plant Biol. 55, 197–223. doi:10.1146/annurev.arplant.55.031903.141753

Suzuki, M., Akihama, T., Ishikawa, M., 2005. Cryopreservation of encapsulated gentian

axillary buds following 2 step-preculture with sucrose and desiccation. Plant Cell. Tissue Organ Cult. 83, 115–121. doi:10.1007/s11240-005-4854-y

Suzuki, M., Ishikawa, M., Akihama, T., 1998. A novel preculture method for the induction of desiccation tolerance in gentian axillary buds for cryopreservation. Plant Sci. 135, 69–76. doi:10.1016/S0168-9452(98)00054-5

Suzuki, M., Ishikawa, M., Okuda, H., Noda, K., Kishimoto, T., Nakamura, T., Ogiwara, I., Shimura, I., Akihama, T., 2006. Physiological changes in gentian axillary buds during two-step preculturing with sucrose that conferred high levels of tolerance to desiccation and cryopreservation. Ann. Bot. 97, 1073–81. doi:10.1093/aob/mcl054

Suzuki, M., Kao, C.Y., McCarty, D.R., 1997. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. Plant Cell 9, 799–807. doi:10.1105/tpc.9.5.799

Suzuki, M., Mccarty, D.R., 2008. Functional symmetry of the B3 network controlling seed development. Curr. Opin. Plant Biol. 548–553. doi:10.1016/j.pbi.2008.06.015

Swindell, W.R., Huebner, M., Weber, A.P., 2007. Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC Genomics 8, 125. doi:10.1186/1471-2164-8-125

Szenker, E., Ray-Gallet, D., Almouzni, G., 2011. The double face of the histone variant H3.3. Cell Res. 21, 421–434. doi:10.1038/cr.2011.14

Tamaru, H., 2010. Confining euchromatin/heterochromatin territory: jumonji crosses the line. Genes Dev. 24, 1465–78. doi:10.1101/gad.1941010

Tan, B.-C., Joseph, L.M., Deng, W.-T., Liu, L., Li, Q.-B., Cline, K., McCarty, D.R., 2003. Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. Plant J. 35, 44–56.

Tanaka, M., Kikuchi, A., Kamada, H., 2008. The Arabidopsis histone deacetylases HDA6 and HDA19 contribute to the repression of embryonic properties after germination. Plant Physiol. 146, 149–61. doi:10.1104/pp.107.111674

Teng, S., Keurentjes, J., Bentsink, L., Koornneef, M., Smeekens, S., 2005. Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the MYB75/PAP1 gene. Plant Physiol. 139, 1840–1852.

Terrasson, E., Buitink, J., Righetti, K., Ly Vu, B., Pelletier, S., Zinsmeister, J., Lalanne, D., Leprince, O., 2013. An emerging picture of the seed desiccome: confirmed regulators and newcomers identified using transcriptome comparison. Front. Plant Sci. 4, 497. doi:10.3389/fpls.2013.00497

Thomas, S.G., Sun, T., 2004. Update on gibberellin signaling. A tale of the tall and the short. Plant Physiol. 135, 668–76. doi:10.1104/pp.104.040279

Thompson, A.J., Jackson, A.C., Parker, R.A., Morpeth, D.R., Burbidge, A., Taylor, I.B., 2000. Abscisic acid biosynthesis in tomato: regulation of zeaxanthin epoxidase and 9-cis-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid. Plant Mol. Biol. 42, 833–845. doi:10.1023/A:1006448428401

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., 2006. A Network of Local and Redundant Gene Regulation Governs Arabidopsis Seed Maturation. Society 18, 1642–

1651. doi:10.1105/tpc.105.039925.served

Toldi, O., Tuba, Z., Scott, P., 2009. Vegetative desiccation tolerance: Is it a goldmine for bioengineering crops? Plant Sci. 176, 187–199. doi:10.1016/j.plantsci.2008.10.002

Tuba, Z., 2008. Notes on the poikilochlorophyllous desiccation-tolerant plants. Acta Biol. Szeged. 52, 111–113.

Tuba, Z., Lichtenthaler, H.K., Csintalan, Z., Nagy, Z., Szente, K., 1996. Loss of chlorophylls, cessation of photosynthetic CO2 assimilation and respiration in the poikilochlorophyllous plant Xerophyta scabrida during desiccation. Physiol. Plant. 96, 383–388. doi:10.1111/j.1399-3054.1996.tb00448.x

Tuba, Z., Lichtenthaler, H.K., Maroti, I., Csintalan, Z., 1993. Resynthesis of Thylakoids and Functional Chloroplasts in the Desiccated Leaves of the Poikilochlorophyllous Plant Xerophyta scabrida upon Rehydration. J. Plant Physiol. 142, 742–748. doi:10.1016/S0176-1617(11)80912-0

Tuba, Z., Protor, C.F., Csintalan, Z., 1998. Ecophysiological responses of homoiochlorophyllous and poikilochlorophyllous desiccation tolerant plants: a comparison and an ecological perspective. Plant Growth Regul. 24, 211–217. doi:10.1023/A:1005951908229

Tyler, L., Thomas, S.G., Hu, J., Dill, A., Alonso, J.M., Ecker, J.R., Sun, T.-P., 2004. Della proteins and gibberellin-regulated seed germination and floral development in Arabidopsis. Plant Physiol. 135, 1008–19. doi:10.1104/pp.104.039578

Ueguchi-Tanaka, M., Nakajima, M., Katoh, E., Ohmiya, H., Asano, K., Saji, S., Hongyu, X., Ashikari, M., Kitano, H., Yamaguchi, I., Matsuoka, M., 2007. Molecular interactions of a soluble gibberellin receptor, GID1, with a rice DELLA protein, SLR1, and gibberellin. Plant Cell 19, 2140–55. doi:10.1105/tpc.106.043729

Umezawa, T., Sugiyama, N., Anderson, J.C., Takahashi, F., Ishihama, Y., Peck, S.C., Shinozaki, K., 2013. Protein Phosphorylation Network in Abscisic Acid Signaling, in: Plant and Microbe Adaptations to Cold in a Changing World. Springer, pp. 155–164.

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., Ishihama, Y., Hirayama, T., Shinozaki, K., 2009. Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 106, 17588–93. doi:10.1073/pnas.0907095106

Umezawa, T., Yoshida, R., Maruyama, K., Yamaguchi-Shinozaki, K., Shinozaki, K., 2004. SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 101, 17306–11. doi:10.1073/pnas.0407758101

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K., Yamaguchi-Shinozaki, K., 2000. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc. Natl. Acad. Sci. U. S. A. 97, 11632–7. doi:10.1073/pnas.190309197

Van Buren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E., Freeling, M., Bartels, D., Ten Hallers, B., Hastie, A., Michael, T.P., Mockler, T.C., 2015. Single-molecule sequencing of the desiccation-tolerant grass

Oropetium thomaeum. Nature 527, 508–511. doi:10.1038/nature15714

van Zanten, M., Carles, A., Li, Y., Soppe, W.J.J., 2012. Control and consequences of chromatin compaction during seed maturation in Arabidopsis thaliana. Plant Signal. Behav. 7, 338–41. doi:10.4161/psb.19281

van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., Soppe, W.J.J., 2011. Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proc. Natl. Acad. Sci. U. S. A. 108, 20219–24. doi:10.1073/pnas.1117726108

Verslues, P.E., Juenger, T.E., 2011. Drought, metabolites, and Arabidopsis natural variation: a promising combination for understanding adaptation to water-limited environments. Curr. Opin. Plant Biol. 14, 240–245. doi:10.1016/j.pbi.2011.04.006

Vicente-Carbajosa, J., Carbonero, P., 2005. Seed maturation: developing an intrusive phase to accomplish a quiescent state. Int. J. Dev. Biol. 49, 645–651. doi:10.1387/ijdb.052046jc

Vicré, M., Farrant, J.M., Driouich, A., 2004. Insights into the cellular mechanisms of desiccation tolerance among angiosperm resurrection plant species. Plant. Cell Environ. 27, 1329–1340.

Vieira, C.V., Amaral da Silva, E.A., Alvarenga, A.A., Castro, E.M., Toorop, P.E., 2010. Stress-associated factors increase after desiccation of germinated seeds of Tabebuia impetiginosa Mart. Plant Growth Regul. 62, 257–263. doi:10.1007/s10725-010-9496-3

Vierling, E., 1991. The Roles of Heat Shock Proteins in Plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42, 579–620. doi:10.1146/annurev.pp.42.060191.003051

Vlad, F., Rubio, S., Rodrigues, A., Sirichandra, C., Belin, C., Robert, N., Leung, J., Rodriguez, P.L., Laurière, C., Merlot, S., 2009. Protein phosphatases 2C regulate the activation of the Snf1-related kinase OST1 by abscisic acid in Arabidopsis. Plant Cell 21, 3170–84. doi:10.1105/tpc.109.069179

Voelkerding, K. V, Dames, S.A., Durtschi, J.D., 2009. Next-generation sequencing: from basic research to diagnostics. Clin. Chem. 55, 641–58. doi:10.1373/clinchem.2008.112789

von Bubnoff, A., 2008. Next-generation sequencing: the race is on. Cell 132, 721–3. doi:10.1016/j.cell.2008.02.028

von Koskull-Döring, P., Scharf, K.-D., Nover, L., 2007. The diversity of plant heat stress transcription factors. Trends Plant Sci. 12, 452–7. doi:10.1016/j.tplants.2007.08.014

Walley, J.W., Coughlan, S., Hudson, M.E., Covington, M.F., Kaspi, R., Banu, G., Harmer, S.L., Dehesh, K., 2007. Mechanical Stress Induces Biotic and Abiotic Stress Responses via a Novel cis-Element. PLoS Genet. 3, e172. doi:10.1371/journal.pgen.0030172

Walters, C., Farrant, J.M., Pammenter, N.W., Berjak, P., 2002. Desiccation Stress and Damage, in: Desiccation and Survival in Plants: Drying without Dying. p. 263.

Wang, F., Perry, S.E., 2013. Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol. 161, 1251–64. doi:10.1104/pp.112.212282

Wang, T.-W., Balsamo, R.A., Ratnayake, C., Platt, K.A., Ting, J.T.L., Huang, A.H.C., 1997. Identification, subcellular localization, and developmental studies of oleosins in the anther of Brassica napus. Plant J. 11, 475–487. doi:10.1046/j.1365-313X.1997.11030475.x

Wang, X.-S., Zhu, H.-B., Jin, G.-L., Liu, H.-L., Wu, W.-R., Zhu, J., 2007. Genome-scale identification and analysis of LEA genes in rice (Oryza sativa L.). Plant Sci. 172, 414–420. doi:10.1016/j.plantsci.2006.10.004

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. doi:10.1038/nrg2484

Waters, E.R., 2013. The evolution, function, structure, and expression of the plant sHSPs. J. Exp. Bot. 64, 391–403. doi:10.1093/jxb/ers355

Webb, M.A., Arnott, H.J., 1982. Cell Wall Conformation in Dry Seeds in Relation to the Preservation of Structural Integrity During Desiccation. Am. J. Bot. 69, pp. 1657–1668.

Weber, H., Borisjuk, L., Wobus, U., 1997. Sugar import and metabolism during seed development. Trends Plant Sci. 2, 169–174. doi:10.1016/S1360-1385(97)85222-3

Wehmeyer, N., 1996. Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. PLANT Physiol. 112, 747–757. doi:10.1104/pp.112.2.747

Werner, O., Ros Espín, R.M., Bopp, M., Atzorn, R., 1991. Abscisic-acid-induced drought tolerance in Funaria hygrometrica Hedw. Planta 186, 99–103. doi:10.1007/BF00201503

Wierzbicki, A.T., 2005. Suppression of Histone H1 Genes in Arabidopsis Results in Heritable Developmental Defects and Stochastic Changes in DNA Methylation. Genetics 169, 997–1008. doi:10.1534/genetics.104.031997

Wilson, K., Long, D., Swinburne, J., Coupland, G., 1996. A Dissociation insertion causes a semidominant mutation that increases expression of TINY, an Arabidopsis gene related to APETALA2. Plant Cell 8, 659–71. doi:10.1105/tpc.8.4.659

Wise, M.J., Tunnacliffe, A., 2004. POPP the question: what do LEA proteins do? Trends Plant Sci. 9, 13–7. doi:10.1016/j.tplants.2003.10.012

Wobus, U., Weber, H., 1999a. Sugars as signal molecules in plant seed development. Biol. Chem. 380, 937–944.

Wobus, U., Weber, H., 1999b. Seed maturation: genetic programmes and control signals. Curr. Opin. Plant Biol. 2, 33–38. doi:10.1016/S1369-5266(99)80007-7

Wolkers, W.F., McCready, S., Brandt, W.F., Lindsey, G.G., Hoekstra, F.A., 2001. Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1544, 196–206. doi:10.1016/S0167-4838(00)00220-X

Wollmann, H., Berger, F., 2015. Epigenetic Modifications at Developmental Transitions in Arabidopsis, in: Nuclear Functions in Plant Transcription, Signaling and Development. Springer, pp. 119–131.

Wood, A.J., Oliver, M.J., 1999. Translational control in plant stress: the formation of messenger ribonucleoprotein particles (mRNPs) in response to desiccation of Tortula ruralis gametophytes. Plant J. 18, 359–370. doi:10.1046/j.1365-313X.1999.00458.x

Wu, C.-Y., Suzuki, A., Washida, H., Takaiwa, F., 1998. The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. Plant J. 14, 673–683. doi:10.1046/j.1365-313x.1998.00167.x

Xiao, L., Yang, G., Zhang, L., Yang, X., Zhao, S., Ji, Z., Zhou, Q., Hu, M., Wang, Y., Chen, M., Xu, Y., Jin, H., Xiao, X., Hu, G., Bao, F., Hu, Y., Wan, P., Li, L., Deng, X., Kuang, T., Xiang, C., Zhu, J.-K., Oliver, M.J., He, Y., 2015. The resurrection genome of Boea hygrometrica: A blueprint for survival of dehydration. Proc. Natl. Acad. Sci. U. S. A. 112, 5833–7. doi:10.1073/pnas.1505811112

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G.K.-S., Wang, J., 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660–6. doi:10.1093/bioinformatics/btu077

Xie, Z., Li, X., Glover, B.J., Bai, S., Rao, G.-Y., Luo, J., Yang, J., 2008. Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. Mol. Biol. Evol. 25, 1581–92. doi:10.1093/molbev/msn105

Xiong, L., Lee, H., Ishitani, M., Zhu, J.-K., 2002. Regulation of osmotic stress-responsive gene expression by the LOS6/ABA1 locus in Arabidopsis. J. Biol. Chem. 277, 8588–96. doi:10.1074/jbc.M109275200

Xiong, L., Zhu, J.-K., 2003. Regulation of abscisic acid biosynthesis. Plant Physiol. 133, 29–36.

Xu, D., Li, J., Gangappa, S.N., Hettiarachchi, C., Lin, F., Andersson, M.X., Jiang, Y., Deng, X.W., Holm, M., 2014. Convergence of Light and ABA signaling on the ABI5 promoter. PLoS Genet. 10, e1004197. doi:10.1371/journal.pgen.1004197

Xu, F., Tan, X., Wang, Z., 2010. Effects of Sucrose on Germination and Seedling Development of Brassica Napus. Int. J. Biol. 2, p150. doi:10.5539/ijb.v2n1p150

Xu, Z.-Y., Kim, S.Y., Hyeon, D.Y., Kim, D.H., Dong, T., Park, Y., Jin, J.B., Joo, S.-H., Kim, S.-K., Hong, J.C., Hwang, D., Hwang, I., 2013. The Arabidopsis NAC transcription factor ANAC096 cooperates with bZIP-type transcription factors in dehydration and osmotic stress responses. Plant Cell 25, 4708–24. doi:10.1105/tpc.113.119099

Xue, G.-P., Loveridge, C.W., 2004. HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. Plant J. 37, 326–339. doi:10.1046/j.1365-313X.2003.01963.x

Yamaguchi-Shinozaki, K., Shinozaki, K., 1994. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell 6, 251–64. doi:10.1105/tpc.6.2.251

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., Hattori, T., 2009. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. 58, 843–56. doi:10.1111/j.1365-313X.2009.03817.x

Yamamoto, A., Kagaya, Y., Usui, H., Hobo, T., Takeda, S., Hattori, T., 2010. Diverse roles and mechanisms of gene regulation by the Arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. Plant Cell Physiol. 51, 2031–46. doi:10.1093/pcp/pcq162

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., Calonje, M., 2013. VAL- and

AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in Arabidopsis. Curr. Biol. 23, 1324–9. doi:10.1016/j.cub.2013.05.050

Yoshida, H., Ueguchi-Tanaka, M., 2014. DELLA and SCL3 balance gibberellin feedback regulation by utilizing INDETERMINATE DOMAIN proteins as transcriptional scaffolds. Plant Signal. Behav. 9, e29726. doi:10.4161/psb.29726

Yoshida, T., Fujita, Y., Maruyama, K., Mogami, J., Todaka, D., Shinozaki, K., Yamaguchi-Shinozaki, K., 2015. Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. Plant. Cell Environ. 38, 35–49. doi:10.1111/pce.12351

Yoshida, T., Fujita, Y., Sayama, H., Kidokoro, S., Maruyama, K., Mizoi, J., Shinozaki, K., Yamaguchi-Shinozaki, K., 2010. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant J. 61, 672–85. doi:10.1111/j.1365-313X.2009.04092.x

Yotsui, I., Saruhashi, M., Kawato, T., Taji, T., Hayashi, T., Quatrano, R.S., Sakata, Y., 2013. ABSCISIC ACID INSENSITIVE3 regulates abscisic acid-responsive gene expression with the nuclear factor Y complex through the ACTT-core element in Physcomitrella patens. New Phytol. 199, 101–9. doi:10.1111/nph.12251

Yu, Y., Bu, Z., Shen, W.-H., Dong, A., 2009. An update on histone lysine methylation in plants. Prog. Nat. Sci. 19, 407–413. doi:10.1016/j.pnsc.2008.07.015

Zentner, G.E., Henikoff, S., 2013. Regulation of nucleosome dynamics by histone modifications. Nat. Struct. Mol. Biol. 20, 259–66. doi:10.1038/nsmb.2470

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–9. doi:10.1101/gr.074492.107

Zhang, H., Bishop, B., Ringenberg, W., Muir, W.M., Ogas, J., 2012. The CHD3 remodeler PICKLE associates with genes enriched for trimethylation of histone H3 lysine 27. Plant Physiol. 159, 418–32. doi:10.1104/pp.112.194878

Zhang, H., Jin, J., Tang, L., Zhao, Y., Gu, X., Gao, G., Luo, J., 2011. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res. 39, D1114–7. doi:10.1093/nar/gkq1141

Zhang, H., Rider, S.D., Henderson, J.T., Fountain, M., Chuang, K., Kandachar, V., Simons, A., Edenberg, H.J., Romero-Severson, J., Muir, W.M., Ogas, J., 2008. The CHD3 remodeler PICKLE promotes trimethylation of histone H3 lysine 27. J. Biol. Chem. 283, 22637–48. doi:10.1074/jbc.M802129200

Zhang, K., Sridhar, V. V, Zhu, J., Kapoor, A., Zhu, J.-K., 2007. Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. PLoS One 2, e1210. doi:10.1371/journal.pone.0001210

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V, Pellegrini, M., Goodrich, J., Jacobsen, S.E., 2007. Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol. 5, e129. doi:10.1371/journal.pbio.0050129

Zhang, Z.-L., Ogawa, M., Fleet, C.M., Zentella, R., Hu, J., Heo, J.-O., Lim, J., Kamiya, Y.,

Yamaguchi, S., Sun, T., 2011. Scarecrow-like 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 108, 2160–5. doi:10.1073/pnas.1012232108

Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X., Hao, P., 2011. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12 Suppl 1, S2. doi:10.1186/1471-2105-12-S14-S2

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., Liu, X., 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9, e78644. doi:10.1371/journal.pone.0078644

Zhu, G., Ye, N., Zhang, J., 2009. Glucose-induced delay of seed germination in rice is mediated by the suppression of ABA catabolism rather than an enhancement of ABA biosynthesis. Plant Cell Physiol. 50, 644–51. doi:10.1093/pcp/pcp022

Zhu, Y., Dong, A., Shen, W.-H., 2012. Histone variants and chromatin assembly in plant abiotic stress responses. Biochim. Biophys. Acta - Gene Regul. Mech. 1819, 343–348. doi:10.1016/j.bbagrm.2011.07.012

Appendix

Some of the data pertaining to the desiccation tolerance of *X. viscosa* seedlings as shown in Chapter 2 was included in an earlier publication on which I was first author (Lyall R, Ingle RA, Illing N (2014) *The Window of Desiccation Tolerance Shown by Early-Stage Germinating Seedlings Remains Open in the Resurrection Plant, Xerophyta viscosa.* PLoS ONE 9(3): e93093), which has been reproduced here.

OPEN ACCESS Freely available online

PLOS ONE

The Window of Desiccation Tolerance Shown by Early-Stage Germinating Seedlings Remains Open in the Resurrection Plant, *Xerophyta viscosa*

Rafe Lyall, Robert A. Ingle, Nicola Illing*

Department of Molecular and Cell Biology, University of Cape Town, Rondebosch, South Africa

Abstract

Resurrection plants are renowned for their vegetative desiccation tolerance (DT). While DT in vegetative tissues is rare in angiosperms, it is ubiquitous in mature orthodox seeds. During germination, seedlings gradually lose DT until they pass a point of no return, after which they can no longer survive dehydration. Here we investigate whether seedlings of the resurrection plant *Xerophyta viscosa* ever lose the capacity to establish DT. Seedlings from different stages of germination were dehydrated for 48 hours and assessed for their ability to recover upon rehydration. While a transient decline in the ability of *X. viscosa* seedlings to survive dehydration was observed, at no point during germination was the ability to re-establish DT completely lost in all seedlings. Pre-treatment of seedlings with PEG or sucrose reduced this transient decline, and improved the survival rate at all stages of germination. Additionally, we observed that the trait of poikilochlorophylly (or loss of chlorophyll) observed in adult *X. viscosa* leaves can be induced throughout seedling development. These results suggest that the window of DT seen in germinating orthodox seeds remains open in *X. viscosa* seedlings and that vegetative DT in *Xerophyta* species may have evolved from the ability to retain this program through to adulthood.

Citation: Lyall R, Ingle RA, Illing N (2014) The Window of Desiccation Tolerance Shown by Early-Stage Germinating Seedlings Remains Open in the Resurrection Plant, *Xerophyta viscosa*. PLoS ONE 9(3): e93093. doi:10.1371/journal.pone.0093093

Editor: Bob Goldstein, University of North Carolina at Chapel Hill, United States of America

Received October 4, 2013; **Accepted** March 1, 2014; **Published** March 25, 2014

Funding: This research was supported by incentive funding from the National Research Foundation of South Africa (http://www.nrf.ac.za/index.php), and the University of Cape Town. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Nicola.Illing@uct.ac.za

Introduction

Desiccation tolerance (DT) is the ability of an organism to revive unharmed after almost complete loss of cellular water from its tissues [1]. While vegetative DT is relatively common in mosses and liverworts, it is extremely rare in angiosperms, having been reported from only 135 taxonomically diverse species from 44 genera collectively known as resurrection plants [2]. Though rare in vegetative tissue, DT is common in the reproductive tissues of angiosperms, with the overwhelming majority of species surveyed to date producing DT (orthodox) seed embryos and pollen [2]. Vegetative DT is a polyphyletic trait in the angiosperms, having evolved independently at least thirteen times [2], with some species retaining chlorophyll (homoiochlorophyllous), whilst other species break chlorophyll down (poikilochlorophyllous) during dehydration. It has been argued that vegetative DT in resurrection plants arose through co-option of the genetic network responsible for the acquisition of DT in orthodox seeds [3–5], however the mechanism by which this might have occurred is unknown.

In orthodox seeds, embryonic DT is acquired late in the development process, during seed maturation. Maturation occurs subsequent to embryo and endosperm morphogenesis, and is characterised by embryo growth arrest, reserve accumulation and the acquisition of stress tolerance [6]. In many angiosperm species seed maturation is not required for the generation of a viable embryo: excised, un-matured seeds will still germinate under favourable conditions [7,8]. Rather, the process of maturation prepares the seed for survival in unfavourable conditions outside of the parent plant by inducing a stress-tolerant, quiescent state.

Embryonic DT occurs as a pre-programmed event rather than as a physiological response to water loss [9,10]. The exact mechanisms by which orthodox seeds become DT are not yet fully understood, however a number of factors have been shown to correlate strongly with the ability of a seed to survive desiccation. These include the up-regulation of LEA (late embryogenesis abundant) proteins, and the accumulation of sucrose and several oligosaccharides. The appearance of LEA proteins in maturing seeds coincides with the acquisition of DT, and their levels fall once germination has been initiated and tolerance is lost [11]. LEA proteins are hypothesised to function, at least in part, as a "water replacement" mechanism: as water is lost from the cell, the hydrophobic LEAs form a protective layer around exposed chemical groups, preventing potential damage or loss of conformation [11–13]. Carbohydrates - primarily sucrose, but also oligosaccharides such as raffinose and stachyose - show a marked increase in concentration in embryonic tissues during maturation [14]. In addition to acting as a readily-available energy source for the embryo during germination, these sugars are thought to function in a similar fashion to LEAs by acting as water replacement molecules [12,15].

Mature orthodox seeds can remain quiescent and DT for extremely long periods of time until growth is re-initiated by imbibition. Germination is induced in non-dormant imbibed seeds by gibberellin-regulated gene networks which reduce ABA

concentrations in the seed and silence maturation-specific genes, thereby reversing the maturation-induced embryo quiescence [16]. DT is progressively lost during the germination of orthodox seeds, until the seedlings reach a so-called "point of no return" past which they can no longer survive dehydration [17]. For many species, including *Arabidopsis thaliana*, this point is reached early during the imbibition stage prior to visible protrusion of the radicle through the testa [18,19]. In other species, however, the point of no return can be extended until well after radicle emergence. This phenomenon is often seen in desert plant species, where the ability for a seedling to survive sudden and unexpected drought conditions unharmed would confer a significant selective advantage [20–22].

Exposure to moderate osmotic stress, such as incubation in high molecular weight polyethylene glycol (PEG), can extend the developmental window during which DT can be re-established in germinating seedlings of many species [23], albeit narrowly in many cases. For example, in *Medicago truncatula* DT is rapidly lost following germination; only 20% of seedlings with radicles >2 mm and 0% with >3 mm in length were able to re-establish DT [24]. However, following pre-treatment with -1.7 MPa PEG these values increased to 100% and 72% respectively [24]. Notably, the survival rate of seedlings with radicles >4 mm in length remained at 0%. Similarly, in *Tabebuia impetiginosa*, untreated seedlings are unable to re-establish DT once radicle length reaches 2.5 mm, while >50% of seedlings pre-treated with -1.7 MPa PEG can do so at this developmental stage [25]. In *A. thaliana*, while loss of DT occurred at the time of testa rupture in untreated seedlings, pre-treatment with -2.5 MPa PEG allowed re-establishment of DT in >95% of seedlings with radicles 0.3–0.5 mm in length, and in 38% of seedlings displaying root hair formation [18].

Transcriptome analysis of *M. trunculata* seedlings undergoing re-establishment of DT following PEG treatment revealed a significant overlap with changes in gene expression observed during the establishment of DT during seed maturation [26]. Similarly, analysis of gene expression in *A. thaliana* seedlings revealed that genes encoding LEA, seed storage and dormancy related proteins were strongly up-regulated following PEG treatment, while those involved in energy metabolism and cell-wall modification were repressed [18]. Together these data suggest that re-establishment of DT in seedlings involves a return to a quiescent state similar to that occurring in mature orthodox seeds.

If vegetative DT has indeed evolved in resurrection plants via co-option of the genetic network responsible for the acquisition of DT in seeds, one possible mechanism by which this may have occurred is through extension of the developmental window during which seedling re-establishment of DT can occur. If this is so, we would predict that the seedlings of resurrection plants would not display a point of no return, and would instead be able to re-establish DT throughout germination and beyond. In order to test this prediction, we characterised the germination process of seeds of the resurrection plant *Xerophyta viscosa* (Velloziaceae), and compared the ability of seedlings of *X. viscosa* and the model plant *A. thaliana* to re-establish DT following germination.

Materials and Methods

Plant material

X. viscosa (Baker) seeds were obtained from Silverhill seeds (http://www.silverhillseeds.co.za/default.asp), from a batch of seeds collected from Witsieshoek (Free State province, South Africa). Seeds were stored in the dark at room temperature. *Arabidopsis thaliana* seeds of the Columbia (Col-0) ecotype were obtained from the European Arabidopsis Stock Centre (http://arabidopsis.info) and stored at 4°C in the dark.

Germination and growth conditions

Seeds of both species were surface sterilised with 75% (v/v) ethanol for 2 min and air-dried in a laminar flow hood. Seeds were germinated on half-strength Murashige and Skoog (MS) medium containing 0.7% (w/v) bacteriological agar. Unless specified otherwise, seeds of *X. viscosa* were germinated immediately while *A. thaliana* seeds were cold-stratified at 4°C for 48 h in the dark prior to germination. Plants were grown under a long-day photoperiod (16 h light, 8 h dark) at 22°C, and cool white fluorescent light of 250 μmol $m^{-2}s^{-1}$.

Photography and seedling measurements

Seed and seedling photographs were obtained using a digital colour camera (JVC) mounted on a dissecting stereomicroscope (Olympus). Seedlings were photographed on a single piece of damp filter paper in order to prevent premature dehydration of their tissues due to strong lighting. Cotyledon, root and seed measurements were determined for each seedling using the ImageJ imaging software (http://rsbweb.nih.gov/ij/). High resolution photographs of seeds and seedlings were obtained using a Nikon Stereoscope Zoom Microscope (SMZ1500) and NIS-Elements (Nikon) digital 3D imaging software.

Response of *X. viscosa* seeds to various dormancy-breaking treatments

Batches of 25-30 seeds were treated with one of a number of dormancy breaking treatments previously described for various plant species [27]: 4°C or 37°C incubation in the dark for 72 h on half strength MS medium plates; 72 h soak in the dark at 4°C or 37°C in either a 1% (w/v) thio-urea solution or 0.25% (w/v) potassium nitrate solution; 72 h stored at −20°C; acid scarification with a 50% (v/v) H_2SO_4 solution for 1 minute; 60°C soak in water for either 1 min or 1 h; twelve weeks 4°C moist chilling in constant dark on half-strength MS plates. The treated seeds were subsequently transferred to half-strength MS medium plates (if not already on such a plate) and germinated under standard growth conditions.

The effect of developmental stage on seedling survival after rapid dehydration

A. thaliana seedlings were developmentally staged according to the system previously described by Maia *et al.* [18]. The stages of *X. viscosa* germination have not previously been described, and so this developmental process is described in the results section. For the dehydration experiments, germinated seeds at specific developmental stage were transferred to damp filter paper in an open petri dish and dehydrated under constant air flow in a laminar flow hood. After 48 h, seedlings were rehydrated on the filter paper with 2.5 ml sterile water and the plates transferred back to standard growth conditions. Seedlings were tracked over a 5 d period by daily inspection under an Olympus dissecting light stereo microscope, and considered DT if they resumed normal development.

The impact of PEG or sucrose treatment on seedling revival after dehydration

For the PEG-treatment, germinated seeds at appropriate developmental stages were moved to filter paper saturated with 1.2 ml of a PEG-8000 solution calculated to have an osmotic potential of -2.5 MPa at 22°C [28], on which they were incubated

under standard growth conditions for 48 h. Following PEG incubation, the seedlings were rinsed in sterile water to remove excess PEG and transferred to fresh damp filter paper. Dehydration and rehydration proceeded as described above. For the sucrose treatment experiments, seeds were germinated on half-strength MS in the presence or absence of 3% (w/v) sucrose under standard growth conditions. Germinated seeds at each developmental stage were then transferred to damp filter paper and subjected to dehydration as described above.

Data analysis

Due to the asynchronous and variable nature of *X. viscosa* germination, coupled with the difficulty in acquiring the seeds of this plant, the experiments described here occurred over the course of multiple weeks in a small number of large-scale experiments involving several hundred seeds. Data from multiple independent experiments was pooled and the variation within the entire pool of samples was determined via bootstrapping. The seedlings in each pooled experiment were resampled with replacement and placed into bins of pre-defined cotyledon lengths. This was repeated 10000 times, and the mean survival rate for the resampled seeds in each bin was calculated for each iteration. The standard error of the mean (SEM) of the sample pools was calculated from the standard deviation of the bootstrapped mean survival values. The same procedure was applied to the data generated for the *Arabidopsis* germination experiments to facilitate comparison.

Results

Germination and growth of *X. viscosa* seedlings

The germination and subsequent development of *X. viscosa* seedlings has not been previously been reported. We attempted to optimize germination conditions and developed a staging system to describe the growth of seedlings to aid further study on these plants. The seeds of *X. viscosa* are generally elliptical in shape, and surrounded by a triangular, dry, transparent husk that extends from one side of the seed (Figure 1A). When imbibed the seed swells with absorbed water and the embryo, located opposite the origin of the husk, is clearly visible (Figure 1B). The vast majority of the seed volume consists of nutrient endosperm, in which the embryo is embedded, and is surrounded by a thin, brown seed coat. Germination occurs over a period of 4 to 14 days thereafter. Germination rate and synchronicity were only marginally improved by various common dormancy breaking treatments, such as moist chilling (Figure S1). The germination of *X. viscosa* seedlings was found to visibly resemble that of the onion, *Allium cepa*, albeit on a much smaller scale [29]. Initially, the cotyledon lengthens within the seed and forces the developing radicle through the testa (Figure 1C). The embryo is reliant on nutrients from the endosperm for the first few days of germination and embryos excised during this stage are not viable. As germination progresses, the cotyledon continues to lengthen and rapidly turns green (Figure 1D). The radicle develops into a single primary root which develops root hairs once the cotyledons are between 1–3 mm in length (Figure 1E). At this stage, the embryo can be divided into four different zones: the cotyledon tip (a pale translucent ball of tissue that absorbs food stores from the endosperm), photosynthetic cotyledon tissue, the meristem above the roots, and a pointed root tip (Figure 1F). A central protoxylem is visible in cross-section of the germinating seedling, connecting the meristem, above the root zone to the endosperm (Figure 1F). The cotyledon undergoes a period of rapid extension (to 4–5 mm in length), before a split in the leading end of the cotyledon (Figure 1G) heralds the appearance of the first true leaf (Figure 1H). Undifferentiated, pale meristem tissue visible in cross-sections of seedlings above the root is no longer connected to the endosperm via the protoxylem (Figure 1I).

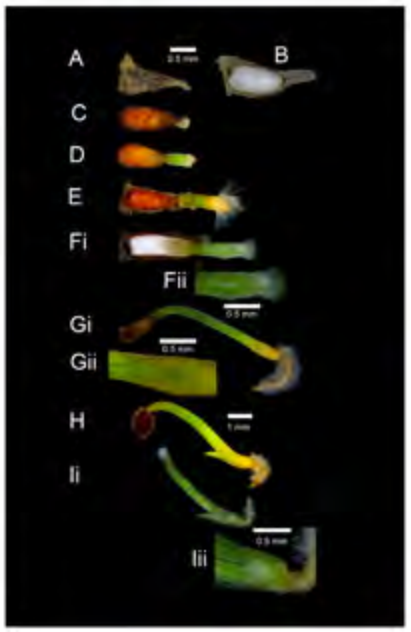


Figure 1. The germination process in *X. viscosa*. (A) dry seed; (B) cross-section of an imbibed seed; (C) radicle emergence; (D) chlorophyll is present in the emerging cotyledon; (E) root hairs appear once the cotyledon is 1 to 3 mm in length; (Fi) cross section of seedling showing tip of the photosynthetic cotyledon in contact with the endosperm, and the pale meristem; (Fii) a central protoxylem connecting the meristem to the tip of the cotyledon is visible; (Gi) the cotyledon and root tip elongate rapidly, and a slit appears above the root through which the first true leaf emerges (Gii); (H) the emerging leaf is clearly visible and the cotyledon ceases to grow. A cross section at this stage (Ii and Iii) shows that the meristem is no longer connected to the endosperm via the protoxylem. Seedlings at this stage are no longer dependent on the endosperm for survival. The 0.5 mm scale bar at the top of the figure is applicable to images A to Fi, and the 1 mm scale bar to images Gi, H and Ii. Individual scale bars are provided for Fii, Gii and Iii.
doi:10.1371/journal.pone.0093093.g001

The effect of developmental stage on seedling survival after rapid dehydration

Mature leaves of adult *X. viscosa* are desiccation tolerant [30]. We set out to determine whether germinating seedlings of *X. viscosa* are initially DT and whether this property was lost as seedlings matured through to the appearance of the first true leaves. We found that *X. viscosa* seedling survival was strongly associated with the seedling size at the time of drying (Figure 2A). Seedlings at an early stage of germination (i.e. cotyledon <0.4 mm) showed a near-100% survival rate. Survival rate decreased as seedlings increased in size, dropping to only 6% in seedlings with cotyledons 1.6–2 mm in length. Survival rate began to increase again, however, in larger seedlings (cotyledons >2 mm), rising to over 40% survival in the oldest tested seedlings (cotyledon >4.4 mm, around the stage at which the primary leaf emerges).

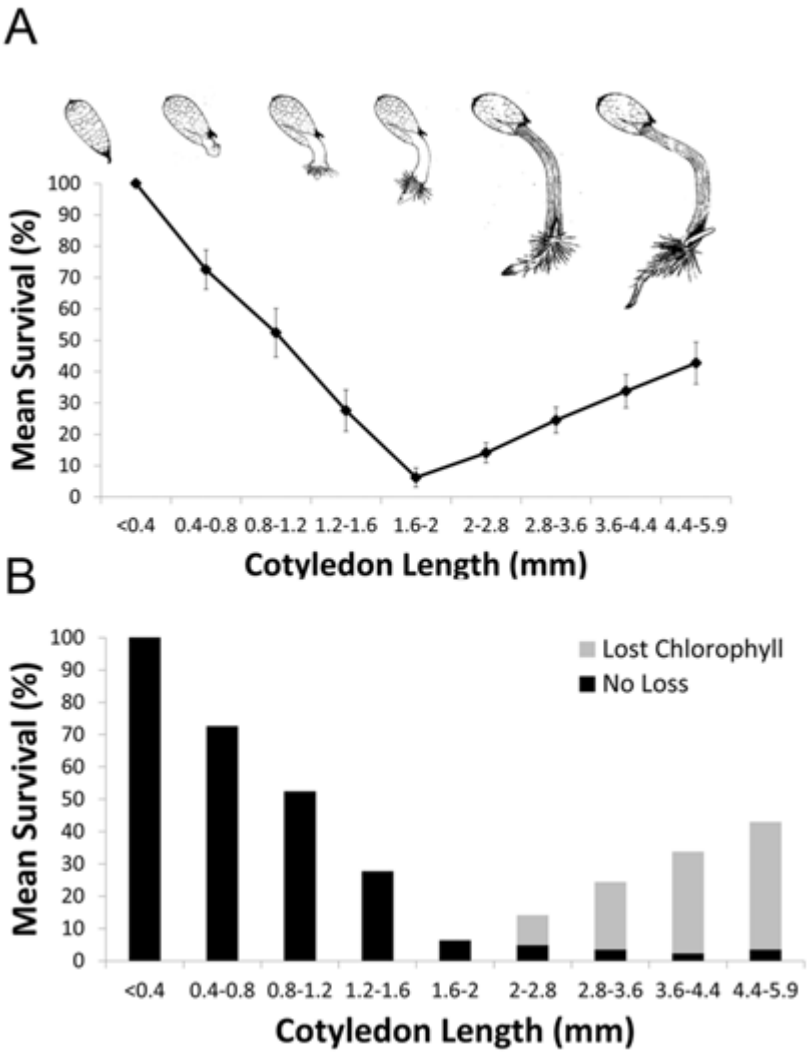


Figure 2. Desiccation tolerance in germinating seedlings of *X. viscosa* and *A. thaliana*. (A) Survival of germinating *X. viscosa* seedlings after 48 h desiccation, grouped by cotyledon length. Data shows the mean survival rate ±SEM as determined by bootstrapping. Representative line drawings of germinating seedlings are shown above. (B) Incidence of chlorophyll degradation in the surviving seedlings from Figure 2A. Black bars: survived and did not lose chlorophyll. Grey bars: survived and lost chlorophyll. The numbers of seeds per bin are provided in Table S1.
doi:10.1371/journal.pone.0093093.g002

Small seedlings (cotyledons <2 mm) displayed no visible predictors of survival: the cotyledons of desiccated seedlings remained green, and cotyledon and root tissues withered rapidly from lack of water (Figure 2B & 3). In contrast, survival of seedlings with cotyledons >2 mm was associated with the loss of chlorophyll either wholly or partially from the cotyledon (Figure 2B & 3). Of seedlings with cotyledons >2 mm, a total of 30% lost chlorophyll of which 76% survived dehydration, whereas only 5% of seedlings that did not lose chlorophyll at these stages survived. In a number of cases, tissue survival was restricted to regions that had degraded chlorophyll, while surrounding tissues failed to revive (Figure S2). The primary root of surviving seedlings usually also survived the desiccation process unharmed; in cases where it did not, a secondary root would erupt next to or through the dead root tissue (Figure 3). As has been previously reported [18], we observed that *A. thaliana* seedlings lost DT rapidly once the seed coat had ruptured. Less than 10% of seedlings survived 48 h of dehydration once the radicle had visibly protruded through the seed coat, with the survival rate dropping to 0% at all subsequent developmental stages (Figure 4A).

The impact of PEG or sucrose pre-treatment on seedling survival after dehydration

The transient decline in the ability of *X. viscosa* seedlings to survive dehydration during early germination (Figure 2A) might represent a developmental stage at which DT generally cannot be re-established. Alternatively, these young seedlings may be particularly vulnerable as they do not have sufficient resources to mount a successful response against or recover from very rapid dehydration. For example, depletion of the endosperm-derived nutrients upon which the germinating seedlings are dependent during initial growth, or a failure to protect the protoxylem connecting the endosperm to the meristem, may be lethal. The point of no return can be delayed in many species by priming germinating seedlings with moderate osmotic stress prior to dehydration, for example by incubating in high molecular weight

Figure 3. Recovery of *X. viscosa* seedlings from dehydration. Representative images of three untreated *X. viscosa* seedlings 1 mm, 2.2 mm and 3 mm in length before and after dehydration and rehydration. The 1 mm seedling dehydrated without chlorophyll loss, but recovered and resumed growth rapidly after rehydration. A 2.2 mm seedling likewise failed to degrade chlorophyll, but did not survive the rehydration process. A 3 mm seedling displayed partial chlorophyll loss from its cotyledon during dehydration. The seedling survived with its meristem, primary leaf and leading end of the cotyledon intact; the primary root failed to survive, but two additional secondary roots emerged from the meristem. Anthocyanins accumulated in the photosynthetic tissue during rehydration of this seedling.
doi:10.1371/journal.pone.0093093.g003

PEG. Sucrose has also been implicated in the survival of both germinating seedlings and resurrection plants in response to desiccation, and the presence of additional sucrose may alleviate resource depletion in germinating seedlings that have not yet become self-sufficient. We thus used PEG and sucrose pre-treatment to investigate whether priming of *X. viscosa* or *A. thaliana* seedlings improved their ability to survive dehydration.

In agreement with previous reports, incubation in a -2.5MPa PEG-8000 solution for 48 h prior to dehydration improved the survival rate of *A. thaliana* seedlings (Figure 4A), most noticeably at the stages of radicle protrusion (radicle 0.2 mm; 9% survived untreated vs. 74% PEG-treated) and radicle extension (radicle 0.35 mm; 1% vs. 20%). Nonetheless, mean survival rate dropped to 5% by the time the root hairs were first visible (radicle 0.45 mm). Growth on 3% sucrose resulted in delayed germination in *A. thaliana* and led to the accumulation of anthocyanins, as previously reported [31,32]. However, the presence of exogenous sucrose also improved seedling survival after dehydration. Whereas only 9% and 1% of untreated seedlings survived at the stages of radicle protrusion and extension respectively, supplementary sucrose improved survival to 59% and 42% at these stages (Figure 4A). Mean survival dropped to less than 5% in seedlings with a radicle longer than 0.45 mm.

PEG treatment also increased seedling survival rate after dehydration in *X. viscosa* (Figure 4B). Total seedling survival averaged 65%, and the transient decline in survival rate observed in untreated plants was not apparent in PEG-treated seedlings. Notably, the majority (83%) of seedlings of all sizes showed complete loss of chlorophyll after incubation in PEG (Figure 5). None of the 17% of PEG-treated seedlings that failed to degrade chlorophyll were able to survive dehydration. In contrast to its effects on delaying germination in *A. thaliana*, 3% sucrose did not have any noticeable effect on *X. viscosa* germination rate, nor did it induce anthocyanin accumulation in seedlings prior to dehydration (Figure 5). However, it drastically improved seedling survival irrespective of cotyledon length: average survival of sucrose-treated seedlings was 91% across all developmental stages, with the lowest survival rate of 79% occurring in seedlings 2–2.9 mm in size (Figure 4B). As with untreated plants, chlorophyll loss was only observed in seedlings with larger cotyledons (>1.5 mm), and occurred most often (>90%) in seedlings >3 mm in length (Figure 4C).

Discussion

The seeds of the desiccation-sensitive plant *A. thaliana* can halt the process of germination under adverse conditions, with a window of opportunity of 48–60 h post imbibition [33]. Current evidence suggests that this process involves resetting the seedling to a quiescent state that resembles that of mature, dry orthodox seeds [18]. A similar window is seen in many other angiosperm plant species. The acquisition of vegetative DT in angiosperms may have evolved through the co-option of the networks regulating angiosperm seed maturation genes into adult tissues [2,4]. A possible means by which this could have occurred would be through extending the temporary window during which angiosperm seedlings are able to re-establish DT indefinitely. If this is the case, we would predict that the seedlings of a resurrection plant would not display a point of no return, but would instead be desiccation tolerant throughout germination. Our aim was to determine the extent to which this window of opportunity existed in the monocot resurrection plant *X. viscosa*, a poikilochlorophyllous resurrection plant from Southern Africa. We investigated the DT of the seedlings of both *X. viscosa* and the desiccation sensitive *A. thaliana*, when untreated or either incubated in PEG or supplemented with exogenous sucrose.

In agreement with previous reports, we found that the seedlings of *A. thaliana* rapidly lost DT once the seed coat had ruptured (Figure 4A), but could be temporarily rescued by pre-treatment in a high molecular weight PEG solution prior to desiccation [18]. Growth on medium supplemented with 3% sucrose, though it delayed germination and retarded seedling development, was also capable of increasing survival rates in germinating seedlings to a similar degree. However, neither PEG nor sucrose treatment could substantially extend the window beyond the point at which radicle length >0.45 mm (Figure 4A).

X. viscosa seedlings grown on half-strength MS media displayed a transient decline in DT during germination. Although freshly imbibed seedlings were desiccation tolerant, seedlings steadily lost DT as germination progressed, dropping to a minimum survival

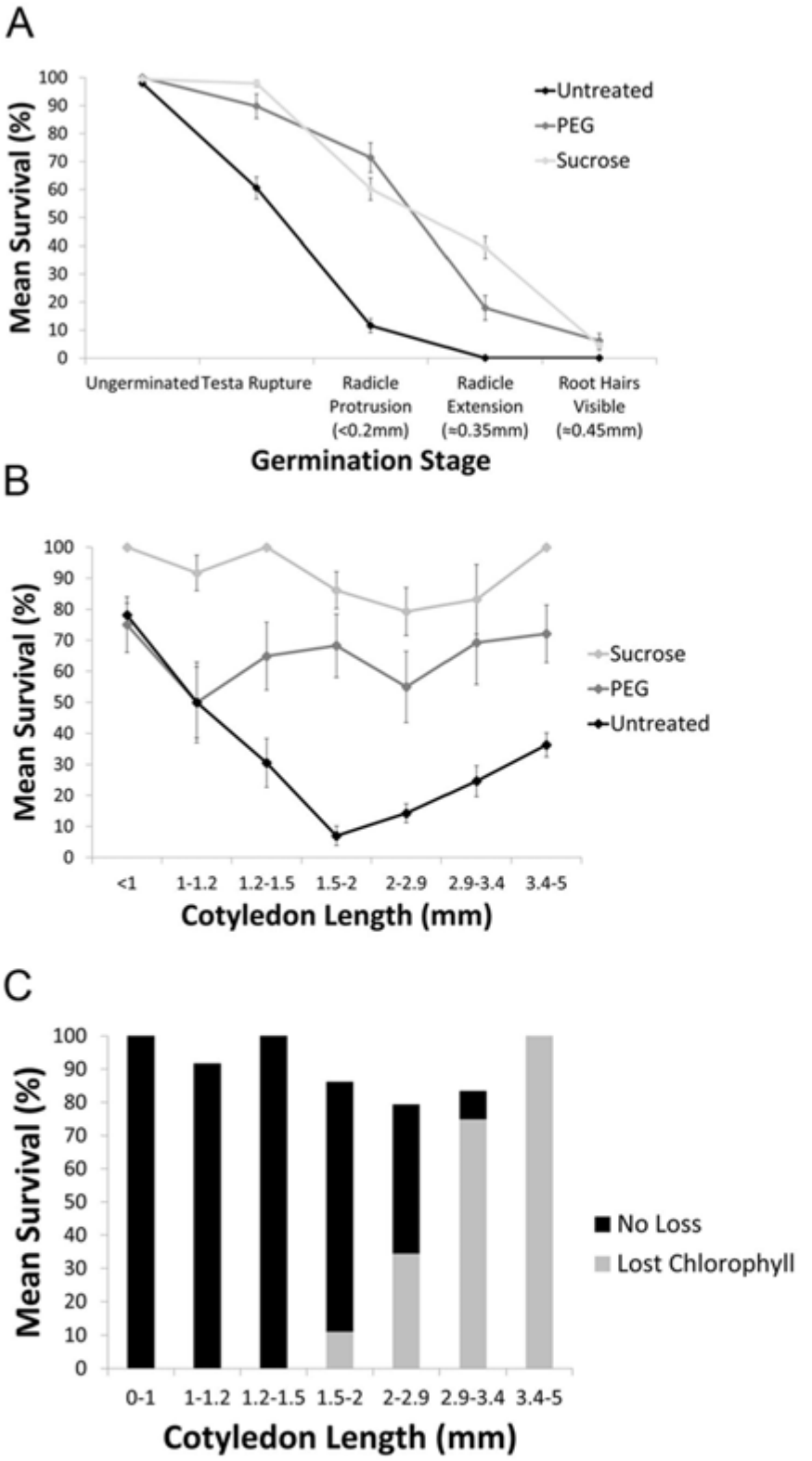


Figure 4. Pre-treatment with PEG or sucrose improves desiccation tolerance in germinating seedlings of *A. thaliana* and *X. viscosa*. (A) Mean survival rate (±SEM) of *A. thaliana* seedlings after 48 h desiccation at various stages of germination. Although survival improves with PEG and sucrose treatment, *A. thaliana* seedlings are unable to survive desiccation past the ''root hairs visible'' seedling stage. (B) Survival of germinating *X. viscosa* seedlings with or without PEG or sucrose pre-treatment, grouped by cotyledon size (±SEM). Seedlings pre-treated with either sucrose or PEG did not show a decline in desiccation tolerance compared to untreated seedlings. (C) Chlorophyll loss in *X. viscosa* sucrose-treated seedlings. Although the mean survival rate of sucrose-treated seedlings was relatively consistent across most stages, the incidence of chlorophyll loss increased steadily with increasing seedling size. Seeds per bin can be found in Table S1.
doi:10.1371/journal.pone.0093093.g004

Figure 5. Pre-treatment of *X. viscosa* seedlings with PEG results in chlorophyll degradation. Representative images of an untreated, PEG-treated and sucrose-treated seedling undergoing dehydration and rehydration. An untreated seedling (MS media) lost chlorophyll and accumulated anthocyanins in response to dehydration, and survived rehydration. A PEG treated seedling completely degraded its chlorophyll in response to PEG priming (prior to dehydration), and recovered well after rehydration. A sucrose-treated seedling failed to lose chlorophyll during dehydration; however, the meristem, primary leaf and a small portion of the cotyledon recovered after rehydration nonetheless.
doi:10.1371/journal.pone.0093093.g005

rate of 6% in seedlings between 1.6–2 mm in length (Figure 2A). However, this was a temporary phenomenon, and a greater proportion of seedlings recovered from dehydration at later stages, rising to over 40% survival in seedlings >4.4 mm. The majority of surviving seedlings greater than 2 mm in length also lost chlorophyll from their photosynthetic tissues (Figure 2B), a trait reminiscent of the mechanism of poikilochlorophylly employed by mature *X. viscosa* plants to prevent build-up of reactive oxygen species during dehydration [34].

Pre-treatment of germinating *X. viscosa* seedlings with PEG for 48 h prior to desiccation reduced the transient decline in DT (Figure 4B), and improved the average survival rate over that of untreated seedlings by a large margin (minimum survival rate was over 50%, compared to 6% in untreated seedlings). Interestingly, all surviving seedlings showed a complete loss of chlorophyll from the PEG treatment alone, irrespective of seedling size. Sucrose-supplemented *X. viscosa* seedlings displayed an even more marked increase in survival: on average 91% of total seedlings recovered completely after desiccation, with a minimum survival rate of 79% in seedlings 2–2.9 mm in length (Figure 4B). As with untreated seedlings, only sucrose-treated seedlings with a cotyledon >2 mm displayed any degradation of chlorophyll during dehydration, with the incidence of chlorophyll loss increasing as seedlings increased in size (Figure 4C).

Moderate osmotic stress during the early stages of germination has been shown to improve the survival of desiccated seedlings in a number angiosperm species [23,24]. The exact mechanism through which this occurs is not clear, but it seems likely that an extended period of moderate stress primes the seedlings to activate the seed maturation genes and desiccation response mechanisms prior to the period of rapid desiccation. On the other hand, treatment with exogenous sucrose has to our knowledge not been previously shown to induce DT in germinating seedlings, although it has been reported to induce DT in axillary buds and shoot tips of various species for the purposes of cryopreservation [35,36].

As PEG treatment alone was sufficient to substantially rescue *X. viscosa* seedlings at the stages at which they were most sensitive to dehydration, it is unlikely that the transient loss of DT in untreated seedlings was due to an inherent inability of these seedlings to induce DT. Some other factor, such as depletion of endosperm-derived nutrients, or increased dehydration rate due to a larger surface-area-to-volume ratio, may have played a role in reducing the survival rate of seedlings at these sizes. Whilst it is well known that lower plants, such as the moss *Tortula ruralis*, can withstand very rapid water loss, this is not possible for angiosperm resurrection plants which need time to lay down protective measures [37]. We hypothesise that pre-treatment with PEG primed the desiccation response prior to the rapid dehydration treatment. Consistent with this hypothesis, all surviving PEG-treated seedlings showed a loss of chlorophyll from the cotyledon, indicating that *X. viscosa* is capable of poikilochlorophylly at all stages of germination. This suggests either a direct relationship between the *X. viscosa* seedling and vegetative desiccation response mechanisms, or alternatively that the adult response mechanisms can be induced in seedlings under the appropriate conditions. None of the PEG-treated seedlings that failed to lose chlorophyll (17% of the total) survived the dehydration treatment, strong evidence that chlorophyll degradation is an indicator of the preparedness of a seedling to respond to rapid, severe water loss.

In contrast, small (<2 mm) untreated and sucrose-treated seedlings failed to degrade chlorophyll during dehydration. This observation may be explained by the rate of water loss in these seedlings being too rapid for this process to occur without PEG priming due to the seedlings' small size. However, the survival of these small seedlings without chlorophyll loss suggests that the mechanisms that control vegetative DT and poikilochlorophylly do not occur simultaneously; rather DT protection is activated prior to visible loss of chlorophyll from these tissues. Presumably, in non-PEG-treated seedlings >2 mm in length, their increased size sufficiently delayed the rate of dehydration for chlorophyll degradation to begin to occur in these tissues in the absence of any priming.

Sucrose treated seedlings showed an even greater survival rate than that of PEG treated seedlings in both species, despite the fact that they should have had a similar dehydration rate to untreated seedlings. Thus, the rate at which seedlings lose water is not the only factor influencing the probability of recovery. The exact role exogenous sucrose plays in this process cannot be determined from the results in this study. It seems unlikely that the relatively minor osmotic effects of sucrose alone (−0.22 MPa) [38], would be

sufficient to prime the DT response, as osmotically stressed PEG-treated seedlings (−2.5 MPa) did not recover to a similar degree. Exogenous sucrose may have acted as a nutrient source that was independent of seedling stage, reducing the total number of seedlings that failed to survive purely due to a lack of sufficient resources to mount a successful desiccation response. Sucrose has also been predicted to act as an osmoprotectant that, together with LEA proteins, forms the cellular glass that protects tissues from desiccation [39]. For example, it is possible that the exogenous sucrose protected the protoxylem which serves as a lifeline between the endosperm reserves and the meristem in small seedlings. Alternatively, high background levels of intra- or extracellular sucrose could have also reduced the time needed to induce glass formation during the desiccation process. Treatment with exogenous sucrose has been shown to have a synergistic effect with ABA in inducing DT in seeds or somatic embryos, if applied during the early stages of the maturation process [40–42]. If the window of seedling DT does involve an ABA-mediated reactivation of maturation genes, high sucrose levels may have had a similar synergistic signalling role with ABA. None of these roles are mutually exclusive, and it is likely that exogenous sucrose functions through multiple pathways to improve survival rates of desiccating seedlings.

Our data suggest that *X. viscosa* seedlings do not inherently lose DT during germination, as other angiosperm species do. A particularly striking result is the potential for poikilochlorophylly at all seedling stages, suggesting a direct relationship between the seedling and adult desiccation response. One explanation for these results is that *X. viscosa* never closes the window of ABA responsiveness that allows germinating angiosperm seeds to revert to a quiescent, mature-dry state, but has instead evolved the ability to activate this response throughout its life, even in adult tissues. The acquisition of additional molecular data and functional assays will be informative in determining whether the vegetative desiccation response in *Xerophyta* is indeed an extension of the germinating seedling desiccation response, or whether these are different survival programmes.

Supporting Information

Figure S1 The effect of various common dormancy-breaking treatments on seeds of *X. viscosa*. Data shown are the percentage germination over time from batches of 25 to 50 seedlings per experiment. (A) Dormancy-breaking treatments coupled with cold (4°C or freezing) temperature. (B) Dormancy-breaking treatments coupled with warm (37°C) temperature. (C) Extended 4°C stratification coupled with germination in light or dark conditions. Not shown: acid scarification (50% H_2SO_4, 1 minute) and 60°C soak, which resulted in 100% mortality. While the majority of treatments advanced the germination process by 24 h (possibly due to advanced period of imbibition compared to untreated seedlings), neither the rate of germination nor the total percentage of seeds germinating was greatly enhanced.
(TIF)

Figure S2 Chlorophyll loss is a predictor of tissue survival in older seedlings. Chlorophyll degradation and subsequent recovery (indicated by black arrows) of seedling tissues was a strong predictor of survival after dehydration. A) A desiccated seedling shows chlorophyll loss around the meristem, a small emerging leaf, and the upper cotyledon. B) These tissues recover 5 days post rehydration, the meristem survives, and the emerging leaf is substantially larger. C) A desiccated seedling showing chlorophyll loss from only the upper cotyledon D) Whilst chlorophyll is restored in this section of the cotyledon, the meristem and lower cotyledon fail to survive.
(TIF)

Table S1 Seedling bin sizes for figures 2A, 4A and 4B. The seedlings from independent experimental repeats were pooled and divided into bins based on cotyledon length (*X. viscosa*) or germination stage (*A. thaliana*).
(XLSX)

Author Contributions

Conceived and designed the experiments: RL RAI NI. Performed the experiments: RL NI. Analyzed the data: RL RAI NI. Contributed reagents/materials/analysis tools: RL RAI NI. Wrote the paper: RL RAI NI.

References

1. Bewley J (1979) Physiological aspects of desiccation tolerance. Ann Rev Plant Physiol 54: 195–238.
2. Gaff DF, Oliver M (2013) The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. Funct Plant Biol 40: 315–328.
3. Oliver M, Tuba Z, Mishler B (2000) The evolution of vegetative desiccation tolerance in land plants. Plant Ecol 151: 85–100.
4. Illing N, Denby KJ, Collett H, Shen A, Farrant JM (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.
5. Bewley JD, Oliver MJ (1992) Desiccation tolerance in vegetative plant tissues and seeds: Protein synthesis in relation to desiccation and a potential role for protection and repair mechanisms. In: Somero GN, Osmond CB, Bolis CL, editors. Water and Life. Berlin: Springer. pp. 141–160.
6. Bewley JD, Black MJ (1994) Seeds. Physiology of development and germination. New York: Plenum Press. 367 p.
7. Vicente-Carbajosa J, Carbonero P (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. Int J Dev Biol 49: 645–651.
8. Kermode AR (1990) Regulatory mechanisms involved in the transition from seed development to germination. Crit Rev Plant Sci 9: 155–195.
9. Vertucci CW, Farrant JM (1995) Acquisition and loss of desiccation tolerance. In: Kigel J, Galili G, editors. Seed development and germination. New York: Marcel Dekker Press. pp. 237–271.
10. Kermode AR, Finch-Savage BE (2002) Desiccation sensitivity in orthodox and recalcitrant seeds in relation to development. In: Black M, Pritchard HW, editors. Desiccation and survival in plants: drying without dying. CABI Publishing. pp. 149–184.
11. Cuming AC (1999) LEA proteins. In: Shewry PR, Casey R, editors. Seed proteins. Berlin: Springer. pp. 753–780.
12. Hoekstra F, Golovina E, Buitink J (2001) Mechanisms of plant desiccation tolerance. Trends Plant Sci 6: 431–438.
13. Berjak P (2006) Unifying perspectives of some mechanisms basic to desiccation tolerance across life forms. Seed Sci Res 16: 1–15.
14. Baud S, Boutin J, Miquel M (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. Plant Physiol Bioch 40: 151–160.
15. Leprince O, Bronchart R, Deltour R (1990) Changes in starch and soluble sugars in relation to the acquisition of desiccation tolerance during maturation of *Brassica campestris* seed. Plant Cell Environ 13: 539–546.
16. Finch-Savage W, Leubner-Metzger G (2006) Seed dormancy and the control of germination. New Phytol 171: 501–523.
17. Evenari M, Shanan L, Tadmor N (1971) The Negev: The challenge of a desert. Cambridge: Harvard University Press. 345 p.
18. Maia J, Dekkers BJW, Provart NJ, Ligterink W, Hilhorst HWM (2011) The re-establishment of desiccation tolerance in germinated *Arabidopsis thaliana* seeds and its associated transcriptome. PloS One 6: e29123.
19. Negbi M, Evenari M (1962) The means of survival of some desert summer annuals. Plant water relationships in arid and semiarid conditions (Proceedings Madrid Symposium UNESCO) pp. 249–259.
20. Gutterman Y (2000) Environmental factors and survival strategies of annual plant species in the Negev Desert, Israel. Plant Spec Biol 15: 113–125.
21. Huang Z, Dong M, Gutterman Y (2004) Factors influencing seed dormancy and germination in sand, and seedling survival under desiccation, of *Psammochloa villosa* (Poaceae), inhabiting the moving sand dunes of Ordos, China. Plant Soil 259: 231–241.
22. Friedman J, Stein Z, Rushkin E (1981) Drought tolerance of germinating seeds and young seedlings of *Anastatica hierochuntica*. Oecologia 51: 400–403.

23. Bruggink T, van der Toorn P (1995) Induction of desiccation tolerance in germinated seeds. Seed Sci Res 5: 1–4.
24. Buitink J, Vu BL, Satour P, Leprince O (2003) The re-establishment of desiccation tolerance in germinated radicles of *Medicago truncatula* Gaertn. seeds. Seed Sci Res 13: 273–286.
25. Vieira CV, Amaral da Silva EA, Alvarenga AA, Castro EM, Toorop PE (2010) Stress-associated factors increase after desiccation of germinated seeds of *Tabebuia impetiginosa* Mart. Plant Growth Regul 62: 257–263.
26. Buitink J, Leger JJ, Guisle I, Vu BL, Wuillème S, et al. (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. Plant J 47: 735–750.
27. Emery DE (1988) Seed propagation of native California plants. Santa Barbara Botanic Garden. 115 p.
28. Michel BE (1983) Evaluation of the water potentials of solutions of polyethylene glycol 8000 both in the absence and presence of other solutes. Plant Physiol 72: 66–70.
29. Gray A (1887) The elements of botany for beginners and for schools. American Book Company.
30. Mundree SG, Farrant JM (2000) Some physiological and molecular insights into the mechanisms of desiccation tolerance in the resurrection plant *Xerophyta viscosa* Baker. In: Cherry JH, Locy RD, Rychter A, editors. Plant tolerance to abiotic stresses in agriculture: Role of genetic engineering. Netherlands: Kluwer Academic Publishers. pp. 201–222.
31. Gibson SI (2005) Control of plant development and gene expression by sugar signaling. Curr Opin Plant Biol 8: 93–102.
32. Teng S, Keurentjes J, Bentsink L, Koornneef M, Smeekens S (2005) Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the MYB75/PAP1 gene. Plant Physiol 139: 1840–1852.
33. Lopez-Molina L, Mongrand S, McLachlin DT, Chait BT, Chua N-H (2002) ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. Plant J 32: 317–328.
34. Ingle RA, Schmidt UG, Farrant JM, Thomson JA, Mundree SG (2007) Proteomic analysis of leaf proteins during dehydration of the resurrection plant *Xerophyta viscosa*. Plant Cell Environ 30: 435–446.
35. Suzuki M, Ishikawa M, Akihama T (1998) A novel preculture method for the induction of desiccation tolerance in gentian axillary buds for cryopreservation. Plant Sci 135: 69–76.
36. Suzuki M, Akihama T, Ishikawa M (2005) Cryopreservation of encapsulated gentian axillary buds following 2 step-preculture with sucrose and desiccation. Plant Cell Tiss Org 83: 115–121.
37. Oliver MJ, Bewley DJ (1997) Desiccation-tolerance of plant tissues: a mechanistic overview. Hort Rev 18: 171–213.
38. Sarkar A (2009) Plant stem cells. New Delhi: Discovery Publishing House. 324 p.
39. Buitink J, Leprince O (2008) Intracellular glasses and seed survival in the dry state. C R Biol 331: 788–795.
40. Senaranta T, McKersie BD, Bowley SR (1989) Desiccation tolerance of alfalfa (*Medicago sativa* L.) somatic embroys. Influence of abscisic acid, stress pretreatments and drying rates. Plant Sci 65: 253–259.
41. Saranga Y, Rhodes D, Janick J (1990) Factors inducing desiccation tolerance in celery somatic embryos. HortScience 25: 1088.
42. Meurs C, Basra AS, Karssen CM, Van Loon LC (1992) Role of abscisic acid in the induction of desiccation tolerance in developing seeds of *Arabidopsis thaliana*. Plant Physiol 3: 1484–1493.