CENTRO DE INVESTIGACION Y ESTUDIOS AVANZADOS DEL INSTITUTO POLITECNICO NACIONAL

UNIDAD IRAPUATO

UNIDAD DE GENOMICA AVANZADA

"Análisis molecular de la tolerancia a la desecación en semillas de *Arabidopsis thaliana"*

*"*Molecular analysis of desiccation tolerance in *Arabidopsis thaliana* seeds"

Tesis que presenta:

IBQ. Sandra Isabel González Morales

como requisito para obtener el grado de

Doctorado en Ciencias

en la especialidad de Biotecnología de Plantas

Director de tesis:

Dr. Luis Rafael Herrera Estrella

Irapuato, Gto., México Septiembre, 2016

ACKNOWLEDGEMENTS

I would like to thanks to CONACYT, for the scholar fellowship 220090 to complete my Ph. D.

I would like to thanks to Dr. Luis Herrera Estrella for allowing me the opportunity to work in his laboratory, for his great support, advices and guide through this project

To Drs. Alfredo Herrera, Stefan de Folder, Octavio Martinez de la Vega and Luis González de la Vara to be part of my committee and their support with their comments, advices and suggestions during these years, as well as to Dra. Alejandra Covarrubias, to be part of my thesis committee.

To Dr. Ricardo Chavez for his contribution in the network analysis as well as all discussions related to that, also for his friendship and advices.

To Dr. Corina Hayano for her contribution at beginning of the project and also for her friendship.

To M.C. Gerardo Alejo for his valuable contribution with the mutant characterization also to believe on this new research line, as well as his friendship during these years.

To IBQ. Thelma Rico for his valuable contribution with the transgenic lines characterization also to show high interest on this new research line, as well as his friendship during these years

To all member of the Physiology and Metabolic Engineering of Plants Laboratory from LANGEBIO to share with me nice moments.

I would like to thanks to Araceli Oropeza and Rosy Gutierrez for providing me with material and several tasks in the laboratory.

I would like to thanks to administrative and technical members of CINVESTAV and LANGEBIO who have been kind enough to advise and help in different tasks.

To my family for the support provided thru this amazing road to obtain the Ph. D. Grade.

To my husband for his love and support during these years.

Index

Index of figures

Index of tables

LIST OF NOMENCLATURES

DT	Desiccation Tolerace
TF	Transcription Factor
ROS	Reactive Oxygen Species
DAF	Days After Flowering
TAG	Triacylglycerol
LEA	Late Embryogenesis Abundant
HSP	Heat Shock Proteins
SOD	Superoxide Dismutase
GSH	Glutathione
CAT	Catalase
T-DNA	Transfer DNA
GR	Glutathione Reductase
ABA	Abscisic Acid
GA	Gibberellin
SSP	Seed Storage Protein
RNA-seq	RNA Sequencing
RFO	Raffinose-Family Oligosaccharides
AP	Ascorbate Peroxidase
RH	Relative Humidity
cDNA	Complementary DNA
WTA	Whole Transcriptome Analysis
RPKM	Reads Per Kilobase Of Exon Sequence, Per Million Reads
TMM	Trimmed Mean M
DEG	Differentially Expressed Genes
FDR	False Discovery Rate
GLM	Generalized Linear Model
GO	Gene Ontology

RESUMEN

La tolerancia a la desecación (TD) es un extraordinario proceso que permitió a las plantas con semillas conquistar ecosistemas que presentan baja disponibilidad de agua por periodos prolongados. Este proceso permite mantener a las semillas secas y viables por siglos o incluso por milenios. En *Arabidopsis*, LEC1, LEC2, FUS2 and ABI3 son factores de transcripción (FTs) los cuales son importantes en la maduración de la semilla incluyendo la TD, sus mutantes tienen defectos en la TD debido a la ausencia o niveles reducidos de algunos componentes como proteínas LEA y de choque térmico y la acumulación de oligosacáridos. Sin embargo, existe muy poca información sobre los componentes de señalización requeridos para lograr la TD o cómo los FTs modulan los procesos globales de TD. Entender las redes que regulan la TD en semillas en plantas modelo nos proveerá las herramientas para comprender el proceso evolutivo el cual es muy importante en la diversificación de las plantas con flores. A través, de un análisis comparativo de RNA-seq y perfiles metabólicos de las mutantes *lec1, lec2, fus3* y *abi3* así como sus correspondientes silvestres durante el periodo de desecación de la semilla identificamos genes específicamente relacionado con los procesos de TD, estos datos nos permitieron integrar a los procesos metabólicos, rutas de señalización y FTs. Simultáneamente, la ingeniería inversa de redes de regulación de la TD reveló módulos transcripcionales que activan a los genes de TD en semillas. Muy notablemente, dos subredes transcripcionales fueron identificadas, una relacionada a la síntesis compuestos de reserva y la otra a mecanismos de protección celular. La expresión ectópica de PLATZ1, AGL67 y DREB2D, identificados en estas subredes, fueron suficiente para activar genes de TD porque parcialmente rescata el fenotipo de intolerancia a la desecación en la mutante *abi3-5*, mientras que las mutantes estos FTs mostraron una reducida TD en semillas. Finalmente, encontramos que la expresión constitutiva de PLATZ1 en plantas silvestres confiere tolerancia a la baja disponibilidad de agua en tejidos vegetativos.

ABSTRACT

Desiccation tolerance (DT) is a remarkable process that allowed seed plants to conquer ecosystems with long periods of limited water availability. This process allows seeds in the dry state to remain viable for centuries or even thousands of years. In Arabidopsis, LEC1, LEC2, FUS3 and ABI3 are transcription factors (TFs) that are key players in seed maturation including DT. Their mutants lack DT, due to the absence or have reduced levels of some components like LEA, heat shock proteins, and accumulation of oligosaccharides. However, there is little information about the signaling components required to achieve DT and on how TFs modulate global DT processes. Understanding the networks that regulate seed DT in model plant systems will provide the tools to understand an evolutionary process that played a crucial role in the diversification of flowering plants. Through comparative analysis of RNA-seq and metabolic profiles of *lec1*, *lec2*, *fus3* and *abi3* mutants as well as their corresponding wild types during the seed desiccation period, we identified genes specifically involved in the DT process. These data enabled us to integrate metabolic processes, signaling pathways, and specific TF activity. Additionally, reverse engineering of a DT-specific regulatory network revealed transcriptional modules that activate the DT genes in seeds. Notably, two major transcriptional networks were identified related to storage of reserve compounds and cell protection mechanisms, respectively. Ectopic expression of PLATZ1, AGL67 and DREB2D identified in these subnetworks are sufficient to activate genes that contribute to DT, because they partially rescue the desiccation intolerance phenotype of the *abi3* mutant, whereas the elimination of these TFs showed reduced seed DT. Finally, we found that constitutive expression of PLATZ1 in WT plants confers tolerance to low water availability in vegetative tissues.

1. INTRODUCTION

1.1. Introduction to desiccation tolerance

Water is the most limiting resource in living systems and most abundant compound in all active cells, it is essential for metabolism and organism need take water from the environment to survive. Water molecules are critical components of chemical reactions and contribute to the stability of proteins, lipids, DNA and membranes. The survival of organisms in the absence or under very limited amounts of water is still an open question. A major challenge for all organisms is to grow and reproduce in an atmosphere that is very dry in many terrestrial regions. Indeed, upon drying below 90 % relative humidity, desiccation-sensitive cells and organisms suffer damage to all cell components. Under such condition proteins lose their correct conformation and tend to form toxic aggregates: membranes undergo phase transitions causing their leakiness and fusion. RNA and DNA undergo structural rearrangements and fragmentation, and chromatin stability is compromised. Also the stress imposed by the loss of water leads to oxidative damage because drying occurs in the presence of oxygen in cells where high rates of electron transport activities are conducive of generating reactive oxygen species (ROS). Nevertheless, biological mechanisms that allow a small proportion of organisms to survive essentially complete desiccation, which results in dry but viable tissues, containing as little as 0.1 % of their hydrated water content. When water becomes available again, they rapidly imbibe and resume metabolism (Leprince and Buitink 2015).

Desiccation tolerance (DT) can be operationally defined as the ability of an organism to dry to equilibrium with moderately dry air (50-70% relative humidity at 20-30 °C) and then resume normal function when rehydrated. Almost all species known to recover from complete drying at 80% relative humidity also recover from drying at 50% (Alpert 2002). DT can be conceptually defied as reversible interruption of metabolism in response to water loss. This suggests that the mechanism of DT must involve at least two important elements. First, the shutdown of metabolism during desiccation, for example metabolic pathways that lead to accumulation of intermediates and generation of free radical, minimizing photo-oxidation may explain why some desiccation-tolerant plants stop photosynthesis at relatively high water contents during drying. Second, essential preserve cells enough cellular organization and functional enzymes so that metabolism can resume after rewetting. Preserving a skeletal machinery for metabolism must involve both protection and repair (Alpert 2002).

1.2. The occurrence of desiccation tolerance

Desiccation-tolerant organisms are present in all domains, both in prokaryotes and eukaryotes. In bacteria, DT has been encountered among many different forms such as akinetes, exospores, myxospores, cysts as well as vegetative cells, the most DT forms tending to be *Gram* positive. One the best studied species as a model is *Nostoc commune*, a filamentous cyanobacterium that is conspicuous on exposed limestone of karst regions. In fungi, studies on DT have focused mainly on *Saccharomyces cerevisiae* which has turned out to be a powerful genomic model to test *in vivo* mechanisms of DT that were previously characterized *in*

*vitro(*Leprince and Buitink 2015). In the animal kingdom, DT commonly occurs in three phyla: nematodes (e.g., *Aphelenchus avena* and the dauer stage of the genetic model species *Caenorhabditis elegans*), rotifers and tardigrades. It is also present in juvenile stages of two additional phyla: the encysted embryo stage of some crustaceans, including the brine shrimp *Artemia sp* and the larvae of a non-biting midge, *Polypedilum vanderplanki*.(Leprince and Buitink 2015; Alpert 2002). In plants, DT is considered to be a complex trait that is present in reproductive structures (pollen and seeds) of vascular plants. DT in vegetative tissues is also relatively common in less complex plants such as bryophytes and lichens but rare in pteridophytes and angiosperms and absent in gymnosperms (Alpert 2002; Gaff and Oliver 2013; Oliver et al. 2005).

1.3. Evolution of desiccation tolerance in plants

Early land plants had to evolve mechanisms to survive harsh drying environments to successfully exploit different ecosystems in dry land. Therefore, it has been postulated that the initial evolution of vegetative DT, in both vegetative and reproductive stages, was a crucial step required for land colonization by primitive plants from a fresh water origin (Mishler and Churchill 1985). Recent synthetic phylogenetic analyses suggest that vegetative DT was primitively present in less complex plants such as bryophytes, but was then lost in the evolution of vascular plants (tracheophytes) (Figure1). The initial evolution of constitutive vegetative DT came at a cost, metabolic rates in tolerant plants are low compared to those in desiccation-sensitive plants, therefore, loss of vegetative DT was probably favored

in vascular plants. Ancestral vegetative DT was then probably rewired and perfected to generate DT in seed and pollen (Oliver et al. 2000).

Successful radiation of vascular plants on land was probably a direct consequence of the evolution of DT in seeds, in parallel to the evolution of structural and morphological modifications in vegetative tissue that permitted greater control of water status. Seed DT as an important adaptive trait, that allow plants to survive long periods of lack of water until favorable conditions are present for germination, is probably part of the answer to the Darwin's " abominable mystery": Why are the angiosperms so species-rich and ecologically successful?.

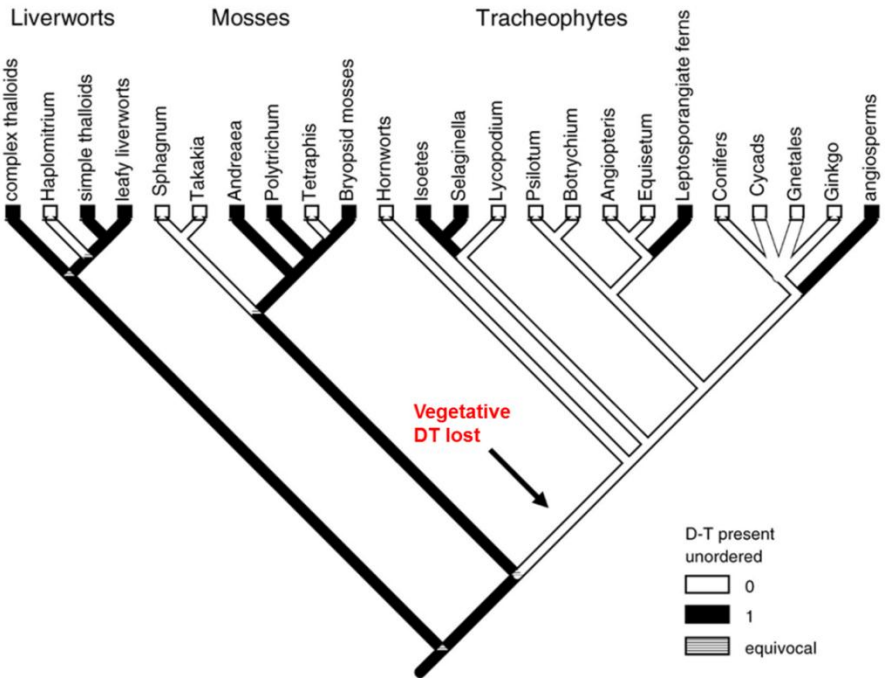

Figure 1. Simplified phylogeny of the land plants. The figure shows a reconstruction of presence of vegetative desiccation tolerance (in at least some members of indicated clade) in black, non-tolerance in white, and equivocal reconstruction by hatching. The arrow marks where the ancestral condition can unequivocally be reconstructed as non-tolerant (Oliver et al. 2005).

1.4. Desiccation tolerance in vegetative tissues from Angiosperms

DT vegetative tissue occurs in thirteen unrelated angiosperm families, suggesting that vegetative DT in angiosperms subsequently re-evolved independently at least thirteen times as an adaptation of seed DT. To answer the question, why was it an evolutionary advantage to re-introduce vegetative DT, instead of surviving unfavorable conditions as a seed, still remains a question.

Table 1. The angiosperm families that contain species with DT vegetative tissue and the number of genera in each family with one or more desiccation-tolerant species(Gaff and Oliver 2013).

Family	Number of genera with DT species	Number of all genera in the family (prone to taxonomic revisions)
Dicotyledoneae		
Cactaceae	1	100
Gesneriaceae	9	133
Lamiaceae	1	212
Myrothamnaceae	1[a]	1
Ranunculaceae	2	51
Scrophulariaceae	4	292
Stylidiaceae	1	5
Tamaricaceae	1	4
Monocotyledoneae		
Anthericaceae	1	29
Cyperaceae	7	102
Philydraceae	1	3
Poaceae	9	657
Velloziaceae	6[b]	10

[a]The two species in the genus and in the family have DT foliage.
[b]All tested species in the family have DT foliage.

DT in seeds and vegetative tissues of angiosperms is different from that in extant lower orders. In the former, it is based on induction of several relatively complex protection mechanisms during drying, with minimal reliance on repair of desiccation-induced damage during rehydration. In the lower order plants, DT is

constitutive, less dependent on complex cellular protection and more on repair during rehydration. Consequently, although DT in seeds appears to include some mechanisms present in DT vegetative tissues of less-complex plants, additional, more complex mechanisms of protection have evolved. Whereas DT is developmentally regulated in seeds with the putative mechanisms of tolerance being accumulated only at precise times after fertilization, drought is stochastic and DT vegetative tissues must respond to environmental signals to activate protective mechanisms for the whole plant. It is possible that environmental signals for extreme water loss activate an existing repertoire of protective seed-specific genes in vegetative tissue(Gaff and Oliver 2013). A hypothesis by Iling et al., 2005 is that DT evolved from the response of desiccation-sensitive plants to abiotic stresses such as cold, salt and drought . Desiccation sensitive plants use an interconnected signaling network to activate a common repertoire of responses to abiotic stress. These responses appear to overlap with those described for DT plants during extreme water loss as they include the accumulation of anti-oxidants and anti-oxidant enzymes (Illing et al. 2005). Hence, it has been suggested that there should be significant overlap between genes that are induced in response to abiotic stresses and DT .

1.5. Desiccation tolerance in seeds

The DT of seeds can vary greatly between and within species and between tissues within individuals; and there is a continuum degree of tolerance across species. Most of angiosperms (~95% species) produce seeds that tolerate desiccation and long-term dry storage traditionally known as “orthodox”. Most cultivated crops, such

rice, wheat, corn, barley, soybean and beans produce desiccation-tolerant seeds (Kermode 2002). Additionally, DT allows seed to remain viable for centuries or even thousands of years, for example, it has been documented the germination of 1300-years-old *Nelumbo nucifera* seeds (Shen-Miller et al. 2002) and 1900-years-old *Phoenix dactilifera L.* seeds (Sallon et al. 2008). However, a significant number of wild species, particularly from wet climate areas, produce desiccation-sensitive seeds. Desiccation-sensitive seeds are known as "recalcitrant" ~ 4% of angiosperms do not tolerate drying and are hardly storable(Gaff and Oliver 2013). Consequently, the use and conservation of recalcitrant-seed species which include some economically important crops such as avocado, cacao, banana, coconut, mango and coffee, remain a challenge (Kermode 2002).

Table 2. The numbers of Spermatophyta species with DT or with desiccation-sensitive seed and the proportions of the total number categorized in the list of (Gaff and Oliver 2013).

Seed type	No. of species listed in category	Category as % of the total species
Recalcitrant seed (desiccation-sensitive)	281	4%
Orthodox seed (desiccation-tolerant)	6703	95%

1.6. Seed development

Seed development in higher plants begins with a double fertilization process that occurs within the ovule and ends with a dormant seed primed to become the next plant generation and consists of two main phases, embryogenesis and maturation. Embryogenesis comprises tissue specification and patterning, which is obtained via a well-organized series of cell divisions and cell differentiation. In *Arabidopsis thaliana* the complete seed development cycle is of about 20 days. The embryogenesis stage takes place until day 7 after flowering, total lipid content is low, in contrast with soluble sugars and starch. Starch level increasing sharply at 6 days after flowering (DAF). At 7 DAF, the heart shape stage was completed and the embryo takes the shape of a torpedo (Figure 2A and 2C)(Baud et al. 2002).

After completion of embryogenesis, seed development switches to the maturation phase that can be divided into early and late maturation. In this stage the seed accumulates storage compounds that include both the synthesis of nitrogen (proteins) and carbon storage compounds (lipids and carbohydrates). Proteins are stored in protein storage vacuoles, and lipids in oil bodies, both fill up the cells and offer resistance against cellular collapse upon drying (Figure 2C) (Angelovici et al. 2010)

During early maturation, the embryo is characterized by a period of cellular elongation and differentiation. At the end of this first period, the embryo occupies all the space in-between the integuments and the overall composition of the seed is similar to that of the mature seed. In *Arabidopsis*, the early maturation phase takes place from 7 to 16 DAF(Figure 2A) (Baud et al. 2002). During this stage, the

seed dry weight increases sharply to reach a maximum at 16 DAF, while a regular loss of water occurrs (Figure 2C). This maturation phase can be conveniently divided into two periods. Between 7 and 12 DAF, the starch concentration declines markedly and the synthesis of storage products, TAG and proteins, began. A second period of the early maturation ranges from 12 to 16 DAF which is characterized by the onset of sucrose storage. After a short lag phase, the storage compound synthesis continued, but the seed global composition does not evolve markedly, with equal and unchanged proportions of oil, protein (approximately 40% of each compound), and a stable fatty acid composition(Figure 2A y 2C) (Baud et al. 2002).

During the last stage (late maturation), the embryo becomes metabolically quiescent and tolerant to desiccation, seeds dry out while considerable changes occur at both transcriptome and metabolome levels. In *Arabidopsis* seeds this stage comprise between 17 and 20 DAF (Figure 2A), the water content in seeds declines sharply from 32% to less than 10% (Figure 2C). This drastic loss of water led to the rupture of the trophic connections between the plant and quiescent seeds. Synthesis continues, with raffinose and stachyose being specifically accumulated during this late maturation stage, as well as some proteins. Despite of a severe dehydration state, metabolic activities persisted in the embryo (Figure 2C) (Baud et al. 2002; Angelovici et al. 2010).

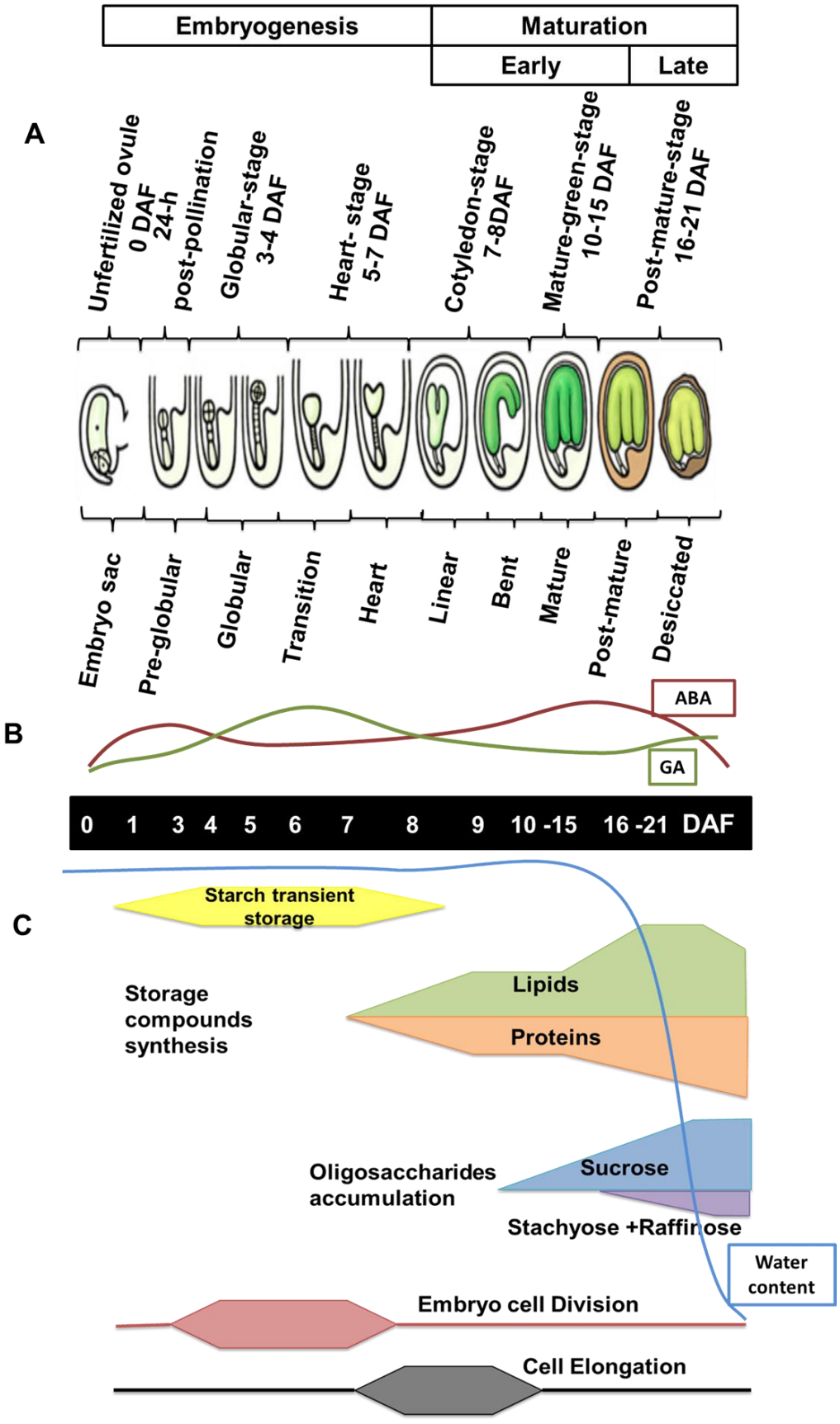
Embryogenesis
Maturation
Early
Late
A
Unfertilized ovule 0 DAF
24-h post-pollination
Globular-stage 3-4 DAF
Heart- stage 5-7 DAF
Cotyledon-stage 7-8DAF
Mature-green-stage 10-15 DAF
Post-mature-stage 16-21 DAF
Embryo sac
Pre-globular
Globular
Transition
Heart
Linear
Bent
Mature
Post-mature
Desiccated
B
ABA
GA
0 1 3 4 5 6 7 8 9 10 -15 16 -21 DAF
C
Starch transient storage
Storage compounds synthesis
Lipids
Proteins
Oligosaccharides accumulation
Sucrose
Stachyose +Raffinose
Water content
Embryo cell Division
Cell Elongation

Figure 2. Seed development. A) Schematic representation of Arabidopsis seed development, cartoons were adapted from (Le et al. 2010) and are not drawn to scale. Double fertilization in angiosperms involves two phases, embryogenesis and maturation. During the embryogenesis phase, the basic body plan is established and generates the different morphological domains of the embryo, the embryonic tissue and organ systems. After transition to the maturation phase, the embryo undergoes typical seed filling, growth arrest, acquisition of desiccation tolerance and entry into quiescence. B) Balance of ABA/GA ratio. C) Major reserves synthesized and accumulated during maturation phase include starch, storage proteins and lipids.

Approximately one third of the *Arabidopsis* genome (6,963 genes) was identified as regulated at the maturation stage. (Fait et al. 2006; Angelovici et al. 2010). Metabolic and transcriptomic changes at this stage coincide with a gradual increase in seed longevity (Verdier et al. 2013). Several genes which transcription is activated during seed maturation are involved in a number of mechanisms that influence seed survival in the dry state such as the production of late embryogenesis abundant (LEA) proteins (Manfre et al. 2009; Delahaie et al. 2013), heat shock proteins (HSPs) (Wehmeyer and Vierling 2000; Prieto-Dapena et al. 2008), oligosaccharides (Baud et al., 2002; Verdier et al., 2013), the synthesis of enzymes that participate in the scavenging of ROS such as catalases (CAT) and superoxide dismutase (SOD) and glutathione (GSH) reductases(Bailly 2004), as well as the antioxidants metabolites, tocopherols (Mene-Saffrane et al. 2010) and flavonoids present in the seed coat (Chen et al. 2012).

1.7. Mechanisms of desiccation tolerance in seeds

1.7.1. Late embryogenesis abundant proteins

Late embryogenesis abundant (LEA) proteins were first identified in seeds, when mRNA and proteins that occur in large quantities were investigated (Dure and Chlan 1981). It has been proposed that LEA proteins could protect cell structures, membranes and other proteins by acting as a hydration buffer, sequestering ions and renaturing unfolded proteins, however, just some of these hypotheses have been tested. Depending on their identity, LEA proteins can be located in the cytoplasm, nucleus, mitochondria, vacuoles, near the cellular membrane, as well as in amyloplasts, but primarily they accumulate in the cytoplasm and nucleus (Battaglia et al. 2008). After synthesis they are transported to cell organelles and membranes, where they stabilize cell structures and molecules. The partition of LEA proteins into various cell compartments determines their possible function in the protection or regulation of essential biochemical processes.

LEA proteins are a group of proteins that have been associated with plant responses to water deficit. They accumulate principally in dry seeds but also in vegetative tissues when plants experience water deficit such as drought, cold and high salinity and are classified into seven groups (Cuevas-Velazquez et al. 2014). The classification is based on sequence similarity and biochemical properties. Proteins of group 1 contain a 20 amino acid region, which was first identified in the wheat named “Em protein”. The role of LEAs from this group has been demonstrate in *Arabidopsis* mutants with a T-DNA insertion allele in the *AtEm6,* which was found to alter the processes of seed dehydration and maturation, which

suggests that ATEM6 protein might be required for normal seed development and the acquisition of desiccation tolerance (Manfre et al. 2006; Manfre et al. 2009).

LEA proteins of group 2 (D11) also known as dehydrins have been widely analyzed regarding of water stress. To date they have been found only in plants (Cuevas-Velazquez et al. 2014).

LEA proteins of group 3 are involved in response to cold stress. Usually they are located in the cytoplasm, nucleus and mitochondria, where they can be associated with membranes. Interestingly, proteins similar to plant group 3 LEA proteins accumulate in several non-plant organisms in response to dehydration, in fungi, microbial and animal kingdoms(Battaglia et al. 2008).

LEA proteins of group 4 are of widespread occurrence in the plant kingdom, including nonvascular plants and vascular plants and are possibly involved in the protection of membrane stability, the microarray data analysis indicates that group 4 LEA homologues in many plant species accumulate under drought in shoot meristems and in developing and dry seeds (Kalemba and Pukacka 2007).

LEA proteins of group 5 were first described from cotton. They represent an atypical LEA subgroup because they contain a higher proportion of hydrophobic residues. These proteins are not soluble after boiling, suggesting that they adopt a globular conformation(Battaglia et al. 2008).

LEA proteins of group 6 (PvLEA18) were first described from bean. To date, 36 genes of this family have been described from vascular plants. The proteins in this

group are characterized by their small size (approximately 7 - 14 kD). A distinctive characteristic of thegroup 6 LEA protein characterized is its accumulation in the elongation or growing regions of common bean roots and hypocotyls, which showed lower water potentials than the non-growing regions (Amara et al. 2014)

LEA proteins of group 7 (ASR proteins) are small, heat-stable, and intrinsically unstructured proteins. The group of this gene has been identified from various species of dicotyledonous and monocotyledonous plants as well as from gymnosperm species. However, no ASR-like genes are found in *Arabidopsis*. They share physiochemical properties with other LEA proteins and they accumulate in seeds during late embryogenesis and in response to water-limiting conditions(Battaglia et al. 2008).

1.7.2. Heat shock proteins

Heat shock proteins (HSP) and other stress proteins are known to protect cells against deleterious effects of stress. The major HSP synthesized by eukaryote cell have different molecular weights and belong to six structurally distinct classes: HSP100, HSP90, HSP70, HSP60 (or chaperonins), ~ 17-30 kDa small HSP (sHSP) and ubiquitin (8,5 kDA) (Al-Whaibi 2011).

In plants, sHSPs are generally undetectable in vegetative tissues in the absence of stress, but are among the most abundant proteins synthesized in response to high temperature. The plant sHSPs can be divided into five nuclear-encoded gene families based on DNA sequence analysis, immunological cross-reactivity, and intracellular localization(Kalemba and Pukacka 2007). There are two classes of

sHSPs that localize to the cytosol (classes I and II), and distinct classes of organelle-localized sHSPs found in the endoplasmic reticulum, the mitochondrion, or the chloroplast. Although *in vivo* data are still lacking, experiments *in vitro* suggest that cytosolic sHSPs function as molecular chaperones by preventing the thermal aggregation of substrate proteins and facilitating their subsequent reactivation (Al-Whaibi 2011).

It is now well established that in addition to being synthesized in response to stress, sHSPs are also expressed during specific stages of plant development. Induction of sHSPs in the absence of stress has been seen in a variety of plant species at several different developmental stages. sHSP expression during seed development is the most extensively characterized example of this non-stress regulation. In *Arabidopsis* embryos, cytosolic class I sHSPs begin to accumulate at midmaturation and are abundant throughout the late maturation program and in the dry seed (Wehmeyer et al. 1996). A reporter-gene transcription assay showed that sHPS expression exhibits little tissue specificity but instead spreads throughout the embryo during development until essentially all cells are stained in the mature seeds prior to complete desiccation. The activity of the reporter gene was strongly reduced in *fus3-3*, *lec1-2* and was almost null in *abi3-6*, all desiccation intolerant mutants of *Arabidopsis* seeds (Wehmeyer and Vierling 2000). This suggests an overall protective effect of HSPs in the seed during drying.

1.7.3. Carbohydrates

Changes in carbohydrates also occur during seed development and germination. Generally in seeds, the monosaccharide content decreases and the oligosaccharides content increases, the reverse trend is observed during germination. In a variety of seeds, sucrose, raffinose-family oligosaccharides (RFOs) and monsaccharides contribute between 1 and 12 % of the dry mass. In mature seeds, sucrose contents range between 15 and 90% of the soluble carbohydrates and monosaccharides present there in trace amounts (Baud et al. 2002). Correlation between high concentration of nonreducing sugars and desiccation tolerance of seeds has led to the suggestion that they may play a role in the mechanism of desiccation tolerance. There are several ways in which sugars have been proposed to act in conferring tolerance (Kermode 2002).

The “water replacement hypothesis”, suggest that hydroxyl groups of sugars substitute for water and provide the required hydrophilic interaction form membrane and protein stabilization(Crowe 2002). In artificial bilayer systems, the sugar bind to the polar head groups, separating the individual lipid molecules and weakening the Van der Walls attraction among the fatty acyl chains. This results in the maintenance of the bilayer liquid-crystalline structure of the membrane even at very low hydration levels (Crowe et al. 1987). The disaccharide trehalose appears to be most effective at maintaining conformation after desiccation, perhaps because it is sterically compatible with the arrangement of phosphatidylcholine head groups in lipid bilayers. Trehalose does not occur in most angiosperms seed, but the high sucrose content in many of these seeds has led to the suggestion that

sucrose can also supply the hydrogen bonding required to prevent lipid phase transitions. The protective effects of sucrose during drying may be obtained through nonspecific interaction between sugars, and/or polar lipids so that the interference between the two phases is sufficiently stable (Crowe 2002).

The “glass formation hypothesis”, sets down that ,a glass is a thermodynamically unstable solid state with an extremely high viscosity characterized by the glass-to-liquid transition temperature which depends on water content, temperature and its chemical composition(Buitink et al. 2000). Carbohydrates in particular have the susceptibility to form a glass by hydrogen bonding interactions at an appropriate temperature and water content. Tri- and tetra-saccharides such as raffinose and stachyose often occur in considerable quantities in dry seeds of many plant species. The presence and amount of these oligosaccharides have been found to correlate with DT and longevity. Oligosaccharides are thought to contribute to the stabilization of intracellular glasses by increasing viscosity and the glass-to-liquid transition temperature (Buitink et al. 2000).

1.7.4. Antioxidant system

During dry storage, seeds are subjected to progressive damage of cell constituents (proteins, lipids, nucleic acids, sugars, etc.) by autooxidation processes such as Amadori and Maillard reactions, lipid peroxidation or protein carbonylation(Kranner and Birtic 2005). The seeds require efficient antioxidant systems which play a key role in acquisition of DT of developing seeds, completion of seed germination and seed storability. There are a range of protective mechanisms that prevent

excessive oxidation of macromolecules, mechanisms including enzymatic ROS and non-enzymatic ROS scavenging systems detoxification (Sano et al. 2016).

ROS scavenging enzymes include superoxide dismutase (SOD), ascorbate peroxidase (AP) and other peroxidases, mono- and dehydroascorbate reductases, glutathione reductase (GR) and catalase (CAT) (Bailly 2004). Several factors involved in redox changes have been proposed as seed viability markers. Although, several other proteins that act as antioxidants or related to ROS signaling, such as thioredoxins, peroxiredoxins and glutaredoxins, have been identified in seeds, their roles in seed longevity remain to be described (Sano et al. 2016). In bean seeds acquisition of DT seems to be associated with a reorientation of the enzymatic antioxidant defense systems. Dried, mature, desiccation-tolerant seeds display high CAT and GR activities and low SOD and AP activities, whereas the reverse is the case in immature, desiccation-intolerant seeds. Interestingly, it has also been demonstrated that desiccation of developing sunflower seeds is associated with an increase in CAT activity, thus leading to decreased H_2O_2 content and lipid peroxidation damage (Bailly 2004).

The amount of non-enzymatic ROS scavenging systems is controlled during seed development and maturation, and can be influenced by the growth environment of the mother plant. Non-enzymatic ROS scavenging systems involve low molecular weight antioxidants including GSH, tocopherols (Vitamin E), ascorbate (Vitamin C), and flavonoids. Seed aging is correlated with a decline in the cellular antioxidant potential. The GSH system probably plays a major role. ROS oxidize GSH to its dimer GSSG, which accumulates during seed storage(Sano et al. 2016) .In

addition, the lipophilic antioxidant tocopherol, which protects membranes from lipid peroxidation, is essential for seed longevity and germination characteristics, as it was shown in *vte1* mutants from *Arabidopsis* (Mene-Saffrane et al. 2010). Ascorbate, is only present in small amounts in dry seeds and therefore may play a minor role in regulating the redox situation in the dry state (Weitbrecht et al. 2011). Antioxidant polyphenols such as flavonoids that are present not only in the seed coat, but also in the embryo and endosperm, have been shown to increase seed longevity of *Arabidopsis* seeds (Debeaujon et al. 2000).

1.8. Regulation of seed desiccation tolerance

The expression of protective mechanisms can be observed during both early and late seed maturation. In *Arabidopsis*, embryo development and seed maturation, including the acquisition of DT, is orchestrated by a set of four master regulators: LEAFY COTYLEDON 1 (LEC1), a CCAAT-box binding factor, and three B3 domain-containing proteins(Lotan et al. 1998) ABSCISIC ACID INSENSITIVE 3 (ABI3), FUSCA 3 (FUS3) and LEC2. In addition to controlling embryo formation and seed maturation, these master regulators also repress the expression of genes required for the transition from embryonic to vegetative development (Nambara et al. 1992; Giraudat et al. 1992; Luerßen et al. 1998; Stone et al. 2001). Although the role of these master regulators during seed maturation is globally similar, some of their functions are very specific. For example, in contrast to mutations in *LEC1*, *ABI3* or *FUS3* that drastically affect DT(Keith et al. 1994; Meinke et al. 1994), *lec2* mutants do not present this effect (Meinke et al. 1994; To et al. 2006). Interestingly, ectopic expression of *LEC1*, *FUS3* or *ABI3* in single or double mutant backgrounds

of the other two regulators activated some processes of seed maturation, such as lipid and seed storage protein accumulation, but not DT, suggesting that all three regulators are required to activate DT.

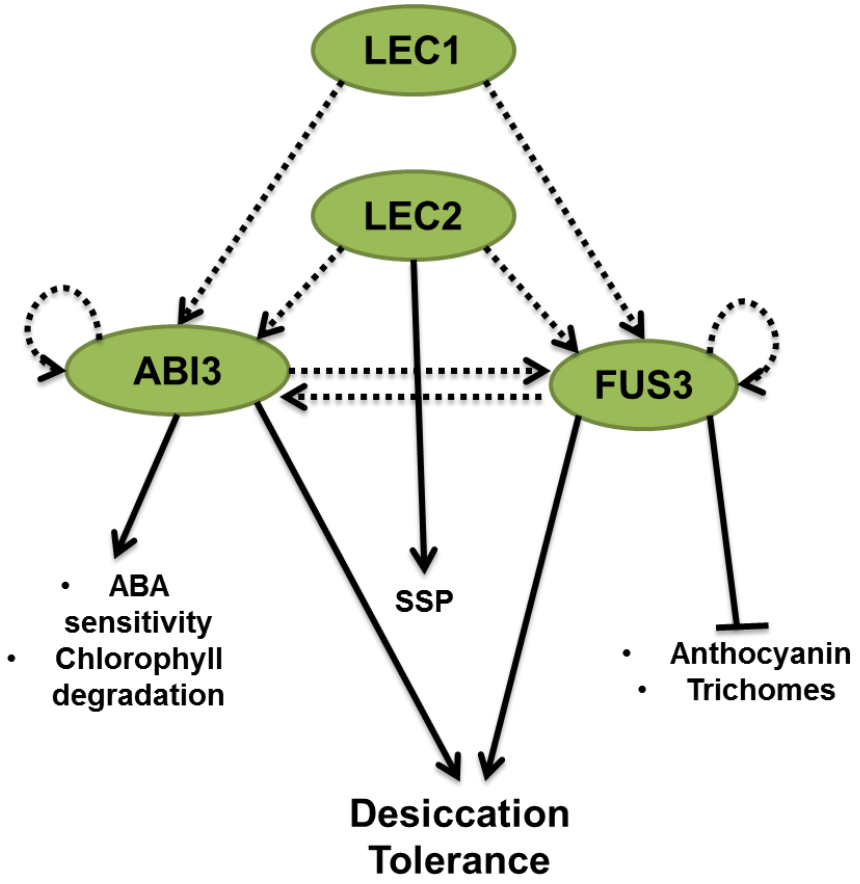

Figure 3. LEC1, FUS3, ABI3 and LEC2 network. Relations between the regulators are depicted as dashed lines, whereas proposed downstream actions are show as solid lines(To et al. 2006).

1.9. Hormonal signaling, role of Abscisic acid

1.9.1. Abscisic acid biosynthesis

Abscisic acid (ABA), derived from epoxycarotenoid cleavage, is the key hormone required as positive regulator of induction of dormancy throughout the process of

seed maturation. Its role also has been shown with reduction of ABA activity in a switch from seed maturation program to a germination program(Finkelstein 2013).

ABA is first produced in maternal tissues and subsequently in the embryo, where its accumulation peaks during maturation(Figure 2B). During the late-maturation phase, dormancy induction is dependent on a strong ABA signal as a high ABA to gibberellin (GA) ratio, which is mainly regulated by the dynamic balance between ABA biosynthesis and degradation(Figure 2B) (Finkelstein 2013).

Most seed ABA biosynthetic genes have been identified and described in seeds, include zeaxanthin epoxidase (ABA1), the 9-cisepoxycarotenoid dioxygenase (NCEDs), short chain dehydrogenase/reductase-like (ABA2/GIN1/SDR1), molybdenum cofactor sulfurase (ABA3) and an enzyme involve in neoxanthin synthesis (ABA4)(Figure 4) (Endo et al. 2014).

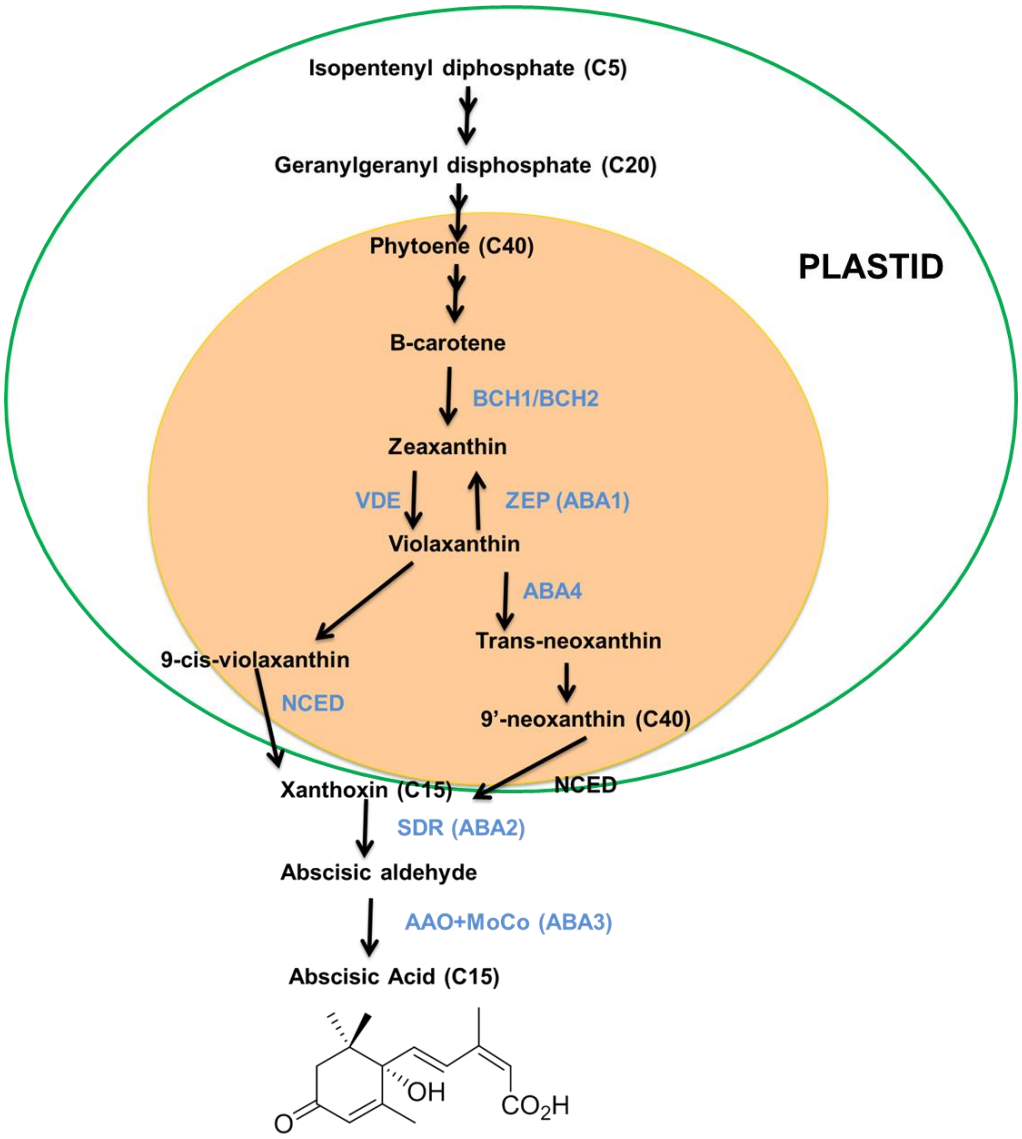

Figure 4. Abscisic acid (ABA) biosynthesis pathway. ABA is synthesized from C40 carotenoid precursors, and the pathway shown is from the xanthophyll zeaxanthin ahead. Enzymes responsible for the different reactions are in blue. AAO, abscisic aldehyde oxidase; MoCo/ABA3, molybdenum co-factor sulfurase; NCED, 9-cis-epoxycarotenoid dioxygenase; SDR/ABA2, short chain dehydrogenase/reductase; ZEP/ ABA1, zeaxanthin epoxidase. Carotenoid intermediates are highlighted in orange.

1.9.2. ABA signaling

There are approximately 100 mutants identified as being involved in ABA signaling which have been associated with developmental processes, including the DT and

induction of dormancy(Cutler et al. 2010), this mutants including the ABA-insensitive mutants *abi1*, *abi2*, *abi3*, *abi4* and *abi5*(Cutler et al. 2010). ABI1 and ABI2 encode type 2C protein phosphatases (PP2Cs) and ABI3, ABI4 and ABI5 are TFs of the B3, APETALA2 (AP2) and basic leucine zipper (bZIP) TF classes, respectively. ABI5 plays an important role in *Arabidopsis* seeds. Dry seeds of *abi5* mutants show reduced transcript levels of ABA-responsive genes, leading to the hypothesis that ABI5 is necessary to bring germinated embryos into a quiescent state under drought stress, thereby protecting young seedlings from the loss of water (Lopez-Molina et al. 2002). ABA regulates ABI5 accumulation and activity during a limited window between 12 and 48 h of germination. The ABI proteins are part of a recently discovered cascade of events involving ABA receptors, protein phosphatases and protein kinases (Cutler et al. 2010).The core of this pathway consists of three protein families: the PYR/PYL/RCAR receptor family consisting of PYRABACTIN RESISTANCE1 (PYR1)-LIKE REGULATORY COMPONENTS OF ABA RECEPTORS, the TYPE 2C PROTEIN PHOSPHATASES (PP2Cs) and the SUCROSE-NON-FERMENTING KINASE1- RELATED PROTEIN KINASE2 (SnRK2) family (Figure 5) (Cutler et al. 2010). Together, these three protein families form a double- negative regulatory pathway. In the absence of ABA, the PP2Cs inactivate SnRK2s by dephosphorylation (Figure 5). Conversely, when ABA is present, it binds to the PYL/ PYR/RCAR receptors, thus creating a complex which interacts with the PP2Cs (Figure 5). Via this interaction, the dephosphorylation of the SnRK2s by the PP2Cs is inhibited. The active kinases subsequently phosphorylate different proteins, including membrane proteins and

TFs (e.g. ABI5), eventually leading to an ABA response (Figure 5) (Cutler et al. 2010).

ABA-responsive element (ABRE) and the RY/Sph sequences are regulatory cis-acting elements required for specific expression of genes during seed maturation. The ABRE is one of the best-studied components of ABA signalling. A typical ABRE contains an ACGT nucleotide core motif which is recognized by bZIP transcription factors. Interestingly, none of the major regulators of seed maturation (ABI3, FUS3, LEC1 and LEC2) correspond to a transcription factor that interacts with the ABRE boxes (Figure 5). However, the B3 domain, which is present in ABI3, FUS3 and LEC2 proteins, binds the RY element .In combination with closely associated ABRE like elements, this module acts as an enhancer of seed-specific transcription. The Arabidopsis ABI5 bZIP transcription factor binds to ABRE and activates ABA-mediated transcription in seeds. Promoters of genes down-regulated in *abi5* dry seeds show an over-representation of t ABA regulates ABI5 accumulation and activity during a limited window between 12 and 48 h of germination ABRE, suggesting that ABI5 contributes to the activation of ABRE-mediated transcription. . Evidence has been reported indicating that ABI3, which activates RY element mediated transcription, interacts physically with ABI5, and that this physical interaction appears to be important for the synergistic activation of gene expression by these two TFs (Holdsworth et al. 2008).

Many genes encoding LEA, SSP and HSP proteins were downregulated in this cascade signaling, including AtEM1and AtEM6, which are under the control of the bZIP transcription factor ABI5, a phosphorylation target of SnRK2 kinases (Sano et al. 2016). Aquaporins have essential functions in the transport of water and other

small molecules, such as H_2O_2, and several are regulated by ABA. In seeds, ABI3 is involved in the regulation of genes encoding tonoplast intrinsic proteins TIP3;1 and TIP3;2.Defects in both genes suggest that ABA may control water relations and H_2O_2 accumulation via ABI3 modulation of aquaporins, thereby contributing to seed longevity(Mao and Sun 2015).

Other hormones in addition to ABA are involved in seed developmental processes. In addition to its important role during morphogenesis, IAA affects seed ABA sensitivity during germination. Gibberellin is required to induce seed germination, and mutants defective in gibberellin biosynthesis germinate poorly (Locascio et al. 2014). ABA negatively regulates gibberellin biosynthesis in developing and imbibed seeds, and suppression of gibberellin biosynthesis by ABA is required for the induction and/or maintenance of seed dormancy (Seo et al. 2006).

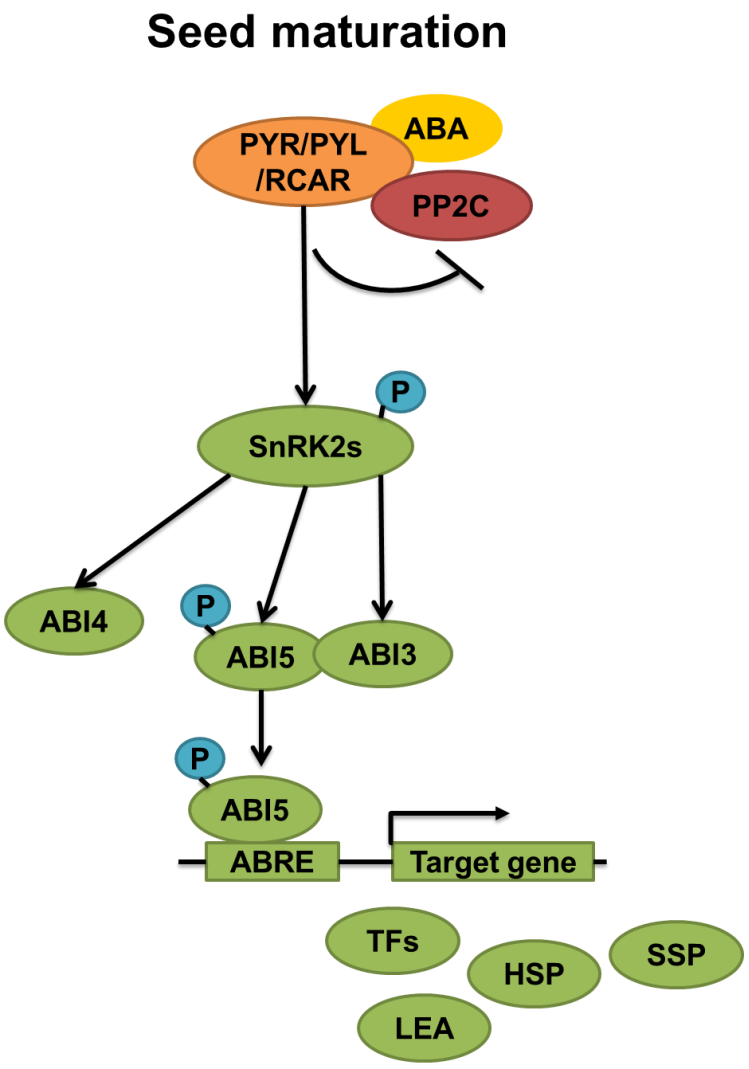

Figure 5. Overview of the molecular events of ABA in seed maturation. ABA accumulation activates PYR/PYL/RCAR ABA receptors to inhibit group A PP2Cs. PP2C inhibition in turn allows SnRK2 activation through autophosphorylation. Active SnRK2s mediate the ABA response through the phosphorylation of downstream targets. In seeds, ABI5 phosphorylation by SnRK2s leads to the inhibition of seedling growth. The phosphorylation of ABI5 activates the transcription of target genes such as Late Embryogenesis Abundant (LEA)-class genes as well as transcription factors (TF) involved in seed maturation. Positive regulators of the ABA signaling pathway are shown in green, whereas negative regulators are shown in red.

2. Hypothesis

There is a common subsets of genes that fail to be activated in desiccation intolerant mutants when compared to their expression patterns in tolerant lines, that should include most of the genes that are essential for the acquisition of seed DT, and that their activation patterns should allow the construction of regulatory networks to identify TFs that constitute the main nodes controlling the DT regulatory subnetworks.

3. Objectives

Main Objective

Identify molecular components involved in seed desiccation tolerance of Arabidopsis thaliana

Specific Objectives

- Establish the conditions and stages of development suitable for analysis of desiccation tolerance
- Identify genes differentially expressed in tolerant and intolerant seeds to desiccation at various stages of development
- Identify differential metabolites in tolerant and intolerant seeds to desiccation at various stages of development
- Identify the processes involved with desiccation tolerance in seeds
- Evaluate some mutants that affect in desiccation tolerance

4. Material and Methods

4.1. Plant materials and growth conditions

Arabidopsis thaliana ecotype Col-0, Ler and Ws-2 were used as wild type in this study. The following mutant lines were used: *lec1-1*, *lec2-1*(Meinke et al. 1994), in Ws-2 background; *abi3-1* and *abi3-5* (Ooms et al. 1993) in Ler background; *fus3-3*(Keith et al. 1994) in Col-0 background. Plants were grown in a sterile mix of vermiculite and soil in a growth chamber at 22 °C with a 16-h photoperiod at 200 mmol m^2 s^{-1}.

For RNAseq and carbohydrate profiles, flowers were marked and at specific times after flowering (15, 17 and 21 DAF), siliques were harvested from 24 plants and seeds collected and immediately frozen in liquid nitrogen, and stored at -80°C. All samples were collected in triplicates

Arabidopsis T-DNA insertion mutant lines were obtained from the Nottingham Arabidopsis Stock Centre (Table 3). Homozygous lines were identified by PCR genotyping to corroborate site of insertion and identify homozygous plants (Table 4).

Table 3. T-DNA insertion seeds (SALK Line) examined in this work.

Locus	Aliases	Description	SALK lines
AT1G21000	PLATZ1	PLATZ transcription factor family protein	SALK_005374
AT1G76590	PLATZ2	PLATZ transcription factor family protein	SALK_016183
AT4G26050	PIRL8	plant intracellular ras group-related LRR 8	SALK_001827
AT1G01720	ATAF1	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	SALK_067648
AT2G47180	GOLS1	galactinol synthase 1	SALK_013103 SALK_121059

AT1G56600	GOLS2	galactinol synthase 2	SALK_075769 SALK_101144
AT1G77950	AGL67	AGAMOUS-like 67	SALK_050367 SALK_013790
AT5G01670	NRS1	NAD(P)-linked oxidoreductase superfamily protein	SALK_150592 SALK_111609
AT4G01970	STS	stachyose synthase	SALK_021300
AT1G01250	ERF23	Integrase-type DNA-binding superfamily protein	SALK_044761
AT1G75490	DREB2D	Integrase-type DNA-binding superfamily protein	SALK_023056
AT3G54510	ERD4	Early-responsive to dehydration stress protein (ERD4)	SALK_055548
AT4G04840	MSRB6	methionine sulfoxide reductase B6	SALK_039712
AT2G19320		unknown protein	SALK_092667
AT5G54070	HSF9	heat shock transcription factor A9	SALK_063950
AT1G05510	OBAP1A	OBAP1A, oil body-associated protein 1A	SALK_017397 SALK_011689

Table 4. Primers used to SALK T-DNA genotyping.

Primers	5´-3'
Genotyping SALK lines	
SALK_005374	Fw TGATTGTGGTTTTCGATAGCC
	Rv CACGTTAAGTAACGTGGCAGC
SALK_016183	Fw ATCATGAGCATTTCAAATGCC
	Rv GAATGAAGCGTGATCCAAGTC
SALK_001827	Fw ACCATTTGATTCGGCCTATTC
	Rv TTTGAGTTGACGGAGAGCTTC
SALK_067648	Fw TCCCAGGGACAGAAAATATCC
	Rv AAATATTAAATTGATTGCGGCAC
SALK_013103	Fw TTCGAAACAAAAATTGAACCG
	Rv TTGTTTGACCTACCAGATGG
SALK_121059	Fw ATGAATGTATGTTTCGCAGGC
	Rv TTGTTTGACCTACCAGATGG
SALK_075769	Fw TCATAGACTTGATTGGTTTCCG
	Rv AACAACACAGCCTTGATCCAC
SALK_101144	Fw TCATAGACTTGATTGGTTTCCG
	Rv AACAACACAGCCTTGATCCAC
SALK_050367	Fw AAAGGTATCGGATGTTTTCGG
	Rv TTGGTAACAGCTTATGGACCC
SALK_013790	Fw TACAGTCCAGAAGCCACCAAC
	Rv ATTTCTCGTGGCAACGTTATG
LB3.1	ATTTTGCCGATTTCGGAAC

Plants were grown in the same conditions as described above for the production of seed. To determine the phenotypes of the T-DNA insertion lines, experiments were performed on three biological replicates of 100 seeds, harvested from a pool of eight plants. The seed storage conditions were 25°C with 10% of relative humidity (RH) for 1, 2, 4 and 8 weeks. Viability of the dry seeds was determined by germination on 0.1X MS media after imbibition of seeds for 2d at 4°C to remove dormancy. To determine ABA sensitivity, seeds were germinated in MS medium in the presence of different concentrations of ABA (mixed isomers; Sigma-Aldrich) 0, 3, 5, 7 and 10 µM.

4.2. RNAseq library preparation and sequencing analysis

Total RNA was isolated using Concert Plant RNA Purification reagent (Invitrogen) and then re-purified with the TRIZOL reagent (Invitrogen). To ensure high quality RNA samples, RNeasy MinElute Cleanup kit (Qiagen) was used following the manufacturer's instructions. First and second strand cDNA synthesis was performed using 3 µg of the total RNA mixture using Message Amp-II kit (Ambion) following the manufacturers instruction. 10-12 ng of the synthesized cDNA was transcribed by *in vitro* transcription and the resulting 70-90 µg of antisense RNA (aRNA) were purified using RNAeasy columns (Qiagen). A second round of cDNA synthesis was performed using 20 µg of mRNA as template. cDNA synthesis was performed as described above except that random primers (mostly hexamers) were used for first strand synthesis. This procedure yielded approximately 10 µg of cDNA that was purified using the DNA Clear Kit for cDNA purification (Ambion). Samples were barcoded for multiplexing using the SOLiD Barcoding Kit. Libraries

were sequenced for 50 bps on the SOLID 4 platform (Applied Biosystems). The raw data have been deposited in the Gene Expression Omnibus (GEO) database accession number GSE76015.

4.3. Transcriptome data analysis

The SOLiD BioScope Whole Transcriptome Analysis (WTA) version 1.2.1 pipeline for single reads was used to align the reads to the TAIR10 Col-0 reference genome available at The Arabidopsis Information Resource (Figure 6A). Two perfect matches per location were allowed aligned reads per exon were counted, and per base coverage was calculated. The WTA pipeline CountTags module provides normalize RPKM (reads per kilobase of exon sequence, per million reads) values along with read counts per exon(Figure 6B). Custom bash and MySQL scripts were then used to calculate the number of counts per gene model.

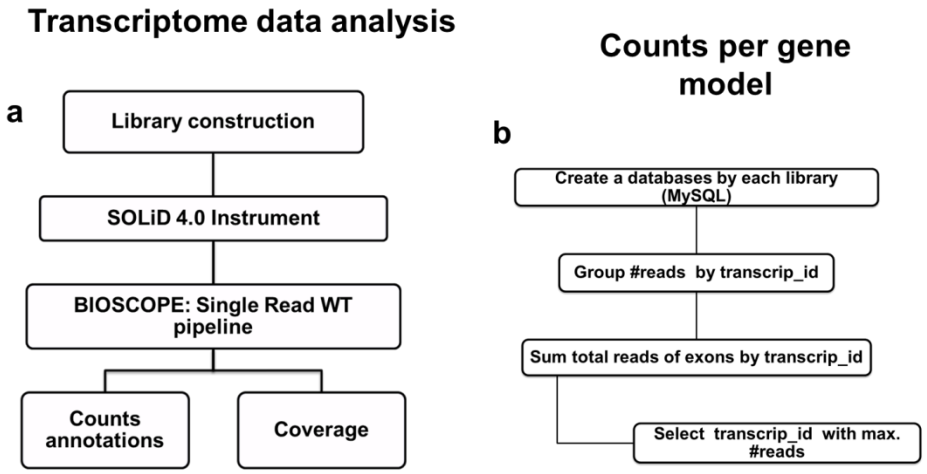

Figure 6. Pipeline of transcriptome data analysis. (a) General strategy. (b) Strategy to obtain counts by gene

4.4. Statistical analyses

Gene counts were normalized using edgeR's (v2.9.16)(McCarthy et al. 2012) TMM [trimmed mean M (¼ log fold-change gene expression)] algorithm. For the analysis of differentially expressed genes only genes with at least 5 reads across all samples were included. For the tolerance designation, we tested for differential expression of genes using the multifactor generalized linear models (*glms*) approach in edgeR. We fit negative binomial *glms* with Cox-Reid tagwise dispersion estimates to models that included intolerant and tolerant lines and mutants as factors.

Table 5. Generalized linear model fitted to desiccation tolerance seeds data.

Comparison Model
[(*lec1-1* - Ws) - (*fus3-3* - Col-0) - (*abi3-5* - Ler] vs [(*lec2-1* - Ws) - (*abi3-1*- Ler)]
[(*lec1-1* - Ws) - (*fus3-3* - Col-0) - (*abi3-5* - Ler] vs [(*lec2-1* - Ws) - (*abi3-1*- Ler)]
[(*lec1-1* - Ws) - (*fus3-3* - Col-0) - (*abi3-5* - Ler] vs [(*lec2-1* - Ws) - (*abi3-1*- Ler)]

To determine differential expression, we performed likelihood ratio tests by dropping one coefficient from the design matrix (i.e., the "null" model) and

comparing it to the full model (Table 5) (McCarthy et al. 2012). Differentially expressed genes (DEGs) between samples were considered based on a 5% false discovery rate (FDR) cutoff.

4.5. Categorization and functional analysis

GO annotation analysis on gene clusters was performed using the BiNGO 2.3 plugin tool in Cytoscape version 2.6 with GO_full and GO_slim categories, as described in (Maere et al. 2005). Over-represented GO_Full categories were identified using a hypergeometric test with a significance threshold of 0.05 after a Benjamini and Hochberg FDR correction(Benjamini and Hochberg 1995).

4.6. Network analysis

4.6.1. Microarray data

A list of all microarray experiments using the Affymetrix GeneChip ATH1 was downloaded from the EBI ArrayExpress database. The CEL files were manipulated as previously described (Chavez Montes et al. 2014). Also, in order to obtain a high quality, homogeneous dataset, the arrayQualityMetrics Bioconductor package was ran on each experiment and low quality CEL files were excluded from further analysis. To avoid possible perturbations of the underlying gene regulatory network, all CEL files corresponding to transgenic samples (mutants, overexpressions, promoter constructions) were excluded. This resulted in 168 CEL files that were normalized using gcRMA under R. CEL files were transformed as previously described (Chavez Montes et al. 2014).

For TFs-only networks, the selected seed CEL files were transformed to ASCII format using the celutil utility and the ATH1 array name was replaced with a custom name. A TFs-only CDF file was created using a modified version of the XSpecies use_ME.pl Perl script and the Affymetrix ATH1 probe_ tab file was renamed to match the custom CDF name. Both files were packaged for R using the makecdfenv and AnnotationDbi packages, respectively, which allowed us to normalize the modified CEL files with gcRMA. The resulting normalized data was used as input for the ARACNe algorithm using the previously reported ARACNe pipeline and list of transcription factors (Chavez Montes et al. 2014).

4.6.2. Transcription factors list

The -e or Data Processing Inequality ARACNe parameter uses a list of TF in order to preserve interactions including one or more TFs. There are three Arabidopsis TF databases, Agris, RARTF and DATF. We further added all AGI IDs from the TAIR10 ATH_GO_GOSLIM file that were annotated with the Gene Ontology entry GO:0003700, “sequence-specific DNA binding transcription factor activity. The final TF list contains 2575 AGI IDs corresponding to 2088 probesets.

4.6.3. Network inference using ARACNe

Normalized data was used to calculate the config_kernel.txt and config_threshold.txt parameters required by ARACNe using the author provided Matlab scripts. Interactions were inferred for the 2088 TF probesets using the Linux command-line ARACNe 32 bit program at three DPI values, 0.0, 0.1 and 0.2, with the 2088 probeset list as the -l parameter for the complete dataset or without the -l

parameter for the TF-only dataset. The command-line execution of the ARACNe program was in the form: aracne –H config_parameters -i normalized_ data [−l TF-list] -e 0.0/0.1/0.2 -o output_file.

4.6.4. Adjacency files transformation

The TAIR10 array_elements and aliases tables were combined in order to obtain a single table linking probesets to AGI IDs and, when available, their corresponding symbols.The resulting combined table was used to transform the ARACNe output adjacency files to Cytoscape compatible, tab-delimited tables using a custom Perl script.

4.7. Carbohydrate analysis

50 µg of seeds from 15, 17 and 21 DAF were ground in liquid nitrogen, and soluble sugars extracted twice for 15 min at 75°C with metanol 80% (v/v); then samples were centrifuged at 5000 rpm for 10 min. The supernatant was dried under vacuum and dissolved in acetonitrile:water (1:1). Carbohydrates were quantified on an LC-MS system composed of the ACQUITY ultra-performance liquid chromatography (UPLC) system (Waters Corporation) fitted to a Q-TOF Premier mass spectrometer (LCT XE Premier, Waters). Conditions for chromatography were as follows: column, ACQUITY HILIC BEH column (2.1×100mm i.d., 1.7 µm, Waters); solvent A, ACN:H_2O (90:10) + 0.1% NaOH ; solvent B, ACN:H_2O (30:70) + NaOH t, 0 to 55 min, initial, 100% A and 0% B; 20 min, 40% A and 60% B; 55 min 100% A and 0% B; flow rate, 0.20 mL/min; injection volume 5 µL and column temperature, 35 °C. Conditions for mass spectrometry under negative mode were as follows, capillary voltage, 2.8 kV; cone voltage, 40 V; source temperature, 350°C; desolvation temperature, 100°C; cone gas flow, 30 L/h; desolvation gas flow, 750 L/h; nebulizer

and curtain gas, N_2.The amount of carbohydrates were determined by spectrometer software (MassLynxTM v. 4.1, Micromass) using calibration curves prepared with carbohydrates standards (D-glucose, D-fructose, sucrose, raffinose and stachyose) purchased from Sigma-Aldrich.

4.8. Construction of overexpression lines

For pCaV35S::cDNA constructs, the corresponding coding sequence was amplified with the primers indicated for each gene in Table 6.

Table 6. Primers used to perform construction.

	cDNA (CDS)
PLATZ1	Fw GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGATAAGAACAGAGGAAGAAGAAGA
	Rv GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGAACGGAGCACGGTGAG
PLATZ2	Fw GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGACCGATGATGATGAGAGC
	Rv GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGAAAGGAGCACGGTGTGGGA
AGL67	Fw GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGTCGGGTTAAATTGGAGT
	Rv GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATTTTGGATTTTGGTTCGAA
DREB2G	Fw GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAAGAAGAGCAACCTCCG
	Rv GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGAACCAATTCCATGGATGTTGA
DREB2D	Fw GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACATGTCATCCATAGAGCCAAA
	Rv GGGGACCACTTTGTACAAGAAAGCTGGGTCGTGACACCTCAAAGTGGGGA

PCR products were cloned in pDONR221 and transferred into T-DNA binary vector pFastG02(Shimada et al. 2010). The resultant vectors were used for Agrobacterium-mediated transformation of Arabidopsis Col-0 and *abi3-5* plants using the floral dip transformation method as previously reported (Martinez-Trujillo et al. 2004). Harvested seeds were spread on MS medium containing 20 µM of Phosphinothricin (PPT) for selection of transgenic plants. After selfing, plants with 3:1 segregation rates corresponding to single insertions were selected to produce homozygous lines.

Desiccation tolerance of *abi3-5* TF-overexpressing lines was performed using three biological replicates of 100 seeds, harvested from a pool of 8 plants. Seeds at 25 DAF were harvested and stored at 25°C and 20% RH for 0, 2, 3 and 4 weeks. Viability of the dry seeds was determined by germination on 0.1x MS media after imbibition of seeds for 2d at 4°C to remove dormancy.

4.9. Gene expression analysis

Gene-specific primer pairs (Table 7) designed using the NCBI/ Primer-BLAST tool (Ye et al. 2012) were used for real-time PCR.

Table 7. Primers used to RT PCR analysis.

Real Time PCR	
ABI5	Fw AAACATGCATTGGCGGAGTT Rv TCAATGTCCGCAATCTCCCG
ATAF1	Fw TCAGGCTGGATGATTGGGTT Rv GGCGGAGGCATAACCATCTC
UGE3	Fw GGAGCTCACGAGAGTGGAAG Rv CTAACCGCGCTACCATCCAT
TTP5	Fw GCCAAAGAGCTTATGGAACACC Rv CCTTTGTTCACACCCTGTGG
LEA4	Fw GAGAAACGCGTCAGCACAAC Rv ATCTGATGTGTCCCAGTGCC
ChlADR	Fw CCAATTGAAGTGGCACCATCC Rv CCACAGCTCCTCCATTAGGG
PLATZ2	Fw TCGGCAAAGGAGTCACCAAT Rv TCCTCCAAGTTTGCAACCGA
ABI3	Fw GAACACCGGCGATTTTGTGA Rv TTTTGTCCGCTCGGTTGTCT
GOLS1	Fw GTGTACAATCTCGTCCTTGCG Rv AACCCGCTGCACAGTAGTGA
PLATZ1	Fw TGCAAGCTTGGAGGAATGAGG Rv AGCTTCATCTGATTCCGACCC
UBQ10	Fw GCGTCTTCGTGGTGGTTTCTA Rv GTCGAGTCACTTTGCAGGC
TIP4L	Fw AGTCATGCCAAGCTCATGGT Rv TCAACTGGATACCCTTTCGCA

A total of 10 µg of RNA was reverse transcribed using SuperScript® III Reverse Transcriptase (Life Technologies) according to the manufacturer's instructions.

Reactions were performed with the SYBR Green PCR Master Mix in an ABI 7500 Fast Real-time system. UBQ10 and TIP4L were used as standards for cDNA content normalization. The thermal cycling program was set to 95 °C for 5 min, 40 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min. Results were analyzed using the ABI 7500 on-board software, version 2.0.5 (Applied Biosystems). The real-time PCR was conducted with at least three experimental replicates for each biological sample.

5. Results

5.1. Overview of the experimental comparative strategy

In order to identify important mechanisms related to seed DT, we designed a comparative transcriptomic analysis between seed desiccation intolerant mutants *lec1-1, abi3-5* and *fus3-3*, the desiccation tolerant mutant *lec2-1* and the desiccation tolerant weak allele of *abi3 (abi3-1*) and their respective wild type controls (To et al. 2006) to identify genes that are differentially expressed in the desiccation intolerant mutants respect to tolerant mutants and WT controls. Since *lec1* and *lec2* have similar phenotypes in terms of morphological alterations in cotyledons, reduced accumulation of storage compounds, etc, but differ in desiccation tolerance a comparative analysis should allow the identification of genes that are involved in desiccation tolerance that are not activated in *lec1* respect to *lec2*. To confirm the identity of the mutants obtained from the ABRC stock center, we tested the genotype and phenotype of each mutant, including desiccation tolerance, presence of trichomes in cotyledon and ABA sensitivity (Table 8 and Figure 7).

Table 8. Phenotypes of different lines used

Phenotype	***Lines***					
	wildtype	***fus3-3***	***lec1-1***	***lec2-1***	***abi3-5***	***abi3-1***
Chlorophyll accumulation in dry seed	**NO**	**NO**	**YES**	**YES**	**YES**	**NO**

Anthocyanin in cotyledons	NO	YES	YES	YES	NO	NO
Storage protein expression	Normal	Reduced	Reduced	Reduced	Reduced	Normal
Ectopic trichomes on Cotyledons	NO	YES	YES	YES	NO	NO
Seed ABA sensitivity	Normal	Normal	Normal	Normal	Reduced	Reduced
Desiccation-tolerant seeds	YES	NO	NO	YES	NO	YES
Genetic Backgroud	-	Col-0	Ws	Ws	Ler	Ler

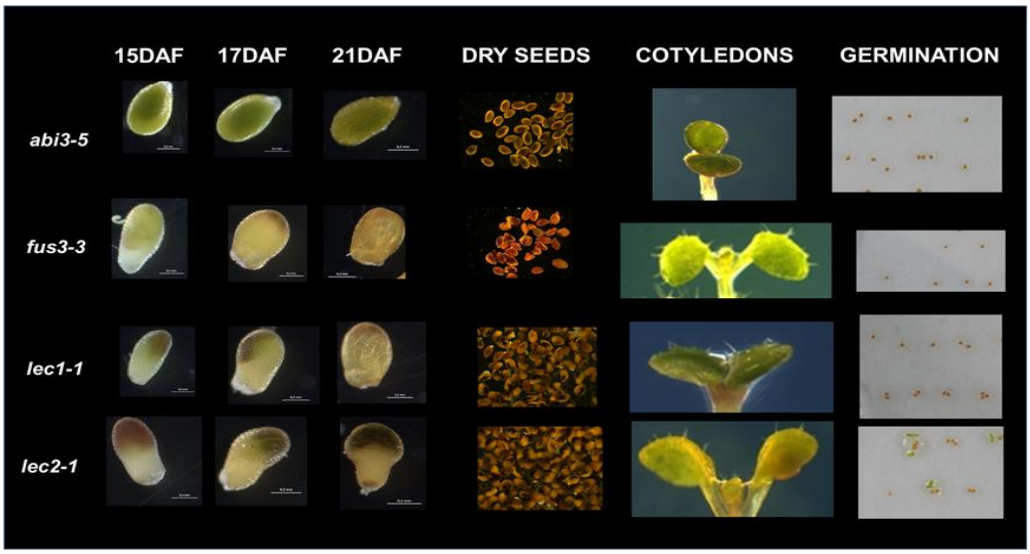

Figure 7. Phenotype of the different lines. Seed phenotypes of each mutant at 15,17 ,21 DAF, dry seeds, trichome cotyledons and germination.

5.2. Global transcriptional analysis of desiccation tolerant and intolerant *Arabidopsis* seeds

To obtain a global view of the transcriptional differences between DT and desiccation intolerant lines during seed maturation, we constructed RNA sequencing (RNAseq) libraries from RNA extracted from each mutant, and their corresponding wild type, at three different developmental stages namely 15 DAF (day after flowering) that corresponds to a developmental stage previous to drastic water loss, 17 DAF where drastic water loss starts and 21 DAF when the seed is completely dried (Figure 7). Since *lec1-1, fus3-3* and *abi3-5* have highly pleiotropic phenotypes, comparisons of their transcriptome at these seed developmental stages with those of *lec2-1* and *abi3-1* should allow the identification of differentially expressed genes related to loss of desiccation tolerance (Figure 8).

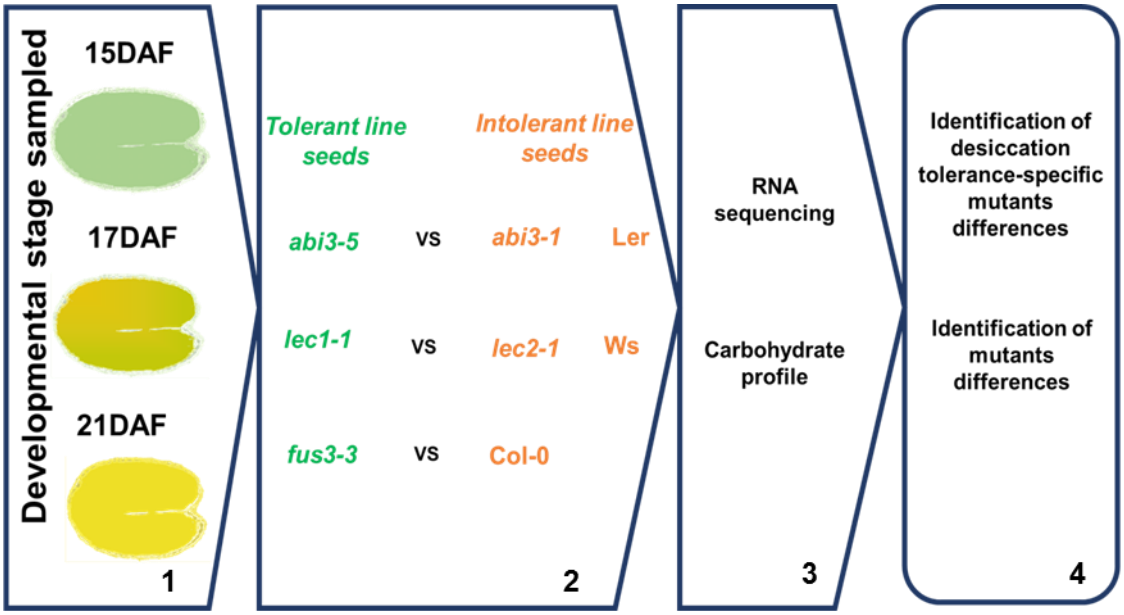

Figure 8. General approach used to identify desiccation tolerance differences.1) Seeds were sampled at 15, 17 and 21DAF 2) Different lines: desiccation intolerant mutants, tolerant mutants and wildtypes were collected. Contrasts were performed tolerant line seeds vs intolerant line seeds at 15, 17 and 21 DAF. 3) RNA sequencing and

carbohydrate profiles were performed from each lines sampled. 4) To RNA sequencing, the DT-specific differences were identified by generalized linear model and to carbohydrates analysis, the DT–specific differences were identified by pair comparision (mutant *vs* wildtype)

Table 9. Overview of sequencing and mapping statistics on Arabidopsis genomes

library	Total reads	Reads mapped (%)	Reads uniquely (%)	Genes detected
***lec1-1*_15daf**	8,907,428	8,907,428 (73.6)	5,909,235 (92.4)	22,011
***lec1-1*_17daf**	6,664,483	5,000,413 (75.0)	4,413,624 (92.0)	22,011
***lec1-1*_21daf**	6,317,389	4,560,674 (72.2)	3,977,475 (91.4)	22,011
***lec2-1*_15daf**	6,469,212	4,983,250 (77.0)	4,601,405 (93.3)	22,011
***lec2-1*_17daf**	6,895,879	5,119,705 (74.2)	4,703,369 (93.3)	22,011
***lec2-1*_21daf**	8,499,125	6,305,852 (74.2)	5,668,640 (93.4)	22,011
Ws_15daf	6,441,380	4,784,939 (74.3)	4,468,431 (93.7)	22,011
Ws_17daf	5,562,550	4,242,351 (76.3)	3,934,393 (93.0)	22,011
Ws_21daf	7,275,259	5,402,772 (74.3)	5,002,734 (93.5)	22,011
***abi3-5*_15daf**	5,112,140	3,941,394 (77.1)	3,572,798 (91.6)	22,011
***abi3-5*_17daf**	5,180,272	3,989,412 (77.0)	3,652,498 (92.4)	22,011
***abi3-5*_21daf**	5,678,982	4,019,785 (70.8)	3,630,208 (91.5)	22,011
***abi3-1*_15daf**	5,963,995	4,355,175, (73.0)	3,993,489 (92.2)	22,011
***abi3-1*_17daf**	4,621,130	3,629,414 (78.5)	3,368,136 (93.0)	22,011
***abi3-1*_21daf**	3,632,793	2,806,398 (77.3)	2,568,817 (92.3)	22,011
Ler_15daf	8,408,818	6,327,513, (75.2)	5,860,618 (92.9)	22,011
Ler_17daf	5,538,070	4,125,875 (74.5)	3,777,419 (92.5)	22,011
Ler_21daf	5,241,362	4,017,069 (76.6)	3,695,767 (92.5)	22,011
***fus3-3*_15daf**	11,058,194	8,442,157 (76.3)	7,717,931 (93.2)	22,011
***fus3-3*_17daf**	13,368,320	10,544,256 (78.9)	9,452,736 (92.3)	22,011

***fus3-3*_21daf**	7,491,466	5,859,888 (78.2)	5,103,317 (91.9)	22,011
Col-0_15daf	6,365,373	4,862,900 (76.4)	4,495,715 (92.8)	22,011
Col-0_17daf	3,510,513	2,693,072 (76.7)	2,464,782 (92.4)	22,011
Col-0_21daf	8,101,628	6,195,345 (76.5)	5,796,084 (93.8)	22,011
Total reads	**162,305,761**	**125,117,037**	**111,829,621**	

To determine differentially expressed genes (DEGs), RNASeq data was analyzed using two types of model analysis. First, we used a Generalized Linear Model (GLM) based on an interaction model, which comprises genes for DT-specific mutants difference (Table 10). Using GLM and a stringency level of false discovery rate (FDR) of < 0.05 and $\log_2$ fold change ($\log_2$FC) <1 and >-1, we identified 3,781 DEGs between tolerant and intolerant lines in at least one of the three developmental stages sampled for this analysis. Among the DEGs, 2,320 were upregulated and 1,461 downregulated (Table 10).

Table 10. DEGs in generalized linear model fitted to desiccation tolerance seeds data.

Comparison Model	Interpretation	Developmental stage sampled	Differentially Expressed Genes
[(*lec1-1* - Ws) - (*fus3-3* - Col-0) - (*abi3-5* - Ler] vs [(*lec2-1* - Ws) - (*abi3-1*- Ler)]	Desiccation tolerance-specific mutants differences intolerant lines VS tolerant lines	15 DAF	Upregulated 434 Downregulated 548
[(*lec1-1* - Ws) - (*fus3-3* - Col-0) - (*abi3-5* - Ler] vs [(*lec2-1* - Ws) - (*abi3-1*- Ler)]		17 DAF	Upregulated 893 Downregulated 432
[(*lec1-1* - Ws) - (*fus3-3* - Col-0) - (*abi3-5* - Ler] vs [(*lec2-1* - Ws) - (*abi3-1*- Ler)]		21 DAF	Upregulate 993 Downregulated 481

The second GLM analysis was based on pairwise comparisons between each mutant line and their corresponding wild type ecotype (Figure 8, Table 11).

Table 11. DEGs from pairwise comparison.

Comparison	Upregulated DEGs	Downregulated DEGs
***lec1-1* vs Ws 15 DAF**	1535	992
***lec1-1* vs Ws 17 DAF**	2164	1441
***lec1-1* vs Ws 21 DAF**	2397	1057
***lec2-1* vs Ws 15 DAF**	418	98
***lec2-1* vs Ws 17 DAF**	450	23
***lec2-1* vs Ws 21 DAF**	883	50
***abi3-5* vs Ler 15 DAF**	476	538
***abi3-5* vs Ler 17 DAF**	946	1128
***abi3-5* vs Ler 21 DAF**	2104	1660
***abi3-1* vs Ler 15 DAF**	2	0
***abi3-1* vs Ler 17 DAF**	425	397
***abi3-1* vs Ler 21 DAF**	0	0
***fus3-3* vs Col-0 15 DAF**	1610	1807
***fus3-3* vs Col-0 17 DAF**	2522	1429
***fus3-3* vs Col-0 21 DAF**	2066	1766

A Heatmap of DEGs (Figure 9) shows that there is a large subset of genes that is upregulated in *lec2-1*, *lec1-1*, *abi3-5* and *fus3-3* respect to their WT controls, which probably represent genes that are involved in the direct transition of the embryo to vegetative growth rather that entering into dormancy and DT. This group of genes

is not activated in *abi3-1,* a weak allele of *abi3* that does not show transition to vegetative growth as indicated by the lack of ectopic trichomes on cotyledons or presence of chlorophyll in the mature embryo. A second subset of genes is evident in the heatmap which is drastically repressed only in the DT intolerant mutants and which probably represents genes that involved in the acquisition of DT.

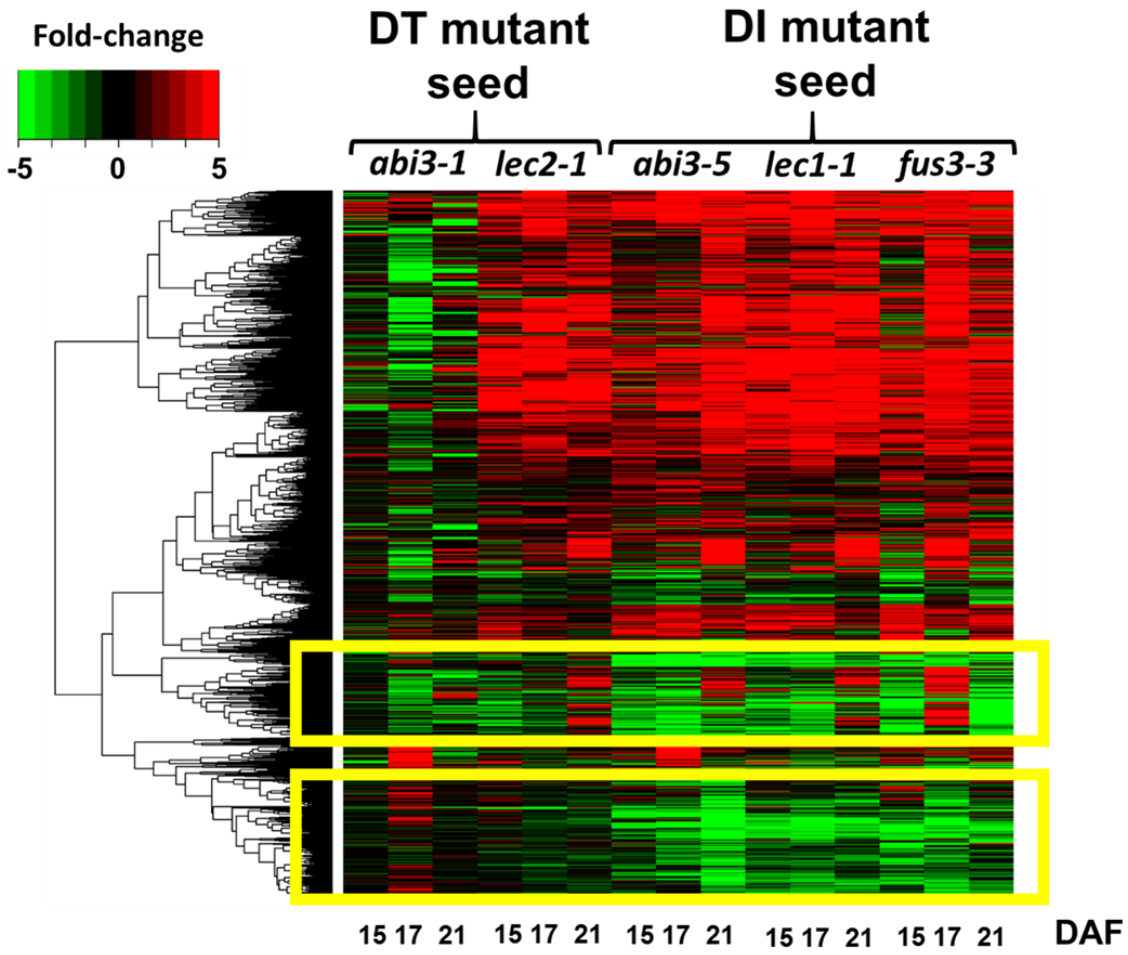

Figure 9. Global gene expression profile in desiccation intolerant seed mutants. (a) Heat map from hierarchical clustering of differentially expressed genes. Each mutant was compared against its corresponding wild type (*lec1-1* and *lec2-1* versus Ws; *abi3-5* versus Ler; *fus3-3* versus Col-0) at three developmental stages, 15, 17 and 21 DAF. Green indicates downregulated values, red upregulated values and black unchanged values. The yellow squares represent the common downregulated genes between intolerant mutants, DI desiccation intolerant, and DT desiccation tolerant.

The Venn diagrams presented in Figure 10 show the relationship between upregulated DEGs in *lec1-1*, *abi3-5, fus3-3* and *lec2-1* with respect to their corresponding wild type. Comparison between tolerant and intolerant mutant pairs such as *lec1-1* and *lec2-1* or *abi3-5* and *abi3-1* with their corresponding WT controls produced a common set of DEGs that are relevant to the pleiotropic phenotype of the mutants. For example, as *lec2* has a similar phenotype to *lec1* but it is still desiccation tolerant, genes that are differentially expressed in both mutants when compared to the WT should have no direct relevance to the tolerance process. Comparison of differentially expressed genes of all desiccation intolerant mutants yielded a set of genes that differentially expressed in all desiccation intolerant mutants respect to the DT mutants and corresponding WT. At 15 DAF, 89 upregulated genes were common to all intolerant mutants (*lec1, abi3-5* and *fus3-3)* (Figure10). As the level of water loss increased, the number of DEGs also increased; at 17 and 21 DAF a total of 129 and 310 upregulated genes were common to intolerant mutants(Figure 10). The expression of these genes was altered in all desiccation-intolerant mutants respect to tolerant mutants with related phenotypes or their wild type counterparts. The highest fold change values of DEGs were reached at 21 DAF, when water loss in the seed is complete.

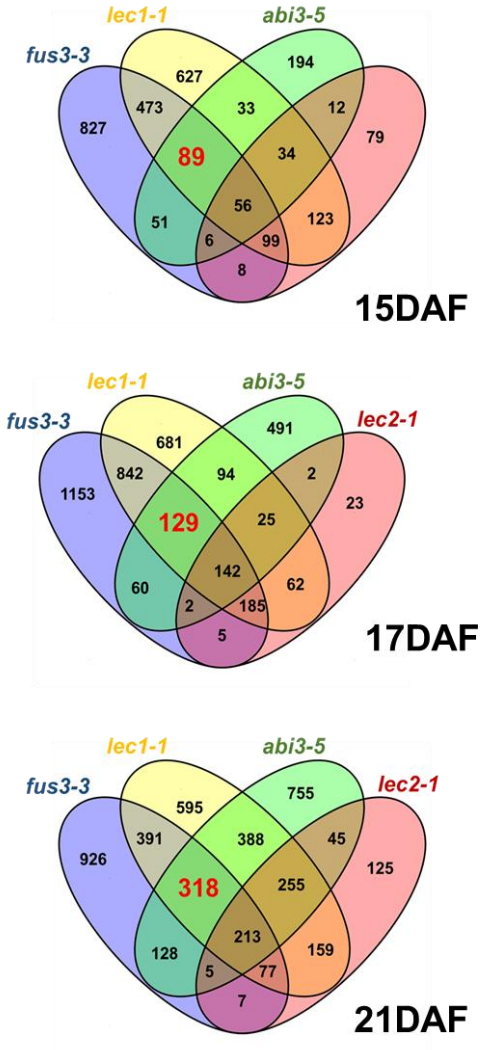

Figure 10. Venn diagrams of upregulated genes. Venn diagrams showing the number and distribution of differentially upregulated genes across the tested mutants lines at 15, 17 and 21 DAF, respectively. The number of shared in desiccation intolerant mutants genes is indicated in red.

A subset of DEGs was identified as specific for each mutant. For instance, *fus3-3* showed the highest number of specific DEGs in the three time-points analyzed (827, 1153 and 926 upregulated and 1187, 779 and 587 downregulated genes at 15, 17 and 21 DAF, respectively) (Figure 10). The mutant specific DEGs most

probably encode genes that are independently regulated by each of these TFs and are related to the specific phenotypes of the corresponding mutant. For example, genes unique to *fus3-3* were mostly related to cell cycle, cell division and ethylene responses, in agreement with previous reports reporting that the *fus3-3* mutation induces ectopic cell divisions in the embryo (Gazzarrini et al. 2004) together with an increased expression of a subset of ethylene responsive genes (Lumba et al. 2012) (Table 12) .Since the reduced expression of these genes is not detected in the other two DT mutants it suggests that these processes are independent of seed DT.

Table 12. Expression pattern of upregulated genes related with different process.

Locus	Description	Fold-change *abi3-5/Ler*			Fold-change *fus3-3/Col-0*			Fold-change *lec1-1/Ws*			Fold-change *lec2-1/Ws*			Fold-change *abi3-1/Ler*		
		15 DAF	17 DAF	21 DAF	15 DAF	17 DAF	21 DAF	15 DAF	17 DAF	21 DAF	15 DAF	17 DAF	21 DAF	15 DAF	17 DAF	21 DAF
Anthocyanins genes related																
AT5G13930	Chalcone and stilbene synthase family protein	0.97	-1.42	5.49	0.71	5.70	2.71	1.26	6.34	5.67	1.70	5.70	4.11	0.59	-4.69	1.85
AT5G07990	Cytochrome P450 superfamily protein	-1.12	-3.87	1.95	-0.57	1.79	-0.50	-0.77	0.41	1.66	-0.71	-0.18	1.03	1.04	-2.18	0.08
AT5G42800	dihydroflavonol 4-reductase	-7.40	-2.27	0.76	5.96	5.61	0.56	0.67	2.32	6.85	3.84	4.56	9.33	-2.70	-1.68	-4.87
AT3G51240	flavanone 3-hydroxylase	-1.29	-2.57	1.40	-0.78	1.62	2.59	-0.72	1.98	1.21	0.54	1.30	0.97	0.43	-1.73	0.82
AT5G17220	glutathione S-transferase phi 12	-2.32	-0.73	7.56	8.07	10.48	5.75	8.58	9.78	3.07	10.62	10.62	5.82	-0.82	-4.57	0.00
AT3G29590	HXXXD-type acyl-transferase family protein	0.00	0.00	7.07	7.76	9.18	-0.16	6.48	8.40	7.00	8.01	8.01	7.30	0.00	0.00	0.00
AT3G55970	jasmonate-regulated gene 21	3.20	-4.57	4.27	3.42	4.73	0.40	3.83	2.90	0.00	6.67	6.67	6.62	0.00	-4.57	4.18
AT4G22880	leucoanthocyanidin dioxygenase	-3.11	3.84	6.70	5.47	9.28	1.57	3.50	8.96	1.29	5.73	9.92	3.58	-6.31	0.00	0.00
AT1G66390	myb domain protein 90	3.20	3.84	3.78	3.80	1.16	1.10	6.32	1.87	5.58	8.72	3.49	9.09	3.76	0.00	-3.92
AT5G54060	UDP-glucose:flavonoid 3-o-glucosyltr	0.00	0.00	7.47	7.40	8.90	3.67	8.00	10.17	7.21	10.10	10.10	8.63	3.76	0.00	0.00

	ansferase															
AT4G14090	UDP-Glycosyltransferase superfamily protein	0.94	0.74	3.44	1.46	4.49	1.96	1.71	4.60	1.50	3.17	4.39	2.06	-1.23	-2.09	0.59
AT5G49690	UDP-Glycosyltransferase superfamily protein	1.78	6.21	-1.25	4.75	-2.57	-0.31	1.30	-0.69	-1.45	4.52	0.53	-0.50	0.15	7.37	0.44
Chlorophyll synthesis genes related																
AT1G08520	ALBINA 1	2.58	3.87	3.12	2.51	1.46	0.51	9.28	5.35	2.84	9.26	5.00	1.67	0.48	0.43	-1.04
AT1G69740	Aldolase superfamily protein	1.48	0.91	1.64	1.67	1.47	2.12	1.92	1.93	0.68	1.08	1.38	0.07	0.35	-0.64	0.34
AT1G03475	Coproporphyrinogen III oxidase	0.67	0.92	1.65	1.90	1.67	1.06	1.53	1.95	0.49	0.90	1.42	-0.25	-0.49	0.83	0.09
AT3G56940	dicarboxylate diiron protein, putative (Crd1)	1.93	2.31	7.57	1.56	6.11	2.68	4.67	6.08	5.90	5.31	5.51	5.97	0.66	-4.34	0.68
AT3G59400	enzyme binding;tetrapyrrole binding	1.09	1.60	3.15	0.18	6.45	2.99	7.57	8.43	6.46	8.62	8.62	7.69	-0.95	-5.53	1.22
AT2G30390	ferrochelatase 2	1.58	1.47	3.17	1.12	1.88	0.09	0.94	1.18	0.88	1.01	1.20	0.84	1.23	0.59	0.27
AT1G48520	GLU-ADT subunit B	1.87	3.01	2.29	2.73	2.45	0.03	1.71	-0.36	0.85	-0.22	-0.61	0.92	0.89	1.64	0.08
AT3G48730	glutamate-1-semialdehyde 2,1-aminomutase 2	1.87	2.13	2.78	3.44	2.44	2.22	3.66	4.60	1.61	1.96	2.82	0.39	-0.18	0.25	0.38
AT5G63570	glutamate-1-semialdehyde-2,1-aminomutase	0.99	1.09	1.57	3.19	2.28	0.68	2.54	2.35	1.36	1.41	1.07	0.50	-0.09	0.33	0.25
AT1G58290	Glutamyl-tRNA reductase family protein	1.44	3.43	3.76	1.25	8.43	1.54	5.79	4.93	4.99	5.46	4.43	5.35	-0.69	-1.11	-4.87
AT2G26550	heme oxygenase 2	0.98	5.76	1.72	2.40	0.73	-0.16	1.73	7.79	2.69	-1.17	5.55	2.28	0.52	5.97	-5.85
AT5G08280	hydroxymethylbilane synthase	1.71	1.79	2.50	2.66	1.04	3.22	2.56	4.32	1.37	1.38	2.71	0.28	-0.74	0.68	0.42
AT5G45930	magnesium chelatase i2	2.09	4.78	8.00	2.63	7.56	6.16	7.39	7.65	6.35	5.55	5.55	5.92	-4.55	0.00	0.00
AT5G13630	magnesium-chelatase subunit chlH, chloroplast, putative / Mg-protoporphyrin IX chelatase, putative (CHLH)	0.87	2.17	10.14	0.66	7.77	0.53	4.01	9.88	4.95	5.55	9.75	6.39	0.60	-6.52	0.00
AT4G25080	magnesium-protoporphyrin IX methyltransferase	1.71	2.51	2.78	1.22	1.98	2.27	3.53	4.14	1.61	2.60	3.30	0.94	0.64	-3.44	0.27
AT5G18660	NAD(P)-binding Rossmann-fold	1.89	3.03	9.36	1.62	9.19	3.34	3.59	5.85	8.86	2.20	4.39	7.57	-0.16	-4.57	0.00

	superfamily protein															
AT4G18480	P-loop containing nucleoside triphosphate hydrolases superfamily protein	1.44	1.60	10.47	1.90	4.67	3.49	9.69	9.92	3.06	9.10	9.10	3.08	-0.42	-1.93	0.00
AT5G54190	protochlorophyllide oxidoreductase A	-0.43	0.91	5.79	4.91	4.51	8.82	3.29	6.17	3.85	4.27	4.27	4.81	-4.55	-5.53	-4.87
AT4G27440	protochlorophyllide oxidoreductase B	2.41	4.20	8.66	2.53	6.79	1.84	4.83	6.37	3.44	5.55	5.73	4.20	-2.23	-1.13	2.52
AT1G03630	protochlorophyllide oxidoreductase C	3.14	6.21	9.52	3.50	8.38	4.13	8.23	3.70	3.51	8.69	3.46	4.00	1.09	4.43	4.18
AT5G63290	Radical SAM superfamily protein	0.28	5.92	-0.22	1.34	-1.80	1.80	6.67	5.72	-0.91	5.80	5.80	-0.74	0.56	6.95	-0.31
AT1G45110	Tetrapyrrole (Corrin/Porphyrin) Methylases	2.04	1.36	1.88	0.72	1.30	3.06	5.86	2.14	0.34	3.34	-0.92	-0.09	0.15	-4.57	2.20
AT3G14110	Tetratricopeptide repeat (TPR)-like superfamily protein	1.34	1.31	1.00	1.81	2.15	2.76	2.14	3.21	2.05	1.15	1.43	0.93	-0.24	-0.44	0.52
AT3G51820	UbiA prenyltransferase family protein	1.84	1.89	4.33	2.62	2.47	1.87	3.15	4.35	2.71	3.69	4.60	2.65	0.85	-1.35	0.05
AT2G40490	Uroporphyrinogen decarboxylase	0.93	0.23	2.77	2.04	0.76	1.48	1.11	3.01	1.53	0.62	1.66	0.60	-0.57	-2.29	0.40
AT3G14930	Uroporphyrinogen decarboxylase	0.51	0.36	2.93	1.50	1.67	1.05	1.37	2.44	1.39	0.92	1.45	0.98	0.29	-0.39	1.20
Leaf development genes related																
AT1G63650	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	0.54	6.74	6.86	0.39	1.97	-0.16	1.26	2.06	6.22	-0.11	0.79	0.00	-5.52	0.00	0.00
AT5G41315	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	0.00	0.00	5.23	2.11	0.51	-1.67	4.52	0.00	5.58	3.34	3.34	0.00	0.00	0.00	0.00
AT1G79840	HD-ZIP IV family of homeobox-leucine zipper protein with lipid-binding START domain	1.39	0.34	1.70	0.89	1.70	3.36	3.54	5.37	2.42	2.37	3.37	1.25	0.01	-1.58	0.36
AT3G61150	homeodomain GLABROUS 1	-0.03	-1.05	1.13	1.52	0.61	0.54	-0.06	7.87	1.40	-2.28	5.55	0.49	0.16	-1.94	1.27
AT1G17920	homeodomain GLABROUS 12	-1.23	3.84	5.95	2.24	5.41	1.57	5.63	-0.48	3.31	0.00	-5.23	-0.48	-1.61	0.00	4.18

AT1G05230	**homeodomain GLABROUS 2**	-1.03	0.12	1.50	0.59	0.55	0.87	7.47	0.90	0.89	7.11	0.10	0.93	-0.16	0.22	0.97
AT2G46410	**Homeodomain-like superfamily protein**	4.12	4.39	3.72	3.42	5.87	5.58	1.30	6.17	2.43	-0.85	3.34	-0.48	3.76	4.43	5.14
AT5G53200	**Homeodomain-like superfamily protein**	4.68	4.39	2.82	-0.32	5.30	-1.04	1.17	1.27	5.37	0.39	1.28	0.00	3.76	0.00	0.00
AT5G11060	**KNOTTED 1-like homeobox gene 4**	-0.43	-0.73	5.44	4.35	0.33	-2.10	5.34	1.77	2.93	7.20	2.94	3.50	0.72	0.83	0.00
AT4G08150	**KNOTTED-like from Arabidopsis thaliana**	3.14	7.60	4.01	6.29	0.01	-0.16	5.86	1.07	-0.24	5.55	0.32	-1.00	1.09	4.43	2.20
AT3G27920	**myb domain protein 0**	0.00	0.00	2.82	2.90	2.88	0.00	4.52	0.00	4.40	4.27	4.27	0.00	0.00	0.00	0.00
AT2G26580	**plant-specific transcription factor YABBY family protein**	1.30	-1.32	6.79	-2.37	7.24	-0.81	-0.21	1.08	6.67	-1.42	-0.50	6.26	0.08	-2.27	0.00
AT2G45190	**Plant-specific transcription factor YABBY family protein**	-0.09	-0.36	3.25	-1.84	2.39	0.11	-0.77	1.04	4.38	-1.04	0.13	3.69	0.18	-1.59	-5.44

The heatmap show a second subset of DEGs, which appeared as drastically repressed in all desiccation intolerant mutants when compared to DT lines, could represent genes that are directly or indirectly relevant for the acquisition of DT in *Arabidopsis* seed (Figure 9). The number of downregulated genes specific for desiccation intolerant mutants increased as the level of water content decreased in the seed, from 147 genes at 15 DAF to 380 at 21 DAF (Figure 11).

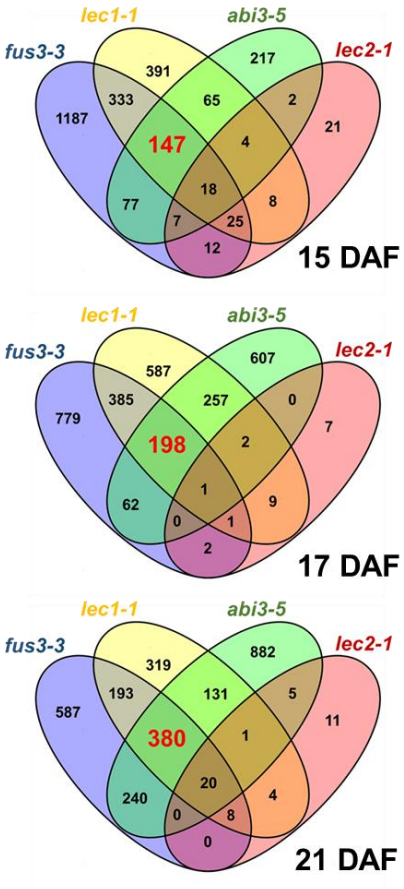

Figure 11. Venn diagrams showing the number and distribution of differential genes across the comparison among mutants representing downregulated genes at 15, 17 and 21 DAF, respectively. Genes potentially related with desiccation tolerance are indicated in red.

5.3. General responses to desiccation tolerances

To determine global responses involved in DT, the set of DEGs specific for desiccation-intolerant mutants was classified into functional categories according to gene ontology (GO) enrichment using the BiNGO tool (Maere *et al*., 2005). BiNGO-derived graph illustrated the most highly significantly enriched specific function categories (Figure 12 and 13).

Upregulated genes were significantly enriched for categories associated to metabolic processes, particularly photosynthesis and carbohydrate and amino acid metabolism. Among this category we found overrepresentation of genes involved in chloroplast activity such as photosystem (PS) I and II, carbon fixation, chlorophyll (tetrapyrrole) biosynthesis and amino acid metabolisms (Figure 12). Other functions were enriched for genes involved in vegetative development like post-embryonic development. Upregulated expression of these genes was expected as intolerant mutants transit directly from embryo to vegetative development (or remain green as in the case of *abi3-5*), processes that are shut down during embryo maturation in the WT.

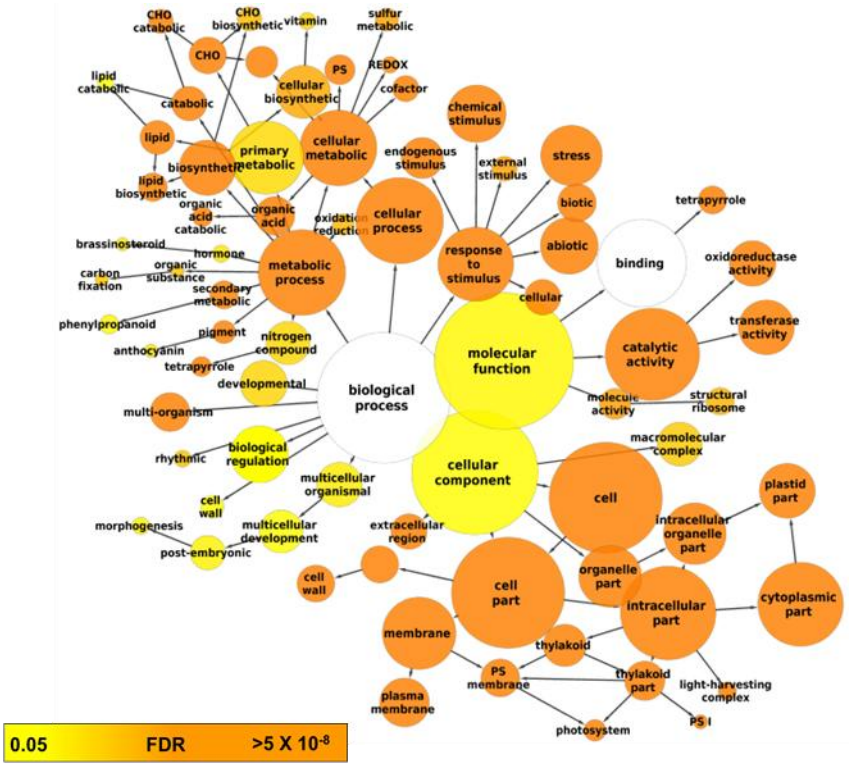

Figure 12. Enriched GO Terms of upregulated genes in *lec1*, *lec2*, *fus3* and *abi3-5*. The graph is a BiNGO visualization of overrepresented GO terms for upregulated (2,218 genes). Uncolored nodes are not

overrepresented, but they may be the parents of overrepresented terms. Colored nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected P value < 0.05), the node color indicates significance as shown in the color bar. PS, Photosynthesis; CHO carbohydrates.

In contrast, downregulated DEGs specific for desiccation intolerant mutants reveal a very different picture, clearly illustrating the role of protective responses and nutrient storage during normal seed formation (Figure 13). These downregulated genes were enriched (FDR < 0.05) in the following GO categories: Molecular Function: oxidoreductase activity and nutrient reservoir and Biological Process: lipid and carbohydrate biosynthesis, seed development and ABA and stress responses, such as water, oxidative and temperature (Figure 13).

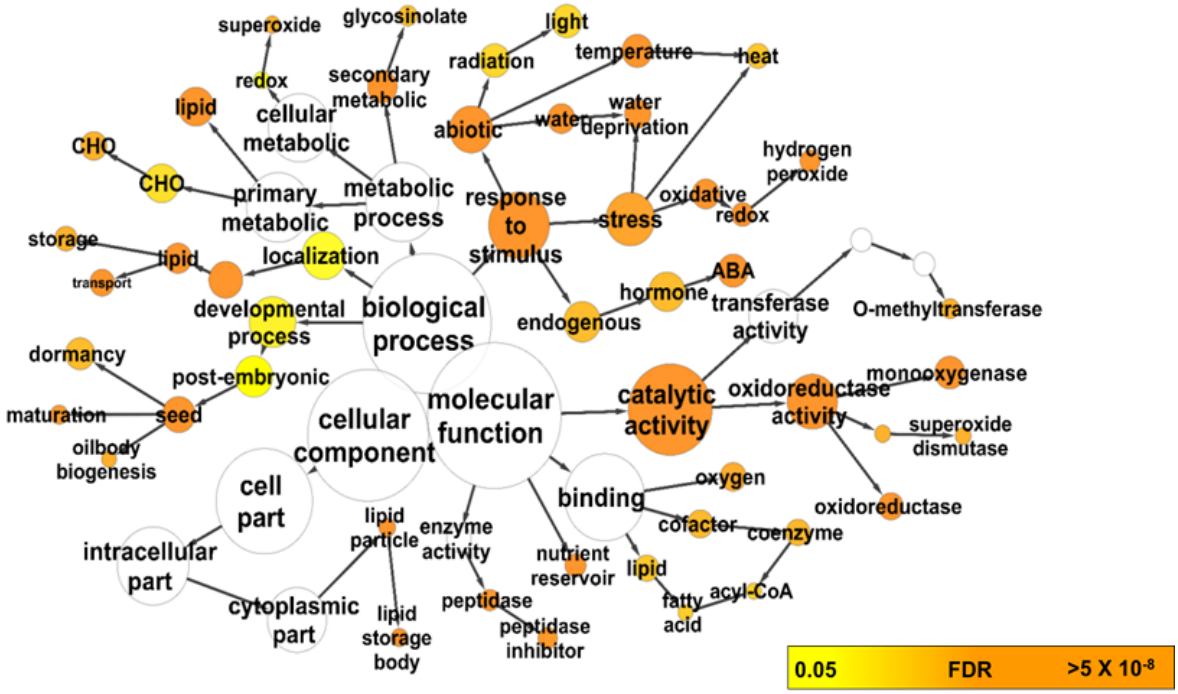

Figure 13. Enriched GO Terms of downregulated genes in *lec1, lec2, fus3* and *abi3-5*. a) The graph is a BiNGO visualization of overrepresented GO

terms for downregulated for desiccation tolerance-specific genes (1,461 genes). Uncolored nodes are not overrepresented, but they may be the parents of overrepresented terms. Colored nodes represent GO terms that are significantly overrepresented (Benjamini and Hochberg corrected P value < 0.05), the node color indicates significance as shown in the color bar. PS, Photosynthesis; CHO carbohydrates; ABA, abscisic acid.

Although a global analysis of this type is informative and the enrichment analysis by GO terms can help identifying categories of genes involved in processes that are relevant for DT, a more specific analysis of each developmental stage should facilitate the understanding of the processes activated during the crucial phase of DT. To identify specific pathways involved in seed desiccation, we used a no-redundant classification system of Mapman using the web-tool Superviewer (Provart et al., 2003). Upregulated genes in desiccation intorelant mutants were mainly classified in metabolic pathways, in particular a large number of DEGs were related to photosynthesis, starch biosynthesis, glycolysis and light signaling that correlate with the high metabolic activity, as well as secondary metabolism specially anthocyanin synthesis (Figure 14).

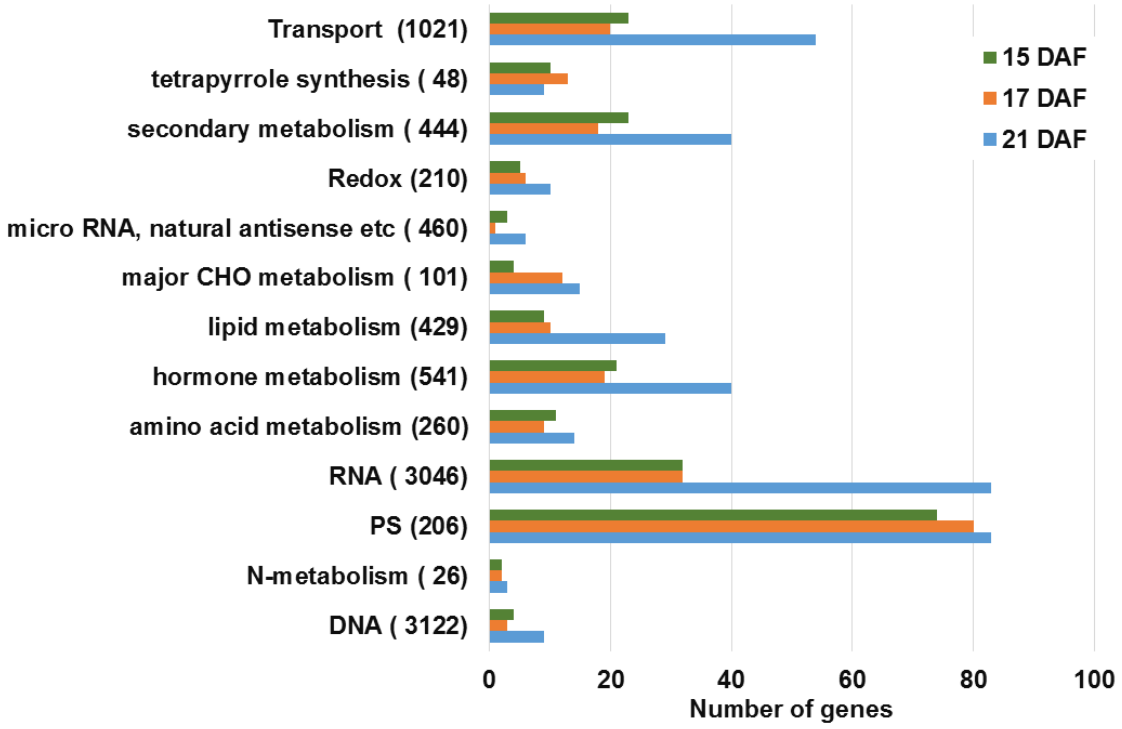

Figure 14. MAPMAN-based functional classification of upregulated DEGs in *lec1, lec2, fus3 and abi3-5* during seed desiccation tolerance. (a) MAPMAN classification of the transcripts using the web-tool Classification Superviewer (http://bar.utoronto.ca). All categories are significant (P value <0.05). CHO, carbohydrate; PS, photosynthesis.

This Mapman classification of DEGs is in accordance to classification of GO enriched terms for upregulated genes; these categories were mainly represented at 17 and 21 DAF. The subset of genes common between DI mutants and *lec2-1* is also overrepresented in gene functions related to photosynthesis and metabolic activity (Figure 15).

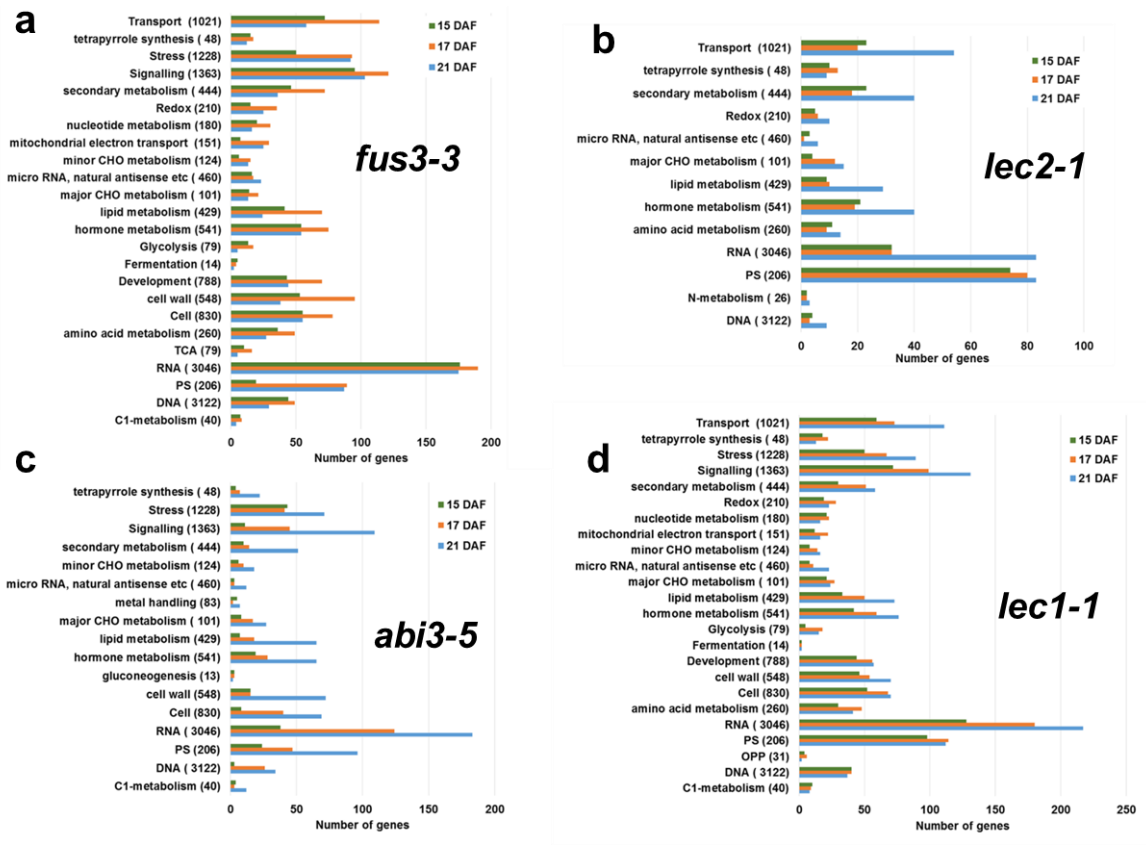

Figure 15. MAPMAN-based functional classification of upregulated DEGs in each mutant line during seed desiccation tolerance. (a) *fus3-3* vs Col-0 (b) *lec2-1* vs Ws (c) *abi3-5* vs Ler and (d) *lec1-1* vs Ws contrasts. MAPMAN classification of the transcripts using the web-tool Classification Superviewer (http://bar.utoronto.ca). All categories are significant (P value <0.05). CHO, carbohydrate; OPP, oxidative pentose phosphate; PS, photosynthesis; TCA, tricarboxylic

A more detailed analysis of the set of downregulated genes of using the MAPMAN system (Thimm et al. 2004) showed enrichment in processes such as abiotic stress, LEA protein synthesis, and metabolic pathways including raffinose, stachyose and trehalose biosynthesis (Figure 16) . We observed that the number of genes for these stress categories increased at 17 and 21 DAF when rapid water loss occurs. (Figure 17),

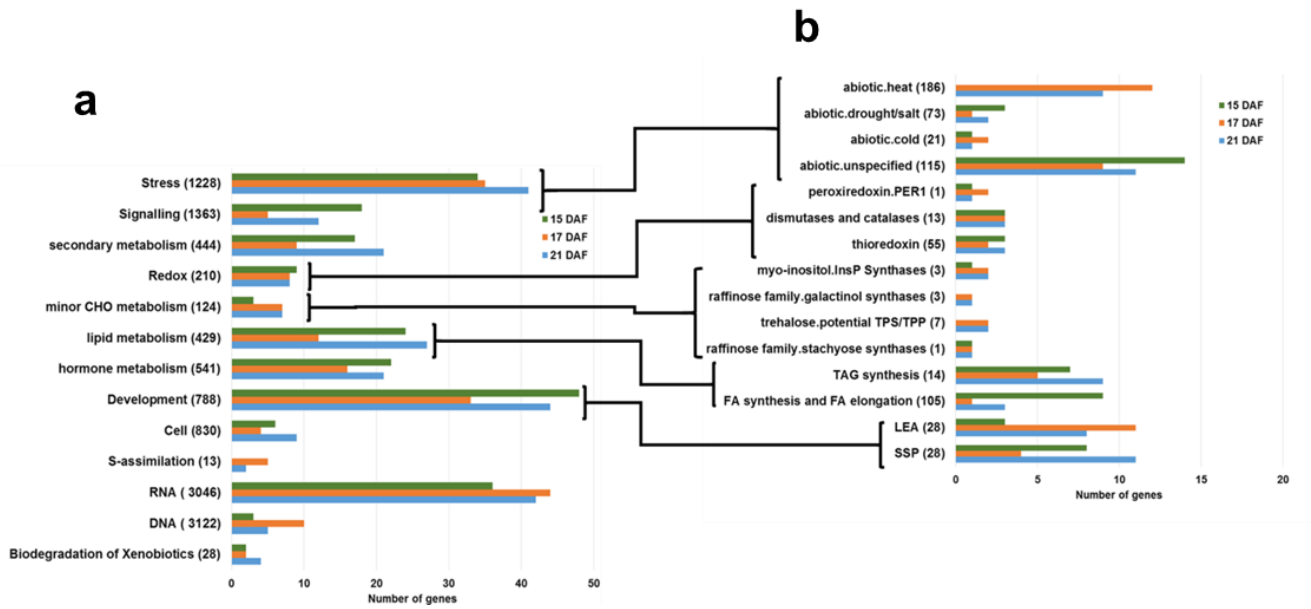

Figure 16. **MAPMAN-based functional classification of downregulated DEGs in desiccation intolerant lines during seed desiccation tolerance.** MAPMAN (a) categories and (b) subcategories classification of the downregulated transcripts using the web-tool Classification Superviewer (http://bar.utoronto.ca). All categories are significant (P value <0.05). CHO, carbohydrate; TPS, Trehalose-6-Phosphate Synthase; TPP, Trehalose-6-Phosphate Phosphatase; TAG, triacylglycerol; LEA, late embryogenesis abundant; SSP, storage seed proteins.

The finding that genes that are not activated in desiccation intolerant mutants during seed maturation belong to water stress responses and cell protection mechanisms confirmed that desiccation intolerant mutants fail to activate mechanisms required to acquire DT in the seed. As expected, *lec2-1* and *abi3-1* had a much lower number of downregulated genes respect to the WT controls than *lec1-1*, *fus3-1* and *abi3-5,* which correlates well with the fact that these mutant is still capable of acquiring DT (Figure 17).

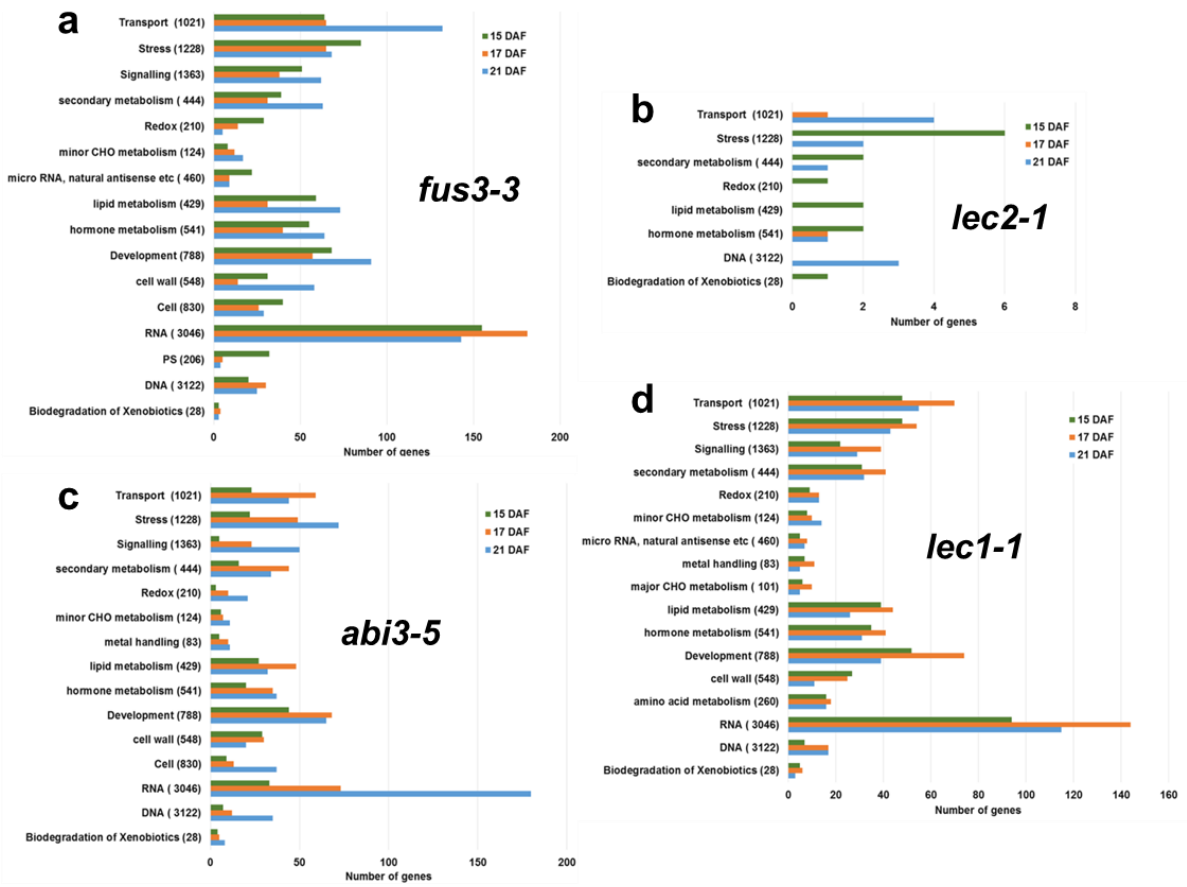

Figure 17. MAPMAN-based functional classification of downregulated DEGs in each mutant line during the acquisition of seed desiccation tolerance. (a) *fus3-3* vs Col-0 (b) *lec2-1* vs Ws (c) *abi3-5* vs Ler and (d) *lec1-1* vs Ws contrasts. MAPMAN classification of the transcripts using the web-tool Classification Superviewer (http://bar.utoronto.ca). All categories are significant (P-value <0.05). CHO, carbohydrate.

5.4. Carbohydrates profiling of seeds from desiccation tolerant and intolerant mutants.

It has been reported that the raffinose family of oligosaccharides (RFO), which accumulate during the last phase of seed maturation, can act as 'water replacement' molecules that provide the hydrogen bonds required for membrane and protein stabilization, as well as for protecting DNA against hydrolytic damage

(Crowe 2002; Obendorf 1997). Therefore, to complement our transcriptional study, we performed a carbohydrate profiling of DT and desiccation intolerant mutants using the same seed developmental stages chosen for RNAseq analysis (15, 17 and 21DAF). We specifically analyzed the sucrose to raffinose pathway (Figure 18), as well as the invertase products D-glucose and D-fructose (Figure 19). This analysis showed that in *lec1-1*, *abi3-5, fus3-3* and *lec2-1* sucrose levels at 15 DAF were one-fold higher than in their wild-type controls and later decreased to almost wild-type levels at 21 DAF.

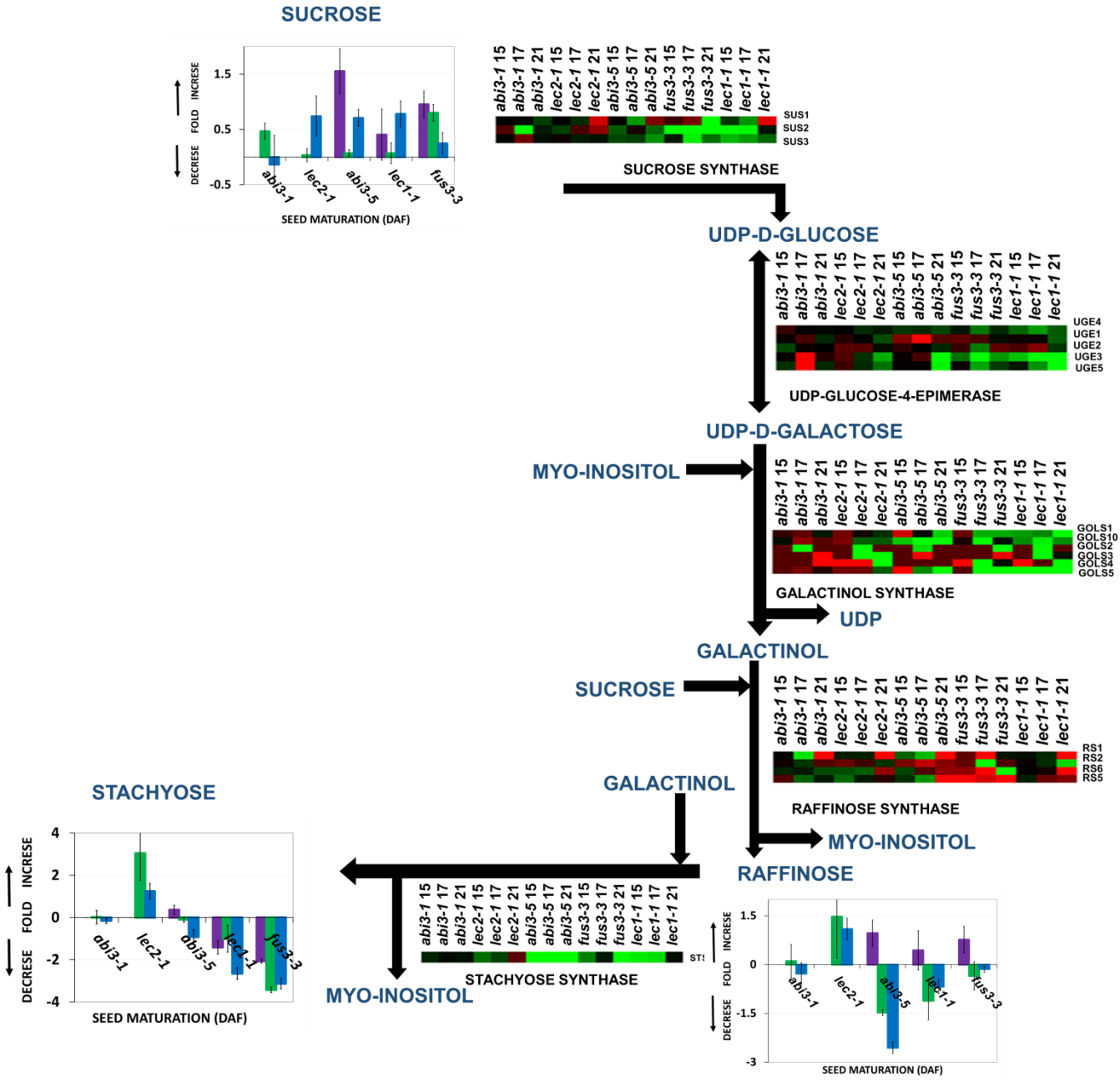

Figure 18. Metabolic pathway of stachyose biosynthesis in Arabidopsis seeds. Graphs represents changes in the contents of sugars and oligosaccharides for each mutant during the desiccation period. Bar graphs indicate the $\log_2$ fold change relative in content relative to wildtype controls. Values from two independently grown set of plants with three technical replicates each. Heatmaps represent expression profiles of putative enzymes involved in raffinose and stachyose synthesis. Red and green represent upregulated and downregulated genes, respectively. The biochemical and physiological pathways were classified according to the KEGG database

(http://www.genome.jp/kegg/).*SUCROSE SYNTHASE* (*SUS1*, AT5G20830; *SUS2*, AT5G49190; *SUS3*, AT4G02280), *UDP-D-GLUCOSE/UDP-D-GALACTOSE 4-EPIMERASE* (*UGE1*, AT1G12780; *UGE2*, AT4G23920; *UGE3*, AT1G63180 ; *UGE4*, AT1G64440; *UGE5*, AT4G10960), *GALACTINOL SYNTHASE* (*GOLS1*, AT2G47180; *GOLS2*, AT1G56600; *GOLS3*, AT1G09350; *GOLS4*, AT1G60470 ; *GOLS5*, AT5G23790; *GOLS10*, AT5G30500), *RAFFINOSE SYNTHASE* (*RS1*, AT1G55740; *RS2*, AT3G57520; *RS5*, AT5G40390; *RS6*, AT5G20250), *STACHYOSE SYNTHASE* (*STS*, AT4G01970).

Raffinose levels decreased 1.5- and 2.5-fold in *abi3-5* and 1- and 0.7-fold in *lec1-1* at 17 and 21 DAF, respectively, while in *lec2-1*, raffinose increased two-fold at 21 DAF. Stachyose levels decreased three-fold in *lec1-1* and *fus3-3,* whereas for *lec2-1* an increase of three-fold respect to the wild type control was determined. (Figure 18). In agreement with the observed reduction in oligosaccharide accumulation, transcript levels of genes encoding key enzymes in the raffinose pathway, such as sucrose synthases (AT5G4919, *SUS2*, and AT4G02280, *SUS3*), UDP-D-galactose-4-epimerases (AT1G63180, *UGE3*, and AT4G10960, *UGE5*), galactinol synthases (AT2G4718, *GOLS1* and AT1G09350, *GOLS2*) and stachyose synthase (AT4G01970, *STS*) were strongly repressed in desiccation intolerant mutants (Figure 18). Additionally, in agreement with the higher levels of D-glucose and D-fructose levels present in the seed of desiccation intolerant mutants, invertase genes are upregulated in these mutants (Figure 19). This suggests that sucrose is degraded to D-glucose and D-fructose via invertase enzymes and is not metabolized to stachyose and raffinose (Figure 18), suggesting that the accumulation of stachyose and raffinose appears to be a key, common factor for the acquisition of DT in *Arabidopsis* seeds.

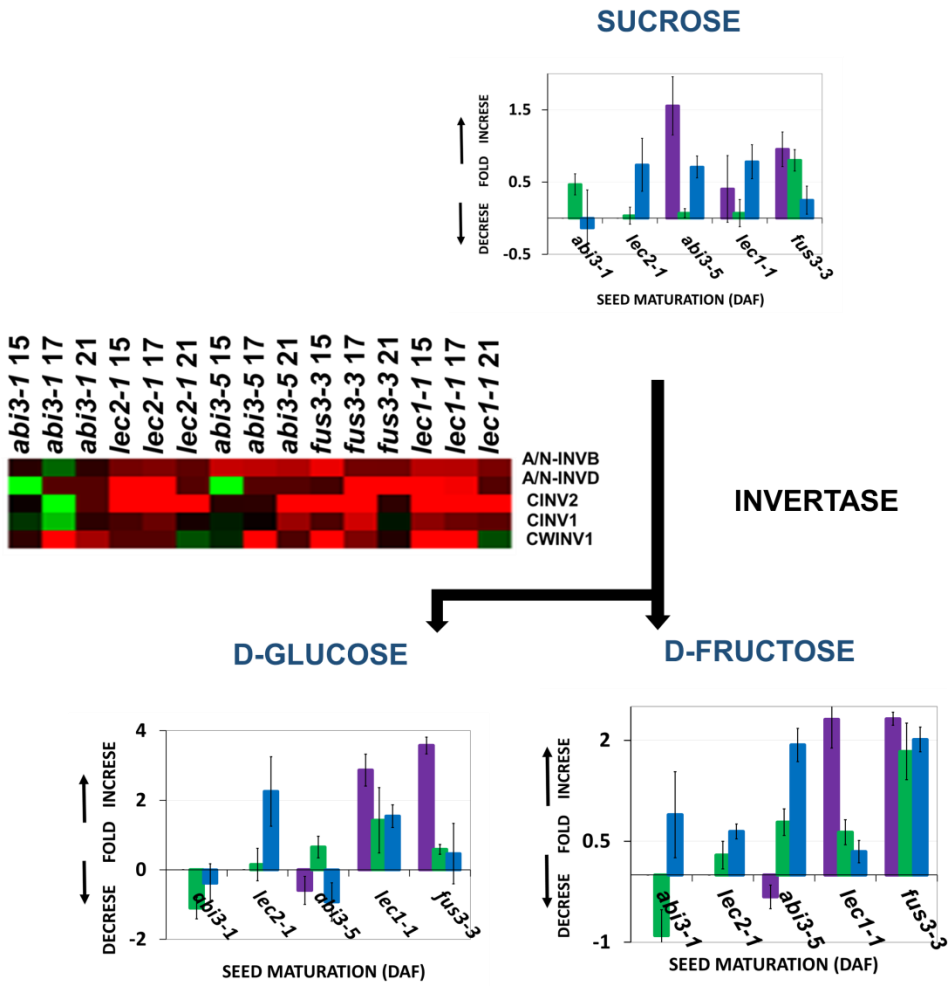

Figure 19. Metabolic pathway sucrose degradation in Arabidopsis seeds. Graphs represents changes in the contents of sugars for each mutant during the desiccation period. Bar graphs indicate the $\log_2$ fold change relative in content relative to wildtype controls. Values from two independently grown set of plants with three technical replicates each. Heatmaps represent expression profiles of putative enzymes involved in sucrose degradation pathway. Red and green represent upregulated and downregulated genes, respectively. The biochemical and physiological pathways were classified according to the KEGG database (http://www.genome.jp/kegg/). *ALKALINE/NEUTRAL INVERTASE* (*A/N-INVB*, AT4G34860; *A/N-INVD,* AT1G22650), *CYTOSOLIC INVERTASE* (*CINV2*, AT4G09510 ; *CINV1*, AT1G35580), *CELL WALL INVERTASE* 1 (*CWINV1*, AT3G13790).

5.5. Inferring novel transcription factors involved in desiccation tolerance of acquisition *Arabidopsis* seed

Our transcriptomics data provided information on the genes possibly involved in DT acquisition. However, transcriptomic data by themselves do not unveil the regulatory networks that control complex processes, nor the key TFs that coordinate the expression of genes involved in such regulatory networks. In order to identify these regulatory pathways and predict novel TF genes involved in DT, we constructed two ARACNE-based (Basso et al. 2005; Chavez Montes et al. 2014) co-expression networks using two carefully curated datasets obtained from 169 seed-specific CEL format files from 24 Affymetrix ATH1 microarray experiments (Table 13).

Table 13 . CEL. Files used in this study

Accession	Description of experiment	CEL_file
E-ATMX-1	Transcription profiling of Arabidopsis wild type seeds grown under sulfur-deficient conditions	CEL_C1.CEL
E-ATMX-1	Transcription profiling of Arabidopsis wild type seeds grown under sulfur-deficient conditions	CEL_C2.CEL
E-ATMX-1	Transcription profiling of Arabidopsis wild type seeds grown under sulfur-deficient conditions	CEL_DS1.CEL
E-ATMX-1	Transcription profiling of Arabidopsis wild type seeds grown under sulfur-deficient conditions	CEL_DS2.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10445.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10448.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10449.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10451.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10453.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10477.CEL
E-GEOD-1051	Transcription profiling by array of Arabidopsis seeds mutant for lec1	GSM10481.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284384.CEL

E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284385.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284386.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284387.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284389.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284391.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284392.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284394.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284395.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284396.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284397.CEL
E-GEOD-11262	Transcription profiling by array of Arabidopsis seed compartments at the globular embryo stage	GSM284398.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299322.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299323.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299324.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299325.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299326.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299333.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299335.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299336.CEL
E-GEOD-11852	Transcription profiling by array of Arabidopsis mutant for pickle after treatment with uniconazole	GSM299337.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311273.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311274.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311275.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311277.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311278.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311279.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311280.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311281.CEL

E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311282.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311283.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311284.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311285.CEL
E-GEOD-12402	Transcription profiling by array of Arabidopsis seed compartments at the pre-globular stage	GSM311286.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311288.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311289.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311290.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311291.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311292.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311293.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311294.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311295.CEL
E-GEOD-12403	Transcription profiling by array of Arabidopsis seed compartments at the linear-cotyledon stage	GSM311296.CEL
E-GEOD-14229	Transcription profiling by array of Arabidopsis mutant for myb5	GSM356564.CEL
E-GEOD-14229	Transcription profiling by array of Arabidopsis mutant for myb5	GSM356565.CEL
E-GEOD-14229	Transcription profiling by array of Arabidopsis mutant for myb5	GSM356566.CEL
E-GEOD-14229	Transcription profiling by array of Arabidopsis mutant for myb5	GSM356567.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378647.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378649.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378650.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378651.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378653.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378654.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378655.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378656.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378657.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378658.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed compartments at the heart stage	GSM378659.CEL
E-GEOD-15160	Transcription profiling by array of Arabidopsis seed	GSM378660.CEL

	compartments at the heart stage	
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378733.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378734.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378735.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378736.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378737.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378738.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378739.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378740.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378741.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378742.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378743.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378744.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378745.CEL
E-GEOD-15165	Transcription profiling by array of Arabidopsis seed compartments at the mature green stage	GSM378746.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501157.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501158.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501159.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501160.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501161.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501162.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501163.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501164.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501165.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501166.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501167.CEL
E-GEOD-20039	Expression data from Arabidopsis seed compartments at the bending coyledon stage	GSM501168.CEL
E-GEOD-24887	Transcription profiling by array of Arabidopsis wild-type and dcl1-15 torpedo-staged embryos	GSM612043.CEL
E-GEOD-24887	Transcription profiling by array of Arabidopsis wild-type and dcl1-	GSM612044.CEL

	15 torpedo-staged embryos	
E-GEOD-24887	Transcription profiling by array of Arabidopsis wild-type and dcl1-15 torpedo-staged embryos	GSM612045.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712386.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712387.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712388.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712392.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712393.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712394.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712395.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712396.CEL
E-GEOD-28747	Transcription profiling by array of Arabidopsis ft-1 mutants	GSM712397.CEL
E-GEOD-30223	Transcription profiling by array of Arabidopsis seeds during the germination ripening, stratification and germination	GSM748466_Har_rep_1_Arab.CEL
E-GEOD-30223	Transcription profiling by array of Arabidopsis seeds during the germination ripening, stratification and germination	GSM748467_Har_rep_2_Arab.CEL
E-GEOD-30223	Transcription profiling by array of Arabidopsis seeds during the germination ripening, stratification and germination	GSM748468_Har_rep_3_Arab.CEL
E-GEOD-30223	Transcription profiling by array of Arabidopsis seeds during the germination ripening, stratification and germination	GSM748469_0_h_dry_rep_1_Arab.CEL
E-GEOD-30223	Transcription profiling by array of Arabidopsis seeds during the germination ripening, stratification and germination	GSM748470_0_h_dry_rep_2_Arab.CEL
E-GEOD-30223	Transcription profiling by array of Arabidopsis seeds during the germination ripening, stratification and germination	GSM748471_0_h_dry_rep_3_Arab.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131694.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131695.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131696.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131697.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131698.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131699.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131700.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131701.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131702.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131703.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131705.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131706.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131707.CEL
E-GEOD-5634	Transcription profiling by array of Arabidopsis seeds and siliques at different developmental stages	GSM131708.CEL

E-GEOD-5641	Transcription profiling by array of Arabidopsis mutant for mdh	GSM131848.CEL
E-GEOD-5641	Transcription profiling by array of Arabidopsis mutant for mdh	GSM131849.CEL
E-GEOD-5687	Transcription profiling by array of Arabidopsis seeds after incubation at different temperatures	GSM133117.CEL
E-GEOD-5687	Transcription profiling by array of Arabidopsis seeds after incubation at different temperatures	GSM133118.CEL
E-GEOD-5687	Transcription profiling by array of Arabidopsis seeds after incubation at different temperatures	GSM133119.CEL
E-GEOD-5687	Transcription profiling by array of Arabidopsis seeds after incubation at different temperatures	GSM133120.CEL
E-GEOD-5700	Transcription profiling by array of Arabidopsis after treatment with abscisic acid	GSM133304.CEL
E-GEOD-5700	Transcription profiling by array of Arabidopsis after treatment with abscisic acid	GSM133310.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133312.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133314.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133315.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133316.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133318.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133320.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133321.CEL
E-GEOD-5701	Transcription profiling by array of Arabidopsis after treatment with gibberellin	GSM133322.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133758.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133759.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133761.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133762.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133763.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133767.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133769.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133770.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133771.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133772.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133775.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133777.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133779.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133780.CEL
E-GEOD-5730	Transcription profiling by array of Arabidopsis embryo tissues	GSM133781.CEL
E-MEXP-1607	Transcription profiling by array of Arabidosis seeds after treatment with EL6 and GR24 to investigate the effect on germination	DMSO_rep1.CEL
E-MEXP-849	Transcription profiling of Arabidopsis seed and flowers to identify DELLA regulated transcripts	PJR-Dongni-seed-la-er-210404.CEL

E-MEXP-849	Transcription profiling of Arabidopsis seed and flowers to identify DELLA regulated transcripts	PJR-Dongni-seed-la-er-230304.CEL
E-TABM-1007	Transcription profiling by array of Arabidopsis mutant for fis2	caquinof_20080327_wt_2-v4.CEL
E-TABM-1007	Transcription profiling by array of Arabidopsis mutant for fis2	caquinof_20080327_wt_3-v4.CEL
E-TABM-17	Transcription profiling by array of organism parts from different strains of Arabidopsis	ATGE_79_A.CEL
E-TABM-17	Transcription profiling by array of organism parts from different strains of Arabidopsis	ATGE_79_B.CEL

The first co-expression network contained only the TF known to be expressed in Arabidopsis seed, which we named TFsSeedNet, which consists in a matrix of 169 columns by 2088 rows that correspond to the 2088 TF table probe sets present in the ATH1 chip. (Basso et al. 2005) (Figure 20).

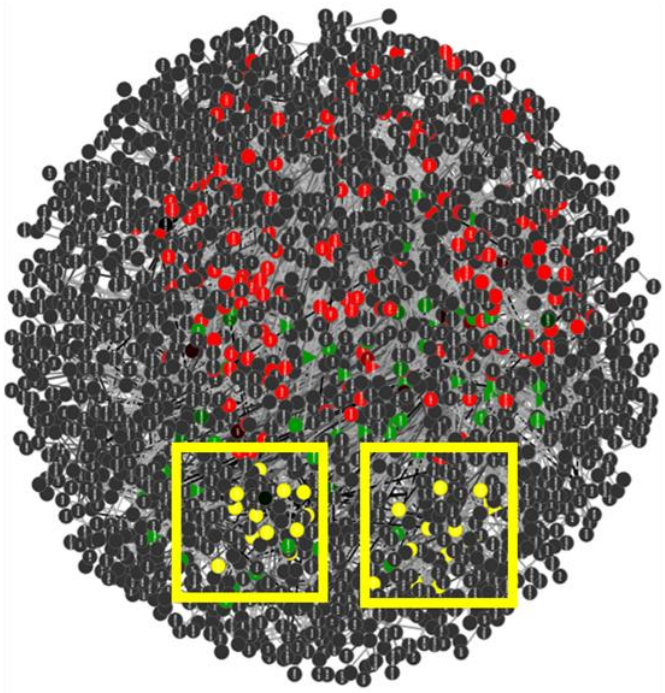

Figure 20.The whole gene regulatory network TFsSeedNet. Overview of the TFsSeedNet obtained at DPI0.0. Genes are represented as nodes and inferred interactions as edges (lines). Nodes are colored grey, except upregulated TFs in red and downregulated TFs in green from DT analysis. Edge width and color intensity is proportional to the Mutual Information (MI)

value of the interaction, with higher MI values corresponding to thicker and darker edges.

An second co-expression networks was a global co-expression network containing all genes expressed in the seed that are represented in the ATH1 microarray, which we named FullSeedNet which consist in a matrix of 169 by 22810 that contain all 228810 probe sets present in the ATH1 chip (Figure 21). We used both dataset as input to construct regulatory networks using the ARACNe software (Basso et al. 2005).

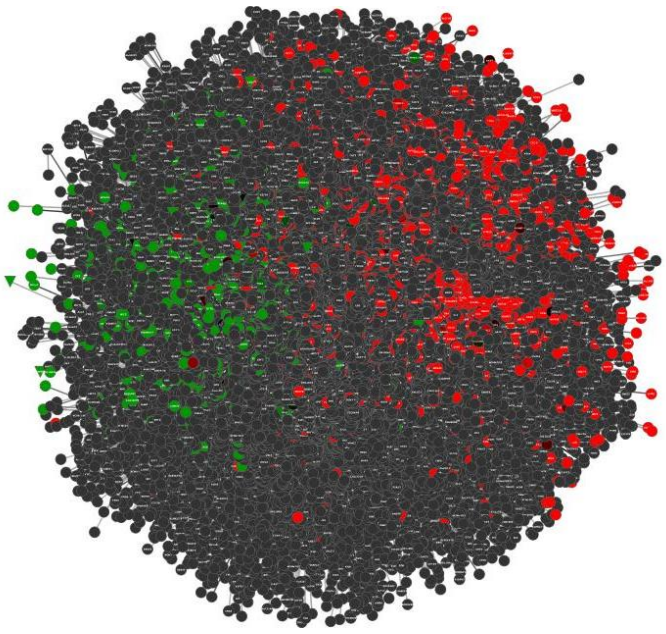

Figure 21. **The whole gene regulatory network FullSeedNet**. Overview of the FullSeedNet obtained at DPI 0.0. Genes are represented as nodes and inferred interactions as edges. Nodes are colored grey, except upregulated genes in red and downregulated genes in green from DT analysis. Edges width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges.

The ARACNe output is a list of interaction probeset pairs ranked through a Mutual Information value (MI) and its associated p-value (Margolin et al. 2006). Detail for a theoretical background and practical use of ARACNe have been previously

published (Margolin et al. 2006; Chavez Montes et al. 2014). In a biological context, an interaction between gene A and gene B would indicate that gene A and gene B participate in the same physiological process and, if gene A is a TF and gene B is not TF, the interaction would suggest that the gene A is a transcriptional regulator of gene B (Chavez Montes et al. 2014).

Network construction was concentrated on the 2088 TF probeset and obtained at three data processing inequality (DPI) values, 0.0, 0.1, and 0.2 (Table 14 and Figure 21) . DPI is a known information property explained in (Margolin et al. 2006). A DPI 0.0 means when a three-node triangle is present, the interaction with lowest MI will be removed,as this interaction is considered to represent an indirect interaction. At DP1 1.0 interactions are removed (Margolin et al. 2006) .

Table 14. Number of nodes and edges in the TFsSeedNet and FullSeedNet obtained at DPI 0.0, 0.1 and 0.2

DPI	Nodes	Edges
TFsSeedNet		
0.0	2068	3745
0.1	2068	16608
0.2	2068	59079
FullSeedNet		
0.0	22553	89045
0.1	22553	372424
0.2	22553	2570778

Nodes and edges present at DPI 0.0 are part of the network obtained at DPI 0.1 and both are part of the network obtained at DPI 0.2.

After transforming the ARACNe output adjacency files into Cytoscape compatible tables, we obtained the TFs-only (TFsSeedNet) and complete (FullSeedNet) databases. Table 14 shows that the number of edges in the network increases dramatically from DPI 0.0 to DPI 0.1 to DPI 0.2. For this work, a graphical representation of TFSeedNet was obtained at DPI 0.0 while FullSeedNet was obtained at DPI 0.1. The DPI value of FullSeedNet was used to preserve interaction that are removed in DPI 0.0 in order to have a wider view of the genes that could interact with different TFs (Figure21).

Using the Cytoscape tools, we compared the FullSeedNet with DEGs specific to desiccation tolerance mutants at each sampled time point to construct interaction subnetworks. We observed that these subnetworks are clearly divided in two sections, corresponding to genes that are upregulated in desiccation intolerant mutants respect to the tolerant controls and vicervesa (Figure 21).

When TF genes that fail to be activated in desiccation intolerant mutants were mapped to the TFsSeedNet at 15, 17 and 21 DAF, we found that they formed two main co-expression subnetworks, which we termed TFsSeed-sNetDT1 and TFsSeed-sNetDT2 (Figure 22 and 23)

TFsSeed-sNetDT1 was mainly composed of members of the APETALA2/Ethylene Responsive Factor (AP2-ERF) TF gene family (*DREB2D*, AT1G75490; *RAP2-13*, AT1G22190; *ERF15*, AT4G31060; *ERF23,* AT1G01250 and *ERF8*, AT1G53170), but also contained members of the Basic Leucine Zipper family (*bZIP12*, AT2G41070), MADS-BOX (*AGL67, AT1G77950*), C2H2 Zinc-Finger (*PEI1*;

AT5G07500) and Delay of Germination-Like (DOG-like) (AT4G18650, that we named *DOGL4*;) TF families(Figure 22).

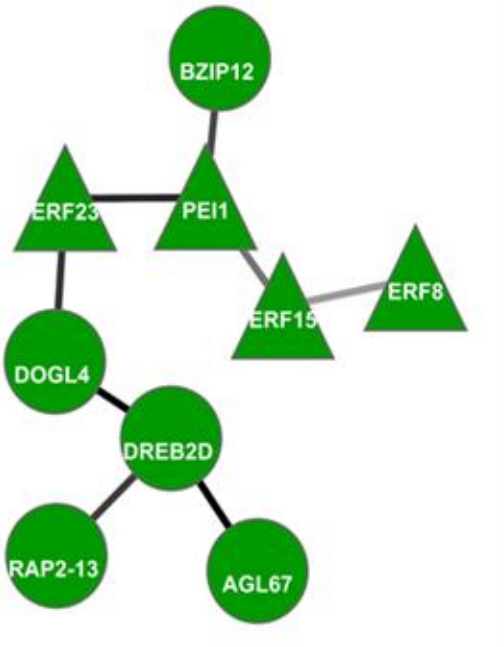

Figure 22. Overview of the TFsSeed-sNetDT1 obtained at DPI0.0. Subnetwork of the most representative downregulated in desiccation intolerant lines. Genes are represented as nodes and inferred interactions as edges (lines). Nodes are colored grey, except upregulated TFs in red and downregulated TFs in green from DT analysis. Edge width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges. Triangular nodes represent genes expressed exclusively at 15DAF and circular nodes represent genes expressed at more than one time point.

TFsSeed-sNetDT2 includes three members of the NAC family (*ATAF1*, AT1G01720; *ANAC032*, AT1G77450 ; *ANAC089*, AT5G22290), two from the AP2-ERP family (*DREB2G*, AT5G18450 and *ERF12*, AT1G28360), two from the C3H Zinc-Finger family (*SOM*, AT1G03790 and *TZF5*, AT1G03790), two from the Plant

AT-rich sequence- and Zinc-binding protein family (PLATZ) (AT1G21000, that we named PLATZ1 and AT1G76590, that we named PLATZ2), one to the bZIP family (*ABI5*, AT2G36270) and one from the Ethylene Insensitive 3 family (*EIL5*, AT5G65100) (Figure 23).

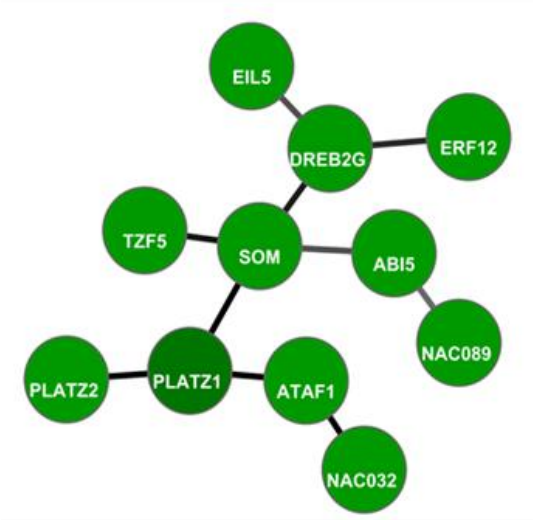

Figure 23. Overview of the TFsSeed-sNetDT2 obtained at DPI0.0. Subnetwork of the most representative downregulated in desiccation intolerant lines. Genes are represented as nodes and inferred interactions as edges (lines). Nodes are colored grey, except upregulated TFs in red and downregulated TFs in green from DT analysis. Edge width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges. Triangular nodes represent genes expressed exclusively at 15DAF and circular nodes represent genes expressed at more than one time point.

5.6. Co-expression subnetworks of differentially expressed genes and potential targets

We then searched for genes that are co-expressed with the TFs in TFsSeed-sNetDT1 and TFsSeed-sNetDT2 in the FullSeedNet; these coexpression

subnetworks were denominated FullSeed-sNetDT1 (Figure 24A, B and C) and FullSeed-sNetDT2 (Figure 24D and E).

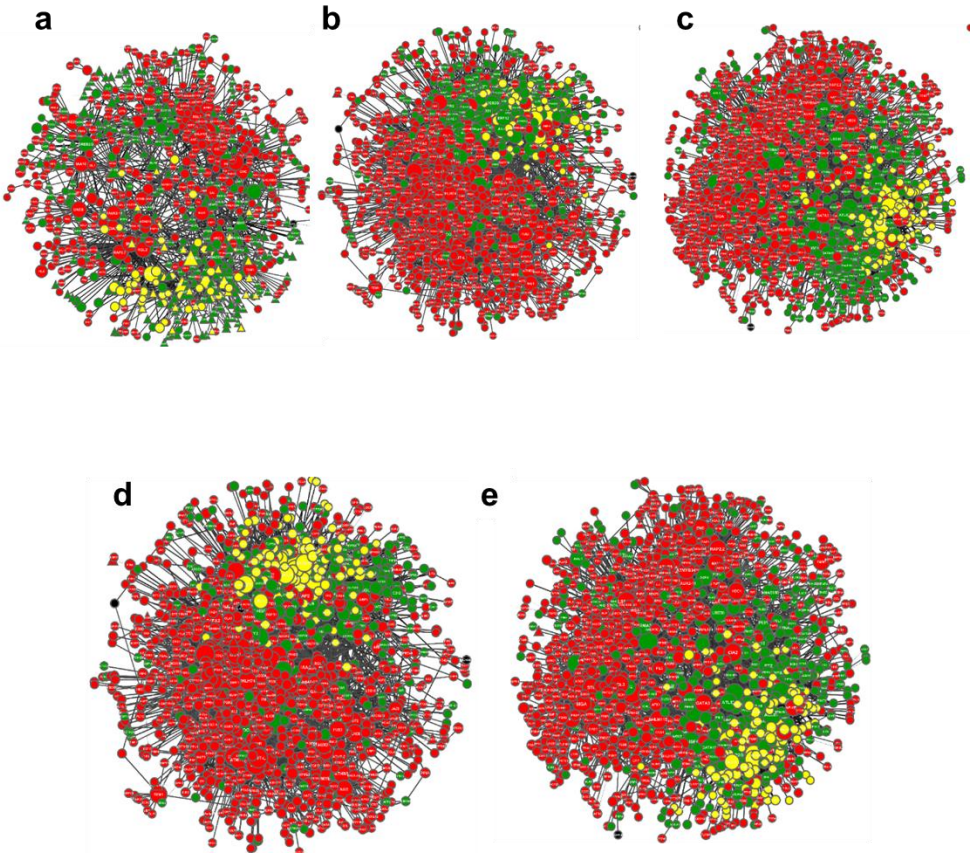

Figure 24. The whole temporal gene regulatory network FullSeedNet. (a), (b) and (c) represent the FullSeed-sNetDT1 related to nutrient storage. (d) and (e) represent the FullSeed-sNetDT2 related to cellular protection mechanism. Overview of the FullSeedNet obtained at DPI 0.1 at (a) 15, (b) and (d) 17 and (c) and (e) 21 DAF. Genes are represented as nodes and inferred interactions as edges. Nodes are colored grey, except upregulated genes in red and downregulated genes in green from DT analysis. Edges width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges. Yellow nodes in represent the FullSeed-sNetDT1 and FullSeed-sNetDT2 at 15, 17 and 21 DAF.

These genes represent the potential effector genes activated by the TF that are part of the two regulatory subnetworks that failed to be activated in desiccation

intolerant mutants. FullSeed-sNetDT1 was composed of a total of 280 genes (Figure 25), for which the most significantly enriched categories included nutrient reservoir activity and lipid storage (Figure 29).

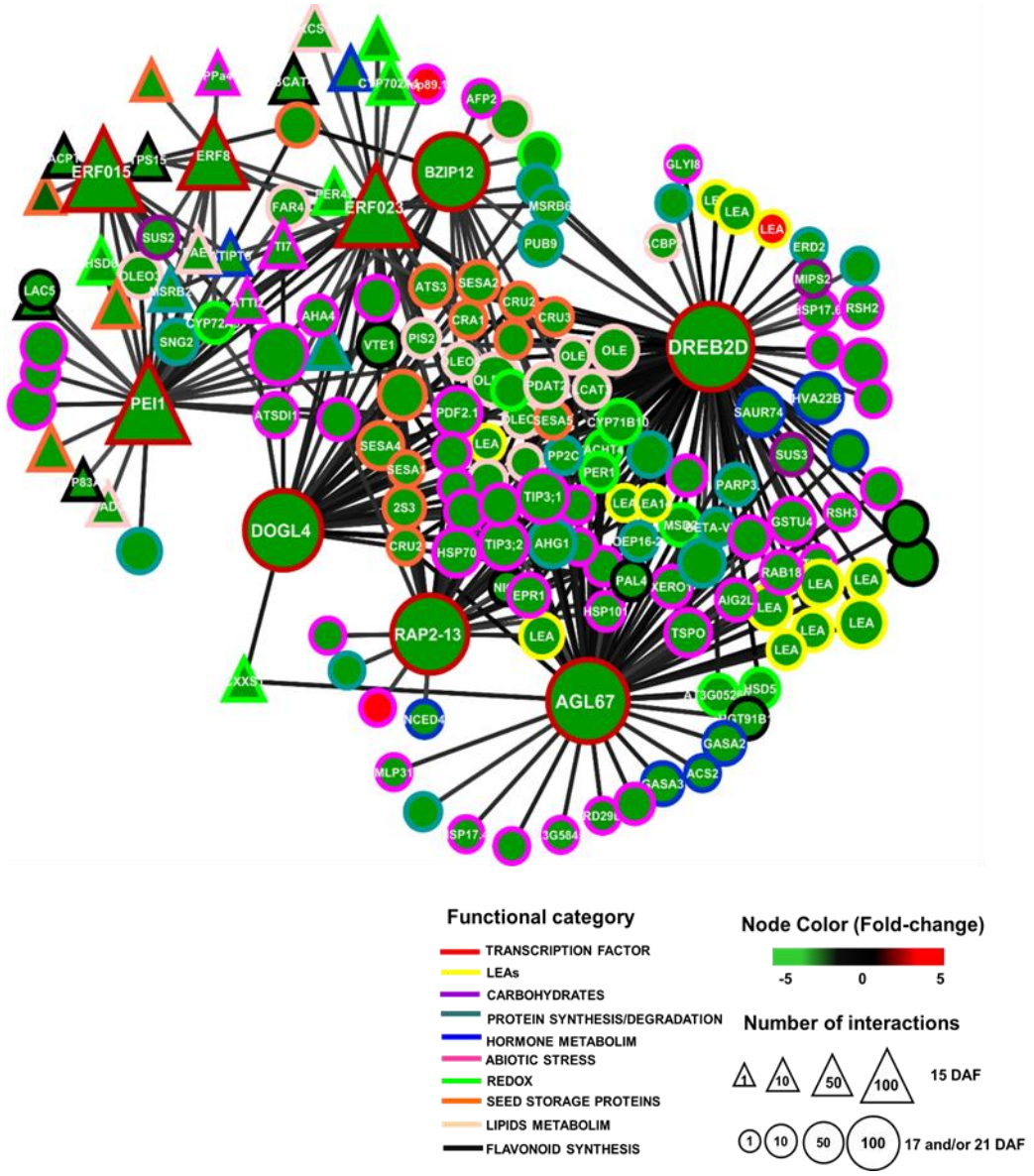

Figure 25. snetFullDT1 related to stress and nutrient storage. Subnetworks were obtained from FullSeedNet at DPI 0.1. Genes are represented as nodes and inferred interactions as edges. Information box shows attributes in the networks. Triangular nodes represent genes expressed exclusively at 15DAF and circular nodes represent genes

expressed at more than one time point. The border node color represent the functional category see Supplementary Table 18 and 19. Edge width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges.

When the FullSeed-subNetDT1 was analyzed at each developmental stage (15, 17 and 21 DAF), we found that at 15 DAF it is composed of five TF nodes, DOGL4, PEI1 ERF15, ERF8 and ERF23, that interact (MI > 0.5 and pvalue<1e-35) (Figure 26) with effector genes mainly involved in fatty acid, TAG biosynthesis and storage protein genes such as those encoding the albumin proteins SESA1, SESA2, SESA3, SESA4 and SESA5, cruciferin proteins CRU2 and CRU3, and oleosin proteins OLE1, OLE2 and OLE4. At 17 DAF, TFsSeed-sNetDT1 significantly changed maintaining only DOGL4 as a major node and three other TFs (DREB2D, AGL67 and RAP2-13) were incorporated as nodes in this subnetwork. DREB2D appears to be a key TF in TFsSeed-sNetDT1 because it has the largest number of interactions, and it is strongly co-expressed (MI >0.6 and p-value <1E-40) with AGL67 and DOGL4. These three TFs have as predicted targets genes related to stress tolerance, including LEA genes (*LEA4-1, LEA14*), heat shock proteins (*HSP17.4, HSP101, HSP70*), dehydrins (*XERO1, RAB18*) and aquaporins (*TIP3;1*) (Figures 25 and 26). The evolution of FullSeed-subNetDT1 during seed maturation suggest that this subnetwork is initially involved in regulating genes involved in the accumulation of reserve compounds and then in DT acquisition.

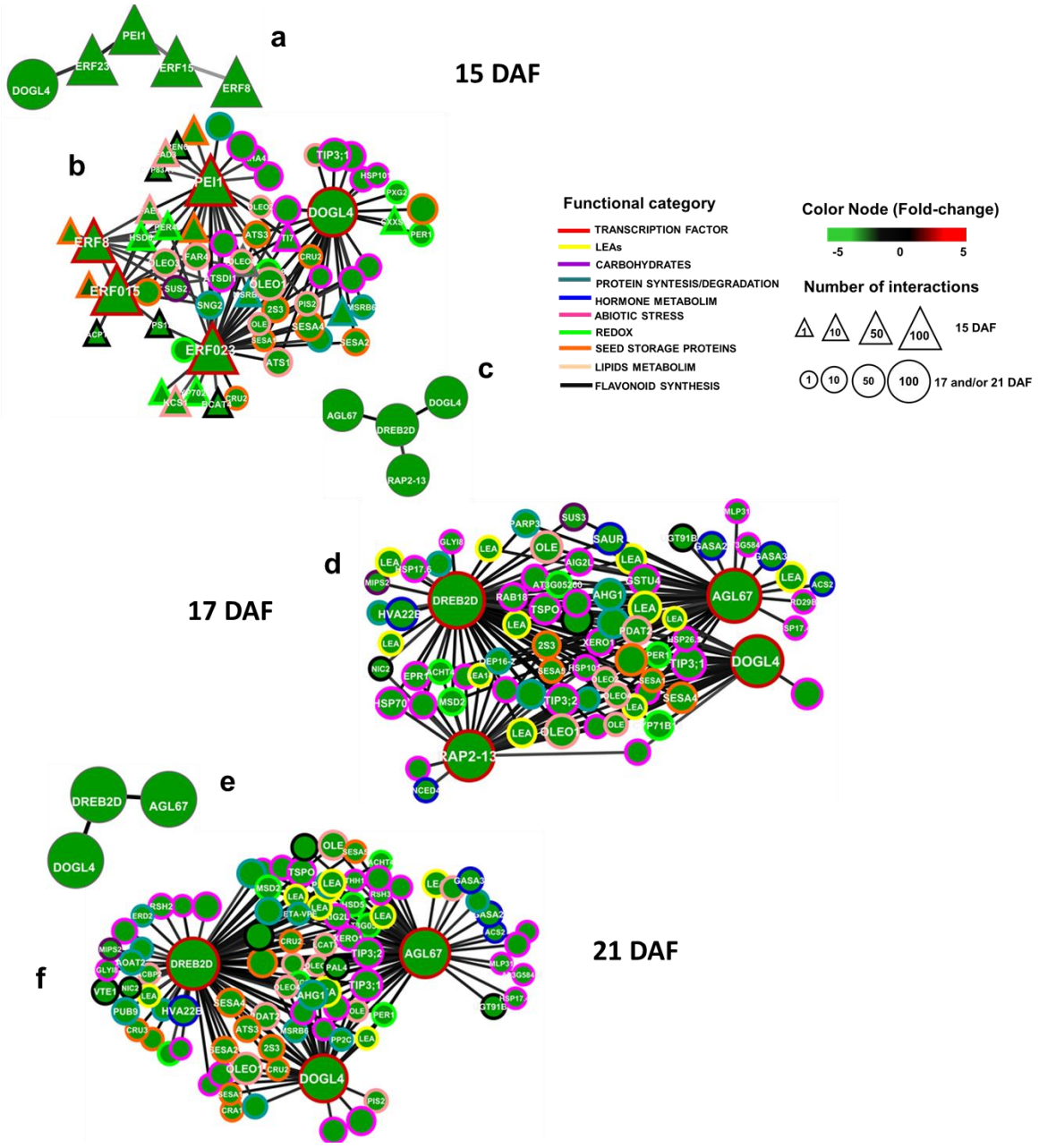

Figure 26. Temporal FullSeed-subNetDT1, TFsSeed-subNetDT1. In each panel (a,c,e) graphs represent TFsSeed-sNetDT1 and (b,d,f) graphs represent the interactions transcription factors in (a,c,e) and non transcription factors from the FullSeedNet. Graphs correspond to genes for FullSeed-sNetDT1, involved in nutrient storage, at 15, 17 and 21 DAF. Genes are represented as nodes and inferred interactions as edges. Network attributes are presented in the information box. Nodes with a triangle shape represent genes that are expressed only at 15DAF and circles genes that are expressed at least at two of the sampled stages. Node borders represent functional categories.Edge width and color

intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges.

FullSeed-subNetDT2 was composed by 317 genes, which represent 17% of all downregulated genes from tolerance-specific mutant differences (Figure 27). This subnetwork is enriched in effector genes in the following categories: redox activity, LEA protein and response to abiotic stimulus (Figure 29).

In FullSeed-subNetDT2, we identified ATAF1 as a TF with an important (81) number of interactions, and is strongly co-expressed (MI > 0.5 and pvalue < 1E-35) with other TFs in the subnetwork such as PLATZ1, SOM and DREB2G. These four TFs are central in FullSeed-subNetDT2 because each has a large number of interactions and they share common nodes that correspond to genes involved in cellular protection. For example, among the genes involved in stress resistance that are predicted to be activated by these TFs we found 3 LEA group 4 genes (*LEA4-1*, *LEA4-5* and *LEA4-2*), two LEA group 1 genes (*ATEM1* and *ATEM6*), three genes involved in raffinose synthesis (*UGE3, SUS3, GOLS1*), two catalase genes (*CAT2, CAT3*), one superoxide dismutase gene (*SOD*), two small heat shock protein genes (*HSP17.4* and *HSP17.6*), and two genes involved in ABA signaling (*PP2C* and *ABI5*) (Figure 27 and 28).

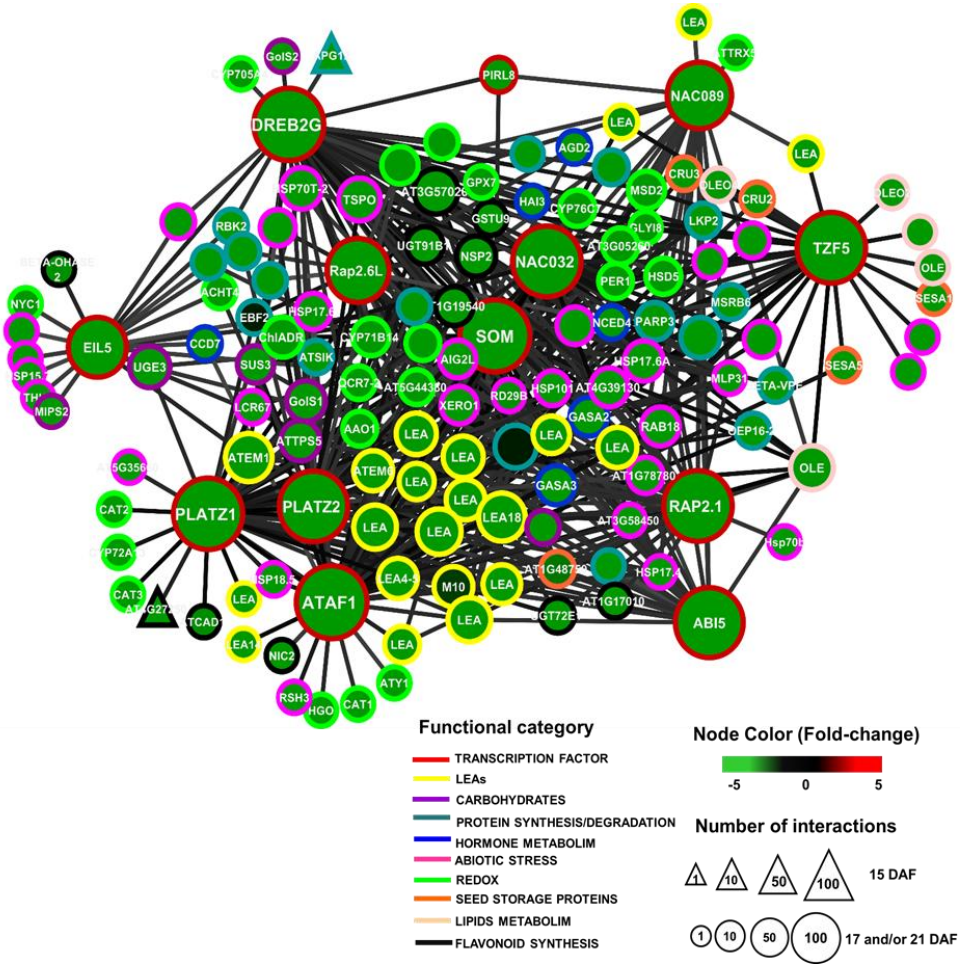

Figure 27. **FullSeed-sNetDT2 related to cellular protection mechanisms.** Subnetworks were obtained from FullSeedNet at DPI 0.1. Genes are represented as nodes and inferred interactions as edges. Information box shows attributes in the networks. Triangular nodes represent genes expressed exclusively at 15DAF and circular nodes represent genes expressed at more than one time point. The border node color represents the functional category. Edge width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges.

Interestingly, FullSeed-subNetDT2 was specifically activated at 17 DAF and became more complex at 21 DAF, which corresponds to the developmental stages at which rapid and total water loss occurs (Figure 28). ATAF1, a NAC TF, was the

node with the highest number of interactions in FullSeed-subNetDT2, and its potential targets are stress protection genes. In FullSeed-subNetDT2 we identified four other TFs with a high number of interactions, namely SOM, DREB2G, PLATZ1 and PLATZ2, which also preferentially interacted with genes associated with stress protection processes (Figure 28).

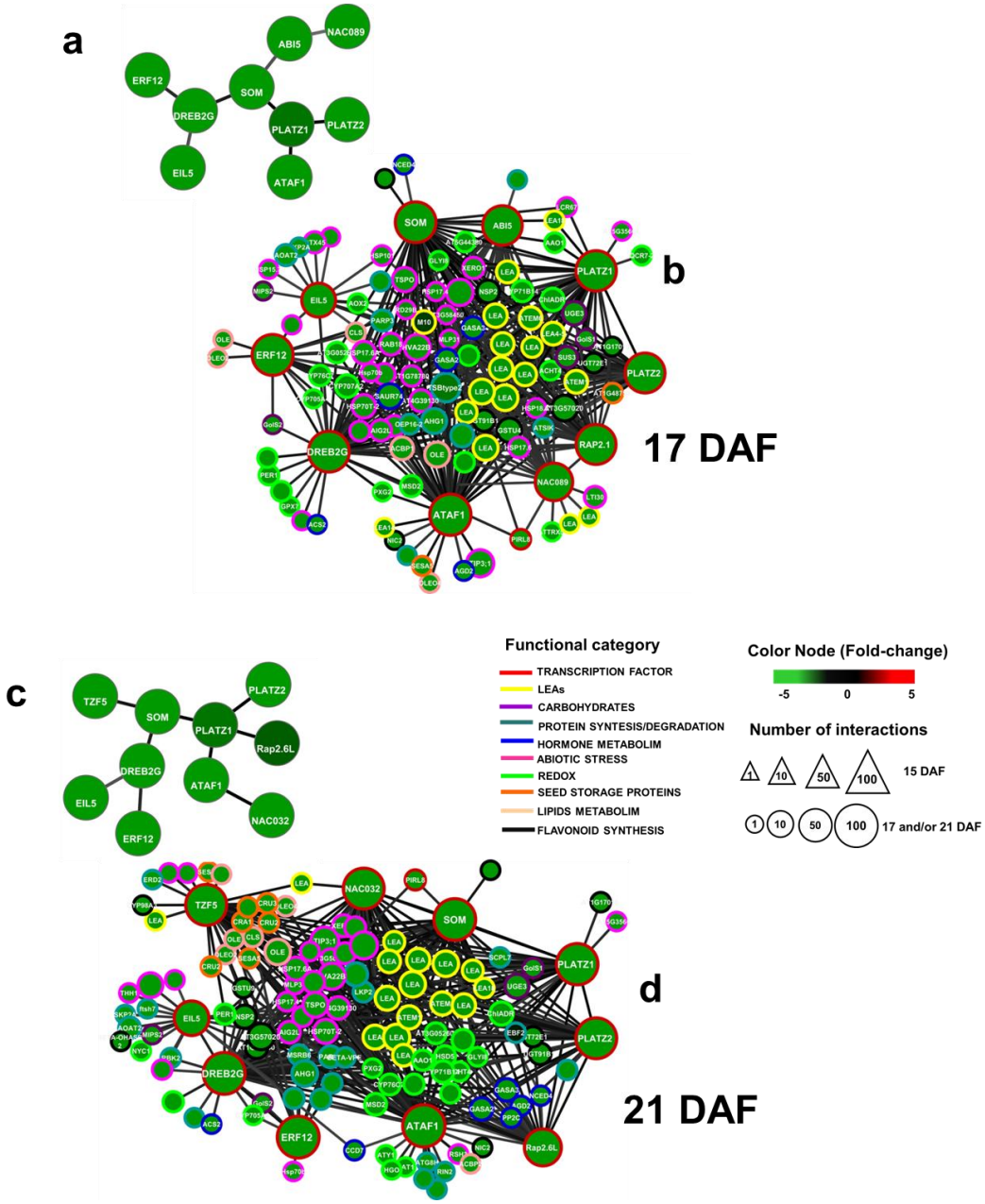

Figure 28. Temporal FullSeed-subNetDT2, TFsSeed-subNetDT2. Subnetworks were obtained from FullSeedNet at DPI 0.1. Genes are represented as nodes and inferred interactions as edges. Information box shows attributes in the networks. Triangular nodes represent genes expressed exclusively at 15DAF and circular nodes represent genes expressed at more than one time point. The border node color represents the functional category. Edge width and color intensity is proportional to the Mutual Information (MI) value of the interaction, with higher MI values corresponding to thicker and darker edges.

In summary, our network analysis supports a model in which at early stages FullSeed-subNetDT1 regulates nutrient production genes and at later stages FullSeed-subNetDT1 and FullSeed-subNetDT2 regulates DT genes (Figure 29).

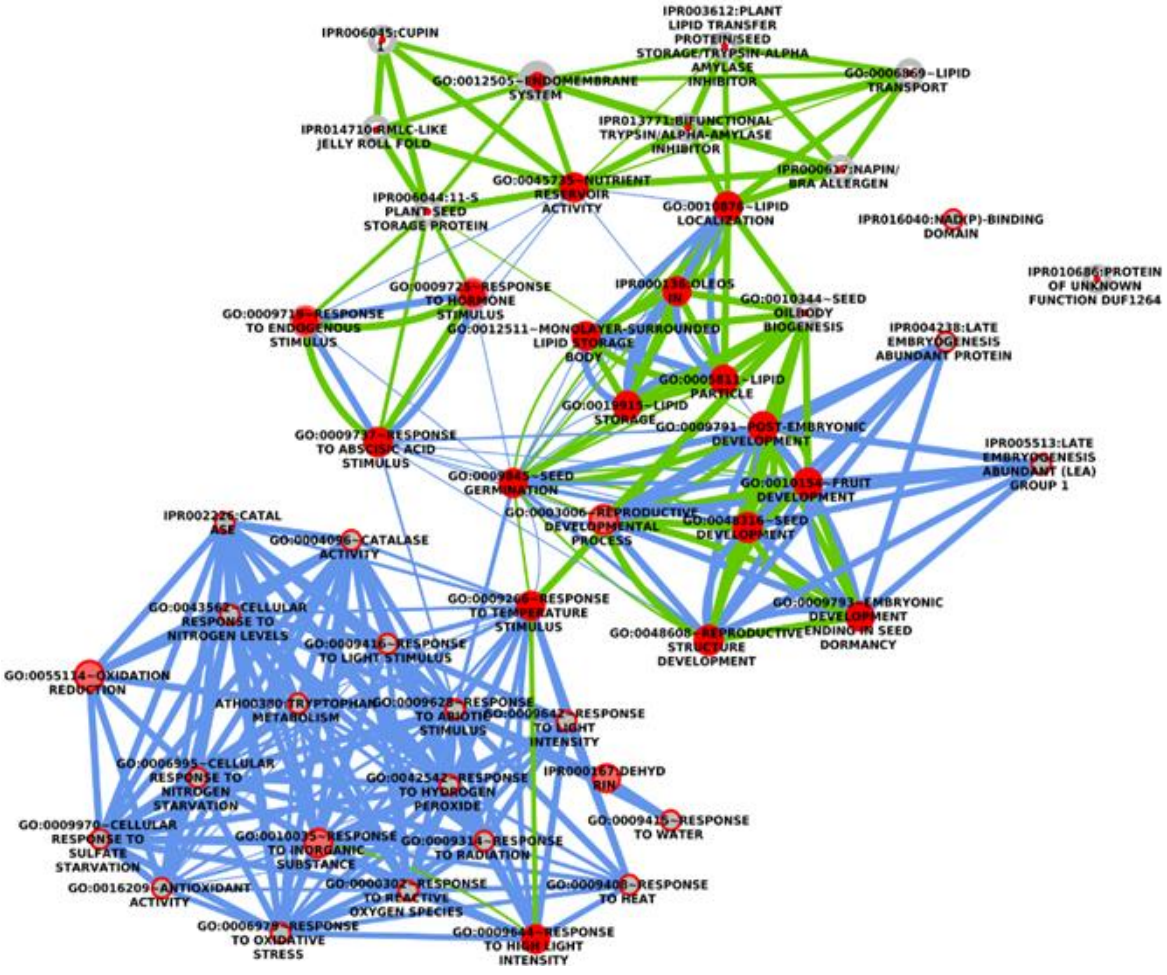

Figure 29. Comparative analysis of the functional enriched category between FullSeed-sNetDT1 and FullSeed-sNetDT2 . Enrichment Map

graph of category enrichments calculated with David for genes in snetFullDT1 and snetFullDT2. Node (inner circle) size corresponds to the number of genes in snetFullDT1 within the geneset. Node border (outer circle) size corresponds to the number of genes in snetFullDT2 within the geneset Nodes in red belong to genes categories that are only enriched in snetFullDT1**.** Nodes border red represent gene categories that are enriched snetFullDT2. Edge size corresponds to the number of genes that overlap between the two connected genesets. Green edges corresponds to snetFullDT1 and blue corresponds to snetFullDT2.

We then searched for enriched cis-regulatory elements in the promoters of target genes in each subnetwork. In general, the enriched motifs were ABA signaling-related (ABF binding site motif, ABRE binding site motif, ACGT ABRE motif A2OSE). The seed specific motif RY-repeat was found in the three stages of FullSeed-subNetDT1, whereas dehydration and drought responses (ABRE-like binding site motif, DRE core motif, CBF1 BS in COR15a, AtMYC2 BS in RD22) were enriched in FullSeed-subNetDT2 (Table 15). This finding also supports our model for distinct but complementary roles for FullSeed-subNetDT1 and FullSeed-subNetDT2 in the acquisition of seed DT.

Table 15. Motifs enrichment in DT subnetworks.

Genes of subnetworks	Motif identified	No. Genes with this element	No. of elements	P-value
snetFullDT1 at 15DAF	ABRE-like binding site motif	28	42	< 10 - 3
	AtMYC2 BS in RD22	42	59	< 10 - 4
	GADOWNAT	17	21	< 10 - 3
	RY-repeat promoter motif	9	20	< 10 - 3
	ACGTABREMOTIFA2OSEM	24	33	< 10 - 4
	CACGTGMOTIF	25	66	< 10 - 4

	MYCATERD1	42	59	< 10 - 4
snetFullDT1 at 17DAF	ABFs binding site motif	12	15	< 10 - 5
	ABRE-like binding site motif	44	85	< 10 - 10
	ACGTABREMOTIFA2OSEM	36	70	< 10 - 10
	DRE core motif	30	41	< 10 - 4
	GBF1/2/3 BS in ADH1	7	14	< 10 - 4
	RY-repeat promoter motif	11	24	< 10 - 4
	ABRE binding site motif	19	23	< 10 - 9
	ABREATRD22	10	12	< 10 - 5
	CACGTGMOTIF	36	108	< 10 - 10
	GADOWNAT	28	42	< 10 - 10
	GBOXLERBCS	12	15	< 10 - 6
	Z-box promoter motif	8	8	< 10 - 3
snetFullDT1 at 21 DAF	ABFs binding site motif	17	22	< 10 - 8
	ABRE-like binding site motif	59	113	< 10 - 10
	ACGTABREMOTIFA2OSEM	52	89	< 10 - 10
	CACGTGMOTIF	54	148	< 10 - 10
	GADOWNAT	35	46	< 10 - 10
	GBOXLERBCS	16	21	< 10 - 8
	RAV1-B binding site motif	21	23	< 10 - 3
	ABRE binding site motif	25	31	< 10 - 10
	ABREATRD22	12	14	< 10 - 5
	AtMYC2 BS in RD22	51	69	< 10 - 4
	DRE core motif	33	41	< 10 - 3
	GBF1/2/3 BS in ADH1	10	20	< 10 - 5
	MYCATERD1	51	69	< 10 - 4
	RY-repeat promoter motif	10	22	< 10 - 3
snetFullDT2 at 17DAF	ABFs binding site motif	20	24	< 10 - 10
	ABRE-like binding site motif	74	158	< 10 - 10
	ACGTABREMOTIFA2OSEM	67	129	< 10 - 10
	CBF1 BS in cor15a	5	5	< 10 - 3
	GADOWNAT	49	74	< 10 - 10
	GBOXLERBCS	19	23	< 10 - 10
	RY-repeat promoter motif	12	26	< 10 - 3
	ABRE binding site motif	32	40	< 10 - 10
	ABREATRD22	18	23	< 10 - 10
	CACGTGMOTIF	56	178	< 10 - 10
	DRE core motif	43	59	< 10 - 4
	GBF1/2/3 BS in ADH1	10	20	< 10 - 5

	Hexamer promoter motif	22	25	< 10 - 3
	Z-box promoter motif	18	18	< 10 - 10
snetFullDT2 at 21DAF	ABFs binding site motif	24	31	< 10 - 10
	ABRE-like binding site motif	78	169	< 10 - 9
	ACGTABREMOTIFA2OSEM	72	136	< 10 - 10
	CACGTGMOTIF	64	192	< 10 - 10
	GADOWNAT	51	72	< 10 - 9
	GBOXLERBCS	22	29	< 10 - 9
	Z-box promoter motif	15	15	< 10 - 6
	ABRE binding site motif	36	47	< 10 - 9
	ABREATRD22	19	24	< 10 - 9
	AtMYC2 BS in RD22	63	87	< 10 - 4
	DRE core motif	41	53	< 10 - 3
	GBF1/2/3 BS in ADH1	12	24	< 10 - 6
	MYCATERD1	63	87	< 10 - 4

Survey was performed in 1000 bp maximum upstream range cutting off at adjacent genes. Data obtained using the Athena Web tools (http://www.bioinformatics2.wsu.edu/cgibin/Athena/cgi/home.pl).

5.7. Mutants of the major nodes of FullSeed-subNetDT1 and FullSeed-subNetDT2 have reduced desiccation tolerance.

If the genes identified as major nodes in FullSeed-subNetDT2 indeed play roles in DT, it would be expected that the corresponding mutants should present some degree of desiccation intolerance. We therefore used T-DNA insertion mutants for the TFs *PLATZ1*, *PLATZ2*, *AGL67*, *DREB2D*, *ERF23*, and *ATAF1* to determine whether they had a reduced germination percentage as a consequence of decreased DT. We observed that the germination percentage of dried *ataf1*, *agl67*, *platz1* and *platz2* seeds respect to their WT was reduced to 76, 75, 77, and 53%, respectively (Figure 30) while *dreb2d-1* and *erf23-2* did not show any phentyope.

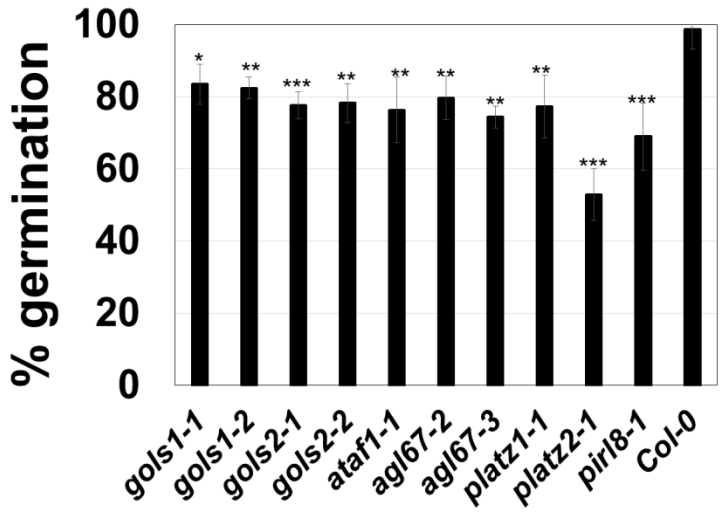

Figure 30. TF node mutants with reduced DT. Germination percentages of seeds six days after sowing. Values are the means ±SD of four biological replicates n=100 seeds per replica. Bars with asterisks are significantly different from the control (Student's t-test, * = P<0.05, ** = P < 0.01, *** P<0.0001).

platz1-1 and *plaltz2-1* showed a significantly reduced germination percentage when stored for seven and eight weeks at 10 % relative humidity (Figure 31), suggesting PLATZ1, PLATZ2 are important for DT in the dry seed stage.

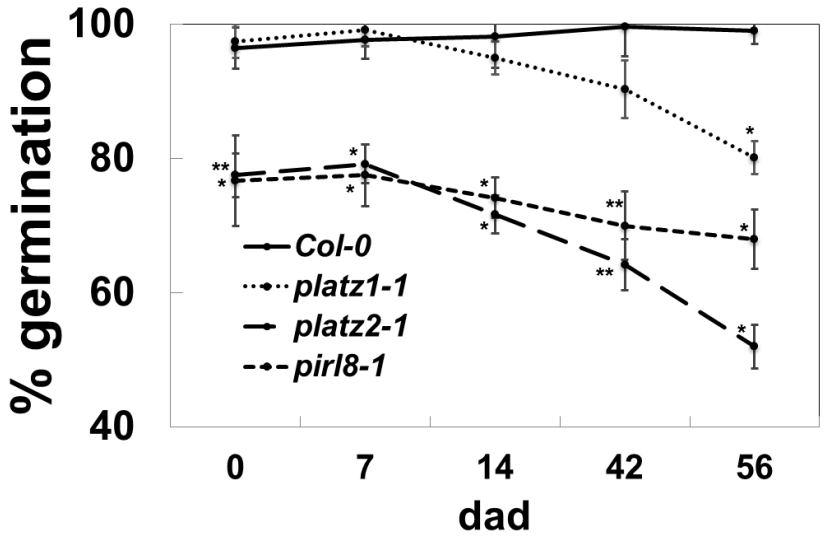

Figure 31. Germination percent of Col-0, *platz1-1*, *platz2-1* and *pirl8-1* seeds stored for 0,7,14, 43 and 56 days after desiccation (dad). Values are mean ± SD of three biological replicates. Bars with asterisks are significantly different from the control (Student's t-test, * = P<0.05, ** = P < 0.01, *** P<0.0001).

TFs identified as major nodes in our DT subnetworks should regulate the expression of target genes directly involved in DT, such as those involved in raffinose biosynthesis or encoding LEA proteins. To determine whether FullSeed-subNetDT2 effector genes play a role in DT, we determined the germination capacity of T-DNA insertion mutants for some of these effector genes. A mutant for the non-TFs gene *PIRL8* had decreased germination and ABA insensibility (Figure 32).

PIRL8 belongs to a plant-specific class of intracellular LRRs that likely mediate protein interaction, possibly in the context of signal transduction (Forsthoefel et al. 2005). *PIRL8* is neighbor of *SOM*, *DREB2G* and *ANAC089* and these TFs are coexpressed with an important number of genes involved in cellular protection mechanisms (Figure 27).

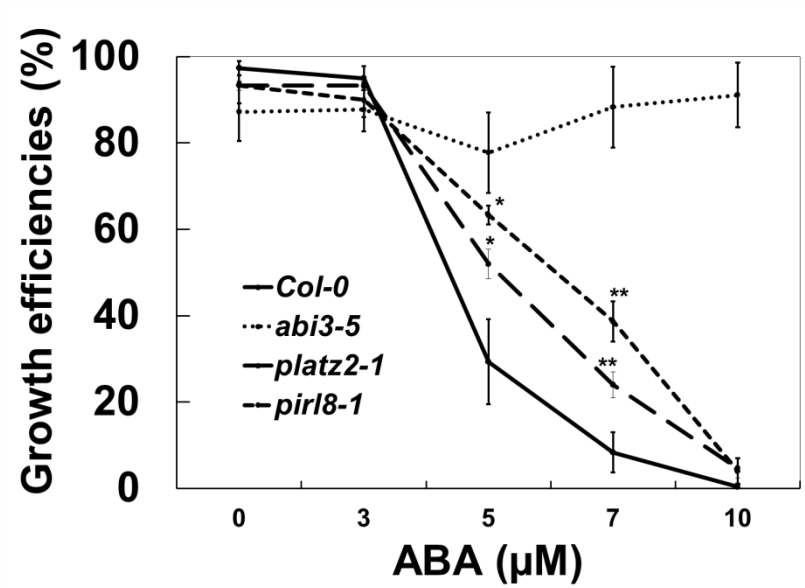

Figure 32. Effects of ABA on growth efficiency. Scored as the % of seedlings with green cotyledons in each treatment, note that the highest percentage is detected for *abi3-5* and the lowest for the WT control. Seedlings were grown on MS medium with 0, 3, 5, 7 and 10 µM ABA and their phenotypes scored 15 days after sowing. Values are the means ±SD of three replicates n=50 per replicate. Bars with asterisks are significantly different from the control (Student's t-test, * = P<0.05, ** = P < 0.01, *** P<0.0001).

We also found that mutants for the *GOLS1* and *GOLS2* genes (which encode enzymes involved in raffinose biosynthesis) had a reduction in germination of 20 to 30% with respect to their wild-type control (Figure 30). These results suggest that target genes of the major nodes of FullSeed-subNetDT2 do indeed play important roles in DT. Lower germination efficiency in these mutants compared with their WT could be due to the formation of defective embryos rather than to a defect in seed DT. Therefore, to determine whether the lower germination phenotype observed in effector mutants is due to developmental defects during embryo formation or to a

reduced DT, the morphology of green and dry seeds for each mutant was evaluated. We did not detect any apparent embryo defect in all the TF or effector genes mutants that were tested in this work (Figure 33).

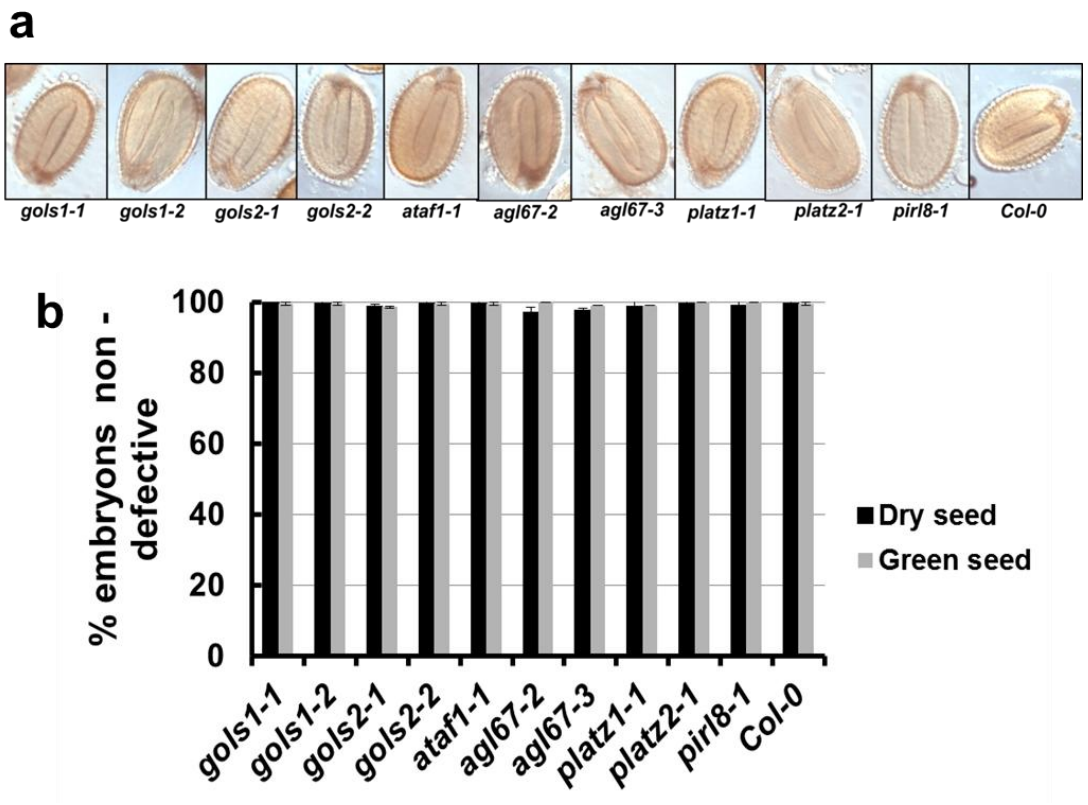

Figure 33 Embryos defective in T-DNA mutants seeds. (a) The visualization and (b) counting defective seeds were performed clearing dry and mature green seeds with Hoyer's solution overnight and visualized with microscope equipped with Nomarski optics. Values are means ± SD of two biological replicates for 100 seeds.

5.8. Overexpression of *PLATZ1*, *AGL67* or *DREB2D* partially rescues the desiccation intolerance phenotype of *abi3-5*.

If the TFs identified as major nodes in FullSeed-subNetDT1 and FullSeed-subNetDT2 act downstream of ABI3, FUS3 and LEC1 and play important roles in activating effector genes involved in DT, then overexpression of some these TFs in

a desiccation intolerant background, for example *abi3-5*, should partially revert the desiccation intolerance phenotypes of these mutants. To test this, we expressed *AGL67*, *DREB2D* and *ERF23* from TFsSeed-subNetDT1 and *PLATZ1*, *PLATZ2* and *DREB2G* from TFsSeed-subNetDT2 under the control of the 35S promoter in the *abi3-5* background (Figure 34-36). Dry seeds were collected and stored for 0, 1, 2 and 4 weeks at 10 % of RH and then tested for germination efficiency. As previously reported (Ooms et al. 1993) (Figure 34-36), *abi3-5* seeds rapidly lose viability after desiccation, and after one week of storage germination was reduced to less than 5%. In contrast, seeds from *abi3-5/35S::PLATZ1* (Figure 34) *abi3-5/35S::AGL67* (Figure 35) and *abi3-5/35S::DREB2G* (Figure 36) lines showed a germination rate of 25, 30 and 12%, respectively after 4 weeks of storage. These results demonstrate that overexpression of some of the TFs identified as major nodes in TFsSeed-subNetDT1 and TFsSeed-subNetDT2 can partially rescue DT in *abi3-5* seeds.

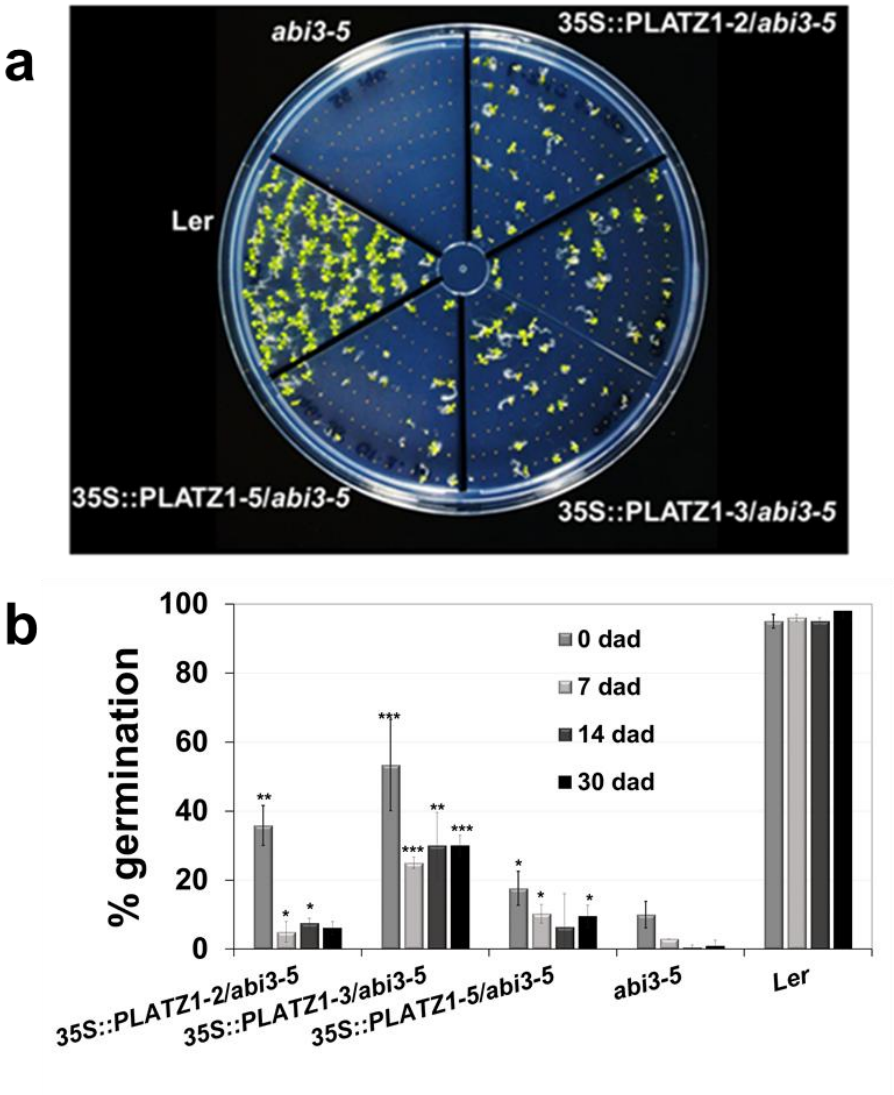

Figure 34. PLATZ1 overexpression partially rescues *abi3-5* DI phenotype. (a) Seeds were grown on MS medium and survival was scored 10 days after sowing. (b) Germination percentage of WT (Ler), *abi3-5* and 35S::PLATZ1/*abi3-5* seed stored for 0, 7 , 14 and 30 days after desiccation. Values are mean ± SD of three biological replicates. Bars with asterisks are significantly different from the control (Student's t-test, * = P<0.05, ** = P < 0.01, *** P<0.0001).

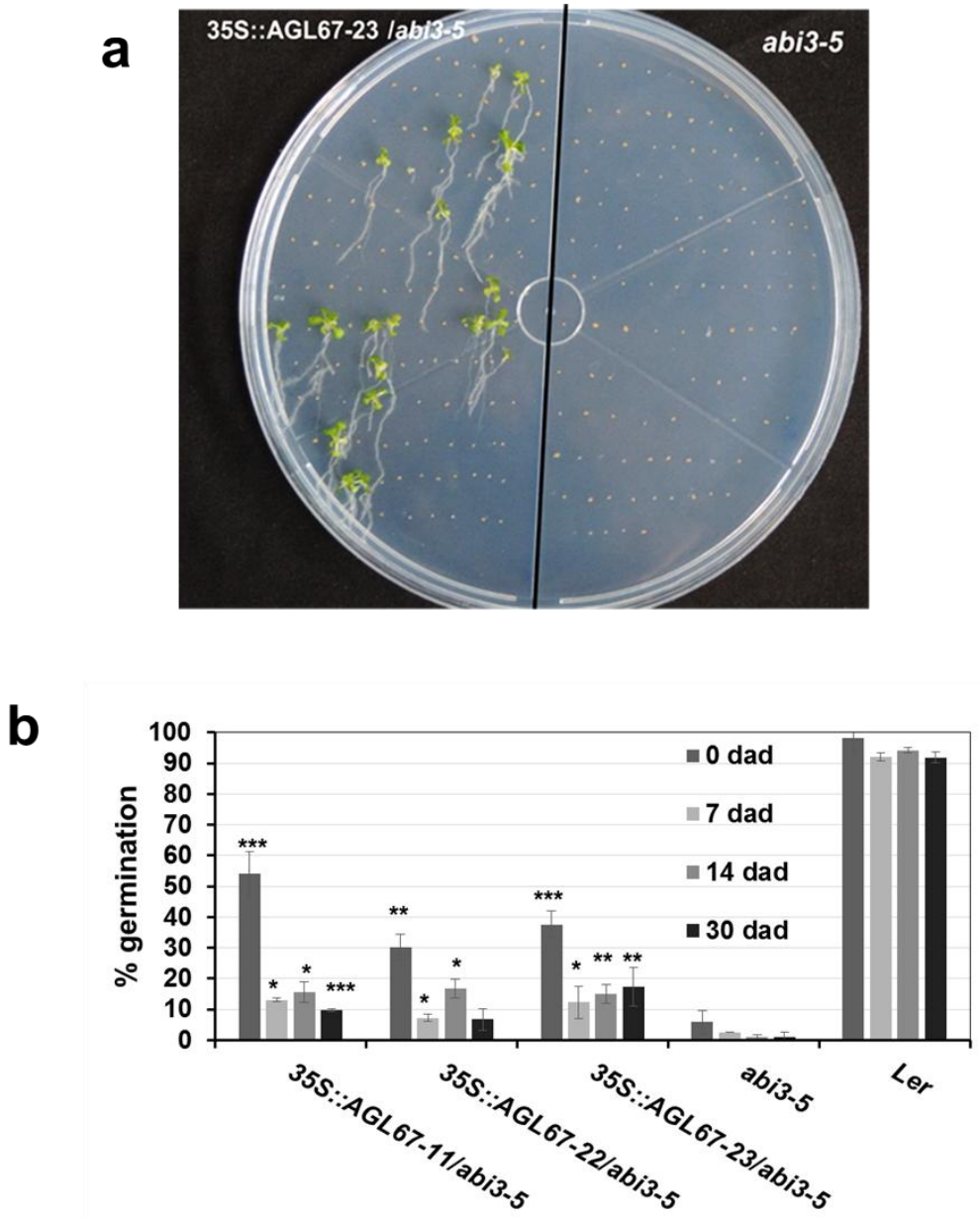

Figure 35. Overexpression of AGL67 rescues *abi3-5* intolerance desiccation phenotype. (a) Germination efficiency of 35S::AGL67-23/*abi3-5*; (b) Germination percentage of Ler, *abi3-5* and 35S::AGL67-23/*abi3-5* lines. For all experiments seeds at 21 DAF were stored for 0, 7, 14 and 30 days after desiccation (dad) to assess DT. Values are means ± SD of three biological replicates for 100 seeds. Statistical analysis of the data was done using a student t-test. Bars with asterisks are significantly different from the control (* = P<0.05, ** = P < 0.01, *** P<0.0001).

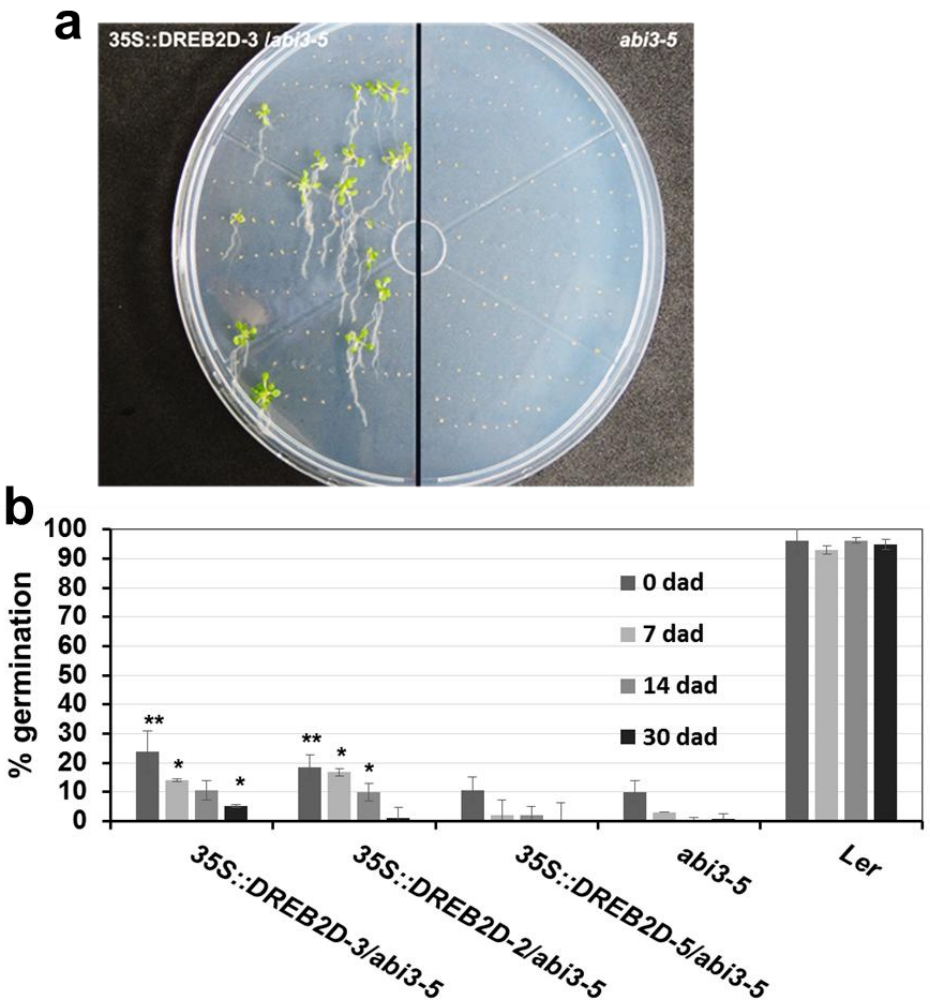

Figure 36. Overexpression of DREB2D rescues *abi3-5* intolerance desiccation phenotype. (a) Germination efficiency of 35S::DREB2D-3/*abi3-5;* (b) Germination percentage of Ler, *abi3-5* and 35S::DREB2D-3/*abi3-5* lines. For all experiments seeds at 21 DAF were stored for 0, 7, 14 and 30 days after desiccation (dad) to assess DT. Values are means ± SD of three biological replicates for 100 seeds. Statistical analysis of the data was done using a student t-test. Bars with asterisks are significantly different from the control (* = P<0.05, ** = P < 0.01, *** P<0.0001).

The capacity of PLATZ1, AGL67 and DREBG2 to partially complement the desiccation intolerance phenotype of *abi3-5* seeds, should in principle involved the activation of the effector target genes with which these transcription factors interact in TFsSeed-subNetDT1 and TFsSeed-subNetDT2. To confirm that the corresponding predicted target genes are indeed transcriptionally activated by these TFs, we evaluated the effect of the expression of the *35S::PLATZ1* gene construct (Figure 37) in the *abi3-5* genetic background, on the expression of some of the putative PLATZ1 target genes with a high MI value, such as LEA related genes, LEA4, XERO1; antioxidants, NRS; oligosaccharides synthesis, UGE3, GOLS1 and TTP5; oleosins, OLE3 and TFs, ABI5, PLATZ2 and ATAF1(Figure 38A). Using two independent *PLATZ1* overexpressing lines (Figure 38) we found that in most cases the putative target genes had a 1.5- to 4-fold higher expression in the 35::*PLATZ1* overexpressing *abi3-5* lines than in the *abi3-5* control line. These results suggest that the expression of the genes predicted to be regulated by PLATZ1 is indeed directly or indirectly activated by this TF.

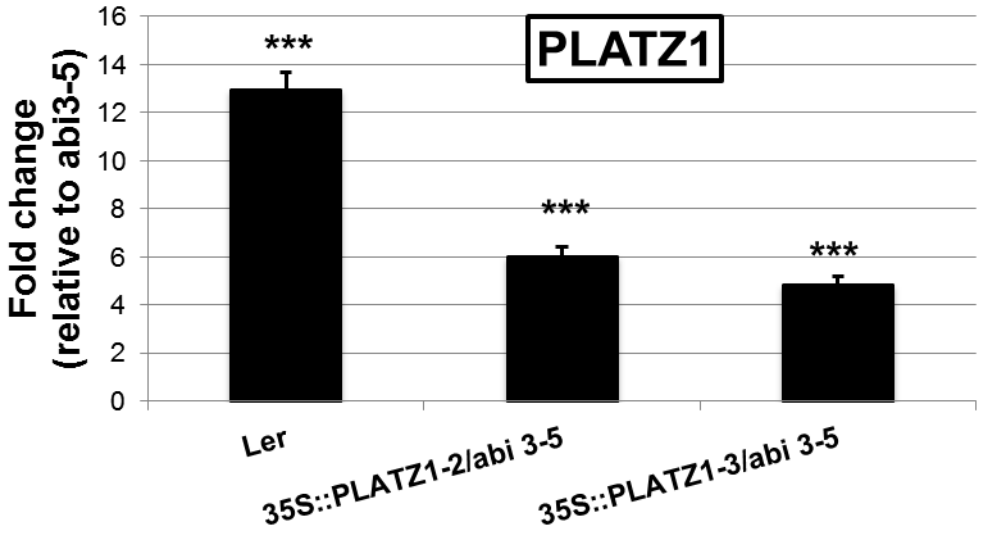

Figure 37. Real Time PCRs of PLATZ1 in 35S::PLATZ1/abi3-5 lines seeds at 21DAF. *abi3-5* was used as the reference sample. UBQ10 and TIP4L served as internal controls. Data are the means ± SD of two biological replicates and three technical replicates. Values are means and ±SD of two biological replicates .Bars with asterisks are significantly different from the control (Student's t-test, * = P<0.05, ** = P < 0.01, *** P<0.0001).

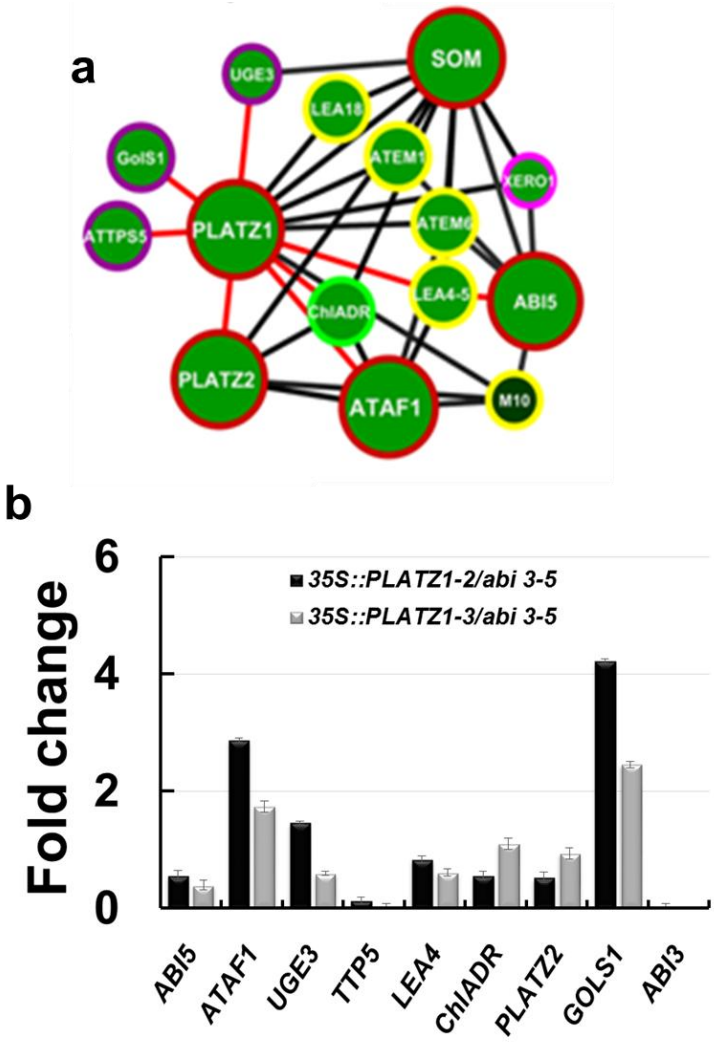

Figure 38. PLATZ1 interactor genes (FullSeed-sNetDT2) related to cellular protection at DPI 0.1. (a) Subnetwork of PLATZ1 interactor genes (b) Real Time PCRs of *ABI5, ATAF1, UGE3, TTP3, LEA4, ChIADR, PLATZ2, GOLS1* (neighboring genes of

PLATZ1 from panel (f) represented with red edges) and *ABI3* in 35S::PLATZ1/*abi3-5* lines seeds at 21DAF; *abi3-5* was used as the reference sample. *UBQ10* and *TIP4L* served as internal controls. Data are the means ± SD of two biological replicates and three technical replicates. Values are means and ±SD of three biological replicates.

5.9. Constitutive expression of *PLATZ1* confers tolerance to low water availability in vegetative tissues.

Our results suggest that PLATZ1 is capable of activating a subset of TFsSeed-subNetDT2 genes that appears to be important for the acquisition of DT in *Arabidopsis* seeds. To test whether *PLATZ1* indeed plays a relevant role in DT, we introduced the *35S::PLATZ1* gene construct into the WT *Arabidopsis* Col-0 ecotype and tested whether its constitutive overexpression had any effect on the phenotype of vegetative tissue subjected to water stress. With this aim, seedlings from four independent *35S::PLATZ1* transgenic lines were grown under full irrigation in the greenhouse, then subjected to a period of seven days without irrigation, and then the number of surviving plants four days after a recovery irrigation treatment was scored. The four tested lines showed a 70 to 80% plant survival rate compared to 10% survival recorded for WT Col-0 (Figure 39).

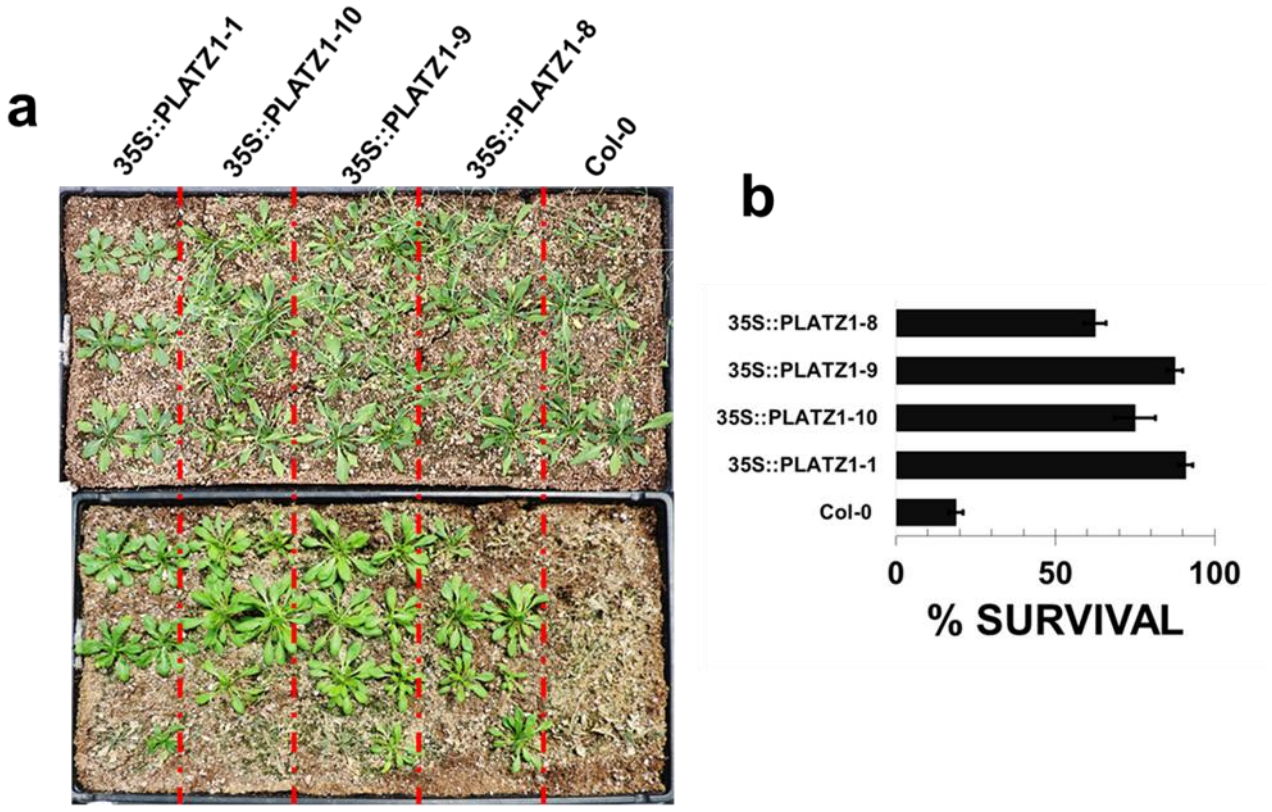

Figure 39. Constitutive expression of PLATZ1 confers tolerance to low water availability in vegetative tissues in greenhouse condition. 35S::PLATZ1/Col-0 overexpressing lines display increased to low water availability tolerance in greenhouse condition. Images of Col-0, 35S::PLATZ1 lines throughout the low water availability assay. (a) Upper row shows plants subjected to a low water availability stress in which water was withheld for 15 days and the lower row shows the plants five days after a recovery irrigation (flowering stems were cut to show the rosette leaves) (b) Percentage of survival plants subjected to low water availability for 15 days and 5 days after recovery irrigation. Values are means and ±SD of three biological replicates (n ≥ 18).

Similar results were obtained in experiments carried out in growth chambers in which temperature and light conditions were controlled (Figure 40). These results show that transgenic plants ectopically expressing PLATZ1 have an enhanced tolerance to low water availability when compared to the WT. Although additional experiments are required to determine whether enhance tolerance to low water

availability is due to the activation of protective mechanisms, a reduced cell metabolism or a reduced growth rate, the finding that plants expressing PLATZ1 under the 35S promoter are capable of better surviving low water availability in the soil suggests that the DT regulatory sub-networks are relevant to seed DT, as activation of any of these mechanisms or their combinations could play important role in allowing cell in the *Arabidopsis* embryo to enter into a low metabolic dormant state in which the cell are protected against water stress and become more tolerant to desiccation.

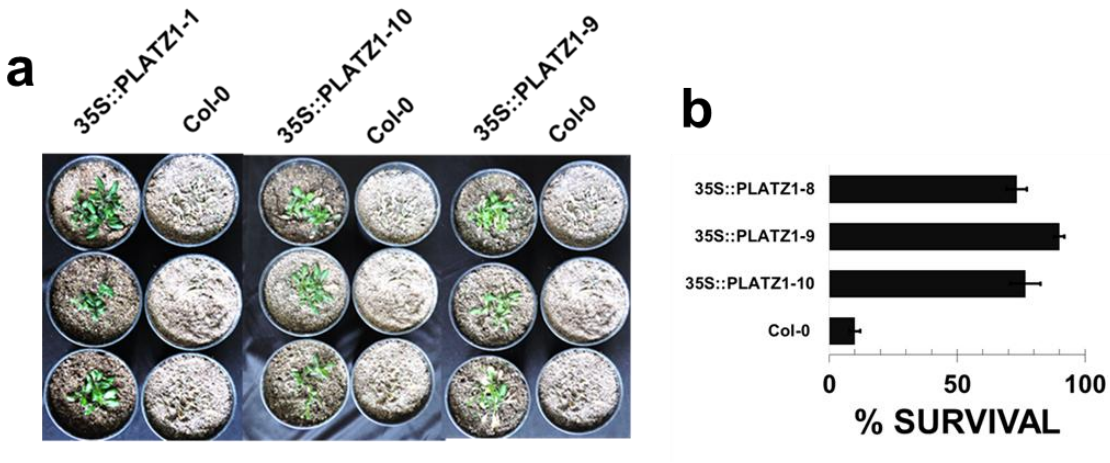

Figure 40. Constitutive expression of PLATZ1 confers tolerance to low water availability in vegetative tissues in controlled condition. 35S::PLATZ1/Col-0 overexpressing lines display increased to low water availability tolerance growth in chamber at 22 °C with a 16-h photoperiod at 200 mmol/m^2s. (a) Independent flowerpots with *Arabidopsis* plants were subjected to a low water availability stress in which water was withheld for 15 days, images of Col-0, 35S::PLATZ1 lines represent five days after a recovery irrigation (flowering stems were cut to show the rosette leaves). (b) Percentage of survival plants subjected to low water availability for 15 days and 5 days after recovery irrigation. Values are means and ±SD of three biological replicates (n ≥ 18).

6. Discussion

DT allowed seed plants to conquer ecosystems with long periods of limited water availability by providing a state in which the embryo is maintained viable until favorable conditions for germination are encountered. This adaptive feature allows seeds to remain dried for hundreds or even thousands of years without losing their ability to germinate (Oliver et al. 2000; Oliver et al. 2005; Illing et al. 2005). Significant advances have been achieved in our understanding of processes involved in seed maturation and their regulatory mechanisms in *Arabidopsis* (Fait et al. 2006) and *Medicago* (Verdier et al. 2013; Righetti et al. 2015). However, the molecular networks that regulate the DT process remain largely unknown. Our results provide novel clues as to the molecular networks acting downstream of the master genes *LEC1*, *ABI3*, *FUS3* and *LEC2* genes that modulate seed development including DT during seed maturation.

ABI3, FUS3, LEC1 and LEC2 have been shown to control most aspects of seed maturation and to repress the transition from embryo to vegetative development (Nambara et al. 1992; Keith et al. 1994; Meinke et al. 1994). Although some aspects of their function are specific, i.e. chlorophyll breakdown controlled by ABI3 (Delmas et al. 2013) and production of somatic embryos controlled by LEC1 and LEC2 (Lotan et al. 1998; Stone et al. 2001), their role on seed maturation is globally similar. A genetic analysis of single and double mutants suggests that these genes integrate an important regulatory network of genes with overlapping functions, which could be explained by synergism or redundancy (To et al. 2006). Interestingly, ectopic expression of LEC2, FUS3 or ABI3 in the single or double

mutant backgrounds of the other two regulators was able to initiate seed lipid and storage protein accumulation but not DT, suggesting that all three regulators are required to activate the expression of TFs directly involved in DT(Roscoe et al. 2015). Our comparative transcriptomic analysis confirms that ABI3, FUS3, LEC1 and LEC2 repress the transition from embryo to vegetative development, because upregulated genes in the corresponding null mutants were mainly associated with high metabolic activity including genes involved in photosynthesis and anthocyanin synthesis that are probably responsible for the green and purple seeds phenotype observed in these mutants ,as well as genes involved in leaf development and trichrome formation. These results are in agreement with the notion that ABI3, FUS3, LEC1 and LEC2 are key genes determining seed identity and that mutation in any of these genes transit directly from embryo to vegetative development.

Our comparative transcriptional global analysis from the interaction model analysis allowed us identify DT-specific differentially expressed genes independent of other genes which altered expression contribute to the pleiotropic phenotype of *lec1-1*, *lec2-1*, *fus3-3* and *abi3-5*.In order to remove the background related with the development transition, we used the comparison of *lec1-1, fus3-3, abi3-5,* which have similar phenotypes and *lec2-1* and *abi3-1* to identify specific genes of DT. Therefore, in our model the downregulated genes represent specific genes that fail to be activated in desiccation intolerant mutants, which should include TF genes that coordinate the establishment of DT and effector genes that are directly responsible for the different components that result in DT.

One of the most intriguing questions about how seeds in the desiccated state can remain viable for periods of time that can exceed centuries, particularly how is the integrity of DNA preserved to prevent permanent damage making the seed unviable. Three types of DNA damage under physiological conditions have been reported: hydrolysis of the N-glycosyl bond, hydrolytic deamination of cytosine to form uracil and DNA damage by oxidation. The first two types of DNA damage are catalyzed by water, therefore removal from the last layers of water that interact with DNA need to be removed or decreased to prevent damage, the third is mediated mainly by ROS (Kranner and Birtic 2005; Rajjou and Debeaujon 2008). In *Arabidopsis,* some oligosaccharides are accumulated very late during seed development and they have been proposed as possible key factor in the acquisition of DT, acting as protective agents via a 'water replacement' mechanism(Baud et al. 2002). This hypothesis suggests that hydroxyl groups of these sugars substitute for water and provide the required hydrophilic interaction for membrane and protein stabilization, as well as DNA protection against hydrolytic damage. In agreement with the notion that oligosaccharides play an important role in preserving protein, membranes and DNA from damage during seed desiccation and the dry seed stage, we found that among the DT-specific gene categories unveiled by our transcriptome analysis one of the most affected was oligosaccharide synthesis (mainly raffinose, stachyose and trehalose synthesis). In transcriptome analysis, we also found a significant number of genes related the control of cellular damage by free radicals. These detoxification genes protective enzymes against ROS damage such as SOD, CAT, ascorbate peroxidase AP, glutathione peroxidases (GSHPx), and glutathione reductase

(GSSGR) could play an important role in the context of dry quiescent seeds in which reactive oxygen species are generated. Therefore, accumulation of antioxidant components during the late maturation stage contributes to control their storage potential, and helps to prevent damage from accumulated ROS during seed maturation. Accumulation of oligosaccharides and activation of mechanisms that prevent damage by ROS allow maximum metabolism reduction to decrease the production of toxic compounds and to prevent membrane, DNA, RNA and protein damages.

Some studies have proposed that the accumulation of oligosaccharides is not sufficient for the acquisition of DT in *Arabidopsis* seed (Ooms et al. 1993; Bentsink et al. 2000). However, the finding that the level of oligosaccharides was found to be reduced in desiccation intolerant mutants, but not in *lec2* and the WT, suggest a clear correlation between DT and raffinose accumulation. Moreover, the reduction in germination displayed by *gols1* and *gols2* T-DNA insertion mutants provides direct evidence supporting the role of oligosaccharides in the acquisition of DT.

Our comparative transcriptomic analysis of DT and desiccation intolerant mutants identified biological processes important for DT acquisition in seeds, and coincides with findings from previous reports (Angelovici et al. 2010; Verdier et al. 2013). However, there is no information available on TFs that control the regulatory networks involved in the acquisition of seed DT.

In this work, we identified TFs genes that are major nodes in the subnetworks related to DT. The importance of these TFs in the acquisition of seed DT was confirmed by the observation that seed viability was reduced in T-DNA insertion

mutants of some of these TFs, such as PLATZ1, PLATZ2, AGL67 and ATAF1. PLATZ1 and PLATZ2 from FullSeed-subNetDT2 which are TFs with unknown function(Nagano et al. 2001), were shown in our study suggest to be essential regulators of the DT process, because 1) PLATZ1 and PLATZ2 are central TFs with high connection to genes highly related to protection of cellular mechanism (LEA proteins, antioxidants and oligosaccharides) activated only at 17 and 21 DAF, 2) the *platz1-1* and *platz2-1* mutants showed a low germination when the seed is dried, 3) the ectopic expression of PLATZ1 in *abi3-5* mutants was able to at least partially recover the germination even up to 30 days of storage and 4) the subnetwork of PLATZ1 was confirmed by the observation that the expression of several putative targets of PLATZ1 were indeed upregulated in *PLATZ1* overexpression lines.

Nevertheless, AGL67 and DREB2D from FullSeed-subNetDT1 also appear to have importance in DT because their ability to complement in *abi3-5* and to recovery the germinability suggesting their importance in the switch activating the DT genes mentioned above. AGL67 belongs to MADS-box TF family that regulate several kinds of transitions (Bassel et al. 2011). In this work, we hypothesized that this TF is important component in the transition from nutrient reservoir activity to protection cellular process because it was main expressed at 17 and 21DAF and showed a switch between the two processes. The absence of this TF showed low germination and, its ectopic expression to restores the germination in the intolerant mutant *abi3-5* (Figure 35). The DREB2D mutant did not show any DT phenotype, probably due to the redundancy that exists in the family. However, when this TF

was overexpressed in the *abi3-5* mutant background, only a slight recovery of the desiccation intolerance by short time desiccation was observed. This means that DREB2D is not sufficient to drive DT but could be a part of the components involved in the regulation of DT (Figure 36). These data support the notion that TFsSeed-subNetDT1 and TFsSeed-subNetDT2 downstream of LEC1, FUS3 and ABI3 are essential for the acquisition of DT in seeds.

It remains to determined whether the regulatory sub-networks involved, at least in part, in DT in *Arabidopsis* seeds are or not conserved in other angiosperm species. In this regard, it is quite suggestive that several of the nodes in TFsSeed-subNetDT1 and TFsSeed-subNetDT2 have seed-specific expression in different plants species. For example, *PLATZ1* orthologous are specifically expressed during seed maturation in rice, soybean and maize. Interestingly, the maize *PLATZ1* ortholog, in addition to being expressed during seed maturation, it is also strongly induced by drought stress (Figure 41).

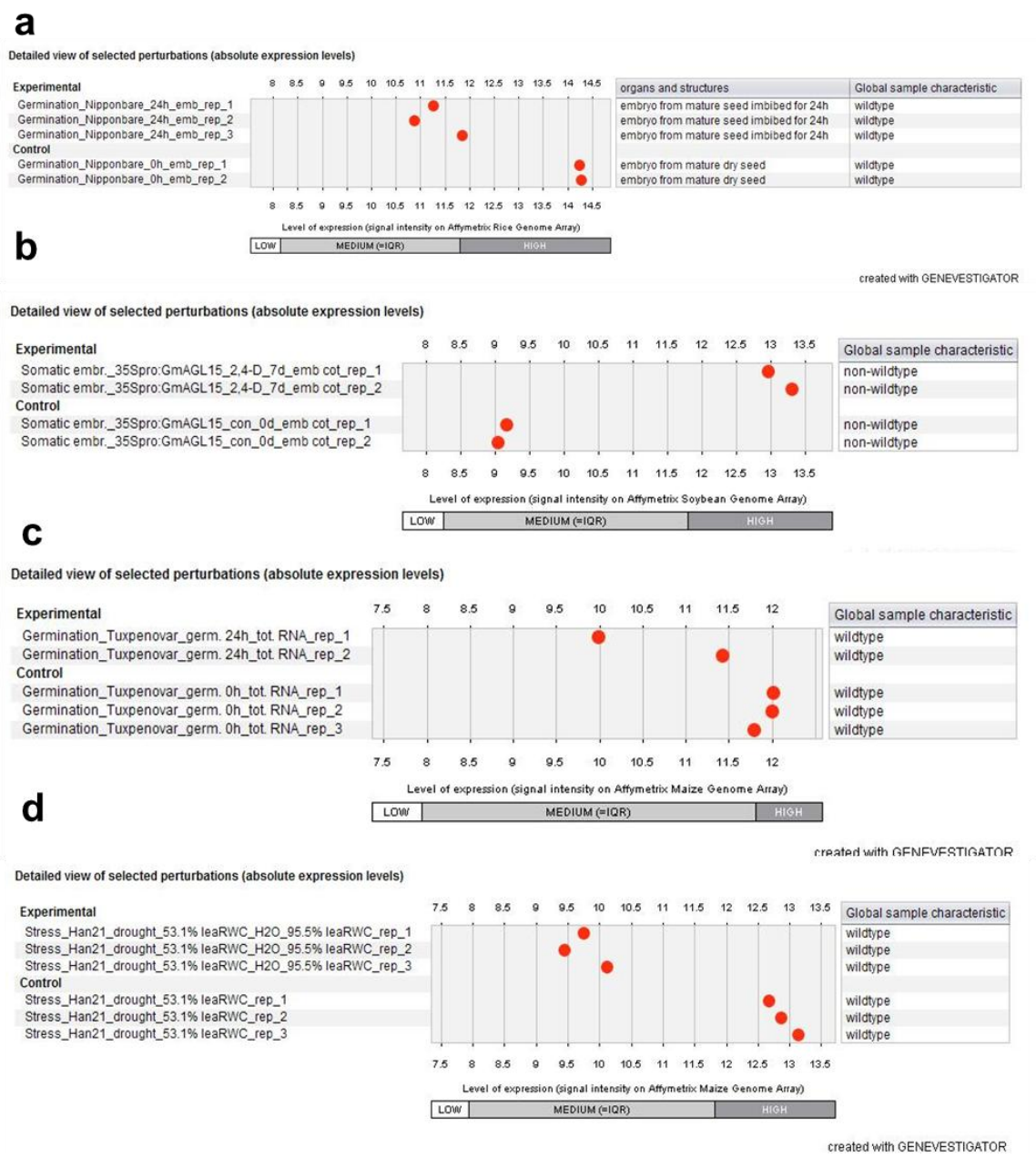

Figure 41. Expression pattern of possible orthologues of PLATZ1. Expression analysis was determined using the Genevestigator database**.** (a) LOC Os06g41930.5 (*Oryza sativa japonica*) (b) Glyma.01G234000.2.Wm82.a2.v1 (*Glycine Max*), (c) and (d) GRMZM2G004548 T01.v6a (*Zea mays).*

A number of previously published reports further support our results on the prediction of regulatory subnetworks involved in seed DT: 1) Some of the TFs that

are major nodes in these subnetworks seem to be directly activated by LEC1; for example, it has been shown that *PEI1* and *ERF23* are activated by ectopic expression of *LEC1* (Mu et al. 2008) 2) Over-expression in *Arabidopsis* of *ATAF1* (one of the TFsSeed-subNetDT2 nodes with a high number of interactions with effector genes) was previously reported to enhance drought tolerance in *Arabidopsis* (Wu et al. 2009); 3) SOM is a TF belonging to the CCCH-type zinc finger family that has been reported to negatively regulate seed germination by activating ABA biosynthesis and inhibiting GA biosynthesis (Kim et al. 2008); 4) Although the precise functions of DREB2G and DREB2D are still unknown, they belong to a TF family that is generally involved in abiotic stress tolerance (Hu and Xiong 2014). 5) AGL67 is a gene associated with the transition to the dormancy process(Bassel et al. 2011). In this study, we suggest its possible role in the transition between seed filling and DT.

Since LEC1 and ABI3 are highly conserved from bryophytes to angiosperms(Cagliari et al. 2014), an important question is whether vegetative DT was originally controlled by these TFs, which were then recruited to control seed development or whether the networks responsible for DT were controlled by other master regulators responsible for the activation of DT genes in an inducible manner in vegetative tissues. The study of basal plants that have vegetative DT will allow to answer this question that has significant importance to understand the evolution of land plants. Moreover, as vegetative DT is an ancestral feature of plants and has re-evolved at least eight times in angiosperms, suggesting that a conserved set of regulatory genes has been recruited several times to activate

vegetative desiccation during plant evolution. Thus, another interesting question is whether the regulatory subnetworks that control vegetative DT in basal plants are similar to those controlling DT in seeds. We propose that recruitment of the DT regulators, that act downstream of LEC1 and ABI3, for expression in vegetative tissues might be responsible for the independent re-evolution of vegetative DT during plant evolution. As a first attempt to answer this question, we performed a phylogenetic analysis of the conservation of the key nodes of TFsSeed-subNetDT1 and TFsSeed-subNetDT2 networks. We found that PLATZ1, PLATZ2 (Figure 42), AGL67 (Figure 43), DREB2D and DREB2G (Figure 44) are conserved in the bryophytes *Physcomitrella patents*, in vascular DT basal plants such as *Selaginella moellendorffii*, in the basal angiosperm *Amborella trichopoda*, in the monocotyledonous species *Oryza sativa* (rice)*, Zea mays* (corn) and DT plant *Oropetium thomaeum,* and in the dicotyledonous species *Glycine max* (soybean) and *Solanum lycopersicum* (tomato).

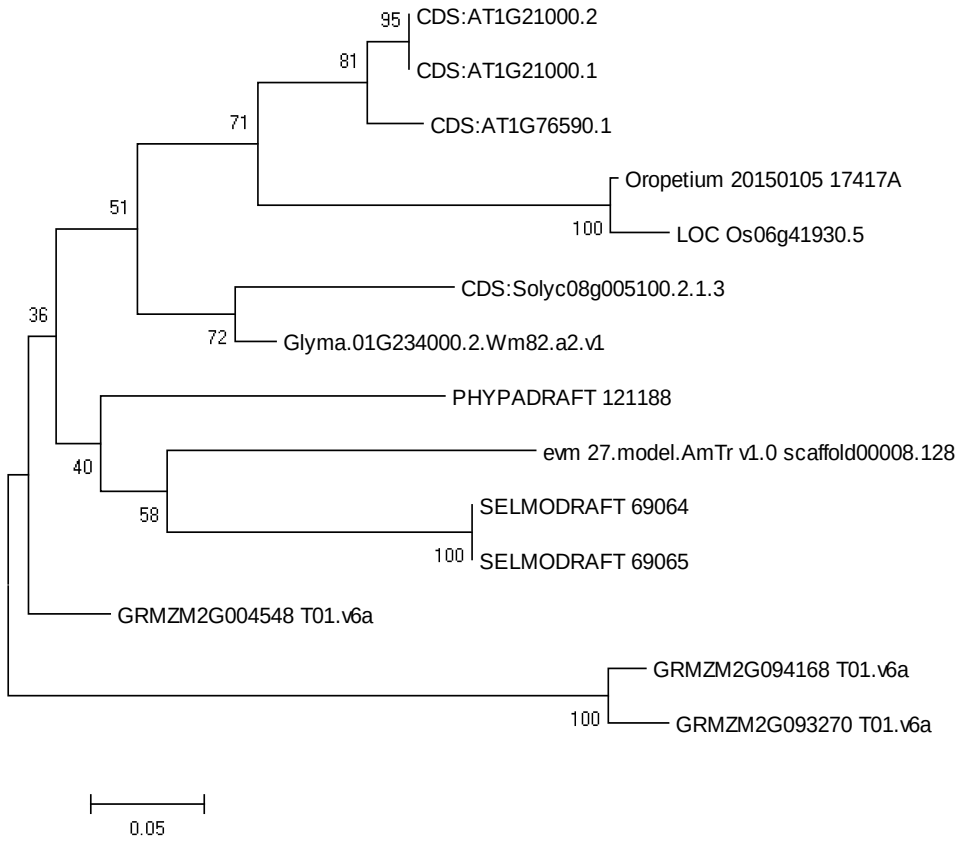

Figure 42. Phylogenetic analysis of PLATZ1 and PLATZ2.The NJ tree was constructed from the amino acid sequences of the PLATZ1 and PLATZ2 domain using the MEGA 6 program based on the JTT model. A consensus tree (after 1000 bootstrap samplings) is shown, and support values are indicated on the sides of important nodes. The BLAST was performed with CoGe Blast tool (https://genomevolution.org/coge/CoGeBlast.pl). PHYPADRAFT 121188 (*Physcomitella pattens* genome v.1 from NCBI database), SELMODRAFT 69064 and SELMODRAFT 69065 (*Selaginella mollendorffi* genome v.1 from NCBI database), evm 27.model.AmTr v1.0 scaffold00008.128 (*Amborella trichopoda* genome v.2 from Amborella Genome Consortium), LOC Os06g41930.5 (*Oryza sativa japonica* from MSU: Rice Genome Annotation v.7), Oropetium 20150105 17417A (*Oropetium thomaeum* genome v.1 from PacBio), Solyc08g005100.2.1.3 (*Solanum Lycopersicum* genome v.2 from SGN), Glyma.01G234000.2.Wm82.a2.v1 (*Glycine Max* genome v.1 from JGI), AT1G21000 and AT1G76590 (Arabidopsis thaliana genome TAIR10 from Ensembl Plants) GRMZM2G004548 T01.v6a, GRMZM2G094168 T01.v6a and GRMZM2G093270 T01.v6a (*Zea mays* genome v.6 from Ensembl Plants).

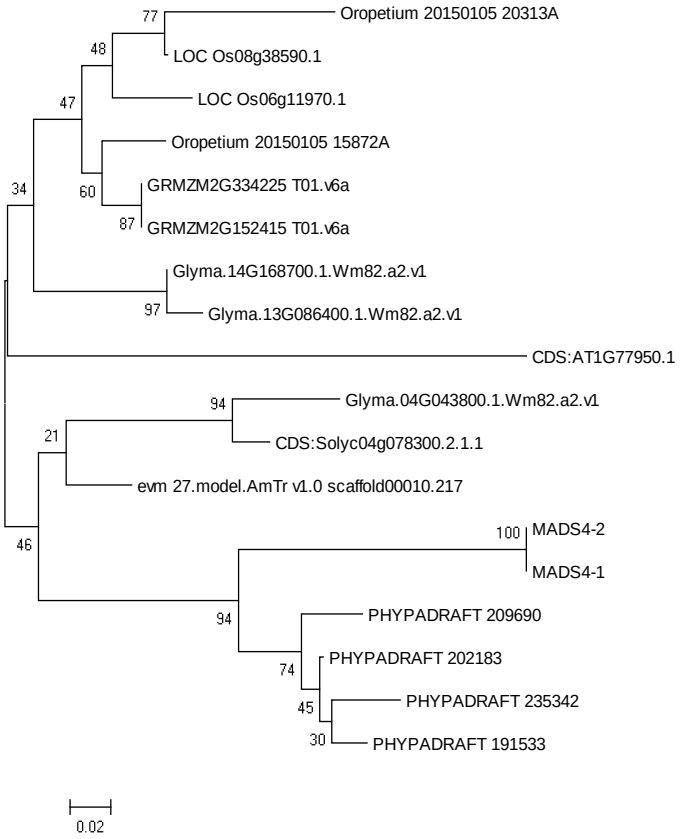

Figure 43. Phylogenetic of AGL67. The NJ tree was constructed from the amino acid sequences of the AGL67 domain using the MEGA6 program based on the JTT model. A consensus tree (after 1000 bootstrap samplings) is shown, and support values are indicated on the sides of important nodes. The BLAST was performed with CoGe Blast tool (https://genomevolution.org/coge/CoGeBlast.pl). PHYPADRAFT 202183, PHYPADRAFT 235342, PHYPADRAFT 191533 and PHYPADRAFT 209690 (*Physcomitella pattens* genome v.1 from NCBI database), MADS4-2 and MADS4-1 (*Selaginella mollendorffi* genome v.1 from NCBI database), evm 27.model.AmTr v1.0 scaffold00010.217 (*Amborella trichopoda* genome v.2 from Amborella Genome Consortium), LOC Os08g38590.1 and LOC Os06g11970.1 (*Oryza sativa japonica* from MSU: Rice Genome Annotation v.7), Oropetium 20150105 20313A and Oropetium 20150105 15872Ac (*Oropetium thomaeum* genome v.1 from PacBio), Solyc04g078300.2.1.1 (*Solanum Lycopersicum* genome v.2 from SGN), Glyma.04G043800.1.Wm82.a2.v1, Glyma.14G168700.1.Wm82.a2.v1 and Glyma.13G086400.1.Wm82.a2.v1 (*Glycine Max* genome v.1 from JGI), AT1G77950 (*Arabidopsis thaliana* genome TAIR10 from Ensembl Plants) GRMZM2G334225 T01.v6a and GRMZM2G152415 T01.v6a (*Zea mays* genome v.6 from Ensembl Plants).

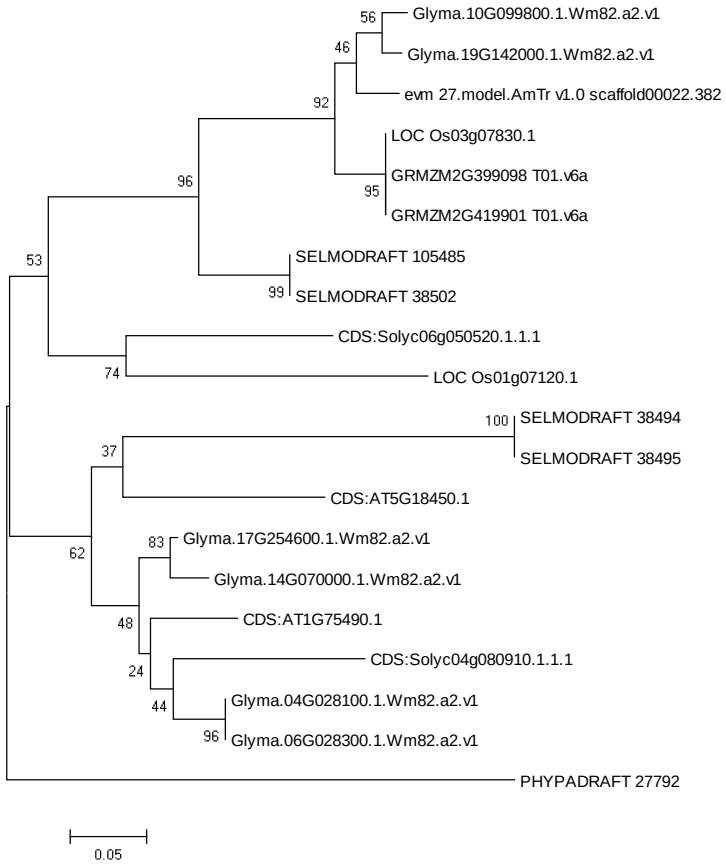

Figure 44. Phylogeny of DREB2D and DREB2G. The NJ tree was constructed from the amino acid sequences of the DREB2D and DREB2G domain using the MEGA6 program based on the JTT model. A consensus tree (after 1000 bootstrap samplings) is shown, and support values are indicated on the sides of important nodes. The BLAST was performed with CoGe Blast tool (https://genomevolution.org/coge/CoGeBlast.pl), PHYPADRAFT 27792 (*Physcomitella pattens* genome v.1 from NCBI database), SELMODRAFT 69064 and SELMODRAFT 105485, SELMODRAFT 38502, SELMODRAFT 38494 and SELMODRAFT 38495 (*Selaginella mollendorffi* genome v.1 from NCBI database), evm 27.model.AmTr v1.0 scaffold00022.382 (*Amborella trichopoda* genome v.2 from Amborella Genome Consortium), LOC Os03g07830.1 (*Oryza sativa japonica* from MSU: Rice Genome Annotation v.7) Solyc06g050520.1.1.1 and Solyc04g080910.1.1.1 (*Solanum Lycopersicum* genome v.2 from SGN), Glyma.10G099800.1.Wm82.a2.v1, Glyma.19G142000.1.Wm82.a2.v1, Glyma.17G254600.1.Wm82.a2.v1, Glyma.14G070000.1.Wm82.a2.v1, Glyma.04G028100.1.Wm82.a2.v1 and Glyma.06G028300.1.Wm82.a2.v1 (*Glycine Max genome* v.1 from JGI), AT5G18450.1 and AT1G75490.1

(*Arabidopsis thaliana* genome TAIR10 from Ensembl Plants) GRMZM2G399098 T01.v6a and GRMZM2G419901 T01.v6a (*Zea mays* genome v.6 from Ensembl Plants).

This opens the possibility that a core DT regulatory network could indeed have been conserved throughout plant evolution. Further research will be needed to explore whether DT is orchestrated by regulatory networks in which at least a common core of TFs has been conserved during plant evolution and to determine how it has been rewired several times to be activated in seeds and in vegetative tissues.

Finally, we propose a hypothetical model of DT seed regulation during seed maturation: TFs identified are downstream of the master regulators LEC1, ABI3 and FUS3 (Figure 45A). The regulatory subnetworks related to storage of reserve compounds are active at 15, 17 and 21 DAF while that subnetwork related to cell protection mechanism is mainly activated at 17 and 21 DAF. At 15 DAF, PEI, DOGL4 and ERF23 activate the expression of genes related to storage of reserve compounds such as fatty acid, TAG and SPP biosynthesis which have cognate sites for these TFs are present in the promoter of maturation-specific genes (RY-elements) (Figure 45A). At 17 and 21DAF the expression of genes related to storage of reserve compounds is maintained, and a the expression of AGL67, DREB2D and DOGL4 is activated which in turn regulate the expression of effector genes involved in cell protection mechanisms responsible for the acquisition of DT . The cognate binding sites (ABREs, RYs and DREB elements) for these second set of TFs is present in dehydration responsive genes. PLATZ1 together with PLATZ2 and ATAF1 activate genes related to stress response and cell protection

mechanisms, including genes belonging to LEA protein families, genes involved in the production of antioxidants and protective oligosaccharides (Figure 45A and 45B). PLATZ1 is a novel regulator of DT, which is transcriptionally activated by ABI3 and/or ABI5 in a network conserved during the evolution of land plants (Figure 45C).

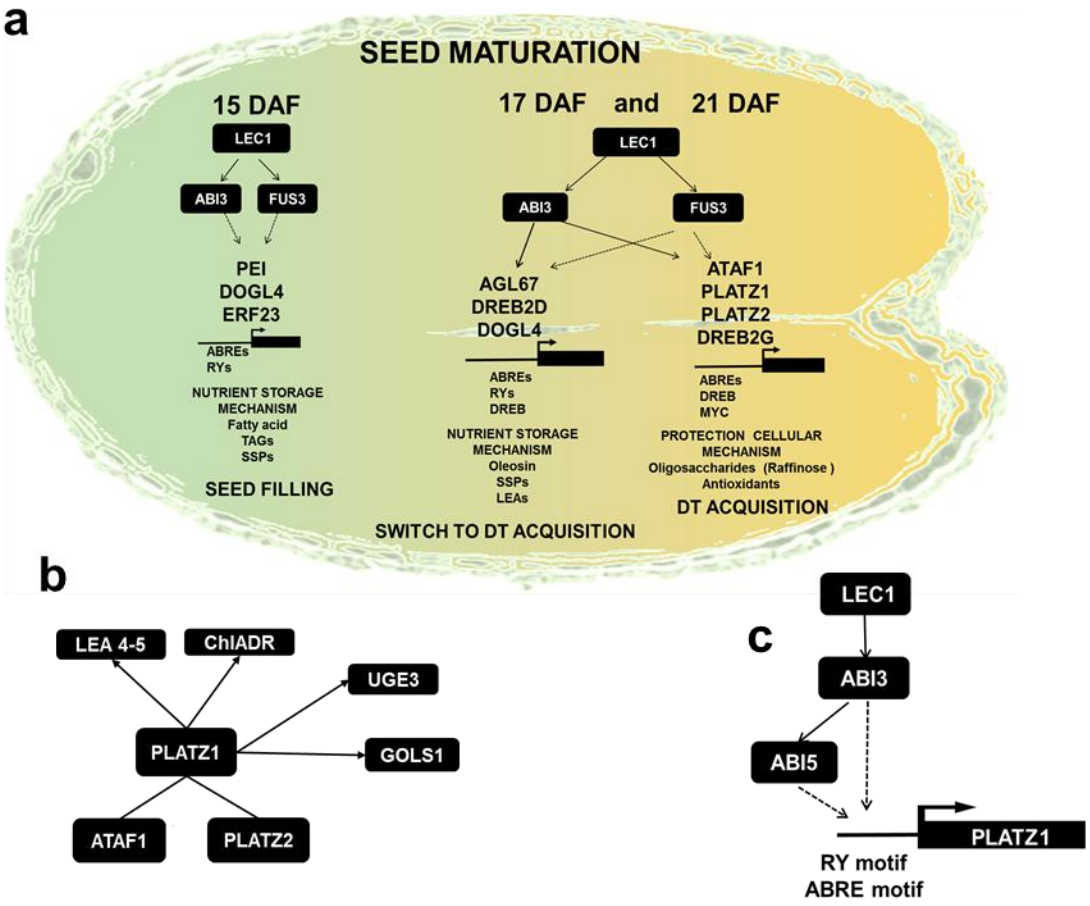

Figure 45. Hypothetical model of desiccation tolerance seed regulation during seed maturation. (a) TFs downstream from LEC1, FUS3 and ABI3 (b) Targets of PLATZ1 (c) Regulation network conserved in plants

7. Conclusions

- Our work confirms that constructing co-expression networks and TF regulatory networks based on transcriptome data allows the identification of TFs that are relevant for specific biological process.
- Application of this strategy for the analysis of differentially regulated genes between desiccation tolerant and intolerant Arabidopsis lines allowed the identification of novel TFs that regulate different regulatory nodes important for the phenotype of seed desiccation tolerance.
- The relevance of the identified TFs was confirmed because knock-out mutants in these genes had reduce seed desiccation tolerance.
- PLATZ1, AGL67 and DREB2D are sufficient to activate nodes that contribute to desiccation tolerance because its constitutive expression partially rescues the desiccation intolerance phenotype of *abi3-5*.
- PLATZ1 is involved in desiccation tolerance and has the potential to increase drought tolerance in Arabidopsis vegetative tissues.

8. Perspectives

- Complete metabolome analysis in order to identify novel metabolites involved in seed DT and correlate with regulatory networks.
- Detailed analysis of drought responses in *35S::PLATZ1* transgenic *Arabidopsis* plants.
- Drought analysis of other TFs identified as key in DT seed through constitutive and inducible expression.
- Reconstruct the subnetworks identified in vegetative tissues of model plants through co-expression of TFs identified as key in DT seed.
- Functional analysis of orthologous TFs in resurrection plants in vegetative tissue in order to evaluate the conservation in function of subnetworks proposed.
- Functional analysis of orthologous TFs in recalcitrant seeds in order to design new conservation methods.

9. References

Al-Whaibi MH (2011) Plant heat-shock proteins: A mini review. Journal of King Saud University - Science 23 (2):139-150. doi:http://dx.doi.org/10.1016/j.jksus.2010.06.022

Alpert P, Oliver, M.J. (2002) Drying without dying. In: M. Black HP (ed) Desiccation and Survival in Plants: Drying without dying. CABI Publishing,

Amara I, Zaidi I, Masmoudi K, Ludevid M, Pagès M, Goday A, Brini F (2014) Insights into Late Embryogenesis Abundant (LEA) Proteins in Plants: From Structure to the Functions. American Journal of Plant Sciences 5:3440-3455. doi:10.4236/ajps.2014.522360

Angelovici R, Galili G, Fernie AR, Fait A (2010) Seed desiccation: a bridge between maturation and germination. Trends in plant science 15 (4):211-218. doi:10.1016/j.tplants.2010.01.003

Bailly C (2004) Active oxygen species and antioxidants in seed biology. Seed Science Research 14 (2):93-107. doi:10.1079/ssr2004159

Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provart NJ (2011) Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proceedings of the National Academy of Sciences of the United States of America 108 (23):9709-9714. doi:10.1073/pnas.1100958108

Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human B cells. Nature genetics 37 (4):382-390. doi:10.1038/ng1532

Battaglia M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias AA (2008) The Enigmatic LEA Proteins and Other Hydrophilins. Plant physiology 148 (1):6-24. doi:10.1104/pp.108.120725

Baud S, Boutin J, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in Arabidopsis thaliana ecotype WS. Plant Physiology and Biochemistry 40 (2):151-160. doi:http://dx.doi.org/10.1016/S0981-9428(01)01350-X

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57 (1):289-300

Bentsink L, Alonso-Blanco C, Vreugdenhil D, Tesnier K, Groot SPC, Koornneef M (2000) Genetic Analysis of Seed-Soluble Oligosaccharides in Relation to Seed Storability of Arabidopsis. Plant physiology 124 (4):1595-1604. doi:10.1104/pp.124.4.1595

Buitink J, Hemminga MA, Hoekstra FA (2000) Is There a Role for Oligosaccharides in Seed Longevity? An Assessment of Intracellular Glass Stability. Plant physiology 122 (4):1217-1224. doi:10.1104/pp.122.4.1217

Cagliari A, Turchetto-Zolet AC, Korbes AP, Maraschin Fdos S, Margis R, Margis-Pinheiro M (2014) New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. Genomics 103 (5-6):380-387. doi:10.1016/j.ygeno.2014.03.005

Crowe JH, Crowe LM, Carpenter JF, Aurell Wistrom C (1987) Stabilization of dry phospholipid bilayers and proteins by sugars. Biochemical Journal 242 (1):1-10

Crowe LM (2002) Lessons from nature: the role of sugars in anhydrobiosis. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 131 (3):505-513. doi:http://dx.doi.org/10.1016/S1095-6433(01)00503-7

Cuevas-Velazquez CL, Rendón-Luna DF, Covarrubias AA (2014) Dissecting the cryoprotection mechanisms for dehydrins. Frontiers in plant science 5:583. doi:10.3389/fpls.2014.00583

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic Acid: Emergence of a Core Signaling Network. Annual review of plant biology 61 (1):651-679. doi:10.1146/annurev-arplant-042809-112122

Chavez Montes RA, Coello G, Gonzalez-Aguilera KL, Marsch-Martinez N, de Folter S, Alvarez-Buylla ER (2014) ARACNe-based inference, using curated microarray data, of Arabidopsis thaliana root transcriptional regulatory networks. BMC plant biology 14:97. doi:10.1186/1471-2229-14-97

Chen M, Wang Z, Zhu Y, Li Z, Hussain N, Xuan L, Guo W, Zhang G, Jiang L (2012) The effect of transparent TESTA2 on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in Arabidopsis. Plant physiology 160 (2):1023-1036. doi:10.1104/pp.112.202945

Debeaujon I, Léon-Kloosterziel KM, Koornneef M (2000) Influence of the Testa on Seed Dormancy, Germination, and Longevity in Arabidopsis. Plant physiology 122 (2):403-414. doi:10.1104/pp.122.2.403

Delahaie J, Hundertmark M, Bove J, Leprince O, Rogniaux H, Buitink J (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. Journal of experimental botany 64 (14):4559-4573. doi:10.1093/jxb/ert274

Delmas F, Sankaranarayanan S, Deb S, Widdup E, Bournonville C, Bollier N, Northey J, McCourt P, Samuel M (2013) ABI3 controls embryo degreening through Mendel's I locus. Proceedings of the National Academy of Sciences 110 (40):E3888-E3894. doi:10.1073/pnas.1308114110

Dure L, Chlan C (1981) Developmental Biochemistry of Cottonseed Embryogenesis and Germination: XII. PURIFICATION AND PROPERTIES OF PRINCIPAL STORAGE PROTEINS. Plant physiology 68 (1):180-186. doi:10.1104/pp.68.1.180

Endo A, Okamoto M, Koshiba T (2014) ABA Biosynthetic and Catabolic Pathways. In: Zhang D-P (ed) Abscisic Acid: Metabolism, Transport and Signaling. Springer Netherlands, Dordrecht, pp 21-45. doi:10.1007/978-94-017-9424-4_2

Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie A, Galili G (2006) Arabidopsis Seed Development and Germination Is Associated with Temporally Distinct Metabolic Switches. Plant physiology 142 (3):839-854. doi:10.1104/pp.106.086694

Finkelstein R (2013) Abscisic Acid Synthesis and Response. The Arabidopsis Book:e0166. doi:10.1199/tab.0166

Forsthoefel NR, Cutler K, Port MD, Yamamoto T, Vernon DM (2005) PIRLs: a novel class of plant intracellular leucine-rich repeat proteins. Plant & cell physiology 46 (6):913-922. doi:10.1093/pcp/pci097

Gaff DF, Oliver M (2013) The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. Functional Plant Biology 40 (4):315-328. doi:http://dx.doi.org/10.1071/FP12321

Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P (2004) The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. Developmental cell 7 (3):373-385. doi:10.1016/j.devcel.2004.06.017

Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. The Plant cell 4 (10):1251-1261. doi:10.1105/tpc.4.10.1251

Holdsworth MJ, Bentsink L, Soppe WJ (2008) Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. The New phytologist 179 (1):33-54. doi:10.1111/j.1469-8137.2008.02437.x

Hu H, Xiong L (2014) Genetic Engineering and Breeding of Drought-Resistant Crops. Annual review of plant biology 65 (1):715-741. doi:10.1146/annurev-arplant-050213-040000

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. Integrative and comparative biology 45 (5):771-787. doi:10.1093/icb/45.5.771

Kalemba E, Pukacka S (2007) Possible roles of LEA proteins and sHSPs in seed protection: a short review. Biol Lett 44 (1):3-16

Keith K, Kraml M, Dengler NG, McCourt P (1994) fusca3: A Heterochronic Mutation Affecting Late Embryo Development in Arabidopsis. The Plant cell 6 (5):589-600. doi:10.1105/tpc.6.5.589

Kermode AR, Finch-Savage, B. E (2002) Desiccation sensitivity in orthodox and recalcitrant seeds in relation to development. In: Black M, Pritchard, H. W. (ed) Desiccation and survival in plants: drying without dying. . p 149. doi:10.1079/9780851995342.0149

Kim DH, Yamaguchi S, Lim S, Oh E, Park J, Hanada A, Kamiya Y, Choi G (2008) SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5. The Plant cell 20 (5):1260-1277. doi:10.1105/tpc.108.058859

Kranner I, Birtic S (2005) A modulating role for antioxidants in desiccation tolerance. Integrative and comparative biology 45 (5):734-740. doi:10.1093/icb/45.5.734

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences of the United States of America 107 (18):8063-8070. doi:10.1073/pnas.1003530107

Leprince O, Buitink J (2015) Introduction to desiccation biology: from old borders to new frontiers. Planta 242 (2):369-378. doi:10.1007/s00425-015-2357-6

Locascio A, Roig-Villanova I, Bernardi J, Varotto S (2014) Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. Frontiers in plant science 5:412. doi:10.3389/fpls.2014.00412

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. The Plant Journal 32 (3):317-328. doi:10.1046/j.1365-313X.2002.01430.x

Lotan T, Ohto M, Yee K, West M-, Lo R, Kwong R, Yamagishi K, Fischer R, Goldberg R, Harada J (1998) Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. Cell 93 (7):1195-1205. doi:http://dx.doi.org/10.1016/S0092-8674(00)81463-4

Luerßen H, Kirik V, Herrmann P, Miséra S (1998) FUSCA3encodes a protein with a conserved VP1/ABI3-like B3 domain which is of functional importance for the regulation of seed maturation inArabidopsis thaliana. The Plant Journal 15 (6):755-764. doi:10.1046/j.1365-313X.1998.00259.x

Lumba S, Tsuchiya Y, Delmas F, Hezky J, Provart NJ, Shi Lu Q, McCourt P, Gazzarrini S (2012) The embryonic leaf identity gene FUSCA3 regulates vegetative phase transitions by negatively modulating ethylene-regulated gene expression in Arabidopsis. BMC biology 10:8. doi:10.1186/1741-7007-10-8

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics 21 (16):3448-3449. doi:10.1093/bioinformatics/bti551

Manfre AJ, LaHatte GA, Climer CR, Marcotte WR, Jr. (2009) Seed dehydration and the establishment of desiccation tolerance during seed maturation is altered in the Arabidopsis thaliana mutant atem6-1. Plant & cell physiology 50 (2):243-253. doi:10.1093/pcp/pcn185

Manfre AJ, Lanni LM, Marcotte WR (2006) The Arabidopsis Group 1 LATE EMBRYOGENESIS ABUNDANT Protein ATEM6 Is Required for Normal Seed Development. Plant physiology 140 (1):140-149. doi:10.1104/pp.105.072967

Mao Z, Sun W (2015) Arabidopsis seed-specific vacuolar aquaporins are involved in maintaining seed longevity under the control of ABSCISIC ACID INSENSITIVE 3. Journal of experimental botany 66 (15):4781-4794. doi:10.1093/jxb/erv244

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC bioinformatics 7 Suppl 1:S7. doi:10.1186/1471-2105-7-S1-S7

Martinez-Trujillo M, Limones-Briones V, Cabrera-Ponce J, Herrera-Estrella L (2004) Improving transformation efficiency ofArabidopsis thaliana by modifying the floral dip method. Plant Mol Biol Rep 22 (1):63-70. doi:10.1007/BF02773350

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic acids research 40 (10):4288-4297. doi:10.1093/nar/gks042

Meinke DW, Franzmann LH, Nickle TC, Yeung EC (1994) Leafy Cotyledon Mutants of Arabidopsis. The Plant cell 6 (8):1049-1064. doi:10.1105/tpc.6.8.1049

Mene-Saffrane L, Jones AD, DellaPenna D (2010) Plastochromanol-8 and tocopherols are essential lipid-soluble antioxidants during seed desiccation and quiescence in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 107 (41):17815-17820. doi:10.1073/pnas.1006971107

Mishler BD, Churchill SP (1985) Transition to a land flora: Phylogenetic relationships of the green algae and bryophytes. Cladistics 1 (4):305-328. doi:10.1111/j.1096-0031.1985.tb00431.x

Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang X, Zuo J (2008) LEAFY COTYLEDON1 Is a Key Regulator of Fatty Acid Biosynthesis in Arabidopsis. Plant physiology 148 (2):1042-1054. doi:10.1104/pp.108.126342

Nagano Y, Furuhashi H, Inaba T, Sasaki Y (2001) A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. Nucleic acids research 29 (20):4097-4105

Nambara E, Naito S, McCourt P (1992) A mutant of Arabidopsis which is defective in seed development and storage protein accumulation is a new abi3 allele. The Plant Journal 2 (4):435-441. doi:10.1111/j.1365-313X.1992.00435.x

Obendorf RL (1997) Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. Seed Science Research 7 (02):63-74

Oliver MJ, Tuba Z, Mishler BD (2000) The evolution of vegetative desiccation tolerance in land plants. Plant Ecology 151:85-100

Oliver MJ, Velten J, Mishler BD (2005) Desiccation tolerance in bryophytes: a reflection of the primitive strategy for plant survival in dehydrating habitats? Integrative and comparative biology 45 (5):788-799. doi:10.1093/icb/45.5.788

Ooms J, Leon-Kloosterziel KM, Bartels D, Koornneef M, Karssen CM (1993) Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants). Plant physiology 102 (4):1185-1191. doi:10.1104/pp.102.4.1185

Prieto-Dapena P, Castano R, Almoguera C, Jordano J (2008) The ectopic overexpression of a seed-specific transcription factor, HaHSFA9, confers tolerance to severe dehydration in vegetative organs. The Plant journal : for cell and molecular biology 54 (6):1004-1014. doi:10.1111/j.1365-313X.2008.03465.x

Rajjou L, Debeaujon I (2008) Seed longevity: survival and maintenance of high germination ability of dry seeds. Comptes rendus biologies 331 (10):796-805. doi:10.1016/j.crvi.2008.07.021

Righetti K, Vu JL, Pelletier S, Vu BL, Glaab E, Lalanne D, Pasha A, Patel RV, Provart NJ, Verdier J, Leprince O, Buitink J (2015) Inference of Longevity-Related Genes from a Robust Coexpression Network of Seed Maturation Identifies Regulators Linking Seed Storability to Biotic Defense-Related Pathways. The Plant cell 27 (10):2692-2708. doi:10.1105/tpc.15.00632

Roscoe TT, Guilleminot J, Bessoule JJ, Berger F, Devic M (2015) Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis. Plant & cell physiology 56 (6):1215-1228. doi:10.1093/pcp/pcv049

Sallon S, Solowey E, Cohen Y, Korchinsky R, Egli M, Woodhatch I, Simchoni O, Kislev M (2008) Germination, genetics, and growth of an ancient date seed. Science 320 (5882):1464. doi:10.1126/science.1153600

Sano N, Rajjou L, North HM, Debeaujon I, Marion-Poll A, Seo M (2016) Staying Alive: Molecular Aspects of Seed Longevity. Plant and Cell Physiology 57 (4):660-674. doi:10.1093/pcp/pcv186

Shen-Miller J, Schopf JW, Harbottle G, Cao RJ, Ouyang S, Zhou KS, Southon JR, Liu GH (2002) Long-living lotus: germination and soil {gamma}-irradiation of centuries-old fruits, and cultivation, growth, and phenotypic abnormalities of offspring. American journal of botany 89 (2):236-247. doi:10.3732/ajb.89.2.236

Shimada TL, Shimada T, Hara-Nishimura I (2010) A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. The Plant Journal 61 (3):519-528. doi:10.1111/j.1365-313X.2009.04060.x

Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proceedings of the National Academy of Sciences 98 (20):11806-11811. doi:10.1073/pnas.201413498

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal 37 (6):914-939. doi:10.1111/j.1365-313X.2004.02016.x

To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. The Plant cell 18 (7):1642-1651. doi:10.1105/tpc.105.039925

Verdier J, Lalanne D, Pelletier S, Torres-Jerez I, Righetti K, Bandyopadhyay K, Leprince O, Chatelain E, Vu BL, Gouzy J, Gamas P, Udvardi MK, Buitink J (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. Plant physiology 163 (2):757-774. doi:10.1104/pp.113.222380

Wehmeyer N, Hernandez LD, Finkelstein RR, Vierling E (1996) Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. Plant physiology 112 (2):747-757

Wehmeyer N, Vierling E (2000) The Expression of Small Heat Shock Proteins in Seeds Responds to Discrete Developmental Signals and Suggests a General Protective Role in Desiccation Tolerance. Plant physiology 122 (4):1099-1108. doi:10.1104/pp.122.4.1099

Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: early seed germination. Journal of experimental botany 62 (10):3289-3309. doi:10.1093/jxb/err030

Wu Y, Deng Z, Lai J, Zhang Y, Yang C, Yin B, Zhao Q, Zhang L, Li Y, Yang C, Xie Q (2009) Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. Cell research 19 (11):1279-1290. doi:10.1038/cr.2009.108

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC bioinformatics 13:134. doi:10.1186/1471-2105-13-134

List of publications

Publication arising from this thesis:

Gonzalez-Morales SI, Chávez Montes RA, Hayano-Kanashiro C, Alejo-Jacuinde G, Rico-Cambron TY, Folter S and Herrera-Estrella L. "**Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana".* Proc Natl Acad Sci U S A**. 2016 Aug 30;113(35):E5232-41. doi: 10.1073/pnas.1610985113. Epub 2016 Aug 22.

Co-director of bachelor thesis of IBQ. Gerardo Alejo Jacuinde**. "Caracterización del factor de transcripción PLATZ involucrado en la tolerancia a la desecación en semillas de *Arabidopsis thaliana*"** Instituto Tecnológico de Celaya. October 2014.

Other publications generated as a PhD student related with transcriptome analysis:

Yong-Villalobos,L. Cervantes-Perez,S.A. Gutierrez-Alanis,D. **Gonzales-Morales,S**. Martinez,O. Herrera-Estrella,L. 2016. Phosphate starvation induces DNA methylation in the vicinity of cis-acting elements known to regulate the expression of phosphate-responsive genes Plant Signaling and Behavior, 11, e1173300.

Yong-Villalobos L, **González-Morales SI**, Wrobel K, Gutiérrez-Alanis D, Cervantes-Peréz SA, Hayano-Kanashiro C, Oropeza-Aburto A, Cruz-Ramírez A, Martínez O, Herrera-Estrella L. Methylome analysis reveals an important role for epigenetic changes in the regulation of the Arabidopsis response to phosphate starvation. *Proc Natl Acad Sci U S A*. 2015 Dec 29;112(52):E7293-302.

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y, Wu Z, Mao L, Wu H, Ling-Hu C, Zhou H, Lin H, **González-Morales S**, Trejo-Saavedra DL, Tian H, Tang X, Zhao M, Huang Z, Zhou A, Yao X, Cui J, Li W, Chen Z, Feng Y, Niu Y, Bi S, Yang X, Li W, Cai H, Luo X, Montes-Hernández S, Leyva-González MA, Xiong Z, He X, Bai L, Tan S, Tang X, Liu D, Liu J, Zhang S, Chen M, Zhang L, Zhang L, Zhang Y, Liao W, Zhang Y, Wang M, Lv X, Wen B, Liu H, Luan H, Zhang Y, Yang S, Wang X, Xu J, Li X, Li S, Wang J, Palloix A, Bosland PW, Li Y, Krogh A, Rivera-Bustamante RF, Herrera-Estrella L, Yin Y, Yu J, Hu K, Zhang Z.Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc. Natl Acad Sci USA*.2014 Apr 8;111(14):5135-40.

Alatorre-Cobos F, Calderón-Vázquez C, Ibarra-Laclette E, Yong-Villalobos L, Pérez-Torres CA, Oropeza-Aburto A, Méndez-Bravo A, **González-Morales SI**, Gutiérrez-Alanís D, Chacón-López A, Peña-Ocaña BA, Herrera-Estrella L. An

improved, low-cost, hydroponic system for growing Arabidopsis and other plant species under aseptic conditions. *BMC Plant Biol*. 2014 Mar 21;14:69

López-Arredondo D, **González-Morales SI**, Bello-Bello E, Alejo-Jacuinde G, Herrera L. Engineering food crops to grow in harsh environments. F1000Res. 2015 Sep 2;4(F1000 Faculty Rev):651

López-Arredondo DL, Leyva-González MA, **González-Morales SI**, López-Bucio J, Herrera-Estrella L. Phosphate nutrition: improving low-phosphate tolerance in crops. Annu Rev Plant Biol. 2014;65:95-123.