SeedMatExplorer: The transcriptome atlas of Arabidopsis seed maturation

Mariana A. S. Artur[1*†], Robert A. Koetsier[2†], Leo A. J. Willems[1], Lars L. Bakermans[1], Annabel D. van Driel[1], Joram A. Dongus[1], Bas J. W. Dekkers[1], Alexandre C. S. S. Marques[1], Asif Ahmed Sami[1], Harm Nijveen[2], Leónie Bentsink[1], Henk Hilhorst[1,3], Renake Nogueira Teixeira[1,4*]

1 Laboratory of Plant Physiology, Wageningen Seed Science Centre, Wageningen University & Research, Wageningen, 6708 PB, The Netherlands.

2 Bioinformatics Group, Wageningen University & Research, Wageningen, 6708 PB, The Netherlands.

3 Current address: Department of Molecular and Cell Biology, University of Cape Town, Cape Town, South Africa.

4 Current address: Hazera Seeds B.V., Schanseind 27, 4921 PM Made.

*Correspondence: Mariana A. S. Artur (mariana.silvaartur@wur.nl); Renake Nogueira Teixeira (renakent@gmail.com)

†Equal contribution

Short title: Arabidopsis seed maturation atlas

One sentence summary: This study provides a comprehensive physiological and transcriptome atlas of *Arabidopsis thaliana* seed maturation, revealing the complex gene expression and regulation of key seed quality traits, including desiccation tolerance and longevity, through identification of candidate regulatory genes and their physiological validation.

Abstract

Seed maturation is a critical developmental phase during which seeds acquire essential traits for nutritional value and desiccation survival. Abscisic acid (ABA) signalling has long been recognized as a key positive regulator of this process, coordinating the expression of genes that drive the acquisition of vital seed quality traits (SQTs) crucial for survival, resilience, and storage. However, the precise molecular regulation of many SQTs remains elusive. To address this gap, we conducted an extensive transcriptome analysis of seed maturation in *Arabidopsis thaliana*, focusing on Col-0 wild-type and ten mutant lines affecting ABA biosynthesis, signaling, and catabolism from 12 days after pollination up to the dry seed stage. Our study employed Weighted Gene Co-Expression Network Analysis (WGCNA) and module-trait correlation to uncover specific gene groups linked to the regulation of multiple SQTs. We identified *Late Embryogenesis Abundant* (*LEA*) genes uniquely associated with desiccation tolerance and three *DREB AND EAR MOTIF PROTEIN* (*DEAR*) transcription factors involved in seed longevity regulation. The value of our dataset lies not only in its depth and detail, but also in its integration into SeedMatExplorer, a web tool specifically designed to advance seed maturation research. SeedMatExplorer serves as a unique resource, enabling researchers to explore, visualize, and generate new hypotheses about gene expression regulation during seed maturation and the intricate control of SQTs. This physiological and transcriptomic atlas offers unprecedented insights into the complex, coordinated gene expression that drives SQT acquisition in Arabidopsis seeds, making it an invaluable asset for the scientific community.

Keywords: seed maturation, seed quality, co-expression network, transcriptomics, desiccation tolerance, longevity.

Introduction

Seeds are a crucial stage of a plant's life cycle, allowing the plant species to survive in space and time, and through severe environmental challenges. In most plant species, seed development can be divided into three main phases: embryogenesis, maturation, and drying. During embryogenesis (also called morphogenesis or histodifferentiation) cell division takes place to form the embryo body plan (Mansfield et al., 1991, Baud et al., 2002). In the maturation phase (also called seed filling), cell division ceases, an increase in cell expansion is observed because of nutrient reserve deposition, and embryonic chlorophyll is gradually degraded (Jalink et al., 1998, Leprince et al., 2017). Finally, in most plant species, a drying phase (or late seed maturation) takes place, which is marked by a decrease in the seed water content reaching a metabolically inactive, quiescent state (Leprince et al., 2017, Angelovici et al., 2010, Kermode, 1990, Kermode and Bewley, 1985, Kermode et al., 1986). The seed maturation phase has gained increasing attention in the past years not only because this is a phase in which the seeds gain nutritional value, but also because this is the moment in which several seed quality traits (SQTs) important for seed survival, resilience, and storage are acquired (Leprince et al., 2017). For example, seed longevity, defined as the capacity of seeds to remain alive during storage, is a critical SQT for *ex situ* conservation of plant genetic resources and for seed persistence in agricultural and ecological contexts (Pritchard, 2020, Zinsmeister et al., 2020b, Leprince et al., 2017). Dormancy and desiccation tolerance (DT) are two other SQTs of great agronomical and ecological importance that have been associated with optimising seedling establishment (Dekkers et al., 2015) and plant evolution and adaptation on land, respectively (Nonogaki et al., 2022).

During seed maturation, the acquisition of the above-mentioned SQTs requires the coordinated activity of multiple genes. The hormone abscisic acid (ABA) has been identified as a crucial positive regulator of these genes (Groot and Karssen, 1992, Finch-Savage and Leubner-Metzger, 2006, Finkelstein et al., 2008, Holdsworth et al., 2008, Yan and Chen, 2017, Sajeev et al., 2024). ABA accumulates during seed maturation and declines during seed drying (Karssen et al., 1983, Nambara and Marion-Poll, 2003, Kanno et al., 2010). With the isolation of mutants deficient in and insensitive to ABA in the early '80s and '90s, *Arabidopsis thaliana* became one of the main models to study the genetic and molecular control of seed maturation (Karssen et al., 1983, Koornneef et al., 1982, Koornneef et al., 1984, Léon-Kloosterziel et al., 1996, Meurs et al., 1992, Nambara et al., 1998). ABA deficiency has been associated with the absence of primary dormancy, and plants with overexpression or loss of ABA biosynthesis genes show increased and decreased seed dormancy, respectively (Karssen et al., 1983, Léon-

Kloosterziel et al., 1996, Nambara et al., 1998, Lefebvre et al., 2006). On the other hand, mutations of ABA catabolism genes, such as the enzyme *CYP707A2*, lead to ABA over-accumulation and enhanced dormancy (Kushiro et al., 2004, Okamoto et al., 2006). ABA is also directly involved in the control of seed DT, as mutants with strong defects in ABA sensitivity fail to acquire DT during maturation (Ooms et al., 1993, Ooms et al., 1994). *ABSCISIC ACID INSENSITIVE 3* (*ABI3*), encoding a transcription factor of the B3-domain family, is one of the main master regulators of seed maturation affecting the acquisition of multiple SQTs (Karssen et al., 1983, Koornneef et al., 1982, Koornneef et al., 1984, Koornneef et al., 1989, Giraudat et al., 1992, Nambara et al., 1995, Nambara et al., 1994). ABI3 is one of the main transducers of the ABA signal during seed maturation, and strong mutant alleles of *ABI3* (e.g., *abi3-4*; *abi3-5* and *abi3-6*) fail to degrade chlorophyll and to acquire dormancy, DT and longevity, and also fail to accumulate seed storage proteins (SSPs) and late embryogenesis abundant proteins (LEAs) (Koornneef et al., 1989, Ooms et al., 1993, Ooms et al., 1994, Bies-Etheve et al., 1999, Sugliani et al., 2009).

Advances on gene expression analyses, particularly transcriptomics, facilitated genetic studies aimed at further understanding the ABI3-mediated transcriptional control of seed maturation. For example, using a strong *ABI3* allele mutant (*abi3-6*), Delmas et al. (2013) discovered that ABI3 is involved in embryo degreening during seed maturation by affecting the expression of enzymes involved in the chlorophyll degradation pathway. Furthermore, these authors discovered that ABI3 is involved in two independent seed maturation programs: one that regulates seed degreening through *STAY GREEN* (*SGR*) genes which is partially coupled to ABA sensitivity, and another that regulates seed maturation and acquisition of DT. Similarly, studies on the weak *ABI3* mutant allele (*abi3-1*) identified the *green-seed* (*GRS*) locus and *DELAY OF GERMINATION* (*DOG*) 1 as novel enhancers of the *ABI3*-dependent green seed phenotype (Clerkx et al., 2003, Dekkers et al., 2016). *DOG1* is a key regulator of seed dormancy acquisition, depth, and release in Arabidopsis (Alonso-Blanco et al., 2003, Bentsink et al., 2006, Nakabayashi et al., 2012). Interestingly, genetic, molecular, and transcriptomic evidence indicates that *DOG1* likely functions independently of ABA, but the pathways converge to regulate multiple downstream steps of seed maturation (Nakabayashi et al., 2012, Née et al., 2017, Dekkers et al., 2016). DOG1 mutant seeds (*dog1-1)* show reduced longevity and lower expression of several seed maturation genes such as *LEA*s and seed storage proteins (SSPs), while *abi3-1dog1-1* double mutants produce ABA insensitive, green seeds (Bentsink et al., 2006, Dekkers et al., 2016). Despite the importance of chlorophyll degradation for proper seed maturation (Clerkx et al., 2003, Teixeira et al., 2016, Li et al., 2017, Nakajima et al., 2012), it remains largely elusive why chlorophyll is produced and then degraded during seed maturation and how this process is temporally regulated (Leprince et al., 2017). Furthermore, the mechanisms that explain how *DOG1* (as well as *SGR*) enhances the *abi3-1* phenotype and how ABA levels and sensitivity are involved in embryo degreening are still largely unknown (Delmas et al., 2013, Dekkers et al., 2016). The integration of transcriptome data with advanced molecular assays like chromatin immunoprecipitation (ChIP) enabled the identification of direct and indirect targets of *ABI3*, providing additional evidence for the role of ABI3 in directly regulating the expression of genes associated with seed maturation, DT, dormancy, and longevity (Mönke et al., 2012, Tian et al., 2020). Direct targets of ABI3 include several *LEA*s involved in DT, genes responsible for the biosynthesis of Raffinose Family Oligosaccharides (RFOs) associated with seed longevity, genes involved in chlorophyll degradation, and the dormancy regulator *DOG1* (Tian et al., 2020). Together, the insights gained from these approaches highlighted how intricate and multifaceted the regulation governing seed maturation and the acquisition of SQTs is.

Gene co-expression network analyses have emerged as a commonly used method to unravel the complex regulation of processes related to seed maturation in multiple species (Basnet et al., 2013, Sekhon et al., 2014, Qi et al., 2018, Serin et al., 2016, Cheng et al., 2018, Niu et al., 2022, Song et al., 2022, Guo et al., 2023, Wang et al., 2023). Such approaches, combined with thorough temporal physiological analyses of SQT acquisition during maturation, allowed the identification of key regulators (hub genes) of specific SQT in multiple model and crop species, such as *Medicago truncatula*, maize, tomato, and soybean (Bizouerne et al., 2021, Righetti et al., 2015, Verdier et al., 2013, Zinsmeister et al., 2020a, Zinsmeister et al., 2016, Garg et al., 2017, Pereira Lima et al., 2017, Niu et al., 2022, Wang et al., 2023, Yao et al., 2023). Despite the widespread use of transcriptomics to study seed development in *A. thaliana* (Santos-Mendoza et al., 2008, Le et al., 2010, Belmonte et al., 2013, Holdsworth et al., 2008, González-Morales et al., 2016, Verma et al., 2022), there remains a scarcity of studies using large-scale co-expression network analysis to explore the specific regulation of SQT acquisition during Arabidopsis seed maturation, even with the advancements in co-expression network analysis tools (Cantó-Pastor et al., 2021). One reason for the limited focus on using Arabidopsis as a model to study the regulation of SQTs is the challenge in identifying their specific regulators. This difficulty is likely caused by the short length of the seed maturation phase and the overlap of processes, such as seed filling and late maturation drying (Verdier et al., 2013, Leprince et al., 2017). This complicates the dissection of temporal regulatory modules and genes associated with specific SQTs.

To overcome this challenge, we developed a comprehensive integrative approach aimed at identifying novel, specific regulators of SQTs in Arabidopsis. For this, we used wild-type (Columbia-0) and ten mutant combinations of known regulators of seed maturation and ABA signalling for a detailed temporal physiological analysis of SQT acquisition. We generated high-resolution transcriptome datasets to construct a gene co-expression network through gene clustering analysis that allowed us to perform correlation analysis of the physiological data to identify biological processes, gene modules and key regulatory genes specifically associated with different SQTs. To validate our transcriptome datasets and co-expression network, we identified candidate genes involved in DT and longevity. Using T-DNA insertion and multiple CRISPR-Cas9 mutants we functionally validated their role *in planta*. Finally, we generated an open access, user-friendly web-based tool, SeedMatExplorer (http://www.bioinformatics.nl/SeedMatExplorer/login), which offers experimental data, gene expression pattern analysis, gene module identification, and correlation analysis to allow further exploitation of our broad datasets by the scientific community.

Results

Physiological characterization of Arabidopsis seed maturation

We characterised temporal changes in physiological parameters (fresh weight, dry weight, and water content) and acquisition of SQTs (germination, dormancy, desiccation tolerance (DT), longevity, and chlorophyll content) associated with Arabidopsis maturation. We used seeds from wild-type Columbia-0 (WT, Col-0) and of single gene mutants and mutant combinations of genes involved in ABA biosynthesis (*ABA2*), signalling (*ABI3, DOG1*), and catabolism (*CYP707A2*) (Supplemental Table 1) were harvested and phenotyped every two days from 12 to approximately 26 (dry seed, DS) days after pollination (DAP) in two experimental rounds (Supplemental Table 2).

In WT Col-0 seeds at 12 DAP, the total seed weight (fresh weight) was composed of approximately 75% water (moisture content) and approximately 25% of dry matter (dry weight) (Figure 1A). From 12 DAP to 20 DAP the water content decreased to about 44% and the dry weight increased by approximately 80%, while the total seed weight slightly fluctuated. From 20 DAP until the DS stage a significant decrease in the water content to approximately 10% was observed, while the dry weight remained unchanged. Together, these observations suggest that Col-0 Arabidopsis seeds are in the seed filling phase from 12 DAP to 20 DAP and that the maturation drying phase starts at 20 DAP. Despite showing slightly distinct patterns of water loss, all mutants and the WT reached a low water percentage (approximately 9-12%) at 26 DAP (Supplemental Figure 1A).

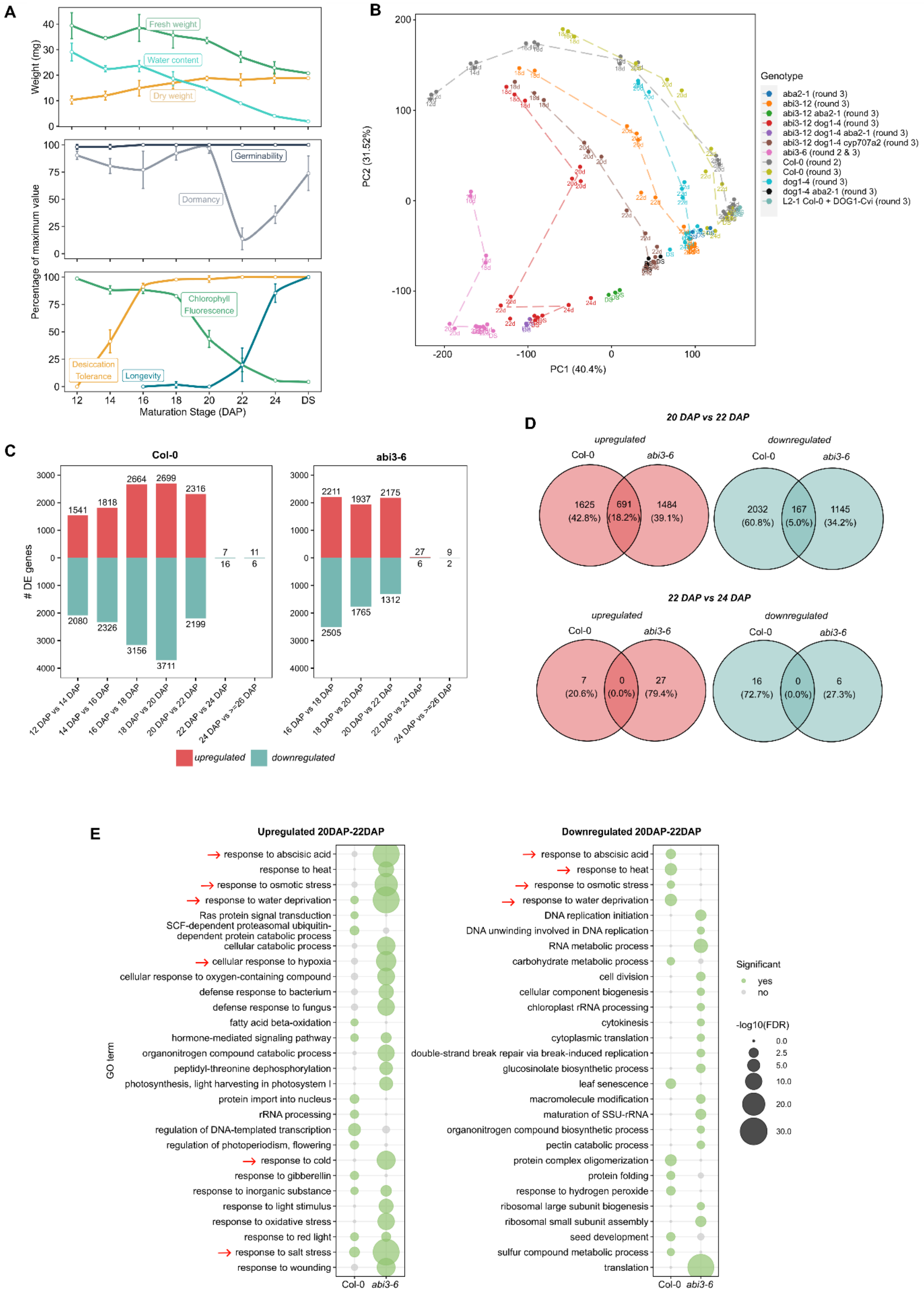


Figure 1. Physiological characterisation of *Arabidopsis thaliana* seed maturation. A - Physiological traits (top: seed fresh weight, dry weight, and water content) and seed quality

traits (SQTs) acquisition (middle: germinability and dormancy, bottom: desiccation tolerance, longevity, and chlorophyll fluorescence) for Arabidopsis (Col-0) seeds at different days after pollination (DAP) and in the dry seed (DS) stage. B - Global transcriptome relationships among different stages of seed maturation and mutants using principal component analysis (PCA) of the RNA-seq read counts (normalised with variance stabilising transformation) for 123 seed samples. C - Number of differentially expressed (DE) genes of Col-0 (round 2) and *abi3-6* (round 2 and 3) between consecutive days after pollination (DAP). D - Numbers of shared and unique up and downregulated genes (p-adj. < 0.01) for Col-0 and *abi3-6* seed maturation between 22 DAP and 20 DAP, and between 24 DAP and 22 DAP. E - Gene ontology (GO) term enrichment of shared and uniquely upregulated and downregulated genes in Col-0 and *abi3-6* between 20 DAP to 22 DAP (p-adj. < 0.05). The top 20 representative enriched GO categories among the upregulated (left) and downregulated (right) genes are shown. The green colour indicates significance, and size of the dots represent enrichment level ($-\log_{10}$p-adj.). Red arrows indicate biological processes related to stress response.

The germination ability of Col-0 seeds remained at 100% from 12 DAP until the end of the maturation drying phase (Figure 1A). However, seed dormancy fluctuated during the seed filling phase and showed a steep decrease from 20 DAP to 22 DAP, going from 100% to approximately 15% and increasing again to approximately 75% at the DS stage. As expected, the germinability of all genotypes with a mutation in the *DOG1* gene (*dog1-4*, *dog1 aba2-1*, *abi3-12 dog1-4*, *abi3-12 dog1-4 aba2-1*, *abi3-12 dog1-4 cyp707a2*) was higher and their dormancy percentage was overall lower than Col-0 throughout the maturation, while the opposite pattern was observed only in the *abi3-12* mutant (Supplemental Figure 1B-C).

We also observed a temporal separation in the acquisition of other SQTs in the Col-0 WT seeds. In Col-0, a dramatic decrease in chlorophyll fluorescence was observed from 18 DAP onwards, suggesting that chlorophyll degradation initiates during the seed filling phase and progresses until the end of seed drying (Figure 1A). The chlorophyll fluorescence of all mutants also decreased during maturation; however, *abi3-6* and *abi3-12 dog1-4 aba2-1* showed a slower decrease in chlorophyll fluorescence when compared to Col-0 (Supplemental Figure 1D). DT was acquired between 12 and 18 DAP during the seed filling phase, while longevity acquisition occurred in the maturation drying phase from 20 DAP to the DS stage (Figure 1A). Considering the definition of DT as the ability to survive the loss of almost all cellular water without irreversible cellular damage (Leprince and Buitink, 2010), the seeds of all mutants, except for *abi3-6*, acquired DT since they reached around 10% moisture content at the end of seed maturation (DS stage, Supplemental Figure 1A) and were still able to germinate with or without dormancy breaking treatment (Supplemental Figure 1B-C). However, at the DS stage, all mutants showed a quicker decrease in germination percentage during storage at an elevated partial pressure of oxygen (EPPO, 200 bar) when compared to Col-0 (Supplemental Figure 1E), suggesting that their longevity acquisition might have been impaired during seed maturation. The viability of Col-0 started to decrease after 80 days of EPPO storage, followed by *abi3-12*, *aba2-1,* and *dog1-4*. A higher reduction in storability was observed for *abi3-6* and for the double and triple mutants that combined the *abi3-12* allele with other mutations in ABA biosynthesis or signalling. To investigate the contribution of ABA accumulation to SQT acquisition during maturation, we quantified ABA at 18 DAP, the timepoint preceding the transition from seed filling and maturation drying, as well as in dry seeds of all mutants and wild-type (Supplemental Figure 2A-B). As expected, the single and combinations of the *aba2-1* mutant (*abi3-12 aba2-1* and *abi3-12 dog1-4 aba2-1*) had significantly lower ABA when compared to Col-0 at 18 DAP, indicating that decreased *ABA2* expression is sufficient to decrease the levels of ABA during seed maturation in these mutants (Supplemental Figure 2A). On the other hand, *abi3-12* and *abi3-12 dog1-4 cyp707a2* showed significantly higher ABA content. The higher accumulation of ABA in *abi3-12* could explain why this mutant displays much higher dormancy than all other mutants and the wild-type during seed maturation (Supplemental Figure 1C). On the other hand, loss of ABA catabolism enzyme CYP707a2 in the double *abi3-12 dog1-4* background explains the higher ABA content in its seeds at 18 DAP. At the dry stage all mutants had similar seed ABA content as the wild-type, except the *abi3-12 dog1-4 cyp707a2* triple mutant which had significantly higher ABA content (Supplemental Figure 2B) suggesting that, as long as *CYP707a2* is fully functional, seeds will accumulate wild-type levels of ABA at the end of maturation.

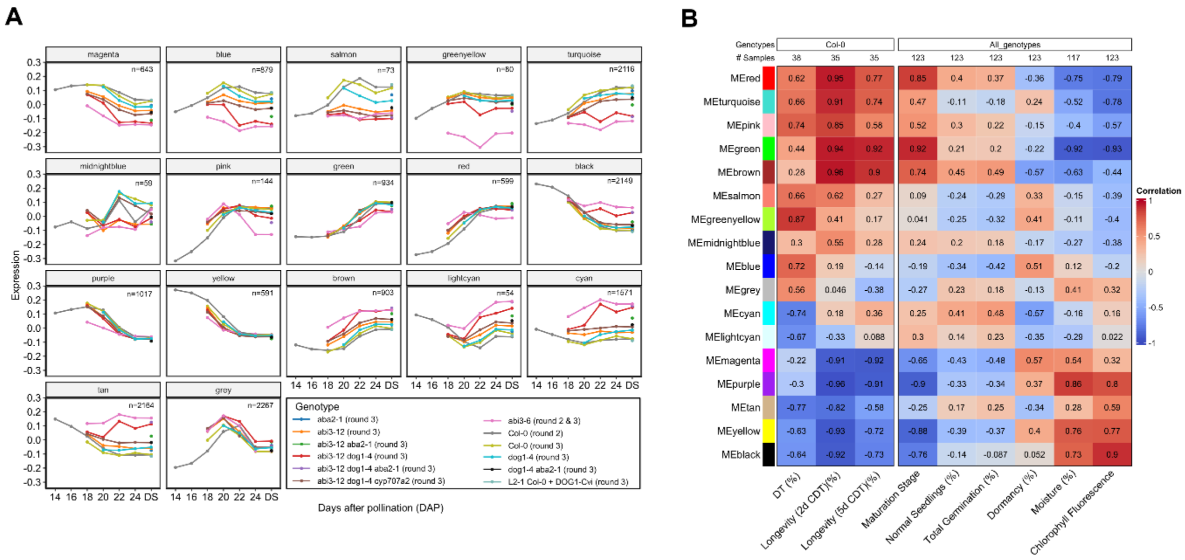
SeedMatExplorer: A transcriptome atlas of Arabidopsis seed maturation

To systematically identify transcriptional signatures associated with Arabidopsis seed maturation, we performed RNA sequencing (RNA-seq) of whole Col-0 seeds at eight stages throughout maturation, every two days from 12 DAP to dry seeds (DS, ~26DAP) (Supplemental Table 2, Supplemental Figure 3). We included multiple Arabidopsis seed maturation mutants at different seed maturation stages to obtain a global view of the transcriptomes associated with the acquisition of specific SQTs. Using principal component analysis (PCA) a clear separation between the different genotypes was observed in the first principal component (PC1), explaining 40.4% of the total variability, while PC2 shows a clear progression of the maturation time for each genotype, contributing with 31.52% of the variability (Figure 1B). We observed a gradual separation in the maturation time course among the various mutants compared to the maturation samples of Col-0. The *abi3-6* maturation time course was notably the furthest from the Col-0 samples while the *abi3-12* had a less extreme transcriptome profile change, suggesting a significant change in the transcriptional profile of seed maturation when *ABI3* expression is severely impaired. The transcriptional profiles of the maturation time course for *dog1-4* were closer to the respective time points in Col-0, suggesting that *DOG1* also influences the seed maturation transcriptome, although not to the same extent as the severe *abi3-6* mutation. As the *abi3-12* mutation was introduced into double and triple mutant combinations, we observed a greater separation between their samples compared to the same time points in Col-0 (Figure 1B). In Col-0, the number of up and downregulated genes between consecutive time points increased until 20 DAP (Figure 1C, Supplemental Data Set 1). From 22 DAP to 24 DAP a striking decrease in both up and downregulated genes indicates a potential shutdown of the transcriptional program as the seeds reach the dry state. In *abi3-6*, the number of upregulated genes remained constant until 22 DAP, while the number of downregulated genes gradually decreased. From 22 DAP to 24 DAP a striking decrease in both up and downregulated genes, similar to Col-0, was observed (Figure 1C, Supplemental Data Set 2). A very small percentage of upregulated (18.2%) and downregulated (5.0%) genes were shared between Col-0 and *abi3-6* before the transcriptome shutdown (20 - 22 DAP), and no up or downregulated genes were shared between both genotypes after the transcriptome shutdown (22 - 24 DAP) (Figure 1D). This suggests that, despite a similar transcriptome shutdown before reaching the dry seed stage, Col-0 and *abi3-6* have very distinct transcriptome profiles. Gene Ontology (GO) enrichment analysis revealed distinct enrichment of biological processes between Col-0 and *abi3-6* before the transcriptome shutdown (22 - 24 DAP) (Figure 1E, Supplemental Data Set 3). Overall, *abi3-6* showed upregulation of processes related to stress response, e.g. response to abscisic acid, response to osmotic stress, response to water deprivation, cellular response to hypoxia, response to cold, and response to salt stress. On the other hand, several of these stress-related processes were significantly downregulated in Col-0, which could suggest a stress response towards the end of seed maturation in *abi3-6*.

To allow easy visualisation and exploration of our seed maturation dataset for multiple *A. thaliana* mutants, we generated a user-friendly web interface, SeedMatExplorer (https://www.bioinformatics.nl/SeedMatExplorer/app/). It allows comprehensive visualisation of the SQTs with phenotyping data, individual gene expression analyses in the different mutants, differential expression (DE), gene ontology (GO) analyses, and weighted correlation network (WGCNA) analysis (described below).

Multiple co-expressed gene modules are associated with specific SQTs

Using dynamic time warping (DTW) (Giorgino, 2009) we aligned the RNA-seq sample time-series from two experimental rounds (Supplemental Figure 4A-B), allowing us to combine the datasets for further analysis. This provided a powerful basis for our weighted correlation network analysis (WGCNA), through which we were able to identify modules of co-expressed genes that could be linked to the SQTs from our extensive phenotypic data. We identified 17 modules with contrasting expression patterns, and we visualised their expression profiles as module eigengene, which represents the weighted average expression profile of the genes in the module (Figure 2A, Supplemental Data Set 4).



Genotypes	Col-0			All_genotypes					
# Samples	38	35	35	123	123	123	123	117	123
	DT (%)	Longevity (2d CDT)(%)	Longevity (5d CDT)(%)	Maturation Stage	Normal Seedlings (%)	Total Germination (%)	Dormancy (%)	Moisture (%)	Chlorophyll Fluorescence
MEred	0.62	0.95	0.77	0.85	0.4	0.37	-0.36	-0.75	-0.79
MEturquoise	0.66	0.91	0.74	0.47	-0.11	-0.18	0.24	-0.52	-0.78
MEpink	0.74	0.85	0.58	0.52	0.3	0.22	-0.15	-0.4	-0.57
MEgreen	0.44	0.94	0.92	0.92	0.21	0.2	-0.22	-0.92	-0.93
MEbrown	0.28	0.96	0.9	0.74	0.45	0.49	-0.57	-0.63	-0.44
MEsalmon	0.66	0.62	0.27	0.09	-0.24	-0.29	0.33	-0.15	-0.39
MEgreenyellow	0.87	0.41	0.17	0.041	-0.25	-0.32	0.41	-0.11	-0.4
MEmidnightblue	0.3	0.55	0.28	0.24	0.2	0.18	-0.17	-0.27	-0.38
MEblue	0.72	0.19	-0.14	-0.19	-0.34	-0.42	0.51	0.12	-0.2
MEgrey	0.56	0.046	-0.38	-0.27	0.23	0.18	-0.13	0.41	0.32
MEcyan	-0.74	0.18	0.36	0.25	0.41	0.48	-0.57	-0.16	0.16
MElightcyan	-0.67	-0.33	0.088	0.3	0.14	0.23	-0.35	-0.29	0.022
MEmagenta	-0.22	-0.91	-0.92	-0.65	-0.43	-0.48	0.57	0.54	0.32
MEpurple	-0.3	-0.96	-0.91	-0.9	-0.33	-0.34	0.37	0.86	0.8
MEtan	-0.77	-0.82	-0.58	-0.25	0.17	0.25	-0.34	0.28	0.59
MEyellow	-0.63	-0.93	-0.72	-0.88	-0.39	-0.37	0.4	0.76	0.77
MEblack	-0.64	-0.92	-0.73	-0.76	-0.14	-0.087	0.052	0.73	0.9

Figure 2. Module eigengenes (MEs) generated by WGCNA. A - ME expression profiles, representing the expression of the module's genes for each genotype throughout the maturation time course. These profiles are the first principal component of expression read counts that have been normalised with variance stabilising transformation. The numbers in the upper right corner indicate gene count in individual clusters. DS = Dry Seed. B - Module eigengene (ME) associations with the indicated seed physiological traits. The scale on the right indicates the Pearson correlation coefficient (PCC). The numbers at the top indicate the number of samples used for the analysis. DT = desiccation tolerance, CDT = controlled deterioration test.

The largest modules contained around 2200 genes (grey, tan, and black modules) and the smallest module contained 54 genes (lightcyan). Interestingly, the expression pattern of some modules, for example magenta, blue, turquoise, and cyan, clearly follow a similar genotype separation pattern as we have observed in the PCA (Figure 1C), suggesting that these gene modules significantly contribute to the underlying transcriptional differences between the multiple genotypes (Figure 2A). The greenyellow module showed larger differences in expression profile in *abi3-6*, with much lower gene expression levels in *abi3-6* compared to the other genotypes, which could indicate that this module contains genes whose expression underlies the extreme phenotypic differences of *abi3-6* (Supplemental Figure 1B-E).

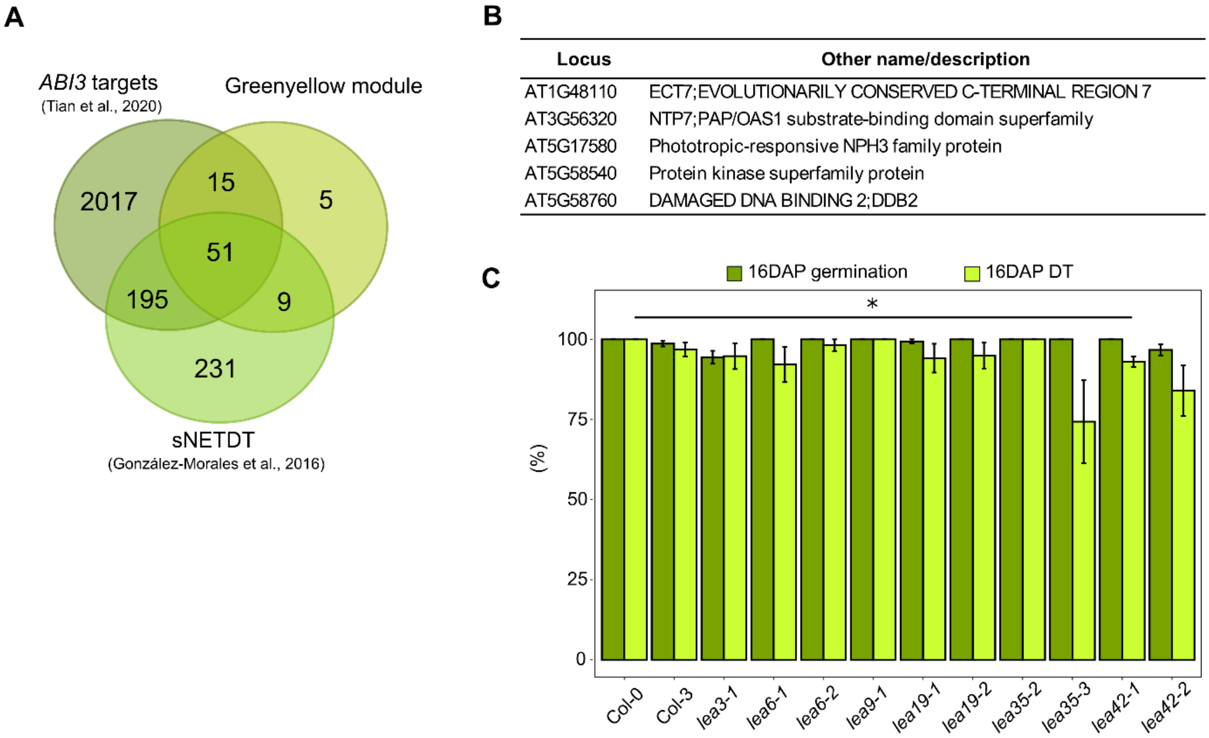
To further explore our gene modules and identify genes and transcription factors (TFs) associated with the regulation of SQTs, we calculated Pearson's correlation coefficient (PCC) to determine which of the 17 gene modules identified with WGCNA were significantly correlated with the trait data available for each of the genotypes (Figure 2B). Multiple distinct modules positively and negatively correlated with the same traits, indicating that the control of multiple SQTs investigated here has complex genetic regulation through multiple gene groups with distinct expression patterns. For the Col-0 SQTs, acquisition of DT had the highest correlation with the greenyellow module (PCC=0.87). The tan module showed the strongest negative correlation with DT (PCC=-0.77). Both longevity acquisition datasets for Col-0 (2d and 5d CDT) showed multiple strong positive and negative correlations, with the highest positive correlations (PCC above 0.9) with modules brown, red, green, and turquoise, and the highest negative correlation (PCC below -0.9) with modules purple, yellow, magenta and black. Using the phenotypic data acquired from all genotypes (wild-type and mutants), and the maturation stage (DAP) as a time component, we also identified several modules specifically associated with other seed traits. For example, the seed maturation stage (DAP) was highly positively correlated with module green (PCC=0.92) and module red (PCC=0.85), and highly negatively correlated with modules purple, yellow, and black. As expected, the opposite correlation trend between these modules was observed for water content and chlorophyll fluorescence, since these two traits are an indication of the seed maturation stage. The other traits (normal seedlings, germination, and dormancy) showed correlations below 0.51 and above -0.57 with all gene modules.

Given the importance of DT and longevity for seed resilience and survival in dry environments and under commercial (artificial) storage, we focused on the modules that could be most relevant for these traits to better understand their relationship and identify candidate regulators. For that, we made use of an approach that incorporates Pearson correlation, Gene Ontology (GO) enrichment, and enrichment of *LEA*s. Initially, we evaluated the Pearson correlation between the eigengenes of each module and the DT and longevity trait values. Correlation analysis was restricted to Col-0 as this genotype had the most complete time-series trait data (Supplemental Table 2). We considered modules surpassing a correlation threshold of 0.8, to be relevant to longevity or DT (Figure 2B). For the remaining genotypes, the data used for longevity and DT was obtained exclusively for fully mature seeds. We used this information on the relative differences between genotypes (e.g., higher versus lower longevity, Supplemental Figure 1) to filter out modules that, although correlated based on Col-0 data, showed contradicting differences between genotypes for trait and expression data. To address the limitations of a correlation-based approach, specifically that gene expression may precede trait acquisition and the fact that our correlation calculations did not include all genotypes, we employed additional measures to refine our selection of relevant modules. Namely, modules that approached or surpassed the correlation threshold were deemed more relevant if they showed enrichment of GO terms related to an abiotic stress response (Supplemental Figure 5). Additionally, we assessed the enrichment of the modules with *LEA*s using Fisher's exact test using the total pool of genes included in WGCNA as a background (Supplemental Table 3). The enrichment was considered significant if the resulting p-value was less than 0.05.

Based on the output of these analyses we highlight below the identification of candidate gene modules associated with DT and longevity.

A *LEA*-enriched gene module is associated with DT acquisition in Arabidopsis

Of all modules from our WGCNA analysis, the greenyellow module had the highest correlation with the DT acquisition trait. We observed a much lower expression of this gene module in the *abi3-6* genotype (Figure 2A), which correlates with the known desiccation sensitive phenotype of this mutant. Since ABI3 is a master regulator of DT acquisition in seeds, we selected genes from the greenyellow module based on existing ABI3 ChIP-seq experiments or transcriptome experiments (Mönke et al., 2012, González-Morales et al., 2016, Tian et al., 2020). We found that, from 80 genes in the greenyellow module, 66 genes (82.5%) are associated with *ABI3* regulation based on ChIP-seq experiments in embryonic tissues, of which 52 genes (65%) are directly, positively regulated by *ABI3* (Tian et al., 2020; Supplemental Table 4 and 5). To further corroborate that our analysis identified relevant regulators and to investigate if we found novel genes associated with DT, we also compared our greenyellow module genes with genes from two DT co-expression modules identified by González-Morales et al. (2016) (Figure 3A, Supplemental Table 6). We found that 60 genes (75%) of the greenyellow module genes have been identified as DT-associated in the co-expression network analysis of González-Morales et al. (2016), while 5 genes have been identified only in our analysis. These potentially novel DT-associated genes encode for an *EVOLUTIONARILY CONSERVED C-TERMINAL REGION 7* protein (*ECT7*, AT1G48110) involved in N6-methyladenosine (m6A) mRNA reading (Cai et al., 2024), a *NUCLEOTIDYL TRANSFERASE PROTEIN 7* (*NTP7*, AT3G56320) involved in microRNA (miRNA) cytidilation (Song et al., 2019); an NPH3 family protein (AT5G17580) found to be a target of miR415 (Zheng et al., 2012) and miR159/319 (Zhou et al., 2007), a protein kinase (AT5G58540) of unknown function, and a *DAMAGED DNA BINDING 2* (*DDB2*, AT5G58760) protein involved in DNA methylation and repair (Schalk et al., 2016) (Figure 3B).



Locus	Other name/description
AT1G48110	ECT7;EVOLUTIONARILY CONSERVED C-TERMINAL REGION 7
AT3G56320	NTP7;PAP/OAS1 substrate-binding domain superfamily
AT5G17580	Phototropic-responsive NPH3 family protein
AT5G58540	Protein kinase superfamily protein
AT5G58760	DAMAGED DNA BINDING 2;DDB2

Figure 3. A - Venn diagram of genes identified as direct or indirect targets of ABI3 through ChIP-seq and transcriptome analysis (Tian et al., 2020), genes from two Arabidopsis seed DT co-expression subnetworks (sNETDT, González-Morales et al. (2016), and genes from the greenyellow module. B – Functional descriptions of the 5 genes specifically identified as associated with DT in this study. C - Phenotypic analysis of 10 independent *A. thaliana LEA* T-DNA insertion lines belonging to DT modules for desiccation tolerance (DT) and

germination percentage at 16 DAP (Supplemental Table 7). Bars indicate standard error of the mean; asterisks indicate significant difference ($p<0.05$) when comparing mutant and respective wild type background (Col-3 as the wild-type background of *lea35-2* and *lea35-3*, and Col-0 as the wild-type background of all other mutants) for germination percentage or DT percentage according to Student's t-test ($n = 3$).

Interestingly, the greenyellow module also showed the highest enrichment of *LEAs* of all modules, with 11 genes (13.8%), of which 9 genes are direct targets of ABI3 (Supplemental Table 6, Tian et al. (2020)). Based on the high correlation of this module with DT acquisition, the high enrichment of known ABI3-regulated genes (Figure 2B, Supplemental Table 5), and the fact that *LEAs* have been strongly associated with DT in seeds, we hypothesised this module is the strongest candidate to contain genes involved in DT acquisition in Arabidopsis seeds. To validate the findings of our WGCNA and module-trait correlation analysis we used ten independent T-DNA mutant lines of six *LEA* genes identified in the greenyellow module and analysed their DT acquisition at 16 DAP, since at this time point wild-type Arabidopsis seeds acquired DT (Figure 1A, Supplemental Table 7). As expected, at 16 DAP, seeds of the wild-types were able to germinate at approximately 100% after desiccation (Figure 3C). Although the DT percentage for most LEA mutants did not show a significant difference compared to the wild-types, there was generally a greater variability in DT percentage observed at this time point. One LEA mutant line (*lea42.1*, group LEA_4, Supplemental Table 7) showed significantly reduced DT at 16 DAP. This gene is expressed in seeds and has previously been described to form droplet-like structures in imbibed seeds (Ginsawaeng et al., 2021, Hundertmark and Hincha, 2008), but its involvement with DT acquisition has not yet been described. Together, our co-expression and module-trait correlation analyses identified potentially novel genes involved with the acquisition of DT during seed maturation in Arabidopsis.

Transcription factor binding site analysis identifies regulators of longevity acquisition

Longevity is a complex quantitative trait regulated by multiple genetic factors and by the seed maturation and storage environments (Zinsmeister et al., 2020b). Accordingly, in our co-expression analysis we found that multiple gene modules were positively correlated with longevity acquisition using module-trait correlation analysis using the CDT data (Figure 2B). The brown (PCC=0.96), red (PCC=0.95), green (PCC=0.94), and turquoise (PCC=0.91) modules had the highest positive correlations, while magenta (PCC=-0.92), purple (PCC=-0.96), yellow (PCC=-0.93) and black (PCC=-0.92) had the highest negative correlations. We then decided to focus on identifying transcription factors (TFs) that might work as master regulators of the expression of these longevity-related genes. For this, we tested longevity-associated gene modules for enrichment in TF binding sites using TF2Network (Kulkarni et al., 2018). By selecting only genes that were included in the WGCNA analysis and by removing all TFs with a low expression or coefficient of variance, we filtered the list down to 332 TFs potentially associated with regulation of seed longevity (Supplemental Data Set 5). Enrichment analysis showed that nine different TF families were significantly overrepresented in our list:

ARR-B (p=0.0329), BES1 (p=0.0143), bZIP (p=5.36e-06), Dof (p=0.000188), ERF (p=2.89e-13), G2-like (p=5.34e-05), GATA (p=1.62e-08), MIKC_MADS (p=0.0413) and TCP (p=0.000335) (Figure 4A).

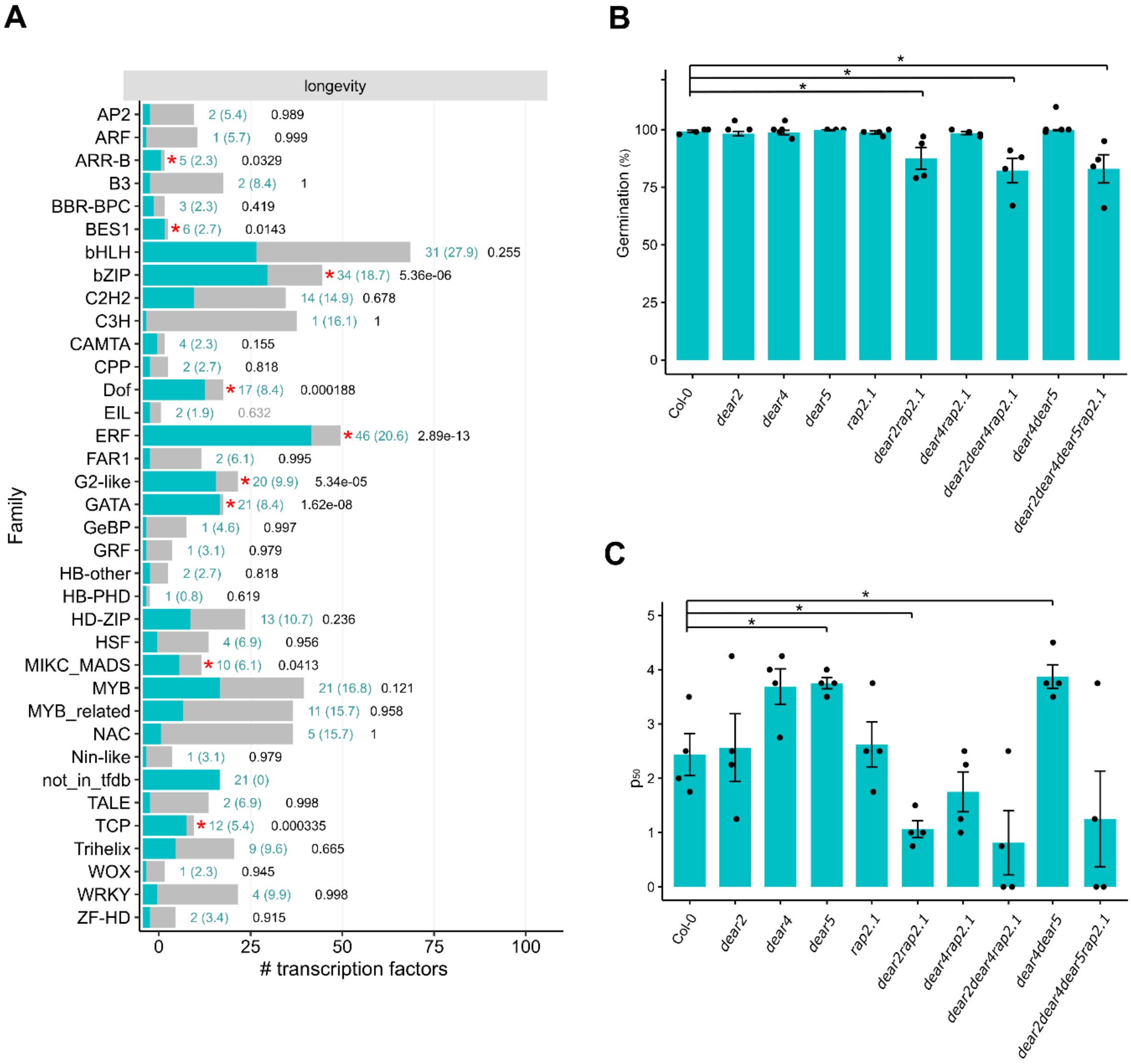


Figure 4. A - In turquoise: TF families identified by motif enrichment analysis with TF2Network for longevity modules. TF families were determined based on the plantTFDB list; the background distribution is shown as grey bars. The numbers adjacent to the bars indicate the number of TFs that were found in TF2Network analysis, and parentheses show the number of TFs that would have been expected by chance. Black values show the BH adjusted p-values (Benjamini and Hochberg, 1995) for Fisher's exact test. The result includes all transcription factors from TF2Network analysis that passed filtering for low expression and low variation, using the same thresholds as in the WGCNA analysis (see Materials & Methods). Red asterisks indicate significant enrichment of binding motifs in the dataset ($p<0.05$). B - Percentage of germination of mature dry seeds stored under natural conditions (50%RH, 20°C) for 47 days from Col-0 and multiple CRISPR-Cas9 mutants germinated in the presence of 10 µM GA_{4+7}

and 10 mM KNO_3. C - Half viability time (number of days to lose 50% of seed viability) of seeds submitted to CDT. Asterisks in B and C indicate significant differences ($p<0.05$) based on a Wilcoxon Sum Rank test and error bars indicate standard error ($n = 4$). The different CRISPR alleles and backgrounds are described in Supplemental Figure 7 and Supplemental Table 9.

We used multiple filtering steps (see Material & Methods) to carefully select a set of 25 TFs as the top candidates for the regulation of longevity acquisition during maturation (Supplemental Table 8). For further validation of the role of these TFs, we focused on four *DREB AND EAR MOTIF PROTEIN* (*DEAR*) from the DREB subfamily A-5 of ERF/AP2 transcription factor family as it contains stress-inducible TFs that act as transcriptional repressors of multiple abiotic stresses (Mizoi et al., 2012). We selected *DEAR4* (*RAP2.10*, AT4G36900), *DEAR5* (*RAP2.9*, AT4G06746), and *RAP2.1* (AT1G46768), and the closely related homolog *DEAR2* (AT5G67190) from the turquoise module and we generated single, double, triple, and quadruple mutants using CRISPR-Cas9 (Supplemental Table 9). *DEAR4* overexpression has been reported to lead to lower germination under stress conditions (ABA and NaCl treatments) and knock-down mutants display delayed leaf senescence (Zhang et al., 2020). *DEAR5* has been shown to be downregulated in response to heat and drought stresses, suggesting a role as a repressor under these conditions in Arabidopsis (Prasch and Sonnewald, 2013). *RAP2.1* has been described as a transcriptional repressor under cold and drought, as *rap2.1* T-DNA insert alleles showed reduced sensitivity to these stresses (Dong and Liu, 2010). To investigate the role of these candidate TFs on longevity we assessed germination after storage under natural conditions (50%RH, 20°C, 47 days) and after CDT (Figure 4B-C). We observed that mutants *dear2rap2.1*, *dear2dear4rap2.1*, and *dear2dear4dear5rap2.1* showed significantly lower germination than Col-0, which could indicate that the seed maturation processes, including longevity acquisition, are impaired in these mutants. We found that *dear5* and *dear4dear5* were significantly less sensitive to CDT compared to Col-0, while *dear4* was not significantly different from the wild-type despite showing a slightly higher viability after CDT. These results suggest that DEAR5 is a potential negative regulator of seed longevity in Arabidopsis. *dear2rap2.1* was significantly more sensitive to CDT while their respective single mutants did not differ from the wild-type, indicating a redundant role for *DEAR2* and *RAP2.1* on seed longevity. Together, these findings support a combinatorial and antagonistic function of DEARs in *Arabidopsis thaliana* seed longevity regulation.

Discussion

In Arabidopsis, the seed maturation and late maturation phases are shorter compared to other plant species (Leprince et al., 2017), which has been thought to lead to an overlap of events such as seed filling, acquisition of DT and longevity, and to complicate the identification of molecular regulators of specific SQT (Leprince et al., 2017, Verdier et al., 2013). With an extensive phenotypic and transcriptional analysis of the *Arabidopsis thaliana* wild-type (Col-0) and of mutants of known regulators of seed maturation, including *ABI3* and ABA biosynthesis and catabolism genes, we identified (groups of) genes involved in the control of the acquisition of different SQTs.

Seed maturation is characterised by an increase in seed dry mass caused by the accumulation of storage reserves and ends when seeds enter a quiescent stage caused by drying (Ali et al., 2022, Kozaki and Aoyanagi, 2022). Under the growth condition used in this study, the complete development of Arabidopsis until the dry seed (DS) stage lasted about 26 DAP, the seed filing occurred from 12 DAP to 20 DAP and drying occurred from 20 DAP onwards (Figure 1A). Changes in these physiological parameters were also accompanied by a dynamic change in the acquisition of SQTs. The start of chlorophyll degradation has been proposed as an indicator of the maturity level of seeds and the transition to the late seed maturation phase (Jalink et al., 1998, Teixeira et al., 2016, Leprince et al., 2017, Barboza da Silva et al., 2021, Zinsmeister et al., 2020b). Our data shows a sharp increase in chlorophyll degradation from 18 DAP onwards, while seed filling is still on the way (Figure 1 A, Supplemental Figure 1B). This indicates that the start of chlorophyll degradation in Arabidopsis does not overlap with the transition to the late seed maturation phase and, therefore, cannot be used as a parameter to indicate the maturity level of Arabidopsis seeds.

DT and longevity are two interdependent traits that are essential for long term dry seed survival and storability (Dekkers et al., 2015). Similar to species such as tomato (Bizouerne et al., 2021), soybean (Pereira Lima et al., 2017), and Medicago (Verdier et al., 2013), we observed a temporal separation between DT and longevity acquisition. DT acquisition started early during the seed filling phase, and at 16 DAP the seed population was fully desiccation tolerant (Figure 1A). On the other hand, the start of longevity acquisition overlapped with the start of the late seed maturation phase at 20 DAP. Contrary to what has been previously proposed, our findings show that it is possible to temporally separate the seed filling and late seed maturation events in Arabidopsis, as well as the onset of DT and longevity acquisition.

We then generated a comprehensive gene expression atlas of Col-0 and of ABA biosynthesis, catabolism, and signalling mutants to study the underlying transcriptional control of SQT acquisition (Supplemental Table 1). *ABI3* mutation strongly affects overall seed development, especially the accumulation of seed storage and LEA proteins (Nambara et al., 1992, Santos-Mendoza et al., 2008), chlorophyll degradation (Delmas et al., 2013); DT and longevity acquisition (Ooms et al., 1993) and dormancy acquisition (Léon-Kloosterziel et al., 1996). In line with this, our RNA-seq data analysis revealed that mutation on the strong allele of *ABI3* (*abi3-6*) leads to an extreme transcriptome shift during seed maturation when compared to Col-0 (Figure 1B), although both genotypes showed a transcriptome shutdown from 22 DAP onwards, after the onset of drying (Figure 1C). We found a significant upregulation of stress-related biological processes during *abi3-6* maturation contrary to Col-0, just before the global transcriptome shutdown at 22 DAP (Figure 1D). *ABI3* mutation has been shown to directly affect the expression of stress responsive genes in Medicago seeds under osmotic stress (Terrasson et al., 2013) and in Arabidopsis seedlings under dehydration stress (Bedi et al., 2016). It is possible that, with the start of seed drying at 20 DAP, the functional impairment of *ABI3* function in *abi3-6* leads to an enhanced expression of stress-responsive genes instead of the expected expression of DT-related genes. In fact, previous transcriptome studies have identified desiccation-associated cell-protective genes, such as *LEA*s, as direct targets of *ABI3* (Mönke et al., 2012, González-Morales et al., 2016, Tian et al., 2020).

By combined analysis of the RNA-seq data and the phenotypic data of SQT acquisition for Col-0 and ABA-related mutants, we identified seventeen distinct gene expression modules which could be correlated with specific SQTs (Figure 2). We found that acquisition of DT, acquisition of longevity, and the maturation stage have a similar module-trait correlation pattern (Figure 2B), indicating that DT and longevity are not only interdependent (Dekkers et al., 2015), but are also similarly influenced by the seed developmental stage. On the other hand, the correlation similarity between seed moisture and chlorophyll fluorescence was the opposite of what we observed for DT acquisition, longevity, and maturation stage. Although the physiological relationship between seed moisture content variation, especially drying, and chlorophyll degradation has been previously described for crop species (Adams et al., 1983, Green et al., 1998, Gomes et al., 2003, Sinnecker et al., 2005), the effects of seed drying on chlorophyll degradation during seed maturation have not yet been explored at the molecular level.

An in-depth analysis of the gene module with the highest correlation with DT (greenyellow module) showed a high enrichment of genes previously described as being part of the *ABI3* regulon and involved with DT (González-Morales et al., 2016, Tian et al., 2020) (Figure 3, Supplemental Table 3-5). Among these, we found several *LEA*s and five genes that, as of our knowledge, have not yet been associated with DT and that can be further explored in future research (Figure 3B). To validate the findings of our module-trait correlation analysis, we used T-DNA lines to assess the contribution of the greenyellow module-*LEA*s on DT. Overall, we found that mutation of individual *LEA*s led to a higher variation in DT acquisition in the seed population at 16 DAP, and mutation of *AtLEA42* led to significant impairment in seed DT acquisition. *LEA*s are known to belong to a group of low protein sequence similarity and highly redundant functions (Hundertmark & Hincha, 2008; Artur et al., 2019), therefore mutation in single *LEA* genes might not be sufficient to induce a quantifiable DT phenotype during seed maturation. Nevertheless, our study confirms the participation of *LEA*s on DT acquisition in Arabidopsis and indicates that certain LEAs may have a more prominent role in seed maturation than others.

Longevity is a complex, quantitative trait which is regulated by the combinatorial interaction of multiple TFs (Sano et al., 2016, Zinsmeister et al., 2020b, Alizadeh et al., 2021). Therefore, we turned our attention to the four gene modules with the highest positive correlation with the longevity trait in our dataset (brown, red, green, and turquoise, Figure 2B) to identify TFs involved in the control of this complex SQT. In our approach, we combined genes from the longevity-associated modules to identify enriched TF binding sites in their promoter regions and to identify candidate TFs that may bind to these motifs for validation. As expected, we identified several TF binding motifs enriched in the promoter of longevity-associated genes, including four *DEAR* transcription factors from the DREB subfamily A-5 of the ERF/AP2 family (Figure 4, Supplemental Table 9). Using CRISPR-Cas9 mutants we identified a repressing role for *DEAR5* and a potentially redundant positive role for *DEAR2* and *RAP2.1* on seed longevity. These findings disentangle part of the complex and multifactorial regulation of seed longevity and provide a list of potentially novel TFs that can be prioritised for further validations.

In conclusion, our study offers a comprehensive dataset that sheds light on the regulatory mechanisms of seed maturation in Arabidopsis, with a particular focus on the acquisition of seed quality traits (SQTs) critical for storage and survival. All this information is seamlessly integrated into SeedMatExplorer (https://www.bioinformatics.nl/SeedMatExplorer/app/), a unique resource designed to facilitate the testing and generation of new hypotheses about gene expression regulation during seed maturation and the control of multiple SQTs.

Materials and Methods

Plant materials and growth conditions

Arabidopsis thaliana (L.) Heynh. accession Columbia (Col-0) and several mutants known to display maturation defects (Supplemental Table 1) were used. *cyp707a2-1* (Salk_072410, Okamoto et al. (2006)), *aba2-1* (Léon-Kloosterziel et al., 1996), *abi3-6* (Nambara et al., 1994), *dog1-4* (SM_3_20808,Tissier et al. (1999)), *abi3-12* (Nambara et al., 2002) were obtained in the Nottingham Arabidopsis Stock Centre (NASC). The *dog1-4* is an induced transposon insertion mutant (SM_3_20808, Tissier et al. 1999) and was obtained as described by Bentsink et al. (2006). Genotyping of T-DNA mutants was performed by standard PCR using primers provided by the Salk Institute Genomic Analysis Laboratory (http://signal.salk.edu/tdnaprimers.2.html). The *abi3-12* described by Nambara et al. (2002) was genotyped using (d)CAPS markers forward: ATGGCGAGACAGAGGAGGTTCTTGTC and Reverse: CCGAGGTTACCCACGTCGCTTTGCT developed using dCAPS Finder 2.0 (http://helix.wustl.edu/dcaps/dcaps.html) (Neff et al., 2002) and wild-type fragments were digested with Ava II restriction enzyme (NEB, R0152). Double and triple mutants were obtained after crossing their respective parents. F2 and F3 individuals were genotyped using PCR and CAPS markers to confirm their double or triple mutant genotype. To generate dormant Col-0 lines (Col-0+*DOG1*-Cvi), Col-0 plants were transformed with a 5 kb construct containing the *DOG1*-Cvi gene (Bentsink et al., 2006). Tail-PCR located the insertion site in line 2-1-2, RT-qPCR on DNA confirmed the single insertion in this line. Plants were grown on 4 x 4 cm Rockwool blocks in a growth chamber at 20°C/18 °C (day/night) under a 16-h photoperiod of artificial light (150 µmol m^{-2} s^{-1}) and 70% relative humidity (RH). Plants were watered three times per week with the standard nutrient solution Hyponex (He et al., 2014). Several flowers were tagged in every plant of all genotypes just before anthesis to allow follow up on days after pollination (DAP). For each round of experiments plants were grown together and seeds were harvested from 12 DAP to 26 DAP in 2 days intervals. For each sample, four biological replicates were used, and each biological replicate consisted of 4 to 6 siliques. Fresh siliques were immediately dissected, and the seeds were carefully removed for seed quality traits (SQTs) phenotyping. The samples for RNA extraction were immediately frozen in liquid nitrogen and stored at -80 °C. The experiment was repeated 3 times, hereby called rounds 1, 2, and 3. Round 1 consisted of pre-tests to assess scalability for phenotyping and RNA-seq sampling and rounds 2 and 3 consisted of phenotyping for moisture content, fresh weight, dry weight, germination, dormancy, desiccation tolerance (DT), chlorophyll fluorescence, and longevity. RNA-seq sampling was also done for Col-0 and the mutants used in this study in both rounds 2 and 3 (Supplemental Table 2).

Seed quality trait phenotyping

The water content of the seed samples was determined gravimetrically on a fresh weight basis using two replicates of 3 siliques per time point for each genotype. The seeds were excised from the siliques using fine tweezers, the fresh weight was determined within 10 minutes after excision and the dry weight was determined after 24 hours of drying at 105 ºC. To assess germination capacity and dormancy level, four replicates of seeds immediately excised from 4 to 6 siliques were sown onto two layers of blue germination paper soaked with 40 ml H_2O and incubated at 22 °C in continuous light for 7 days. Pictures were taken twice or thrice every day to calculate the maximum germination percentage (gMax%) and time to reach 50% of germination (T50), the latter as a dormancy-related parameter using the Germinator program (Joosen et al., 2010). To obtain a measure of chlorophyll content we used a non-destructive method based on chlorophyll fluorescence. The seeds we analysed with the PathoScreen system developed by Wageningen University and Research and commercialised by PhenoVation B.V. (Kastelein et al., 2014). The same four replicates of seeds from 4 to 6 siliques used for germination phenotyping were analysed in the PathoScreen and the chlorophyll fluorescence signals were recorded as the ratio between the variable fluorescence (fv) and the maximum fluorescence. DT was assessed according to Ooms et al. (1993). In summary, four replicates of seeds from 4 to 6 siliques at different DAPs were dissected and artificially desiccated for 48 hours at 22 °C and 30% relative humidity (RH) in the dark. Desiccated and non-desiccated seeds were sown onto two layers of blue paper soaked with 10 µM GA_{4+7} and 10 mM KNO_3 to break possible residual seed dormancy. DT percentage was calculated as the difference in seed germination percentage after 5 days and the percentage of normal seedling formation after 7 days between desiccated and non-desiccated samples. Longevity assessment was performed using a controlled deterioration test (CDT) for Col-0 from 16 – 26 DAP (Delouche and Baskin, 2021, Powell and Matthews, 2005) (Delouche and Baskin, 2016; Powell and Matthews, 1984) and Elevated Partial Pressure of Oxygen (EPPO) assay for mature seeds of all genotypes (Groot et al., 2012). For CDT, four replicates of Col-0 seeds were placed in an opened 1.5ml tube and stored above a saturated KCl solution in a closed ventilated tank at 80−85 % relative humidity (RH) and 40 °C for 2 and 5 days. RH was monitored using a data logger (EL-USB-1-LCB, Lascar Electronics). For the EPPO assay three replicates of seeds from the total bulk harvest from all genotypes were placed in 2 ml tubes perforated with a hole of approximately 1 mm diameter closed with an oxygen-permeable polyethylene membrane, placed in 1.5 L steel tanks, filled with air at a rate of 0.4 MPa per minute from buffer tanks until the tank pressure reached approximately 20 MPa (Buijs et al., 2018). The tanks were stored at 22°C and 55% RH, and the seeds were collected after 21, 42, 62, 77, 91, 105 and 119 days. After CDT and EPPO, seed germination data analysis was performed as described previously using Germinator (Joosen et al., 2010). ABA quantification was performed on 18 DAP and dry seeds. In summary, five milligrams of seeds were ground to fine powder and extracted with 1 mL of 10% (v/v) methanol containing 100 nM deuterated ABA as an internal standard. The extraction was further carried out according to Floková et al. (2014) with minor modifications, namely the use of a Strata-X 30 mg/3 mL SPE column (phenomenex). For the detection and quantification of ABA by LC-MS/MS, we used the same procedure as Zhang et al. (2024) with minor modifications. Multiple reaction monitoring (MRM) mode was used for the quantification and identification of ABA, by comparing the retention time and MRM transitions (263.25>153.15 + 263.25>219.15). The final content of each sample was normalised using the internal standard.

RNA extraction, sequencing, and analysis

Total RNA was extracted according to the hot borate protocol modified from Wan and Wilkins (1994). Two to three replicates of a pool of seeds from 4 siliques for each treatment (genotypes and time points) to a total of123 samples (Supplemental Table 2) were used. The samples were homogenised and mixed with 800 µL of extraction buffer heated to 80 ºC (0.2 M Na borate decahydrate (Borax), 30 mM EGTA, 1% SDS, 1% Na deoxy-cholate (Na-DOC), 1.6 mg DTT and 48 mg PVP40). One milligram of proteinase K was added to this suspension and incubated for 15 min at 42ºC. After adding 64 µL of 2 M KCl, the samples were incubated on ice for 30 min and subsequently centrifuged for 20 min at 12,000 g. 270 µL of ice-cold 8 M LiCl was added to the supernatant in a final concentration of 2 M and the tubes were incubated overnight on ice. After centrifugation for 20 min at 12,000 g at 4 ºC, the pellets were washed with 750 µL ice-cold 2 M LiCl. The samples were centrifuged for 10 min at 10,000 g at 4 ºC and the pellets were resuspended in 100 µL DEPC treated water. The samples were cleaned with phenol chloroform and treated with DNAse (RQ1 DNase, Promega). The RNA quality and concentration were assessed by agarose gel electrophoresis and UV spectrophotometry. RNA was processed for use in RNA-seq with mRNA enrichment polyA capture (Illumina Incorporated, San Diego, CA, USA). The sequencing of a total of 123 samples was performed strand specific on Illumina Hiseq 2500, using cDNA and random hexamer priming, and generating single-end 125 nt reads.

The raw reads were mapped to the TAIR10 Arabidopsis reference genome with the Araport11 annotation (Cheng et al., 2017) using the HISAT2 software v2.1.0 (Kim et al., 2015). The alignment rate was about 99%. Transcript expression was quantified from uniquely mapped reads using the StringTie program v1.3.2 (Pertea et al., 2015). StringTie's prepDE.py script was used to extract read counts from the StringTie output. Fragments per kilobase of exon per million fragments mapped (FPKM) were calculated from StringTie output with the R package Ballgown (Frazee et al., 2015). Pearson correlation coefficient was calculated between biological replicates with the normalised expression levels of log2 (FPKM+1). DESeq2 (Love et al., 2014) was used for differential expression (DE) analysis. For exploring gene functions during analyses, the Araport11 annotation (Cheng et al., 2017) .gff file was downloaded from arabidopsis.org. Custom R scripts (R v4.0.3) were used to parse the information and link it to the relevant genes during the analyses. To analyse sample variation, seed gene expression profiles were investigated using principal component analysis (PCA) and hierarchical clustering. For PCA, we first normalised the RNA-seq read counts of all genes with a variance stabilising transform (VST) using the R package 'DESeq2' V1.30.0 (Love et al., 2014). We then conducted PCA using the R function prcomp with default settings and visualised the results using ggplot2.

Dynamic time warping

To align time series data of round 2 and 3 samples from Col-0 and *abi3-6*, dynamic time warping was performed by using functions from the R package 'dtw' (V1.22.3) (Giorgino, 2009). The means of replicates were used for time warping to prevent the assignment of different time labels to replicates. Since the phenotype data for round 3 was more complete and could be used for both correlations with round 2 and 3 expression data, the expression data of Col-0 round 3 samples was used as a template and the expression of round 2 was used as a query for the dynamic time warping alignment. In summary, the time axis alignment was done with the option to allow open beginnings or ends for the alignment enabled. This restricted the choice of step-patterns to 'asymmetric', 'rigid' and 7 types of Rabiner- Juang step-patterns and the most common alignment was selected as the final result. The majority of step-patterns obtained suggested adding 2 days to the labels of early Col-0 round 2 expression data (12-22 DAP) and not changing the labels of the latest time points. For round 2 samples of the *abi3-6* genotype, the same transformation was applied, since the lack of overlap with the round 3 time points made another dynamic time warping impossible. The effect of dynamic time warping is illustrated by comparison of the WGCNA module eigengenes with or without time warping of Col-0 and *abi3-6* round 2 samples, in the Supplemental Figure 4A-B.

Gene co-expression network analysis

We constructed a co-expression network using the WGCNA package V1.69 (Langfelder and Horvath, 2008) in R, including all 123 samples for the 11 available genotypes in the analysis. Three different criteria were used to filter the genes: having an expression of above 25 counts in at least 25% of the samples, at least 9 samples with an expression above 100 counts, or a coefficient of variation above 0.3. The expression of the remaining genes was normalized using the varianceStabilizingTransformation function from the R package DESeq2 V1.30.0 (Love et al., 2014). We next checked for outliers among the samples by conducting hierarchical clustering using R's hclust function with average linkage and Euclidean distance. For constructing the co-expression network calculated the adjacency of genes with a soft-thresholding power of 8 for a scale-free topology fit of a 'signed hybrid' network. The adjacency matrix was then transformed into a topological overlap matrix (TOM) to derive the weighted network. Genes were clustered hierarchically with average linkage, using 1-TOM as a distance metric. To separate the clustering result into modules, the function cutreeDynamic from the dynamicTreeCut package V1.63-1 was used with deepsplit=0 and a minimum required module size of 30 genes. Highly similar modules were merged with mergeCloseModules from the WGCNA package (Langfelder and Horvath, 2008), combining all modules with a dissimilarity of less than 0.14 to a final 17 modules with distinct patterns.

We determined the functions associated with our WGCNA modules through Gene Ontology (GO) enrichment analysis of biological processes, which we conducted using topGO R package v2.42.0. To map GO terms to genes, we used the annotations available in the org.At.tair.db database. The significance of GO term enrichments in the modules was tested with the runTest function of topGO, selecting the 'weight01' algorithm and Fisher's exact test to calculate p-values. All genes included in the WGCNA analysis served as the background for these tests. For multiple testing correction, we applied the BH-FDR correction method (Benjamini and Hochberg, 1995).

Motif enrichment and transcription factor selection

Motif enrichment was performed with the selected modules to find potential transcription factors (TFs) relevant to the traits. All genes from the selected modules were used for motif enrichment, except for the longevity-related modules turquoise and black because the large number of genes in these modules (>2000) would decrease the specificity of the analysis. Here, the EPPO data from all genotypes allowed for subsetting the genes based on ranked correlation. Genes falling in the top right quadrant of the module membership versus gene significance plot of the modules were selected to be used in motif enrichment (Supplemental Figure 6). The threshold gene significance (GS) and module membership (MM) were chosen to reduce the genes by about half. For the turquoise module: MM>0.78, GS>0.65, and black module MM>0.8, GS>0.69. The genes were submitted to TF2Network (Kulkarni et al., 2018) and the TFs found based on enriched position weight matrices were exported for further filtering. TFs of interest were manually selected based on their expression profiles for the different genotypes. The following two selection criteria were used: The first criterion was based on selecting TFs that showed an increase or decrease in the expression profile for Col-0 at 18 or 20 DAP for longevity acquisition. With the second criterion, TFs that had different expression profiles for the mutants were prioritised, separating genotypes that acquire longevity from those that do not. Some TFs could be further prioritised by showing that they are directly regulated by *ABI3* based on the research of Tian et al. (2020), or by finding that they have genes as nearest neighbours that were previously shown to be important for the trait of interest, like *LEA*s for DT and longevity (full list provided in Supplemental table 10 and 11). We considered genes nearest neighbors if they showed a topological overlap score of 0.2 or greater in the WGCNA TOM matrix. The motivation for selecting each transcription factor is described in Supplemental Table 8.

T-DNA and CRISPR-Cas9 mutants for phenotypic validations

Ten *Arabidopsis thaliana* T-DNA lines of six *LEA*s identified in the DT greenyellow module (Supplemental Table 7) were obtained from the Nottingham Arabidopsis Stock Centre (NASC) and genotyped using standard primers from SIGnAL-SALK (Supplemental Table 12). Out of the ten T-DNA lines, eight were SALK lines having a Col-0 background, while two (*lea35*.2 and *lea35.3*) were SAIL lines with a Col-3 background. To generate CRISPR-Cas9 mutants, genes were selected based on the list of TFs related to longevity acquisition and filtered as described previously. To build the constructs, closely related genes (based on protein sequence) were targeted simultaneously in one construct (Supplementary Table 13). For each gene, we designed one or two guides using CHOPCHOP V3 (Labun et al., 2019) (Supplemental Figure 7, Supplementary Table 13). Target sites were selected for the lowest mismatch value and closest proximity to the translational start site, with a preference for the first exon, if specific enough. The guide-RNAs and subsequent expression vectors were cloned using the GoldenGate cloning system described by Stuttmann et al. (2021). As an expression vector, we chose the pDGE885 (pNOS:nptII_Ubq:FCY-UPP_pRPS5a:zCas9io_ccdB_CmR) to select transformants for kanamycin resistance and to later on select against transformants using the fungicide 5-fluorocytosine (5-FC) (Stuttmann et al., 2021). Col-0 plants were transformed using floral dipping and T1 seeds were harvested in bulk for each construct and immediately germinated on ½ MS plates with 100 µg/ml kanamycin to select only successfully transformed seedlings. For each construct, 11 to 35 surviving seedlings were transferred to rockwool under greenhouse conditions and genotyped for signs of mutagenesis. Gene specific primers surrounding the target sites of each locus were used to genotype all lines (Supplemental Table 13). Primers were designed ~150 bp upstream and ~300 bp upstream of the target sites close to the ATG, and for genes targeted with two target sites an additional primer was designed ~150 bp downstream of the distal target site. T1 seedlings showing a PCR product size differing from Col-0 were selected for further propagation. Seeds from each T1 parent were propagated separately on 1 mM 5-FC ½ MS plates to select against the Cas9 constructs. Eight surviving T2 seedlings per plate were transferred to rockwool for further selection and were genotyped as a pool for signs of mutagenesis. Pools of T2 plants that showed signs of mutagenesis were further genotyped by PCR as single plants. The individual T2 plants with mutations visible by PCR were selected for further propagation and, in addition, plants with a homozygous mutation were sequenced. Homozygous plants were then propagated and used for trial phenotyping experiments and T2 plants with one or more non-homozygous mutations were propagated for another round of selection for homozygosity in the next generation.

Mutant phenotypic analysis

T-DNA mutants were phenotyped for DT as described previously (Ooms et al., 1993) with modifications. Plants were grown under controlled conditions as described previously and siliques were harvested at 16 days after pollination (DAP) and dissected using tweezers with the help of a stereomicroscope to remove the seeds. Three replicates of seeds pooled from three siliques were immediately desiccated for 8 hours at 22°C and 30% RH in dark or directly sown onto two layers of blue germination paper soaked with 40 ml dH_2O. After seven days, non-germinated seeds were transferred to blue germination paper containing 10 µM GA_{4+7} and 10 mM KNO_3 and incubated for another 5 days to break residual dormancy. DT percentage was calculated as described previously. The total number of seeds at the start of the assay and the number of germinated seeds with and without desiccation after seven days were determined and statistically analysed using the Germinator software (Joosen et al., 2010). The CRISPR-Cas9 mutant plants were grown under the conditions described above. For each line, four replicates containing four plants were used and seeds were harvested from each replicate individually. After three months of after-ripening (50% RH, 20° C) seeds were subjected to CDT. For this, seeds were stored in the dark at 89% Relative Humidity (RH) and 40°C. After 0, 2, 5, 7, 8, 9, and 10 days of storage, germination on dH_2O was assessed. Maximum germination percentage ($G_{max(\%)}$) was determined after seven days, making use of the

Germinator software (Joosen et al., 2010). For each replicate, G_{max} was fitted to a logistic curve to determine the half-viability period (p_{50}). To assess if mutant lines had longevity phenotypes, p_{50} of each mutant line was compared to Col-0. Statistical analyses were performed using a Wilcoxon rank-sum test ($\alpha = 0.05$).

Accession numbers

Identifiers for all genes mentioned in this work are provided in Supplemental Table 14.

Supplementary Data

Supplemental Figures

Supplemental Figure 1. Seed maturation traits of the *Arabidopsis thaliana* mutants *aba2-1, abi3-6, abi3-12, aba2-1 abi3-12, dogl-4*, *dogl aba2-1*, *abi3-12 dogl-4*, *abi3-12 dogl-4 aba2-1*, *abi3-12 dogl-4 cyp707a2 and* L2-1 Col-0 + DOG1-Cvi. A - Water content percentage, B - Total germination percentage, C - Dormancy percentage, D - Chlorophyll fluorescence (fv/fm), and E - Germination percentage after Elevated Partial Pressure of oxygen (EPPO) treatment. Each mutant genotype is represented by the same colours in all panels. A-D was performed using seeds from different maturation stages (days after pollination, DAP), n=4. EPPO treatment was performed in mature seeds (>26DAP), n = 3.

Supplemental Figure 2. ABA content in (A) 18 DAP and (B) dry seeds. Values are normalised by 5 mg of seeds and internal standard (IS) by LC-MS/MS, (n=4). ANOVA was conducted to test for differences in mean ABA content across genotypes, pairwise comparisons were made using Tukey's HSD post-hoc test. In the figure, a compact letter display illustrates statistically similar means: genotypes sharing the same letter in panel A or B have means that are not significantly different. Specifically, a shared letter indicates a Tukey HSD p-value above the significance threshold of 0.05.

Supplemental Figure 3. Hierarchical clustering of RNAseq samples using R's hclust function with average linkage and Euclidean distance.

Supplemental Figure 4. Effect of dynamic time warping on module eigengenes (MEs). A. MEs without warping time axis. B. MEs with warped time axis. Time warping was used to adjust time point labels of round 2 to round 3, only Col-0 and *abi3-6* samples are affected. Adjustment of time axis labels for round 2 samples was as follows: +2 days for each sample from 12 to 22 days. This figure shows a better agreement of the ME profiles of Col-0 round 2 (grey) and Col-0 round 3 (lightgreen), clearly noticeable when comparing for the black module, before and after time-warping. Improved alignment of time-series for abi3-6 round 2 samples is noticeable when comparing the ME of *abi3-6* (pink) with the ME of *abi3-12 dog1-4 (red)*, for example for the turquoise module, before and after time-warping.

Supplemental Figure 5. Overview of significantly enriched GO terms and trait correlations for modules obtained from WGCNA analysis. The selection of GO terms includes all biological processes with a significant enrichment (padj <0.05). A column with grey text displays the number of genes in each module annotated with a GO term, as well as the total number of genes annotated with the term in the background population. The expected number of genes associated with each GO term is shown based on the total number of genes in the module and the prevalence of the term in the background population. At the bottom of each graph, a heatmap is presented depicting the Pearson correlation between the eigengene of each module and a set of traits. The p−value corresponding to the correlation shown in parentheses.

Supplemental Figure 6. Subsetting of the genes from WGCNA modules 'black' and 'turquoise' to select those most relevant to longevity acquisition. The genes are separated on the scatter plots' x-axis based on module membership (MM), the correlation of a gene's expression profile with the module eigengene of its respective WGCNA module, and on the y-axis by the correlation of their dry stage expression across genotypes with longevity ranks for those genotypes from the EPPO experiment. Only genes that fall in the area shaded green (top-right) were included in longevity motif enrichment analysis.

Supplemental Figure 7. Overview of CRISPR-Cas9 mutant construct generation. Sg1 - Single guide RNA 1.

Supplemental Tables

Supplemental Table 1. Overview of genotypes used in this study and their respective phenotypes.

Supplemental Table 2. RNA-seq sample overview. Number of samples sequenced per time point (DAP = days after pollination, DS = Dry Seed).

Supplemental Table 3. Enrichment of LEAs amongst the ABI3 regulated genes for WGCNA modules. P-values were obtained with Fisher's exact test.

Supplemental Table 4. Description and ABI3 regulation of genes belonging to the greenyellow module (DT). LEA genes are indicated in bold.

Supplemental Table 5. Enrichment of ABI3 regulated genes for WGCNA modules. P-values were obtained with Fisher's exact test.

Supplemental Table 6. Genes that are direct or indirect targets of ABI3 based on Tian et al. (2020), genes from the greeyellow module (this work) and genes of two desiccation-tolerance subnetworks based on González-Moralez et al. (2016).

Supplemental Table 7. The T-DNA lines of Arabidopsis LEA (Atlea) from the greenyellow module used for desiccation tolerance assessment.

Supplemental Table 8. Filtered lists of candidate transcription factors involved in longevity regulation in Arabidopsis. Genes selected for CRISPR-Cas9 mutant validation are indicated in bold.

Supplemental Table 9. Summary of CRISPR-Cas9 lines generated in this study.

Supplemental Table 10. List of genes with described association with seed longevity in multiple plant species.

Supplemental Table 11. Primers used for genotyping LEA T-DNA lines.

Supplemental Table 12. Overview of CRISPR-Cas9 mutants of longevity-related TFs.

Supplemental Table 13. List of guide RNAs and genotyping primers used for CRISPR-Cas9 mutants generated in this study.

Supplemental Table 14. Gene names and IDs mentioned in the text.

Supplemental Data Sets

Supplemental Data Set 1. Differentially expressed (DE) genes between individual subsequent days after pollination (DAP) for Col-0. AGI – gene identifier, log2FC – log2 fold- change, stat - statistics, p-adj – false discovery rate adjusted p-value <0.01.

Supplemental Data Set 2. Differentially expressed (DE) genes between individual subsequent days after pollination (DAP) for abi3-6. AGI – gene identifier, log2FC – log2 fold- change, stat - statistics, p-adj – false discovery rate adjusted p-value <0.01.

Supplemental Data Set 3. Gene Ontology (GO) term enrichment for downregulated and upregulated differentially expressed (DE) genes at 20 and 22 days after pollination (DAP)

Supplemental Data Set 4. List of TFs identified using TF2Network and in the WGNA analysis.

Supplemental Data Set 5. Genes belonging to 17 modules created with WGCNA.

Acknowledgements

We thank all members of the Wageningen Seed Science (WSSC) for their contribution with the experiments, Francel Verstappen from the Laboratory of Plant Physiology of Wageningen University and Research for his help with hormone measurements and Dr. Parvinderdeep Kahlon from Laboratory of the Plant Physiology from Wageningen University and Research for his insightful feedback on the manuscript.

Funding

This work was supported by The São Paulo Research Foundation and The Netherlands Organization for Scientific Research (FAPESP/NWO, project 17/50211-9) to R.N.T. and H.H., NWO-ENW Veni (project Fine Drying VI.Veni.202.038) to M.A.S.A., NWO-TTW Open Technology Program (project HEAT 19932) to L.L.B., NWO VICI (project Seeds4Ever 17047) to L.B., A.D.D and L.A.J.W., and NWO-VICI (project VI.C.192.033) to J.A.D, NWO-ALW Open (project Green Seeds) to B.J.W.D., and The National Council for Scientific and Technological Development Brazil (CNPq 246220/2012-0) to A.C.S.S.M.

Data availability

The sequencing data generated in this study are available in the Gene Expression Omnibus (GEO) under accession number GSE270988. Processed data can be found in the SeedMatExplorer browser https://www.bioinformatics.nl/SeedMatExplorer/app/. Requests for access links should be directed to seedmatexplorer@bioinformatics.nl.

Author contributions

B.J.W.D., A.C.S.S.M. and R.N.T. designed the research.

M.A.S.A., L.A.J.W., L.L.B., A.D.D., J.A.D., B.J.W.D., A.C.S.S.M.., A.A.S. and R.N.T. performed seed physiological experiments.
L.L.B., A.D.D., J.A.D. and R.N.T. designed and generated CRISPR mutants.
R.A.K., H.N. and R.N.T. performed bioinformatics analysis.
M.A.S.A., R.A.K., L.L.B., A.A.S. and R.N.T. analysed experimental data.
M.A.S.A., R.A.K. and R.N.T. wrote the manuscript.
B.J.W.D., L.B., H.H. and R.N.T. conceived and supervised the project.
All authors commented on the article and agreed upon the final version of the manuscript.

References

ADAMS, C. A., FJERSTAD, M. C. & RINNE, R. W. 1983. Characteristics of Soybean Seed Maturation: Necessity for Slow Dehydration 1. *Crop Science,* 23**,** 265-267.

ALI, F., QANMBER, G., LI, F. & WANG, Z. 2022. Updated role of ABA in seed maturation, dormancy, and germination. *Journal of Advanced Research,* 35**,** 199-214.

ALIZADEH, M., HOY, R., LU, B. & SONG, L. 2021. Team effort: Combinatorial control of seed maturation by transcription factors. *Current Opinion in Plant Biology,* 63**,** 102091.

ALONSO-BLANCO, C., BENTSINK, L., HANHART, C. J., VRIES, H. B.-D. & KOORNNEEF, M. 2003. Analysis of natural allelic variation at seed dormancy loci of Arabidopsis thaliana. *Genetics,* 164**,** 711-729.

ANGELOVICI, R., GALILI, G., FERNIE, A. R. & FAIT, A. 2010. Seed desiccation: a bridge between maturation and germination. *Trends in plant science,* 15**,** 211-218.

BARBOZA DA SILVA, C., OLIVEIRA, N. M., DE CARVALHO, M. E. A., DE MEDEIROS, A. D., DE LIMA NOGUEIRA, M. & DOS REIS, A. R. 2021. Autofluorescence-spectral imaging as an innovative method for rapid, non-destructive and reliable assessing of soybean seed quality. *Scientific reports,* 11**,** 17834.

BASNET, R. K., MORENO-PACHON, N., LIN, K., BUCHER, J., VISSER, R. G., MALIEPAARD, C. & BONNEMA, G. 2013. Genome-wide analysis of coordinated transcript abundance during seed development in different Brassica rapa morphotypes. *BMC genomics,* 14**,** 1-19.

BAUD, S., BOUTIN, J.-P., MIQUEL, M., LEPINIEC, L. & ROCHAT, C. 2002. An integrated overview of seed development in Arabidopsis thaliana ecotype WS. *Plant Physiology and Biochemistry,* 40**,** 151-160.

BEDI, S., SENGUPTA, S., RAY, A. & CHAUDHURI, R. N. 2016. ABI3 mediates dehydration stress recovery response in Arabidopsis thaliana by regulating expression of downstream genes. *Plant Science,* 250**,** 125-140.

BELMONTE, M. F., KIRKBRIDE, R. C., STONE, S. L., PELLETIER, J. M., BUI, A. Q., YEUNG, E. C., HASHIMOTO, M., FEI, J., HARADA, C. M. & MUNOZ, M. D. 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proceedings of the National Academy of Sciences,* 110**,** E435-E444.

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**,** 289-300.

BENTSINK, L., JOWETT, J., HANHART, C. J. & KOORNNEEF, M. 2006. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proceedings of the national academy of sciences,* 103**,** 17042-17047.

BIES-ETHEVE, N., DA SILVA CONCEICAO, A., GIRAUDAT, J., OME, KOORNNEEF, M., LÉON-KLOOSTERZIEL, K., VALON, C. & DELSENY, M. 1999. Importance of the B2 domain of the Arabidopsis ABI3 protein for Em and 2S albumin gene regulation. *Plant molecular biology,* 40**,** 1045-1054.

BIZOUERNE, E., BUITINK, J., VU, B. L., VU, J. L., ESTEBAN, E., PASHA, A., PROVART, N., VERDIER, J. & LEPRINCE, O. 2021. Gene co-expression analysis of tomato seed maturation reveals tissue-specific regulatory networks and hubs associated with the acquisition of desiccation tolerance and seed vigour. *BMC plant biology,* 21**,** 1-23.

BUIJS, G., KODDE, J., GROOT, S. P. & BENTSINK, L. 2018. Seed dormancy release accelerated by elevated partial pressure of oxygen is associated with DOG loci. *Journal of experimental botany,* 69**,** 3601-3608.

CAI, Z., TANG, Q., SONG, P., TIAN, E., YANG, J. & JIA, G. 2024. The m6A reader ECT8 is an abiotic stress sensor that accelerates mRNA decay in Arabidopsis. *The Plant Cell***,** koae149.

CANTÓ-PASTOR, A., MASON, G. A., BRADY, S. M. & PROVART, N. J. 2021. Arabidopsis bioinformatics: tools and strategies. *The Plant Journal,* 108**,** 1585-1596.

CHENG, B., FURTADO, A. & HENRY, R. J. 2018. The coffee bean transcriptome explains the accumulation of the major bean components through ripening. *Scientific Reports,* 8**,** 11414.

CHENG, C. Y., KRISHNAKUMAR, V., CHAN, A. P., THIBAUD-NISSEN, F., SCHOBEL, S. & TOWN, C. D. 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal,* 89**,** 789-804.

CLERKX, E. J., VRIES, H. B.-D., RUYS, G. J., GROOT, S. P. & KOORNNEEF, M. 2003. Characterization of green seed, an enhancer of abi3-1 in Arabidopsis that affects seed longevity. *Plant physiology,* 132**,** 1077-1084.

DEKKERS, B. J., COSTA, M. C. D., MAIA, J., BENTSINK, L., LIGTERINK, W. & HILHORST, H. W. 2015. Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. *Planta,* 241**,** 563-577.

DEKKERS, B. J., HE, H., HANSON, J., WILLEMS, L. A., JAMAR, D. C., CUEFF, G., RAJJOU, L., HILHORST, H. W. & BENTSINK, L. 2016. The Arabidopsis DELAY OF GERMINATION 1 gene affects ABSCISIC ACID INSENSITIVE 5 (ABI 5) expression and genetically interacts with ABI 3 during Arabidopsis seed development. *The Plant Journal,* 85**,** 451-465.

DELMAS, F., SANKARANARAYANAN, S., DEB, S., WIDDUP, E., BOURNONVILLE, C., BOLLIER, N., NORTHEY, J. G., MCCOURT, P. & SAMUEL, M. A. 2013. ABI3 controls embryo degreening through Mendel's I locus. *Proceedings of the National Academy of Sciences,* 110**,** E3888-E3894.

DELOUCHE, J. C. & BASKIN, C. C. 2021. Accelerated aging techniques for predicting the relative storability of seed lots.

DONG, C.-J. & LIU, J.-Y. 2010. The Arabidopsis EAR-motif-containing protein RAP2. 1 functions as an active transcriptional repressor to keep stress responses under tight control. *BMC plant biology,* 10**,** 1-15.

FINCH-SAVAGE, W. E. & LEUBNER-METZGER, G. 2006. Seed dormancy and the control of germination. *New phytologist,* 171**,** 501-523.

FINKELSTEIN, R., REEVES, W., ARIIZUMI, T. & STEBER, C. 2008. Molecular aspects of seed dormancy. *Annu. Rev. Plant Biol.,* 59**,** 387-415.

FLOKOVÁ, K., TARKOWSKÁ, D., MIERSCH, O., STRNAD, M., WASTERNACK, C. & NOVÁK, O. 2014. UHPLC–MS/MS based target profiling of stress-induced phytohormones. *Phytochemistry,* 105**,** 147-157.

FRAZEE, A. C., PERTEA, G., JAFFE, A. E., LANGMEAD, B., SALZBERG, S. L. & LEEK, J. T. 2015. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology,* 33**,** 243-246.

GARG, R., SINGH, V. K., RAJKUMAR, M. S., KUMAR, V. & JAIN, M. 2017. Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *The Plant Journal,* 91**,** 1088-1107.

GINSAWAENG, O., HEISE, C., SANGWAN, R., KARCHER, D., HERNÁNDEZ-SÁNCHEZ, I. E., SAMPATHKUMAR, A. & ZUTHER, E. 2021. Subcellular localization of seed-expressed LEA_4 proteins reveals liquid-liquid phase separation for LEA9 and for LEA48 homo-and LEA42-LEA48 heterodimers. *Biomolecules,* 11**,** 1770.

GIORGINO, T. 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. *Journal of statistical Software,* 31**,** 1-24.

GIRAUDAT, J., HAUGE, B. M., VALON, C., SMALLE, J., PARCY, F. & GOODMAN, H. M. 1992. Isolation of the Arabidopsis ABI3 gene by positional cloning. *The Plant Cell,* 4**,** 1251-1261.

GOMES, M. S. O., SINNECKER, P., TANAKA, R. T. & LANFER-MARQUEZ, U. M. 2003. Effect of harvesting and drying conditions on chlorophyll levels of soybean (Glycine max L. Merr). *Journal of agricultural and food chemistry,* 51**,** 1634-1639.

GONZÁLEZ-MORALES, S. I., CHÁVEZ-MONTES, R. A., HAYANO-KANASHIRO, C., ALEJO-JACUINDE, G., RICO-CAMBRON, T. Y., DE FOLTER, S. & HERRERA-ESTRELLA, L. 2016. Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences,* 113**,** E5232-E5241.

GREEN, B. R., SINGH, S., BABIC, I., BLADEN, C. & JOHNSON-FLANAGAN, A. M. 1998. Relationship of chlorophyll, seed moisture and ABA levels in the maturing Brassica napus seed and effect of a mild freezing stress. *Physiologia Plantarum,* 104**,** 125-133.

GROOT, S., SURKI, A., DE VOS, R. & KODDE, J. 2012. Seed storage at elevated partial pressure of oxygen, a fast method for analysing seed ageing under dry conditions. *Annals of botany,* 110**,** 1149-1159.

GROOT, S. P. & KARSSEN, C. M. 1992. Dormancy and germination of abscisic acid-deficient tomato seeds: studies with the sitiens mutant. *Plant physiology,* 99**,** 952-958.

GUO, X., YAN, N., LIU, L., YIN, X., CHEN, Y., ZHANG, Y., WANG, J., CAO, G., FAN, C. & HU, Z. 2023. Transcriptomic comparison of seeds and silique walls from two rapeseed genotypes with contrasting seed oil content. *Frontiers in Plant Science,* 13**,** 1082466.

HE, H., DE SOUZA VIDIGAL, D., SNOEK, L. B., SCHNABEL, S., NIJVEEN, H., HILHORST, H. & BENTSINK, L. 2014. Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *Journal of experimental botany,* 65**,** 6603-6615.

HOLDSWORTH, M. J., BENTSINK, L. & SOPPE, W. J. 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New phytologist,* 179**,** 33-54.

HUNDERTMARK, M. & HINCHA, D. K. 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC genomics,* 9**,** 1-22.

JALINK, H., VAN DER SCHOOR, R., FRANDAS, A., VAN PIJLEN, J. G. & BINO, R. J. 1998. Chlorophyll fluorescence of Brassica oleracea seeds as a non-destructive marker for seed maturity and seed performance. *Seed Science Research,* 8**,** 437-443.

JOOSEN, R. V., KODDE, J., WILLEMS, L. A., LIGTERINK, W., VAN DER PLAS, L. H. & HILHORST, H. W. 2010. GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. *The Plant Journal,* 62**,** 148-159.

KANNO, Y., JIKUMARU, Y., HANADA, A., NAMBARA, E., ABRAMS, S. R., KAMIYA, Y. & SEO, M. 2010. Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. *Plant and Cell Physiology,* 51**,** 1988-2001.

KARSSEN, C., BRINKHORST-VAN DER SWAN, D., BREEKLAND, A. & KOORNNEEF, M. 1983. Induction of dormancy during seed development by endogenous abscisic acid: studies on abscisic acid deficient genotypes of Arabidopsis thaliana (L.) Heynh. *Planta,* 157**,** 158-165.

KASTELEIN, P., KRIJGER, M., CZAJKOWSKI, R., VAN DER ZOUWEN, P., VAN DER SCHOOR, R., JALINK, H. & VAN DER WOLF, J. 2014. Development of Xanthomonas fragariae populations and disease progression in strawberry plants after spray-inoculation of leaves. *Plant pathology,* 63**,** 255-263.

KERMODE, A. R. 1990. Regulatory mechanisms involved in the transition from seed development to germination. *Critical Reviews in Plant Sciences,* 9**,** 155-195.

KERMODE, A. R. & BEWLEY, J. D. 1985. The role of maturation drying in the transition from seed development to germination: I. acquisition of desiccation–tolerance and germinability during development of Ricinus communis L. seeds. *Journal of experimental botany,* 36**,** 1906-1915.

KERMODE, A. R., BEWLEY, J. D., DASGUPTA, J. & MISRA, S. 1986. The transition from seed development to germination: a key role for desiccation?

KIM, D., LANGMEAD, B. & SALZBERG, S. L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods,* 12**,** 357-360.

KOORNNEEF, M., HANHART, C. J., HILHORST, H. W. & KARSSEN, C. M. 1989. In vivo inhibition of seed development and reserve protein accumulation in recombinants of abscisic acid biosynthesis and responsiveness mutants in Arabidopsis thaliana. *Plant physiology,* 90**,** 463-469.

KOORNNEEF, M., JORNA, M., BRINKHORST-VAN DER SWAN, D. & KARSSEN, C. 1982. The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in non-germinating gibberellin sensitive lines of Arabidopsis thaliana (L.) heynh. *Theoretical and applied genetics,* 61**,** 385-393.

KOORNNEEF, M., REULING, G. & KARSSEN, C. 1984. The isolation and characterization of abscisic acid-insensitive mutants of Arabidopsis thaliana. *Physiologia Plantarum,* 61**,** 377-383.

KOZAKI, A. & AOYANAGI, T. 2022. Molecular aspects of seed development controlled by gibberellins and abscisic acids. *International Journal of Molecular Sciences,* 23**,** 1876.

KULKARNI, S. R., VANEECHOUTTE, D., VAN DE VELDE, J. & VANDEPOELE, K. 2018. TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information. *Nucleic acids research,* 46**,** e31-e31.

KUSHIRO, T., OKAMOTO, M., NAKABAYASHI, K., YAMAGISHI, K., KITAMURA, S., ASAMI, T., HIRAI, N., KOSHIBA, T., KAMIYA, Y. & NAMBARA, E. 2004. The Arabidopsis cytochrome P450 CYP707A encodes ABA 8′-hydroxylases: key enzymes in ABA catabolism. *The EMBO journal,* 23**,** 1647-1656.

LABUN, K., MONTAGUE, T. G., KRAUSE, M., TORRES CLEUREN, Y. N., TJELDNES, H. & VALEN, E. 2019. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic acids research,* 47**,** W171-W174.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics,* 9**,** 1-13.

LE, B. H., CHENG, C., BUI, A. Q., WAGMAISTER, J. A., HENRY, K. F., PELLETIER, J., KWONG, L., BELMONTE, M., KIRKBRIDE, R. & HORVATH, S. 2010. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences,* 107**,** 8063-8070.

LEFEBVRE, V., NORTH, H., FREY, A., SOTTA, B., SEO, M., OKAMOTO, M., NAMBARA, E. & MARION-POLL, A. 2006. Functional analysis of Arabidopsis NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *The Plant Journal,* 45**,** 309-319.

LÉON-KLOOSTERZIEL, K. M., VAN DE BUNT, G. A., ZEEVAART, J. A. & KOORNNEEF, M. 1996. Arabidopsis mutants with a reduced seed dormancy. *Plant Physiology,* 110**,** 233-240.

LÉON-KLOOSTERZIEL, K. M., GIL, M. A., RUIJS, G. J., JACOBSEN, S. E., OLSZEWSKI, N. E., SCHWARTZ, S. H., ZEEVAART, J. A. & KOORNNEEF, M. 1996. Isolation and characterization of abscisic acid-deficient Arabidopsis mutants at two new loci. *The Plant Journal,* 10**,** 655-661.

LEPRINCE, O. & BUITINK, J. 2010. Desiccation tolerance: from genomics to the field. *Plant Science,* 179**,** 554-564.

LEPRINCE, O., PELLIZZARO, A., BERRIRI, S. & BUITINK, J. 2017. Late seed maturation: drying without dying. *Journal of experimental botany,* 68**,** 827-841.

LI, Z., WU, S., CHEN, J., WANG, X., GAO, J., REN, G. & KUAI, B. 2017. NYEs/SGRs-mediated chlorophyll degradation is critical for detoxification during seed maturation in Arabidopsis.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology,* 15**,** 1-21.

MANSFIELD, S., BRIARTY, L. & ERNI, S. 1991. Early embryogenesis in Arabidopsis thaliana. I. The mature embryo sac. *Canadian Journal of Botany,* 69**,** 447-460.

MEURS, C., BASRA, A. S., KARSSEN, C. M. & VAN LOON, L. C. 1992. Role of abscisic acid in the induction of desiccation tolerance in developing seeds of Arabidopsis thaliana. *Plant physiology,* 98**,** 1484-1493.

MIZOI, J., SHINOZAKI, K. & YAMAGUCHI-SHINOZAKI, K. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1819**,** 86-96.

MÖNKE, G., SEIFERT, M., KEILWAGEN, J., MOHR, M., GROSSE, I., HÄHNEL, U., JUNKER, A., WEISSHAAR, B., CONRAD, U. & BÄUMLEIN, H. 2012. Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic acids research,* 40**,** 8240-8254.

NAKABAYASHI, K., BARTSCH, M., XIANG, Y., MIATTON, E., PELLENGAHR, S., YANO, R., SEO, M. & SOPPE, W. J. 2012. The time required for dormancy release in Arabidopsis is determined by DELAY OF GERMINATION1 protein levels in freshly harvested seeds. *The Plant Cell,* 24**,** 2826-2838.

NAKAJIMA, S., ITO, H., TANAKA, R. & TANAKA, A. 2012. Chlorophyll b reductase plays an essential role in maturation and storability of Arabidopsis seeds. *Plant physiology,* 160**,** 261-273.

NAMBARA, E., KAWAIDE, H., KAMIYA, Y. & NAITO, S. 1998. Characterization of an Arabidopsis thaliana mutant that has a defect in ABA accumulation: ABA-dependent and ABA-independent accumulation of free amino acids during dehydration. *Plant and cell physiology,* 39**,** 853-858.

NAMBARA, E., KEITH, K., MCCOURT, P. & NAITO, S. 1994. Isolation of an internal deletion mutant of the Arabidopsis thaliana ABI3 gene. *Plant and cell physiology,* 35**,** 509-513.

NAMBARA, E., KEITH, K., MCCOURT, P. & NAITO, S. 1995. A regulatory role for the ABI3 gene in the establishment of embryo maturation in Arabidopsis thaliana. *Development,* 121**,** 629-636.

NAMBARA, E. & MARION-POLL, A. 2003. ABA action and interactions in seeds. *Trends in plant science,* 8**,** 213-217.

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**,** 435-441.

NAMBARA, E., SUZUKI, M., ABRAMS, S., MCCARTY, D. R., KAMIYA, Y. & MCCOURT, P. 2002. A screen for genes that function in abscisic acid signaling in Arabidopsis thaliana. *Genetics,* 161**,** 1247-1255.

NÉE, G., KRAMER, K., NAKABAYASHI, K., YUAN, B., XIANG, Y., MIATTON, E., FINKEMEIER, I. & SOPPE, W. J. 2017. DELAY OF GERMINATION1 requires PP2C phosphatases of the ABA signalling pathway to control seed dormancy. *Nature communications,* 8**,** 72.

NEFF, M. M., TURK, E. & KALISHMAN, M. 2002. Web-based primer design for single nucleotide polymorphism analysis. *TRENDS in Genetics,* 18**,** 613-615.

NIU, L., DU, C., WANG, W., ZHANG, M., WANG, W., LIU, H., ZHANG, J. & WU, X. 2022. Transcriptome and co-expression network analyses of key genes and pathways associated with differential abscisic acid accumulation during maize seed maturation. *BMC Plant Biology,* 22**,** 359.

NONOGAKI, M., YAMAZAKI, S., NISHIYAMA, E., OHSHIMA, K. & NONOGAKI, H. 2022. Seed traits and phylogenomics: prospects for the 21st century. *Seed Science Research,* 32**,** 137-143.

OKAMOTO, M., KUWAHARA, A., SEO, M., KUSHIRO, T., ASAMI, T., HIRAI, N., KAMIYA, Y., KOSHIBA, T. & NAMBARA, E. 2006. CYP707A1 and CYP707A2, which encode abscisic acid 8′-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. *Plant physiology,* 141**,** 97-107.

OOMS, J. J., LEON-KLOOSTERZIEL, K. M., BARTELS, D., KOORNNEEF, M. & KARSSEN, C. M. 1993. Acquisition of desiccation tolerance and longevity in seeds of Arabidopsis thaliana (a comparative study using abscisic acid-insensitive abi3 mutants). *Plant physiology,* 102**,** 1185-1191.

OOMS, J. J., VAN DER VEEN, R. & KARSSEN, C. M. 1994. Abscisic acid and osmotic stress or slow drying independently induce desiccation tolerance in mutant seeds of Arabidopsis thaliana. *Physiologia Plantarum,* 92**,** 506-510.

PEREIRA LIMA, J. J., BUITINK, J., LALANNE, D., ROSSI, R. F., PELLETIER, S., DA SILVA, E. A. A. & LEPRINCE, O. 2017. Molecular characterization of the acquisition of longevity during seed maturation in soybean. *PloS one,* 12**,** e0180282.

PERTEA, M., PERTEA, G. M., ANTONESCU, C. M., CHANG, T.-C., MENDELL, J. T. & SALZBERG, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology,* 33**,** 290-295.

POWELL, A. A. & MATTHEWS, S. 2005. Towards the validation of the controlled deterioration vigour test for small seeded vegetables. *Seed Testing International,* 129**,** 21-24.

PRASCH, C. M. & SONNEWALD, U. 2013. Simultaneous application of heat, drought, and virus to Arabidopsis plants reveals significant shifts in signaling networks. *Plant physiology,* 162**,** 1849-1866.

PRITCHARD, H. W. 2020. Diversity in seed longevity amongst biodiverse seeds. *Seed Science Research,* 30**,** 75-80.

QI, Z., ZHANG, Z., WANG, Z., YU, J., QIN, H., MAO, X., JIANG, H., XIN, D., YIN, Z. & ZHU, R. 2018. Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development. *Plant, Cell & Environment,* 41**,** 2109-2127.

RIGHETTI, K., VU, J. L., PELLETIER, S., VU, B. L., GLAAB, E., LALANNE, D., PASHA, A., PATEL, R. V., PROVART, N. J. & VERDIER, 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**,** 2692-2708.

SAJEEV, N., KOORNNEEF, M. & BENTSINK, L. 2024. A commitment for life: Decades of unraveling the molecular mechanisms behind seed dormancy and germination. *The Plant Cell,* 36**,** 1358-1376.

SANO, N., RAJJOU, L., NORTH, H. M., DEBEAUJON, I., MARION-POLL, A. & SEO, M. 2016. Staying alive: molecular aspects of seed longevity. *Plant and Cell Physiology,* 57**,** 660-674.

SANTOS-MENDOZA, M., DUBREUCQ, B., BAUD, S., PARCY, F., CABOCHE, M. & LEPINIEC, L. 2008. Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *The Plant Journal,* 54**,** 608-620.

SCHALK, C., DREVENSEK, S., KRAMDI, A., KASSAM, M., AHMED, I., COGNAT, V., GRAINDORGE, S., BERGDOLL, M., BAUMBERGER, N. & HEINTZ, D. 2016. DNA damage binding protein2 shapes the DNA Methylation Landscape. *The Plant Cell,* 28**,** 2043-2059.

SEKHON, R. S., HIRSCH, C. N., CHILDS, K. L., BREITZMAN, M. W., KELL, P., DUVICK, S., SPALDING, E. P., BUELL, C. R., DE LEON, N. & KAEPPLER, S. M. 2014. Phenotypic and transcriptional analysis of divergently selected maize populations reveals the role of developmental timing in seed size determination. *Plant physiology,* 165**,** 658-669.

SERIN, E. A., NIJVEEN, H., HILHORST, H. W. & LIGTERINK, W. 2016. Learning from co-expression networks: possibilities and challenges. *Frontiers in plant science,* 7**,** 444.

SINNECKER, P., BRAGA, N., MACCHIONE, E. L. & LANFER-MARQUEZ, U. M. 2005. Mechanism of soybean (Glycine max L. Merrill) degreening related to maturity stage and postharvest drying temperature. *Postharvest Biology and Technology,* 38**,** 269-279.

SONG, J., PEI, W., WANG, N., MA, J., XIN, Y., YANG, S., WANG, W., CHEN, Q., ZHANG, J. & YU, J. 2022. Transcriptome analysis and identification of genes associated with oil accumulation in upland cotton. *Physiologia Plantarum,* 174**,** e13701.

SONG, J., WANG, X., SONG, B., GAO, L., MO, X., YUE, L., YANG, H., LU, J., REN, G. & MO, B. 2019. Prevalent cytidylation and uridylation of precursor miRNAs in Arabidopsis. *Nature Plants,* 5**,** 1260-1272.

STUTTMANN, J., BARTHEL, K., MARTIN, P., ORDON, J., ERICKSON, J. L., HERR, R., FERIK, F., KRETSCHMER, C., BERNER, T. & KEILWAGEN, J. 2021. Highly efficient multiplex editing: one-shot generation of 8× Nicotiana benthamiana and 12× Arabidopsis mutants. *The Plant Journal,* 106**,** 8-22.

SUGLIANI, M., RAJJOU, L., CLERKX, E. J., KOORNNEEF, M. & SOPPE, W. J. 2009. Natural modifiers of seed longevity in the Arabidopsis mutants abscisic acid insensitive3-5 (abi3-5) and leafy cotyledon1-3 (lec1-3). *New Phytologist,* 184**,** 898-908.

TEIXEIRA, R. N., LIGTERINK, W., FRANÇA-NETO, J. D. B., HILHORST, H. W. & DA SILVA, E. A. 2016. Gene expression profiling of the green seed problem in soybean. *BMC Plant Biology,* 16**,** 1-15.

TERRASSON, E., BUITINK, J., RIGHETTI, K., LY VU, B., PELLETIER, S., ZINSMEISTER, J., LALANNE, D. & LEPRINCE, O. 2013. An emerging picture of the seed desiccome: confirmed regulators and newcomers identified using transcriptome comparison. *Frontiers in Plant Science,* 4**,** 497.

TIAN, R., WANG, F., ZHENG, Q., NIZA, V. M., DOWNIE, A. B. & PERRY, S. E. 2020. Direct and indirect targets of the arabidopsis seed transcription factor ABSCISIC ACID INSENSITIVE3. *The Plant Journal,* 103**,** 1679-1694.

TISSIER, A. F., MARILLONNET, S., KLIMYUK, V., PATEL, K., TORRES, M. A., MURPHY, G. & JONES, J. D. 1999. Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. *The Plant Cell,* 11**,** 1841-1852.

VERDIER, J., TORRES-JEREZ, I., WANG, M., ANDRIANKAJA, A., ALLEN, S. N., HE, J., TANG, Y., MURRAY, J. D. & UDVARDI, M. K. 2013. Establishment of the L otus japonicus Gene Expression Atlas (L j GEA) and its use to explore legume seed maturation. *The Plant Journal,* 74**,** 351-362.

VERMA, S., ATTULURI, V. P. S. & ROBERT, H. S. 2022. Transcriptional control of Arabidopsis seed development. *Planta,* 255**,** 90.

WAN, C.-Y. & WILKINS, T. A. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (Gossypium hirsutum L.). *Analytical biochemistry,* 223**,** 7-12.

WANG, Y.-C., HSIEH, W.-H., HUANG, C.-J., JHAN, Y.-T., ZHAN, J., CHANG, C.-C., HSIEH, T.-F. & LIN, J.-Y. 2023. Temporal Transcriptome Analysis Uncovers Regulatory Modules Programming Embryo Development from Embryonic Morphogenesis to Post-Germination. *bioRxiv***,** 2023.12. 26.573376.

YAN, A. & CHEN, Z. 2017. The pivotal role of abscisic acid signaling during transition from seed maturation to germination. *Plant Cell Reports,* 36**,** 689-703.

YAO, Y., XIONG, E., QU, X., LI, J., LIU, H., QUAN, L., LU, W., ZHU, X., CHEN, M. & LI, K. 2023. WGCNA and transcriptome profiling reveal hub genes for key development stage seed size/oil content between wild and cultivated soybean. *BMC genomics,* 24**,** 494.

ZHANG, Y., LI, Y., DE ZEEUW, T., DUIJTS, K., KAWA, D., LAMERS, J., MUNZERT, K. S., LI, H., ZOU, Y. & MEYER, A. J. 2024. Root branching under high salinity requires auxin-independent modulation of LATERAL ORGAN BOUNDARY DOMAIN 16 function. *The Plant Cell,* 36**,** 899-918.

ZHANG, Z., LI, W., GAO, X., XU, M. & GUO, Y. 2020. DEAR4, a member of DREB/CBF family, positively regulates leaf senescence and response to multiple stressors in Arabidopsis thaliana. *Frontiers in Plant Science,* 11**,** 367.

ZHENG, Y., LI, Y.-F., SUNKAR, R. & ZHANG, W. 2012. SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic acids research,* 40**,** e28-e28.

ZHOU, X., WANG, G. & ZHANG, W. 2007. UV-B responsive microRNA genes in Arabidopsis thaliana. *Molecular systems biology,* 3**,** 103.

ZINSMEISTER, J., BERRIRI, S., BASSO, D. P., LY-VU, B., DANG, T. T., LALANNE, D., DA SILVA, E. A. A., LEPRINCE, O. & BUITINK, J. 2020a. The seed-specific heat shock factor A9 regulates the depth of dormancy in Medicago truncatula seeds via ABA signalling. *Plant, cell & environment,* 43**,** 2508-2522.

ZINSMEISTER, J., LALANNE, D., TERRASSON, E., CHATELAIN, E., VANDECASTEELE, C., VU, B. L., DUBOIS-LAURENT, C., GEOFFRIAU, E., SIGNOR, C. L. & DALMAIS, M. 2016. ABI5 is a regulator of seed maturation and longevity in legumes. *The Plant Cell,* 28**,** 2735-2754.

ZINSMEISTER, J., LEPRINCE, O. & BUITINK, J. 2020b. Molecular and environmental factors regulating seed longevity. *Biochemical Journal,* 477**,** 305-323.