Document downloaded from:

http://hdl.handle.net/10251/176441

This paper must be cited as:

Renard, J.; Niñoles Rodenes, R.; Martínez-Almonacid, I.; Gayubas, B.; Mateos-Fernández, R.; Bissoli, G.; Bueso Rodenas, E.... (2020). Identification of novel seed longevity genes related to oxidative stress and seed coat by genome wide association studies and reverse genetics. Plant Cell & Environment. 43(10):2523-2539. https://doi.org/10.1111/pce.13822

The final publication is available at

https://doi.org/10.1111/pce.13822

Copyright Blackwell Publishing

Additional Information

Identification of novel seed longevity genes related to oxidative stress and seed coat by genome-wide association studies and reverse genetics

Joan Renard, Regina Niñoles, Irene Martínez-Almonacid, Beatriz Gayubas, Rubén Mateos-Fernández, Gaetano Bissoli, Eduardo Bueso, Ramón Serrano* and José Gadea*

Instituto de Biología Molecular y Celular de Plantas (IBMCP), Universitat Politècnica de València-C.S.I.C., Camino de Vera, 46022, Valencia, Spain

*corresponding authors (e-mail: jgadeav@ibmcp.upv.es and rserrano@ibmcp.upv.es)

This work was funded by grant BIO2017-88898-P from the "Ministerio de Ciencia, Innovación y Universidades", Madrid, Spain.

ABSTRACT

Seed longevity is a polygenic trait of relevance for agriculture and for understanding the effect of environment on the ageing of biological systems. In order to identify novel longevity genes we have phenotyped the natural variation of 270 ecotypes of the model plant *Arabidopsis thaliana* for natural ageing and for three accelerated ageing methods. Genome-wide analysis, using publicly available single-nucleotide polymorphisms (SNPs) data sets, identified multiple genomic regions associated with variation in seed longevity. Reverse genetics of twenty candidate genes in Columbia ecotype resulted in seven genes positive for seed longevity (*PSAD1, SSLEA, SSTPR, DHAR1, CYP86A8, MYB47* and *SPCH*) and five negative ones (*RBOHD, RBOHE, RBOHF, KNAT7* and *SEP3*). In this uniform genetic background, natural and accelerated ageing methods provided similar results for seed-longevity in knock-out mutants. The NADPH oxidases (RBOHs), the dehydroascorbate reductase (DHAR1) and the photosystem I subunit (PSAD1) highlight the important role of oxidative stress on seed ageing. The cytochrome P-450 hydroxylase CYP86A8 and the transcription factors MYB47, KNAT7 and SEP3 support the protecting role of the seed coat during seed ageing.

Keywords: Seed longevity, natural variation, reverse genetics, *Arabidopsis thaliana*, seed ageing, accelerated ageing, oxidative stress, NADPH oxidases, seed coat, CYP86A8

INTRODUCTION

Imperfections of metabolism result in accumulation of deleterious and damaged molecules that age cells and organisms (Gladyshev, 2014; López-Otín *et al.*, 2016). According to the oxidative stress theory of ageing, a major metabolic reaction contributing to ageing is the production of reactive oxygen species (ROS) (Harman, 1956). Current views (López-Otín *et al.*, 2016) consider that ageing is controlled by metabolism in a more general way than just ROS production. However, it has been argued that oxidative stress contributes to all the metabolic and genomic alterations associated with ageing (Schmidlin *et al.*, 2019).

One model for the study of biological ageing is the seed of vascular plants. Ageing is slowed in desiccated forms of organisms, where metabolism is reduced because of high viscosity and slow diffusion inside cells. Accordingly, desiccation is the mechanism developed by many life forms to decrease metabolism and endure for long times (Wood and Jenks, 2008). Most organisms die upon desiccation and therefore acquisition of desiccation tolerance is a prerequisite for this longevity mechanism. In the case of plants, it has been suggested that the evolution of desiccation tolerance was a crucial step in the colonization of land by primitive plants living in fresh water (Oliver *et al.*, 2000). Nowadays, very few plants (the so called resurrection plants) are desiccation tolerant in their vegetative form. However, orthodox seeds are tolerant to desiccation, a feature acquired during their late maturation phase. This tolerance results from the accumulation of osmolytes (sucrose and raffinose family oligosaccharides, RFO), late embryogenesis abundant proteins (LEA) and small heat shock proteins (Sano *et al*., 2016; Leprince *et al.*, 2017). Under these conditions cytoplasm enters an immobilized glassy state that protects macromolecules and cellular membranes during desiccation (Hoekstra *et al.*, 2001; Ballesteros and Walters, 2011; Sano *et al.*, 2016; Leprince *et al.*, 2017). This desiccated state of seeds permits to keep the plant alive for long periods of time, until environmental conditions are favourable. However, dry seeds also experience deterioration and ageing and it has been proposed that oxidative damage is a major cause reducing longevity in this system (Harman and Mattick, 1976; Bailly, 2004; Sano et al., 2016; Nagel et al., 2019).

High temperature and humidity are deleterious for seeds because they break the glassy state and promote metabolism, and therefore ROS and cellular damage. Equilibrating mature seeds between 19 and 27% relative environmental humidity (0.04-

0.06 g H_2O/g dry weight in seeds) provides the optimum moisture level for reducing metabolism and maintaining seed longevity during long-term storage (Vertucci and Roos, 1990). Seed deterioration occurs under ambient storage or natural conditions (about 20-25 ºC, 40-60 % relative humidity (RH), 0.08-0.10 g H_2O/g dry weight), where dried seeds still exhibit some metabolism.

Analysis of Quantitative Trait Loci (QTLs) has demonstrated the polygenic nature of seed longevity in different plants such as rice (Jiang *et al.*, 2011), wheat (Landjeva *et al.*, 2010), lettuce (Schwember *et al.*, 2010), tobacco (Agacka-Mołdoch *et al.*, 2015) and barley (Nagel *et al.*, 2016). Several genes important for seed longevity in the model plant *Arabidopsis thaliana* have been identified by reverse genetics and compiled by Sano *et al.* (2016). They refer to three biological functions: anti-oxidative defence, repair of damaged proteins and nucleic acids, and development of seed coat. Some important aspects that need to be clarified include the nature of the ROS producing systems and of antioxidant defences during seed storage, the protective mechanisms of the seed coat and the complete gene networks controlling seed longevity. Many antioxidant defences such as enzymes and chemicals have been proposed to participate in seed longevity but only vitamin E (tocopherol) has been demonstrated to be important by mutant studies (Sattler *et al*., 2004). Also, the seed coat has been proposed to be important, specially the endothelium flavonoids (proanthocyanidins; Clerkx *et al*. 2004) and the palisade suberin (Bueso *et al*., 2014 and 2016, Renard *et al*., 2020) but the transcription factors and enzymes involved in the development of this structure (Haughn and Chaudhury, 2005) are not completely known.

In the present work, we took advantage of natural variation in seed longevity between accessions (ecotypes) of *Arabidopsis thaliana* with well-characterized genomes to perform a Genome Wide Association Study (GWAS) of this trait. GWAS analysis is a powerful tool to dissect polygenic traits (Atwell *et al*., 2010; Korte and Farlow, 2013) and Arabidopsis is an excellent organism for this purpose (Horton *et al.*, 2012; Weigel, 2012). GWAS consists on the association of the molecular genetic information (Single Nucleotide Polymorphisms, SNPs) with the observed variation in a given trait through statistical computation. These associations point to causal loci that may be involved in the trait and may explain part of the natural variation between accessions (Curtin *et al.*, 2017). However, for each candidate gene in these loci empirical validation is required due to the high number of false positives with this genomic approach (Ioannidis *et al.*, 2009;

Korte and Farlow, 2013). *Arabidopsis* is the most studied plant species at the molecular level and a large collection of sequence-indexed T-DNA insertion mutants is available (O'Malley and Ecker, 2010). This valuable resource allows validation of GWAS candidate genes by testing the phenotype of the corresponding loss-of-function mutants in the same cultivar background.

As the study of seed longevity under natural, ambient storage conditions requires long times for ageing, several accelerated ageing procedures have been utilized in this field. The quickest is the accelerated ageing test (AAT), which consists on a high temperature treatment of water imbibed seeds during days (Prieto-Dapena *et al.*, 2006; Bueso *et al.*, 2014). In the controlled deterioration test (CDT) ageing of dry seeds occurs in weeks with a high humidity and high temperature atmosphere. Finally, to target the oxygen-produced damage, there is the elevated partial pressure of oxygen test (EPPO), in which dry seeds are stored for months in a low humidity and high-pressured oxygen atmosphere (ISTA, 2018; Hay *et al.*, 2019; Groot *et al.*, 2012). Although in some studies with few genotypes there is a similar behaviour in both natural and artificial ageing (Rajjou *et al.*, 2008; Bueso *et al.*, 2014), each ageing procedure might affect different physiological and molecular aspects of the seed deterioration-resistance trait. Accordingly, to increase the list of candidate genes, we have performed natural ageing and different artificial ageing treatments.

Our results suggest that the different tests used to evaluate seed longevity are highlighting different aspects of the machinery deployed by the seed to avoid deterioration. However, when seed longevity was determined in mutants from the same genetic background, natural and artificial ageing procedures gave similar results. Validation of candidate genes resulted in several novel participants in seed longevity such as ROS-producing NADPH oxidases (RBOHs), antioxidant dehydroascorbate reductase and, related to seed coat development, a P450 cytochrome hydroxylase and several transcription factors (TFs).

MATERIALS AND METHODS

Plant material and growth conditions

A core set of 360 Arabidopsis thaliana ecotypes was obtained from the Nottingham Arabidopsis Stock Centre (NASC) (reference N76309). Seven plants of each ecotype were grown in 11 cm diameter pots containing a 1:2 vermiculite:soil mixture under growth chamber conditions (16 h light / 8 h dark, at 16-23 ºC and 70-75% RH). T-DNA insertion lines in Col-0 background were obtained from NASC or kindly supplied by original authors. T-DNA lines and other mutants used in this study are listed in Table S 1. Homozygous plants were obtained and confirmed by PCR. Some mutants were previously described: *dhar1*, *dhar2, dhar3* and *dhar1,2,3* (Rahantaniaina *et al.,* 2017), *rbohe* (Xie *et al.*, 2014), *rbohd,f* (Torres *et al.*, 2002), *spch* (MacAlister *et al.*, 2007), *knat7* (Romano *et al.*, 2012), *sep3* (Pelaz *et al.*, 2000) and *sep1,2,4* (Ditta *et al.*, 2004). Control and mutant plants were grown simultaneously and seeds were collected and stored under same conditions.

Artificial and natural ageing assays

Ecotype seed sterilization consisted on eight hours chlorine gas treatment. One week of stratification (4 ºC, 100% RH) was used to ensure complete dormancy break. A control germination assay was used to discard ecotypes with affected germination. Selected ecotypes underwent three different artificial ageing treatments, whose specific conditions were prior adjusted with the Col-0 ecotype. The Accelerated Ageing Treatment (AAT) consisted in imbibing seeds in water and maintain at 39ºC for 48h (Prieto-Dapena *et al.*, 2006). The Controlled Deterioration Treatment (CDT) was set with 14 days at 38 ºC in a 75% RH atmosphere (Righetti *et al,*. 2015). The Elevated Partial Pressure of Oxygen (EPPO) was performed during 5 months at 5 bar O_2 with a 40% RH (Groot *et al.*, 2012 and Nagel *et al,*. 2016). Natural seed Ageing Treatment (NAT) which consisted in dry seed storage at room temperature (20-25 ºC, 40-60% RH) for 18 months. Except in AAT, all ageing treatments were performed prior to seed sterilization and stratification.

Arabidopsis ecotypes were sown on 25 mL Murashige and Skoog (MS) plates with sucrose as described by Alejandro *et al.* (2007). Each plate was planted with 125 seeds on average. Germination ratios were scored after seven days. Radicle extrusion and

seedling establishment with green cotyledons were used as germination indication. The number of established seedlings under our experimental conditions did not increase after 1 week. The corrected germination ratio was calculated dividing the germination ratio in each treatment by the germination ratio in control conditions.

T-DNA and transposon homozygous lines were tested with natural (NAT) and artificial ageing treatments (AAT and CDT) similarly to natural ecotypes. Ethanol and bleach sterilization were used. The NAT time was adjusted to obtain maximum differences with Col-0 seeds. Mutant lines were tested three times in two different seed generations in all different assays. Significant differences with control Col-0 were assessed by Student's t-test ($p < 0.01$).

GWAS analysis

All four GWAS were performed using the online application GWAPP (Seren *et al.*, 2012). Corrected germination ratios were used as GWAPP input. Only NAT data were transformed logarithmically, for Shapiro-wilk score and pseudo-heritability calculation. The dataset used for the GWAS was the Imputed Fullsequence Dataset (Horton *et al.*, 2012; Long *et al.*, 2013; Cao *et al.*, 2011; Gan *et al.*, 2011). GWAS results were filtered by Minor Allele Frequency (MAF) ≥ 0.05. Genes 1.5 kilobases (kb) up and downstream of significant SNPs were mapped using TAIR10 (The Arabidopsis Information Resource version 10). The significance threshold in this study was set at p value $\leq 10^{-5}$. The easyGWAS online tool (Grimm *et al.*, 2017) was used to calculate Spearman correlations (r_s) between our data and different public data. Strong direct correlations were considered if $r_s \geq 0.3$, and strong inverse correlation if $r_s \leq -0.3$.

In silico gene expression analysis

Gene expression analysis was performed with the online tool eFP browser (Winter *et al.*, 2007). Developing seed gene-expression was obtained from Le *et al.* (2010) data. Subcellular location information from Hooper *et al.* (2017) data.

Seed analysis

Mucilage staining of dry seeds was performed as described in Bueso *et al.* (2014). Proantocyanidin (PA) staining in seeds was performed 5 days-after-pollination (DAP), as described in Debeaujon *et al.* (2000). Lipid polyester barriers staining was performed as described in Brundrett *et al.* (1991) and Beisson *et al.* (2007) in dry seeds. Tetrazolium permeability assays were performed as Debeaujon *et al.* (2000). Size calculation was done with Fiji program (Schindelin *et al.*, 2012).

Construction of transgenic plants

Primers used in this work are listed in Table S 2. For expression analysis, 2 kb upstream *MYB47* coding region were amplified from genomic DNA with *proMYB47-F* and *proMYB47-R* primers and inserted into pCR8 plasmid by using the pCR™8/GW/TOPO® TA Cloning® Kit from Thermo Fisher Scientific (Waltham, Massachusetts, USA). LR reaction recombined our construct with the *pMDC107* plasmid (Curtis and Grossniklaus, 2003). For construction of over-expression lines, its coding sequence was amplified with *MYB47-F* and *MYB47-R* primers from cDNA, and cloned into pCR8 as before. The resulting plasmid was recombined with *pUB-DEST* plasmid (Grefen *et al.*, 2010) by LR reaction. Agrobacterium C58 containing final plasmids was used for plant transformation. T_3 homozygous plants and seeds were used to avoid genotype segregation.

Expression analysis

Total RNA of knock-out mutant *myb47* and OE mutants of *MYB47* was obtained from 7 day old seedlings with E.Z.N.A. Plant RNA Kit (Omega Bio-tek, Norcross, Georgia, USA). 2 µg RNA were reverse transcribed using the Maxima first-strand cDNA synthesis kit for RT-qPCR (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Quantitative (real-time) PCR (qRT-PCR) was performed in triplicate using an Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher Scientific) with PyroTaq EvaGreen qPCR Mix Plus (ROX) (Cultek S.L.U., Madrid, Spain) and *MYB47-qRT-F* and *MYB47-qRT-R* primers (Table S 2). Relative mRNA abundance to PP2AA3 was calculated using the comparative ΔCt method (Czechowski *et al.*, 2005).

Confocal microscopy

To visualize driven-promotor GFP expression a confocal ZEISS-LSM710 microscope was used. Laser used was Argon with an excitation lambda of 488 nm, and an emission spectrum of 500-550 nm was registered. Images were processed with Fiji program (Schindelin *et al.*, 2012).

RESULTS

CDT correlates best with natural ageing

To elucidate new genetic components involved in seed longevity, a core-set of 360 Arabidopsis ecotypes was obtained from NASC to perform GWAS. All accessions were grown on soil under long-day conditions, and mature seeds where collected from all accessions that were able to flower and develop seeds in a time-window not wider than 160 days. 270 natural accessions were selected for ageing treatments after the control (non-aged) germination test (Data S1). Four different ageing methods were performed: Natural Ageing Treatment (NAT) and three artificial ageing treatments (AAT, CDT and EPPO). Aged-seeds germination data were corrected using control germination data (Data S2). A wide range of variability was observed among all the accessions in each of the different assays (Figures 1a, b, c,d). The accession Col-0 (Columbia), used as reference to set up all treatments, varied from 40% to 60% of germination in all treatments.

To estimate how each artificial ageing treatment resembles natural ageing, Spearman correlations were calculated (Figure 1e). CDT is the artificial ageing treatment correlating best with NAT ($r_s = 0.68$), followed by EPPO ($r_s = 0.47$). AAT does not correlate with NAT in this study ($r_s = 0.11$). AAT is the quickest (days) artificial ageing method used and at variance with other treatments, it involves immersion of seeds into water.

We next calculated correlations with public datasets and ecotype collection data. Although the coincident ecotypes with our collection was small in most cases (48-83 ecotypes), an inverse correlation between seed longevity and flowering time, rosette leave number and plant life spam could be observed (Table 1). These three traits are linked and suggest the importance of early flowering for seed longevity. There is also an inverse

correlation of iron content in leaves with NAT, CDT and EPPO. No correlation was found with seed dormancy, latitude of collection point or anthocyanin content, among others. All other correlations with public datasets are listed in Data S3.

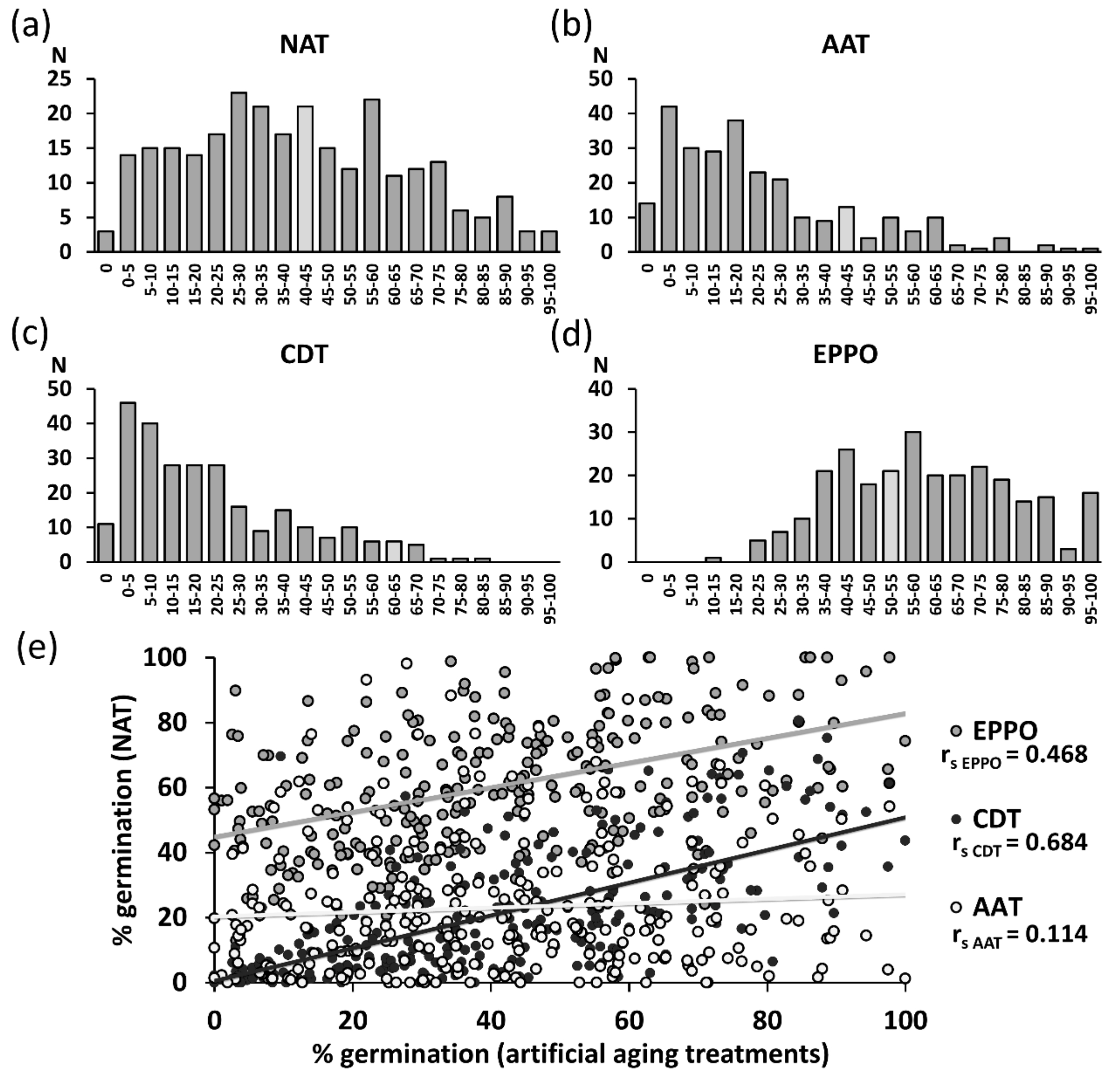


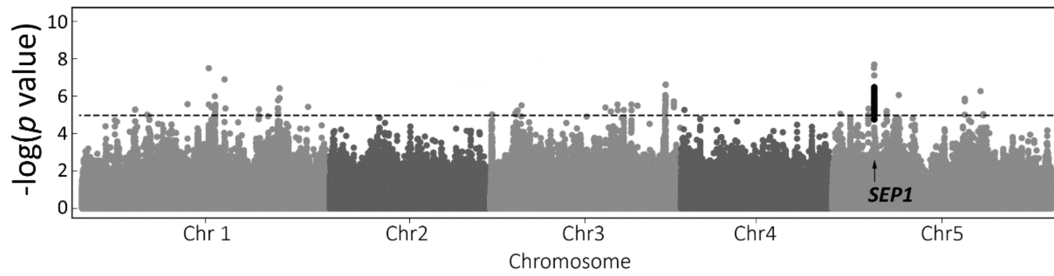
Figure 1. Arabidopsis ecotypes present a wide range of variability in seed deterioration resistance among all four ageing treatments. (a): Histogram distribution of 270 ecotypes after NAT. (b): AAT histogram. (c): CDT histogram. (d): EPPO histogram. Germination data of each treatment is corrected by control germination of each ecotype. Histograms represent the number of ecotypes (y-axis) in the indicated range of germination (x-axis). The bar in light grey corresponds to the range including the reference accession Col-0. (e): CDT correlates better to NAT. Dispersion plots of EPPO, CDT and AAT (y-axis) and NAT (x-axis). Regression lines are plotted and Spearman correlations are indicated for each artificial ageing correlation with NAT.

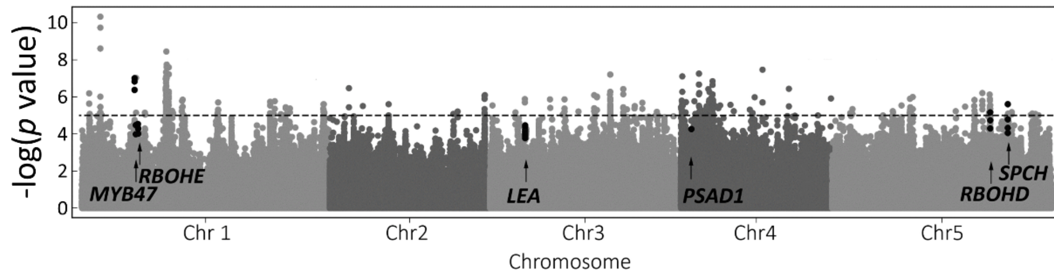
Table 1. Correlation indexes of different phenotypic traits with seed longevity determined with different ageing treatments (NAT, AAT, CDT and EPPO). Indicated correlation indexes correspond to Spearman correlations. Ecotypes column refers to the number of coinciding ecotypes in the correlations. Complete correlation information is listed in the Data S3. Upper part: correlated traits. Down part: non-correlated traits.

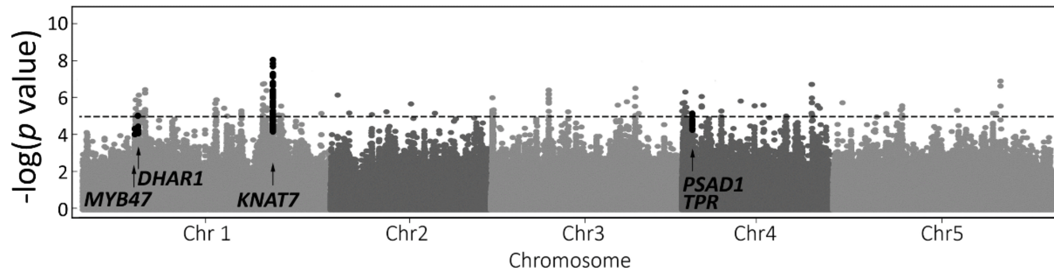
	NAT	AAT	CDT	EPPO	Ecotypes	Reference
Plant life spam	-0.35	-0.33	-0.30	-0.22	270	Data S 3
Days to flower	-0.28	-0.35	-0.29	-0.31	71	Grimm *et al.* (2017)
Rossete leaves number	-0.40	-0.37	-0.36	-0.32	71	Grimm *et al.* (2017)
Fe content in leaves	-0.32	0.14	-0.44	-0.30	51	Atwell *et al.* (2010)
Latitude at collection point	-0.11	0.00	0.03	0.10	256	Horton *et al.* (2012)
Seed dormancy	0.09	-0.28	0.00	-0.09	48	Atwell *et al.* (2010)
Anthocyanin content	0.13	-0.01	-0.07	-0.09	83	Atwell *et al.* (2010)

The different seed ageing treatments highlight multiple genomic areas and candidate genes.

The corrected germination data obtained after each ageing treatment (Data S2) was used to perform four different GWAS analysis, one for each seed ageing treatment. Results can be visualized at https://gwas.gmi.oeaw.ac.at/#/study/5406/overview. We established a relaxed *p* value ($< 10^{-5}$) threshold to consider putative longevity-related genomic areas (Figure 2). Genes enclosed within the 1.5kb-region spanning significant SNPs were listed (Data S 4). Natural ageing (NAT) highlighted 99 genes, the AAT treatment, 280 genes, the CDT, 130 genes and the EPPO, 337 genes. We further filtered for candidate genes following three criteria: (a) biological function based on the state-of-knowledge (Sano *et al.*, 2016); (b) gene expression during seed development (Le *et al.*, 2010; compiled in Figure Ss 1 and 2); and (c) distance to highly scored SNPs. Occasionally, a few genes with *p* value slightly above 10^{-5} were considered for reverse genetic analysis. In some cases, gene family members from candidate genes were also selected. For validation and further experiments, double and triple mutants were obtained. Table 2 shows the list of candidate genes selected for reverse genetic analysis in this study.







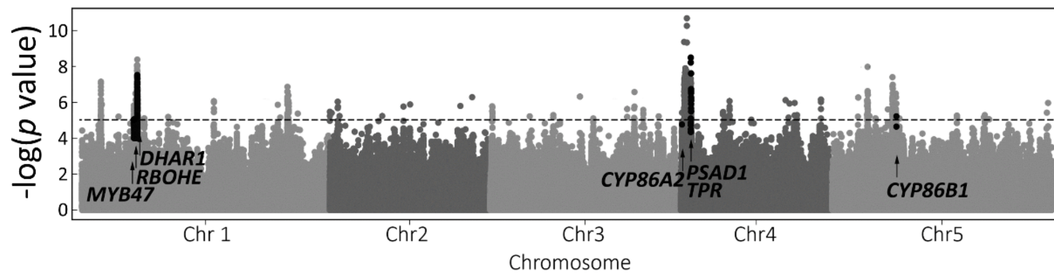


Figure 2. Different seed longevity GWAS highlight different and similar genomic areas. (a): NAT-GWAS Manhattan plot. (b): AAT-GWAS Manhattan plot. (c): CDT-GWAS Manhattan plot. (d): EPPO-GWAS Manhattan plot. Dashed lines indicate the threshold significance (p value $< 10^{-5}$). In black, SNPs showing p value$<10^{-4}$ that are at less than 1.5 kb distance of genes used for reverse genetics. Arrows associate gene alias with SNPs.

Table 2. Highlighted genes used for reverse genetics analysis. GWAS column: ageing treatments where the indicated genes appeared at less than 1.5 kb distance of SNPs with p value ≤ 10^{-4}; Score column or –log(p value): score of most significant SNP close to the indicated gene in first-listed GWAS; AGI column: Arabidopsis gene identifier; Gene alias: gene name used in this paper. Lower part of table correspond to genes whose SNPs were not significant in GWAS (score between 4 and 5), but were used for reverse genetics. Complete GWAS gene-scoring information is listed in Data S4.

GWAS	Score	AGI	Gene alias
EPPO, CDT	8.51	AT4G02750	*SSTPR*
EPPO, AAT, CDT	8.24	AT4G02770	*PSAD1*
CDT	8.14	AT1G62990	*KNAT7*
EPPO, CDT	7.49	AT1G19570	*DHAR1*
AAT, EPPO, CDT	6.99	AT1G18710	*MYB47*
NAT	6.26	AT5G15800	*SEP1*
AAT	5.63	AT5G53210	*SPCH*
EPPO	5.54	AT5G23190	*CYP86B1*
AAT	5.28	AT5G47910	*RBOHD*
EPPO, AAT	*4.85*	*AT1G19230*	*RBOHE*
EPPO	*4.75*	*AT4G00360*	*CYP86A2*
AAT	*4.41*	*AT3G17520*	*SSLEA*

T-DNA insertion mutants identify novel genes involved in seed longevity

Three different seed lots of two different plant generations from every mutant line were aged with two artificial treatments (AAT and CDT; Table 3) and with natural conditions (NAT). Figure 3 describes three seed ageing sensitive mutants: a Late Embryogenesis Abundant (LEA) protein (AT3G17520), a subunit of Photosystem I (PSAD1), and a tetratricopeptide repeat (TPR) protein (AT4G02750). The *LEA* and *TPR* genes were named *SSLEA* and *SSTPR*, with *SS* meaning Seed Storability. *PSAD1* and *SSTPR* were highlighted by two and three artificial ageing treatments, respectively (Table 2). After 18 months of natural ageing, all three mutant seeds present a drastic reduction of germination. *PSAD1* and *LEA* are expressed in mature embryos, while *TPR* is expressed in developing embryos (Figure S1; Le *et al.*, 2010). As indicated in Table 3, mutants in these three genes also exhibited reduced seed longevity with two artificial ageing

treatments (AAT and CDT). Mutant *sstpr* seeds present an increased tetrazolium reduction (Figure S3) indicating greater seed coat permeability.

Table 3. Schematic results of ageing treatments (NAT, AAT and CDT) of loss-of-function mutants in Col-0 background. Reduced seed germination (Susceptible), increased seed germination (Resistant) or non-significant changes in germination after the ageing treatment (n.s.) in comparison to Col-0 seeds. *Significantly differing from wild type seeds at $p < 0.05$ (Student´s t-test). **Significantly differing from wild type seeds at $p < 0.01$ (Student´s t-test). Supporting numerical and statistical data are supplied in Data S5.

Mutant line	NAT	AAT	CDT
psad1	Susceptible**	Susceptible*	Susceptible**
sslea	Susceptible**	Susceptible*	Susceptible*
sstpr	Susceptible**	Susceptible*	Susceptible*
rbohd	Resistant*	Resistant*	Resistant*
rbohe	Resistant*	Resistant*	Resistant**
rbohf	Resistant*	Resistant*	Resistant*
rbohd,f	Resistant**	n.s.	Resistant**
dhar1	Susceptible**	Susceptible*	Susceptible**
dhar2	n.s.	n.s.	n.s.
dhar3	n.s.	n.s.	Susceptible**
dhar1,2,3	Susceptible*	Susceptible*	Susceptible*
cyp86a1	n.s.	n.s.	n.s.
cyp86a2	n.s.	n.s.	n.s.
cyp86a8	Susceptible**	Susceptible**	Susceptible*
cyp86b1	n.s.	n.s.	n.s.
cyp86b2	n.s.	n.s.	n.s.
cyp86c1	n.s.	Susceptible*	n.s.
myb47	Susceptible**	Susceptible*	Susceptible*
spch	Susceptible**	Susceptible*	Susceptible*
knat7	Resistant**	Resistant*	Resistant*
sep3	Resistant**	Resistant**	Resistant**
sep1,2,4	n.s.	Resistant**	n.s.

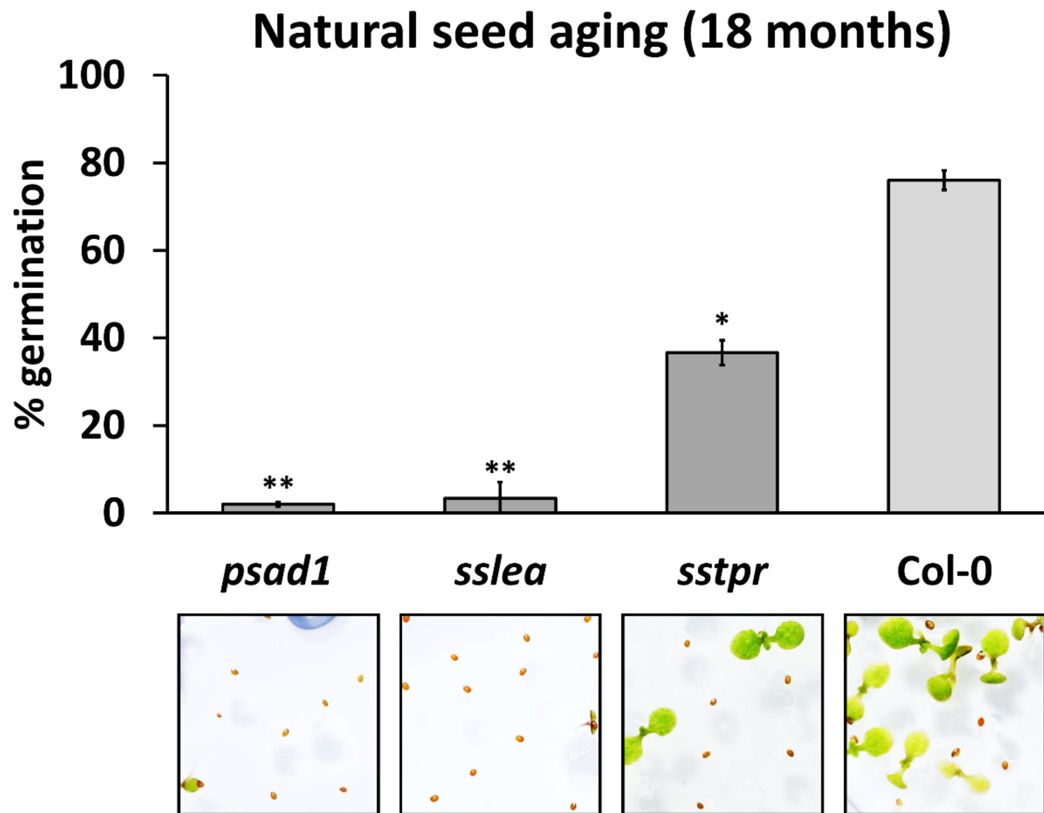


Figure 3. Seeds of *psad1*, *sslea* and *sstrp* mutant exhibit reduced seed longevity. Seed lots were stored for 18 months under natural conditions (see Materials and Methods) and sown on MS plates. Above: The percentage of germination was recorded after one week. The results are the average of three experiments with 50 seeds per line. Bars indicate standard errors. Not-aged seeds from all lines germinated more than 95%. *Significantly differing from wild type seeds at $p < 0.05$ (Student´s t-test). ** Significantly differing from wild type seeds at $p < 0.01$ (Student´s t-test). Below: representative images.

ROS accumulation by NADPH oxidases (RBOHs) is detrimental for seed longevity

Interestingly, two RBOHs genes, *RBOHD* (AT5G47910) and *RBOHE* (AT1G19230) were highlighted after AAT and EPPO treatments, respectively. RBOHs are transmembrane NADPH oxidases important for ROS production. They produce superoxide anion ($O_2^{\cdot-}$), rapidly transformed to H_2O_2, and play important roles in biotic and abiotic stresses (Chang *et al.*, 2016; Qu *et al.*, 2017). We decided to study them given the importance of ROS in seed ageing (Bailly, 2004; Sano *et al.*, 2016). There are ten *RBOH* genes in Arabidopsis named from A to J. Only three of them are abundantly expressed during seed development according to Le *et al.* (2010): *RBOHE* and *RBOHF*, mainly in the seed coat, and *ROBHD*, in the endosperm (Figure S 1). Seeds of *rbohd,*

rbohe and *rbohf* mutant plants exhibited increased longevity and the double mutant *rbohd,f* has even more seed longevity. (Figure 4). RBOH-mutant seeds showed no significant seed coat alterations (Figure S4). As indicated in Table 3, these four mutants also exhibited increased seed longevity with two artificial ageing treatments (AAT and CDT).

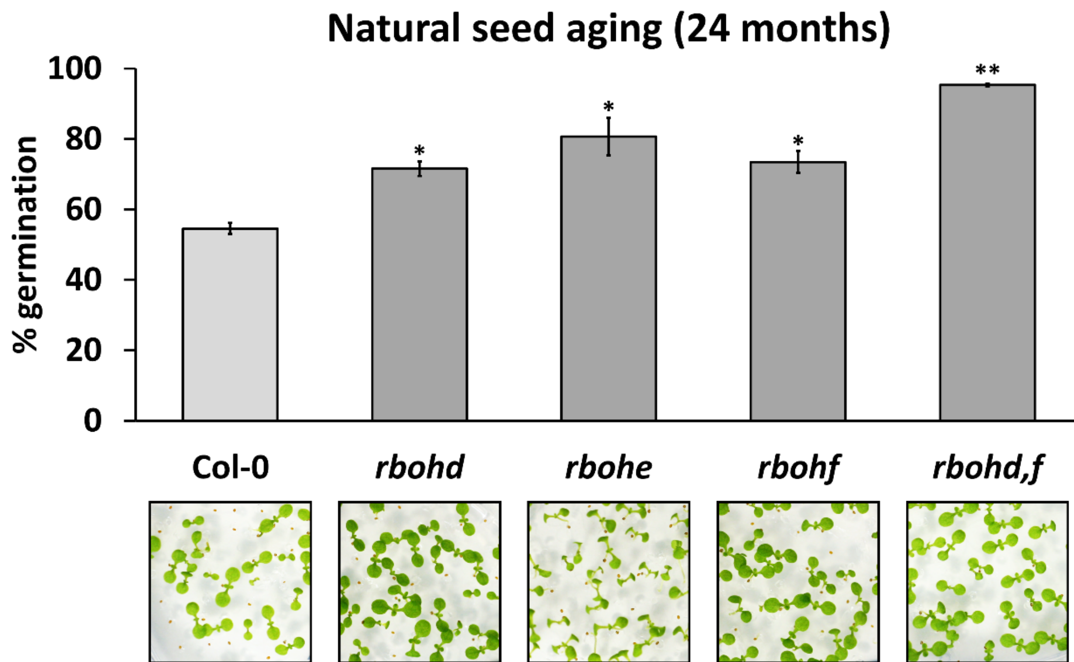


Figure 4. Seeds of *rbohd*, *rbohe*, *rbohf* and double mutant *rbohd,f* exhibit increased seed longevity. Seed lots were stored for 24 months and sown on MS plates. Above: The percentage of germination was recorded after one week. The results are the average of three experiments with 50 seeds per line. Bars indicate standard errors. Not-aged seeds from all lines germinated more than 97%. *Significantly differing from wild type seeds at $P < 0.05$ (Student´s t-test). **Significantly differing from wild type seeds at $p < 0.01$ (Student´s t-test). Below: representative images.

DHAR1 ROS-detoxification system is important for seed longevity

The *DHAR1* (AT1G19570) gene was highlighted after the EPPO and CDT treatments. DHARs are glutathione-dependent dehydroascorbate reductases involved in ROS detoxification. They catalyse the regeneration of ascorbate oxidized during detoxification of H_2O_2 by ascorbate peroxidase (Foyer and Noctor, 2011; Smirnoff, 2011). Given the importance of ROS in seed longevity we decided to study the implication of this gene. Three functional isoforms are described in Arabidopsis (Dixon *et al.*, 2002; Dixon and

Edwards, 2010). We tested all three single mutants and the triple mutant *dhar1,2,3* in order to clarify if there is a major player contributing to seed longevity.

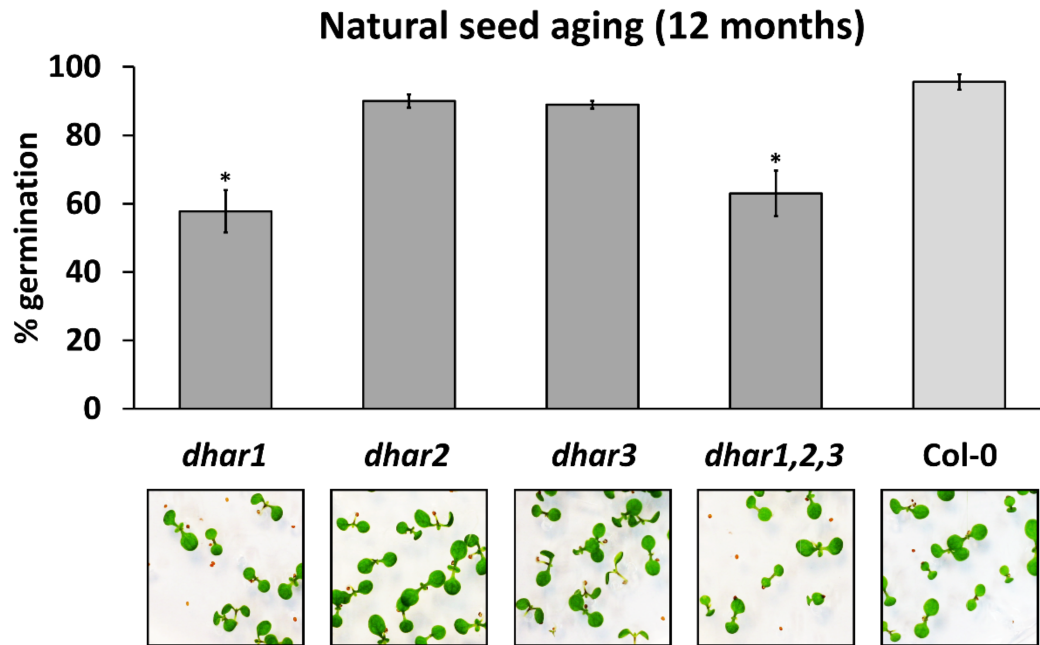


Figure 5. Seeds of *dhar1* and triple mutant *dhar1,2,3* present a reduced seed longevity, but *dhar2* and *dhar3* seeds no phenotype. Seed lots were stored for 12 months and sown on MS plates. Above: The percentage of germination was recorded after one week. The results are the average of three experiments with 50 seeds per line. Bars indicate standard errors. Not-aged seeds from all lines germinated more than 99%. *Significantly differing from wild type seeds at $p < 0.05$ (Student´s t-test). Below: representative images.

After one year of NAT treatment, *dhar1* seeds already lost part of their germination ability, while *dhar2* and *dhar3* mutant seeds did not. The triple mutant *dhar1,2,3* exhibited the same seed longevity reduction as the single *dhar1* mutant, indicating that other isoforms are not needed for seed longevity (Figure 5). According to Le *et al.* (2010) dataset, *DHAR1* is the most abundantly expressed *DHAR* isoform during seed development and in mature seed (Figure S1). This expression pattern is consistent with our natural ageing results. As indicated in Table 3, *dhar1* and *dhar1,2,3* mutants also exhibited reduced seed longevity with both artificial ageing treatments (AAT and CDT).

CYP86A8 is involved in synthesis of seed lipid-polyesters needed for seed longevity

GWAS highlights genomic areas causing the observed natural variation of traits (Curtin *et al.*, 2017). Natural variation, however, may affect less to major player genes due to their important physiological role. Here we present the example of a gene family, whose two members where highlighted after the ageing treatments, although a third member resulted to be more important for seed longevity in a mutant background. After the EPPO treatment, *CYP86A2* (AT4G00360) and *CYP86B1* (AT5G23190) were highlighted. They are related to suberin and cutin biosynthesis (Bak *et al.*, 2011), lipid polymers involved in seed permeability and recent results suggest an important role in seed longevity (Renard, 2020).

We obtained T-DNA mutant lines affecting members of the *CYP86* family abundantly expressed in seeds accordingly to Le *et al.* (2010) data (Figure S2): *CYP86A1* (AT5G58860), *CYP86A2* (AT4G00360), *CYP86A8* (AT2G45970), *CYP86B1* (AT5G23190), *CYP86B2* (*AT5G08250*) and *CYP86C1* (*AT1G24540*). A T-DNA insertion line close to *CYP86C3* was not available. After one year of seed dry storage, only seeds of *cyp86a8* showed a drastic reduction in their germination (Figure 6a), also observed during AAT and CDT treatments (Table 3). Interestingly, *cyp86a8* seeds are notably different to wild type: they were rounder, they present a reduced lipid polyester staining (Figure 6b) and they are more permeable due to the higher rate of tetrazolium reduction (Figure 6c).

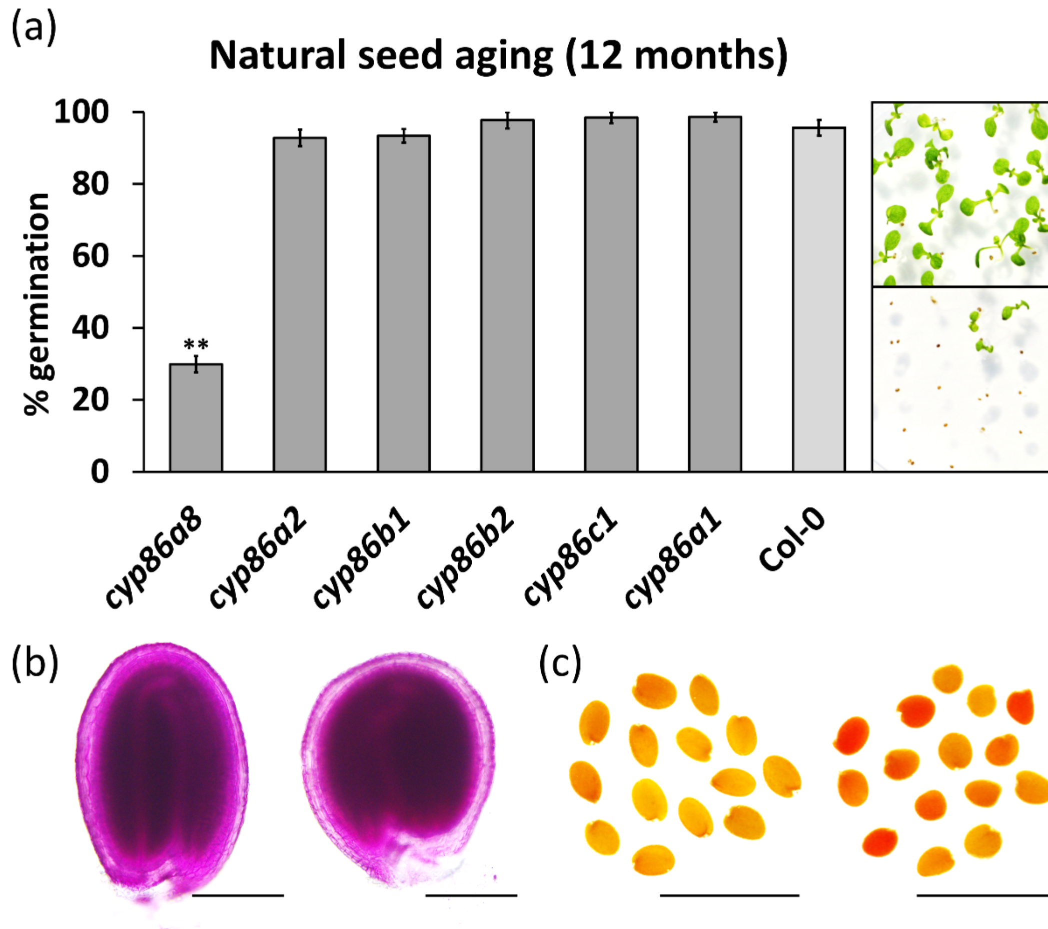


Figure 6. Reduced seed longevity of mutant *cyp86a8* correlates with reduced lipid polyester barriers. (a): Seeds of *cyp86a8* present a drastically reduced seed longevity, while seeds of mutant plants in other family members have no phenotype. Seed lots were stored for 12 months and sown on MS plates. Left: The percentage of germination was recorded after one week. Results are the average of three experiments with 50 seeds per line. Bars indicate standard errors. Not-aged seeds from all lines germinated more than 95%. **Significantly differing from wild type seeds at $p < 0.01$ (Student´s t-test). Right: representative images of Col-0 (up) and *cyp86a8* (down) germination. (b): Mutant *cyp86a8* seeds present a reduced suberin layer and a rounder shape. Representative image of Sudan Red staining of Col-0 (left) and *cyp86a8* (right) seeds. Scale bars: 200 µm. (c): Mutant seeds of *cyp86a8* are more permeable than Col-0 seeds. Tetrazolium reduction assay (24h) in Col-0 (left) and *cyp86a8* (right) seeds. Scale bars: 2 mm.

TFs expressed in seed coat and endosperm are important for seed longevity

Transcription factors (TFs) are essential for seed coat differentiation. We found some transcription factors expressed in seed coat and endosperm (Le *et al.*, 2010) among our candidate gene list. *SEP1* (AT5G15800), highlighted after natural ageing, *KNAT7* (AT1G62990) highlighted in the CDT treatment, *SPCH* (AT5G53210), highlighted in the

AAT, and *MYB47* (AT1G18710) , highlighted in all three artificial ageing treatments.

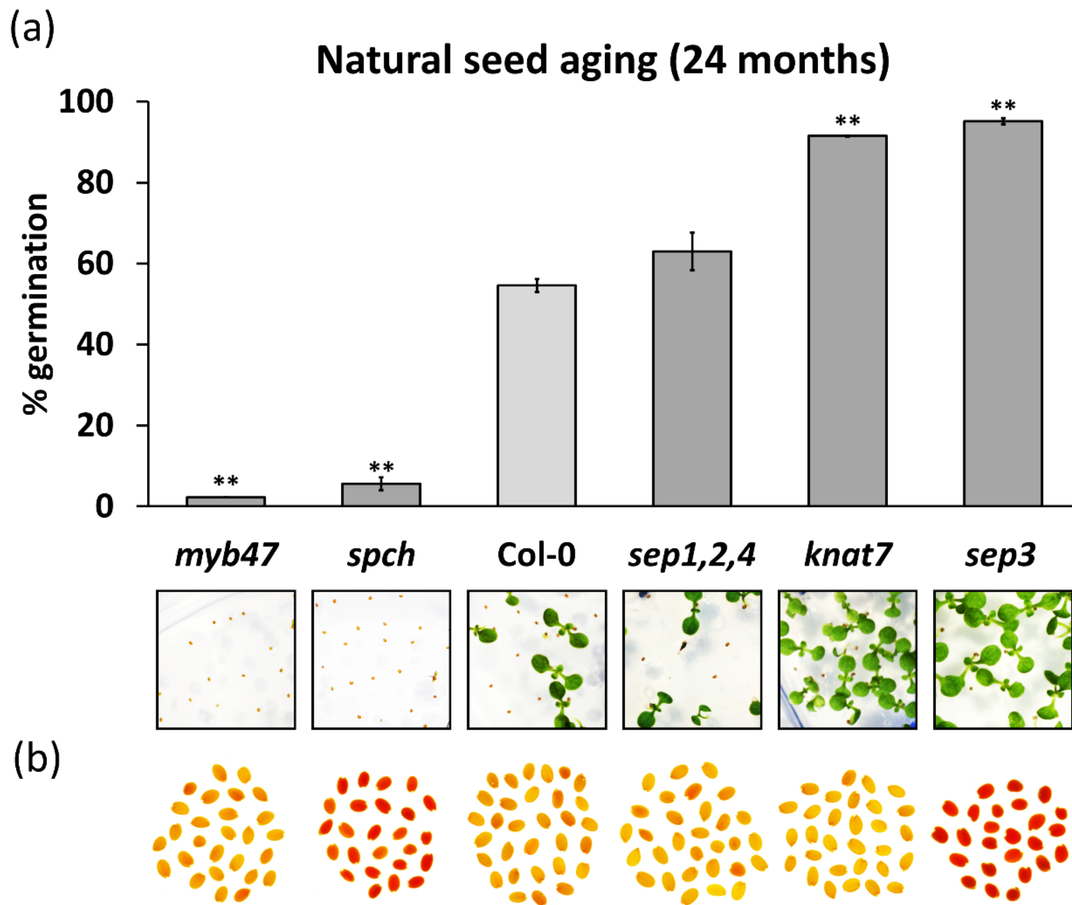


Figure 7. Mutants of different TFs exhibit differences in seed longevity and seed coat permeability. (a): Seeds of *spch* and *myb47* have reduced seed longevity; seeds of *knat7* and *sep3* have increased seed longevity, but triple mutant *sep1,2,4* has no phenotype. Seed lots were stored for 24 months and sown on MS plates. Above: The percentage of germination was recorded after one week. The results are the average of three experiments with 50 seeds per line. Bars indicate standard errors. Not-aged seeds from all lines germinated more than 97%. **Significantly differing from wild type seeds at $p < 0.01$ (Student´s t-test). Below: representative images. (b): Mutant seeds of *spch* and *sep3* are more permeable and *knat7* seeds are less permeable than Col-0 seeds. Tetrazolium reduction assay (24h) in *myb47*, *spch*, Col-0, *knat7* and *sep3* seeds. Scale bars: 2 mm.

Seeds of *myb47* and *spch* showed an important reduction of seed longevity, while seed longevity of *knat7* and *sep3* (but not of *sep1,2,4*) was increased (Figure 7a). Similar behaviour was observed in artificial ageing tests (Table 3). Here, a candidate gene (*SEP1*) was again not the member of the gene family with the strongest seed-phenotype (this was *sep3*). As those TFs were expressed in the seed coat (Figure S1), we performed a

tetrazolium assay to assess their seed coat permeability (Figure 7b). Seeds of *myb47* and *sep1,2,4* do no present significant differences to Col-0 seeds. Interestingly, *spch* and *knat7* seeds present enhanced and reduced tetrazolium salt reduction respectively, in correlation with their seed longevity phenotype (Figure S 6a for *knat7* seeds). Surprisingly, *sep3* seeds present a high reduction to tetrazolium salts. Seed coat assays to visualize proanthocyanidins (PAs), the mucilage halo, the suberin layer, and seed size were performed. Differences were observed only in *sep3* and *knat7* seeds. As described, *knat7* mutant seeds present less mucilage extrusion (Romano *et al.*, 2012), and *sep3* seeds exhibit a similar feature (Figure S5). Seed size is reduced in *knat7* seeds (Figure S 6b and c) and *sep3* seeds presented a rounder shape (Figure 7b).

***MYB47* is expressed in the seed coat and it is a positive regulator of seed longevity.**

To further investigate the MYB47 mechanism, we developed Arabidopsis transgenic lines. Confocal imageing of *proMYB47::GFP* plants locates the expression of *MYB47* in the seed coat during seed development. More precisely, GFP expression was visualized during first days of seed development in the seed coat, and in latter days, it was localized at the chalaza (Figure 8a). This expression pattern fits with Le *et al.* (2010) data (Figure S1).

Over-expression of *MYB47* driven by the UBQ10 promotor conferred resistance to AAT and CDT treatments compared to *wild-type* plants in three independent lines, highlighting the importance of *MYB47* in seed longevity as a positive regulator (Figure 8b). Expression analysis in seedlings demonstrates differences in MYB47 gene expression among different over-expressing lines. The *myb47* mutant has 30-fold reduction of *MYB47* gene expression, while over-expression lines increased *MYB47* expression 35 and 60-fold compared to wild-type plants (Figure 8c). Although *MYB47* expression is located in seed coat during seed development, we did not observe significant differences in seed coat component abundance (mucilage, PAs, or lipid polyester barriers) among over-expressing lines, compared to Col-0 (Figure S7).

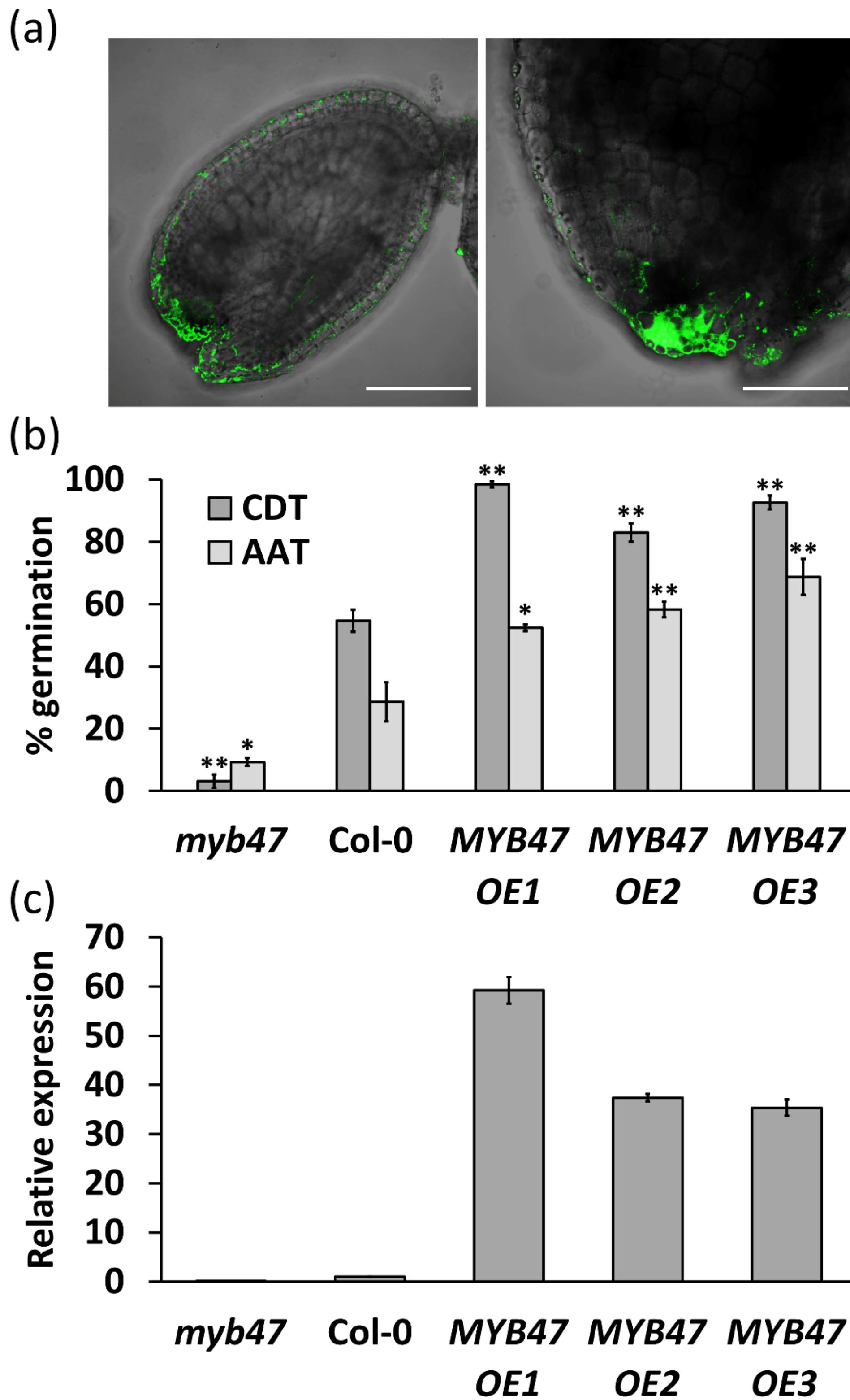


Figure 8. ***MYB47*** **is a positive TF gene for seed longevity and it is expressed in the chalaza during seed development.** (a): *MYB47* expression during seed development is localized initially at the seed coat (left, DAP3) and later in the chalaza (right, DAP7). Confocal imageing of developing seeds of *proMYB47::GFP* plants. Scale bar: 100 µm. (b): Seeds of three independent lines over-expressing *MYB47* are more resistant to accelerated ageing treatments CDT (dark bars) and AAT (light bars). The percentage of germination was recorded after one week. The results are the average of three experiments with 50 seeds per line. Bars indicate standard errors. Not-aged seeds from all lines germinated more than 99%.

*Significant differences from wild type seeds at P < 0.05 (Student´s t-test). **Significantly differing from wild type seeds at p < 0.01 (Student´s t-test). (c): qRT-PCR gene expression analysis of *MYB47* in 7 days-old seedlings from *myb47* mutant and three *MYB47* over-expressing lines. Expression values are relative to housekeeping gene *PP2AA3* and the resulting ratios are normalized to wild type, taken as 1. Results are the average of three determinations with bars corresponding to standard errors.

DISCUSSION

Many of the genes required for seed longevity are still unknown. In this work we have utilized a combination of GWAS, rational filtering and reverse genetics in *Arabidopsis thaliana* Col-0 to identify twelve novel genes involved in this trait. Seven are positive for longevity (*PSAD1, SSLEA, SSTPR, DHAR1, CYP86A8, MYB47* and *SPCH*) as their knock-out mutants have less seed longevity than their corresponding wild type. Five are negative ones (*RBOHD, RBOHE, RBOHF, KNAT7* and *SEP3*) as their knock-out mutants have more seed longevity than their corresponding wild type. These results contribute to the understanding of this polygenic trait.

Our data demonstrates a wide variation in seed longevity among Arabidopsis ecotypes. Correlations between artificial and natural ageing treatments indicate that, with exception of the AAT case, CDT and EPPO are valuable treatments to test seed longevity (Figure 1). Bueso *et al.* (2014) reported a good correlation between AAT and NAT. However, their plant material consisted of mutant lines in the same *Arabidopsis* background. In our case, each ecotype presents a different background genome, affecting the expression of multiple genetic components, probably causing this miss-correlation. In our T-DNA mutant approach, lines have the same genetic background (Col-0) and behave similarly with natural, AAT and CDT ageing procedures (Table 3), as in Bueso *et al.* (2014) experiments.

One important aspect observed in natural variation of seed longevity is that it inversely correlates with plant life spam, flowering time and rossette leave number (Table 1). These there phenotypes are linked, as rosette leave number increases as flowering delays, and plant life spam is dependent on flowering time to ensure plant reproduction. Flowering initiates the developmental program of tissues that will develop into seeds. In some way,

the rapid flowering initiation signals must be important for seed development and seed longevity acquisition, although the connection is not straightforward. There are no previous reports concerning this trait association. No correlation with ecotype latitude-collection point discard an effect of temperature or photoperiod on adaptation of seed-longevity. The negative correlation found between seed longevity and iron content has been already suggested (Murgia et al., 2015). It is plausible that high amounts of iron drives oxidative stress during seed storage. Seed dormancy, described to be inversely correlated with seed longevity (Nguyen *et al.*, 2012), was did not correlated with our longevity data, according to our data. Anthocyanin content influences PAs deposition in seed coats, a component putatively implicated in seed longevity (Debeaujon *et al.*, 2000) but was not correlated.

New mechanisms regulating seed longevity

The AAT treatment highlighted a LEA protein (AT3G17520) that we named SSLEA after the observed reduced seed storability of the mutant. There are 71 putative LEA proteins in Arabidopsis and their functions are usually associated with desiccation tolerance. A link between seed longevity and LEA proteins has been proposed based on the RNA interference targeting of LEA14 of Arabidopsis resulting in decreased seed longevity (Hundermarkt *et al*., 2011). This LEA protein belongs to group 2 (dehydrins) and SSLEA belongs to group 6 according to Jaspard *et al.* (2012). Dry seeds of the *sslea* mutant are able to germinate completely after ripening, but germination ability decays rapidly in months. This fact discards that *sslea* mutant seeds cannot cope with desiccation.
The EPPO treatment highlighted a tetratricopeptide-repeat (TPR) protein (AT4G02750) that we named SSTPR after the reduced seed storability of the mutant. These mutant seeds exhibit a high permeability by the tetrazolium test. SSTPR belongs to the TPR-PTR (tetratricopeptide-pentatricopeptide) family of repeat domain proteins (Sharma and Pandey, 2016). Not much is known about this TPR protein, or other family members. PTRs are described to play constitutive and essential roles in mitochondria and chloroplasts, probably binding to RNA (Lurin *et al.*, 2004). Indeed, SSTPR protein is sublocalized in the mitochondria according to Hooper *et al*. (2017).

ROS generation and detoxification are important in seed longevity

All three artificial ageing treatments highlighted genes involved in the ROS metabolism and detoxification. It has been widely accepted that this oxidative damage is one of the

major causes of seed ageing. This ROS negative effect has been remarked by the ROS-producing enzymes RBOHD, RBOHE and RBHOF, the ROS detoxification system involving DHAR1 and the photosystem I subunit PSAD1. ROS signalling is essential for diverse cellular processes and developmental programs. It is involved in the fine-tuning of the hypersensitive response (HR) upon pathogen infection, abiotic stress signalling and in hormone and developmental signalling, including programmed cell death (PCD) (Miller *et al.*, 2008; Suzuki *et al.*, 2011; Marino *et al.*, 2012). In seeds, ROS are required for embryogenesis, programmed cell death of the seed coat and endosperm and during seed germination, and act as protection against pathogens (Murphy *et al.*, 1998; Bailly *et al.*, 2008; Jeevan Kumar *et al.*, 2015). Here we present evidence that RBOH-produced ROS are a major cause of deterioration during ageing. The enhanced seed longevity of RBOH mutants could indicate RBOH-ROS activity through seed storage. RBOHE and RBOHF are expressed in the seed coat and RBOHD in the endosperm and embryo (Figure S 1), pointing to a role during cell death of these tissues. Plants express these proteins in mature seeds because ROS are needed for maturation of the seed coat and for defence against pathogen attack (Jeevan Kumar *et al.*, 2015).

A ROS-detoxification system has been highlighted by means of the *DHAR1* gene. DHARs are glutathione-dependent dehydroascorbate reductases. They catalyse the regeneration of ascorbate oxidized during detoxification of H_2O_2 by ascorbate peroxidase (Foyer and Noctor, 2011; Smirnoff, 2011). This model is supported by the high oxidation of glutathione in the triple mutant *dhar1,2,3* (Rahantaniaina *et al.*, 2017) and the use of glutathione redox equilibrium as a seed viability marker (Kranner *et al.*, 2006). Over-expression of ascorbate peroxidases leads to resistance to oxidative stress (Wang et al., 1999), highlighting the importance of this ROS-detoxification system. We observed that plants lacking *DHAR1,* but not *DHAR2* or *DHAR3*, presented less seed longevity, probably because less ROS detoxification. This is consistent with their expression profiles, with *DHAR1* being the most expressed isoform in seed embryo (Le *et al*., 2010).

Another highlighted ROS-related protein is the Photosystem I subunit D1 or PSAD1 (AT4G02770). It participates in the stability of this photosystem (Ihnatowicz *et al.*, 2004). This hydrophilic protein is exposed in the stroma and interacts directly with ferredoxin in the electron transport chain (Andersen *et al.*, 1992; Merati and Zanetti, 1987; Zilber and Malkin, 1988). Although there is a homologue gene in Arabidopsis (*PSAD2*), *PASD1* loss of function leads to growth defect due to a deficient photosynthesis, indicating that

PSAD2 is not completely redundant to *PSAD1*. Mutant *psad1* plants are pale-green and dwarf, and they have an increased photosensitivity and altered redox state of the stroma (Haldrup et al., 2003). The double mutant *psad1 psad2* is not viable (Ihnatowicz *et al.*, 2004). The importance of PSAD1 in seed longevity points to its role in photosynthesis and the imbalance of photosystems in the mutant plant. An imbalance of photosystems leads to increased ROS production (Pinnola and Bassi, 2018) and this may explain the reduced seed longevity phenotype. *PSAD1* transcripts are highly expressed at the embryo in latter stages of seed development according to Le *et al.* (2010) data (Figure S 1). Chlorophyll develops during embryo development for proper seed filling (Ruuska *et al.*, 2002; Goffman *et al.*, 2005) and latter disappears for seed storage avoiding light-induced ROS accumulation (Nakajima et al., 2012). However, we cannot discard that no direct correlation may exist and that the seed longevity phenotype of *psad1* seeds is due to seed developmental problems due to the strong pale and dwarf phenotype observed in the mutant plant.

Lipid polyester barriers prevents embryo ageing

The physical protection by seed coat, through lipids polyester barriers, has been highlighted in our study by two members of the CYP86 family. The CYP86 family belongs to the Cytochrome P450 superfamily, with 244 members in Arabidopsis and is involved in numerous metabolic pathways in all organisms (Mansuy, 1998; Nelson, 1999). Members of CYP86 gene family catalyse the ω-hydroxylation of fatty acids (Benveniste *et al.*, 1998; Wellesen *et al.*, 2001; Duan and Schuler, 2005), and mutant plants in different CYP86 members show a reduction in ω-hydroxy fatty acids and α,ω-dicarboxylic fatty acids (Höfer *et al.*, 2008; Compagnon *et al.*, 2009; Kai *et al.*, 2009). CYP86 enzymes participate in the synthesis of lipid-polyester barriers, such as cutin and suberin (Watson et al., 2001; Compagnon *et al.*, 2009; Kannangara *et al.*, 2007). Recent studies point to an important role of these lipid polyester barriers in seed longevity and tetrazolium impermeability (Beisson *et al.*, 2007; Yadav *et al.*, 2014; Bueso *et al.*, 2016; Renard *et al.*, 2020).

The drastic reduction of seed longevity, the high tetrazolium reduction rate and the lighter Sudan Red staining suggest that *cyp86a8* seeds are more permeable due to a reduced suberin and/or cutin layer. With this analysis, we can conclude that CYP86A8 is probably

the major cytochrome P-450 of the CYP86 family producing the ω-hydroxylation for seed lipid-polyester synthesis necessary for seed longevity. However, CYP86A8 is also important for diverse developmental and signalling processes as described by Wellesen *et al*. (2001), and we cannot discard that these processes are affecting also seed development.

TFs modulate seed longevity in different ways

The first physical barrier to protect the seed embryo is the seed coat, which requires a precise developmental program of its cell layers. Floral identity TFs regulate the ovule integuments that will develop into the seed coat. Many TFs and different protein-protein interactions important for seed coat development have been described (Sano *et al*., 2016; Golz *et al*., 2018). However, the complete TF cascade is far to be completely known. Here we demonstrate the implication of four TFs in the seed development transcriptional programme whose mutant lines present seeds with altered seed longevity: SEP3, SPCH, KNAT7 and MYB47.

SEP3 is part of the *SEPALLATA* (*SEP*) gene family, formed by four MAD-box TFs (Pelaz *et al*., 2000). They interact with other MAD-box proteins to determine flowering and ovule development (Favaro *et al*., 2003; Hugouvieux *et al*., 2018). They are not completely redundant as they differ in DNA-binding patterns (Jetha *et al*., 2014; Soza *et al*., 2016). SEP1 was highlighted in our analysis, but we found SEP3 to be determinant, due to the enhanced seed longevity and seed coat phenotype of *sep3* mutant seeds. The different seed shape and the reduced mucilage halo of *sep3* seeds suggest that SEP3 has a role in seed coat development. The SEP3 seed-coat expression pattern during seed development (Le *et al*., 2010) supports this idea. The results of the tetrazolium reduction by *sep3* seeds is intriguing. Normally tetrazolium salts reduction is related with a higher seed coat permeability, which is inversely correlated to seed longevity (Debeaujon *et al*., 2000). Mutant seeds of two SEP3 interactors present similar trait effects: *stk* seeds present a round shape as *sep3* seeds do (Mizzotti *et al*., 2014) and *tt16* seeds present high tetrazolium reduction rate (Debeaujon *et al*., 2000). Both of them are involved in inner integument regulation and seed shape (Coen *et al*., 2017). An explanation could be that *sep3* seed coat is physically different, perhaps softer. Internal embryo pressure would be responsible for this round shape. Pressure may be stronger during seed development and

during water imbibition (producing seed coat fissures), but no significant in the dry stage. This would explain the miss-correlation between tetrazolium reduction and their enhanced seed longevity in the dry state.

SPCH is a well-described basic helix-loop-helix (bHLH) TF involved in stomatal lineage determination during the asymmetrical division of future guard cells (MacAlister *et al.*, 2007). Mutant plants have no stomata at all. They have been widely studied for stomata formation but not in relation with seed development and longevity. We found that seed longevity of *spch* mutant seeds was reduced. Tetrazolium assay in *spch* seeds showed enhanced permeability (Figure 7b), indicating a putative function in seed development.

KNAT7 is a Homeobox-TF and seeds of the *knat7* mutant are more tolerant to seed ageing, are notably smaller and present decreased tetrazolium reduction (Figure S 6). *knat7* mutant seeds have been described to have affected epidermis cells responsible for mucilage secretion (Romano *et al.*, 2012). The expression pattern is at seed coat during last maturation stages (Figure S1). Recent studies propose that KNAT7 inhibits cell wall formation by regulating lignin synthesis through interaction with BHL6 and MYB75 (Bhargava *et al.*, 2013; Liu *et al.*, 2014). Cell wall deposition is an important aspect of seed coat development and seed longevity and the absence of this TF must improve it.

MYB47 is the only TF not previously described in other works. It belongs to R2-R3 MYB transcription factor family formed by 126 members involved in the regulation of diverse metabolic pathways, including PAs, suberin and cuticle biosynthesis (Dubos *et al.*, 2010). *MYB47* was highlighted after the three artificial treatments (AAT, CDT and EPPO) and mutant seeds present reduced seed longevity. Over-expression plants had enhanced seed longevity, confirmed with artificial ageing treatments (Figure 8b). This corroborates its positive effect in seed longevity. GFP promoter-driven expression locates *MYB47* expression in the seed coat during seed development, coinciding with published expression dataset (Figure S1). Nevertheless, we did not find differences in mucilage, lipid polyester staining or PAs deposition in the knock out mutant or in the over-expression mutants (Figure S7). Thus, MYB47 participates in an important seed coat mechanism contributing to seed longevity, not related with seed coat components such as PAs, mucilage or lipid polyester barriers, pointing to a new unknown protecting mechanism of the seed coat which remains unclear.

We have provided insights of the participation of four TFs in the seed longevity trait. Observed seed phenotypes suggest a role of this four seed TFs in seed coat development. However, further study is necessary to complete the understanding of these TF and their role in the TF cascade regulating seed development. Seed RNA-seq analysis of these mutants might provide us new players and new molecular mechanisms involved in seed longevity.

Concluding remarks

Our work demonstrates the power of combining GWAS and reverse genetics to identify novel seed longevity genes. We have described 12 novel genes involved in seed longevity. They were found by GWAS and validated by T-DNA insertion mutants. These genes are coding for seed coat TFs, ROS-producing enzymes, a ROS-detoxification enzyme, a mature seed protein, a photosystem component, a cytochrome P-450 and a scaffolding protein. More candidate genes remain to be validated in future studies, increasing the list of genes involved in this highly polygenic trait.

Acknowledgements

This work was funded by grant BIO2017-88898-P from the "Ministerio de Ciencia, Innovación y Universidades", Madrid, Spain. We thank Prof. Julian I. Schroeder and Prof. June M. Kwak (Division of Biological Sciences, University of California at San Diego, La Jolla, USA) for the double mutant *rbohd,f*; Prof. Graham Noctor (Institute of Plant Sciences Paris-Saclay (IPS2), UMR 9213/UMR1403, Université Paris-Sud, CNRS, INRA, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Orsay, France) for the *dhar* mutants (singles and triple) and Dr. Cristina Ferrándiz (IBMCP, Universisat Politècnica de València-CSIC, Valencia, Spain) for the *sep3 and sep1,2,4* mutants. The authors declare no conflicts of interest.

BIBLIOGRAPHY

Agacka-Mołdoch M., Nagel M., Doroszewska T., Lewis R.S. & Börner A. (2015). Mapping quantitative trait loci determining seed longevity in tobacco (*Nicotiana tabacum* L.). *Euphytica,* **202**, 479–486. https://doi.org/10.1007/s10681-015-1355-x

Alejandro, S., Rodríguez, P. L., Bellés, J. M., Yenush, L., García-Sanchez, M. J., Fernández, J. A., & Serrano, R. (2007). An Arabidopsis quiescin-sulfhydryl oxidase regulates cation homeostasis at the root symplast-xylem interface. *The EMBO Journal,* **26**, 3203–3215. https://doi.org/10.1038/sj.emboj.7601757

Andersen B., Koch B. & Scheller H.V. (1992). Structural and functional analysis of the reducing side of photosystem I. *Physiologia Plantarum,* **84**, 154–161. https://doi.org/10.1111/j.1399-3054.1992.tb08778.x

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, **465**, 627–631. https://doi.org/10.1038/nature08800

Bailly C. (2004). Active oxygen species and antioxidants in seed biology. *Seed Science Research*, **14**, 93–107. https://doi.org/10.1079/SSR2004159

Bailly C., El-Maarouf-Bouteau H. & Corbineau F. (2008). From intracellular signaling networks to cell death: the dual role of reactive oxygen species in seed physiology. *Comptes Rendus Biologies,* **331**, 806–814. https://doi.org/10.1016/j.crvi.2008.07.022

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., & Werck-Reichhart, D. (2011). Cytochromes P450. *The Arabidopsis book,* **9**, e0144. https://doi.org/10.1199/tab.0144

Ballesteros D. & Walters C. (2011). Detailed characterization of mechanical properties and molecular mobility within dry seed glasses: relevance to the physiology of dry biological systems. *The Plant Journal,* **68**, 607–619. https://doi.org/10.1111/j.1365-313X.2011.04711.x

Beisson, F., Li, Y., Bonaventure, G., Pollard, M., & Ohlrogge, J. B. (2007). The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of Arabidopsis. *The Plant Cell,* **19**, 351–368. https://doi.org/10.1105/tpc.106.048033

Benveniste I., Tijet N., Adas F., Philipps G., Salaün J.-P. & Durst F. (1998). CYP86A1 from *Arabidopsis thaliana* encodes a Cytochrome P450-dependent fatty acid omega-hydroxylase. *Biochemical and Biophysical Research Communications,* **243**, 688–693. https://doi.org/10.1006/bbrc.1998.8156

Bhargava A., Ahad A., Wang S., Mansfield S.D., Haughn G.W., Douglas C.J. & Ellis B.E. (2013). The interacting MYB75 and KNAT7 transcription factors modulate secondary cell wall deposition both in stems and seed coat in Arabidopsis. *Planta,* **237**, 1199–1211. https://doi.org/10.1007/s00425-012-1821-9

Brundrett M.C., Kendrick B. & Peterson C.A. (1991). Efficient lipid staining in plant material with sudan red 7B or fluoral yellow 088 in polyethylene glycol-glycerol. *Biotechnic & Histochemistry,* **66**, 111–116. https://doi.org/10.3109/10520299109110562

Bueso, E., Muñoz-Bertomeu, J., Campos, F., Brunaud, V., Martínez, L., Sayas, E., Ballester, P., Yenush, L., & Serrano, R. (2014). ARABIDOPSIS THALIANA HOMEOBOX25 uncovers a role for Gibberellins in seed longevity. *Plant physiology,* **164**, 999–1010. https://doi.org/10.1104/pp.113.232223

Bueso, E., Muñoz‐Bertomeu, J., Campos, F., Martínez, C., Tello, C., Martínez‐Almonacid, I., Ballester, P. Simón-Moya, M., Brunaud, V., Yenush, L., Ferrándiz, C. & Serrano, R. (2016). Arabidopsis COGWHEEL1 links light perception and gibberellins with seed tolerance to deterioration. *The Plant Journal,* **87**, 583– 596. https://doi.org/10.1111/tpj.13220

Cao J., Schneeberger K., Ossowski S., Günther T., Bender S., Fitz J., … Weigel D. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nature Genetics,* **43**, 956–963. https://doi.org/10.1038/ng.911

Chang, Y. L., Li, W. Y., Miao, H., Yang, S. Q., Li, R., Wang, X., Li, W. Q., & Chen, K. M. (2016). Comprehensive genomic analysis and expression profiling of the NOX gene families under abiotic stresses and hormones in plants. *Genome Biology and Evolution,* **8**, 791–810. https://doi.org/10.1093/gbe/evw035

Clerkx, E.J.M., Blankestijn-De Vries, H., Ruys, G.J., Groot, S.P.C. and Koornneef, M. (2004) Genetic differences in seed longevity of various Arabidopsis mutants. *Physiologia Plantarum* **121**, 448–461. https://doi.org/10.1111/j.0031-9317.2004.00339.x

Coen O., Fiume E., Xu W., Vos D.D., Lu J., Pechoux C., … Magnani E. (2017). Developmental patterning of the sub-epidermal integument cell layer in *Arabidopsis* seeds. *Development,* **144**, 1490–1497. https://doi.org/10.1242/dev.146274

Compagnon V., Diehl P., Benveniste I., Meyer D., Schaller H., Schreiber L., … Pinot F. (2009). CYP86B1 Is Required for Very Long Chain ω-Hydroxyacid and α,ω-Dicarboxylic Acid Synthesis in Root and Seed Suberin Polyester. *Plant Physiology*, **150**, 1831–1843. https://doi.org/10.1104/pp.109.141408

Curtin, S. J., Tiffin, P., Guhlin, J., Trujillo, D. I., Burghart, L. T., Atkins, P., Baltes, N. J., Denny, R., Voytas, D. F., Stupar, R. M., & Young, N. D. (2017). Validating genome-wide association candidates controlling quantitative variation in nodulation. *Plant Physiology,* **173**, 921–931. https://doi.org/10.1104/pp.16.01923

Curtis, M. D., & Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiology,* **133**, 462–469. https://doi.org/10.1104/pp.103.027979

Czechowski T., Stitt M., Altmann T., Udvardi M.K. & Scheible W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology,* **139**, 5–17. https://doi.org/10.1104/pp.105.063743

Debeaujon, I., Léon‐Kloosterziel, K. M., & Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiology*, **122**, 403– 414. https://doi.org/10.1104/pp.122.2.403

Ditta G., Pinyopich A., Robles P., Pelaz S. & Yanofsky M.F. (2004). The SEP4 Gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Current Biology* **14**, 1935–1940. https://doi.org/10.1016/j.cub.2004.10.028

Dixon D.P. & Edwards R. (2010). Glutathione transferases. *The Arabidopsis Book,* **2010**. https://doi.org/10.1199/tab.0131

Dixon D.P., Davis B.G. & Edwards R. (2002). Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **277**, 30859–30869. https://doi.org/10.1074/jbc.M202919200

Dubos C., Stracke R., Grotewold E., Weisshaar B., Martin C. & Lepiniec L. (2010). MYB transcription factors in *Arabidopsis*. *Trends in Plant Science,* **15**, 573–581. https://doi.org/10.1016/j.tplants.2010.06.005

Duan H. & Schuler M.A. (2005). Differential Expression and Evolution of the Arabidopsis CYP86A Subfamily. *Plant Physiology,* **137**, 1067–1081. https://doi.org/10.1104/pp.104.055715

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M., & Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant Cell,* **15**, 2603–2611. https://doi.org/10.1105/tpc.015123

Foyer C.H. & Noctor G. (2011). Ascorbate and glutathione: the heart of the redox hub. *Plant Physiology* **155**, 2–18. https://doi.org/10.1104/pp.110.16756

Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., … Mott, R. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature,* **477**, 419–423. https://doi.org/10.1038/nature10414

Gladyshev V. N. (2014). The free radical theory of ageing is dead. Long live the damage theory!. *Antioxidants & Redox Signaling*, *20*(4), 727–731. https://doi.org/10.1089/ars.2013.5228

Goffman, F. D., Alonso, A. P., Schwender, J., Shachar-Hill, Y., & Ohlrogge, J. B. (2005). Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant physiology,* **138**, 2269–2279. https://doi.org/10.1104/pp.105.063628

Golz J.F., Allen P.J., Li S.F., Parish R.W., Jayawardana N.U., Bacic A. & Doblin M.S. (2018). Layers of regulation – Insights into the role of transcription factors controlling mucilage production in the *Arabidopsis* seed coat. *Plant Science,* **272**, 179–192. https://doi.org/10.1016/j.plantsci.2018.04.021

Grefen C., Donald N., Hashimoto K., Kudla J., Schumacher K. & Blatt M.R. (2010). A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *The Plant Journal,* **64**, 355–365. https://doi.org/10.1111/j.1365-313X.2010.04322.x

Grimm, D. G., Roqueiro, D., Salomé, P. A., Kleeberger, S., Greshake, B., Zhu, W., … Borgwardt, K. M. (2017). easyGWAS: A Cloud-Based Platform for Comparing the Results of Genome-Wide Association Studies. *The Plant Cell*, **29**(1), 5–19. https://doi.org/10.1105/tpc.16.00551

Groot, S. P., Surki, A. A., de Vos, R. C., & 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**(6), 1149–1159. https://doi.org/10.1093/aob/mcs198

Haldrup A., Lunde C. & Scheller H.V. (2003). Arabidopsis thaliana Plants Lacking the PSI-D Subunit of Photosystem I Suffer Severe Photoinhibition, Have Unstable Photosystem I Complexes, and Altered Redox Homeostasis in the Chloroplast Stroma. *Journal of Biological Chemistry,* **278**, 33276–33283. https://doi.org/10.1074/jbc.M305106200

Harman D. (1956). Ageing: A Theory Based on Free Radical and Radiation Chemistry. *Journal of Gerontology,* **11**, 298–300. https://doi.org/10.1093/geronj/11.3.298

Harman, G.E. & Mattick, L.R. (1976). Association of lipid oxidation with seed ageing and death. *Nature,* **260,** 323–324. https://doi.org/10.1038/260323a0

Haughn, G. and Chaudhury, A. (2005) Genetic analysis of seed coat development in Arabidopsis. Trends in Plant Sciences **10**, 472–477. https://doi.org/10.1016/j.tplants.2005.08.005

Hay F.R., Valdez R., Lee J.-S. & Sta. Cruz P.C. (2019). Seed longevity phenotyping: recommendations on research methodology. *Journal of Experimental Botany,* **70**, 425–434. https://doi.org/10.1093/jxb/ery358

Hoekstra F.A., Golovina E.A. & Buitink J. (2001). Mechanisms of plant desiccation tolerance. *Trends in Plant Science,* **6**, 431–438. https://doi.org/10.1016/S1360-1385(01)02052-0

Höfer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L., & Franke, R. (2008). The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis. *Journal of experimental botany*, **59**(9), 2347–2360. https://doi.org/10.1093/jxb/ern101

Hooper C.M., Castleden I.R., Tanz S.K., Aryamanesh N. & Millar A.H. (2017). SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Research,* **45**, D1064–D1074. https://doi.org/10.1093/nar/gkw1041

Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., … Bergelson, J. (2012). Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nature Genetics*, **44**(2), 212–216. https://doi.org/10.1038/ng.1042

Hugouvieux V., Silva C.S., Jourdain A., Stigliani A., Charras Q., Conn V., … Zubieta C. (2018). Tetramerization of MADS family transcription factors SEPALLATA3 and AGAMOUS is required for floral meristem determinacy in Arabidopsis. *Nucleic Acids Research,* **46**, 4966–4977. https://doi.org/10.1093/nar/gky205

Ihnatowicz A., Pesaresi P., Varotto C., Richly E., Schneider A., Jahns P., … Leister D. (2004). Mutants for photosystem I subunit D of Arabidopsis thaliana: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *The Plant Journal*, **37**, 839–852. https://doi.org/10.1111/j.1365-313X.2004.02011.x

Ioannidis, J., Thomas, G. & Daly, M. (2009). Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet,* **10,** 318–329. https://doi.org/10.1038/nrg2544

ISTA. 2018. Seed vigour testing. In: International rules for seed testing 2018. Bassersdorf: International Seed Testing Association.

Jaspard, E., Macherel, D., & Hunault, G. (2012). Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. *PloS One*, **7**, e36968. https://doi.org/10.1371/journal.pone.0036968

Jeevan Kumar, S. P., Rajendra Prasad, S., Banerjee, R., & Thammineni, C. (2015). Seed birth to death: dual functions of reactive oxygen species in seed physiology. *Annals of Botany*, **116**, 663–668. https://doi.org/10.1093/aob/mcv098

Jetha, K., Theißen, G., & Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic acids research*, **42**(17), 10927–10942. https://doi.org/10.1093/nar/gku755

Jiang W., Lee J., Jin Y.-M., Qiao Y., Piao R., Jang S.M., … Koh H.-J. (2011). Identification of QTLs for seed germination capability after various storage periods using two RIL populations in rice. *Molecules and Cells,* **31**, 385–392. https://doi.org/10.1007/s10059-011-0049-z

Kai K., Hashidzume H., Yoshimura K., Suzuki H., Sakurai N., Shibata D. & Ohta D. (2009). Metabolomics for the characterization of cytochromes P450-dependent fatty acid hydroxylation reactions in *Arabidopsis*. *Plant Biotechnology,* **26**, 175–182. https://doi.org/10.5511/plantbiotechnology.26.175

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, **9**, 29. https://doi.org/10.1186/1746-4811-9-29

Kranner I., Birtić S., Anderson K.M. & Pritchard H.W. (2006). Glutathione half-cell reduction potential: A universal stress marker and modulator of programmed cell death? *Free Radical Biology and Medicine,* **40**, 2155–2165. https://doi.org/10.1016/j.freeradbiomed.2006.02.013

Landjeva S., Lohwasser U. & Börner A. (2010). Genetic mapping within the wheat D genome reveals QTL for germination, seed vigour and longevity, and early seedling growth. *Euphytica,* **171**, 129–143. https://doi.org/10.1007/s10681-009-0016-3

Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., … Goldberg, R. B. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(18), 8063–8070. https://doi.org/10.1073/pnas.1003530107

Leprince O., Pellizzaro A., Berriri S. & Buitink J. (2017). Late seed maturation: drying without dying. *Journal of Experimental Botany,* **68**, 827–841. https://doi.org/10.1093/jxb/erw363

Liu, Y., You, S., Taylor-Teeples, M., Li, W. L., Schuetz, M., Brady, S. M., & Douglas, C. J. (2014). BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 interact and regulate secondary cell wall formation via repression of REVOLUTA. *The Plant cell*, **26**(12), 4843–4861. https://doi.org/10.1105/tpc.114.128322

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., … Nordborg, M. (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature genetics*, **45**(8), 884–890. https://doi.org/10.1038/ng.2678

López-Otín, C., Galluzzi L., Freije J.M.P., Madeo F. & Kroemer G. (2016). Metabolic Control of Longevity. *Cell,* **166**, 802–821. https://doi.org/10.1016/j.cell.2016.07.031

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., … Small, I. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *The Plant cell*, **16**(8), 2089–2103. https://doi.org/10.1105/tpc.104.022236

MacAlister, C., Ohashi-Ito, K. & Bergmann, D. (2007). Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nature,* **445,** 537–540 https://doi.org/10.1038/nature05491

Mansuy D. (1998). The great diversity of reactions catalyzed by cytochromes P450. *Comparative Biochemistry and Physiology. Part C, Pharmacology, Toxicology & Endocrinology,* **121**, 5–14. https://doi.org/10.1016/s0742-8413(98)10026-9

Marino D., Dunand C., Puppo A. & Pauly N. (2012). A burst of plant NADPH oxidases. *Trends in Plant Science,* **17**, 9–15. https://doi.org/10.1016/j.tplants.2011.10.001

Merati G. & Zanetti G. (1987). Chemical cross-linking of ferredoxin to spinach thylakoids: Evidence for two independent binding sites of ferredoxin to the membrane. *FEBS Letters,* **215**, 37–40. https://doi.org/10.1016/0014-5793(87)80109-6

Miller G., Shulaev V. & Mittler R. (2008). Reactive oxygen signaling and abiotic stress. *Physiologia Plantarum,* **133**, 481–489. https://doi.org/10.1111/j.1399-3054.2008.01090.x

Mizzotti, C., Ezquer, I., Paolo, D., Rueda-Romero, P., Guerra, R. F., Battaglia, R., Rogachev, I., Aharoni, A., Kater, M. M., Caporali, E., & Colombo, L. (2014). SEEDSTICK is a master regulator of development and metabolism in the Arabidopsis seed coat. *PLoS genetics*, **10**(12), e1004856. https://doi.org/10.1371/journal.pgen.1004856

Murgia I., Giacometti S., Balestrazzi A., Paparella S., Pagliano C. & Morandini P. (2015) Analysis of the transgenerational iron deficiency stress memory in Arabidopsis thaliana plants. Frontiers in Plant Science 6, 745. https://doi.org/10.3389/fpls.2015.00745

Murphy T.M., Asard H. & Cross A.R. (1998). Possible Sources of Reactive Oxygen during the Oxidative Burst in Plants. In Plasma Membrane Redox Systems and their Role in Biological Stress and Disease (eds H. Asard, A. Bérczi & R.J. Caubergs), pp. 215–246. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-017-2695-5_9

Nagel M., Seal C.E., Colville L., Rodenstein A., Un S., Richter J., … Kranner I. (2019). Wheat seed ageing viewed through the cellular redox environment and changes in pH. *Free Radical Research,* **53**, 641–654. https://doi.org/10.1080/10715762.2019.1620226

Nagel, M., Kodde, J., Pistrick, S., Mascher, M., Börner, A., & Groot, S. P. (2016). Barley Seed Ageing: Genetics behind the Dry Elevated Pressure of Oxygen Ageing and Moist Controlled Deterioration. *Frontiers in plant science*, **7**, 388. https://doi.org/10.3389/fpls.2016.00388

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**(1), 261–273. https://doi.org/10.1104/pp.112.196881

Nelson D.R. (1999). Cytochrome P450 and the Individuality of Species. *Archives of Biochemistry and Biophysics,* **369**, 1–10. https://doi.org/10.1006/abbi.1999.1352

Nguyen T.-P., Keizer P., Eeuwijk F. van, Smeekens S. & Bentsink L. (2012). Natural Variation for Seed Longevity and Seed Dormancy Are Negatively Correlated in Arabidopsis. *Plant Physiology,* **160**, 2083–2092. https://doi.org/10.1104/pp.112.206649

O'Malley R.C. & Ecker J.R. (2010). Linking genotype to phenotype using the Arabidopsis unimutant collection. *The Plant Journal,* **61**, 928–940. https://doi.org/10.1111/j.1365-313X.2010.04119.x

Oliver M.J., Tuba Z. & Mishler B.D. (2000). The evolution of vegetative desiccation tolerance in land plants. *Plant Ecology,* **151**, 85–100. https://doi.org/10.1023/A:1026550808557

Pelaz S., Ditta G.S., Baumann E., Wisman E. & Yanofsky M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature,* **405**, 200–203. https://doi.org/10.1038/35012103

Pinnola A. & Bassi R. (2018) Molecular mechanisms involved in plant photoprotection. *Biochemical Society Transactions,* **46**, 467–482. https://doi.org/10.1042/BST20170307

Prieto-Dapena P., Castaño R., Almoguera C. & Jordano J. (2006). Improved Resistance to Controlled Deterioration in Transgenic Seeds. *Plant Physiology,* **142**, 1102–1112. https://doi.org/10.1104/pp.106.087817

Qu, Y., Yan, M., & Zhang, Q. (2017). Functional regulation of plant NADPH oxidase and its role in signaling. *Plant Signaling &Behavior*, **12**(8), e1356970. https://doi.org/10.1080/15592324.2017.1356970

Rahantaniaina M.-S., Li S., Chatel-Innocenti G., Tuzet A., Issakidis-Bourguet E., Mhamdi A. & Noctor G. (2017). Cytosolic and Chloroplastic DHARs Cooperate in Oxidative Stress-Driven Activation of the Salicylic Acid Pathway. *Plant Physiology,* **174**, 956–971. https://doi.org/10.1104/pp.17.00317

Rajjou, L., Lovigny, Y., Groot, S. P., Belghazi, M., Job, C., & Job, D. (2008). Proteome-wide characterization of seed ageing in Arabidopsis: a comparison between artificial and natural ageing protocols. *Plant physiology*, **148**(1), 620–641. https://doi.org/10.1104/pp.108.123141

Renard J., Martínez‐Almonacid I., Sonntag A., Molina I., Moya‐Cuevas J., Bissoli G., … Bueso E. (2020). PRX2 and PRX25, peroxidases regulated by COG1, are involved in seed longevity in Arabidopsis. *Plant, Cell & Environment*, **43**(2), 315-326. https://doi.org/10.1111/pce.13656

Righetti K., Vu J.L., Pelletier S., Vu B.L., Glaab E., Lalanne D., … Buitink J. (2015). Inference of Longevity-Related Genes from a Robust Coexpression Network of Seed Maturation Identifies Regulators Linking Seed Storability to Biotic Defense-Related Pathways. *The Plant Cell,* **27**, 2692–2708. https://doi.org/10.1105/tpc.15.00632

Romano J.M., Dubos C., Prouse M.B., Wilkins O., Hong H., Poole M., … Campbell M.M. (2012). AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. *New Phytologist,* **195**, 774–786. https://doi.org/10.1111/j.1469-8137.2012.04201.x

Ruuska S.A., Girke T., Benning C. & Ohlrogge J.B. (2002). Contrapuntal Networks of Gene Expression during Arabidopsis Seed Filling. *The Plant Cell* **14**, 1191. https://doi.org/10.1105/tpc.000877

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. https://doi.org/10.1093/pcp/pcv186

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M. and DellaPenna, D. (2004) Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *The Plant Cell* **16**, 1419–1432. https://doi.org/10.1105/tpc.021360

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., … Cardona A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods,* **9**, 676–682. https://doi.org/10.1038/nmeth.2019

Schmidlin C.J., Dodson M.B., Madhavan L. & Zhang D.D. (2019). Redox regulation by NRF2 in ageing and disease. *Free Radical Biology and Medicine,* **134**, 702–707. https://doi.org/10.1016/j.freeradbiomed.2019.01.016

Schwember A.R. & Bradford K.J. (2010). Quantitative trait loci associated with longevity of lettuce seeds under conventional and controlled deterioration storage conditions. *Journal of Experimental Botany,* **61**, 4423–4436. https://doi.org/10.1093/jxb/erq248

Seren, Ü., Vilhjálmsson, B. J., Horton, M. W., Meng, D., Forai, P., Huang, Y. S., Long, Q., Segura, V., & Nordborg, M. (2012). GWAPP: a web application for genome-wide association mapping in Arabidopsis. *The Plant cell*, **24**(12), 4793–4805. https://doi.org/10.1105/tpc.112.108068

Sharma, M., & Pandey, G. K. (2016). Expansion and Function of Repeat Domain Proteins During Stress and Development in Plants. *Frontiers in plant science*, **6**, 1218. https://doi.org/10.3389/fpls.2015.01218

Smirnoff N. (2011). Chapter 4 - Vitamin C: The Metabolism and Functions of Ascorbic Acid in Plants. In F. Rébeillé & R. Douce (Eds.), *Advances in botanical research*, 59, 107–177. Academic Press. https://doi.org/10.1016/B978-0-12-385853-5.00003-9

Soza V.L., Snelson C.D., Hewett Hazelton K.D. & Di Stilio V.S. (2016). Partial redundancy and functional specialization of E-class SEPALLATA genes in an early-diverging eudicot. *Developmental Biology,* **419**, 143–155. https://doi.org/10.1016/j.ydbio.2016.07.021

Suzuki N., Miller G., Morales J., Shulaev V., Torres M.A. & Mittler R. (2011). Respiratory burst oxidases: the engines of ROS signaling. *Current Opinion in Plant Biology,* **14**, 691–699. https://doi.org/10.1016/j.pbi.2011.07.014

Torres M.A., Dangl J.L. & Jones J.D.G. (2002). Arabidopsis gp91phox homologues AtrbohD and AtrbohF are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proceedings of the National Academy of Sciences,* **99**, 517–522. https://doi.org/10.1073/pnas.012452499

Vertucci, C. W., & Roos, E. E. (1990). Theoretical basis of protocols for seed storage. *Plant physiology*, **94**(3), 1019–

1023. https://doi.org/10.1104/pp.94.3.1019

Wang J., Zhang H. & Allen R.D. (1999). Overexpression of an Arabidopsis Peroxisomal Ascorbate Peroxidase Gene in Tobacco Increases Protection Against Oxidative Stress. *Plant and Cell Physiology,* **40**, 725–732. https://doi.org/10.1093/oxfordjournals.pcp.a029599

Watson C.J.W., Froehlich J.E., Josefsson C.A., Chapple C., Durst F., Benveniste I. & Coolbaugh R.C. (2001). Localization of CYP86B1 in the Outer Envelope of Chloroplasts. *Plant and Cell Physiology,* **42**, 873–878. https://doi.org/10.1093/pcp/pce110

Weigel D. (2012). Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant physiology*, **158**(1), 2–22. https://doi.org/10.1104/pp.111.189845

Wellesen K., Durst F., Pinot F., Benveniste I., Nettesheim K., Wisman E., … Yephremov A. (2001). Functional analysis of the LACERATA gene of Arabidopsis provides evidence for different roles of fatty acid ω-hydroxylation in development. *Proceedings of the National Academy of Sciences,* **98**, 9694–9699. https://doi.org/10.1073/pnas.171285998

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PloS one*, **2**(8), e718. https://doi.org/10.1371/journal.pone.0000718

Wood, A. J., & Jenks, M. A. (2008). Plant desiccation tolerance: Diversity, distribution, and real-world application. M.A. Jenks & A. J. Wood In *Plant desiccation tolerance*. pp. 1–10. Ames: Blackwell Publishing. https://doi.org/10.1002/9780470376881.ch1

Xie H.-T., Wan Z.-Y., Li S. & Zhang Y. (2014) Spatiotemporal Production of Reactive Oxygen Species by NADPH Oxidase Is Critical for Tapetal Programmed Cell Death and Pollen Development in Arabidopsis. The Plant Cell 26, 2007–2023. https://doi.org/10.1105/tpc.114.125427

Yadav V., Molina I., Ranathunge K., Castillo I.Q., Rothstein S.J. & Reed J.W. (2014). ABCG Transporters Are Required for Suberin and Pollen Wall Extracellular Barriers in Arabidopsis. *The Plant Cell,* **26**, 3569–3588. https://doi.org/10.1105/tpc.114.129049

Zilber A.L. & Malkin R. (1988). Ferredoxin Cross-Links to a 22 kD Subunit of Photosystem I. *Plant Physiology,* **88**, 810. https://doi.org/10.1104/pp.88.3.810

SUPPLEMENTARY MATERIAL

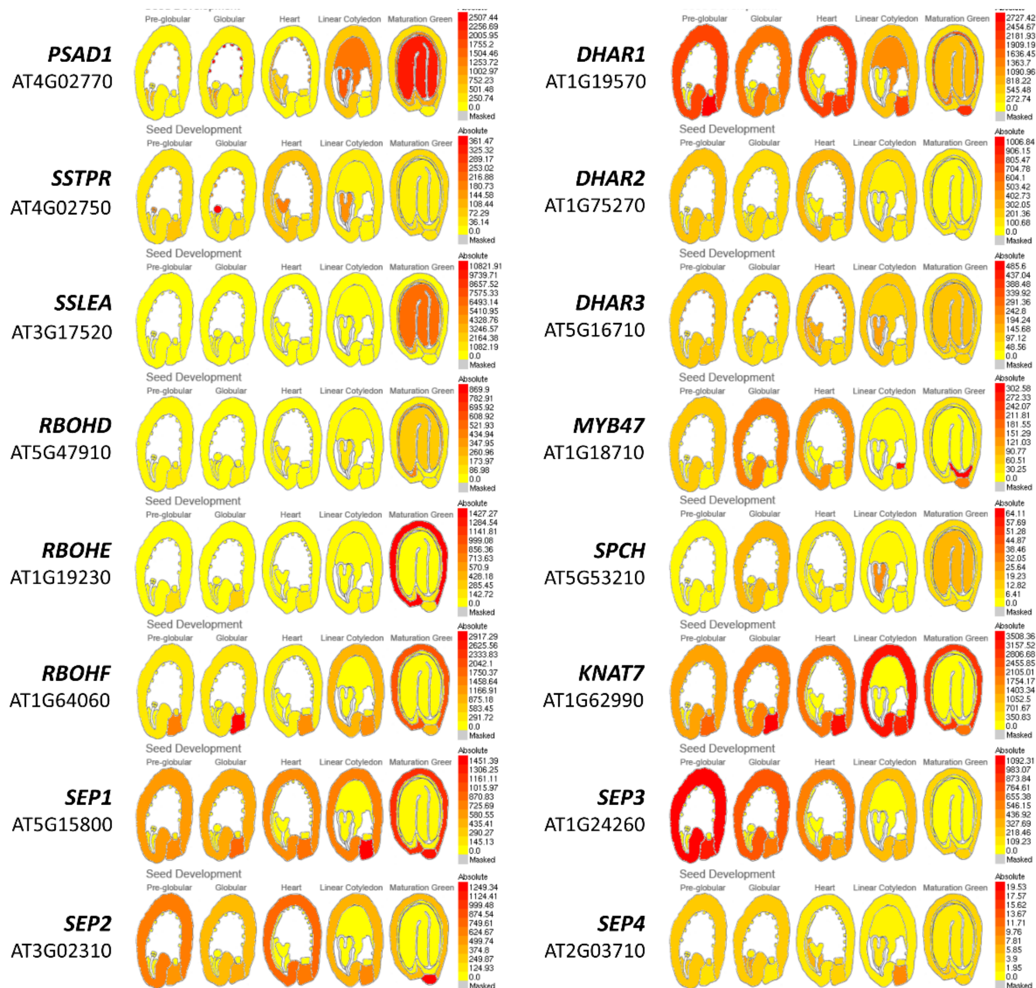


Figure S1. *In silico* expression analysis of genes used in this study (except the CYP86 gene family, Supplemental Figure 2) during seed development based in Le et al . (2010) data. Images were obtained with the eFP browser (Winter et al ., 2007).

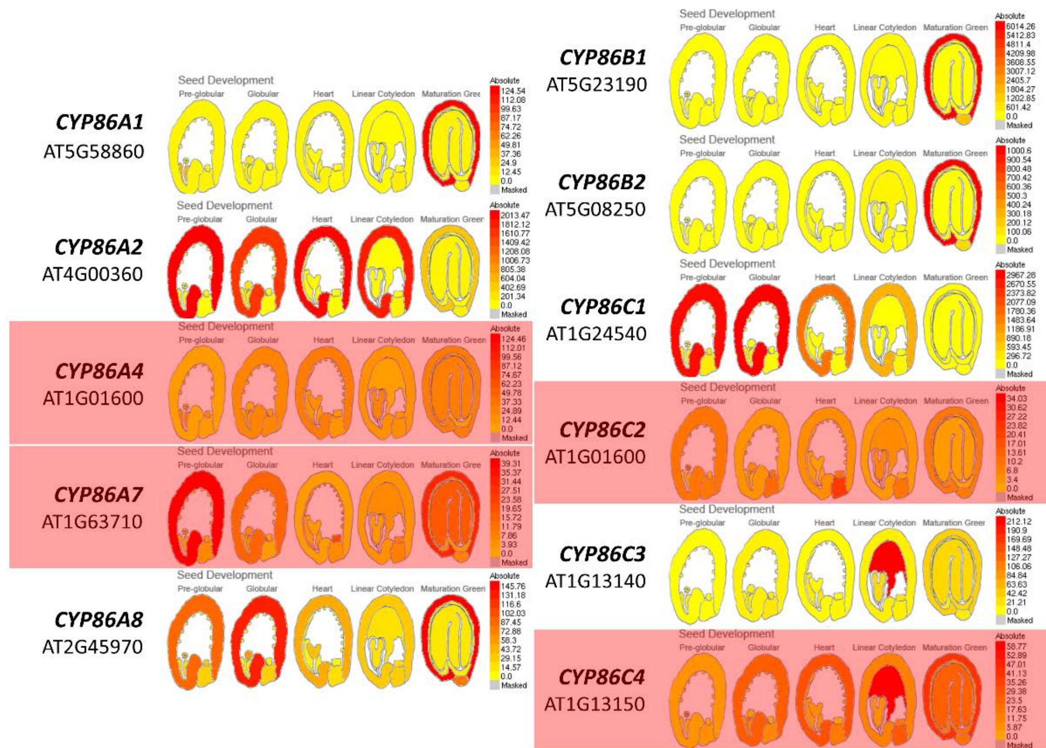


Figure S2. *In silico* expression analysis of the CYP86 gene family during seed development based in Le et al . (2010) data. Images were obtained with the eFP browser (Winter et al ., 2007). Pale red boxes indicate those genes not used in the analysis due to their lower seed-development expression.

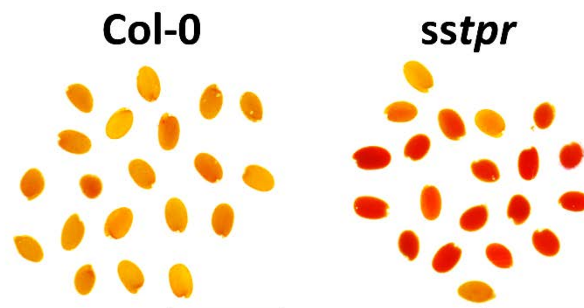


Figure S3. Mutant seeds of sstpr present an increased tetrazolium reduction rate, and thus putatively an increased seed coat permeability, compared to Col-0 seeds. Tetrazolium salt assay performed for 24 h at 28°C. Scale bar: 2 mm.

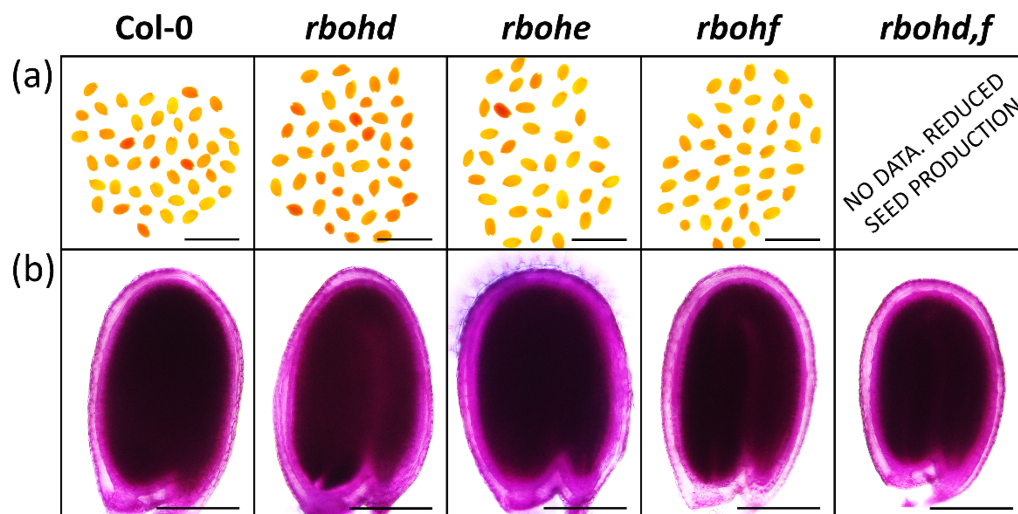


Figure S4. Seed permeability and lipid polyester staining showed no significant differences in rbohd, rbohe, rbohf and double mutant rbohd,f mutant seeds. (a): Tetrazolium salt reduction assay for 24 h. Seeds of rbohd mutant are lightly more permeable than Col-0 seeds, while rbohf seed are lightly less permeable. However, there is not observed a overall behavior in all rboh mutant seeds. Scale bars: 2 mm. (b): Sudan Red staining did not show differences in lipid polyesters barriers that could explain their enhanced seed longevity. Scale bars: 200 µm.

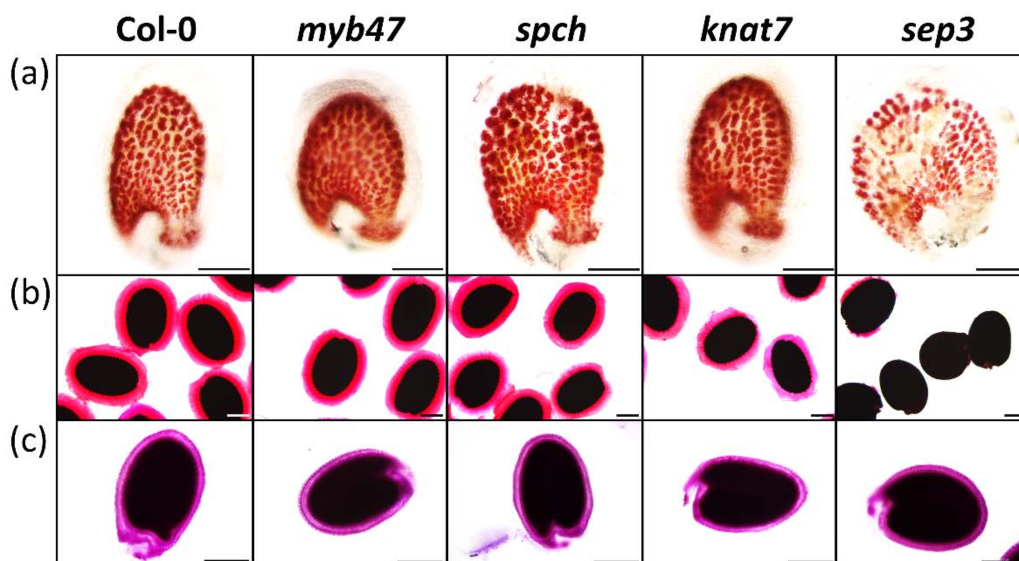


Figure S5. Seed coat analysis in TF mutants seeds of PAs, mucilage halo and lipid polyesters barriers. (a): Vanillin staining in 5 days-after-pollination developing seeds showed no differences in PAs. Little differences observed are related to little differences in pollination times, observed also in same silique seeds. Representative mutant sep3 seed show a rounder shape and a fissure, but there are not highly significant differences in PAs in mutant seeds. Scale bars: 100 µm. (b): Ruthenium red mucilage staining to visualize the mucilage halo. Differences are observed in knat7 and sep3 seeds. As published, knat7 present a reduced halo. Here we can observed a bigger halo compared to previous reports, probably due to a increased incubation time, permitting more mucilage extrusion. Seeds of sep3 mutant present a drastically

reduction of mucilage. Scale bars: 200 µm. (c): Sudan Red staining did not show differences in lipid polyesters barriers that could explain their different seed longevity. Scale bars: 200 µm.

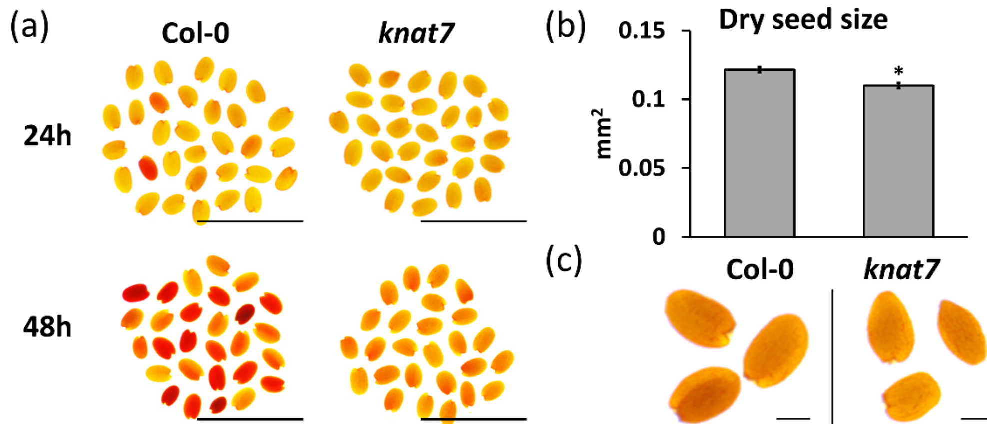


Figure S6. Seeds of knat7 seeds are less permeable and smaller than Col-0 seeds. (a): Tetrazolium salt reduction assay for 24 h and 48 h at 28°C. Scale bars: 2 mm. (b): Seed size analysis. Average measure of projected seed surface of 100 dry seeds. Size analysis was measured with Fiji program. Bars indicate standard errors. *Significantly differing from average wild type seed size at $p < 0.05$ (Student´s t-test). (c): Representative images. Scale bars: 200 µm.

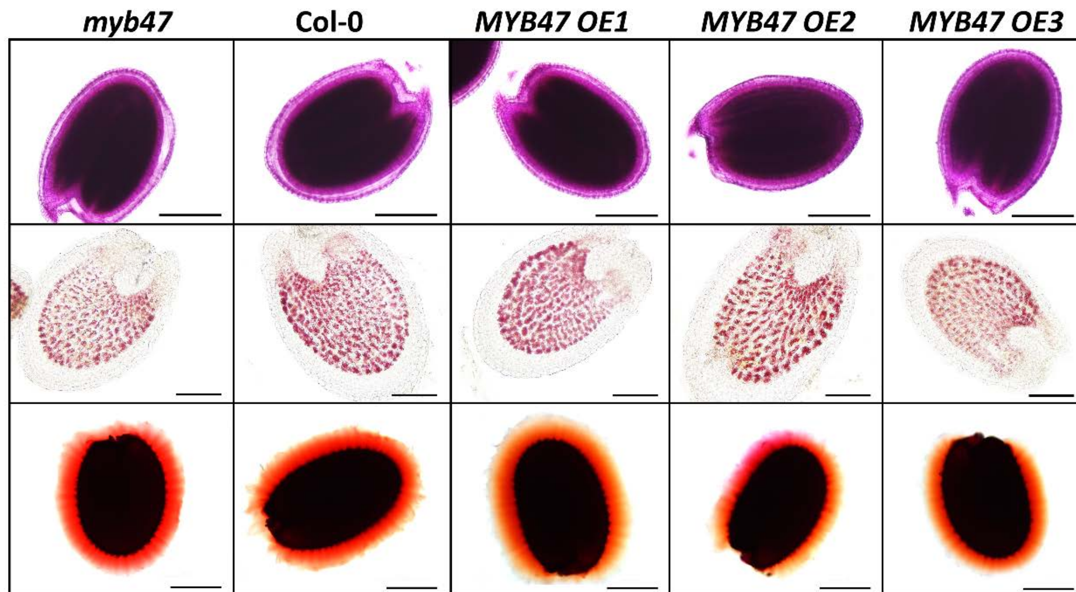


Figure S7. Different dosses of MYB47 do no produced changes in seed coat lipid polyester barriers, PAs or mucilage halo that could explain the different seed longevity observed among different MYB47 mutants. (a): Sudan Red staining did not show differences in lipid polyesters barriers that could explain their different seed longevity. Scale bars: 200 µm. (b): Vanillin staining in 5 days-after-pollination developing seeds showed no differences in PAs. Scale bars: 100 µm. (c): Ruthenium red mucilage staining to visualize the mucilage halo and there are not significant differences. Scale bars: 200 µm.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fpce.13822&file=pce13822-sup-0008-file+1.xlsx

Data S1. Excel file (.xlsx) containing control germination data. Ecotype num and Ecotype name columns provide information about the ecotype (number for GWAS and ecotype name, respectively). Control germ shows the germination ratio in control conditions. Ecotypes rated below 0.6 in control germination ratio were discarded for ageing treatments, indicated in the Discarded column. 270 ecotypes were selected for ageing treatments, indicated at the bottom of the excel sheet.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fpce.13822&file=pce13822-sup-0009-file+2.xlsx

Data S2. Excel file (.xlsx) containing all ecotype ageing treatment data. All treatment data sheet contains all information. Ecotype num and Ecotype name columns provide information about the ecotype (number for GWAS and ecotype name, respectively). Control germ shows the germination ratio in control conditions. uncorrNAT, uncorrAAT, uncorrCDT, uncorrEPPO, columns indicate the observed germination (uncorrected) ratio of each accession under each treatment (NAT, AAT, CDT, and EPPO). NAT , AAT , CDT and EPPO columns show the corrected germination ratio for each ecotype, calculated dividing the uncorrected ratio by the germination ratio of each ecotype. Corrected ratios were used for GWAS analysis. In red letters, there are ecotypes that are not included in the GWAPP Imputed Fullsequence Dataset. Red boxes correspond to missing data of a particular treatment (reduced seed production, plate contamination…). Sheets NAT, AAT, CDT and EPPO contain same information separately for each ageing treatment. Columns 1 and 4 correspond to the GWAPP input data. Sheet No dataset lines list the 11 ecotypes not present in the GWAPP Imputed Fullsequence Dataset.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fpce.13822&file=pce13822-sup-00010-file+3.xlsx

Data S3. Excel file (.xlsx) containing all treatment correlation data. All correlations sheet contains information of correlations of NAT, AAT, CDT and EPPO data with our seed longevity traits (NAT, AAT, CDT, EPPO and Plant life spam), ecotype collection data (longitude and latitude at collection point) and public datasets. Traits are described in four columns: PhenotypeID (unique identificator), Phenotype Name (trait), Scoring (how it has been measured) and variation type (indicates if the variable is continuous, categorical or binary data). The Samples column indicates the number of ecotypes scored for each trait, but it is not an indicative of coinciding ecotype data between published datasets and our data. NAT , AAT , CDT and EPPO columns show the calculated Spearman correlation Rho (rs) number. We consider a strong correlation values above and below 0.3 and − 0.3, respectively. In last column (no named) is indicated the range of coinciding plants of each correlation clade (in rows). As number of ecotypes vary in each seed ageing treatment, we indicate this data approximately. Notice that between our data and ecotype collection data, ecotypes coincides, while in correlation with public datasets ecotypes sets used were different and

there is only an average of 50–100 coinciding ecotypes. Correlation analysis were performed with the easyGWAS comparison tool. In Plant life spam data sheet, we have included the seed collection time data scored in days from showing which was measured each week.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fpce.13822&file=pce13822-sup-00011-file+4.xlsx

Data S4. Excel file (.xlsx) showing all GWAS highlighted genes. Selected genes for reverse genetics study were obtained from this file. Excel file contains four sheets, NAT-GWAS , AAT-GWAS , CDT-GWAS and EPPO-GWAS , to the GWAS analysis of the corresponding seed ageing treatments. GWAPP SNP scoring information was filtered (score ≥ 4 or p value ≤10−4; MAF [minor allele frequency] ≥ 0.05). Genes distanced up to 1.5 kB of sorted SNPs were annotated. Finally, list is ordered by score (−log[p value]). Number of SNPs scoring the same gene has been annotated. In columns: SNP num (number of SNPs close to this gene), max SNP score (score of most significant SNP close to the indicated gene), Gen ID (gene AGI number), Gene alias (short gene name if available) and Gene description (short gene description). NAT, AAT , CDT and EPPO columns indicate if there are filtered SNPs in other GWAS. The number correspond to the score of the most significant SNP from the filtered list. Threshold significance (stablished as p value ≤10−5) is indicated with a black line. Genes associated with non-significant SNPs (only those with p-value between 10−4 and 10−5) are also listed.

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fpce.13822&file=pce13822-sup-00012-file+5.xlsx

Data S5. Excel file (.xlsx) containing ageing treatment assay data (control germination, NAT, CDT and AAT) performed in mutant lines. In columns: Line (indicate the Col-0 background mutant line), Control (control germination, in percentage of germination), error (standard error) and then the different treatments. They follow a similar structure with an additional column for the treatments: the p value column calculated with the t-Student test from three biological replica in comparison to Col-0. Green color correspond to statistical-significant seed ageing resistant lines, red color correspond to statistical-significant seed ageing susceptible lines, and boxes with no color indicate that the mutant has not a statistical-significant altered seed longevity. Significance has been stablished with a $p \leq 0.05$. In addition, in the NAT treatment there is an extra column indicating the number of months of seed dry storage called monthsNAT . Notice that different ageing treatment affect similar to different mutant lines.

Supplemental table 1. Arabidopsis mutant lines used for the reverse genetics study. It is indicated: the AGI number of the affected gene, the line name used in this study, the line reference, the T-DNA insertion location (in T-DNA lines) or the type (in transposon insertion lines), and if the mutant line has been previously reported in a scientific publication.

AGI	Line name	Arabidopsis line	Insertion type	Published
AT4G02770	*psad1*	SALK_036224	300-UTR5	No
AT3G17520	*sslea*	SAIL_681_A07	300-UTR5	No
AT4G02750	*sstpr*	SAIL_208_C05	Exon	No
AT5G47910	*rbohd*	SALK_109396	Exon	No
AT1G19230	*rbohe*	SALK_146126	Exon	No
AT1G64060	*rbohf*	SAIL_1242_C07	Exon	No
Double mutant	*rbohd,f*	Double mutant	Transposons	Torres *et al.* (2002)
AT1G19570	*dhar1*	SALK_005238	Exon	Rahantaniaina *et al.* (2017)
AT1G75270	*dhar2*	SALK_026089	Exon	Rahantaniaina *et al.* (2017)
AT5G16710	*dhar3*	SAIL_435_A09	Intron	Rahantaniaina *et al.* (2017)
Triple mutant	*dhar1,2,3*	Triple mutant	Triple mutant	Rahantaniaina *et al.* (2017)
AT5G58860	*cyp86a1*	GK-055C08	Exon	No
AT4G00360	*cyp86a2*	SALK_005826	Exon	No
AT2G45970	*cyp86a8*	GK-719C03	Exon	No
AT5G23190	*cyp86b1*	SALK_203846	Exon	No
AT5G08250	*cyp86b2*	SALK_070150	Exon	No
AT1G24540	*cyp86c1*	SALK_050565	Exon	No
AT1G18710	*myb47*	SALK_200360	Exon	No
AT5G53210	*spch*	SAIL_36_B06	Exon	MacAlister *et al.* (2007)
AT1G62990	*knat7*	SALK_002098	Intron	Romano *et al*. (2012)
AT1G24260	*sep3-2*	Transposon	Transposon	Pelaz *et al.* (2000)
Triple mutant	*sep1,2,4*	Transposon	Tansposons	Ditta *et al.* (2004)

Supplemental table 2. Primers used in this study.

Primer name	Sequence
psad1-LP	ACCAACACCTTCTCTTCCTGG
psad1-RP	CCAACCCAACATTCAAACAAC
sslea-LP	CATCTCCTGCTTTGCGTTTAG
sslea-RP	TTCCAAATCGTGTGGAAGATC
sstpr-LP	AACCTGCAATCATTGTGTTCC
sstpr-RP	AGTTGTTCGATGAAATGCCTG
rbohd-LP	TTTGATGCCAAACTCCAAGTC
rbohd-RP	GCCGCCTAAGACTTTCTAAGC
rbohe-LP	ATTTTGGCGGCTTAAACTTTG
rbohe-RP	ATTGACTTACCAACGCACTCG
rbohf-LP	ATTCGGTTCGACTGCTTAAGG
rbohf-RP	TTACCCGTTGGTCAAGTTCTG
cyp86a1-LP	AAGTCCTTTCCCAAGCAAGTC
cyp86a1-RP	GGATAATCCCTTCTCTGTCGC
cyp86a2-LP	ATCGAACACATGCTCAAGACC
cyp86a2-RP	GAATTCCAAGCAATCCTCTCC
cyp86a8-LP	TTGCTCCGGTACGTATCAGAC
cyp86a8-RP	TCCGGTAACACATCGTCTTTC
cyp86b1-LP	TTCTGGTCCAAACACAGGTTC
cyp86b1-RP	GTTTCTTGGGATGCTTCCTTC
cyp86b2-LP	TAATCCCCTCCAAAAGATTCG
cyp86b2-RP	CCTCGTCCGGATTACTTCTTC
cyp86c1-LP	ATGTTGGCGCTAAACAAACAC
cyp86c1-RP	AATCCTAGCGCCTTTCTTCAC
myb47-LP	CCTCTTTTGCTTGATCACTGC
myb47-RP	AAAATGATTTCCATCTTTTAACGG
spch-LP	TTTCCCTTTGCAATATGCAAC
spch-RP	GGACTTCGGCGTAGGTTTTAC
knat7-LP	GAGATTAGTGTTTGCGCTTGG
knat7-RP	TATGCGTAAGGGCATATCAGG
sep3-RP	GTTGGGAAAATCGTACGAGGCTTCACCTAGT
sep3-LP	CGTCACTTGCCTATTGATCTTGTT
SALK-LP	ATTTTGCCGATTTCGGAAC
SAIL-LP	GCTTCCTATTATATCTTCCCAAATTACCAATACA
GABI-LP	ATATTGACCATCATACTCATTGC
proMYB47-F	ATTATATTTAATCTCCATGTGGTCG
proMYB47-R	TTTCTCCAGCCAACCAAAC
MBY47-F	ATGGGGAGGACGACAT
MYB47-R	TCAAAAGAGATGATCAAGTATGTC
MYB47-qRT-F	TCTCGGAAACAGGTGGGCAG
MYB47-qRT-R	GATGATGGGCTCGTGGGTCA