Desiccation sensitive seeds:
Understanding their evolution, genetics and physiology

Alexandre Correia Silva de Santana Marques

Thesis committee

Promotor
Prof. Dr Richard G.H. Immink
Special professor Physiology of Flower Bulbs
Wageningen University & Research

Co-promotors
Dr Henk W.M. Hilhorst
Associate professor, Laboratory of Plant Physiology
Wageningen University & Research

Dr Wilco Ligterink
Assistant professor, Laboratory of Plant Physiology
Wageningen University & Research

Other members
Prof. Dr Bas J. Zwaan, Wageningen University & Research
Dr Koen Verhoeven, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen
Dr Peter E. Toorop, Monsanto Vegetable Seeds Division, Bergschenhoek
Dr Wouter Kohlen, Wageningen University & Research

Desiccation sensitive seeds:
Understanding their evolution, genetics and physiology

Alexandre Correia Silva de Santana Marques

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 22 May 2018
at 11 a.m. in the Aula.

Preface

In tropical rainforests we can find a huge diversity of fruits hanging from the trees. Mango, lychee, cacao, avocado, rambutan, mangosteen, jackfruit, coconut, citrus, inga, durian, iamb, and nectandra are just a few examples. Humanity seems not to notice that these forests are shrinking daily. For example, there is only 5% left of the original Brazilian Atlantic forest. The Amazon forest is still there because it is more isolated, but it has been severely deforested and 80% of its original vegetation is left. The main cause of deforestation is farming, by which forest is replaced by grass fields. This consumes huge areas of land. I hope that the technologies we are currently developing will change the system of food production soon and bring a halt to deforestation.

What can be done after deforestation? Re-forestation. There are seeds stored in seed banks. We can simply grow various plant species back. However, almost up to 50% of the tree species in tropical rain forest develop desiccation sensitive and therefore non-storable seeds. Without an effective method for storing these seeds, we will lose much of plant biodiversity. In the worst case scenario, we and next generations will lose for example, the possibility of eating the fruits I cited above, because these plants produce all non-storable seeds.

We must investigate this problem, to understand the physiology and genetics of these plants in order to support the development of methods to overcome these storability issue. The work described in this PhD thesis is a step towards the understanding of the seed desiccation sensitivity phenomenon.

I sincerely hope that I managed to raise your attention for the beauty, richness and fragility of plant biodiversity on our planet.

Please, let's all do our best to protect biodiversity!

"When the last tree is cut down, the last fish eaten, and the last stream poisoned, you will realize that you cannot eat money" (Cree Indian proverb).

Contents

1

General introduction

Time travel is a fascinating quest explored by many sci-fi writers. There is an itchy desire in each of us to discover what will happen in the future. Gipe et al. (1985) invested in physics to develop a machine capable to distort time and take us "back to the future", although without concrete success till the present time. Other strategies have been conjured up much earlier, consisting e.g. of maintaining the human body desiccated by mummification. This strategy is still not successful in calling dead bodies to life, but freezing approaches are under investigation (Merkle, 1992). Although freezing and calling to life of complete human bodies is still not feasible, freezing cells like spermatozoids, ovules and embryos are medical routines. Thus, time travel is already possible for humans, at least from a solely biological point of view.

Similar to freezing and capturing vital cells, many organisms travel through time using their own strategies. Desiccation tolerance[1] of plant seeds is one of these strategies, which consists of the capacity of seeds to survive extreme water losses (below 10% of the fresh weight or 0.1 g H_2O/g dry weight).

Desiccation tolerance is widespread across the propagules of diverse life forms, e.g. bacteria, fungi, protozoa, animals and plants (Crowe, 2014). It assures that these organisms will flourish as soon as water availability returns. In contrast to desiccation tolerance, dormancy is a genetically programmed mechanism that arrests the development of the propagules even when optimal conditions are present and this is also found in diverse life forms. desiccation tolerance, together with dormancy, aims to guarantee the survival of species by spreading propagules through long periods of time and thus avoiding extinction.

Evolutionary history of seed desiccation tolerance and sensitivity

Desiccation tolerance capacitated the first land plants to conquer terrestrial habitats and it is largely present in vegetative tissues of basal plants and in desiccation tolerant seeds (also termed 'orthodox') of spermatophytes (Black and Pritchard, 2002). The evolution of terrestrial plants is believed to be monophyletic and originating from charophycean algae (Zhong et al., 2015). Desiccation tolerance was already present in these algae and is also believed to be carried from these algae towards the vegetative tissues of

---

[1] One needs to distinguish desiccation from drought. Drought characterizes the lack of moisture in the soil (Iljin, W. (1957) Drought resistance in plants and physiological processes. Annual Review of Plant Physiology, **8**, 257-274. while desiccation denotes the lack of moisture in the organism (Moore, J.P., Vicre-Gibouin, M., Farrant, J.M., Driouich, A. (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol Plant, **134**, 237-245.. The lack of available water suspends virtually every metabolic activity in the cells, such as respiration and transcription. Thus, it is also different from dormancy or lethargy that reduces in a wet environment the progress of development or of metabolic activity.

bryophytes and pteridophytes and later on confined to the seeds of the spermatophytes (Oliver et al., 2000). The monophyletic origin of desiccation tolerance hypothesis is supported by the fact that the regulation of desiccation tolerance involves the phytohormone abscisic acid (ABA) and the transcription factor ABI3 in vegetative tissues of the moss *Physcomitrella patens,* just as in seeds of numerous angiosperm plants (Khandelwal et al., 2010).

Although desiccation tolerance is present in seeds of 95% of the angiosperms, desiccation sensitivity is present in almost half of those present in tropical rainforests (Tweddle et al., 2003). Among them, there is a broad phenotypical variation that has been categorized in 'recalcitrant' and 'intermediate' seeds with the latter losing viability at lower water contents than the former (Berjak and Pammenter, 2013).

It has been suggested that desiccation sensitive (DS)-seeded species evolved from desiccation tolerant (DT) ones adapting to environments where the conditions were conducive to immediate germination, e.g. tropical rainforests with abundant and constant water regimes. In these environments, desiccation tolerance is not an adaptive trait and it can eventually get lost in these species (Berjak and Pammenter, 2013; Farrant et al., 1993a).

Depending on the environment, DS-seeded species may show a gradient of dehydration tolerance. For example, the seeds of *Acer pseudoplatanus* and *Coffea arabica* display varying dehydration tolerance levels across their native plant distribution ranges (Daws and Pritchard, 2008; Ellis et al., 1991). The intraspecific variability observed for seed desiccation tolerance reflects the high plasticity of this trait and its possible role in species adaptation to different environments. However, there is no report, to our knowledge, demonstrating the presence of DT and DS seeds in the same species. However, the presence of both phenotypes in species of the same genus has been clearly demonstrated (Dussert et al., 2000). We thus expect that genetic changes associated with seed desiccation sensitivity occurred during evolution of these species.

The long standing categorization of seed storage behaviour in 'orthodox' (DT) or 'recalcitrant' (DS) suggests that it is a qualitative phenotype (Roberts, 1973). However, the 'intermediate' seed storage category was suggested to make sense based on the observation of a wide range of desiccation sensitivity in nature (Ellis et al., 1990). Since then, researchers included one more qualitative category describing seed storability and classified seeds as 'orthodox', 'intermediate' or 'recalcitrant'. Nevertheless, new studies have shown large continuity of seed desiccation tolerance across different species and some authors have suggested that we should regard seed storage behaviour as a continuum (Berjak and Pammenter, 2002).

Studies regarding the regulation of seed desiccation tolerance have been executed with the model plants *Arabidopsis thaliana* and *Medicago truncatula* (Dekkers et al., 2015). To date, three transcription factors (TFs) essential for desiccation tolerance acquisition

have been identified, namely ABSCISIC ACID INSENSITIVE 3 (ABI3), LEAFY COTYLEDON 1 (LEC1) and FUSCA 3 (FUS3). These TFs are regulators of embryogenesis and seed maturation and elimination of their activities, e.g. through knock-out mutations, results in DS seeds (To et al., 2006). These mutants also show a lack of dormancy and display precocious germination, that are also widespread phenotypes among wild DS-seeded species (Farnsworth, 2000; Raz et al., 2001). One could expect, based on these studies, that wild DS-seeded species would also display similar mutations and show an abrupt phenotypical transition. In this way, desiccation tolerance could be addressed as a qualitative trait.

The seeds of the Arabidopsis *abi3-5* mutant are short-lived, losing viability after one week, and thus considered DS. Overexpression of key mis-expressed TFs, such as DREB2A, PLATZ, AGL67 and EIL5, partially rescued desiccation tolerance in *abi3-5* seeds improving their longevity (González-Morales et al., 2016). The partial rescue of desiccation tolerance by overexpression of these TFs supports the hypothesis that desiccation tolerance is built by factors that additively increase seed desiccation tolerance resulting in better longevity. In this respect, we could indeed address seed desiccation tolerance as a quantitative phenotype.

The level of seed desiccation tolerance in F1 hybrids of *Coffea pseudozanguebariae* (DT) and *Coffea liberica* (DS) corresponded to that of the mid-parent, indicating an additive inheritance of seed desiccation tolerance in this genus (Dussert et al., 2004).

DS seeds represent an obstacle for the conservation of these species as they are difficult to be stored in seed banks (Berjak and Pammenter, 2013). Because of the importance of seed desiccation sensitivity for adaptation of these plants to their native environments and for conservation of these plants, we pose the following questions.

Does lack of developmental arrest, maturation drying and metabolic quiescence result in seed desiccation sensitivity?

It has been hypothesized that DS seeds lack the phases of maturation drying and metabolic quiescence, progressing directly towards germination phase (Ellis et al., 1991; Farnsworth, 2000; Farrant et al., 1993b). This hypothesis is supported by the fact that DT seeds lose desiccation tolerance upon radicle protrusion (Dekkers et al., 2015). Similarly, wild viviparous species are predominantly DS (Farnsworth, 2000).

Low dormancy levels in DT seeds may cause vivipary when subjected to moist conditions on the mother plant. Due to the low dormancy level of these seeds, development is not arrested, which results in precocious germination when still attached to the mother plant at the end of the maturation (Finch-Savage, 2013).

Although the majority of the non-dormant seeded species are DT (Tweddle et al., 2003), desiccation during maturation arrests their development due to metabolic quiescence and they will germinate immediately once imbibed. Alternatively, DS seeds do not dry, and can progress towards germination due to an active metabolism and lack of dormancy. Arabidopsis knockout mutants in the transcription factors ABI3, FUS3 and LEC1 also have DS seeds, a lack of dormancy and premature germination inside the siliques (Raz et al., 2001). Furthermore, the *abi3-5* Arabidopsis knockout mutant has embryos with an enlarged apical meristem, revealing the continuation of embryo growth (Rohde et al., 2000).

DS is more frequent in seeds that are non-dormant upon shedding (c. 31%), than dormant (c. 9%) (Tweddle et al., 2003). However, deep seed dormancy can be observed in DS seeded-species, such as *Aesculus hippocastanum* and *Garcinia cowa* (Liu et al., 2005; Obroucheva et al., 2004). Thus, developmental arrest and seed DS may co-exist. It is unclear and it demands further investigation at which stage development is arrested in these seeds. Similarly, seedlings may display development arrest after exogenous ABA treatment (Lopez-Molina et al., 2002) but desiccation tolerance is no longer inducible at this point. Exogenous ABA treatment can arrest the development of protruded Arabidopsis radicles and re-induce desiccation tolerance but only prior to root hair appearance (Maia et al., 2014). Thus, likewise, wild DS dormant seeds may be developmentally arrested at later developmental stages where desiccation tolerance is no longer inducible.

*** What are the genetic alterations among DS-seeded species?***

Since Roberts (1973) first described the phenomenon of seed desiccation sensitivity, which he named 'recalcitrant' seeds, authors have formulated hypotheses to explain its evolutionary, physiological and genetic origin(s). One of the early hypotheses attempting to address the molecular cause of seed desiccation sensitivity was proposed by Farrant et al. (1993b). The authors suggested that DS seeds lack the genes or competence to produce LEA proteins and, as a result, remain sensitive to desiccation.

Since then, several studies have demonstrated the role of proteins involved in the protection against desiccation in DS-seeded species: the accumulation of dehydrins was associated with the induction of desiccation tolerance in the DS seeds of *Acer saccharinum* (Beardmore and Whittle, 2005). An under-representation of LEAs was observed in the transcriptome of the DS seeds of *Castanospermum australe* as compared to taxonomically related, DT-seeded, *Medicago truncatula* (Delahaie et al., 2013). DS-associated genomic features were studied in the DS-seeded seagrass *Zostera marina,* revealing underrepresentation of LEA genes in its genome (Olsen et al., 2016).

Desiccation tolerance could be induced by hormonal manipulations in the DS seeds of *Acer saccharinum* (Beardmore and Whittle, 2005). Thus, desiccation tolerance associated genes were not lost in this species. It seems that the desiccation tolerance mechanism was skipped during seed development and that the hormonal treatment was able to recruit it back.

At first glance, the most remarkable morphological differences that set DS-seeded plants apart from DT ones, are generally larger fruits, larger seeds, higher seed water content at shedding and thinner seed coats in relation to seed size (Daws et al., 2006; Hamilton et al., 2013). Furthermore, the majority of species with green embryos in East-Australian rainforest are DS and have lower oil content (Hamilton et al., 2013). There is a generally significant correlation between these traits and seed DS. However, none of these are unique and true determinants of DS seeds as there are DT seeds that are big, green and have low oil content.

The acquisition of desiccation tolerance in seeds has been associated with the following biochemical/molecular mechanisms: (1) protection of proteins by late embryogenesis abundant (LEA) proteins and small heat shock proteins (sHSPs), and immobilization of the cytoplasm in a glassy state by accumulation of non-reducing sugars; (2) prevention of oxidative damage by a range of antioxidant compounds and enzymes such as tocopherols and glutathione; (3) prevention of structural damage by cell wall relaxation and partitioning of amphiphilic compounds into membranes, reorganization of intracellular membranes and cytoskeleton, vacuolization and chromatin condensation; (4) Regulation of pathways controlling the induction of these protective mechanisms (Crowe, 2014; Hoekstra et al., 2001).

It is currently unclear which of the above-mentioned mechanisms are altered in DS seeds. DS-seeded species are not phylogenetically related (Pammenter and Berjak, 2000). Thus, each DS-seeded species is probably a result of unique unrelated genetic events. However, a genetic bottleneck is possible and thus we could expect convergent evolutionary events taking place due to a restricted number of possible genetic alterations.

Are hormonal biosynthesis or signalling altered in desiccation sensitive seeds?

The cause of the gradual loss of desiccation tolerance during the progress of germination of DT seeds has been associated with the loss of ABA sensitivity of the dedifferentiating growing tissues (Maia et al., 2014). Furthermore, seeds of the Arabidopsis mutant *aba2-1*, which have very low ABA levels, are not dormant (Léon-Kloosterziel et al., 1996) and the chemical inhibition of ABA biosynthesis can even cause vivipary in DT seeds of maize (White et al., 2000). Likewise, the vivipary of *Avicenia marina* seeds was associated with their low ABA levels (Farrant et al., 1993b).

An important transducer of the ABA signal, and a key regulator of seed dormancy and desiccation tolerance, is the B3 domain containing transcriptional regulator known as VIVIPAROUS 1 (VP1) in maize (McCarty et al., 1989) or ABI3 in Arabidopsis (Giraudat et al., 1992; Koornneef et al., 1984). VP1/ABI3 is conserved among many plant species (Graeber et al., 2010) and is implicated in dormancy and desiccation tolerance in, e.g. Arabidopsis, maize and wheat. Thus, the roles of ABA and ABI3 in seed dormancy and desiccation tolerance are conserved among diverse species and its origin dates back to mosses where it is also important for desiccation tolerance acquisition (Koornneef et al., 1984; Marella et al., 2006; McCarty, 1995; McKibbin et al., 2002).

The incubation in ABA plus tetcyclacis, an inhibitor of gibberellin (GA) biosynthesis and ABA catabolism, induced desiccation tolerance in the DS seeds of *Acer saccharinum* (Beardmore and Whittle, 2005). The tetcyclacis induction of desiccation tolerance in these seeds indicates that not only the lack of ABA is responsible for its desiccation sensitivity, but also relatively high GA levels play an important role. This is in agreement with the hypothesis of the ABA/GA balance as a mode of action for regulation of physiological processes (Shu et al., 2015). The ABA/GA balance has also been shown to regulate the germination versus maturation programs in maize. The seeds of the maize mutant *vp5* have very low ABA levels, are DS and viviparous but with the suppression of GA-biosynthesis, viviparity was avoided and desiccation tolerance recovered (White et al., 2000).

Although relevant research has been done to address the molecular mechanisms involved in seed desiccation sensitivity, we are far from understanding this complex process and various research questions remained unanswered. In this thesis, we aimed at shedding light on the physiological and molecular mechanisms underlying the seed desiccation sensitivity phenomenon.

Layout of this thesis

DS seeds represent an obstacle for the conservation of these species as they are difficult (recalcitrant) to store in seed banks (Berjak and Pammenter, 2013). Furthermore, Brazil, which is the main sponsor of this project, has the largest biodiversity of trees in the world (Beech et al., 2017). Thus, I decided to study the phenomenon of seed desiccation sensitivity, to understand it and pave the way for a feasible solution to the difficulty of long-term storage of DS seeds.

We explored the broad phenotypical variation of DS seeds. The DT-, intermediate and DS-seeded species *Arabidopsis thaliana*, *Citrus limon* and *Castanospermum australe*, respectively, were the models targeted in the present thesis. The primary theme of this

thesis entails the molecular mechanisms associated with seed desiccation sensitivity. In **Chapter 2** I review the literature concerning the genetics, physiology and evolutionary aspects of DS seeds, aiming to integrate the results of previous studies in order to learn and develop new perspectives. Based on meta-analysis of available data on the seed's lowest safe water content, I revealed two remarkable limits for dehydration survival, which sheds light on the controversial separation of the categories 'recalcitrant' and 'intermediate' seeds. In this chapter, I also assembled the available knowledge to draw a generalized picture of the evolution of DS-seeded species, placing it in an ecological context and zooming in on the molecular mechanisms underlying it.

Observing a lack of knowledge on transcriptomes of seed development and maturation in relation with desiccation tolerance, I decided to compare the Arabidopsis *abi3-6* mutant, as a model for seed desiccation sensitivity, with DT-seeded wild-type Arabidopsis. Arabidopsis has a long list of attributes that makes it an ideal model for this study. It has for example a very fast life cycle, small size, and a small and well annotated genome. Moreover, it is the most studied plant which allows diverse comparative analyses. Together with *fus3* and *lec1* mutants, the *abi3-6* mutant is one of a few known single mutants that produce DS seeds. It resembles phenotypically most wild DS seeds, due to its green seeds, lack of dormancy, precocious germination and low oil content. In **Chapter 3** we investigated relevant associations between physiology and gene expression alterations throughout late seed development of *abi3-6* in comparison with wild-type seeds.

I asked if the gene expression alterations observed during seed development of the Arabidopsis *abi3-6* mutant could also be observed in a wild DS-seeded species. For this, I studied, in **Chapter 4**, the DS seeds of the basal legume *Castanospermum australe*. In this chapter I not only explored the transcriptome of development of *C. australe* seeds, but I also went further, looking into the origins of seed desiccation sensitivity using comparative genomics of this species with diverse other DS- and DT-seeded species. We decided to sequence the *C. australe* genome and use it as our model for DS seed due to its interesting phylogenetic position in the legume family. Legumes encompass several crop and model species with sequenced genomes and multiple seed-based studies.

In **Chapter 5**, I made attempts to induce desiccation tolerance in the DS seeds of *Citrus limon*. I describe a protocol of hormonal manipulation for induction of desiccation tolerance and studied the associated transcriptomes.

In **Chapter 6** I discussed and integrated all the results obtained and suggest future strategies to further explore and develop research on seed desiccation sensitivity.

Beardmore, T. and Whittle, C.-A. (2005) Induction of tolerance to desiccation and cryopreservation in silver maple (Acer saccharinum) embryonic axes. Tree physiology, **25**(8), 965-972.

Beech, E., Rivers, M., Oldfield, S., Smith, P. (2017) GlobalTreeSearch: the first complete global database of tree species and country distributions. Journal of Sustainable Forestry, 1-36.

Berjak, P. and Pammenter, N. (2002) Orthodox and recalcitrant seeds. Tropical tree seed manual, **721**, 137-147.

Berjak, P. and Pammenter, N.W. (2013) Implications of the lack of desiccation tolerance in recalcitrant seeds. Front Plant Sci, **4**, 478.

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

Crowe, J.H. (2014) Anhydrobiosis: an unsolved problem. Plant, Cell & Environment, **37**(7), 1491-1493.

Daws, M.I., Garwood, N.C., Pritchard, H.W. (2006) Prediction of desiccation sensitivity in seeds of woody species: a probabilistic model based on two seed traits and 104 species. Ann Bot, **97**(4), 667-674.

Daws, M.I. and Pritchard, H.W. (2008) The development and limits of freezing tolerance in Acer pseudoplatanus fruits across Europe is dependent on provenance. CryoLetters, **29**(3), 189-198.

Dekkers, B.J., Costa, M.C.D., Maia, J., Bentsink, L., Ligterink, W., Hilhorst, H.W. (2015) Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. Planta, **241**(3), 563-577.

Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H., Buitink, J. (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. J Exp Bot, **64**(14), 4559-4573.

Dussert, S., Chabrillange, N., Engelmann, F., Anthony, F., Louarn, J., Hamon, S. (2000) Relationship between seed desiccation sensitivity, seed water content at maturity and climatic characteristics of native environments of nine Coffea L. species. Seed Science Research, **10**(03), 293-300.

Dussert, S., Engelmann, F., Louarn, J., Noirot, M. (2004) Inheritance of seed desiccation sensitivity in a coffee interspecific cross: evidence for polygenic determinism. Journal of experimental botany, **55**(402), 1541-1547.

Ellis, R., Hong, T., Roberts, E. (1990) An intermediate category of seed storage behaviour? I. Coffee. Journal of experimental botany, **41**(9), 1167-1174.

Ellis, R., Hong, T., Roberts, E. (1991) An intermediate category of seed storage behaviour? II. Effects of provenance, immaturity, and imbibition on desiccation-tolerance in coffee. Journal of experimental botany, **42**(5), 653-657.

Farnsworth, E. (2000) The ecology and physiology of viviparous and recalcitrant seeds. Annual Review of Ecology and Systematics, 107-138.

Farrant, J.M., Berjak, P., Cutting, J., Pammenter, N. (1993a) The role of plant growth regulators in the development and germination of the desiccation-sensitive (recalcitrant) seeds of Avicennia marina. Seed Science Research, **3**(01), 55-63.

Farrant, J.M., Pammenter, N.W., Berjak, P. (1993b) Seed development in relation to desiccation tolerance: A comparison between desiccation-sensitive (recalcitrant) seeds of Avicennia marina and desiccation-tolerant types. Seed Science Research(3), 1-13.

Finch-Savage, B. (2013) Seeds: Physiology of development, germination and dormancy, Cambridge University Press.

Gipe, G., Nicolaas, T.T.M., Zemeckis, R., Gale, B. (1985) *Back to the Future* Luitingh.

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F., Goodman, H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. The Plant Cell Online, **4**(10), 1251-1261.

González-Morales, S.I., Chávez-Montes, R.A., Hayano-Kanashiro, C., Alejo-Jacuinde, G., Rico-Cambron, T.Y., de Folter, S., Herrera-Estrella, L. (2016) Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 201610985.

Graeber, K., Linkies, A., Muller, K., Wunchova, A., Rott, A., Leubner-Metzger, G. (2010) Cross-species approaches to seed dormancy and germination: conservation and biodiversity of ABA-regulated mechanisms and the Brassicaceae DOG1 genes. Plant Mol Biol, **73**(1-2), 67-87.

Hamilton, K.N., Offord, C.A., Cuneo, P., Deseo, M.A. (2013) A comparative study of seed morphology in relation to desiccation tolerance and other physiological responses in 71 Eastern Australian rainforest species. Plant Species Biology, **28**(1), 51-62.

Hoekstra, F.A., Golovina, E.A., Buitink, J. (2001) Mechanisms of plant desiccation tolerance. Trends in plant science, **6**(9), 431-438.

Iljin, W. (1957) Drought resistance in plants and physiological processes. Annual Review of Plant Physiology, **8**(1), 257-274.

Khandelwal, A., Cho, S., Marella, H., Sakata, Y., Perroud, P.-F., Pan, A., Quatrano, R. (2010) Role of ABA and ABI3 in desiccation tolerance. Science, **327**(5965), 546-546.

Koornneef, M., Reuling, G., Karssen, C. (1984) The isolation and characterization of abscisic acid-insensitive mutants of Arabidopsis thaliana. Physiologia Plantarum, **61**(3), 377-383.

Léon-Kloosterziel, K.M., Gil, M.A., Ruijs, G.J., Jacobsen, S.E., Olszewski, N.E., Schwartz, S.H., Zeevaart, J.A., Koornneef, M. (1996) Isolation and characterization of abscisic acid-deficient Arabidopsis mutants at two new loci. The Plant Journal, **10**(4), 655-661.

Liu, Y., Qiu, Y.P., Zhang, L., Chen, J. (2005) Dormancy breaking and storage behavior of Garcinia cowa Roxb.(Guttiferae) seeds: implications for ecological function and germplasm conservation. Journal of Integrative Plant Biology, **47**(1), 38-49.

Lopez-Molina, L., Mongrand, S., McLachlin, D.T., Chait, B.T., Chua, N.H. (2002) ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. The Plant Journal, **32**(3), 317-328.

Maia, J., Dekkers, B.J., Dolle, M.J., Ligterink, W., Hilhorst, H.W. (2014) Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated Arabidopsis seeds. New Phytol, **203**(1), 81-93.

Marella, H.H., Sakata, Y., Quatrano, R.S. (2006) Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from Physcomitrella patens. The Plant Journal, **46**(6), 1032-1044.

McCarty, D.R. (1995) Genetic control and integration of maturation and germination pathways in seed development. Annual review of plant biology, **46**(1), 71-93.

McCarty, D.R., Carson, C.B., Stinard, P.S., Robertson, D.S. (1989) Molecular analysis of viviparous-1: an abscisic acid-insensitive mutant of maize. Plant Cell, **1**(5), 523-532.

McKibbin, R.S., Wilkinson, M.D., Bailey, P.C., Flintham, J.E., Andrew, L.M., Lazzeri, P.A., Gale, M.D., Lenton, J.R., Holdsworth, M.J. (2002) Transcripts of Vp-1 homeologues are misspliced in modern wheat and ancestral species. Proceedings of the National Academy of Sciences, **99**(15), 10203-10208.

Merkle, R.C. (1992) The technical feasibility of cryonics. Medical hypotheses, **39**(1), 6-16.

Moore, J.P., Vicre-Gibouin, M., Farrant, J.M., Driouich, A. (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol Plant, **134**(2), 237-245.

Obroucheva, N., Antipova, O., Azarkovich, M., Gumilevskaya, N. (2004) Analysis of the growth capacity of embryo axes during seed dormancy and dormancy release. In Doklady Biological Sciences, Springer: pp 218-220.

Oliver, M.J., Tuba, Z., Mishler, B.D. (2000) The evolution of vegetative desiccation tolerance in land plants. Plant Ecology, **151**(1), 85-100.

Olsen, J.L., Rouzé, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., Dattolo, E., De Paoli, E., Dittami, S., Maumus, F. (2016) The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. Nature, **530**(7590), 331-335.

Pammenter, N. and Berjak, P. (2000) Some thoughts on the evolution and ecology of recalcitrant seeds. Plant Species Biology, **15**(2), 153-156.

Raz, V., Bergervoet, J., Koornneef, M. (2001) Sequential steps for developmental arrest in Arabidopsis seeds. Development, **128**(2), 243-252.

Roberts, E.H. (1973) Predicting the storage life of seeds. In Proceedings.

Rohde, A., De Rycke, R., Beeckman, T., Engler, G., Van Montagu, M., Boerjan, W. (2000) ABI3 affects plastid differentiation in dark-grown Arabidopsis seedlings. The Plant Cell, **12**(1), 35-52.

Shu, K., Chen, Q., Wu, Y., Liu, R., Zhang, H., Wang, P., Li, Y., Wang, S., Tang, S., Liu, C. (2015) ABI4 mediates antagonistic effects of abscisic acid and gibberellins at transcript and protein levels. The Plant Journal.

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., Parcy, F. (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. Plant Cell, **18**(7), 1642-1651.

Tweddle, J.C., Dickie, J.B., Baskin, C.C., Baskin, J.M. (2003) Ecological aspects of seed desiccation sensitivity. Journal of Ecology, **91**(2), 294-304.

White, C.N., Proebsting, W.M., Hedden, P., Rivin, C.J. (2000) Gibberellins and seed development in maize. I. Evidence that gibberellin/abscisic acid balance governs germination versus maturation pathways. Plant physiology, **122**(4), 1081-1088.

Zhong, B., Sun, L., Penny, D. (2015) The origin of land plants: a phylogenomic perspective. Evolutionary bioinformatics online, **11**, 137.

2

Evolutionary ecophysiology of seed desiccation sensitivity

Alexandre Marques, Gonda Buijs, Wilco Ligterink and Henk Hilhorst

Abstract. Desiccation sensitive (DS) seeds do not survive dry storage due to their lack of desiccation tolerance. Almost half of the plant species in tropical rainforests produce DS seeds and therefore the desiccation sensitivity of these seeds represents a problem for and long-term biodiversity conservation. This phenomenon raises questions as to how, where and why DS (desiccation sensitive)-seeded species appeared during evolution. These species evolved probably independently from desiccation tolerant (DT) seeded ancestors. They adapted to environments where the conditions are conducive to immediate germination after shedding, e.g. constant and abundant rainy seasons. These very predictable conditions offered a relaxed selection for desiccation tolerance that eventually got lost in DS seeds. These species are highly dependent on their environment to survive and they are seriously threatened by deforestation and climate change. Understanding of the ecology, evolution and molecular mechanisms associated with seed desiccation tolerance can shed light on the resilience of DS-seeded species and guide conservation efforts. In this review, we survey the available literature for ecological and physiological aspects of DS-seeded species and combine it with recent knowledge obtained from DT model species. This enables us to generate hypotheses concerning the evolution of DS-seeded species and their associated genetic alterations.

Additional keyword: dormancy, intermediate seeds, orthodox seeds, seed size

Introduction

Seeds are important for plants as means of regeneration and to humans as the primary basis for nutrition. Each seed contains an embryo which is usually formed by sexual reproduction and thus is the means by which a new individual is dispersed, playing a crucial role in the life history of higher plants (Bewley *et al.* 2013). The biological and economic relevance of seeds make it evident that it is of significant importance to conserve them and, hence, to conserve biodiversity. The current most efficient and feasible way to conserve plant genetic resources is by storing desiccated seeds at low temperatures (Li and Pritchard 2009).

The capacity to survive desiccation was a decisive evolutionary step that allowed the first plants to colonize the dry terrestrial environment. Many of the mosses and ferns have vegetative desiccation tolerance whereas gymnosperms and angiosperms possess desiccation tolerance that is generally expressed only in seeds and/or pollen (Alpert and Oliver 2002). It is estimated that 95% of the spermatophytes have DT seeds, which allows the effective preservation of their genetic diversity in seed banks (Tweddle *et al.* 2002).

However, not all seeds can be stored *ex situ* in a dry state. Termed 'recalcitrant' for this reason (Roberts, 1973), such seeds are DS and consequently do not survive desiccation (and freezing) required for germplasm conservation (Pammenter and Berjak 2000). They may represent up to 45% of the plant species in tropical evergreen rain forests (Tweddle *et al.* 2003; Hamilton *et al.* 2013).

Plant species that produce DS seeds evolved and adapted to environments in which the conditions are conducive to immediate germination in the same season when seeds are shed, e.g. in rainforests with a constant and abundant rainy season, or in temporarily or permanently flooded plains. Evidence suggests that such seeds are programmed to initiate germination upon or shortly after shedding (Farrant *et al.* 1993; Berjak and Pammenter 2013) and thus the loss of desiccation tolerance in such species during evolution was presumably to favour faster seedling establishment. This adaptive trait, however, has made such seeds highly dependent on their habitat for survival. These environments are shrinking worldwide due to deforestation and climate change (Achard *et al.* 2002). As a result, plant species with DS seeds face imminent extinction risk. Despite this, to date, the phenomenon of seed desiccation sensitivity is poorly understood (Berjak *et al.* 2008; Berjak and Pammenter 2013). Therefore, a better understanding of seed desiccation sensitivity is of great importance to guide priorities in conservation efforts. Furthermore, the knowledge acquired from these studies can help to shed light on components that are necessary for the acquisition of DT, on how and where it evolved and was lost from the DS seeds.

Ecological aspects of desiccation-sensitive seeds

Figure 1 gives a broad ecological view of DS seeds and summarizes climate patterns, ecological interactions and seed features relevant for their adaptation. Seed desiccation tolerance is an important mechanism involved in the maintenance and establishment of plant communities. It is a key strategy for seed dispersal, stress avoidance and persistence in soil seed banks. Furthermore, desiccation tolerance is essential for seed persistence in environments with dry seasons. However, in tropical or subtropical evergreen rain forests, plant species with DS seeds commonly occur (Tweddle *et al.* 2003; Wyse and Dickie 2016). Such DS-seeded species span a broad range of plant life forms including shrubs (12%), palms (5%), lianas/vines (2%), herbs (9%), and epiphytes (1%), with canopy trees (45%) and small understory trees (26%) representing the majority (Farnsworth 2000). Wyse and Dickie (2016) also showed the prevalence of species with DS seeds in woody genera.

DS seeds can maintain their high water content due to their commonly large size which results in a large water reservoir, the loss of this being minimized by canopy shading (Joët *et al.* 2016). The increased seed size also results in an increased accumulation of storage reserves. This allows for a longer period for seedling establishment, as being metabolically active they are prepared for quick germination (Walters 2013).

While DT seeds normally form seed banks in the soil, DS seeds germinate fast and usually form seedling banks under shaded forest canopy and take advantage of an eventual light gap for faster establishment (Murdoch and Ellis 1992). Furthermore, DS-seeded species mainly shed seeds during the rainy season (Farnsworth 2000; Pammenter and Berjak 2000; Pritchard *et al.* 2004).

Most tropical tree pioneers, which are typical of early forest succession, have small DT and dormant seeds. Easily dispersed, these seeds may form seed banks in the forest soil and germinate when stimulated by environmental alterations or disturbances, such as natural or artificial clearings in the canopy (Bazzaz 1991). In contrast, a large number of climax species produce sizeable DS seeds (Whitmore 1990). Similarly, biennial and perennial DT species have significantly heavier seeds than annual species (Silvertown 1981). DS species that disperse during periods of high moisture availability have bigger seeds and invest proportionally less in seed coat thickness (Hill *et al.* 2012). Furthermore, seeds of climax tree species in tropical rain forest are bigger, less dormant and more persistent than pioneers (Thompson *et al.* 1993; Funes *et al.* 1999; Cerabolini *et al.* 2003; Baskin and Baskin 2005; de Souza *et al.* 2015; Jaganathan *et al.* 2015). Thus, large-seeded species appear to be more successful in seedling establishment than small-seeded species (Moles and Westoby 2004). This is likely the main driver for the phenotypical variation observed between DS and DT seeds.

Depending on the environment, DS-seeded species may show a gradient of dehydration tolerance. For example, the seeds of *Acer pseudoplatanus* display varying desiccation tolerance across their native plant distribution ranges (Daws and Pritchard 2008). These studies show the effect of provenance on shaping seed desiccation tolerance, a very important mechanism in the species life cycle. The seasonal and intraspecific variability observed for seed desiccation tolerance reflects the high plasticity of this trait and its possible role in species adaptation to different environments.


Primary forests, savannas, deserts

Low pluviosity and unpredictable; exposed to droughts and soil exposed to sun radiation with larger fluctuations of humidity

Plant species are smaller and have shorter life cycle

Seeds smaller, dormant, desiccation tolerant, longer longevity, shade intolerant.

Seed banks

Evergreen climax rain forests

High pluviosity and predictable; no or short droughts; closed canopy buffers soils humidity

Predominantly bigger and long-lived trees

Seeds bigger, non-dormant, desiccation sensitive, shorter longevity, shade tolerant

Seedling banks

Figure 1. Ecological interactions associated with DS seeds

Seed life-span and seed longevity in relation to desiccation tolerance

Seed life-span is the period between maturity and when the seed is not dormant but has lost all germination capacity in a favourable environment (Long *et al.* 2015). Seed dormancy can be released by e.g. prolonged dry storage or cold imbibition, but the level of dormancy and dormancy release depend greatly on the species and genotype (Koornneef *et al.* 2002). Seed life span is a concept that envelops multiple seed related traits such as dormancy, longevity and desiccation tolerance (Thompson *et al.* 1997;

Leck 2012). The loss of germination capacity can be associated with multiple intrinsic and extrinsic factors. Intrinsically, seeds possess a range of anatomical and physiological mechanisms that help them to remain viable until they are able to germinate and resume plant development. Extrinsically, biotic and abiotic factors will cause deterioration if the seed is not appropriately prepared for it (Long *et al.* 2008).

Seed longevity is usually defined as the storability of a seed, i.e. how long a seed remains viable during dry storage (Rajjou and Debeaujon 2008). This definition is suitable for studying seed longevity under laboratory conditions or for seed companies, where seed batches are stored before selling. However, in nature, seeds are almost never continuously exposed to dry environments. Therefore, the seed life-span needs to be considered also in wet and changing environments. This "wet longevity" is studied as seed life-span in a seed soil bank. Thus, seed life-span in a soil bank is an important concept but not much studied in what are considered 'non-model' species, this being particularly evident in forest species. However, it has been studied for weed species that affect the yield of economically important species, e.g. Striga (Ayongwa *et al.* 2011).

Seeds with long life-spans, together with timing of flowering and seed shed, propensity for desiccation tolerance and dormancy, orchestrate an optimal strategy for initiation of germination under suitable conditions for ultimate species survival. These mechanisms help the plant to survive unpredictable and unfavourable germination and growing conditions, such as pioneer species exposed to the environmental fluctuations of primary forests (Donohue *et al.* 2010).

A persistent seed bank buffers a plant population against environmental variability thereby extending the time to extinction. Thus, a long life span supports a species' population growth in unstable environments (Fenner 2000). Species with delayed germination may show a reduced population growth rate in stable environments (Rees 1994; Silvertown and Charlesworth 2009). Thus, we should expect that DS-seeded species face higher risk of extinction as their seeds have short life-span in *ex situ* storage. However, DS seeds from some species can survive for several months in moist conditions (Greggains *et al.* 2000; Le Tam *et al.* 2004; Pina-Rodrigues and Figliolia 2005; Jayasuriya *et al.* 2009) or resist water losses throughout dry seasons (Vaz *et al.* 2016). Depending on the environment into which their seeds are shed, some species, such as *Aesculus hippocastanum*, may have dense wax-covered seed coats that minimize water loss so as to extend the period a DS seed can hold onto its water (Obroucheva and Antipova 2004). Exceptionally (Pammenter and Berjak 2000), seeds of *Podocarpus angustifolius*, a DS-seeded species, do form a persistent soil seed bank (Ferrandis *et al.* 2011), indicating that desiccation tolerance is not a strict prerequisite for a relatively long seed life-span.

Interestingly, the DT seeds of *Lactuca sativa* and *Fraxinus americanus* survive storage much better when fully imbibed, as compared to the dried state (Villiers 1974). Similarly, the DT seeds of rainforest trees (e.g. *Piper hispidum*, *Piper umbelatum*, *Piper*

yzabalanum and *Urera caracasana)* display better longevity when kept dark-imbibed (Orozco-Segovia and Vazquez-Yanes 1990). Thus, some DT seeds do not need to remain in the dry state to warrant good longevity.

The mechanisms underlying the longer longevity of imbibed seeds in these species are not fully understood as imbibed versus dry seed storage studies have been underappreciated and are very scarce. The effectiveness of DNA repair has been related to this longer longevity (Villiers 1974). The repair mechanisms upon seed imbibition of DNA, RNA and proteins in model species and their importance for seed longevity have been reviewed by Sano *et al.* (2015). Similar mechanisms are likely to be involved in the longer longevity of wet-stored seeds.

We hypothesize that in environments with very predictable seasons and abundant and constant water regimes, there is lower selective pressure for seed longevity, dormancy and DT. This because they are not essential mechanisms for such environments and cost energy and, hence, there are evolutionary advantages to loose these mechanisms, resulting in DS seeds.

Physiological aspects of seed desiccation sensitivity

'Drought tolerance' can be considered as the tolerance of moderate dehydration, down to a moisture content of $\sim$0.3 g H_2O.(g dry weight)$^{-1}$, at which there is no bulk cytoplasmic water present. (Hoekstra *et al.* 2001). 'Desiccation tolerance' refers to the tolerance of further dehydration to below 0.1 g H_2O.(g dry weight)$^{-1}$, when the hydration shell of molecules is lost, without accumulation of lethal damage (Bewley 1979).

The lowest safe moisture content (LSWC), below which germination percentage begins to decline, was collected for 98 DS and 31 intermediate categorized seeded species (Figure 2 and supplementary file 1). The data was extracted from the Seed Information Database: Royal Botanic Gardens, Kew (SID) (Kew 2017). A bimodal distribution can be observed in which the first group, with higher LSWC, contains the DS- and the second the intermediate-categorized species.

The DS group is spread over a much broader range with an average of 0.71 ± 0.3 g H_2O.(g dry weight)$^{-1}$. The threshold water content for loss of DS seed viability has been shown to be similar to the permanent wilting point in vegetative tissues, about -2 MPa ($\sim$0.67 g H_2O/g dry weight) (Levitt 1980; Roberts and Ellis 1989; Andrade *et al.* 2003; Born *et al.* 2014). The lower limit for LSWC of 0.33 g H_2O.(g dry weight)$^{-1}$ includes 95% of all DS-seeded species. This lower limit for DS-seeded species is immediately above the water content at which there is no bulk cytoplasmic water present. Below 0.3 g H_2O.(g dry weight)$^{-1}$, the lack of bulk cytoplasmic water implies that the mechanism of preferential hydration by compatible solutes would fail to work

and are unable to protect proteins and membranes against further dehydration (Crowe *et al.* 1990). On further drying sugar molecules and amphiphilic compounds accumulates in the membrane spacing between phospholipid molecules, allowing DT seeds to dry below 0.1 g H_2O.(g dry weight)$^{-1}$, when cytoplasmic components are immobilized in a glassy matrix.

The intermediate group has an LSWC average of 0.13 ± 0.02 g H_2O.(g dry weight)$^{-1}$. The lower limit for LSWC of 0.9 g H_2O.(g dry weight)$^{-1}$ includes 95% of all intermediate-seeded species Dehydration below this water content causes severe loss of viability, indicating that intermediate seeds fail to desiccate to the point when the glassy state is formed.

In DS seeds, the removal of water from the hydration shell of the molecules in the absence of sugars results replace water in the hydration shell of the membranes, thereby maintaining the packing of the phospholipid molecules, which leads to a phase transition into the gel phase. This might lead to lateral phase separations and irreversible membrane damage (Hoekstra *et al.* 2001).
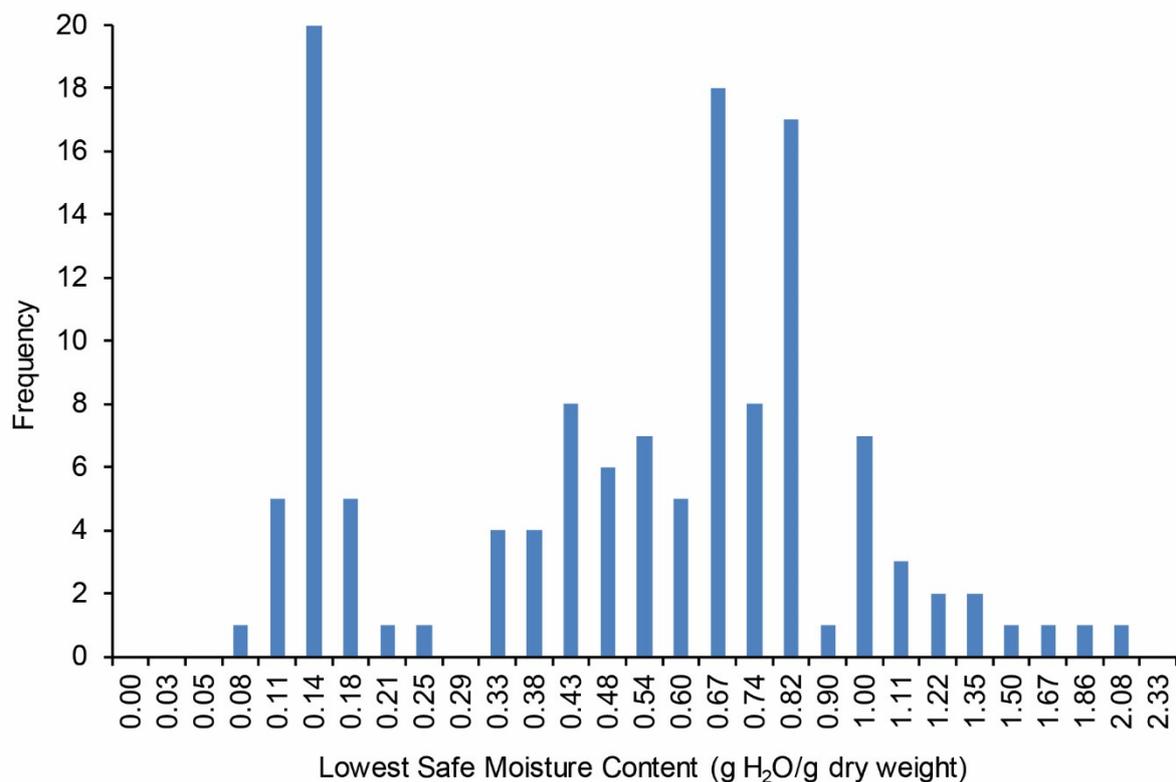

Figure 2. Distribution of seed lowest safe moisture content for seeds of species with sensitivity to dehydration (Data collected from Kew Seed Information Database).

Altogether, seed storage behaviour may be defined as DS for those species that show loss of viability above water contents of total loss of bulk cytoplasmic water (0.33

g H_2O/g dry weight) and intermediate for those that withstand further dehydration but not below ~0.9 g H_2O.(g dry weight)$^{-1}$, when the hydration shell of molecules is lost and DT those that survive further water losses without lethal damage.

The acquisition of desiccation tolerance and longevity

During the development of DT seeds, the orchestration of maturation drying by hormone- induced, as well as hormone-independent transcription factors includes the accumulation of mechanisms that enable desiccation tolerance through a series of subcellular and physiological processes such as, suppression of metabolism, vacuole filling, and cell wall relaxation (Kermode and Finch-Savage 2002; Bewley *et al.* 2012; González-Morales *et al.* 2016).

DT and longevity are acquired consecutively during seed maturation (Verdier *et al.* 2013). Seeds excised at early maturation can survive desiccation but lose viability sooner than those that undergo desiccation at later maturation stages (Chatelain *et al.* 2012; Verdier *et al.* 2013). Many studies describe the ability of DT seeds to remain viable during dry storage or artificial ageing and have shown that seed longevity is controlled by several genetic factors (Clerkx *et al.* 2004a; Clerkx *et al.* 2004b; Nguyen *et al.* 2012; Sano *et al.* 2015). However, seed longevity is a highly adaptive trait and is also influenced by both maternal and prevailing environmental conditions (He *et al.* 2014).

Among the genetic factors that contribute towards seed longevity are the genes *DOG1*, *TT*, *ATS* and *AP2* and *ABI5,* which cause reduction of longevity when knocked-out (Debeaujon *et al.* 2000; Bentsink *et al.* 2006; Zinsmeister *et al.* 2016). The testa permeability caused by the mutations *ats* and *ap2* demonstrates its importance for seed longevity (Debeaujon *et al.* 2000). Likewise, mutations in the genes encoding vitamin E synthesis enzymes (*vte1,vte2*) cause reduction of seed longevity, demonstrating the importance of ROS scavenging (Sattler *et al.* 2004). The incapacity of DS seeds to tolerate desiccation has been related to a deficient ROS scavenging system (Pukacka and Ratajczak 2006; Roach *et al.* 2010).

The acquisition of seed desiccation tolerance involves the participation of the transcription factors *LEC1*, *LEC2*, *FUS3*, *ABI3*, *ABI4*, *ABI5*, *MYB118*, *PLATZ1*, *AGL67*, *DREB2D* and protective proteins, such as *LEAs* and *HSPs*; hormones (abscisic acid (ABA) and gibberellins (GAs)) and metabolites (antioxidants and sugars) (Figure 3) (Bäumlein *et al.* 1994; Beardmore and Whittle 2005; To *et al.* 2006; Khandelwal *et al.* 2010; Delahaie *et al.* 2013; Costa *et al.* 2015; González-Morales *et al.* 2016).

Seed desiccation tolerance and longevity have been regarded and studied as separated phenotypes. However, recently, González-Morales *et al.* (2016) overexpressed transcription factors in the DS seeds of the Arabidopsis mutant *abi3-5*. These seeds

remained viable during dry storage of up to one month. It was argued that overexpression of *PLATZ1*, *AGL67* or *DREB2D* partially rescued the DS phenotype of *abi3-5*. The extended desiccated viability (longevity) of these seeds was interpreted as partial rescue of DT. This study brings the concepts of longevity and desiccation tolerance closer together.

In this way, we can regard desiccation tolerance as a complex mechanism built by multiple genetic components. We suggest that each of these genes cited above contribute additively for the establishment of desiccation tolerance and the lack of each one of them will result in reduction of longevity and ultimately in the loss of desiccation tolerance. This model where desiccation tolerance is composed by additive factors can explain the large variation of seed storage observed in nature. Hence, DT long-lived seeds contain a complete set of factors, intermediate seeds have incomplete set and further losses result in seed desiccation sensitivity.

Models for seed desiccation sensitivity

A convergent evolutionary pattern can be observed among DS-seeded species, which generally have large seeds with high water content, are insensitive to ABA, do not display developmental arrest of the embryo in shed seeds and are sometimes viviparous (Farnsworth 2000; Berjak and Pammenter 2008). Developmental arrest and acquisition of desiccation tolerance are common for DT seeds but does not occur in DS-seeded species. These DS phenotypes are similar to those observed in the Arabidopsis mutants *abi3-6, fus-3* and *lec2* (Nambara *et al.* 1992; Bäumlein *et al.* 1994; West *et al.* 1994).

The *abi3-1/aba1-1* and *abi3-5* mutants have been regarded as DS. Their seeds are nonviable when dried at harvest, but they can survive desiccation at the middle of the maturation phase, albeit with very low longevity (Ooms *et al.* 1994; González-Morales *et al.* 2016). This DT window offers the possibility to store these seeds (for a limited time) and paves the way for a better understanding of the dynamics of acquisition and loss of DT. It is possible that a similar pattern occurs in natural DS seeds.

An important model for studies of desiccation tolerance in plants is the re-induction of desiccation tolerance of germinated seeds. Germination leads seeds towards a developmental seedling program during which desiccation tolerance is gradually lost but can be rescued by ABA and/or osmotic treatments (Buitink *et al.* 2003; Vu *et al.* 2003; Faria *et al.* 2005b; Maia *et al.* 2011; Maia *et al.* 2014; Costa *et al.* 2015). The gradual loss of desiccation tolerance by germinated DT seeds might be associated with the loss of desiccation tolerance in DS species, since the developmental program of DS seeds is also characterized by a constant high water content and a growing embryo, reminiscent of germination.
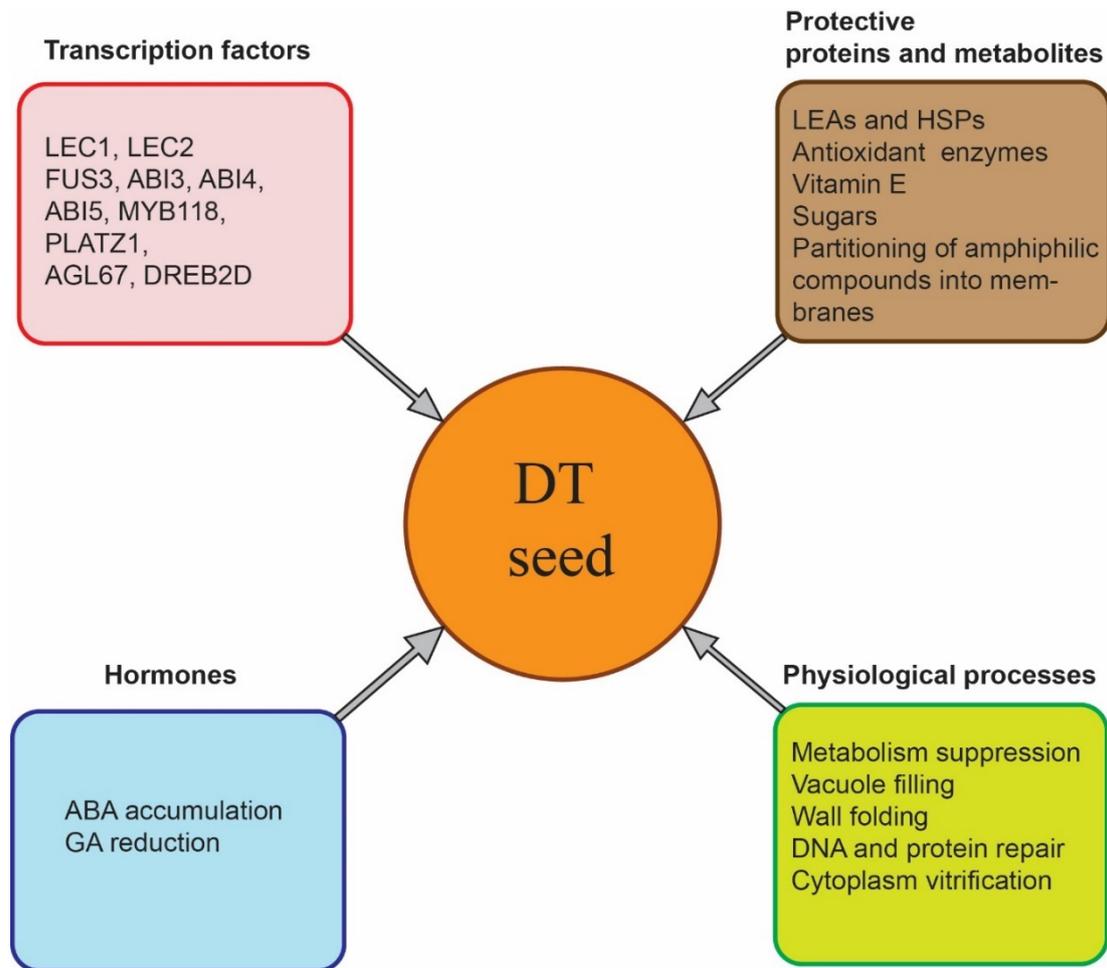
Figure 3. Factors associated with the acquisition of desiccation tolerance in seeds.

Seed size and its role in seed storage

The ratio of seed coat mass to total seed mass was evaluated across 104 species from 37 families and showed a significant correlation with seed desiccation sensitivity. There is a strong evolutionary association of seed desiccation sensitivity with both large seed mass and "thin" seed coats (Daws *et al.* 2006).

The hypothetical evolutionary model (Figure 4) suggests a shift between the seed life-span strategy from dry longevity towards wet longevity in the evolutionary path of DS-seeded species. The wet longevity strategy is supported by a larger seed size that buffers the water loss caused by fluctuations in soil water potential. Furthermore, the larger seed mass also contains more reserves to utilize in protracted ongoing metabolism associated with being hydrated and progressing towards germination.

An important tissue involved in the longevity of seeds is the seed coat. The seed coat performs important functions to protect the embryo and seed reserves from biotic and abiotic stresses during storage (pathogen and predator attacks, UV radiations, moisture, elevated temperature and oxygen) (Rajjou and Debeaujon 2008).

Crop seeds exhibiting testa defects such as yellow-seeded rapeseed (*Brassica napus*) and flax (*Linum usitatissimum*) also exhibit lower longevity (Nesi *et al.* 2008). These data support the hypothesis of the role of the seed coat (thick secondary cell walls and mucilage) in restricting the exchange of gases and fluids between the environment and the seed, prolonging the longevity of the seed, keeping it dormant (Debeaujon *et al.* 2000) and protecting it against predation until it germinates (Haughn and Chaudhury 2005).

DS seeds are generally shed at high water content and metabolic activity (Berjak and Pammenter 2008). Therefore, a gas impermeable seed coat is not suitable for a metabolically active tissue that needs constant gas exchange for respiration. Furthermore, the water impermeability caused by thick seed coats is not desirable when the evolutionary pressure is on seeds that imbibe and germinate faster. The reduction of the thickness of the seed coat as observed in DS seeds (Daws *et al.* 2006) is most likely the result of an evolutionary process caused by selective pressure towards faster germination. This evolutionary process could be associated with the repression of genes controlling seed coat development.
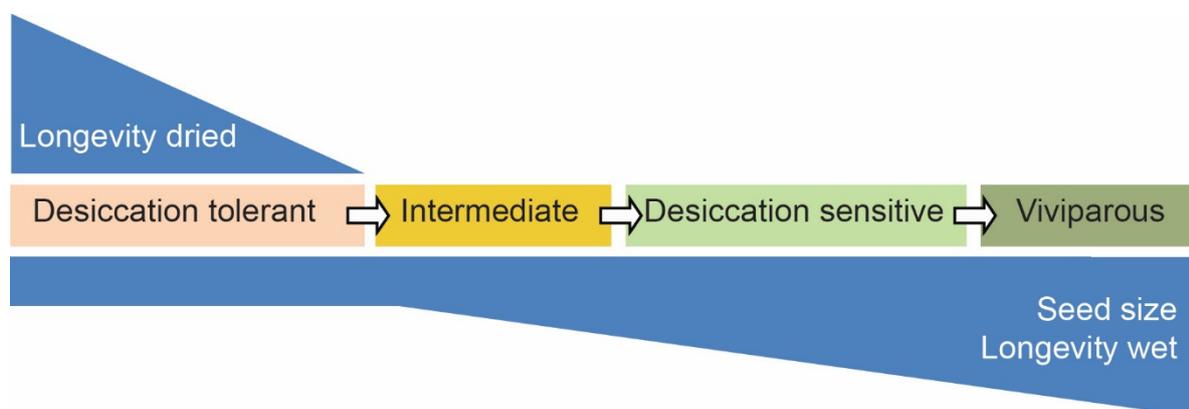

Figure. 4. Hypothetical model of evolutionary events resulting in seed desiccation sensitivity

Desiccation sensitivity and hormones

ABA and GAs are the main phytohormones that have been associated with the acquisition of DT, or lack thereof, in seeds. ABA has been associated mostly with the plant's defence against abiotic stresses. Drought and high salinity increase plant ABA

levels, resulting in changes in gene expression and adaptive physiological responses (Rabbani *et al.* 2003; Christmann *et al.* 2007). During the development of DT seeds, ABA peaks during the onset of maturation in preparation for developmental arrest and desiccation (Kanno *et al.* 2010). GAs regulate growth and various developmental processes, including stem elongation, breaking of dormancy and induction of germination (Silverstone *et al.* 1998; Finch-Savage and Leubner-Metzger 2006). GA is involved in cell division and elongation and its concentration peaks during embryogenesis but drops to very low levels at the start of maturation in DT seeds (Kanno *et al.* 2010).

Incubation in an ABA solution with tetcyclacis, an inhibitor of GA-biosynthesis and ABA catabolism, induced desiccation tolerance in the DS seeds of *Acer sacharinum* (Beardmore and Whittle 2005). It was argued that prolonged incubation in this solution triggered an aspect of the seed maturation phase that was naturally skipped. This changed the long-held concept that seed DS is a static phenotype of the species. The induction of desiccation tolerance in *A. sacharinum* seeds by tetcyclacis indicates that not only the lack of ABA is responsible for DS but also relative high GA levels play an important role. This is in agreement with the hypothesis of the balance of ABA/GA as the mode of action for regulation of physiological process (Shu *et al.* 2015).

The balance ABA/GA has also been shown to regulate the germination versus maturation programs in maize. The seeds of the maize mutant *vp5* have very low ABA levels, are DS and viviparous, but with the suppression of GA synthesis, viviparity was avoided and desiccation tolerance rescued. (White *et al.* 2000).

The importance of ABA accumulation for the establishment of desiccation tolerance in seeds has been observed in several studies (Faria *et al.* 2005a; Maia *et al.* 2011; Dekkers *et al.* 2015). However, DS may also, in certain cases, be associated with a lack of sensitivity to ABA. Germination leads seeds towards a seedling developmental program during which desiccation tolerance is gradually lost and coinciding with progressive loss of ABA sensitivity (Maia *et al.* 2014). Furthermore, embryos of some DS species, display pre-harvest sprouting and viviparous species exhibit reduced sensitivity to ABA (Sussex 1975; Garello and Page‑Degivry 1995; Steinbach *et al.* 1995; Gubler *et al.* 2005).

Viviparous mangrove species tend to have lower ABA contents as compared to DT seeds in where germination is initiated on the parent, or shortly after shedding (Farrant *et al.* 1993; Farnsworth and Farrant 1998). Furthermore, in these species, ABA levels in embryos are significantly lower than that in the seed coat/ maternal tissues, which could be interpreted as a mechanism to slow germination changes (maintain a type of dormancy) until seed shed and coat shedding. However, this might not be universally true of DS seeds and, hence, it might be that sensitivity to ABA or GA content are decisive for the appearance of desiccation sensitivity.

ABA and GA are interconnected by a network of other regulators, such as other hormones, transcription factors and receptors (Kucera *et al.* 2005). Thus, although they are clearly key factors, these two hormones alone cannot explain the complexity of desiccation tolerance acquisition. Taken together, these findings suggest that ABA and GA are important in seed development, controlling embryogenesis and preparing the mature embryo for desiccation.

Evolutionary aspects of seed desiccation sensitivity

Evolution in homogeneous environments will favour specialist species while heterogeneous environments will favour generalists. The more variable the environment, the more plastic a species needs to be (Levins 1968; Lynch and Gabriel 1987; Van Tienderen 1991). There is a cost of maintaining physiological machineries, resulting in a trade-off between tolerance breadth and maximal performance at optimal conditions. Thus, stable environments will favour locally specialized genotypes (Huey and Kingsolver 1989; Richards *et al.* 2006).

Seed dormancy, desiccation tolerance and longevity are important mechanisms for generalists, but not for specialists like DS-seeded species. These species evolved and eventually became specialists in stable ecosystems with predictable rainy seasons. These ecosystems favour faster germination, seedling establishment and disfavour dormancy, desiccation tolerance and longevity (Garwood 1989; Vázquez-Yanes and Orozco-Segovia 1993; de Souza *et al.* 2015). These traits are accompanied by larger seed sizes that can also function as a high-volume water reservoir and aid in emergence of bigger seedlings.

There is no clear phylogenetic pattern among DS-seeded species (Farnsworth 2000; Pammenter and Berjak 2000; Dickie and Pritchard 2002; Berjak and Pammenter 2008). Desiccation sensitivity in seeds is currently considered a secondary evolutionary characteristic and desiccation tolerance ancestral, as the most parsimonious explanation for evolution of DS seeds points to independent losses of desiccation tolerance (Farnsworth 2000). This implies that every DS-seeded species represents an independent event of loss of DT. Thus, we may expect independent genetic and associated phenotypic alterations associated with each DS species. Nevertheless, it is also possible that the evolution of DS seeds needed to follow a convergent path and we might find common genetic/phenotypic alterations, as observed in the Arabidopsis mutants discussed above.

Desiccation tolerance is a complex trait influenced and established by a plethora of genes (Figure 3) (Terrasson *et al.* 2013; Maia *et al.* 2014; Dekkers *et al.* 2015), but the loss of only a single regulatory gene may completely dismantle this mechanism as in the *abi3-6*, *fus3*, *lec1* and *lec2* mutants (Nambara *et al.* 1992; Bäumlein *et al.* 1994;

West *et al.* 1994). These four transcription factors are highly pleiotropic. Mutations in the transcription factor *ABI3,* e.g. the full knock-out *abi3-6*, result in DS seeds, but which also lack dormancy and display absence of chlorophyll degradation, insensitivity to ABA and changes in reserve composition (To *et al.* 2006; Santos-Mendoza *et al.* 2008; Delmas *et al.* 2013). Desiccation sensitivity in seeds of wild species is accompanied by similar phenotypical alterations such as chlorophyll retention and lack of dormancy in mature seeds (Farnsworth 2000; Hamilton *et al.* 2013).

The transcription factor *ABI3/VP1* is strongly conserved across plants and also found in the moss *Physcomitrella patens* which exhibits vegetative DT. The *ABI3* homolog from *P. patens* partially complemented the Arabidopsis *abi3-6* mutant whereas the *Ppabi3* mutant also becomes DS (Zeng and Kermode 2004; Marella *et al.* 2006; Khandelwal *et al.* 2010). This confirms that desiccation tolerance is a mechanism developed in early land plants and its regulation is probably conserved throughout the plant kingdom (Gaff and Oliver 2013).

It is tempting to speculate that wild DS-seeded species have similar mutation(s) in single key regulatory genes instead of extensive small alterations accumulated in multiple structural proteins. This would support a hypothesis of DS evolution by saltatory events instead of Darwinian gradualism. The gradualism hypothesis defends that beneficial changes, fixed by natural selection, are 'infinitesimally' small, so that evolution develops via the gradual accumulation of these tiny modifications (Darwin 1859). Moreover, these gradual changes can accumulate not only by adaptive mutations but through fixation via random genetic drift (Wright 1986). On the other hand, the saltatory theory advocates that evolution can work by large steps and recent evidence suggests that whole genome duplication events represent a dramatic example of saltatory evolution (Freeling 2008).

The suppression of certain key regulator genes, rather than addition of new genes, can be responsible for major and sudden evolutionary steps. The loss of such genes results in a loss of a mechanism or process that was imposing a developmental constraint. With the loss of this constraint, evolution suddenly acquires new possibilities for further adaptation (Bateman and DiMichele 1994; Tucker 2001). We can hypothesize, for example, that it would be impossible for the coconut seed to achieve such sizeable dimensions and be able to adapt to its niches without losing desiccation tolerance considering the association of seed size with desiccation sensitivity.

Genomic alterations associated with seed desiccation sensitivity have not been studied exhaustively in DS-seeded species so far. The only study so far points to the underrepresentation of LEAs and dehydrins in the genome of the seagrass *Zostera marina* (Olsen *et al.* 2016), which produces DS seeds (Pan *et al.* 2012). The genomes of DS-seeded species such as *Theobroma cacao, Hevea brasiliensis, Castanea mollissima* and *Elaeis guineensis* have been sequenced and future studies on them and possibly others will shed light on the genomic alterations associated with desiccation sensitivity.

Only by comparing complete genomes, it will be possible to demonstrate the absence of a particular gene in a genome, to pinpoint gene loss events or other events responsible for the evolution of DS seed (Koonin 2009).

The common definition of seed storability was originally based on a dual concept, namely 'orthodox' vs. 'recalcitrant'. This definition does not appreciate the highly diverse seed storage phenotypes observed in nature. This outdated black and white definition has hampered the generation of hypotheses related to seed storage. In fact, this appears to be a continuum and, thus, the 'intermediate' seed storage category was created (Ellis *et al.* 1991; Berjak and Pammenter 2013). Intermediate seeds can withstand further dehydration as compared to DS seeds but cannot undergo desiccation as in DT seeds (Ellis *et al.* 1990). We hypothesize that desiccation sensitivity in seeds is a consequence of the loss of desiccation tolerance of DT seeds, caused by elimination of the dehydration program driven by maturation. The loss of desiccation tolerance is probably caused by selective pressure towards faster germination, and intermediate seeds are an evolutionary step in between. The continuum of the seed storage phenotype might reflect the various evolutionary histories of each species when adapting to different environments. To support the hypothesis of gradual loss of DT, we assume that DT seeds lost longevity gradually when exposed to ecosystems with predictable and abundant water regimes, culminating in DS seeds.

A concept of the evolutionary history that led to this extent of physiological variation observed in seed storage, proposes an evolutionary path that describes the emergence of these phenotypes. It is tempting to speculate on a single and straightforward sequence of events that can result in these species (Figure 3). However, also completely independent changes converging to it may be expected. Some of the most striking examples of such loss include the absence of eyes in animals living in dark places, the loss of photosynthetic capacity in parasitic plants (Depamphilis and Palmer 1990; McNeal *et al.* 2007) and a highly reduced perianth, a common adaptation to wind pollination in several families of angiosperms (e.g. grasses, oaks) (Culley *et al.* 2002; Jaramillo and Kramer 2007).

Based on these considerations we propose two possible hypotheses to explain the loss of desiccation tolerance in DS seeds.

- The neutral mutation hypothesis (Kimura and Ohta 1971) implies that desiccation tolerance regression is caused by fixation of random deleterious mutations in DT-associated genes, which accumulate in DS-seeded plants under relaxed selective pressure by random drift.
- The adaptation hypothesis suggests that the loss of desiccation tolerance is adaptive and has a selective advantage in an environment with stable water regimes. Regression of desiccation tolerance may be attributed to the high energy cost of these mechanisms, to pleiotropic effects, in which germination traits beneficial to survival in a tropical forest environment are enhanced by desiccation

sensitivity, or even to active elimination of desiccation tolerance because of its potential liability in constant rainy season forests.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest

References

Achard, F, Eva, HD, Stibig, H-J, Mayaux, P, Gallego, J, Richards, T, Malingreau, J-P (2002) Determination of deforestation rates of the world's humid tropical forests. *Science* **297**, 999-1002.

Alpert, P, Oliver, MJ (2002) Drying without dying. *Desiccation and survival in plants: Drying without dying* 3-43.

Andrade, A, Cunha, R, Souza, A, Reis, R, Almeida, K (2003) Physiological and morphological aspects of seed viability of a neotropical savannah tree, Eugenia dysenterica DC. *Seed Science and Technology* **31**, 125-137.

Ayongwa, G, Stomph, T, Belder, P, Leffelaar, P, Kuyper, T (2011) Organic matter and seed survival of Striga hermonthica–mechanisms for seed depletion in the soil. *Crop Protection* **30**, 1594-1600.

Baskin, CC, Baskin, JM (2005) Seed dormancy in trees of climax tropical vegetation types. *Tropical Ecology* **46**, 17-28.

Bateman, R, DiMichele, W (1994) Saltational evolution of form in vascular plants: a neoGoldschmidtian synthesis. *Shape and form in plants and fungi* 63-102.

Bäumlein, H, Miséra, S, Luerßen, H, Kölle, K, Horstmann, C, Wobus, U, Müller, AJ (1994) The FUS3 gene of Arabidopsis thaliana is a regulator of gene expression during late embryogenesis. *The Plant Journal* **6**, 379-387.

Bazzaz, F (1991) Habitat selection in plants. *American Naturalist* S116-S130.

Beardmore, T, Whittle, C-A (2005) Induction of tolerance to desiccation and cryopreservation in silver maple (Acer saccharinum) embryonic axes. *Tree physiology* **25**, 965-972.

Bentsink, L, Jowett, J, Hanhart, CJ, Koornneef, M (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc Natl Acad Sci U S A* **103**, 17042-7.

Berjak, P, Pammenter, NW (2008) From Avicennia to Zizania: seed recalcitrance in perspective. *Ann Bot* **101**, 213-28.

Berjak, P, Pammenter, NW (2013) Implications of the lack of desiccation tolerance in recalcitrant seeds. *Front Plant Sci* **4**, 478.

Berjak, P, Vertucci, CW, Pammenter, NW (2008) Effects of developmental status and dehydration rate on characteristics of water and desiccation-sensitivity in recalcitrant seeds of Camellia sinensis. *Seed Science Research* **3**,

Bewley, JD (1979) Physiological aspects of desiccation tolerance. *Annual Review of Plant Physiology* **30**, 195-238.

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

Bewley, JD, Bradford, KJ, Hilhorst, HW, Nonogaki, H (2013) Longevity, storage, and deterioration. In 'Seeds.' pp. 341-376. (Springer:

Born, N, Behringer, D, Liepelt, S, Beyer, S, Schwerdtfeger, M, Ziegenhagen, B, Koch, M (2014) Monitoring plant drought stress response using terahertz time-domain spectroscopy. *Plant physiology* **164**, 1571-1577.

Buitink, J, Vu, BL, Satour, P, Leprince, O, 2003. The re-establishment of desiccation tolerance in germinated radicles of Medigago truncatula Gaertn. Seed Science Research. 13: 273-286.

Cerabolini, B, Ceriani, RM, Caccianiga, M, De Andreis, R, Raimondi, B (2003) Seed size, shape and persistence in soil: a test on Italian flora from Alps to Mediterranean coasts. *Seed Science Research* **13**, 75-85.

Chatelain, E, Hundertmark, M, Leprince, O, Le Gall, S, Satour, P, Deligny-Penninck, S, Rogniaux, H, Buitink, J (2012) Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ* **35**, 1440-55.

Christina Walters, PB, Norman Pammenter, Kathryn Kennedy, Peter Raven (2013) Preservation of Recalcitrant Seeds. *Science* **339**,

Christmann, A, Weiler, EW, Steudle, E, Grill, E (2007) A hydraulic signal in root to shoot signalling of water shortage. *The Plant Journal* **52**, 167-174.

Clerkx, EJ, El-Lithy, ME, Vierling, E, Ruys, GJ, Blankestijn-De Vries, H, Groot, SP, Vreugdenhil, D, Koornneef, M (2004a) Analysis of natural allelic variation of Arabidopsis seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. *Plant physiology* **135**, 432-443.

Clerkx, EJ, Vries, BD, Ruys, GJ, Groot, SP, Koornneef, M (2004b) Genetic differences in seed longevity of various Arabidopsis mutants. *Physiologia Plantarum* **121**, 448-461.

Costa, MC, Righetti, K, Nijveen, H, Yazdanpanah, F, Ligterink, W, Buitink, J, Hilhorst, HW (2015) A gene co-expression network predicts functional genes controlling the re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds. *Planta* **242**, 435-49.

Crowe, JH, Carpenter, JF, Crowe, LM, Anchordoguy, TJ (1990) Are freezing and dehydration similar stress vectors? A comparison of modes of interaction of stabilizing solutes with biomolecules. *Cryobiology* **27**, 219-231.

Culley, TM, Weller, SG, Sakai, AK (2002) The evolution of wind pollination in angiosperms. *Trends in ecology & evolution* **17**, 361-369.

Darwin, C (1859) On the origin ofspecies by means of natural selection, or the preservation of favoured races in the struggle for life. *New York, NY: D. Appleton and Company*

Daws, MI, Garwood, NC, Pritchard, HW (2006) Prediction of desiccation sensitivity in seeds of woody species: a probabilistic model based on two seed traits and 104 species. *Ann Bot* **97**, 667-74.

Daws, MI, Pritchard, HW (2008) The development and limits of freezing tolerance in Acer pseudoplatanus fruits across Europe is dependent on provenance. *CryoLetters* **29**, 189-198.

de Souza, TV, Torres, IC, Steiner, N, Paulilo, MTS (2015) Seed dormancy in tree species of the Tropical Brazilian Atlantic Forest and its relationships with seed traits and environmental conditions. *Brazilian Journal of Botany* **38**, 243-264.

Debeaujon, I, Léon-Kloosterziel, KM, Koornneef, M (2000) Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant physiology* **122**, 403-414.

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

Delahaie, J, Hundertmark, M, Bove, J, Leprince, O, Rogniaux, H, Buitink, J (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-

regulated LEA protein abundance linked to desiccation tolerance. *J Exp Bot* **64**, 4559-73.

Delmas, F, Sankaranarayanan, S, Deb, S, Widdup, E, Bournonville, C, Bollier, N, Northey, JG, McCourt, P, Samuel, MA (2013) ABI3 controls embryo degreening through Mendel's I locus. *Proceedings of the National Academy of Sciences* **110**, E3888-E3894.

Depamphilis, CW, Palmer, JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **348**, 337-339.

Dickie, JB, Pritchard, HW (2002) Systematic and evolutionary aspects of desiccation tolerance in seeds. *Desiccation and survival in plants: drying without dying. Wallingford, UK: CAB International* 239-259.

Donohue, K, Rubio de Casas, R, Burghardt, L, Kovach, K, Willis, CG (2010) Germination, postgermination adaptation, and species ecological ranges. *Annual Review of Ecology, Evolution, and Systematics* **41**, 293-319.

Ellis, R, Hong, T, Roberts, E (1990) An intermediate category of seed storage behaviour? I. Coffee. *Journal of experimental botany* **41**, 1167-1174.

Ellis, R, Hong, T, Roberts, E (1991) An intermediate category of seed storage behaviour? II. Effects of provenance, immaturity, and imbibition on desiccation-tolerance in coffee. *Journal of experimental botany* **42**, 653-657.

Faria, JM, Buitink, J, van Lammeren, AA, Hilhorst, HW (2005a) Changes in DNA and microtubules during loss and re-establishment of desiccation tolerance in germinating Medicago truncatula seeds. *J Exp Bot* **56**, 2119-30.

Faria, JM, Buitink, J, van Lammeren, AA, Hilhorst, HWM (2005b) Changes in DNA and microtubules during loss and re-establishment of desiccation tolerance in germinating Medicago truncatula seeds. *J Exp Bot* **56**, 2119-30.

Farnsworth, E (2000) The ecology and physiology of viviparous and recalcitrant seeds. *Annual Review of Ecology and Systematics* 107-138.

Farnsworth, E, Farrant, J (1998) Reductions in abscisic acid are linked with viviparous reproduction in mangroves. *American Journal of Botany* **85**, 760-760.

Farrant, JM, Berjak, P, Cutting, J, Pammenter, N (1993) The role of plant growth regulators in the development and germination of the desiccation-sensitive (recalcitrant) seeds of Avicennia marina. *Seed Science Research* **3**, 55-63.

Fenner, M (2000) 'Seeds: the ecology of regeneration in plant communities.' (Cabi:

Ferrandis, P, Bonilla, M, Osorio, LdC (2011) Germination and soil seed bank traits of Podocarpus angustifolius (Podocarpaceae): an endemic tree species from Cuban rain forests. *Revista de Biología Tropical* **59**, 1061-1069.

Finch-Savage, WE, Leubner-Metzger, G (2006) Seed dormancy and the control of germination. *New Phytol* **171**, 501-23.

Freeling, M (2008) The evolutionary position of subfunctionalization, downgraded. In 'Plant genomes.' Vol. 4 pp. 25-40. (Karger Publishers:

Funes, G, Basconcelo, S, Díaz, S, Cabido, M (1999) Seed size and shape are good predictors of seed persistence in soil in temperate mountain grasslands of Argentina. *Seed Science Research* **9**, 341-345.

Gaff, DF, Oliver, M (2013) The evolution of desiccation tolerance in angiosperm plants: a rare yet common phenomenon. *Functional Plant Biology* **40**, 315.

Garello, G, Page Degivry, MT (1995) Desiccation sensitive Hopea odorata seeds: Sensitivity to abscisic acid, water potential and inhibitors of gibberellin biosynthesis. *Physiologia Plantarum* **95**, 45-50.

Garwood, NC (1989) Tropical soil seed banks: a review. *Ecology of soil seed banks* **149**, 210.

González-Morales, SI, Chávez-Montes, RA, Hayano-Kanashiro, C, Alejo-Jacuinde, G, Rico-Cambron, TY, de Folter, S, Herrera-Estrella, L (2016) Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* 201610985.

Greggains, V, Finch-Savage, WE, Quick, WP, Atherton, NM (2000) Putative desiccation tolerance mechanisms in orthodox and recalcitrant seeds of the genus Acer. *Seed Science Research* **10**, 317-327.

Gubler, F, Millar, AA, Jacobsen, JV (2005) Dormancy release, ABA and pre-harvest sprouting. *Current opinion in plant biology* **8**, 183-187.

Hamilton, KN, Offord, CA, Cuneo, P, Deseo, MA (2013) A comparative study of seed morphology in relation to desiccation tolerance and other physiological responses in 71 Eastern Australian rainforest species. *Plant Species Biology* **28**, 51-62.

Haughn, G, Chaudhury, A (2005) Genetic analysis of seed coat development in Arabidopsis. *Trends in plant science* **10**, 472-477.

He, H, de Souza Vidigal, D, Snoek, LB, Schnabel, S, Nijveen, H, Hilhorst, H, Bentsink, L (2014) Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *Journal of experimental botany* **65**, 6603-6615.

Hill, JP, Edwards, W, Franks, PJ, Leishman, M (2012) Size is not everything for desiccation-sensitive seeds. *Journal of Ecology* **100**, 1131-1140.

Hoekstra, FA, Golovina, EA, Buitink, J (2001) Mechanisms of plant desiccation tolerance. *Trends in plant science* **6**, 431-438.

Huey, RB, Kingsolver, JG (1989) Evolution of thermal sensitivity of ectotherm performance. *Trends in ecology & evolution* **4**, 131-135.

Jaganathan, GK, Dalrymple, SE, Liu, B (2015) Towards an understanding of factors controlling seed bank composition and longevity in the alpine environment. *The Botanical Review* **81**, 70-103.

Jaramillo, MA, Kramer, EM (2007) Molecular evolution of the petal and stamen identity genes, APETALA3 and PISTILLATA, after petal loss in the Piperales. *Molecular phylogenetics and evolution* **44**, 598-609.

Jayasuriya, KG, Baskin, JM, Baskin, CC (2009) Sensitivity cycling and its ecological role in seeds with physical dormancy. *Seed Science Research* **19**, 3-13.

Joët, T, Ourcival, J-M, Capelli, M, Dussert, S, Morin, X (2016) Explanatory ecological factors for the persistence of desiccation-sensitive seeds in transient soil seed banks: Quercus ilex as a case study. *Annals of Botany* **117**, 165-176.

Kanno, Y, Jikumaru, Y, Hanada, A, Nambara, E, Abrams, SR, Kamiya, Y, Seo, M (2010) Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. *Plant Cell Physiol* **51**, 1988-2001.

Kermode, AR, Finch-Savage, BE (2002) Desiccation sensitivity in orthodox and recalcitrant seeds in relation to development. *Desiccation and survival in plants: Drying without dying* 149-184.

Kew, RBG, 2017. Seed information database (SID). Version 7.1.

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

Kimura, M, Ohta, T (1971) 'Theoretical aspects of population genetics.' (Princeton University Press:

Koonin, EV (2009) Darwinian evolution in the light of genomics. *Nucleic acids research* **37**, 1011-1034.

Koornneef, M, Bentsink, L, Hilhorst, H (2002) Seed dormancy and germination. *Current opinion in plant biology* **5**, 33-36.

Kucera, B, Cohn, MA, Leubner-Metzger, G (2005) Plant hormone interactions during seed dormancy release and germination. *Seed Science Research* **15**, 281-307.

Le Tam, V, Hong, T, Ellis, R, Ngoc-Tam, B (2004) Seed storage of Avicennia alba Bl. *Seed Science and Technology* **32**, 531-536.

Leck, MA (2012) 'Ecology of soil seed banks.' (Elsevier:

Levins, R (1968) 'Evolution in changing environments: some theoretical explorations.' (Princeton University Press:

Levitt, J (1980) 'Responses of Plants to Environmental Stress, Volume 1: Chilling, Freezing, and High Temperature Stresses.' (Academic Press.:

Li, DZ, Pritchard, HW (2009) The science and economics of ex situ plant conservation. *Trends Plant Sci* **14**, 614-21.

Long, RL, Gorecki, MJ, Renton, M, Scott, JK, Colville, L, Goggin, DE, Commander, LE, Westcott, DA, Cherry, H, Finch Savage, WE (2015) The ecophysiology of seed persistence: a mechanistic view of the journey to germination or demise. *Biological Reviews* **90**, 31-59.

Long, RL, Panetta, FD, Steadman, KJ, Probert, R, Bekker, RM, Brooks, S, Adkins, SW (2008) Seed persistence in the field may be predicted by laboratory-controlled aging. *Weed Science* **56**, 523-528.

Lynch, M, Gabriel, W (1987) Environmental tolerance. *American Naturalist* 283-303.

Maia, J, Dekkers, BJ, Dolle, MJ, Ligterink, W, Hilhorst, HW (2014) Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated Arabidopsis seeds. *New Phytol* **203**, 81-93.

Maia, J, Dekkers, BJ, Provart, NJ, Ligterink, W, Hilhorst, HW (2011) The re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds and its associated transcriptome. *PloS one* **6**, e29123.

Marella, HH, Sakata, Y, Quatrano, RS (2006) Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3 like genes from Physcomitrella patens. *The Plant Journal* **46**, 1032-1044.

McNeal, JR, Kuehl, JV, Boore, JL, de Pamphilis, CW (2007) Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus Cuscuta. *BMC Plant Biol* **7**, 57.

Moles, AT, Westoby, M (2004) Seedling survival and seed size: a synthesis of the literature. *Journal of Ecology* **92**, 372-383.

Murdoch, AJ, Ellis, RH (1992) Longevity, viability and dormancy. *Seeds: the ecology of regeneration in plant communities* 193-229.

Nambara, E, Naito, S, McCourt, P (1992) A mutant of Arabidopsis which is defective in seed development and storage protein accumulation is a new abi3 allele. *The Plant Journal* **2**, 435-441.

Nesi, N, Delourme, R, Brégeon, M, Falentin, C, Renard, M (2008) Genetic and molecular approaches to improve nutritional value of Brassica napus L. seed. *Comptes rendus biologies* **331**, 763-771.

Nguyen, TP, Keizer, P, van Eeuwijk, F, Smeekens, S, Bentsink, L (2012) Natural variation for seed longevity and seed dormancy are negatively correlated in Arabidopsis. *Plant Physiol* **160**, 2083-92.

Obroucheva, N, Antipova, O (2004) The role of water uptake in the transition of recalcitrant seeds from dormancy to germination. *Russian Journal of Plant Physiology* **51**, 848-856.

Olsen, JL, Rouzé, P, Verhelst, B, Lin, Y-C, Bayer, T, Collen, J, Dattolo, E, De Paoli, E, Dittami, S, Maumus, F (2016) The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature* **530**, 331-335.

Ooms, JJ, Veen, R, Karssen, CM (1994) Abscisic acid and osmotic stress or slow drying independently induce desiccation tolerance in mutant seeds of Arabidopsis thaliana. *Physiologia Plantarum* **92**, 506-510.

Orozco-Segovia, A, Vazquez-Yanes, C (1990) Effect of moisture on longevity in seeds of some rain forest species. *Biotropica* **22**, 215-216.

Pammenter, N, Berjak, P (2000) Evolutionary and ecological aspects of recalcitrant seed biology. *Seed Science Research* **10**, 301-306.

Pan, J, Han, H, Jiang, X, Zhang, W, Zhao, N, Song, S, Li, X, Li, X (2012) Desiccation, moisture content and germination of Zostera marina L. seed. *Restoration Ecology* **20**, 311-314.

Pina-Rodrigues, F, Figliolia, M (2005) Embryo immaturity associated with delayed germination in recalcitrant seeds of Virola surinamensis (Rol.) Warb.(Myristicaceae). *Seed Science and Technology* **33**, 375-386.

Pritchard, HW, Daws, MI, Fletcher, BJ, Gaméné, CS, Msanga, HP, Omondi, W (2004) Ecological correlates of seed desiccation tolerance in tropical African dryland trees. *American Journal of Botany* **91**, 863-870.

Pukacka, S, Ratajczak, E (2006) Antioxidative response of ascorbate–glutathione pathway enzymes and metabolites to desiccation of recalcitrant Acer saccharinum seeds. *Journal of plant physiology* **163**, 1259-1266.

Rabbani, MA, Maruyama, K, Abe, H, Khan, MA, Katsura, K, Ito, Y, Yoshiwara, K, Seki, M, Shinozaki, K, Yamaguchi-Shinozaki, K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant physiology* **133**, 1755-1767.

Rajjou, L, Debeaujon, I (2008) Seed longevity: survival and maintenance of high germination ability of dry seeds. *Comptes rendus biologies* **331**, 796-805.

Rees, M (1994) Delayed germination of seeds: a look at the effects of adult longevity, the timing of reproduction, and population age/stage structure. *American Naturalist* 43-64.

Richards, CL, Bossdorf, O, Muth, NZ, Gurevitch, J, Pigliucci, M (2006) Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology letters* **9**, 981-993.

Roach, T, Beckett, RP, Minibayeva, FV, Colville, L, Whitaker, C, Chen, H, Bailly, C, Kranner, I (2010) Extracellular superoxide production, viability and redox poise

in response to desiccation in recalcitrant Castanea sativa seeds. *Plant, Cell & Environment* **33**, 59-75.

Roberts, E, Ellis, R (1989) Water and seed survival. *Annals of Botany* **63**, 39-39.

Sano, N, Rajjou, L, North, HM, Debeaujon, I, Marion-Poll, A, Seo, M (2015) Staying alive: molecular aspects of seed longevity. *Plant and Cell Physiology* pcv186.

Santos-Mendoza, M, Dubreucq, B, Baud, S, Parcy, F, Caboche, M, Lepiniec, L (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J* **54**, 608-20.

Sattler, SE, Gilliland, LU, Magallanes-Lundback, M, Pollard, M, DellaPenna, D (2004) Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *Plant Cell* **16**, 1419-32.

Shu, K, Chen, Q, Wu, Y, Liu, R, Zhang, H, Wang, P, Li, Y, Wang, S, Tang, S, Liu, C (2015) ABI4 mediates antagonistic effects of abscisic acid and gibberellins at transcript and protein levels. *The Plant Journal*

Silverstone, AL, Ciampaglio, CN, Sun, T-p (1998) The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *The Plant Cell* **10**, 155-169.

Silvertown, J, Charlesworth, D (2009) 'Introduction to plant population biology.' (John Wiley & Sons:

Silvertown, JW (1981) Seed size, life span, and germination date as coadapted features of plant life history. *The American Naturalist* **118**, 860-864.

Steinbach, HS, Benech–Arnold, R, Kristof, G, Sánchez, R, Marcucci-Poltri, S (1995) Physiological basis of pre-harvest sprouting resistance in Sorghum bicolor (L.) Moench. ABA levels and sensitivity in developing embryos of sprouting-resistant and-susceptible varieties. *Journal of experimental botany* **46**, 701-709.

Sussex, I (1975) Growth and metabolism of the embryo and attached seedling of the viviparous mangrove, Rhizophora mangle. *American Journal of Botany* 948-953.

Terrasson, E, Buitink, J, Righetti, K, Ly Vu, B, Pelletier, S, Zinsmeister, J, Lalanne, D, Leprince, O (2013) An emerging picture of the seed desiccome: confirmed regulators and newcomers identified using transcriptome comparison. *Front Plant Sci* **4**, 497.

Thompson, K, Bakker, JP, Bekker, RM (1997) 'The soil seed banks of North West Europe: methodology, density and longevity.' (Cambridge university press:

Thompson, K, Band, S, Hodgson, J (1993) Seed size and shape predict persistence in soil. *Functional Ecology* 236-241.

To, A, Valon, C, Savino, G, Guilleminot, J, Devic, M, Giraudat, J, Parcy, F (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* **18**, 1642-51.

Tucker, SC (2001) The ontogenetic basis for missing petals in Crudia (Leguminosae: Caesalpinioideae: Detarieae). *International Journal of Plant Sciences* **162**, 83-89.

Tweddle, J, Turner, R, Dickie, J, 2002. Seed Information Database. Release,

Tweddle, JC, Dickie, JB, Baskin, CC, Baskin, JM (2003) Ecological aspects of seed desiccation sensitivity. *Journal of Ecology* **91**, 294-304.

Van Tienderen, PH (1991) Evolution of generalists and specialist in spatially heterogeneous environments. *Evolution* 1317-1331.

Vaz, TA, Davide, AC, Rodrigues-Junior, AG, Nakamura, AT, Tonetti, OA, da Silva, EA (2016) Swartzia langsdorffii Raddi: morphophysiological traits of a recalcitrant seed dispersed during the dry season. *Seed Science Research* **26**, 47-56.

Vázquez-Yanes, C, Orozco-Segovia, A (1993) Patterns of seed longevity and germination in the tropical rainforest. *Annual Review of ecology and Systematics* **24**, 69-87.

Verdier, J, Lalanne, D, Pelletier, S, Torres-Jerez, I, Righetti, K, Bandyopadhyay, K, Leprince, O, Chatelain, E, Vu, BL, Gouzy, J, Gamas, P, Udvardi, MK, Buitink, J (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. *Plant Physiol* **163**, 757-74.

Villiers, T (1974) Seed aging: chromosome stability and extended viability of seeds stored fully imbided. *Plant physiology* **53**, 875-878.

Vu, BL, Satour, P, Leprince, O (2003) The re-establishment of desiccation tolerance in germinated radicles of <I>Medicago truncatula</I> Gaertn. seeds. *Seed Science Research* **13**, 273-286.

West, MA, Yee, KM, Danao, J, Zimmerman, JL, Fischer, RL, Goldberg, RB, Harada, JJ (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. *The Plant Cell* **6**, 1731-1745.

White, CN, Proebsting, WM, Hedden, P, Rivin, CJ (2000) Gibberellins and seed development in maize. I. Evidence that gibberellin/abscisic acid balance governs germination versus maturation pathways. *Plant physiology* **122**, 1081-1088.

Whitmore, TC (1990) 'An introduction to tropical rain forests.' (Clarendon Press:

Wright, S (1986) 'Evolution: selected papers.' (University of Chicago Press:

Wyse, SV, Dickie, JB (2016) Predicting the global incidence of seed desiccation sensitivity. *Journal of Ecology*

Zeng, Y, Kermode, AR (2004) A gymnosperm ABI3 gene functions in a severe abscisic acid-insensitive mutant of Arabidopsis (abi3-6) to restore the wild-type

phenotype and demonstrates a strong synergistic effect with sugar in the inhibition of post-germinative growth. *Plant molecular biology* **56**, 731-746.

Zinsmeister, J, Lalanne, D, Terrasson, E, Chatelain, E, Vandecasteele, C, Vu, BL, Dubois-Laurent, C, Geoffriau, E, Le Signor, C, Dalmais, M (2016) ABI5 is a Regulator of Seed Maturation and Longevity in Legumes. *The Plant Cell* **28**, 2735-2754.

Definitions of terms

Term	Definition
Orthodox seeds	Orthodox seeds can withstand dehydration to below 0.1g water/g dry weight and stored dry for prolonged periods without losing viability.
Recalcitrant seeds	Seeds that are very sensitive to dehydration. The lowest safe moisture content for such seeds is considered 0.25 g water/g dry weight. Since the term 'recalcitrant' is generic we replace it here with the more accurate term 'desiccation sensitive'.
Intermediate seeds	Seeds that show a storage phenotype in-between orthodox and recalcitrant seeds. Typically, mature seeds are able to tolerate desiccation to about 0.1-0.14 g water/g dry but do not tolerate drying to lower water contents.
Desiccation sensitive seeds	See recalcitrant seeds
Desiccation tolerant seeds	See orthodox seeds
Seed dormancy	Prevention of germination of viable seeds under environmental conditions which are optimal for germination.
Quiescence	A state of metabolic arrest due to a lack of ideal environmental cues, particularly water.
Longevity	Seed longevity is defined as the period of seed viability during *ex situ* dry storage determined by inherited physiological traits.
Seed life span	The time from seeds reaching maturity on the mother plant until they germinate, are predated or have decayed or aged and lost germinability. Seed life span is influenced by intrinsic seed characteristics (e.g. dormancy, longevity and defence) and the environment.
Soil seed bank	The natural storage of seeds, dormant or quiescent, within the soil.

Abbreviations of genes cited

Abbreviation	Description
LEC1	Leafy Cotyledon 1
LEC2	Leafy Cotyledon 2
FUS3	Fusca 3
ABI3	ABA Insensitive 3
ABI4	ABA Insensitive 4
ABI5	ABA Insensitive 5
MYB118	MYB domain protein 118
PLATZ1	-
AGL67	Agamous-like 67
DREB2D	Dehydration Responsive Element Binding 2 D
LEA	Late Embryogenesis Abundant
HSP	Heat Shock Protein
ATS	Aberrant Testa Shape
AP2	Apetala 2
VTE1	Vitamin E
DOG1	Delay of Germination 1
TT	Transparent Testa
VP1	Viviparous-1
VP5	Viviparous-5

3

Understanding seed desiccation tolerance through the desiccation sensitive *Arabidopsis thaliana abi3-6* mutant

Alexandre Marques; Axel Kuil; Renake Nogueira; Harm Nijveen; Charles Somi; Richard G.H. Immink; Wilco Ligterink and Henk Hilhorst

Abstract

Seeds are important as means of plant reproduction. Desiccation tolerance is a mechanism allowing seeds to survive almost total water loss and to withstand long periods in a quiescent state. Dried seeds are able to survive extensive periods of drought and germinate when water is available. Moreover, desiccation tolerance avoids extinction of species and allows their *ex situ* conservation. However, certain species, especially from tropical rainforests, have desiccation sensitive (DS) seeds.

The molecular basis of seed desiccation sensitivity is poorly understood. Molecular studies of DS-seeded species are scarce. Here we aimed at investigating the molecular network associated with seed desiccation sensitivity, using the Arabidopsis mutant *abi3-6*. This mutant resembles the phenotype of natural DS-seeded species and may thus serve as a model, as it shows desiccation sensitivity from seed shedding onwards.

Transcriptomes of *abi3-6* mature seeds appear to be more correlated with 24h-imbibed wild type seeds than mature desiccation tolerant (DT) seeds. Furthermore, *abi3-6* seed development transcriptomes are more similar to non-dormant than to dormant wild type seeds. Altogether, we conclude that Arabidopsis seed maturation, which includes acquisition of desiccation tolerance, longevity and dormancy, is suppressed in the *abi3-6* mutant seeds and that the germination program is precociously activated in these seeds.

Introduction

Desiccation tolerance and seed longevity are fascinating phenomena that allow orthodox seeds to remain quiescent for long periods. The ability to survive long periods in a quiescent state helps the seeds to escape unfavourable periods and thus to avoid species extinction. The oldest viable seed registered is a 1300-year-old lotus seed (Sallon et al., 2008). However, some species bear desiccation sensitive (DS) seeds as a consequence of adaptation to environments where the conditions are conducive to fast germination (Berjak and Pammenter, 2008).

Seed desiccation sensitivity is a major obstacle for plant biodiversity conservation. The long-term storage of these seeds is currently unfeasible. Consequently, DS seed behaviour in a context of climate change, habitat loss due to deforestation and the limited knowledge on DS seed storage, places these species at a high risk of extinction (Li and Pritchard, 2009). Therefore, a deeper understanding of the physiological and genetic alterations responsible for seed desiccation sensitivity is necessary to support conservation efforts.

DS has been associated with increased seed size, thinner seed coats, lack of dormancy, chlorophyll retention, as well as reduction of lipid content and other seed reserves (Daws et al., 2006; Farnsworth, 2000; Hamilton et al., 2013).

The molecular mechanisms underlying seed desiccation tolerance have been investigated, revealing the importance of a number of protectant and regulatory proteins. The protectant proteins include late embryogenesis abundant proteins (LEAs), oleosins, anti-oxidant enzymes and small heat shock proteins (sHSPs) (Pammenter and Berjak, 1999). DS seeds, such as seeds of *Castanospermum australe* and *Camelia sinensis,* fail to accumulate LEAs and to prevent damage caused by reactive oxygen species (ROS) (Chen et al., 2011; Delahaie et al., 2013).

The three transcription factors, *ABI3*, *FUS3* and *LEC2,* are important regulators of acquisition of desiccation tolerance in seeds and knock-outs for these genes are DS (To et al., 2006). A narrow window of desiccation tolerance during seed maturation has been observed previously in the DS seeds of the Arabidopsis null mutant *abi3-5* (Ooms et al., 1993). The double mutant *abi3-1/aba-1* showed more severe symptoms and its seeds did not show any sign of desiccation tolerance when excised and dried, unless when they were dried inside the siliques or dried slowly (Ooms et al., 1994). In the double mutant *abi3-1/aba-1*, the ABA deficiency caused by aba-1 enhances the mild abi3-1 mutation exerting a more pleiotropic and severe phenotype than the *abi3-5* mutation (Ooms et al., 1994). We assume that similar mutations in ABA-related regulatory factors can occur naturally and these might be responsible for the observed similar phenotypical alterations in DS-seeded species.

We selected the null mutant *abi3-6*, which is in a Columbia (Col-0) background, as a model to study desiccation sensitivity of seeds (Nambara et al., 1994). We aimed at further exploration of the molecular mechanisms underlying the phenomenon of desiccation sensitivity and relating it to detailed phenotyping and dynamics of the transcriptome along seed development.

Understanding the molecular orchestration responsible for the acquisition of desiccation tolerance in orthodox seeds, and its absence in DS seeds, is crucial for the development of storage protocols for DS seeds and the improved storage of DS crop seeds, such as coffee, rubber and citrus.

Results

Acquisition of desiccation tolerance during seed development

We assessed the landscape of acquisition of desiccation tolerance and various seed characteristics during seed development in Col-0 and *abi3-6* (Figure 1). Initially, we checked when germinability is acquired by placing fresh seeds of both lines to germinate in a gibberellin solution. We observed that *abi3-6* acquired germinability earlier than Col-0 (Figure 2A). We also investigated how desiccation inside or outside the siliques influenced seed survival. Although *abi3-6* seeds have been reported as being DS, they acquire some extent of desiccation tolerance during maturation. Similarly to *abi3-6* seeds (Figures 2B, C), seeds of a full *ABI3* knock-out in the Landsberg erecta (Ler) accession (*abi3-5*) display a transient increase in desiccation tolerance during maturation (Ooms et al., 1993). Furthermore, we observed a second transient increase until 23 days after flowering (daf) (Figure 2C).

Figure 1. Fully developed seeds of Col-0 (left) and *abi3-6* (right) displaying considerable phenotypical differences.

Applying drying to intact siliques resulted in transient desiccation tolerance at the beginning of seed maturation at 11 daf in both Col-0 and abi3-6 seeds (Figure 2B), however, to a significantly lower extent in the abi3-6 background. To get insight into these transient alterations in DT, we selected developmental stages of both genotypes to investigate the transcriptomic changes possibly associated with the observed acquisition and loss of DT.

Col-0 displayed an almost linear loss of water during seed maturation (Figure 3). The rate of water loss of *abi3-6* was also linear until 18 daf. From this point until 22 daf, *abi3-6* seeds seemed to resist water loss (Figure 3). The water content of ~47% on a fresh weight base, which is reached at 18 daf, has been shown to be harmful causing permanent wilting for non DT tissues (Born et al., 2014).

Wild DS seeds tolerated, on average, water losses until 39.8 ± 9.4 % water content on a fresh weight basis (Chapter 2, Figure 2). This limiting water content for loss of DS seed viability has been shown to be similar to the permanent wilting point in vegetative non-DT tissues, around -2 MPa (~40% water content) (Andrade et al., 2003; Born et al., 2014; Levitt, 1980; Roberts and Ellis, 1989).
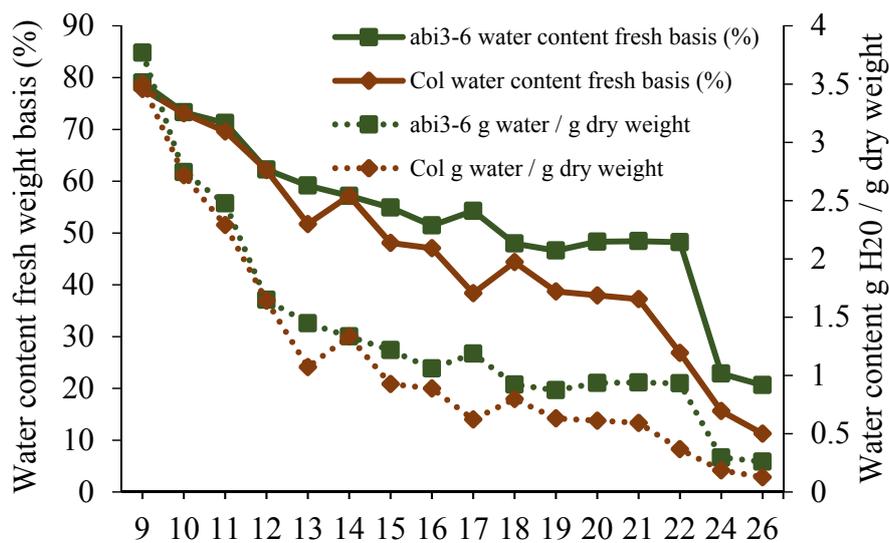
Figure 3. Water content during maturation of Col-0 and *abi3-6* seeds. Solid lines show the water content in fresh seeds and dotted lines the water content per gram of seed dry weight. At each time point, water content was measured in a pool of seeds harvested from 10 siliques.

The transcription factor ABI3 is an orchestrator of seed maturation

Desiccation tolerance is a multifactorial process involving various processes, including cellular protection against oxidation and structural damage, metabolic arrest, compartmentalization of amphiphilic compounds, and cytoplasmic glass formation

(Crowe, 2014; Hoekstra et al., 2001). Here we used transcriptome analysis of developing Col-0 and *abi3-6* seeds to get insight in the gene regulatory network associated with the desiccation sensitivity of *abi3-6* seeds. Based on the desiccation tolerance screening (Figure 2B), we observed that desiccation tolerance in Col-0 wild type seeds is acquired soon after embryogenesis and is established at 11 daf when seeds are dried inside the siliques. Taking this into account, we decided to study the early acquisition of desiccation tolerance monitoring the transcriptome transitions that allowed the survival of *Col-0* from 0% at 10 daf to 91% at 11 daf seeds and the low-acquisition of desiccation tolerance in *abi3-6* in the same time window. We also decided to investigate the loss of desiccation tolerance of *abi3-6* seeds that survived 79% at 18 daf but dropped to 20% at 20 daf, while *Col-0* seeds remained DT at these stages.

At the later stages of Col-0 seed development (11-18 and 18-20 daf), we observed a greater down-regulation of genes compared to *abi3-6* (Supplementary Figure 1). From 18 to 20 daf, when dehydration was most pronounced in Col-0, 2,188 transcripts were down-regulated as compared to only 820 transcripts in *abi3-6* in the same period.
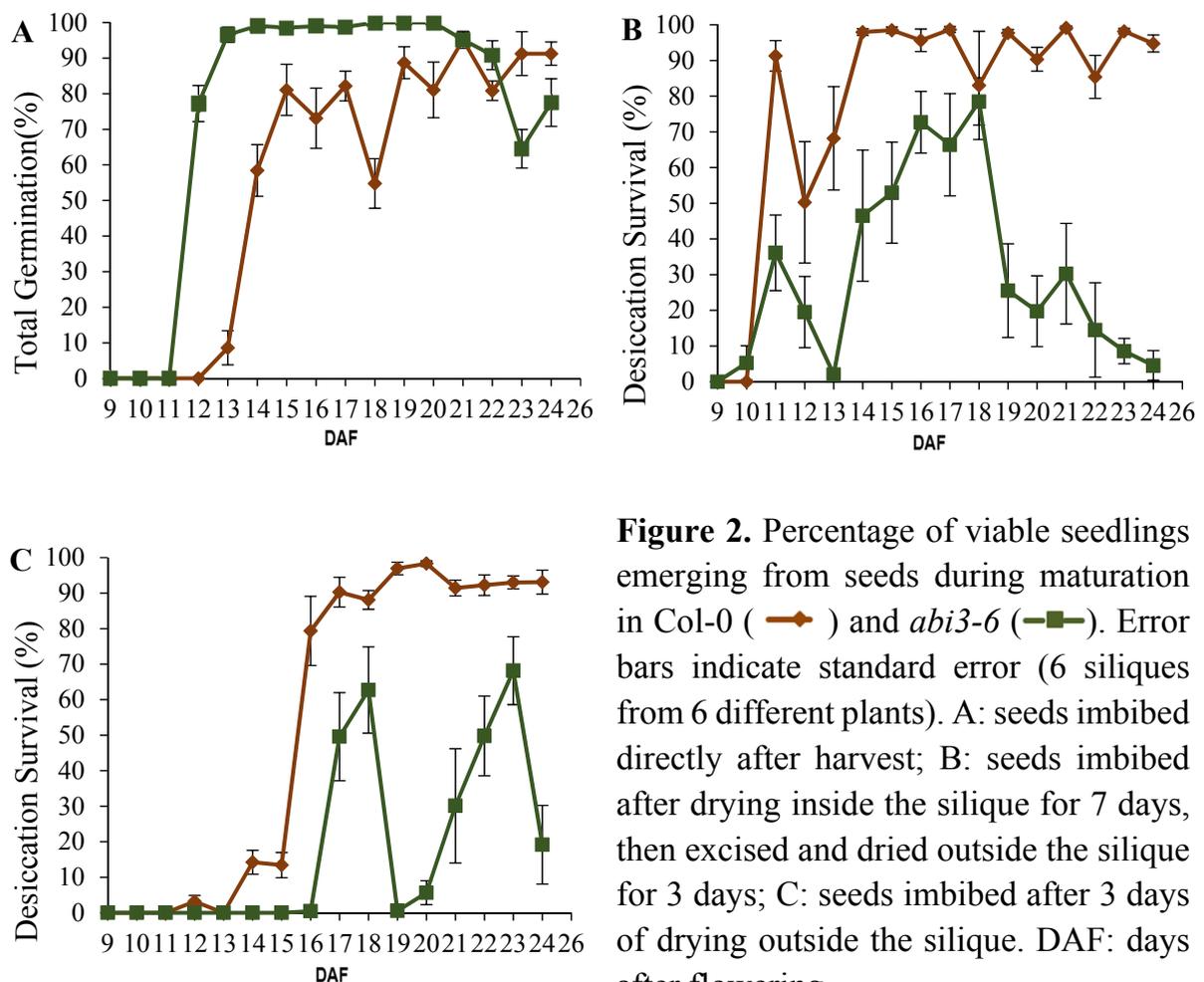
Figure 2. Percentage of viable seedlings emerging from seeds during maturation in Col-0 (◆) and *abi3-6* (■). Error bars indicate standard error (6 siliques from 6 different plants). A: seeds imbibed directly after harvest; B: seeds imbibed after drying inside the silique for 7 days, then excised and dried outside the silique for 3 days; C: seeds imbibed after 3 days of drying outside the silique. DAF: days after flowering.

There was a clear separation of Col-0 and *abi3-6* seed transcriptomes in the PCA plot (Figure 4). The first two principal components explain most of the variation observed between the genotypes and developmental stages (PC1 60% and PC2 26%). Later time points of Col-0, when seeds matured, can be found in the top left region of the PCA plot, whereas similar time points of *abi3-6* are at the bottom left. Roughly, Col-0 late maturation is positively biased towards PC1 and *abi3-6* later time points towards PC2. In order to visualize the genes that contribute for the distribution of these samples on the PCA we plotted the gene loadings based on their PC1 and PC2 values (Supplementary Figure 2) (total gene loadings can be found in Supplementary file 1). Genes with high PC1 are those that contribute mostly to the distribution of Col-0 late maturation samples in the PCA and genes with high values for PC2 are most discriminative for *abi3-6* late time points. To support our analysis, we marked on this plot the genes regulated by ABI3 identified in a previous study by Monke et al. (2012). The gene loadings of this PCA plot showed a clear biased distribution towards the upper right corner of the graph (high in PC2), suggesting that these genes are downregulated by ABI3 in Col-0 late developmental stages.
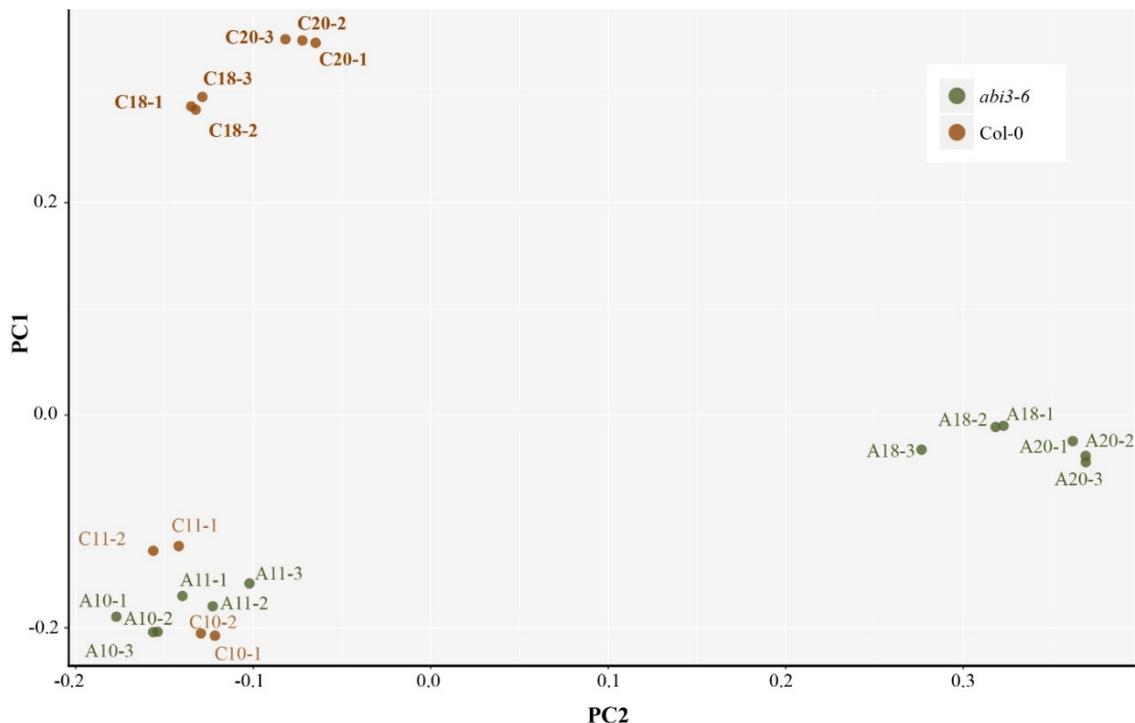

Figure 4. PCA loadings plot of PC1 versus PC2 from RNA-seq samples of different stages of Col-0 and *abi3-6* seed development. Together, PC1 and PC2 explain 86% of the total variance. The coding used for each sample refers to genotype (C=Col-0. A=*abi3-6*), days after flowering (daf) and biological replicate (-1, -2, -3).

Next, we performed a GO-enrichment analysis on two gene lists ranked based on their PC1 or PC2 values and searched enriched GO terms that appear densely at the top of each list (Supplementary file 1) (Eden et al., 2007; Eden et al., 2009). The PC1 showed 46 enriched terms that have coherent association with the seed maturation quiescent status, such as *removal of superoxide radicals*, *lipid storage*, *regulation of seed germination*, *abscission* and *response to water*. On the other hand, PC2 showed 156 enriched terms that resemble an actively developing tissue, *such as photosystem II assembly*, *tRNA processing*, *ribosome biogenesis*, *defense response to fungus*, *defense response to bacterium* and *lateral root formation*. Thirteen overlapping GO categories were found between PC1 and PC2, including *embryo development* and *embryo development ending in seed dormancy*.

To obtain more insight in the transcriptional differences between Col-0 and *abi3-6* during their seed development, we performed a GO-enrichment analysis of the differentially regulated genes comparing genotypes and comparing different time points per genotype (developmental time series). Col-0 and *abi3-6* seeds are still not DT at 10 daf and they are transcriptionally similar in the PCA (Figure 4). However, when we directly compare both transcriptomes (c10 compared to a10), we observe striking differences that reveal that both genotypes have already DT-related transcriptional differences at 10 daf. The Gene Ontology (GO) enrichment analysis of the up-regulated gene set (c10 compared to a10) reveal important DT-related categories such as *seed development* and *embryonic development ending in seed dormancy* (see Supplementary file 2 for the whole list of GO enriched categories). These categories contain transcripts of ten LEAs, one oleosin, one caleosin and one cruciferin gene.

Comparing the different time points for the individual genotypes, we observed various enriched GO categories among the up- and down-regulated genes (Supplementary files 3 and 4, respectively). These categories reflect the phenotypical differences between Col-0 or *abi3-6* seed development, of which some remarkable ones are discussed in detail below (Figure 5). The *abi3-6* enriched GO categories reveals activity of genes linked to active metabolism, photosynthesis, growth and tissue differentiation. Col-0 seeds, conversely, undergo maturation through a series of transcriptomic changes, ultimately resulting in developmental arrest. The *abi3-6* seeds lack a major controller of metabolic suppression and, consequently, express a transcriptional activity linked to an active metabolism which is incompatible with its fate of dehydration.

Photosynthesis in drying seeds

The fact that the *abi3-6* seeds remain green (Figure 1) suggests that its photosynthetic machinery is not dismantled upon maturation. Indeed, in the *abi3-6* up-regulated gene sets the GO categories related to photosynthesis, such as *photosynthetic electron*

transport in photosystem I, photosynthesis light harvesting in photosystem II are enriched (Figure 5).
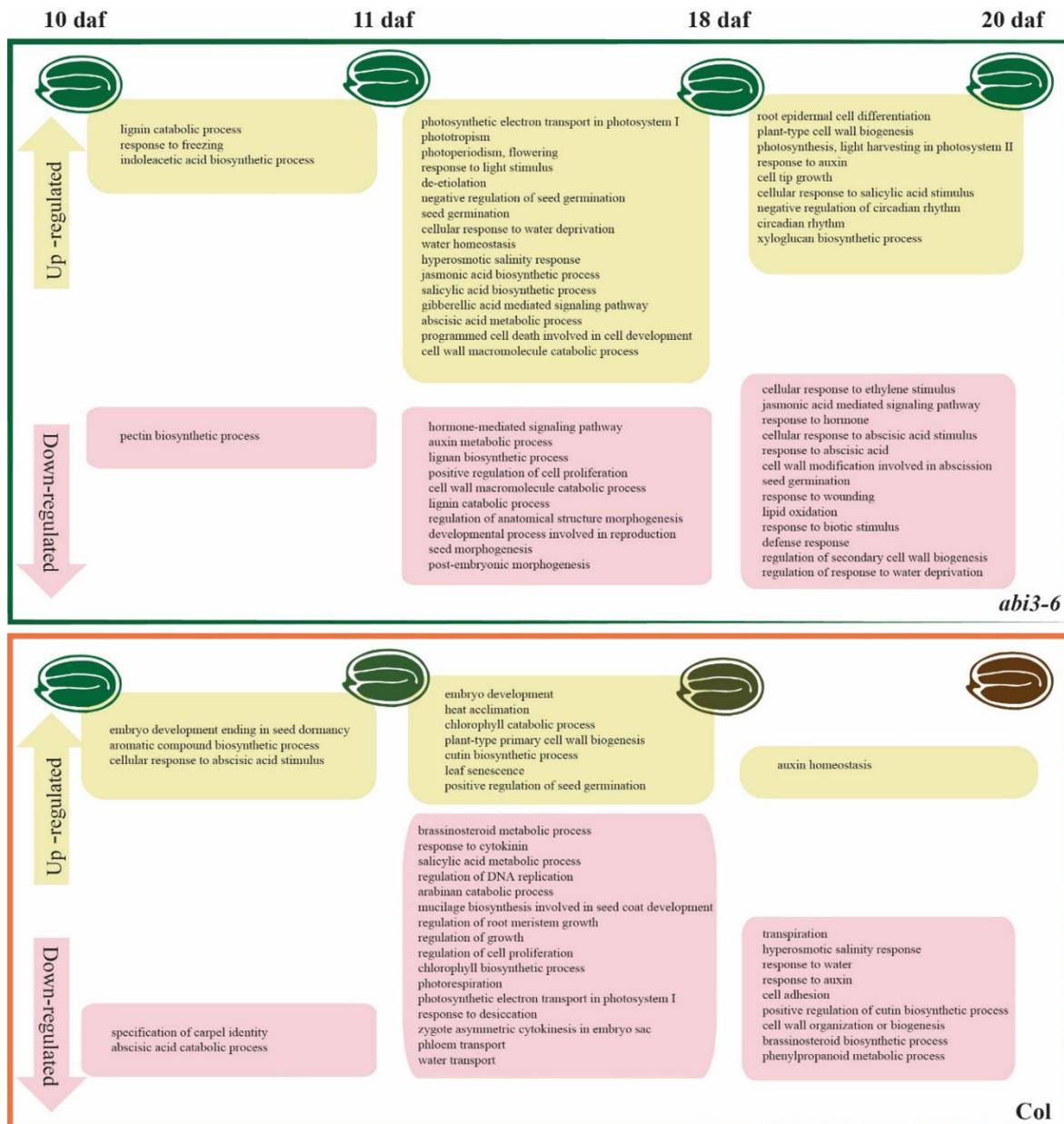
Figure 5. GO categories of differentially-regulated genes enriched in specific transitions. A highlight of specific GO categories associated with seed maturation comparing Col-0 and *abi3-6* is shown. The total list of up-regulated genes in each transition and their GOs can be found in the Supplementary Files 3 and 4.

Arabidopsis seeds normally break down their chlorophyll during maturation. However, some Arabidopsis mutants, e.g. *non-yellow coloring1* (*nyc1*) and *nyc1-like* (*nol*), retain chlorophyll and seed storability is negatively affected in these mutants

(Nakajima et al., 2012). Chlorophyll retention is one of the pleiotropic effects in the Ler *abi3-5* mutant seeds during maturation, which can be rescued by overexpression of *STAY GREEN 1* and *2* (*SGR1/2*), but the seeds remain DS (Delmas et al., 2013b). Thus, chlorophyll retention is not a causal factor of seed desiccation sensitivity in this mutant. Furthermore, the majority of species with green embryos in East-Australian rainforest are DS, however also species with DS seeds without chlorophyll retention have been identified (Hamilton et al., 2013).

Growth and cell-division are not compatible with desiccation

Loss of desiccation tolerance has been associated with the resumption of cell division in germinating seeds (Osborne et al., 2002) and, more specifically, upon the resumption of DNA synthesis prior to cell division (Boubriak et al., 2000; Daphne, 2000; Faria et al., 2005). It has been shown that cells in the G2 phase of the cell cycle, with duplicated DNA, are more sensitive to stress than cells that are still in the G1, pre-synthetic phase (Śliwińska, 2003). Here the down-regulated set in Col-0 is enriched in the GO categories *regulation of meristem growth*, *regulation of DNA replication*, *chromosome segregation*, *regulation of root meristem growth*, *regulation of growth* and *regulation of cell proliferation,* indicating suppression of growth related mechanisms. On the contrary, in the *abi3-6* up-regulated set of transcripts, the enrichment of the GO categories *cell tip growth* and *cell wall biogenesis* is apparent (Figure 5).

We investigated the averaged expression patterns of genes that have been used as markers for cell division and tissue growth, such as tubulins, expansins and GA-biosynthesis (de Castro et al., 1995; Hedden and Phillips, 2000; Li et al., 2003), so we can infer the developmental status of the *abi3-6* and Col-0 embryos. We noticed that these transcripts follow a strikingly divergent pattern between Col-0 and *abi3-6* (Figure 6 a, b and c) reflecting the lack of developmental arrest in *abi3-6* seeds.

Absence of developmental arrest in abi3-6 seeds reflects a transcriptional lack of dormancy

Seed dormancy is acquired early during seed maturation of wild type Arabidopsis seeds (12 daf) (Ooms et al., 1993). This is in accordance with the GO category embryo development ending in seed dormancy enriched exclusively in the up-regulated gene set from 10 to 11 daf in Col-0 seeds.

Desiccation tolerance is acquired during development of Col-0 seeds and to a lower extent in *abi3-6,* but it is eventually lost in *abi3-6* seeds. We correlated gene expression in Col-0 and *abi3-6* with the survival to desiccation inside the siliques in their respective stages (Supplementary file 5).

The expression pattern of *DOG1* is the second most correlating with the level of desiccation tolerance in Col-0 and *abi3-6* during development, after *CYSTEINE PROTEASE 2* (Supplementary file 4). *DOG1* is a gene that explains most of the variation observed for seed dormancy in Arabidopsis and some other species and it is also associated with seed longevity (de Souza Vidigal et al., 2016; Nguyen et al., 2012).

Additionally, expression levels of a protease, cruciferins and albumin also correlated with desiccation tolerance. This suggests an involvement of storage protein and/or amino acid accumulation in the acquisition of desiccation tolerance. This is in accordance with Arabidopsis mutants *abi3, fus3* and *lec2* that have DS seeds and also display a considerable reduction of storage protein accumulation (Kagaya et al., 2005).

Subsequently, we performed a Spearman correlation between our transcriptomes and publicly available datasets (Supplementary Table 1). These datasets represent Arabidopsis seed development from embryogenesis to maturation, desiccation and imbibition. Although the correlations were calculated between different experiments, growth conditions and gene quantification platforms, we expected the log transformed ranked correlation to be useful to find the closest match available. This was confirmed by the fact that our Col-0 transcriptomes (from C10 to C20) matched other published Col-0 transcriptomes from comparable developmental stages (Schmid et al., 2005).

For the *abi3-6* mutant, the transcriptomes correlated best with similar early stages of Col-0 at early development (10 and 11 daf), but, remarkably, for the later stages (18 and 20 daf) the transcriptomes were mostly similar to those of Col-0 24h-imbibed seeds. Genes that have been shown to be up-regulated in dormant seeds (Cadman et al., 2006) followed a steep increasing pattern during Col-0 maturation but they showed a minimal increase in *abi3-6* seeds (Figure 6d). Conversely, the genes up-regulated in after-ripened seeds followed a slight increase during *abi3-6* maturation but in Col-0 seeds they dropped drastically (Figure 6e). The behaviour of these two gene sets, after-ripened and dormant genes, were remarkably contrasting in Col-0 but in *abi3-6* seeds the control of these genes was disrupted.

We also observed a wider maturation window where both genotypes transcriptionally diverged the most, from 11 to 18 daf (Figure 4). The *abi3-6* mutant shares more up-regulated genes (comparing 18 against 11 daf) with imbibed after-ripened seeds than with dormant ones whereas Col-0 showed the opposite pattern (Supplementary Figure 3). Not only the number of shared genes between particular stages are remarkable but also the GO categories that are enriched in these gene sets. The up-regulated genes, comparing 18 against 11 daf, that Col-0 shared with dormant seeds showed diverse GOs related to response to stress and response to abscisic acid. These categories are recurrent among dormancy and desiccation tolerance studies suggesting a transcriptional proximity between these physiological mechanisms. On the other hand, the up-regulated genes, comparing 18 against 11 daf, that *abi3-6* shared with

after-ripened seeds showed diverse GOs related to growth and active development, such as plant-type cell wall modification involved in multidimensional cell growth and unidimensional cell growth.

Induction of desiccation tolerance

The induction of desiccation tolerance has been studied in various model systems. Applying ABA or osmotic (e.g. PEG) solutions to germinated desiccation sensitive orthodox seeds can render them DT and the molecular changes involved in this induction of desiccation tolerance in Arabidopsis have been unravelled (Dekkers et al., 2015).

We investigated how the genes with transcriptomic changes involved in the induction of *desiccation tolerance* behave during Col-0 and *abi3-6* seed maturation (Maia et al., 2011). We detected two very distinguished patterns between Col-0 and *abi3-6* seed maturation of the up and down-regulated genes (Figure 6f,g). The genes up-regulated upon induction of desiccation tolerance accumulated more markedly in Col-0 than in *abi3-6* (Figure 6f). However, the down-regulated genes showed a more dramatic difference. The genes that are down-regulated upon induction of desiccation tolerance were sharply suppressed in Col-0 but in abi3-6 seeds they remained highly expressed (Figure 6g).

We took a closer look to the genes responsible for cell protection against desiccation damage. Protective proteins, such as LEAs and oleosins, have been associated with the acquisition of desiccation tolerance (Hundertmark and Hincha, 2008; Leprince et al., 1998). The transcripts coding for these proteins show divergent patterns between Col-0 and *abi3-6* (Figure 6h,i). Oleosins and LEAs are important desiccation protectant proteins that failed to accumulate in *abi3-6*. Furthermore, both *abi3-6* and Col-0 up-regulate oleosins at 11 daf. This is the developmental stage when desiccation tolerance is acquired. Col-0 seeds acquire desiccation tolerance to much higher levels of survival but *abi3-6* seeds also show some considerable survival at this stage.

Transcriptomic analysis identifies sets of developmentally regulated processes during seed maturation

To determine how transcript abundance changes during seed maturation, we clustered the total transcripts into ten dominant expression patterns (DPs) (Supplemental Figure 4), using the k-means clustering method (Sturn et al., 2002). We set the software to obtain ten dominant DPs since this gave peaking transcriptional modules for each developmental stage and for both genotypes (Supplementary Figure 4).
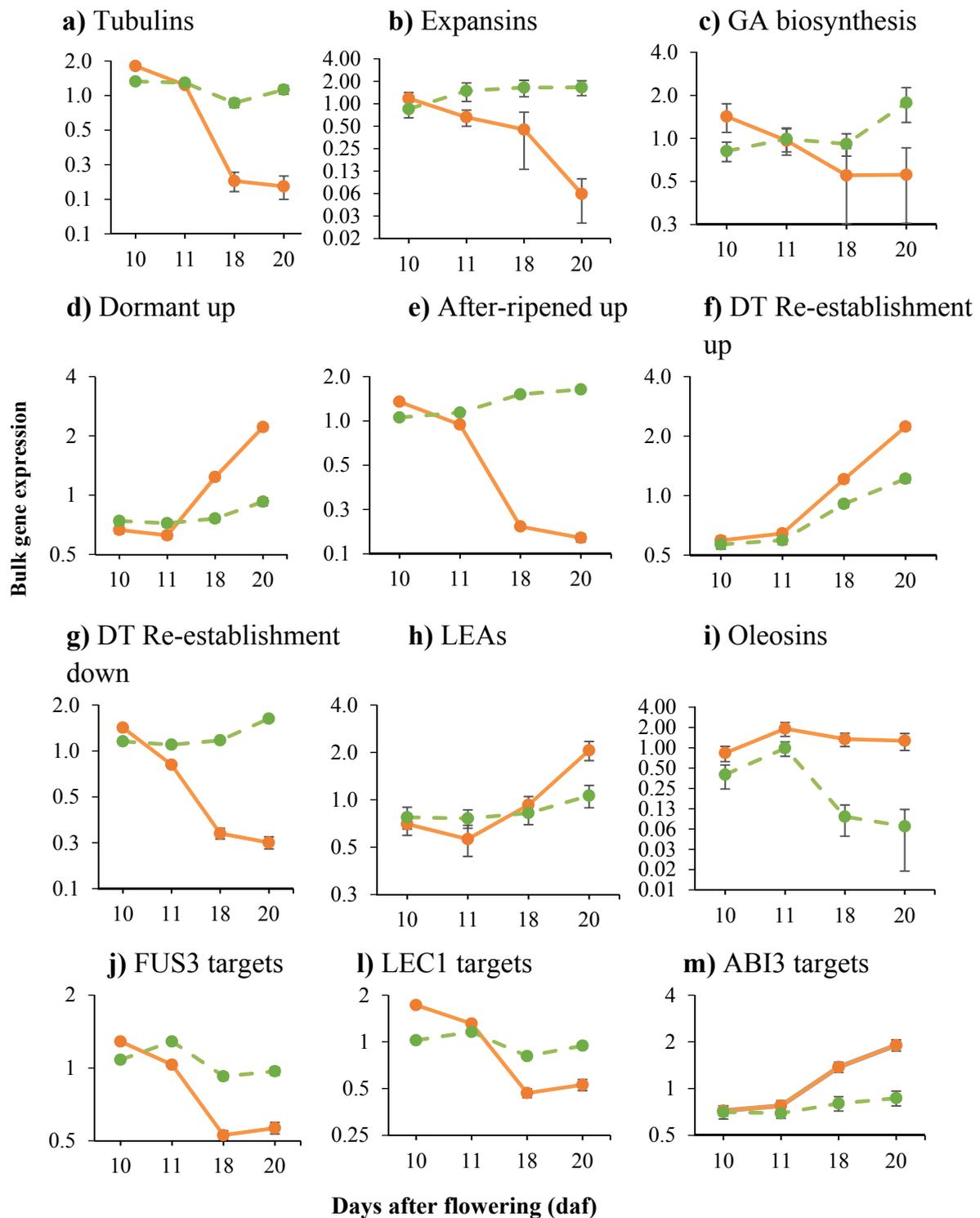
Figure 6. The graphs display the average of target gene groups that are relevant during seed maturation and desiccation tolerance acquisition. Before taking the average on each time point, they were normalized by the average of all the samples for each gene. The green lines represent *abi3-6* samples and the orange ones represent Col-0. The error bars display the SE.

Eight of the co-expressed gene sets consisted of transcripts with high expression at only one developmental stage and specifically in one genotype whereas the other two co-expressed gene sets did not show clear peaking patterns. These eight DPs correspond to gene clusters that peaked specifically in the samples (C10, C11, C18, C20, A10, A11, A18 and A20) (See Supplementary file 6 for the list of genes in the clusters and their enriched GO categories). These expression patterns suggest the occurrence of seed maturation processes related to specific stages and genotypes.

Predictive functions of these processes were determined for each of the DPs by analysis of enriched GO terms (FDR < 0.05). For example, the DP-C10 gene set was overrepresented for genes with GO terms such as *regulation of cell cycle, DNA packaging* and *chromatin assembly* (Supplementary figure 5). The acquisition of desiccation tolerance observed in Col-0 seeds at 11 daf can be a consequence of the DNA protection performed at 10 daf. Seed maturation in Col-0 is characterized by nuclear size reduction and increased chromatin condensation. However, chromatin of *abi3-5* seeds maintains remains uncondensed, and therefore, these seeds have unprotected DNA throughout seed maturation (van Zanten et al., 2011).

Another gene set, DP-A18, was significantly enriched for GO terms associated with *RNA biosynthesis, RNA processing, rRNA metabolism* and *ribosome biogenesis*. Interestingly, the category *post-embryonic development* was enriched only in the early stage 10 daf in Col-0 but in *abi3-6* it was enriched in the late stages 18 and 20 daf. Together, these categories show that the *abi3-6* seeds remain active with characteristics of growth, reflecting the lack of suppression of metabolism and developmental arrest. These features of activity are incompatible with desiccation tolerance and are thus possible main factors that make these seeds start to lose desiccation tolerance from 18 daf onwards.

The developmental stage 20 daf showed a striking phenotypical difference between Col-0 and *abi3-6*, i.e. ~98% and ~5% survival after desiccation of excised seeds, respectively. Therefore, we focused on the divergent GO categories between Col-0 and *abi3-6* at this time point.

The gene set DP-C20 showed overrepresented GO categories mainly associated with abiotic stresses such as *response to reactive oxygen species*, *response to hydrogen peroxide* and *response to heat*.

Conversely, the gene set DP-A20 had very broad enriched GO categories related to abiotic stress: *response to stress* and *response to abiotic stimulus*. Moreover, enrichment of categories related to biotic defence such as *response to stress* and *defence response* were apparent.

Protection against reactive oxygen species has been demonstrated to be important for the acquisition of desiccation tolerance in orthodox- and to be deficient in DS-seeds (Berjak and Pammenter, 2013). However, the response to biotic stress may counteract

the abiotic stress responses (Mauch-Mani and Mauch, 2005). This is possibly one of the factors that hinders acquisition of desiccation tolerance in *abi3-6* seeds.

A gene regulatory network of acquisition of seed desiccation tolerance

To identify regulatory processes that are controlling the acquisition of desiccation tolerance, two unweighted gene co-expression networks were generated, named Col-net and abi36-net. Each network was generated by processing the transcriptomes of all the samples of their respective genotype. The edge adjacency threshold was set at 0.95, resulting in two networks consisting of 1.2 million edges for abi36-net and 1.8 million for Col-net. Network visualization was carried out in Cytoscape using the organic layout (Figure 7).

There are remarkable conformational differences between the two networks. Col-net has a greater number of nodes, is more condensed and presents a thicker transition zone between the two main clusters, which we named "bridge". Roughly, the lower cluster of Col-net contains the early/ DS and the top cluster contains late/DT developmental stages. We assume that the bridge represents an important cluster of regulatory genes that controls the transition between the DS and DT state. There is a large overlap in Col-net between the genes in the bridge and in the gene set DP-C11, which contains genes that peak when desiccation tolerance is acquired in Col-0. However, the abi36-net shows a thinner bridge that does not contain many genes from the DP-A11 gene set (Figure 7).

The Col-net bridge contains 220 genes while the abi36-net contains only 97 (Supplementary file 7). The Col-net bridge contains numerous genes involved in hormonal signalling and seed development with overrepresented genes for *plant hormone signal transduction* (KEGG, p-value $8,9e^{-3}$). This pathway contains regulatory genes involved in seed maturation such as *ATBZIP12* (Bensmihen et al., 2005), *INDOLE-3-ACETIC ACID INDUCIBLE 30* (*IAA30*), (Braybrook et al., 2006) and *PYR1-LIKE 7* (*PYL7*) (Footitt et al., 2011). Moreover, we also found the important regulator of acquisition of seed desiccation tolerance, *FUSCA3* (*FUS3*), present in the Col-net bridge (Wang and Perry, 2013). This TF was not present in the abi36-net bridge but was found in this genetic background in the middle of the upper cluster.

We performed a transcription factor (TF) binding motif over-representation analysis (Jin et al., 2017) of the bridge gene set and we significantly (*p*-value < 0.01) identified *FUS3*, among 12 other TFs, as a regulator of these genes (Supplementary Table 2). LEAFY COTYLEDON 1 and LEAFY COTYLEDON 2 (LEC1 and LEC2), which are also essential for seed development and acquisition of desiccation tolerance, are among the TFs with enriched binding motifs, but the genes encoding these TFs are not present in the Col-net bridge. An identical analysis of the genes in the abi36-net

bridge (Supplementary table 3), revealed an enrichment for FUS3, but not for LEC1 and LEC2.

The interaction between the essential TFs FUS3, LEC1 and ABI3 are important for proper seed development, maturation and acquisition of desiccation tolerance (To et al., 2006). We detected a different expression fate, between *abi3-6* and Col-0, of the genes that are targets of FUS3, LEC1 and, as expected, ABI3 (Figure 6j,l and m, respectively). Complex regulatory interactions have been demonstrated between these TFs. For example, *ABI3* expression is controlled by LEC1 and LEC2 and FUS3 and ABI3 control each other reciprocally (To et al., 2006). The present study demonstrates a strong suppression of the gene targets of FUS3 and LEC1 towards late maturation in Col-0, whereas these TFs appear to have lost this suppressive control in the *abi3-6* mutant.

| DP-C10 | DP-C11 | DP-C18 | DP-C20 |

| DP-A10 | DP-A11 | DP-A18 | DP-A20 |
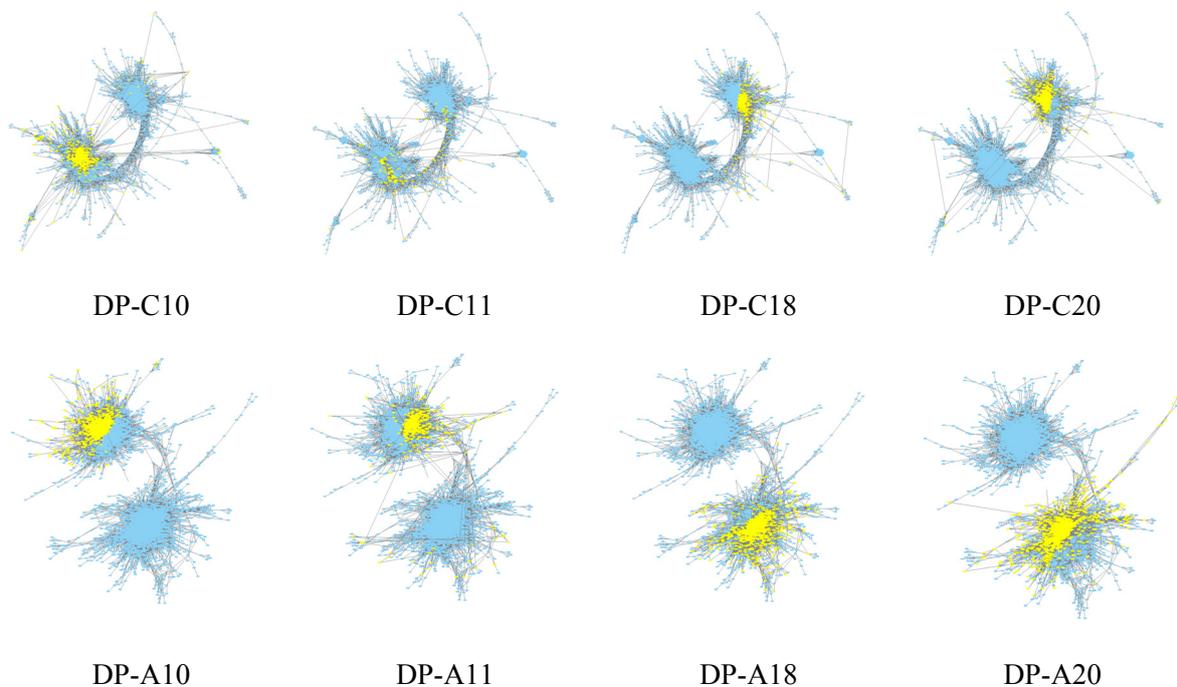
Figure 7. Seed maturation network with nodes (transcripts) coloured according to the dominant patterns (DP) to which they belong, showing expression regions on the network for the different DPs. The dominant patterns are a set of genes that peaks specifically at one stage and genotype. The letter indicates the genotype and the numbers the developmental stage (e.g. C10: Col-0 at 10 daf).

Discussion

Since Roberts (1973) first described the phenomenon of seed desiccation sensitivity, authors have formulated hypotheses in an attempt to explain its evolutionary and physiological origin(s). Pammenter and Berjak (2000) suggested that "Species producing recalcitrant seeds are largely, but not exclusively, confined to habitats continuously conducive to germination and seedling establishment". It implies that desiccation presents a disadvantage for adaptation in such habitats and therefore desiccation tolerance will eventually have been lost resulting in DS seeds. One of the early hypotheses trying to address the cause of desiccation sensitivity of seeds was raised by Farrant et al. (1993). They suggested that DS seeds lack on the genome the competence to produce LEA proteins and as a result remain sensitive to desiccation.

Due to a major impracticality of developing genetic studies with wild DS seeds, we resorted to the use of the Arabidopsis *abi3-6* mutant, which has a DS phenotype. Considering such DS Arabidopsis mutants seeds, Okamoto et al. (2010) defend the hypothesis: "Severely ABA-insensitive mutants, such as *abi3-6* or *snrk2d/e/i* triple mutants, cannot enter seed maturation processes". Similarly, Delmas et al. (2013b) discussed that: "*abi3-6* exhibits pleiotropic effects during seed development that bypass the embryo maturation phase and transition directly into the germination and vegetative phases. This lack of an embryo maturation phase results in seeds that remain green due to failure of embryo degreening in the severe *abi3* alleles". Based on the present study and the large variability of seed desiccation sensitivity phenotypes observed in nature (Chapter 2), we also support the hypothesis that seed desiccation sensitivity is caused by skipping maturation program, or part of it, and proceeding towards germination.

Does seed maturation take place in DS seeds?

Based on the patterns of water loss in seeds and siliques (Figure 3B), it is clear that *abi3-6* seeds start to dehydrate much earlier than their siliques. It indicates that the *abi3-6* seeds dehydrate actively, even without having the capacity to survive desiccation. It seems that there are two mechanisms acting in parallel during *abi3-6* seed development. One is driving the seed towards desiccation and the other towards vegetative development. Studies have demonstrated and supported the participation of both ABA-dependent and ABA-independent pathways in response to dehydration (Maia et al., 2014; Yamaguchi-Shinozaki and Shinozaki, 2006). With the disruption of the ABA-dependent pathway, by the *abi3* knock-out, it is possible that the ABA-independent pathway still induces a dehydration response albeit without engaging the complete desiccation tolerance machinery normally provided by the ABA-dependent pathway.

We observed, for example, the up-regulation of GO category *seed maturation* in Col-0 and *abi3-6* seeds. However, this category contains three cruciferins and one cupin which were upregulated in *abi3-6* only between 10 and 11 daf whereas in Col-0 seeds this GO category was upregulated both from 10 to 11 daf and between 11 and 18 daf. Col-0 seeds displaced expression of the same genes, but also *HISTIDINE KINASE 1* (*HK1*), *SUCROSE SYNTHASE 3* (*SUS3*) and *TEMPERATURE-INDUCED LIPOCALIN* (*TIL*).

Seed dehydration during late maturation appears to be a mechanism triggered and controlled by the seeds themselves and not by the mother plant. The fact that seed dehydration can occur inside moisty fleshy fruits indicates that the route of water loss initially is not by evaporation (Murray, 1988). The path of seed water loss during maturation is currently thought to be mainly by water displacement from the cells by reserve deposition within storage tissues (Black and Pritchard, 2002). Orthodox seeds normally lose water gradually during maturation, while accumulating reserve and protectant proteins and other compounds that enable survival of desiccation. During this process, they shut down their metabolic machinery to prevent the generation of toxic molecules during dehydration (Holdsworth et al., 2008; Verdier et al., 2013).

The Arabidopsis *abi3, fus3,* and *lec2* mutants show defects in the accumulation of seed reserves and have DS seeds (Angeles-Núñez and Tiessen, 2011; Focks and Benning, 1998; Keith et al., 1994; Zeng and Kermode, 2004). The storage of lipids in oil bodies and proteins in vacuoles, is impaired in these mutants and, therefore, they do not resist cellular collapse upon drying, resulting in a wrinkled phenotype (Fig. 1; (Leprince et al., 1998)). Furthermore, many genes of the *ABI3* regulon encode seed storage compounds (Monke et al., 2012; Zeng and Kermode, 2004).

Here we demonstrated a high correlation of storage protein gene expression with acquisition of desiccation tolerance in Col-0 and its considerable reduction in *abi3-6*. Moreover, the gene *CYSTEINE PROTEASE 2* showed an expression that was tightly correlated with desiccation tolerance levels, revealing the importance of protein storage dynamics for acquisition of seed desiccation tolerance.

In the PCA analysis, the GO categories *lipid storage* and *seed oilbody biogenesis* are enriched in PC1 but not in PC2. These principal components represent the gene loadings that contribute the most for the development of Col-0 and *abi3-6,* respectively. The lack of lipid has been evidenced in *abi3-6* and also in DS-seeded wild species (Hamilton et al., 2013). However, the lack of oil is not *per se* the sole determinant of seed desiccation sensitivity because DT seeds have been described with very low oil contents, such as those of *Vigna angularis* with 0.4% (Earle and Jones, 1962) and DS seeds with as high as 47% of oil in *Telfairia occidentalis* (Asiegbu, 1987). The up-regulation of genes involved in lipid biogenesis in *Arabidopsis thaliana* seed maturation is rather related to the characteristics of the species and the family that is known for accumulating high oil content in its seeds (Li et al., 2006). Thus, this dataset with robust

physiological and transcriptional description can also be used to study other aspects of *A. thaliana* seed maturation itself beside seed desiccation sensitivity that was the focus of this work.

In the PCA analysis, PC2 displayed the enrichment of GO terms related to biotic stress responses, such as the categories *defense response to fungus*, *response to external biotic stimulus* and *defense response to bacterium*. Upon the induction of desiccation tolerance in *Citrus limon* seeds, we observed a transcriptomic switch encompassing the suppression of biotic- and induction of abiotic responses (Chapter 5). We hypothesized that the down-regulation of biotic responses, in *C. limon* and Col-0 seeds, prevents the lignification of the seed cell walls and thus make these flexible, allowing their folding and preventing damage upon seed desiccation.

Lignin deposition is a strategic response to combat fungal infection (El Hadrami et al., 2010). However, lignin hardens the cell walls while desiccation tolerance requires flexible and foldable cells (Moore et al., 2008). It has been demonstrated that DS seeds show cell wall damage when desiccated (Chandel et al., 1995; Moore et al., 2008). Further studies are necessary to confirm the cell wall status of *abi3-6*, Col-0 and *C. limon* seeds.

Lack of seed dormancy and, consequently, no developmental arrest

The transcriptomes of late developmental stages in *abi3-6* are the most similar to Col-0 24h imbibed seeds. Based on this analysis, we infer that the *abi3-6* seeds at later maturation stages switch to a genetic program that leads to germination rather than quiescence, as in Col-0 seeds. This explains its enlarged apical meristem (Rohde et al., 2000) and the premature germination inside the siliques of *abi3-5* seeds when exposed to high humidity conditions (Raz et al., 2001).

Dormancy is a mechanism of developmental arrest that prevents fully developed seeds to germinate. The lack of developmental arrest of *abi3-6* seeds can be associated with the lack of expression of dormancy-associated genes. Here we observed the enrichment of dormancy associated GO categories only in Col-0. Furthermore, *DOG1* showed an expression that was second best correlated with desiccation tolerance levels in Col-0 and *abi3-6*. The interaction of *DOG1* with *ABI3* in the regulation of seed development has been demonstrated by Dekkers et al. (2016).

The lack of developmental arrest and suppression of metabolism is also demonstrated by the presence of chlorophyll in late stages of *abi3-6* seed development. The seeds of Arabidopsis *abi3-6* mutants retain chlorophyll after shedding (Figure 1) and lack both desiccation tolerance and dormancy. These phenotypes are established during the maturation phase of the seeds when *ABI3* expression peaks (Delmas et al., 2013a; Koornneef et al., 1989; Leprince et al., 1998; Monke et al., 2012). The majority

of green embryos in the eastern Australian rainforest species are desiccation sensitive (Hamilton et al., 2013).

Taken together, the developmental program of *abi3-6* seeds follows a different transcriptional path during the maturation as compared to Col-0 seeds. This path leads Col-0 seeds towards developmental arrest, metabolic quiescence and desiccation. In contrast, the *abi3-6* transcriptional path leads its seeds towards germination, keeping an active metabolism and, consequently, desiccation sensitive seeds.

Initially, we looked for genes involved in the protection against desiccation damage that showed dissimilar expression patterns between Col-0 and *abi3-6* seed development. We indeed observed the decreased expression of, e.g. LEAs and oleosins in *abi3-6*. They showed dissimilar expression profiles, but at later developmental stages, and, thus, cannot be considered causal for desiccation sensitivity in *abi3-6* seeds since the phenotypical differences were observed already at early stages. Then, we decided to focus our analysis on the regulatory mechanisms that might govern the divergence of these transcriptional paths. We determined when the transcriptional paths between Col-0 and *abi3-6* diverge and what might be the regulators involved.

The co-expression networks of the two genotypes show similar shapes but with a thinner bridge connecting the two major clusters in abi3-net. This bridge may be considered as a cluster of genes responsible for the coordination of the transition from desiccation sensitive to tolerant seeds. The bridge connects two major clusters that mostly contain early/DS and late/DT developmental stages. There is a large overlap in Col-net between the genes in the bridge and in the gene set DP-C11 that contains genes that peak when desiccation tolerance is acquired in Col-0. This pattern was not observed in abi36-net.

We performed a detailed physiological assessment of desiccation tolerance acquisition during Col-0 and *abi3-6* seed development. The robustness of this assessment associated with high quality transcriptomic data allowed us to investigate relevant mechanisms involved in the acquisition of seed desiccation tolerance. The absence of developmental arrest in *abi3-6* seeds is associated with a transcriptional lack of dormancy. Among the dormancy-associated transcripts, *DOG1* was identified as a top correlated one. We also observed the down-regulation of transcripts involved with protection against desiccation damage in *abi3-6* seeds, such as LEAs and oleosins. Finally, co-expression network analysis demonstrated striking disparities between Col-net and abi3-net. Among these the bridge that connects DS and DT clusters in Col-net deserves further investigation.

Many hormonal signalling genes were found in the Col-net bridge, including *FUS3* which is an essential TF for acquisition of seed desiccation tolerance, but which was not found in the abi36-net bridge. We also inferred that FUS3 is a significant potential regulator of several genes in both Col-net and abi36-net bridges. The

significant divergent expression of *FUS3* and its target genes between Col-0 and *abi3-6* during seed maturation strongly suggests that this is an important early regulator underlying desiccation sensitivity of *abi3-6* seeds.

Materials and methods

Plant growth conditions

Arabidopsis thaliana plants, accession Columbia (Col-0, N60000), were grown in a climate room on Rockwool plugs (MM40/40; Grodan), at 22ºC, under 16h day/8h night conditions, watered with Hyponex nutrient solution (1 g.L^{-1}, http://www.hyponex.co.jp) 3 times a week.

Desiccation survival assessment

Col-0 and *abi3-6* (Nambara et al., 1994) mutant plants were grown under the above mentioned conditions. Staging of the developing seeds was done by tagging individual flowers on the day of anthesis. Samples of the various developmental intervals were collected from the same plants at different days in a random order. Immature seeds were collected by opening unripe siliques using a needle and forceps.

Drying was executed at 22ºC, in a cabinet with air flow at 30% relative humidity (RH) in the dark. We applied three seed treatments (1) Seeds were excised from the siliques and placed immediately to germinate; (2) The siliques were first placed in a drying cabinet at 32% RH for 7 days followed by excision of the seeds from the siliques and subsequent drying for three more days; (3) The seeds were excised from the siliques and placed immediately in a drying cabinet at 32% RH for 3 days.

For germination assessment, seeds were placed in Petri dishes (60x15mm) on filter paper, moistened with 1.5 mL of a 50 µM GA$_{4+7}$ (Berelex; Valent Biosciences, Libertyville, USA) solution, at 22ºC, in a germination cabinet (FLOHR Instruments MKKL500, Netherlands) with constant light. After seven days, the percentage of germination (survival) was scored. Each treatment consisted of six biological replicates. Each replicate was represented by one silique from different plants, with on average 28 seeds per silique.

Gravimetric determination of water content

Col-0 and *abi3-6* siliques were collected in micro tubes containing water soaked filter paper. Approximately 100 seeds where excised from siliques, weighed within 10 minutes after excision and after 24h of drying at 105ºC. Each time point contained a single biological replicate represented by a pool of 4 siliques (approx. 200 seeds) from 4 different plants. The water content of the seed samples was calculated on a fresh weight basis. The weight of the fresh samples were subtracted by its dry weight after placing 24h in oven at 100ºC and divided by the initial fresh weight. Water content = (Fresh - Dry)/Fresh.

RNA extraction

Total RNA was extracted according to the hot borate protocol modified from Wan and Wilkins (1994). Three replicates of a pool of seeds from 4 siliques (~200 seeds) for each treatment were homogenized and mixed with 800 µL of extraction buffer (0.2 M Na borate decahydrate (Borax), 30 mM EGTA, 1% SDS, 1% Na deoxy-cholate (Na-DOC)) containing 1.6 mg DTT and 48 mg PVP40 which had been heated to 80ºC. One mg proteinase K was added to this suspension and incubated for 15 min at 42ºC. After adding 64 µl of 2 M KCl the samples were incubated on ice for 30 min and subsequently centrifuged for 20 min at 12,000 g. 270 µL of ice-cold 8 M LiCl was added to the supernatant in a final concentration of 2 M and the tubes were incubated overnight on ice. After centrifugation for 20 min at 12,000 g at 4ºC, the pellets were washed with 750 µL ice-cold 2 M LiCl. The samples were centrifuged for 10 min at 10,000 g at 4ºC and the pellets were re-suspended in 100 µL DEPC treated water. The samples were cleaned with phenol chloroform and treated with DNAse (RQ1 DNase, Promega). The RNA quality and concentration were assessed by agarose gel electrophoresis and UV spectrophotometry.

RNA was processed for use in RNA-seq with mRNA enrichment polyA capture (Illumina Incorporated, San Diego, CA, USA). The sequencing was performed strand specific on Illumina Hiseq 2500, using cDNA and random hexamer priming, and generating paired-end 125 nt reads.

RNA-seq assembly and transcript quantification

Reads were mapped to the TAIR10 Arabidopsis reference genome with the Araport11 annotation (Cheng et al., 2017) using the Hisat2 software (Kim et al., 2015).

Transcript expression was quantified using the StringTie program (Pertea et al., 2015) and further analysed with the R package Ballgown (Frazee et al., 2015). Significant differentially regulated genes were assigned when they show a p-value < 0.05 and a fold change >2 for upregulated or <-2 for down-regulated genes. For the PCA analysis the R prcomp function was used.

RNA-seq assembly is available at National Centre for Biotechnology Information (NCBI) as Bioproject ID SRP112612. Genome assembly, annotation data and all the supplementary data are available for viewing and downloading at http://www.wageningenseedlab.nl/thesis/amarques/SI/chapter3.

Transcriptome correlations and annotation

We performed a Spearman correlation analysis between the average of each gene expression across all samples together (Col-0 and *abi3-6*) and its correspondent survival upon desiccation. We used $\log_2$ transformed gene expression and we exclude the genes with expression values of 0. The GO annotation of differentially regulated genes sets were executed using the DAVID database (Dennis et al., 2003).

Dominant expression patterns

Dominant expression patterns (DPs) were identified as previously described by Sturn et al. (2002). DPs were identified from the total transcripts, corresponding to 29,321 mRNAs. The K choice that yielded the sufficient number of transcriptional modules each peaking at every developmental stage for both genotypes was ten.

Bulk gene expression plots

The graphs in Figure 6 display the bulk expression of gene sets. Firstly, every gene had its expression normalized by the average of the expression of that gene among all samples. Secondly, we took the average of targeted lists of genes expression for each sample. Finally, the graphs were plotted using $\log_2$ scale and standard error bars.

Network analysis

The transcript quantifications as determined before with StringTie were converted to counts per gene using the prepDE.py python script that is available with StringTie. The gene counts were normalized using the edgeR package (Robinson et al., 2010), to log

CPM (counts per million) and genes showing a mean absolute deviation (MAD) of less than 1 were removed. Separately for the Col-0 and the *abi3-6* samples, for each gene pair the Pearson correlation coefficient was calculated. Gene pairs with a coefficient higher than 0.95 were used for an edge list to build the Col-net and abi36-net networks with Cytoscape (Smoot et al., 2010).

References

Andrade, A., Cunha, R., Souza, A., Reis, R., Almeida, K. (2003) Physiological and morphological aspects of seed viability of a neotropical savannah tree, Eugenia dysenterica DC. Seed Science and Technology, **31**(1), 125-137.

Angeles-Núñez, J.G. and Tiessen, A. (2011) Mutation of the transcription factor LEAFY COTYLEDON 2 alters the chemical composition of Arabidopsis seeds, decreasing oil and protein content, while maintaining high levels of starch and sucrose in mature seeds. Journal of Plant Physiology, **168**(16), 1891-1900.

Asiegbu, J.E. (1987) Some biochemical evaluation of fluted pumpkin seed. Journal of the Science of Food and Agriculture, **40**(2), 151-155.

Bensmihen, S., Giraudat, J., Parcy, F. (2005) Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during Arabidopsis thaliana embryo maturation. J Exp Bot, **56**(412), 597-603.

Berjak, P. and Pammenter, N.W. (2008) From Avicennia to Zizania: seed recalcitrance in perspective. Ann Bot, **101**(2), 213-228.

Berjak, P. and Pammenter, N.W. (2013) Implications of the lack of desiccation tolerance in recalcitrant seeds. Front Plant Sci, **4**, 478.

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

Born, N., Behringer, D., Liepelt, S., Beyer, S., Schwerdtfeger, M., Ziegenhagen, B., Koch, M. (2014) Monitoring plant drought stress response using terahertz time-domain spectroscopy. Plant physiology, **164**(4), 1571-1577.

Boubriak, I., Dini, M., Berjak, P., Osborne, D.J. (2000) Desiccation and survival in the recalcitrant seeds of Avicennia marina: DNA replication, DNA repair and protein synthesis. Seed Science Research, **10**(03), 307-315.

Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., Harada, J.J. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proceedings of the National Academy of Sciences of the United States of America, **103**(9), 3468-3473.

Cadman, C.S., Toorop, P.E., Hilhorst, H.W., Finch-Savage, W.E. (2006) Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. Plant J, **46**(5), 805-822.

Chandel, K., Chaudhury, R., Radhamani, J., Malik, S. (1995) Desiccation and freezing sensitivity in recalcitrant seeds of tea, cocoa and jackfruit. Annals of Botany, **76**(5), 443-450.

Chen, Q., Yang, L., Ahmad, P., Wan, X., Hu, X. (2011) Proteomic profiling and redox status alteration of recalcitrant tea (Camellia sinensis) seed in response to desiccation. Planta, **233**(3), 583-592.

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., Town, C.D. (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant Journal, **89**(4), 789-804.

Crowe, J.H. (2014) Anhydrobiosis: an unsolved problem. Plant, Cell & Environment, **37**(7), 1491-1493.

Daphne, j.o. (2000) Hazards of a germinating seed: available water and the maintenance of genomic integrity. Israel journal of plant sciences, **48**(3), 173-179.

Daws, M.I., Garwood, N.C., Pritchard, H.W. (2006) Prediction of desiccation sensitivity in seeds of woody species: a probabilistic model based on two seed traits and 104 species. Ann Bot, **97**(4), 667-674.

de Castro, R.D., Zheng, X., Bergervoet, J.H., De Vos, C.R., Bino, R.J. (1995) [beta]-Tubulin Accumulation and DNA Replication in Imbibing Tomato Seeds. Plant physiology, **109**(2), 499-504.

de Souza Vidigal, D., Willems, L., van Arkel, J., Dekkers, B.J., Hilhorst, H.W., Bentsink, L. (2016) Galactinol as marker for seed longevity. Plant Science, **246**, 112-118.

Dekkers, B.J., Costa, M.C.D., Maia, J., Bentsink, L., Ligterink, W., Hilhorst, H.W. (2015) Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. Planta, **241**(3), 563-577.

Dekkers, B.J., He, H., Hanson, J., Willems, L.A., Jamar, D.C., Cueff, G., Rajjou, L., Hilhorst, H.W., Bentsink, L. (2016) The Arabidopsis DELAY OF GERMINATION 1 gene affects ABSCISIC ACID INSENSITIVE 5 (ABI5) expression and genetically interacts with ABI3 during Arabidopsis seed development. The Plant Journal, **85**(4), 451-465.

Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H., Buitink, J. (2013) LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. J Exp Bot, **64**(14), 4559-4573.

Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J.G., McCourt, P., Samuel, M.A. (2013a) ABI3 controls embryo degreening through Mendel's I locus. Proc Natl Acad Sci U S A, **110**(40), E3888-3894.

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

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A. (2003) DAVID: database for annotation, visualization, and integrated discovery. Genome biology, **4**(9), R60.

Earle, F. and Jones, Q. (1962) Analyses of seed samples from 113 plant families. Economic Botany, **16**(4), 221-250.

Eden, E., Lipson, D., Yogev, S., Yakhini, Z. (2007) Discovering motifs in ranked lists of DNA sequences. PLoS computational biology, **3**(3), e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z. (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC bioinformatics, **10**(1), 48.

El Hadrami, A., Adam, L.R., El Hadrami, I., Daayf, F. (2010) Chitosan in plant protection. Marine drugs, **8**(4), 968-987.

Faria, J.M., Buitink, J., van Lammeren, A.A., Hilhorst, H.W. (2005) Changes in DNA and microtubules during loss and re-establishment of desiccation tolerance in germinating Medicago truncatula seeds. J Exp Bot, **56**(418), 2119-2130.

Farnsworth, E. (2000) The ecology and physiology of viviparous and recalcitrant seeds. Annual Review of Ecology and Systematics, 107-138.

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

Focks, N. and Benning, C. (1998) wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiol, **118**(1), 91-101.

Footitt, S., Douterelo-Soler, I., Clay, H., Finch-Savage, W.E. (2011) Dormancy cycling in Arabidopsis seeds is controlled by seasonally distinct hormone-signaling pathways. Proceedings of the National Academy of Sciences, **108**(50), 20236-20241.

Frazee, A.C., Pertea, G., Jaffe, A.E., Langmead, B., Salzberg, S.L., Leek, J.T. (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. Nature biotechnology, **33**(3), 243-246.

Hamilton, K.N., Offord, C.A., Cuneo, P., Deseo, M.A. (2013) A comparative study of seed morphology in relation to desiccation tolerance and other physiological responses in 71 Eastern Australian rainforest species. Plant Species Biology, **28**(1), 51-62.

Hedden, P. and Phillips, A.L. (2000) Gibberellin metabolism: new insights revealed by the genes. Trends in plant science, **5**(12), 523-530.

Hoekstra, F.A., Golovina, E.A., Buitink, J. (2001) Mechanisms of plant desiccation tolerance. Trends in plant science, **6**(9), 431-438.

Holdsworth, M.J., Bentsink, L., Soppe, W.J. (2008) Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. New Phytologist, **179**(1), 33-54.

Hundertmark, M. and Hincha, D.K. (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC genomics, **9**, 118.

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., Gao, G. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic acids research, **45**(D1), D1040-D1045.

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T. (2005) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant and Cell Physiology, **46**(3), 399-406.

Keith, K., Kraml, M., Dengler, N.G., McCourt, P. (1994) fusca3: A Heterochronic Mutation Affecting Late Embryo Development in Arabidopsis. Plant Cell, **6**(5), 589-600.

Kim, D., Langmead, B., Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods, **12**(4), 357-360.

Koornneef, M., Hanhart, C.J., Hilhorst, H.W.M., Karssen, C.M. (1989) In Vivo Inhibition of Seed Development and Reserve Protein Accumulation in Recombinants of Abscisic Acid Biosynthesis and Responsiveness Mutants in Arabidopsis thaliana. Plant Physiol, **90**(2), 463-469.

Leprince, O., van Aelst, A.C., Pritchard, H.W., Murphy, D.J. (1998) Oleosins prevent oil-body coalescence during seed imbibition as suggested by a low-temperature scanning electron microscope study of desiccation-tolerant and -sensitive oilseeds. Planta, **204**(1), 109-119.

Levitt, J. (1980) *Responses of Plants to Environmental Stress, Volume 1: Chilling, Freezing, and High Temperature Stresses* Academic Press.

Li, D.Z. and Pritchard, H.W. (2009) The science and economics of ex situ plant conservation. Trends Plant Sci, **14**(11), 614-621.

Li, Y., Beisson, F., Pollard, M., Ohlrogge, J. (2006) Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry, **67**(9), 904-915.

Li, Y., Jones, L., McQueen-Mason, S. (2003) Expansins and cell growth. Current opinion in plant biology, **6**(6), 603-610.

Maia, J., Dekkers, B.J., Dolle, M.J., Ligterink, W., Hilhorst, H.W. (2014) Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated Arabidopsis seeds. New Phytol, **203**(1), 81-93.

Maia, J., Dekkers, B.J., Provart, N.J., Ligterink, W., Hilhorst, H.W. (2011) The re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds and its associated transcriptome. PloS one, **6**(12), e29123.

Mauch-Mani, B. and Mauch, F. (2005) The role of abscisic acid in plant–pathogen interactions. Current opinion in plant biology, **8**(4), 409-414.

Monke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hahnel, U., Junker, A., Weisshaar, B., Conrad, U., Baumlein, H., Altschmied, L. (2012) Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. Nucleic Acids Res, **40**(17), 8240-8254.

Moore, J.P., Vicre-Gibouin, M., Farrant, J.M., Driouich, A. (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. Physiol Plant, **134**(2), 237-245.

Murray, D.R. (1988) *Nutrition of the angiosperm embryo* Research Studies Press Ltd.

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.

Nambara, E., Keith, K., McCourt, P., Naito, S. (1994) Isolation of an internal deletion mutant of the Arabidopsis thaliana ABI3 gene. Plant and Cell Physiology, **35**(3), 509-513.

Nguyen, T.P., Keizer, P., van Eeuwijk, F., Smeekens, S., Bentsink, L. (2012) Natural variation for seed longevity and seed dormancy are negatively correlated in Arabidopsis. Plant Physiol, **160**(4), 2083-2092.

Okamoto, M., Tatematsu, K., Matsui, A., Morosawa, T., Ishida, J., Tanaka, M., Endo, T.A., Mochizuki, Y., Toyoda, T., Kamiya, Y., Shinozaki, K., Nambara, E., Seki, M. (2010) Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of Arabidopsis using tiling arrays. Plant J, **62**(1), 39-51.

Ooms, J.J., Leon-Kloosterziel, K.M., Bartels, D., Koornneef, M., Karssen, C.M. (1993) Acquisition of desiccation tolerance and longevity in seeds of Arabidopsis thaliana (a comparative study using abscisic acid-insensitive abi3 mutants). Plant physiology, **102**(4), 1185-1191.

Ooms, J.J., Veen, R., Karssen, C.M. (1994) Abscisic acid and osmotic stress or slow drying independently induce desiccation tolerance in mutant seeds of Arabidopsis thaliana. Physiologia Plantarum, **92**(3), 506-510.

Osborne, D.J., Boubriak, I., Leprince, O. (2002) 1 2 Rehydration of Dried Systems: Membranes and the Nuclear Genome. Desiccation and survival in plants: Drying without dying, 343.

Pammenter, N. and Berjak, P. (1999) A review of recalcitrant seed physiology in relation to desiccation-tolerance mechanisms. Seed Science Research, **9**(01), 13-37.

Pammenter, N. and Berjak, P. (2000) Some thoughts on the evolution and ecology of recalcitrant seeds. Plant Species Biology, **15**(2), 153-156.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature biotechnology, **33**(3), 290-295.

Raz, V., Bergervoet, J., Koornneef, M. (2001) Sequential steps for developmental arrest in Arabidopsis seeds. Development, **128**(2), 243-252.

Roberts, E. and Ellis, R. (1989) Water and seed survival. Annals of Botany, **63**(1), 39-39.

Roberts, E.H. (1973) Predicting the storage life of seeds. In Proceedings.

Robinson, M.D., McCarthy, D.J., Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, **26**(1), 139-140.

Rohde, A., De Rycke, R., Beeckman, T., Engler, G., Van Montagu, M., Boerjan, W. (2000) ABI3 affects plastid differentiation in dark-grown Arabidopsis seedlings. The Plant Cell, **12**(1), 35-52.

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

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., Lohmann, J.U. (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet, **37**(5), 501-506.

Śliwińska, E. (2003) Cell cycle and germination of fresh, dried and deteriorated sugarbeet seeds as indicators of optimal harvest time. Seed Science Research, **13**(02), 131-138.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., Ideker, T. (2010) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, **27**(3), 431-432.

Sturn, A., Quackenbush, J., Trajanoski, Z. (2002) Genesis: cluster analysis of microarray data. Bioinformatics, **18**(1), 207-208.

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., Parcy, F. (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. Plant Cell, **18**(7), 1642-1651.

van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., Soppe, W.J. (2011) Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proceedings of the National Academy of Sciences, **108**(50), 20219-20224.

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., Leprince, O., Chatelain, E., Vu, B.L., Gouzy, J., Gamas, P., Udvardi, M.K., Buitink, J. (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. Plant Physiol, **163**(2), 757-774.

Wan, C.-Y. and Wilkins, T.A. (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (Gossypium hirsutum L.). Analytical biochemistry, **223**(1), 7-12.

Wang, F. and Perry, S.E. (2013) Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol, **161**(3), 1251-1264.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu. Rev. Plant Biol., **57**, 781-803.

Zeng, Y. and Kermode, A. (2004) A gymnosperm ABI3 gene functions in a severe abscisic acid-insensitive mutant of Arabidopsis (abi3-6) to restore the wild-type phenotype and demonstrates a strong synergistic effect with sugar in the inhibition of post-germinative growth. Plant Mol Biol, **56**(5), 731-746.

A)

B)
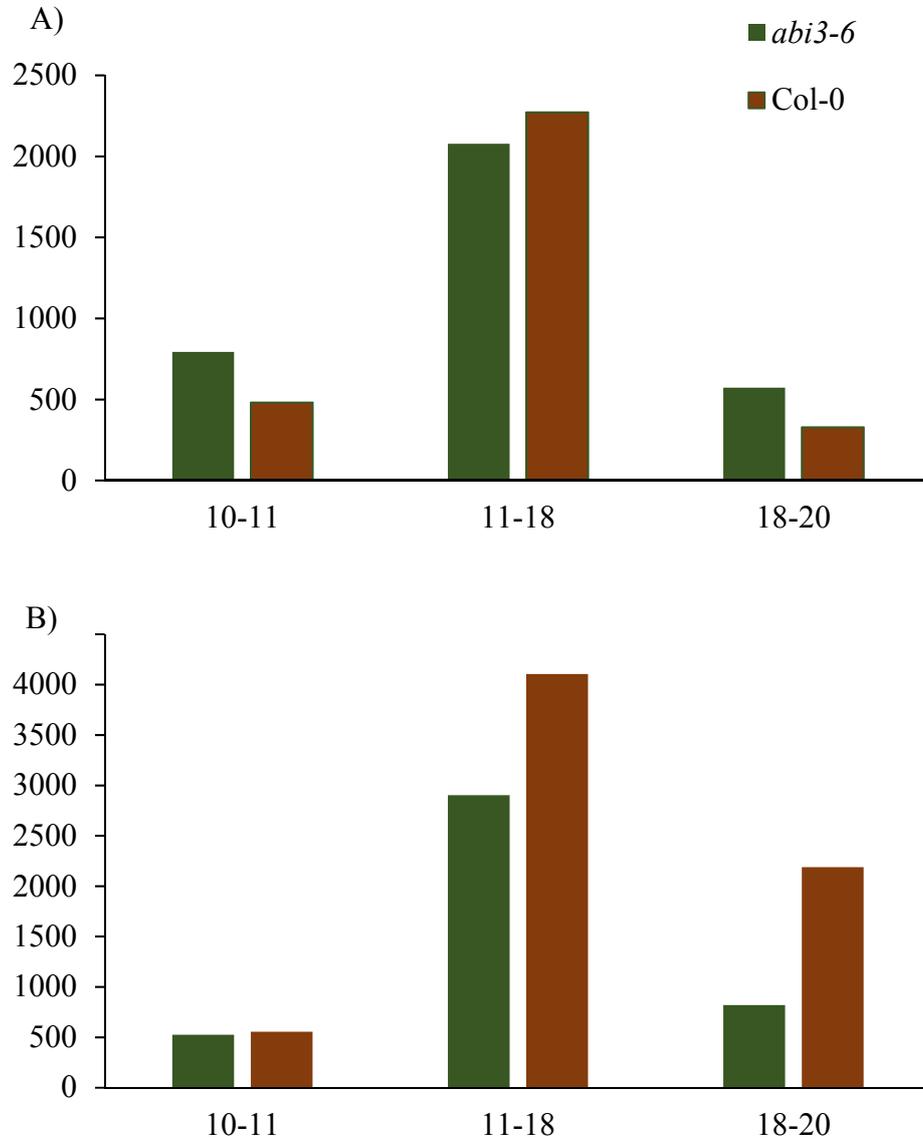
Supplementary Figure 1. The figure depicts on the y-axis the total number of significantly (*p*-value < 0.05 and fold change >2 or <-2) up- (A) and down-regulated genes (B) between different time points. The y-axis represents the values of the comparisons between time points that were always performed by the ratio of the late developmental stage against the previous stage. The letters describe the genotypes (a for *abi3-6* seeds; c for Col-0) and the numbers the analysed time points (10, 11, 18 and 20 days after flowering).
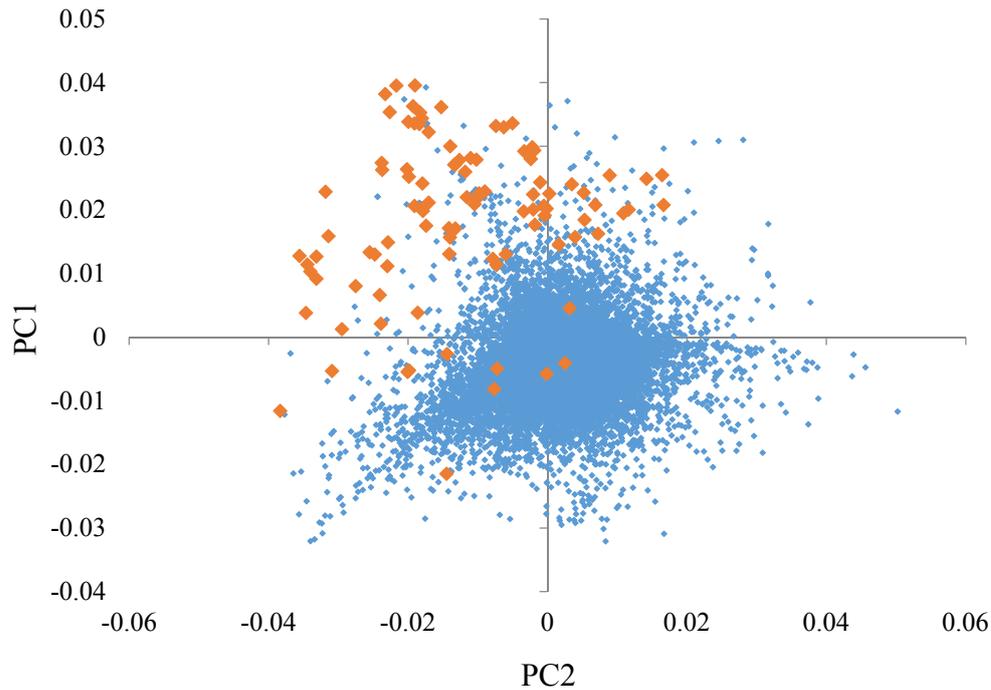
Supplementary Figure 2. Gene loadings from the principal components PC1 and PC2. Genes with high PC1 are those that contributes mostly to the distribution of Col-0 late maturation samples and genes with high PC2 values are the most discriminative for *abi3-6* in the PCA (Figure 4). *ABI3* regulated genes, demonstrated by Monke et al. (2012), are marked in red. The variance explained by the principal components PC1 and PC2 are respectively 60% and 26%.

plant-type cell wall modification involved
  in multidimensional cell growth
response to karrikin
plant-type cell wall loosening
proteolysis involved in cellular protein
  catabolic process
syncytium formation
response to red light
proteolysis
plant-type cell wall organization
unidimensional cell growth
xyloglucan metabolic process
indole glucosinolate biosynthetic process
transcription factor import into nucleus
response to light stimulus

c11-c18
after-ripened
dormant
a11-a18

1343
656
8
2
705
1175
17
0
233
96
0
122
0
77
8

response to high light intensity
response to heat
response to hydrogen peroxide
response to abscisic acid
response to salt stress
positive regulation of translation
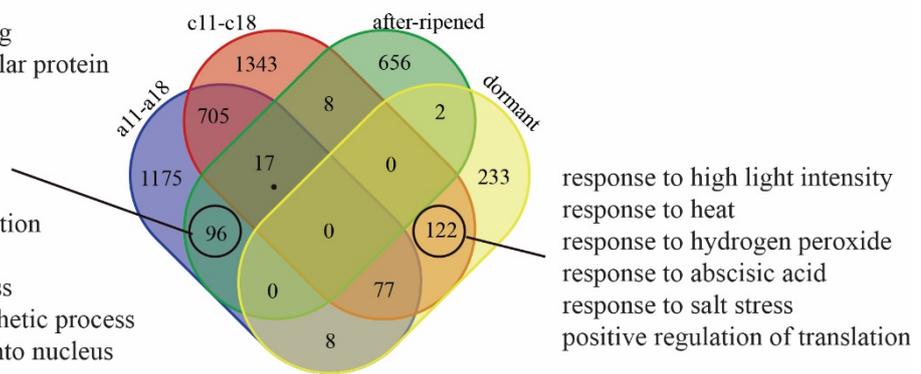
Supplementary Figure 3. The venn diagram depicts the overlap between up-regulated genes during Col-0 and *abi3-6* maturation (18 compared to 11 daf) with genes induced during imbibition of dormant or after-ripened seeds (Cadman et al. 2006). The samples used in this graph are coded with: a for *abi3-6* seeds; c for Col-0 and followed by numbers corresponding to the investigated time points (11 and 18 days after flowering (daf)). GO-enriched categories in two relevant overlapping sets are shown. The 96 neighbouring overlapping set of 17 genes for the common set of 96 identified genes has the enriched GO category *oxidation-reduction process*; The overlapping set with 77 genes, which neighbours the set of 122 genes, has the enriched GO categories *flavonoid glucuronidation*, *flavonoid biosynthetic process*, *response to water deprivation*, *metabolic process*, *response to chitin*, *oxidation-reduction process*, *protein homooligomerization*, *response to salicylic acid* and *response to jasmonic acid*.
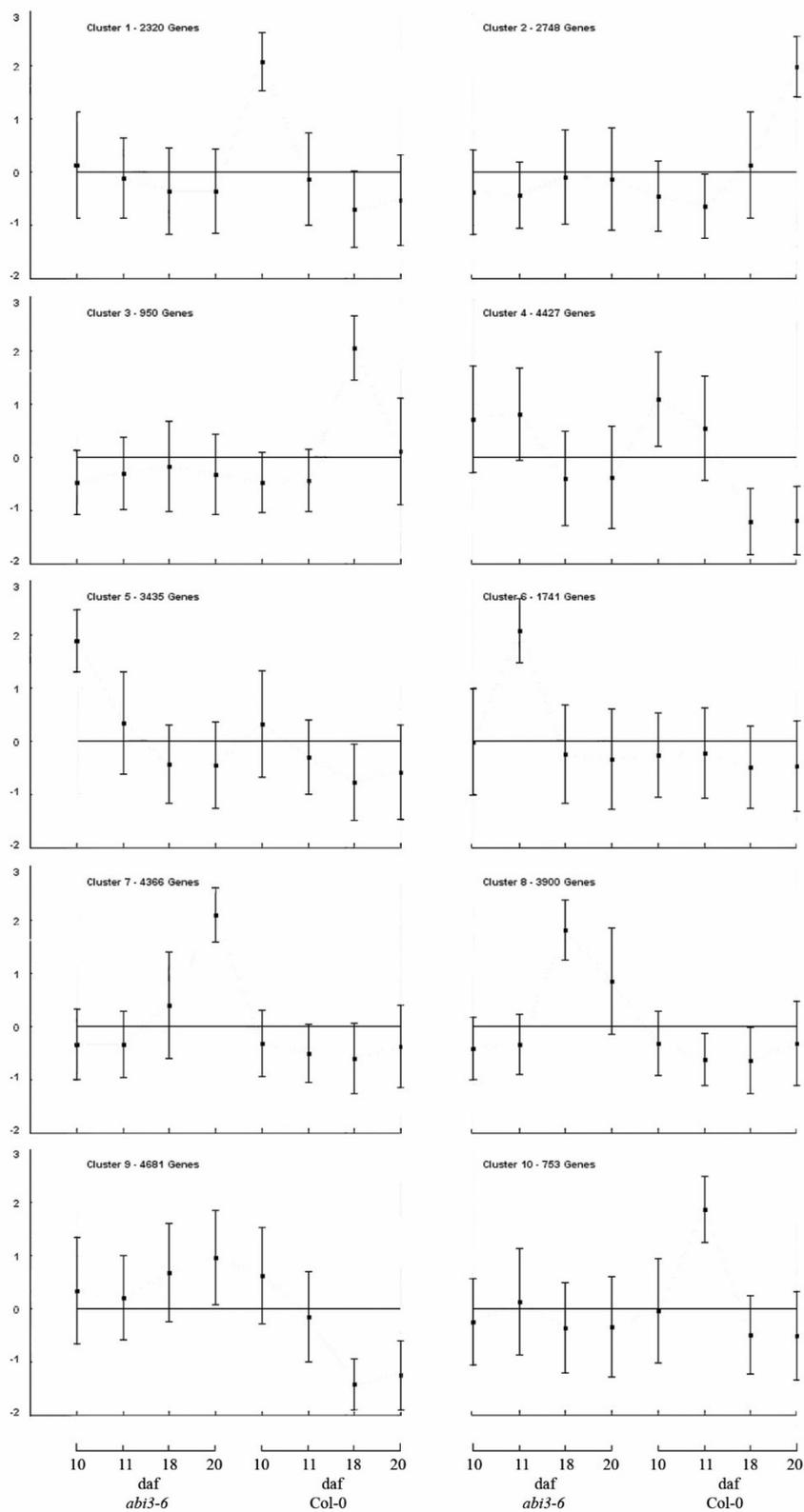
Supplementary Figure 4. Dominant patterns (DP) of gene expression during the seed maturation of Col-0 and *abi3-6* mutant. Ten DPs were found using Fuzzy K-means clustering of the 50% most variant transcripts from the data set of 29,321 transcripts that showed significant expression difference in at least one developmental stage. Y-axis depicts averages of mRNA expression levels in each sample. On the x-axis from left to right the samples: *abi3-6* (10, 11, 18, 20 days after flowering (daf)) and Col-0 (10, 11, 18, 20 daf).

86

Supplementary Figure 5. Heat map showing the FDR significance of enriched GO terms for dominant expression patterns (DPs) (Supplemental figure 4, supplementary file 5). GO terms listed are for biological process or/and cellular components that were overrepresented for each DP gene set. There was no enriched GO category for the sample *abi3-6* on 10 days after flowering (daf) (a10).

Supplementary Table 1. Spearman correlation between the transcriptomes generated in this work and most correlated publicly available* transcriptomes (top 5 shown for each sample).

abi3-6 10 daf		*abi3-6* 11 daf		*abi3-6* 18 daf		*abi3-6* 20 daf	
Col-0 5daf	0.71	Col-0 6daf	0.75	*abi3-6* 16daf	0.79	*abi3-6* 16daf	0.75
Col-0 4daf	0.71	Col-0 5daf	0.72	Col-0 24h water	0.69	Col-0 24h water	0.71
Col-0 6daf	0.69	*abi3-6* 12daf	0.70	Col-0 24h 3 µM ABA	0.68	Col-0 24h 3µM ABA	0.70
abi3-6 12daf	0.62	Col-0 7daf	0.68	Col-0 24h 30µM ABA	0.67	Col-0 24h 30µM ABA	0.69
Col-0 7daf	0.61	Col-0 4daf	0.68	*abi3-6* 12daf	0.66	*abi3-6* 12daf	0.58

Col-0 10 daf		Col-0 11 daf		Col-0 18 daf		Col-0 20 daf	
Col-0 6daf	0.72	Col-0 6daf	0.75	Col-0 16daf	0.77	Col-0 dry	0.79
Col-0 5daf	0.71	Col-0 7daf	0.71	Col-0 10daf	0.77	Col-0 16daf	0.78
Col-0 4daf	0.70	Col-0 5daf	0.69	Col-0 9daf	0.76	Col-0 10daf	0.72
abi3-6 12daf	0.63	*abi3-6* 12daf	0.69	Col-0 8daf	0.75	Col-0 9daf	0.69
Col-0 7daf	0.63	Col-0 12daf	0.66	Col-0 12daf	0.69	Col-0 8daf	0.65

*publicly available arrays used in this analysis: 3 to 10 daf Col-0 seeds (Schmid et al. 2005); 12 and 16 daf Col-0 and *abi3-6* seeds (PRJNA261830); Col-0- dry seeds, seeds imbibed in water for 24 and 48h and seeds imbibed in 3 or 30 µM ABA for 24 or 48h (Okamoto et al. 2010).

Supplementary Table 2. Transcription factors with significant enriched binding motifs among Col-net bridge genes

TF Id	Gene description	Target background	Target Input	p-value	q-value.BH
AT1G21970	LEAFY COTYLEDON 2 (LEC1)	12	2	7.92E-05	4.59E-02
AT1G28300	LEAFY COTYLEDON 2 (LEC2)	500	11	3.19E-04	5.46E-02
AT5G63790	NAC DOMAIN 102 (ANAC102)	309	8	4.58E-04	5.46E-02
AT4G29000	Tesmin/TSO1-like CXC	377	9	4.66E-04	5.46E-02
AT2G26960	MYB DOMAIN PROTEIN 81 (AtMYB81)	276	7	9.73E-04	9.50E-02
AT3G47600	MYB DOMAIN PROTEIN 94 (MYB94)	843	14	1.35E-03	1.13E-01
AT4G16430	JASMONATE ASSOCIATED MYC2 LIKE 3 (JAM3)	75	3	2.20E-03	1.48E-01
AT3G02940	MYB DOMAIN PROTEIN 107 (MYB107)	562	10	2.85E-03	1.67E-01
AT3G08500	MYB DOMAIN PROTEIN 83 (MYB83)	1379	18	5.80E-03	2.64E-01
AT1G08810	MYB DOMAIN PROTEIN 60 (MYB60)	892	13	5.87E-03	2.64E-01
AT1G32640	(MYC2)	102	3	6.60E-03	2.76E-01
AT3G26790	FUSCA3 (FUS3)	644	10	7.74E-03	3.03E-01
AT3G49690	REGULATOR OF AXILLARY MERISTEMS3 (RAX3)	654	10	8.63E-03	3.16E-01

Obs.: 209 Gene input; 28775 Genes background.

Supplementary Table 3. Transcription factors with significant enriched binding motifs among *abi36*-net bridge genes

TF id	Gene description	Target background	Target Input	p_value	q_value
AT5G16600	MYB DOMAIN PROTEIN 43 (MYB43)	650	9	4.01E-05	1.46E-02
AT4G17785	MYB DOMAIN PROTEIN 39 (MYB39)	542	8	5.57E-05	1.46E-02
AT5G08130	(BIM1)	917	10	1.50E-04	2.63E-02
AT1G06180	MYB DOMAIN PROTEIN 13 (MYB13)	957	10	2.17E-04	2.85E-02
AT1G16490	MYB DOMAIN PROTEIN 58 (MYB58)	993	10	2.98E-04	3.13E-02
AT5G56110	(MYB80)	457	6	6.26E-04	5.47E-02
AT3G26790	FUSCA3 (FUS3)	631	7	8.81E-04	6.60E-02
AT4G34000	ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 3 (ABF3)	663	7	1.21E-03	7.93E-02
AT1G09540	MYB DOMAIN PROTEIN 61 (MYB61)	874	8	1.79E-03	9.74E-02
AT3G12820	MYB DOMAIN PROTEIN 10 (MYB10)	883	8	1.92E-03	9.74E-02
AT3G61250	MYB DOMAIN PROTEIN 17 (MYB17)	735	7	2.30E-03	9.74E-02
AT2G42280	FLOWERING BHLH 4 (FBH4)	574	6	2.31E-03	9.74E-02

AT4G38910	BASIC PENTACYSTEINE 5 (BPC5)	7371	35	2.41E-03	9.74E-02
AT1G25340	MYB DOMAIN PROTEIN 116 (MYB116)	316	4	3.33E-03	1.25E-01
AT1G19350	BRI1-EMS-SUPPRESSOR 1 (BES1)	1372	10	3.99E-03	1.28E-01
AT5G14340	MYB DOMAIN PROTEIN 40 (MYB40)	636	6	4.07E-03	1.28E-01
AT3G08500	MYB DOMAIN PROTEIN 83 (MYB83)	1379	10	4.15E-03	1.28E-01
AT5G38860	BES1-INTERACTING MYC-LIKE PROTEIN 3 (BIM3)	1004	8	4.50E-03	1.31E-01
AT4G36930	SPATULA (SPT)	1240	9	5.79E-03	1.44E-01
AT1G79180	MYB DOMAIN PROTEIN 63 (MYB63)	1245	9	5.95E-03	1.44E-01
AT5G45300	BETA-AMYLASE 2 (BMY2)	1052	8	6.07E-03	1.44E-01
AT5G54230	MYB DOMAIN PROTEIN 49 (MYB49)	518	5	6.08E-03	1.44E-01
AT4G28110	MYB DOMAIN PROTEIN 41 (MYB41)	368	4	6.31E-03	1.44E-01
AT1G35460	FLOWERING BHLH 1 (FBH1)	531	5	6.83E-03	1.49E-01
AT1G51700	DOF ZINC FINGER PROTEIN 1 (DOF1)	4867	24	7.77E-03	1.63E-01
AT4G38620	MYB DOMAIN PROTEIN 4 (MYB4)	728	6	8.34E-03	1.68E-01
AT4G18890	BES1/BZR1 HOMOLOG 3 (BEH3)	935	7	9.56E-03	1.86E-01
AT2G18350	HOMEOBOX PROTEIN 24 (HB24)	411	4	9.90E-03	1.86E-01

Obs.: 91 Gene input; 28775 Genes background.

References cited in the supplementary data

Cadman, C. S., P. E. Toorop, H. W. Hilhorst & W. E. Finch-Savage (2006) Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. *Plant J,* 46**,** 805-22.

Monke, G., M. Seifert, J. Keilwagen, M. Mohr, I. Grosse, U. Hahnel, A. Junker, B. Weisshaar, U. Conrad, H. Baumlein & L. Altschmied (2012) Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic Acids Res,* 40**,** 8240-54.

Okamoto, M., K. Tatematsu, A. Matsui, T. Morosawa, J. Ishida, M. Tanaka, T. A. Endo, Y. Mochizuki, T. Toyoda, Y. Kamiya, K. Shinozaki, E. Nambara & M. Seki (2010) Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of Arabidopsis using tiling arrays. *Plant J,* 62**,** 39-51.

Schmid, M., T. S. Davison, S. R. Henz, U. J. Pape, M. Demar, M. Vingron, B. Scholkopf, D. Weigel & J. U. Lohmann (2005) A gene expression map of Arabidopsis thaliana development. *Nat Genet,* 37**,** 501-6.

4

The recalcitrant genome of *Castanospermum australe*

Alexandre Marques, Maria-Cecília D. Costa, Udisha Chathuri, Eef Jonkheer, Tao Zhao, Elio Schijlen, Martijn Derks, Harm Nijveen, Marina Marcet-Houben, Irene Julca, Julien Delahaie, M. Eric Schranz, Toni Gabaldon, Sandra Pelletier, Olivier Leprince, Wilco Ligterink, Julia Buitink, Henk W.M. Hilhorst, Jill M. Farrant

Abstract

Most angiosperm species produce seeds that acquire tolerance to desiccation during their maturation and are dispersed in the environment in a quiescent dehydrated state. These seeds remain viable in seed bank facilities for many years, ranging from decades to millennia. However, there is a spectrum of species which produce seeds that are desiccation sensitive. These seeds represent a challenge for *ex situ* conservation, as they lose viability soon during storage (order of days to months). It has been proposed that desiccation sensitive seeds either lack or do not express the various processes and mechanisms necessary for the acquisition of desiccation tolerance. To address this matter, we sequenced the genome of *Castanospermum australe*, a legume species that produces desiccation sensitive seeds. The comparison of our high-quality genome assembly to other legume species indicated that *C. australe* has a low rate of evolution likely due to its perennial life-cycle and long generation times. Moreover, the *C. australe* genome is syntenic with itself and several *C. australe* orthologs of *Glycine max* and *Medicago truncatula* genes derived from the gamma whole genome duplication event were retained. Analysis of the changes in gene expression during *C. australe* seed development, showed that these seeds remain metabolically active and are ready for germination towards the end of seed maturation, unlike those of *M. truncatula*, which are dispersed dry and quiescent. Our results indicate that the desiccation sensitive phenotype of *C. australe* seeds arose through changes in specific signalling pathways, thus not activating processes and mechanisms necessary for the acquisition of desiccation tolerance. These changes were retained due to the species' natural habitat being favourable to prompt germination after seed dispersal.

Introduction

The seeds of most gymnosperm and angiosperm species are shed in the desiccated state and can be stored dry under sub-zero temperatures for prolonged periods of time, and thus facilitate plant germplasm conservation. Such desiccation tolerant (DT) seeds are termed "orthodox". However, seeds of some species are desiccation-sensitive (DS) and cannot be successfully stored under typical conditions[1]. DS-seeded species are mostly found in the humid tropics and may represent up to 50% of the species present in tropical evergreen rain forests[2]. Seed desiccation sensitivity is considered a derived trait that evolved independently in non-related clades[3]. Such species occur in environments conducive to immediate seed germination and thus selective pressure for desiccation tolerance has been relaxed or is absent. It has been proposed that genes responsible for seed desiccation tolerance have been lost, repressed and/or mutated in DS seeded species[4].

Desiccation tolerance is acquired mid-way during the development of orthodox seeds. This acquisition comprises highly coordinated molecular events including the repression of photosynthesis and energy metabolism, and accumulation of protective components, such as late embryogenesis abundant (LEA) proteins, anti-oxidants, and soluble sugars[5]. These events are tightly regulated by hormones such as abscisic acid (ABA) and transcription factors (TFs) such as *ABSCISIC ACID INSENSITIVE 3* (*ABI3*), *FUSCA 3* (*FUS3*) and *LEAFY COTYLEDON 1* (*LEC1*)[6]. Conversely, in the development of DS seeds the acquisition of desiccation tolerance and accumulation of above mentioned protectants appear to be suppressed and rather they directly progress towards germination[7–10]. However, the genetic makeup underlying the DS seed phenotype is unknown.

The legume family (Fabaceae) contains many important agricultural species, all producing DT seeds. Considerable molecular and physiological work has been reported on what are now considered "model" DT-seeded species, such as *Medicago truncatula* and *Glycine max*. Furthermore, the genomes of 16 legume species have been sequenced, including species producing DT and DS seeds. Thus, there is a large amount of information available, allowing comparative analysis among them, and with species producing DS seeds.

Castanopermum australe A. Cunn & C. Fraser ex Hook, also known as the Moreton Bay Chestnut or Blackbean*,* is a leguminous tropical tree native to the east coast of Australia and west Pacific islands, that produces DS seeds. *C. australe* is the only known species in the genus that forms a separate and early branching clade within the Papilionoideae subfamily[11]. Its important phylogenetic position within the ADA

(Angylocalyceae, Dipterygeae, and Amburaneae) clade make *C. australe* ideal to study both trait evolution and the ancient polyploid history of papilionoid legumes[12].

Here, we provide the detailed genomic sequence information of *C. australe* combined with time-resolved gene expression analysis of seed development of this species, including the comparison with other species producing either DS or DT seeds. Such information is key to understanding mechanisms of desiccation tolerance and, ultimately, to design strategies to improve tolerance of extreme water loss in DS seeds for conservation purposes. We investigated genomic changes associated with seed desiccation sensitivity, including those pertaining to gene deletions, severe mutations and gene mis-expression, as well as their relationship with gene expression patterns during seed development and maturation.

Genome sequencing and assembly

We produced a high-quality assembly of a total length of 382 Mb, with an N50 of 832.6 Kb that covers 96.7% of the predicted genome size consisting of 1,210 contigs and 1027 scaffolds (Supplementary table 1). The GC content was 32.9%. Genome annotation identified 29,124 protein-coding genes of which 98.1% show high sequence similarity to proteins in TrEMBL and 84.4% in Swiss-Prot. An estimation of genome completeness indicated that 96.4% of the BUSCO (Benchmarking Universal Single-Copy Orthologs) genes were present. Transposable elements covered 15.5% of the total genome. Repeat elements comprised 119 Kb of SINE, 383 Kb of LINE, 13 Mb DNA transposon and 42 Mb of annotated LTR (long terminal repeat-retrotransposons) sequence (Supplementary table 2).

Whole genome duplications in papilionoid legumes

Whole-genome duplications (WGD) have played a significant role in speciation and diversification of organisms, these being more common in plants than animals. Such events enable duplicate copies of genes to potentially evolve alternative functions by sub- or neofunctionalization or, more commonly, the loss or pseudogenization of one gene copy (paralog), leaving the other paralog as a single copy gene[13]. However, evolution of novel traits and evolutionary success potentially become apparent or stabilized after millions of years[12]. Histograms of the synonymous rates and average rates of syntenic blocks for six legume species showed distinctive peaks tracing back to the shared papilionoid legume WGD, suggesting that the rate of evolution is very diverse among closely related family members (Supplementary figure 1). The peak in *C. australe* corresponds to a Ks of 0.25 which is less than half of the rate observed in *G.*

max (0.6) and a third of that in *M. truncatula* (0.85). The substitution rate in *C. australe* is so low that a second peak corresponding to the eudicot hexaploidy (gamma WGD event) is still visible. The gamma event was also detected in the histograms of block averages in *G. max* and *Phaseolus vulgaris*. The low synonymous mutation rate in *C. australe* is likely due to this species being the only perennial in this comparative study and the one with the longest generation times.

Phylome reconstruction

To gain further insight into the evolution of *C. australe* and legume diversification, a phylome was constructed which contained the evolutionary histories of all its protein coding genes and their homologues in 20 publicly available sequenced plant species (Figure 1). This set contains species producing DT and DS seeds as well as those producing seeds with 'intermediate' storage characteristics. Such seeds are typically tolerant of extreme water loss but with poor survival of conventional storage conditions[1]. The duplication frequency at the base of the Papilionoideae subfamily is lower (0.25) than that found at the base of the Fabaceae family (1.15, Supplementary figure 2), confirming that the WGD occurred before the divergence of the Fabaceae subfamilies, which is congruent with the WGD event previously predicted. The data further suggest that the WGD happened before *C. australe* diverged from the other species (Supplementary figure 2).
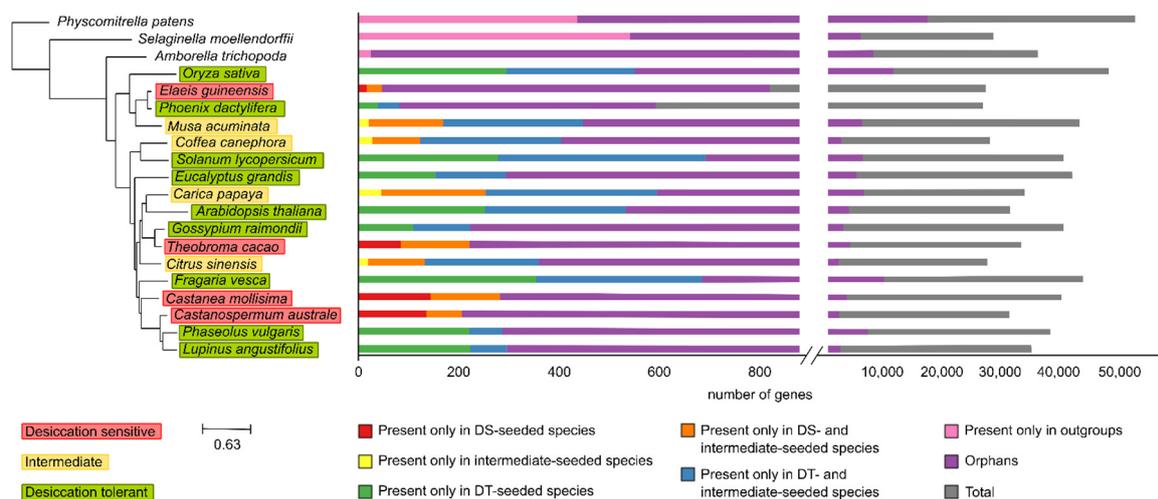

Figure 1. Phylogenetic tree. Reconstruction of the phylogenetic tree was based on a concatenated alignment of 183 single copy proteins that are present in at least 19 out of the 20 species surveyed. Species names are coloured according to their seed storage category.

The phylome analysis also indicated that very few protein-coding genes ($\leq 1\%$) were present in DS species only (Figure 1), confirming that independent evolutionary events gave rise to DS-seeded species. No protein-coding genes were retained in all DT species and lost in all DS species. However, several were lost in all DS and retained in at least half of the DT species (Extended data table 1). Among these were the transcription factors (TFs) *VERDANDI* and *MYB44-like*. *VERDANDI* participates in ovule identity complex and, when mutated, affects embryo sac differentiation in *Arabidopsis thaliana*[14,15]. Interestingly, *VERDANDI* and *MYB44-like* were also retained in intermediate-seeded species.

One gene (*PLAC8*) was lost without retention of paralogs in three out of four DS-seeded species, namely *C. australe, Castanea mollissima* and *Elaeis guineensis*. The knock-out of this gene caused increased seed and fruit size in maize[16]. In addition, the fw2.2 locus containing the *PLAC8* gene has been suggested to be the key to the evolution of tomato fruit size[17]. Large seeds and fruits are common features of DS species and presumably reduce the rate of seed drying and hence the risk of desiccation-induced mortality[18].

There were 59 genes lost in *C. australe* without detected paralogs, of which 11 were shared with other DS-seeded species. Two of these genes, *LEA2* and *FIBRILLIN5* accumulate transcripts in *M. truncatula* during seed maturation and upon re-induction of desiccation tolerance in germinated seeds[19]. The gene *GUN5*, a magnesium chelatase involved in retrograde signalling and ABA signalling to the nucleus[20], was lost in *C. australe* without paralogs. This pathway is strongly affected in *M. truncatula abi5* mutants that produce seeds with strongly reduced longevity[21] and cannot reacquire desiccation tolerance after germination[19].

76 genes lost their ortholog but kept a paralog in *C. australe*. For example, *RETARDED ROOT GROWTH-LIKE* (*RRL)* and *MOTHER OF FT* (*MFT*), involved in ABA- and brassinosteroid (BR) signalling. *RRL* mediates ABA signal transduction through *ABI4*[22] and *MFT* regulates seed germination and fertility involving ABA- and BR-signalling pathways[23].

Next, we identified 3,716 gene family expansions specific for *C. australe* compared to the other species in the phylome (Supplementary figure 3), of which 180 were associated with transposons. Expansion size ranged from 2 to 32 genes and involved 30.5% of the predicted proteome. After removal of expansions associated with transposable elements or viruses, the remaining expanded gene families were enriched for GO (gene ontology) terms such as *defence response, flavonoid biosynthetic process*, *terpene synthase activity* and *nutrient reservoir activity*. GO terms associated with *terpene synthase activity, lyase activity* and *pectinesterase activity* were also enriched at the base of Papilionoideae subfamily. Terpenes are the largest class of plant natural products, with a wide chemical diversity and essential roles in plant growth and development (e.g. gibberellins) and biotic interactions (e.g. floral scents)[24]. The

retention of duplicated genes involved in terpene synthesis at the origin of the papilionoids indicates the possible role of these compounds in the irradiation and adaptation of this sub-family.

To identify the similarities between the DS-seeded species, we investigated the enrichment of functional regions (Pfam domains) and the presence of orthologous genes exclusive to DS-seeded species. No Pfam domain was enriched exclusively in expanded gene families of all the DS-seeded species and 32 domains were enriched exclusively in expanded gene families of DS-seeded species (Supplementary table 4). For example, the Copper/zinc superoxide dismutase (SODC) domain (PF00080.18), is enriched in *E. guineensis* and the AAA_14 (ATPase associated) domain (PF13173.4) is enriched in *C. mollissima* and *Theobroma cacao*. Superoxide dismutases are critical enzymes for the protection against oxidative stress. AAA proteins are functionally diverse and are involved in processes such as DNA replication, protein degradation, signal transduction and regulation of gene expression[25].

Genes with different selective pressures

Natural populations often undergo the weakening or removal of a selective force that had been important in the maintenance of a trait, characterizing a "relaxed selection"[26]. When a DT-seeded species is subjected to an environment where DT is not an adaptive trait, there should be relaxation of its evolutionary constraints that can eventually lead to the loss of desiccation tolerance. DS-seeded species evolved in environments where the conditions favour immediate germination and its seeds are programmed to initiate germination upon, or shortly after shedding[1,18,27]. While DT seeds normally form seed banks in the soil, DS seeds germinate immediately and usually form seedling banks under shaded forest canopy and take advantage of an eventual light gap for faster establishment. Furthermore, the generally increased seed size in these species favours seedling establishment under shaded forest conditions[18]. In both scenarios, the protein coding genes associated with desiccation tolerance would be subjected to different selective pressure that could be estimated by the ratio of the number of nonsynonymous substitutions per non-synonymous site (dN) in a given period of time to the number of synonymous substitutions per synonymous site (dS) in the same period[28].

Genome wide analysis of protein coding genes of *C. australe* in comparison with other sequenced legumes enabled identification of genes with 2-fold higher dN/dS in *C. australe* (Extended data table 2). Among these were the genes associated with hormonal signalling, such as *ACTIVATION-TAGGED BRI1* (*BRASSINOSTEROID-INSENSITIVE1*)-*SUPPRESSOR1* (*ATBS1*) (TF involved in gibberellin- and BR signalling[29]), Ethylene Insensitive 3 family protein (TF involved in ethylene signalling and seed desiccation tolerance[19,30,31]), *BRI1-ASSOCIATED RECEPTOR KINASE*

(*BAK1*) and *SALT TOLERANCE HOMOLOG2* (*STH2*). Other examples are: *SEEDSTICK* (*STK*), *TRANSPARENT TESTA5* (*TT5*), *ENDO-BETA-MANNANASE7* (*MAN7*), *FLOWER FLAVONOID TRANSPORTER* (*FFT*) and *ROTUNDIFOLIA3* (*ROT3*).

Gene displacements

Syntenic relationships across species can be used to investigate the evolution of gene families, such as those involved in the regulation of development. We investigated whether the loss of synteny in *C. australe* genes could be related to the loss of seed desiccation tolerance. There are 169 genes re-arranged in the *C. australe* genome that have syntenic orthologs in 50 angiosperm species. Most noteworthy, among these were the genes *BRASSINOSTEROID INSENSITIVE 3* and *5* (*BIN3* and *BIN5*), which participate in BR signalling and are associated with seed size[32]. Furthermore, genes such as *MINISEED3* (*MINI3*) and *HAIKU1* (*IKU1*), regulators of seed size via the BR pathway in *A. thaliana*[33], also lost synteny in *C. australe*, which could contribute to the large seed size in this species.

The synteny between the genome of *C. australe* and other legume species was evaluated by aligning the genome of *C. australe* against itself and against the genomes of *G. max* and *M. truncatula* (Supplementary figure 4). Most of the *C. australe* genome is syntenic with itself and mostly duplicated after the WGD event. While the duplicates are associated with the most recent WGD, many paralogs derived from the gamma event were also detected. The alignment of *C. australe* against *G. max* indicated a high amount of syntenic orthologs and paralogs, whereas the alignment of *C. australe* against *M. truncatula* indicated that although many syntenic orthologs have been conserved, most of the WGD-derived paralogs were lost. Moreover, most of the duplicated regions retained by *M. truncatula* were also duplicated and retained in *C. australe*.

The expression pattern during seed maturation of *C. australe* genes that lost synteny was compared to their *M. truncatula* orthologs (Supplementary figure 5). Genes involved in carbohydrate metabolism and control of cell proliferation, such as a beta-galactosidase (Mtr.32552.1.S1_at), a xyloglucan galactosyltransferase (Msa.1340.1.S1_at) and a TCP family TF (Mtr.10063.1.S1_at), accumulate transcripts in *C. australe* during seed development but decrease in *M. truncatula*. It is likely that these genes are involved in the germination program, which remains active in *C. australe*. In contrast, indole-3-acetic acid-amido synthetases, involved in auxin homeostasis, accumulated transcripts in *M. truncatula* but decreased in *C. australe*. Auxin has been reported to maintain seed dormancy by interacting with ABA[34].

Seed development in *Castanospermum australe*

Seed development in *C. australe* occurs over a period of 6 months, with reserve accumulation and maximum embryo size completed approximately 3 months after flowering and coincident with pods becoming yellow (Figure 2, Supplementary table 5). Unlike the embryo, however, the seed coat declined in mass until just prior to this stage (yellow green pods), with no further decline once embryos reached full size. During reserve accumulation, there was a decline in water content of all tissues, with this loss stabilizing in all but the seed coat (which continues to lose water) once reserve accumulation was complete at the yellow pod stage. There was no "maturation drying" typical of DT seeds after this stage[7]. While seeds had the capacity to germinate prior to reaching full embryonic size, germination rate was slow. Full germination capacity, typically of newly shed seeds was achieved at the yellow pod stage. The lethal water content of axes below which 50% viability was lost after drying declined from 0.45 gH_2O/g^{-1} dry mass in those extracted from green pods, to 0.23 gH_2O/g^{-1} dry mass in axes from yellow pods, after which there was no significant change. Cotyledons were more sensitive to dehydration, with 50% loss of viability occurring below 0.82 gH_2O/g^{-1} dry mass in those from green pods, this declining slightly with development to 0.70 gH_2O/g^{-1} dry mass in cotyledons from brown pods.

Seed development in *C. australe* vs *M. truncatula*

DT seeds become tolerant of drying midway during seed development, concomitant with reserve accumulation. From this stage onwards, there is progressive loss of water, the final stages thereof termed 'maturation' drying occurring after reserve accumulation is complete and full seed size is attained. Previous work has identified transcripts that accumulate during the acquisition of DT in *M. truncatula*[19,31]. Homologs of 121 of these genes failed to accumulate in *C. australe* cotyledons in comparable seed developmental stages (Extended data table 3). These transcripts are related to sugar metabolism, photosynthesis, seed development, protection against abiotic stress and modulation of plant stress responses. Examples include *ABI3*, *ABI5*, chaperone proteins, heat shock factor proteins, putative LEAs, transparent testa protein, oleosins, 1-CYSTEINE PEROXIREDOXIN, and α-galactosidases.

Next, we identified 269 transcripts with increasing abundance in *M. truncatula* during final maturation and likely to be involved in longevity (life span in the dried state)[31,35], but showing an opposite expression profile in *C. australe* (Extended data table 4). They are related to metabolic and catabolic processes, such as *lipid metabolic*

process, cellular lipid metabolic process and *nitrogen compound metabolic process* (Supplementary table 6), reflecting the kind
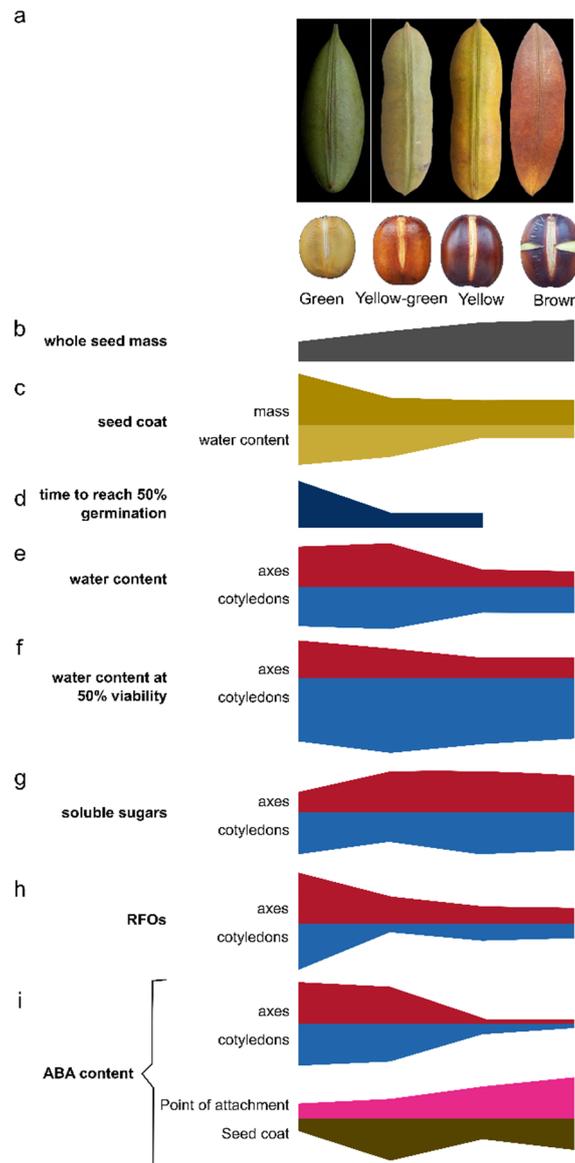
Figure 2. *Castanospermum australe* **late seed development and its associated phenotypical parameters.** (a) Seed develpmental stages based on the colour of the pods, i.e. green, yellow-green, yellow and brown. All the subsequent parameters were based on this categories. (b-i) Qualitative chages several seed-related parameters during development. (c) Mass and water content of the seed coat. (d) Time to reach 50% germination. (e) Water content of axes and cotyledons. (f) Water content at 50% loss of viability in axes and cotyledons. (g) Concentration of soluble sugars in axes and cotyledons. (h) Concentration of raffinose family oligosaccharides (RFOs) in axes and cotyledons. (i) ABA content of different tissues.

of reserves accumulated by *M. truncatula*. Mature seeds of this species predominantly accumulate protein (30-40% DW) but may also contain some lipids (7-9% DW) and a small amount of starch (< 1% DW)[36]. It is likely that the carbon from starch is used either for lipid synthesis or protein synthesis, which provides volume buffering capacity against cell collapse during drying[1,36]. *C. australe* seeds, to the contrary, have starch as main reserve (85% DW) and lipids and proteins constitute only 3-6% and 2.8% DW respectively (Supplementary figure 6). We also identified 296 transcripts with decreasing abundance in *M. truncatula* and increasing abundance in *C. australe*. They are related to *root development, developmental process* and *regulation of localization* (Supplementary table 7), indicating that towards the end of seed development, *C. australe* is metabolically active while *M. truncatula* is entering a phase of low metabolic activity and quiescence. Examples of transcripts in these categories are *β-AMYLASE 1* (*BAM1*), auxin-related genes, phospholipase D, thiamine thiazole synthases, soluble starch synthase and triosephosphate isomerase.

LEA proteins have been related to survival in the dry state[9,37] and to responses to environmental stresses, including desiccation[38]. In *C. australe*, several LEA proteins failed to accumulate in the cotyledons[10]. A genome-wide search for LEAs in the genome of *C. australe* identified 94 LEA motif-containing proteins, a number similar to what has been described for DT-seeded species (for example 94 in *Oropetium thomaeum*, 88 in *A. thaliana* and 99 in *Sorghum bicolor*). A hierarchical clustering of expression of the LEAs in developing seeds of *C. australe* and *M. truncatula* separated the transcripts in two major clusters (Supplementary figure 7). LEAs in the first cluster belong to different families and transcript abundance increased considerably in *M. truncatula* seeds towards late maturation. However, in *C. australe* most of these increased only slightly in early stages of development but declined during the later stages. LEAs in the second cluster belong to the LEA_2 family and do not undergo major changes in transcript abundance in either species. This family encodes 'atypical' LEA proteins because of their more hydrophobic character compared to other LEA families[39]. Functional studies on LEA_2 proteins suggest that they do not act in the protection of membranes in tissues undergoing dehydration, although some proteins of this family were shown to have enzyme protective properties under both freezing and drying conditions[40].

Changes in soluble sugar content and composition have been described as a characteristic of late seed maturation in DT seeds and correlate with the acquisition of longevity and preparation for the dry state[5,41]. The percentage of soluble sugars in seeds of *C. australe* (7-10%) is comparable with the average percentage for legume species (8-10%[36]). However, while raffinose family oligosaccharides (RFOs) are the main sugars in *M. truncatula*, comprising 90% of the total soluble sugar content, in *C. australe*, only minute amounts of RFOs, mainly stachyose, could be detected (0.7% of total soluble sugars at the brown pod stage). Stachyose content was highest in seeds from the green pod stage and decreased with further progress of maturation

(Supplementary figure 8). A similar finding has been reported for the non-viviparous highly DS seeds of *Avicenia marina*[42]. In parallel, sucrose and glucose content increased. The reduction of stachyose content during further maturation suggests hydrolysis, normally occurring in germinating DT seeds. A comparative analysis of transcripts linked to RFO metabolism between *C. australe* and *M. truncatula* identified transcripts of genes related to the synthesis of sucrose from fructose-6 phosphate that remained high in *C. australe* whereas they decreased in *M. truncatula*. Conversely, the transcripts of several genes related to the synthesis of galactinol or raffinose and stachyose accumulated during development of *M. truncatula* seeds while their abundance remained low in *C. australe*. This set of genes might explain the lack of RFO accumulation in *C. australe* seeds. Whereas the specific roles of RFO compared to the nonreducing sucrose remain unclear, the DS Arabidopsis *abi3* mutants as well as Mt-*abi5* are also impaired in the accumulation of RFOs.

ABA and Brassinosteroids

ABA is involved in the production of mature DT seeds, and inhibition of their subsequent germination under conditions unfavorable for seedling growth[43]. During seed development, an increase in ABA content has been related to a transition from growth by cell division to growth by cell enlargement and to cell cycle arrest at the G1/S transition[43]. ABA content of *C. australe* embryos was high during the early developmental stages but declined considerably in both axes and cotyledons in the transition from the yellow-green to the yellow pod stage (Figure 2, Supplementary table 8). ABA content increased considerably in the seed coat in the transition from the yellow to the brown pod stage, especially in the point of attachment (tissue that attaches the embryo to the pod). This increase may be related to the role of ABA in promoting senescence, a process which precedes abscission[43]. Additionally, the higher ABA content in the seed coat could play a role in delaying germination of the embryo until ideal conditions for germination are met or to aid temporal and/or spatial dispersal. Interestingly, the seed coat starts to peel away from the embryo in mature seeds (Figure 2) and this increases markedly once seeds are shed.

ABA is a key regulator of abiotic stress responses, including drought and desiccation, and plant development, especially acquisition of desiccation tolerance during seed development[44]. Disruption of ABA biosynthesis or -signalling leading to lack of or insensitivity to ABA results in loss of seed desiccation tolerance[35]. We observed alterations in genes related to ABA signalling, such as *MOTHER OF FT* (*MFT*), *GPCR-TYPE G PROTEIN 1* (*GTG1*), *STH2*, *ABI3* and *ABI5*. The *MFT* ortholog was lost whereas the paralog was maintained in *C. australe*. Mutations in this gene may cause ABA hypersensitivity at germination and is associated with dormancy[45]. *STH2* is

involved in ABA signalling, is highly expressed during embryogenesis[46] and has a high dN/dS in *C. australe*. *GTG1* was lost in *C. mollissima* and its knock-out causes ABA hyposensitivity in *A. thaliana* seeds[47]. ABI3 and ABI5 showed contrasting expression patterns during *C. australe* seed development compared to *M. truncatula*. These two TFs were shown to play essential roles in seed development and desiccation tolerance acquisition[44].

BRs have been implicated in seed development and are known to antagonize seed dormancy and stimulate germination[48]. In *C. australe,* several genes involved in BR biosynthesis and signaling have undergone genetic changes and among them several have been studied revealing their involvement in seed development (Figure 3). For example, *BRASSINOSTEROID INSENSITIVE-LIKE 3* (*BRL3*) lost an ortholog but kept a paralog; *IKU1* and *MINI3* lost synteny; and *ROT3, DE-ETIOLATED 2* (*DET2*)*, ATBS1* and *BAK1* have higher dN/dS in *C. australe* than in *M. truncatula*. Furthermore, an ortholog of *ATBS1* was highly expressed during late development of *C. australe* contrasting with the decreasing expression in *M. truncatula* seeds. These data support the hypothesis that subcellular metabolism associated with germination is initiated during the late stages of development in DS seeds. Overall, our results support the hypothesis that the evolution of DS was not caused by massive alterations in enzymes and structural proteins but instead by discrete mutations in regulatory genes.
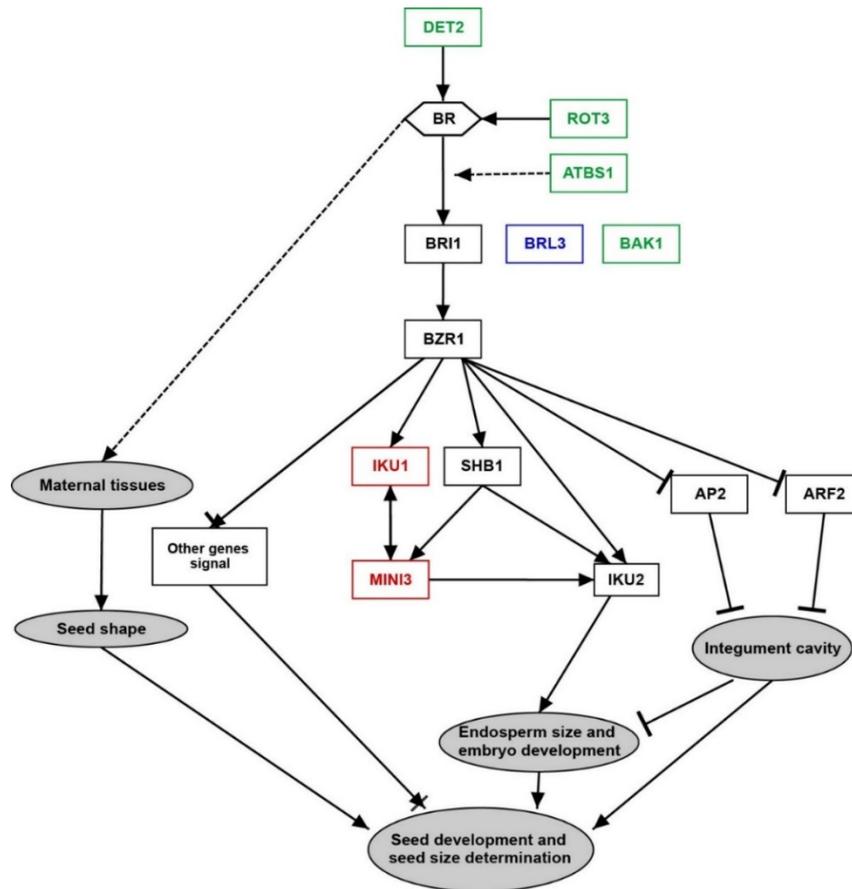
Figure 3. Hypothetical model for brassinosteroid-regulated seed development (adapted from[97]**)**. Red shapes indicate genes that lost synteny in *Castanospermum australe* compared to *Medicago truncatula*. Green shapes indicate genes with higher dN/dS in *C. australe* than in *M. truncatula*. Blue shapes indicate genes that lost an ortholog, but kept a paralog in *C. australe*. BR: brassinosteroid. ATBSI1: *ACTIVATION-TAGGED BRI1 (BRASSINOSTEROID-INSENSITIVE1)-SUPPRESSOR1*. AP2: *APETALA 2*. ARF2: *AUXIN RESPONSE FACTOR 2*. BAK1: *BRI1-ASSOCIATED RECEPTOR KINASE 1*. BRI1: *BRASSINOSTEROID INSENSITIVE 1*. BRL3: *BRASSINOSTEROID INSENSITIVE-LIKE 3*. BZR1: *BRASSINAZOLE RESISTANT 1*. DET2: *DE-ETIOLATED 2*. IKU: *HAIKU*. MINI3: *MINISEED 3*. ROT3: *ROTUNDIFOLIA 3*. SHB1: *SHOEBOX 1*.

Conclusions

Seed desiccation sensitivity evolved multiple independent times mainly in environments where water is highly abundant and predictable across long periods, favouring immediate seed germination. In such environments, the evolutionary pressure for DT seeds is relaxed and the production of DS seeds is not disadvantageous. This was the case for *C. australe*. Among the currently known DS-non-viviparous seeds, *C. australe*

is one of the most sensitive to water loss. We have pinpointed some of the factors behind this sensitivity, namely displacements, loss of synteny and mis-expression of specific genes related to the BR and ABA signalling pathways, carbon metabolism, control of cell proliferation, protection against abiotic stresses and modulation of plant stress responses. These alterations are likely to have led to an increased seed size; high starch and low protein and lipid seed content; low accumulation of LEA proteins and RFOs in the seeds; and failure to start the seed maturation drying phase. Altogether these factors have led to the production of DS seeds.

The low similarity between DS-seeded species confirms the hypothesis that desiccation sensitivity evolved independently. Moreover, it supports the idea that although the evolution of many factors was necessary for the appearance of seed desiccation tolerance, only a few changes in some of these factors are enough for its loss. The understanding of these changes will guide future strategies to improve *ex situ* conservation of DS seeds in seed banks.

Methods

Plant material

A population of trees of *C. australe* growing in Pietermaritzburg (Kwazulu-Natal Province, South Africa) was the source of plant material for this work. Seed development occurs over a 6-month period, during which pods were harvested weekly. Seeds were extracted and, following histodifferentiation, were separated into component tissues (axis, cotyledons and seed coat). The following was determined annually over 2 seasons. Whole seed mass and that of component tissues (n = 60) and water content (n=10-20) was determined gravimetrically by oven drying. The ability of intact seeds to germinate was tested by planting in vermiculite. Once germinable, the amount of water loss tolerated by axes and cotyledons was determined by flash drying[49]. Axis survival was determined as the ability to produce both shoots and roots when cultured in vitro on full MS medium. Survival of cotyledons was assessed by tetrazolium staining followed by spectrometric analysis[50].

Genome sequencing and assembly

Freeze-dried leaf material was used for DNA isolation as described by[51] with modifications. The genomic *C. australe* library consisted of 30x coverage PacBio with a mean read length of 7.8 Kb. In addition, an Illumina paired end library with reads of

100 bp and a 200-400 bp insert size was constructed and sequenced to a 64x coverage. Reads originating for contaminants were removed from all sequence data prior to assembly. Organelle genomes were also removed from the main genome assembly. Illumina reads were error-corrected using Lighter[52] and assembled using SparseAssembler[53]. A hybrid assembly was produced with DBG2OLC[54] and the contigs were reordered and connected into scaffolds using SSPACE-LongRead[55]. The assembly was polished using Sparc[53] and Pilon[56]. PBJelly2[57] was used for gap closure and genome improvement. Alignments due to gene duplication and repeats were filtered out using the delta-filter utility of the MUMmer package[58]. The assembly was validated by mapping the available RNA and DNA libraries to the genome with Bowtie2[59] and Blasr[60]. Assembly statistics were calculated using QUAST[61]. Gene space completeness was measured using BUSCO (Benchmarking Universal Single-Copy Orthologs[62]).

Genome annotation

The *ab initio* predictors AUGUSTUS[63] and SNAP[64] were trained on the transcriptome data using BRAKER1[65]. The MAKER2 annotation pipeline[66] was applied for gene prediction and repeat annotation. Predicted genes were functionally annotated by a consensus approach using InterProScan[67], Gene Ontology, Kyoto Encyclopedia of Genes and Genomes (KEGG[68]), Swiss-Prot with Translated EMBL Nucleotide Sequence Data Library (TrEMBL[69]), and BLAST2GO[70]. Repeatmodeler[71] was applied to build a *de novo* repeat library and identify ribosomal and small nuclear RNAs. Repetitive sequences in the assembly were soft masked using RepeatMasker[72]. SNPs and INDELs were called relative to the genome using Freebayes[73]. Detected polymorphisms from short-read alignments were discarded when the quality was below 20.

Phylome Reconstruction

We reconstructed three phylomes, one for a species set closely related to *C. australe* (phylome 110) and starting in *C. australe*, a second one based on a broader taxonomic focus also starting in *C. australe* (111) and a third one also broad but starting in *Fragaria vesca* (112). The phylome 112 was used to search for lost genes in *C. australe*. It starts in *F. vesca* that is the closest outgroup DT seeded species with a high-quality genome sequence. A complete list of species used in the phylomes can be found in Supplementary Table 9. Both phylomes were reconstructed using the same approach[74].

First, we reconstructed seven neighbour joining trees using BioNJ as implemented in phyML[75], these trees were used to assess the likelihood of seven different evolutionary models (JTT, WAG, MtREV, VT, LG, Blosum62, CpREV and

DCMut). The best model according to the AIC criterion was used to reconstruct the final maximum likelihood tree using phyML v3.0[76], a discrete gamma-distribution model with four rate categories plus invariant positions was used, the gamma parameter and the fraction of invariant positions were estimated from the data. All data generated during the phylome reconstruction has been deposited in phylomeDB (PMID:24275491) under the phylomeID codes 110 and 111. We built a third phylome by adding to Phylome 110 data from six transcriptomes of species belonging to different subfamilies of the Fabaceae family. For each transcript, we obtained the longest ORF and we kept the ORF with a length of 100 amino acids or longer. Then, the proteins of the transcripts were introduced into each tree of this phylome using the following pipeline. First, a homology search using BLASTP was performed from the seed protein against a database that contained the transcriptomes. Results were then filtered based on three thresholds: e-value < $1e^{-05}$, overlap between query with hit had to be at least of 0.3 and an identity threshold > 40.0%. Those hits that passed these filters were incorporated into the raw alignment of the phylome using MAFFT (v 7.222) (--add and --reorder options)[77]. Then RAxML (-f v -G 0.1)[78] was used to incorporate the transcriptome sequenced into the tree without reconstructing it again. Trees were then filtered to remove unreliable transcriptome sequences. Phylomes tend to be redundant, therefore a same transcriptomic sequence can be inserted in many trees. For each inserted transcript, we checked whether it was inserted in the same region in all the trees in which it was present. If this was not the case, the transcript was deemed unreliable and removed from the tree.

Each individual gene tree was automatically analysed with phylogeny-based algorithms that predict orthology and paralogy relationships[79] and detect and date duplication events[74]. The resulting trees, alignments and orthology and paralogy predictions are accessible to browse or download at the PhylomeDB database[80]. A set of 183 one-to-one orthologous proteins present across the compared species was used to reconstruct a species phylogeny. Then, detected duplications were mapped onto the species tree and average duplication frequencies per gene were computed for each branch, divided by the number of trees that contained each node.

Phylomes were analysed using ETE v3.0. Orthology and paralogy relations were extracted using a species overlap algorithm. This algorithm traverses the tree node by node from the seed protein to the root and calls a speciation node when no overlap exists between species at either side of the node. If such overlap exists, then it calls a duplication node. Sequences separated by a speciation node are orthologs while sequences separated by a duplication are paralogs. Species trees were reconstructed using a super-tree approach with the tool duptree[81] in all the cases. A second species tree was reconstructed selecting single copy and widespread proteins. For the first phylome, we first pruned the gene trees to remove the species-specific duplications and then 153 proteins, that were present in at least 19 out of the 20 species included, were chosen, and

a concatenated alignment was reconstructed that had 95,083 amino acid positions. For the second phylome 183 proteins were chosen and an alignment of 134,482 amino acid positions was reconstructed. PhyML was used to reconstruct species trees. All the trees had a congruent topology and *C. australe* adopted a basal position within the Papilionoideae subfamily.

After removing the expansions associated with terms related to transposable elements or viruses, we performed a functional enrichment analysis using FatiGo[82].

Whole genome duplication analysis

The *C. australe* genome was compared to genomic data from five Legumes: *Glycine max*, *Lotus japonicus*, *Medicago truncatula*, *Phaseolus vulgaris*, *Trifolium pratense*. *G. max* and *M. truncatula* are tetraploids and the rest are diploid species. The diploid *Fragaria vesca* species, of the family Rosaceae, was selected as outgroup as it is one of the closest relatives with a completed genome available outside of the Legume family. The assemblies of: *G. max*, *L. japonicus*, and *M. truncatula* are on chromosome level which made it easier to identify genome collinearity and duplication patterns.

Synteny analysis

A synteny network approach[83] was implemented to compare the synteny of *C. australe* and other whole-genome sequenced legume species. Briefly, annotated genome peptide sequences for the following legume species, *Vigna radiate*, *Vigna angularis*, *P. vulgaris*, *G. max*, *Cajanus cajan*, *T. pretense*, *M. truncatula*, *Arachis duranensis*, *L. japonicus*, and *Cicer arietinum*, were downloaded from LIS (https://legumeinfo.org). BLASTP was used for pairwise genome comparison and MCScanX[84] was used for synteny block detection. Infomap algorithm[85] implemented in R igraph package was used for synteny network clustering. Node composition of each network cluster was profiled afterwards. Clusters containing genes/nodes from more than 8 (out of the 10) legume species but no *C. australe* node(s) were screened out for further investigation.

Genomes were loaded into COGE's Synmap2[86] to perform self-alignments, generate syntenic dotplots and identify syntenic regions. Default tools and settings were used for all steps in the Synmap pipeline. The fast Last[87] algorithm was used to create a blastable database. The maximum distance between two matches was 20 genes, a syntenic block consists of minimum 5 genes and no blocks were merged. Quota Align was enabled to determine the syntenic depth (the number of times a genomic region is syntenic). For calculating the fractionation bias, the window size was lowered from 100 to 25 considering the smaller contigs of the *C. australe* genome. Synonymous (Ks) and non-synonymous (Kn) site mutations were calculated for each syntenic gene pair.

Synonymous mutations are changes in DNA coding for protein, but do not change the encoded amino acid. For non-synonymous, the mutations do lead to a different amino acid. Kn and Ks values were represented as histogram to visualize the distribution of the mutation rates. Mutation rates were used to determine if genes were duplicated by the WGD event or not. The distribution of the Ks rate was used to set a different cut-off per species.

Dotplots were generated using a syntenic path assembly algorithm to order contigs based on their synteny. Synfind was used to search for syntenic regions using a gene from one of the three genomes as reference query. The four highest scoring of *G. max* syntelogs, two highest scorings for *C. australe* and two highest scoring for *M. truncatula* were selected. These selected regions represent the orthologous and paralogous regions. Four syntelogs were selected for *G. max* due to independent duplications. The regions were visualized in COGE's genome browser to infer shared synteny and study fractionation. Together with the fractionation bias and syntenic dotplots they were used to infer if *C. australe* shared the polyploidy event with other legumes.

dN/dS analysis

We used the genome of 5 legume species from phytozome (*M. truncatula, G. max, Glycine soja, T. pratense* and *P. vulgaris*) and *C. australe*. The SynMap tool in the online CoGe portal was used to find syntenic gene pairs within these species and calculate their dN/dS[88]. We set the DAGChainer on a maximum distance between two matches for 50 genes and minimum number of aligned pairs on 3 genes. To establish sets of orthologous among the 5 species against *M. truncatula*, the method of reciprocal best hits using Last was used. The global sequence alignment of the protein sequences used the Needleman-Wunsch algorithm in nwalign using the BLOSOM62 scoring matrix.

Codeml in the Phylogenetic Analysis by Maximum Likelihood (PAML) package was used to estimate the dN (the rate of non-synonymous substitutions), dS (the rate of synonymous substitutions) and the ratio of dN/dS[89].

Orthologs with dS>5, dN>2 or dN/dS>2 were filtered. For genes with multiple syntelogs we kept the pair with the lowest dN/dS.

Transcriptome analysis

C. australe seeds were harvested during maturation and at shedding from trees growing in Pietermaritzburg in 2009 and 2011. Total RNA was extracted with phenol from cotyledons or embryonic axes[90].

Signal intensities were extracted and analysed with Nymblegen slides using specifically designed for *C. australe* seeds. For axis data, fold changes were calculated comparing subsequent time-points. For cotyledons data, fold changes were calculated comparing each time-point with the first one (cotyledons between 1,51 g and 2.5 g). A gene was considered differentially expressed if $P \leq 0.05$ in at least one comparison (axis or cotyledon) after the application of linear modelling.

Over-representation analysis (ORA) was used to recover over-represented biological processes using the app BiNGO (default settings)[91] for Cytoscape. Redundant gene ontology (GO) terms were removed applying a cut-off of <0.1 for the term's semantic distance with respect to other semantically close terms ('Dispensability") as calculated by ReviGO[92].

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

Members of the eight LEA protein families were identified uploading Hidden Markov Models (HMM) for each family from the PFAM database[93] to HMMER 3.1b2 package[94] and searching the genome of *C. australe* with 'hmmsearch'. The HMM profiles used are: Dehydrin – PF00257, LEA_1 – PF03760, LEA_2 – PF03168, LEA_3 – PF03242, LEA_4 - PF02987, LEA_5 – PF00477, LEA_6 − PF10714, and Seed Maturation Proteins – PF04927. All proteins with significant hits (E-value ≤ 0.01) were selected.

Sugar and ABA content and determination

Sugars were extracted from frozen and lyophilised seeds and analysed by HPLC on a Carbopac PA-1 column[95] (Dionex Corp., Sunnyvale, CA, USA). Three independent extractions and assays were performed on approx. 100 mg of tissue.

ABA was extracted and quantified as described by Floková et al[96].

Data availability

Sequence and transcriptome data are available from the National Center for Biotechnology Information (NCBI) under BioProject ID PRJNA291138 (SAMN03940282; Sample name: Caus01; SRA: SRS1043905) and GSE109217, respectively.

Supplementary material:

References

1. Berjak, P. & Pammenter, N. W. Implications of the lack of desiccation tolerance in recalcitrant seeds. *Front. Plant Sci.* **4,** 1–9 (2013).

2. Hamilton, K. N., Offord, C. A., Cuneo, P. & Deseo, M. A. A comparative study of seed morphology in relation to desiccation tolerance and other physiological responses in 71 Eastern Australian rainforest species. *Plant Species Biol.* **28,** 51–62 (2013).

3. Berjak, P. & Pammenter, N. W. What ultrastructure has told us about recalcitrant seeds. *Rev. Bras. Fisiol. Veg.* **12,** 22–55 (2000).

4. Berjak, P. & Pammenter, N. W. From Avicennia to Zizania: Seed recalcitrance in perspective. *Ann. Bot.* **101,** 213–228 (2008).

5. Leprince, O., Pellizzaro, A., Berriri, S. & Buitink, J. Late seed maturation: drying without dying. *J. Exp. Bot.* **68,** 827–841 (2017).

6. To, A. *et al.* A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* **18,** 1642–1651 (2006).

7. Farrant, J. M., Pammenter, N. W. & Berjak, P. Seed development in relation to desiccation tolerance: a comparison between desiccation-sensitive (recalcitrant) seeds of *Avicennia marina* and desiccation tolerance types. *Seed Sci. Res.* **3,** 1–13 (1993).

8. Francini, A. *et al.* Enzymatic and non-enzymatic protective mechanisms in recalcitrant seeds of *Araucaria bidwillii* subjected to desiccation. *Plant Physiol. Biochem.* **44,** 556–563 (2006).

9. Han, B., Berjak, P., Pammenter, N. W., Farrant, J. M. & Kermode, A. R. The recalcitrant plant species , *Castanospermum australe* and *Trichilia dregeana*, differ in their ability to produce dehydrin-related polypeptides during seed maturation and in response to ABA or water-deficit-related stresses. *J. Exp. Bot.* **48,** 1717–1726 (1997).

10. Delahaie, J. *et al.* LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. *J. Exp. Bot.* **64,** 4559–4573 (2013).

11. Cardoso, D. *et al.* Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. *Am. J. Bot.* **99,** 1991–2013 (2012).

12. Schranz, M. E., Mohammadin, S. & Edger, P. P. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.* **15,** 147–153 (2012).

13. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10,** 725–732 (2009).

14. Matias-Hernandez, L. *et al.* VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. *Plant Cell* **22,** 1702–1715 (2010).

15. Mendes, M. A. *et al.* Live and let die: a REM complex promotes fertilization through synergid cell death in Arabidopsis. *Development* **143,** 2780–2790 (2016).

16. Libault, M. & Stacey, G. Evolution of FW2.2-like (FWL) and PLAC8 genes in eukaryotes. *Plant Signal. Behav.* **5,** 1226–1228 (2010).

17. Frary, A. fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. *Science.* **289,** 85–88 (2000).

18. Daws, M. I., Garwood, N. C. & Pritchard, H. W. Prediction of desiccation sensitivity in seeds of woody species: a probabilistic model based on two seed traits and 104 species. *Ann. Bot.* **97,** 667–674 (2006).

19. Terrasson, E. *et al.* An emerging picture of the seed desiccome: confirmed regulators and newcomers identified using transcriptome comparison. *Front. Plant Sci.* **4,** 1–16 (2013).

20. Jiang, S.-C., Mei, C., Wang, X.-F. & Zhang, D.-P. A hub for ABA signaling to the nucleus: Significance of a cytosolic and nuclear dual-localized PPR protein SOAR1 acting downstream of Mg-chelatase H subunit. *Plant Signal. Behav.* **9,** e972899 (2014).

21. Zinsmeister, J. *et al.* ABI5 is a regulator of seed maturation and longevity in legumes. *Plant Cell* **28,** 2735–2754 (2016).

22. Park, S.-Y. *et al.* Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* **520,** 545–8 (2015).

23. Sun, Y. *et al.* Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in arabidopsis. *Dev. Cell* **19,** 765–777 (2010).

24. Tholl, D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **9,** 297–304 (2006).

25. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 93–114 (2006).

26. Lahti, D. C. *et al.* Relaxed selection in the wild. *Trends Ecol. Evol.* **24,** 487–496 (2009).

27. Farrant, J. M., Berjak, P., Cutting, J. G. M. & Pammenter, N. W. The role of plant growth regulators in the development and germination of the desiccation sensitive recalcitrant seeds of *Avicennia marina*. *Seed Science Research* **3,** 55–63 (1993).

28. Mugal, C. F., Wolf, J. B. W. & Kaj, I. Why Time matters: codon evolution and the temporal dynamics of dN/dS. *Mol. Biol. Evol.* **31,** 212–231 (2014).

29. Schlereth, A. *et al.* MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* **464,** 913–916 (2010).

30. González-Morales, S. I. *et al.* Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **113,** E5232–E5241 (2016).

31. Righetti, K. *et al.* Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *Plant Cell* **27,** tpc.15.00632 (2015).

32. Yin, Y. *et al.* A crucial role for the putative *Arabidopsis* topoisomerase VI in plant growth and development. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 10191–6 (2002).

33. Luo, M., Dennis, E. S., Berger, F., Peacock, W. J. & Chaudhury, A. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proc. Natl. Acad. Sci.* **102,** 17531–17536 (2005).

34. Liu, X. *et al.* Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. *Proc. Natl. Acad. Sci.* **110,** 15485–15490 (2013).

35. Verdier, J. *et al.* A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* **163,** 757–774 (2013).

36. Djemel, N. *et al.* Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol. Biochem.* **43,** 557–566 (2005).

37. Chatelain, E. *et al.* Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant, Cell Environ.* **35,** 1440–1455 (2012).

38. Tunnacliffe, A. & Wise, M. J. The continuing conundrum of the LEA proteins. *Naturwissenschaften* **94,** 791–812 (2007).

39. Hundertmark, M. & Hincha, D. K. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9,** 118 (2008).

40. Dang, N. X., Popova, A. V., Hundertmark, M. & Hincha, D. K. Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and

in vitro. *Planta* **240,** 325–336 (2014).

41. Wang, M. *et al.* LegumeGRN: A gene regulatory network prediction server for functional and comparative studies. *PLoS One* **8,** e67434 (2013).

42. Farrant, J. M., Pammenter, N. W. & Berjak, P. Development of the recalcitrant (homoiohydrous) seeds of *Avicennia marina*: anatomical, ultrastructural and biochemical events associated with development from histodifferentiation to maturation. *Ann. Bot.* **70,** 75–86 (1992).

43. Finkelstein, R. Abscisic Acid synthesis and response. *Arabidopsis Book* **11,** e0166 (2013).

44. Dekkers, B. J. W. *et al.* Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. *Planta* **241,** 563–577 (2015).

45. Vaistij, F. E. *et al.* Differential control of seed primary dormancy in *Arabidopsis* ecotypes by the transcription factor SPATULA. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10866–71 (2013).

46. Xu, D. *et al.* Convergence of light and ABA signaling on the *ABI5* promoter. *PLoS Genet.* **10,** e1004197 (2014).

47. Pandey, S., Nelson, D. C. & Assmann, S. M. Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. *Cell* **136,** 136–148 (2009).

48. Steber, C. M. A role for brassinosteroids in germination in Arabidopsis. *Plant Physiol.* **125,** 763–769 (2001).

Acknowledgement

Supplemental Material

Supplementary figure 1. Histograms of synonymous mutations (Ks) of duplicated genes (**a-f**) and average Ks of syntenic blocks (**g-l**). (**a-b**) *Castanospermum australe*, (**c-d**) *Lotus japonicus*, (**e-f**) *Medicago truncatula*, (**g-h**) *Glycine max*, (**i-j**) *Phaseolus vulgaris*, (**k-l**) *Trifolium pratense*.

Supplementary figure 2. Phylogenetic tree showing duplication rates of selected species including species of the subfamilies: Mimosoideae (inside yellow rectangle) and Caesalpinioideae (inside orange rectangle) and Papilionoideae (inside green rectangle). Duplication rates are shown in blue and the position of the whole genome duplication is marked by the red star.

Supplementary figure 3. Distribution of cluster size resulting from *Castanospermum australe*-specific duplications

Supplementary figure 4. (**a-c**) Histograms of $\log_{10}$ transformed Ks (rate of synonymous mutations) of syntenic genes pairs identified through the alignment of (**a**) *Castanospermum australe* with itself, (**b**) *C. australe* with *Medicago truncatula*, and (**c**) *C. australe* with *Glycine max*. (**d-f**) Syntenic dotplots of comparisons of (**d**) *C. australe* with itself, (**e**) *C. australe* with *M. truncatula*, and (**f**) *C. australe* with *G. max*. Syntenic dotplots are coloured by their Ks values shown in a-c.

Supplementary figure 5. Hierarchical clustering of expression values of non-syntenic genes exclusively in *Castanospermum australe*. C2.5: cotyledons weighting between 1.51 and 2.5g. C4.5: cotyledons weighting between 3.51 and 4.5g. C7.5: cotyledons weighting between 5.51 and 7.5g. GC: green cotyledons. YGC: yellow-green cotyledons. BC: brown cotyledons. GA: green axes. YGA: yellow green axes. BA: brown axes. DAP: days after pollination.

Supplementary figure 6. Composition of *Castanospermum australe* mature seeds.

Supplementary figure 7. Hierarchical clustering of expression of Late embryogenesis abundant (LEA)-domain containing genes in developing seeds of *Castanospermum australe* and *Medicago truncatula*. C2.5: cotyledons weighting between 1.51 and 2.5g. C4.5: cotyledons weighting between 3.51 and 4.5g. C7.5: cotyledons weighting between 5.51 and 7.5g. GC: green cotyledons. YGC: yellow-green cotyledons. BC: brown cotyledons. GA: green axes. YGA: yellow green axes. BA: brown axes. DAP: days after pollination. ABS: pod abscission. DS: dry seeds. SMP: seed maturation proteins

Supplementary Figure 8. (**a-f**) Sugar content in seeds of *Castanospermum australe* (Caus) and *Medicago truncatula* (Mtru) (**g**) Hierarchical clustering of genes involved in RFO synthesis in developing seeds of *C. australe* and *M. truncatula*. C2.5: cotyledons weighting between 1.51 and 2.5g. C4.5: cotyledons weighting between 3.51 and 4.5g.

116

C7.5: cotyledons weighting between 5.51 and 7.5g. GC: green cotyledons. YGC: yellow-green cotyledons. BC: brown cotyledons. GA: green axes. YGA: yellow green axes. BA: brown axes. GRG: galactinol-raffinose galactosyltransferase. GSS: galactinol synthase. IMP: inositol monophosphatase family protein. RFS / SIP: raffinose synthase or seed inhibition protein. SPP: sucrose phosphate phosphatase. SPS: sucrose-phosphate synthase family protein.

Supplementary table 1. Overview of assembly, annotation and polymorphism of the *Castanospermum australe* genome

Supplementary table 2. Summary of the repeat elements of the *Castanospermum australe* genome

Supplementary table 3. Gene ontology (GO) categories enriched in expanded gene families of *Castanospermum australe.*

Supplementary table 4. List of Pfam domain enriched exclusively in expanded gene families in desiccation-sensitive seeded species.

Supplementary table 5. Phenotypic parameters associated with late seed development of *Castanospermum australe*.

Supplementary table 6. Gene ontology (GO) enrichment analysis of biological processes in relation to the acquisition of tolerance to water loss in *Medicago truncatula* and *Castanospermum australe*. Genes increasing transcript abundance in *M. truncatula* during the phase of acquisition of tolerance to water loss were compared to genes decreasing transcript abundance in *C. australe* in the stage when the cotyledons weight between 1.51 and 2.50g in relation to green pod stage. P-values of the false discovery rate (FDR) are shown together with the number of genes (#) in each GO term. Analysis was performed using AgriGO with the Arabidopsis TAIR10 background.

Supplementary table 7. Gene ontology (GO) enrichment analysis of biological processes in relation to seed maturation in *Medicago truncatula* and *Castanospermum australe*. Genes changing transcript abundance in *M. truncatula* at 28 or 32 days after pollination (DAP) in relation to dry seeds (DS) were compared to genes changing transcript abundance in *C. australe* cotyledons at the green pod stage (GC) in relation to cotyledons at the brown pod stage (BC). P-values of the false discovery rate (FDR) are shown together with the number of genes (#) in each GO term. Analysis was performed using AgriGO with the Arabidopsis TAIR10 background.

Supplementary table 8. ABA concentrations throughout *Castanospermum australe* seed development.

Supplementary table 9. List of species used in the phylome reconstruction. TaxID with asterisk (*) are the species with only transcriptomic data.

Extended data table 1. Protein-coding genes lost in all DS and retained in at least half of the DT species.

Extended data table 2. *Castanospermum australe* protein-coding genes with dN/dS (number of nonsynonymous substitutions per non-synonymous site (dN) in a given period of time divided by the number of synonymous substitutions per synonymous site (dS) in the same period) ratio ≥2. Ratio was calculated by dividing *C. australe* dN/dS by the dN/dS average of other legume species. Caus: *C. autrale*. Gmax: *Glycine max*. Gsoj: *Glycine soja*. Pvul: *Phaseolus vulgaris*. Tpra: *Trifolium pratense*.

Extended data table 3. Genes changing transcript abundance in *Castanospermum australe* and *Medicago truncatula* in comparable seed developmental stages. *C. australe* transcripts significantly changing abundance in the green pod stage compared to the stage when cotyledons weight between 1.51 and 2.50g were used. *M. truncatula* transcripts in the "desiccation tolerance module" identified by Righetti et al. (2015) or increasing abundance during the re-induction of desiccation tolerance in germinated seeds were used (Terrasson et al. 2013). dec: genes decreasing transcript abundance. inc: genes increasing transcript abundance. nc/nd: genes either not changing or not detected in the microarrays.

Extended data table 4: Genes changing transcript abundance in *Castanospermum australe* and *Medicago truncatula* in comparable seed developmental stages during final maturation. *C. australe* transcripts significantly changing abundance in the green pod stage (GC) compared to the brown pod stage (BC) were used. *M. truncatula* transcripts changing abundance in dry seeds (DS) compared to seeds 28 days after pollination (28DAP) were used (Verdier et al. 2013). dec: genes decreasing transcript abundance. inc: genes increasing transcript abundance. nc/nd: genes either not changing or not detected in the microarrays.

References (49-97)

Supplementary figure 1. Histograms of synonymous mutations (Ks) of duplicated genes (**a-f**) and average Ks of syntenic blocks (**g-l**). (**a-b**) *Castanospermum australe*, (**c-d**) *Lotus japonicus*, (**e-f**) *Medicago truncatula*, (**g-h**) *Glycine max*, (**i-j**) *Phaseolus vulgaris*, (**k-l**) *Trifolium pratense*.

Supplementary figure 2. Phylogenetic tree showing duplication rates of selected species including species of the subfamilies: Mimosoideae (inside yellow rectangle) and Caesalpinioideae (inside orange rectangle) and Papilionoideae (inside green rectangle). Duplication rates are shown in blue and the position of the whole genome duplication is marked by the red star.

Supplementary figure 3. Distribution of cluster size resulting from *Castanospermum australe*-specific duplications

Supplementary figure 4. (**a-c**) Histograms of log$_{10}$ transformed *Ks* (rate of synonymous mutations) of syntenic genes pairs identified through the alignment of (**a**) *Castanospermum australe* with itself, (**b**) *C. australe* with *Medicago truncatula,* and (**c**) *C. australe* with *Glycine max*. (**d-f**) Syntenic dotplots of comparisons of (**d**) *C. australe* with itself, (**e**) *C. australe* with *M. truncatula,* and (**f**) *C. australe* with *G. max*. Syntenic dotplots are coloured by their Ks values shown in **a-c**.

Supplementary figure 5. Hierarchical clustering of expression values of non-syntenic genes exclusively in *Castanospermum australe*. C2.5: cotyledons weighting between 1.51 and 2.5g. C4.5: cotyledons weighting between 3.51 and 4.5g. C7.5: cotyledons weighting between 5.51 and 7.5g. GC: green cotyledons. YGC: yellow-green cotyledons. BC: brown cotyledons. GA: green axes. YGA: yellow green axes. BA: brown axes. DAP: days after pollination.

Supplementary figure 6. Composition of *Castanospermum australe* mature seeds.

Supplementary figure 7. Hierarchical clustering of expression of Late embryogenesis abundant (LEA)-domain containing genes in developing seeds of *Castanospermum australe* and *Medicago truncatula*. C2.5: cotyledons weighting between 1.51 and 2.5g. C4.5: cotyledons weighting between 3.51 and 4.5g. C7.5: cotyledons weighting between 5.51 and 7.5g. GC: green cotyledons. YGC: yellow-green cotyledons. BC: brown cotyledons. GA: green axes. YGA: yellow green axes. BA: brown axes. DAP: days after pollination. ABS: pod abscission. DS: dry seeds. SMP: seed maturation proteins

Supplementary figure 8. (**a-f**) Sugar content in seeds of *Castanospermum australe* (Caus) and *Medicago truncatula* (Mtru) (**g**) Hierarchical clustering of genes involved in RFO synthesis in developing seeds of *C. australe* and *M. truncatula.* C2.5: cotyledons weighting between 1.51 and 2.5g. C4.5: cotyledons weighting between 3.51 and 4.5g. C7.5: cotyledons weighting between 5.51 and 7.5g. GC: green cotyledons. YGC: yellow-green cotyledons. BC: brown cotyledons. GA: green axes. YGA: yellow green axes. BA: brown axes. GRG: galactinol-raffinose galactosyltransferase. GSS: galactinol synthase. IMP: inositol monophosphatase family protein. RFS / SIP: raffinose synthase or seed inhibition protein. SPP: sucrose phosphate phosphatase. SPS: sucrose-phosphate synthase family protein.

Supplementary table 1. Overview of assembly, annotation and polymorphism of the *Castanospermum australe* genome

Assembly	Number	N50 (Kb)	L90 (kb)	Total length	Alignment rate
Contigs	1,210	761.1	-	-	-
Scaffolds	1,027	832.6	495	381.7 Mb	97.7%

Annotation	Number	Mean lengh (bp)	Density	Genome percentage
Protein coding genes	29,124	4814.9	-	36.7%
Exons	180,329	232.7	6.2 exons/gene	11.1 %
Introns	141,174	696.1	5.0 introns/gene	25.6%
rRNA	149	623.83	-	0.0%
snRNA	60	121.72	-	0.0%
tRNA	310	75.55	-	0.0%
Transposable elements	110,949	-	-	15.5%

Polymorphisms	Number	Density
SNPs	352,963	0.92 kb^{-1}
INDELS	328,918	0.86 kb^{-1}
Multi-allelic sites	6,328	0.02 kb^{-1}

Supplementary table 2. Summary of the repeat elements on *Castanospermum australe* genome

Repeat genus	Length	Abundant species
SINE	118,770 (0.05%)	-
LINE	3,834,244 (1.6%)	L1, RTE
DNA transposon	11,400,282 (4.8%)	hAT, PIF, CMC
LTR	42,305,056 (17.84)	Gypsy, Copia
Unclassified/Simple	177,648,468 (74.9%)	-

Supplementary table 3. Gene ontology (GO) categories enriched in expanded gene families of *Castanospermum australe*

GO category	Description	Adjusted p-value
GO:0008171	O-methyltransferase activity	1.92E-14
GO:0010333	Terpene synthase activity	1.54E-09
GO:0016705	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	7.34E-08
GO:0020037	Heme binding	3.88E-06
GO:0043531	ADP binding	1.83E-35
GO:0045735	Nutrient reservoir activity	8.03E-04
GO:0006952	Defence response	1.53E-18
GO:0008152	Metabolic process	1.56E-06
GO:0009813	Flavonoid biosynthetic process	3.82E-06
GO:0048544	Recognition of pollen	1.56E-06
GO:0055114	Oxidation reduction	2.89E-07

Supplementary table 4. List of Pfam domain enriched exclusively in expanded gene families in desiccation-sensitive seeded species

Pfam Domain	Family	Summary	C. mollissima	C. australe	E. guineensis	T. cacao
PF00080.18	Sod_Cu	Copper/zinc superoxide dismutase (SODC)			X	
PF00314.15	Thaumatin	Thaumatin family	X			X
PF00413.22	Peptidase_M10	Matrixin	X			X
PF00432.19	Prenyltrans	Prenyltransferase and squalene oxidase repeat	X			
PF00445.16	Ribonuclease_T2	Ribonuclease T2 family				X
PF00689.19	Cation_ATPase_C	Cation transporting ATPase, C-terminus	X			X
PF01471.16	PG_binding_1	Putative peptidoglycan binding domain	X			X
PF01609.19	DDE_Tnp_1	Transposase DDE domain				X
PF02992.12	Transposase_21	Transposase family tnp2				X
PF03004.12	Transposase_24	Plant tranposase (Ptta/En/Spm family)				X
PF03080.13	Neprosin	Neprosin			X	
PF03330.16	DPBB_1	Lytinc transglycolase	X			
PF03492.13	Methyltransf_7	SAM dependent carboxyl methyltranferase				X
PF03911.14	Sec61_beta	Sec61beta family				X
PF04827.12	Plant_tran	Plant transposon protein				X
PF05659.9	RPW8	Arabidopsis broad-spectrum mildew resistance protein RPW8	X			
PF07645.13	EGF_CA	Calcium-binding EGF domain	X		X	
PF08545.8	ACP_syn_III	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III		X		
PF12609.6	DUF3774	Wound-induced protein	X			
PF12776.5	Myb_DNA-bind_3	Myb/SANT-like DNA-binding domain	X			X
PF13173.4	AAA_14	AAA domain	X			X
PF13249.4	SQHop_cyclase_N	Squalene-hopene cyclase N-terminal domain	X			
PF13359.4	DDE_Tnp_4	DDE superfamily endonuclease				X
PF13923.4	zf-C3HC4_2	Zinc finger, C3HC4 type (RING finger)		X		
PF14295.4	PAN_4	PAN domain			X	
PF14299.4	PP2	Phloem protein 2				X
PF14365.4	Neprosin_AP	Neprosin activation peptide			X	
PF14380.4	WAK_assoc	Wall-associated receptor kinase C-terminal			X	
PF14560.4	Ubiquitin_2	Ubiquitin-like domain			X	
PF14576.4	SEO_N	Sieve element occlusion N-terminus		X		
PF14577.4	SEO_C	Sieve element occlusion C-terminus		X		
PF15102.4	TMEM154	TMEM154 protein family	X			

Supplementary table 5. Phenotypic parameters associated with late seed development of Castanospermum australe.

Pod stage		Green		Yellow-green		Yellow		Brown	
		average	SE	average	SE	average	SE	average	SE
Whole seed mass (g)		19.236	0.586	29.618	0.959	38.037	0.867	40.389	0.720
Seed coat	mass (g)	2.496	0.082	1.342	0.039	1.210	0.032	1.228	0.037
	water content (gH$_2$O g^{-1} dry mass)	2.042	0.265	3.065	0.146	1.001	0.055	0.652	0.045
Time (days) to reach 50% germination		25		8		8		8	
Water content (gH$_2$O g^{-1} dry mass)	axes	4.030	0.149	3.458	0.113	2.412	0.061	2.426	0.045
	cotyledons	5.084	0.209	3.255	0.231	1.559	0.073	1.783	0.122
Water content (gH$_2$O g^{-1} dry mass) at which 50% loss of viability occurs	axes	0.448	0.0208	0.342	0.0156	0.228	0.0122	0.221	0.0087
	cotyledons	0.820	0.026	0.735	0.021	0.766	0.021	0.700	0.0267
Soluble sugars (mg g^{-1} dry mass)	axes	13.750	11.798	27.865	23.210	30.622	25.932	27.983	23.786
	cotyledons	14.249	12.936	8.908	7.598	21.298	18.208	19.787	17.094
RFOs (mg g^{-1} dry mass)	axes	1.968	0.248	1.156	0.397	0.569	0.363	0.624	0.302
	cotyledons	2.262	0.767	0.241	0.220	0.376	0.281	0.330	0.216
ABA content (pmol g^{-1} dry mass)	axes	8556.94	62.50	7676.37	42.89	817.89	82.08	444.95	7.65
	cotyledons	18393.12	335.31	16153.81	160.86	4311.68	188.76	1860.64	35.90
	point of attachment	9404.77	101.66	11710.72	265.05	19400.48	181.23	24172.84	574.80
	seed coat	9175.61	330.12	32259.10	115.59	15779.45	125.60	23380.20	216.08

Supplementary table 6. Gene ontology (GO) enrichment analysis of biological processes in relation to the acquisition of tolerance to water loss in *Medicago truncatula* and *Castanospermum australe*. Genes increasing transcript abundance in *M. truncatula* during the phase of acquisition of tolerance to water loss were compared to genes decreasing transcript abundance in *C. australe* in the stage when the cotyledons weight between 1.51 and 2.50g in relation to green pod stage. P-values of the false discovery rate (FDR) are shown together with the number of genes (#) in each GO term. Analysis was performed using AgriGO with the Arabidopsis TAIR10 background.

GO Information		Mtru_increase		Caus_decrease	
GO Term	Description	FDR	#	FDR	#
GO:0009408	response to heat	3.10E-05	8	-	-
GO:0006457	protein folding	2.10E-03	6	-	-
GO:0042221	response to chemical stimulus	3.30E-03	18	5.30E-08	157
GO:0048608	reproductive structure development	0.023	10	0.015	64
GO:0042180	cellular ketone metabolic process	-	-	3.3E-12	114
GO:0006807	nitrogen compound metabolic process	-	-	8.1E-11	218
GO:0044238	primary metabolic process	-	-	8.1E-11	376
GO:0065007	biological regulation	-	-	1.5E-10	232
GO:0005975	carbohydrate metabolic process	-	-	1.7E-09	109
GO:0044106	cellular amine metabolic process	-	-	1.9E-09	62
GO:0006091	generation of precursor metabolites and energy	-	-	6.00E-09	52
GO:0006996	organelle organization	-	-	5.30E-08	97
GO:0044272	sulfur compound biosynthetic process	-	-	2.00E-07	38
GO:0006790	sulfur metabolic process	-	-	2.80E-07	46
GO:0044262	cellular carbohydrate metabolic process	-	-	4.50E-07	85
GO:0009057	macromolecule catabolic process	-	-	4.50E-07	67
GO:0032501	multicellular organismal process	-	-	7.70E-07	153
GO:0032502	developmental process	-	-	8.20E-07	155
GO:0009056	catabolic process	-	-	8.80E-07	97
GO:0048856	anatomical structure development	-	-	1.60E-06	133
GO:0065008	regulation of biological quality	-	-	9.50E-06	76
GO:0051186	cofactor metabolic process	-	-	1.50E-05	44
GO:0019684	photosynthesis, light reaction	-	-	3.80E-05	26
GO:0040007	growth	-	-	6.10E-04	46
GO:0070838	divalent metal ion transport	-	-	0.014	14
GO:0009825	multidimensional cell growth	-	-	0.017	10
GO:0048511	rhythmic process	-	-	0.034	12
GO:0023060	signal transmission	-	-	0.048	62

Supplementary table 7. Gene ontology (GO) enrichment analysis of biological processes in relation to seed maturation in *Medicago truncatula* and *Castanospermum australe*. Genes changing transcript abundance in *M. truncatula* at 28 or 32 days after pollination (DAP) in relation to dry seeds (DS) werecompared to genes changing transcript abundance in *C. australe* cotyledons at the green pod stage (GC) in relation to cotyledons at the brown pod stage (BC). P-values of the false discovery rate (FDR) are shown together with the number of genes (#) in each GO term. Analysis was performed using AgriGO with the Arabidopsis TAIR10 background.

| GO Information | | Mtru_28DAP_vs_DS / Caus_GC_vs_BC | | | | Mtru_32DAP_vs_DS / Caus_GC_vs_BC | | | |
| | | Decrease in Mtru and increase in Caus | | Increase in Mtru and decrease in Caus | | Decrease in Mtru and increase in Caus | | Increase in Mtru and decrease in Caus | |
GO Term	Description	FDR	#	FDR	#	FDR	#	FDR	#
GO:0048364	root development	7.10E-03	18	-	-	-	-	-	-
GO:0050896	response to stimulus	7.10E-03	84	0.041	76	-	-	-	-
GO:0044272	sulfur compound biosynthetic process	0.013	15	-	-	-	-	-	-
GO:0042221	response to chemical stimulus	0.014	58	-	-	-	-	-	-
GO:0005975	carbohydrate metabolic process	0.028	37	1.60E-04	44	-	-	0.019	31
GO:0044237	cellular metabolic process	0.036	127	1.60E-06	146	-	-	1.30E-05	123
GO:0042180	cellular ketone metabolic process	0.041	34	0.026	34	0.032	30	0.018	30
GO:0032502	developmental process	0.043	55	-	-	-	-	-	-
GO:0032879	regulation of localization	0.043	8	-	-	-	-	-	-
GO:0044238	primary metabolic process	-	-	8.60E-06	145	-	-	1.00E-04	121
GO:0009056	catabolic process	-	-	4.00E-04	42	-	-	1.50E-03	35
GO:0006629	lipid metabolic process	-	-	6.90E-04	36	-	-	2.20E-03	30
GO:0044248	cellular catabolic process	-	-	6.90E-04	40	-	-	2.80E-03	33
GO:0009058	biosynthetic process	-	-	0.001	102	-	-	6.40E-03	83
GO:0044255	cellular lipid metabolic process	-	-	0.0012	30	-	-	1.90E-03	26
GO:0046907	intracellular transport	-	-	0.0016	30	-	-	1.30E-03	27
GO:0051179	localization	-	-	0.0034	57	-	-	1.50E-03	51
GO:0051186	cofactor metabolic process	-	-	0.024	17	-	-	0.01	16
GO:0019637	organophosphate metabolic process	-	-	0.024	13	-	-	0.015	12
GO:0006996	organelle organization	-	-	0.026	33	0.033	29	2.90E-04	36
GO:0006508	proteolysis	-	-	0.036	19	-	-	-	-
GO:0006807	nitrogen compound metabolic process	-	-	-	-	-	-	7.10E-03	64
GO:0043412	macromolecule modification	-	-	-	-	-	-	0.012	41
GO:0065007	biological regulation	-	-	-	-	-	-	0.019	66
GO:0043687	post-translational protein modification	-	-	-	-	-	-	0.023	29

Supplementary table 8. ABA concentrations throughout *Castanospermum australe* seed development

Sample	ABA content (ng g⁻¹ dry mass)	SE
Green axis	2261.77	28.63
Green cotyledons	4861.67	153.52
Green point of attachment	2485.87	46.55
Green seed coat	2425.30	151.14
Yellow/Green axis	2029.02	19.64
Yellow/Green cotyledons	4269.78	73.64
Yellow/Green point of attachment	3095.38	121.35
Yellow/Green seed coat	8526.73	52.92
Yellow axis	216.18	37.59
Yellow cotyledons	1139.66	86.41
Yellow point of attachment	5127.93	82.97
Yellow seed coat	4170.82	57.49
Brown axis	117.61	3.49
Brown cotyledons	491.80	16.44
Brown point of attachment	6389.37	263.16
Brown seed coat	6179.85	98.93

Supplementary table 9. List of species used in the phylome reconstruction. TaxID with asterisk (*) are the species with only transcriptomic data.

Species name	Source of protein coding sequences	PhylomeID
*Acacia koa**	NCBI	110
Amborella trichopoda	Uniprot	110, 111, 112
Arabidopsis thaliana	Ensembl Plants - Release 17	110, 111, 112
Arachis duranensis	PeanutBase (http://peanutbase.org)	110
Arachis ipaensis	PeanutBase (http://peanutbase.org)	110
Cajanus cajan	GigaDB	110
Carica papaya	Plaza v3.0	111, 112
Castanea mollissima	hardwoodgenomics.org	111, 112
Castanospermum australe	Castanospermum australe genome project	110, 111, 112
*Cercis gigantean**	NCBI	110
Cicer arietinum	ICRISAT	110
Citrus sinensis	Plaza v3.0	111, 112
Coffea canephora	coffee-genome.org	111, 112
Elaeis guineensis	NCBI	111, 112
Eucalyptus grandis	Plaza v3.0	111, 112
Fragaria vesca	JGI	111,112
*Gleditsia sinensis**	NCBI	110
Glycine max	Ensembl Plants - Release 17	110
Glycine soja	NCBI	110
Gossypium raimondii	Plaza v3.0	111, 112
*Leucaena leucocephala**	NCBI	110
Lotus japonicas	Kazusa DNA Research Institute	110
Lupinus angustifolius	Lupin Genome Portal	110, 111, 112
Medicago truncatula	Plaza - v2.5	110
Musa acuminata	Plaza v3.0	111,112
Oryza sativa	Ensembl Plants - Release 22	110, 111, 112
Phaseolus vulgaris	PhasIbeAM	110, 111, 112
Phoenix dactylifera	NCBI	111, 112
Physcomitrella patens patens	Quest For Orthologs: RELEASE 2011_04	111
*Prosopis alba**	NCBI	110
Selaginella moellendorffii	Phytozome v7.0	111, 112
*Senna alexandrina sona**	NCBI	110
Solanum lycopersicum	ENSEMBL	110, 111, 112
Theobroma cacao	Plaza - v2.5	110, 111, 112
Trifolium pratense	Ensembl Plants	110
Trifolium subterraneum	NCBI	110
Vigna angularis	NCBI	110
Vigna radiata	Seoul National University	110

References:

49. Berjak, P., Farrant, J. M., Mycock, D. J. & Pammenter, N. W. Recalcitrnat (homoiohydrous) seeds: the enigma of their desiccation-sensitivity. *Seed Sci. Technol.* **18,** 297–310 (1990).

50. Sershen, Berjak, P., Pammenter, N. W. & Wesley-Smith, J. Rate of dehydration, state of subcellular organisation and nature of cryoprotection are critical factors contributing to the variable success of cryopreservation: studies on recalcitrant zygotic embryos of Haemanthus montanus. *Protoplasma* **249,** 171–186 (2012).

51. Bernatzky, R. & Tanksley, S. D. Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* **112,** 887–898 (1986).

52. Song, L., Florea, L. & Langmead, B. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol.* **15,** 509 (2014).

53. Ye, C., Ma, Z. S., Cannon, C. H., Pop, M. & Douglas, W. Y. Exploiting sparseness in de novo genome assembly. *BMC Bioinformatics* **13,** S1 (2012).

54. Ye, C., Hill, C. M., Wu, S., Ruan, J. & Ma, Z. S. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci. Rep.* **6,** (2016).

55. Boetzer, M. & Pirovano, W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15,** 211 (2014).

56. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9,** e112963 (2014).

57. English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7,** e47768 (2012).

58. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5,** R12 (2004).

59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

60. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13,** 238 (2012).

61. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1076 (2013).

62. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).

63. Stanke, M. & Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33,** 465–467 (2005).

64. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5,** 59 (2004).

65. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS: Table 1. *Bioinformatics* **32,** 767–769 (2016).

66. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12,** 491 (2011).

67. Mitchell, A. *et al.* The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **43,** D213–D221 (2015).

68. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27,** 29–34 (1999).

69. Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31,** 365–370 (2003).

70. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676 (2005).

71. Smit, A. & Hubley, R. RepeatModeler Open-1.0. 2008–2010

72. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. *Institute for Systems Biology* (2015).

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

74. Huerta-Cepas, J. & Gabaldon, T. Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics* **27,** 38–45 (2011).

75. Gascuel, O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14,** 685–695 (1997).

76. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).

77. Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28,** 3144–3146 (2012).

78. Berger, S. A., Krompass, D. & Stamatakis, A. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* **60,** 291–302 (2011).

79. Gabaldón, T. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol.* **9,** 235 (2008).

80. Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L. P., Marcet-Houben, M. & Gabaldón, T. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* **42,** D897–D902 (2014).

81. Wehe, A., Bansal, M. S., Burleigh, J. G. & Eulenstein, O. DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics* **24,** 1540–1541 (2008).

82. Al-Shahrour, F., Diaz-Uriarte, R. & Dopazo, J. FatiGO: a web tool for finding

significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **20,** 578–580 (2004).

83. Zhao, T., Holmer, R., de Bruijn, S., van der Burg, H. & Schranz, M. E. Phylogenomic synteny network analysis reveals an ancient MADS-Box transcription factor tandem duplication and lineage-specific transpositions. *bioR* (2017).

84. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40,** e49–e49 (2012).

85. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105,** 1118–1123 (2008).

86. Lyons, E., Pedersen, B., Kane, J. & Freeling, M. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Trop. Plant Biol.* **1,** 181–190 (2008).

87. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21,** 487–493 (2011).

88. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53,** 661–673 (2008).

89. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).

90. Bove, J. *et al.* Gene expression analysis by cDNA-AFLP highlights a set of new signaling networks and translational control during seed dormancy breaking in Nicotiana plumbaginifolia. *Plant Mol. Biol.* **57,** 593–612 (2005).

91. Maere, S., Heymans, K. & Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21,** 3448–3449 (2005).

92. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6,** (2011).

93. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39,** W29–W37 (2011).

94. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7,** e1002195 (2011).

95. Rosnoblet, C. *et al.* The regulatory gamma subunit SNF4b of the sucrose non-fermenting-related kinase complex is involved in longevity and stachyose accumulation during maturation of Medicago truncatula seeds. *Plant J.* **51,** 47–59 (2007).

96. Floková, K. *et al.* UHPLC-MS/MS based target profiling of stress-induced phytohormones. *Phytochemistry* **105,** 147–157 (2014).

97. Jiang, W.-B. *et al.* Brassinosteroid Regulates Seed Size and Shape in Arabidopsis. *PLANT Physiol.* **162,** 1965–1977 (2013).

Induction of desiccation tolerance in sensitive *Citrus limon* seeds

Alexandre Marques, Harm Nijveen, Charles Somi,
Wilco Ligterink and Henk Hilhorst

Abstract

Seeds which can be dried and stored for a long time have been described as orthodox. In contrast, recalcitrant seeds have a very limited storage life and die upon drying below a critical moisture content. As a result, the development of long term storage methods for recalcitrant seeds is complicated, yet very valuable. This because many important perennial crops bear recalcitrant seeds, including the three most widely used beverage crops (tea, coffee and cocoa), many tropical fruits (e.g. mango, durian, citrus) and three major industrial crops (rubber, oil palm and coconut). Dried seed storage is the safest, most convenient and cheapest method for conserving plant genetic resources. Therefore, we have studied the induction of desiccation tolerance in the desiccation sensitive seeds of *Citrus limon*. We show that desiccation tolerance can be induced in *Citrus limon* seeds by paclobutrazol treatment and we studied its associated transcriptome to delineate the molecular mechanism underlying it. Paclobutrazol not only interfered with gibberellin related gene expression but also caused extensive changes in expression of genes involved in the biosynthesis and signalling of other hormones. As a result of the paclobutrazol treatment, we observed a transcriptomic switch encompassing the suppression of biotic- and induction of abiotic responses. We hypothesize that this is the main driver of the induction of desiccation tolerance by paclobutrazol in *Citrus limon* seeds.

Introduction

Desiccation tolerance refers to the capacity of an organism to dehydrate to below 10% water on a fresh weight basis (or 0.1g H_2O/g dry weight) without accumulation of lethal damage (Bewley 1979). This mechanism was essential for the colonization of terrestrial habitats by the early land plants. Vegetative is very common in basal plants and is mostly confined to seeds of spermatophytes (Oliver, Tuba and Mishler 2000). However, seed desiccation tolerance can be lost in plants adapted to environments where the conditions are conducive to immediate germination (Baskin and Baskin 2014).

Of the spermatophytes worldwide, 8% display seed desiccation sensitivity (Tweddle, Turner and Dickie 2002). This number can reach up to 50% in tropical evergreen rainforests (Tweddle et al. 2003). The lack of desiccation tolerance places these species on the list of high risk of extinction, as their seeds cannot be stored in seed banks. This severely hampers the long term conservation of their genetic resources.

Although seeds of many citrus cultivars can be dehydrated safely to 10-12% moisture content, further dehydration impairs seed viability (Wen et al. 2010). However, there is considerable variability among seed lowest safe moisture content (LSMC) when comparing the different studies on desiccation tolerance in *C. limon* seeds. This has placed its seeds in different storage categories: from orthodox and intermediate to recalcitrant (Khan, Thompson and Muhammad Usman Fatima 2002, Hong, Ahmad and Murdoch 2001, King, Soetisna and Roberts 1981, Farnsworth 2000, Black and Pritchard 2002, Khan et al. 2003).

The phytohormones abscisic acid (ABA) and gibberellins (GAs) are associated with many physiological processes and often work antagonistically. During seed maturation, ABA is responsible for the acquisition of desiccation tolerance in orthodox seeds (Gutierrez et al. 2007). Low levels of ABA or disruption of its signalling pathway may result in desiccation sensitivity (Ooms et al. 1993). GA is associated with the breaking of dormancy and induction of germination (Hilhorst and Karssen 1992). Furthermore, high levels of GA have been associated with vivipary and, consequently, desiccation sensitivity (White et al. 2000).

The acquisition of desiccation tolerance in orthodox seeds or lack of it in desiccation sensitive seeds is assumed to be a consequence of a changed balance between these phytohormones. The possibility of induction of desiccation tolerance in a desiccation sensitive (DS) seed has been demonstrated in *Acer saccharinum* seeds (Beardmore and Whittle 2005). A combination of ABA and tetcyclacis, an inhibitor of ABA catabolism and GA biosynthesis, (Rademacher et al. 1987), induced desiccation tolerance in these seeds.

Here we report, for the first time, the induction of desiccation tolerance in the DS seeds of a crop by treatment with paclobutrazol (PAC), an inhibitor of GA-biosynthesis.

The induction of desiccation tolerance in *C. limon* seeds and the analysis of the associated transcriptional changes sheds light on a potential mechanism of desiccation tolerance acquisition in otherwise DS seeds. Furthermore, the possibility of induction of desiccation tolerance in seeds may support breeding, germplasm storage and commercialization of *C. limon*. This study may provide useful insights for the development of protocols to induce desiccation tolerance in other DS seeded species.

Materials and Methods

Induction of desiccation tolerance

Citrus limon (variety Primofiore, cultivated in Spain) seeds were collected from ripe fruits, rinsed with water and sterilized in 10% (v/v) bleach for 10 minutes. For the induction of desiccation tolerance, seeds were incubated for 7 days in MS medium containing 1% of agar, 5% of sucrose and 10 mM (2RS,3RS)-1-(4-chlorophenyl)-4,4-dimethyl-2-(1,2,4-triazol-1-yl)pentan-3-ol (paclobutrazol) (Sigma Aldrich).

A preliminary experiment was performed that used the same protocol as described above but using abscisic acid (ABA) (Sigma Aldrich) instead of paclobutrazol. We used different concentrations of ABA (5, 50 and 500 µM) but they failed to arrest germination and induce desiccation tolerance in *C. limon* seeds. All the following experiments performed in this study were based on paclobutrazol treatment.

All drying treatments were performed by placing the seeds in a cabinet for 7 days at 30% relative humidity, with forced ventilation, at 22 °C in the dark.

Seed survival was assessed by placing the desiccated seeds on germination paper soaked with 50 mL of a 100 µM GA solution (GA$_{4+7}$; Berelex, ICI), at 30°C in the dark. Germination and fully developed seedlings were scored after one month.

Total RNA isolation

Total RNA was extracted from non-desiccated seeds after 7 days in paclobutrazol or water (control) according to the hot borate protocol modified from Wan and Wilkins (1994). Three replicates of 25 seeds each for each treatment were homogenized and mixed with 800 µL of extraction buffer (0.2 M Na borate decahydrate (Borax), 30 mM EGTA, 1% SDS, 1% sodium deoxy-cholate (Na-DOC)) containing 1.6 mg DTT and 48 mg PVP40 which had been heated to 80°C. 1 mg proteinase K was added to this suspension and incubated for 15 min at 42°C. After adding 64 µL of 2 M KCL, the samples were incubated on ice for 30 min and subsequently centrifuged for 20 min at 12,000 g. 270 µL of ice-cold 8 M LiCl was added to the supernatant in a final concentration of 2 M and the tubes were incubated overnight on ice. After centrifugation

for 20 min at 12,000 g at 4 °C, the pellets were washed with 750 µL ice-cold 2 M LiCl. The samples were centrifuged for 10 min at 10,000 g at 4 °C and the pellets were re-suspended in 100 µL DEPC treated water. The samples were cleaned with phenol chloroform and treated with DNAse (RQ1 DNase, Promega) according to the manufacturers protocol. The RNA quality and concentration were assessed by agarose gel electrophoresis and UV spectrophotometry.

RNA-seq analysis, assembly and analysis of differential expression

Messenger RNA (mRNA) was isolated from total RNA using mRNA enrichment polyA capture technology (Illumina Incorporated, San Diego, CA, USA) followed by library preparation using random hexamer priming. RNA-seq was performed strand specific with paired-end 125nt reads on an Illumina Hiseq 2500. The reads were quality trimmed using Trimmomatic (Bolger, Lohse and Usadel 2014). The transcriptome was *de novo* assembled using Trinity version 2.1.1 (Grabherr et al. 2011). Blastx (Altschul et al. 1997) with an E-value cut-off of $1e^{-5}$ was used to find homologous Arabidopsis genes (TAIR10 (Lamesch et al. 2011)) for the assembled genes. Expression was quantified using the Kallisto software tool with the *de novo* assembled transcriptome (Bray et al. 2016) and then analysed using Sleuth (Pimentel et al. 2017). All subsequent analyses were performed based on the *de novo* assembly.

The RNA-seq assembly is available at the National Centre for Biotechnology Information (NCBI) as Bioproject ID SRP111791. Genome assembly, annotation data and all the supplementary figures, tables, data are available for viewing and downloading at http://www.wageningenseedlab.nl/thesis/amarques/SI/chapter5

Gene enrichment analysis

Gene Ontology enrichment was performed on significantly up- and down-regulated gene sets using *Arabidopsis thaliana* orthologs and its genome as reference. We used AgriGo and Revigo to summarize the results (Du et al. 2010, Supek et al. 2011). KEGG enrichment pathway analysis was done with DAVID (Dennis et al. 2003).

Results

Induction of desiccation tolerance in desiccation sensitive seeds of Citrus limon

Seeds of different Citrus species vary widely in their level of desiccation tolerance (Khan et al. 2003). *C. limon* seeds have been considered orthodox, intermediate or recalcitrant by different studies (Khan et al. 2002). In this study, we observed that the *Citrus limon (L.) Burm. f.,* cultivar Primofiori, seeds have intermediate storage behaviour as they show severe loss of viability when dried to 10% water content on a fresh weight basis (see Chapter 2). According to this classification, recalcitrant seeds lose viability when dried below ~25% water content on a fresh weight basis.

With the aim to induce desiccation tolerance in *C. limon* seeds, we first placed the seeds on ABA solutions at increasing concentrations (5, 50 and 500 µM). Even on the highest concentration of ABA, high percentages of germination (>90%) were observed and thus failed to induce desiccation tolerance in these seeds. Subsequently, we tested manipulation of GA content. Treatment of whole seeds with PAC, an inhibitor of gibberellin biosynthesis, for 7 days, restored germination to 80% after desiccation, while non-treated control seeds showed less than 1% of survival (Figure 1).

These results imply that a treatment with PAC prevented seeds from germinating and protected them against desiccation damage. PAC-treated dried seeds even displayed a high percentage of germination after 6 months of dry storage. However, after 9 months of dry storage, the survival of PAC-treated seeds had dropped sharply (Table 1).
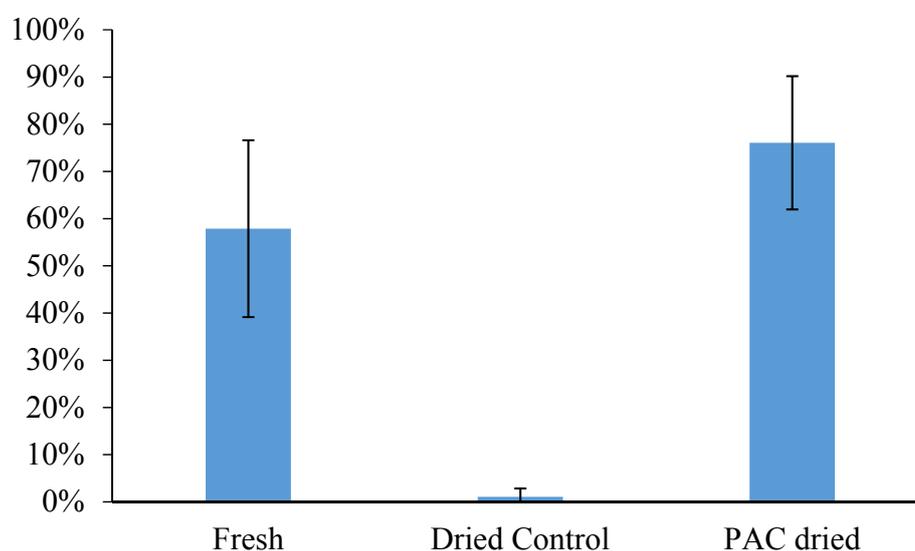

Figure 1. Total germination of *C. limon* seeds. 'Fresh' germination of seeds immediately after removal from the fruits; germination of 'Dried Control', seeds after desiccation to 10% water content; 'PAC Dried', seeds treated by PAC and dried subsequently to 10% water content. Bars represent the average of four replicates of 25 seeds each. Error bar represents the standard deviation.

Table 1. Survival assessment of desiccated PAC-treated seeds of *C. limon* after storage

Months of dry storage	Survival (%)	Water content (% fresh weight)
0	83 $\pm$ 9	10 $\pm$ 0.8
6	69 $\pm$ 13	7.9 $\pm$ 0.5
9	5 $\pm$ 3	7.1 $\pm$ 0.3

Transcriptomic changes triggered by PAC

Next we performed a transcriptome analysis on non-desiccated seeds incubated for 7 days in PAC and in water (control). We used RNA-seq to identify the genes of which the expression had changed in desiccation tolerant PAC-treated seeds in comparison to non-treated desiccation sensitive fresh seeds, using three biological replicates per condition. Since there is no genome sequence available for *C. limon*, we executed a *de novo* assembly of the RNA-seq reads using the Trinity software (Haas et al. 2013). This resulted in 103,841 predicted 'genes' and 179,588 transcripts (Supplementary tables 1 and 2). Using blastx we found hits with 16,003 Arabidopsis genes for 45,425 *C. limon* assembled transcripts.

Seed samples treated with PAC showed significant (fold change > 2 or < -2, FDR < 0.05) up-regulation of 613 genes and down-regulation of 950 genes in comparison to the water control. Based on the GO enrichment analysis of the up-regulated set, we found 8 biological process (BP) categories significantly enriched (FDR < 0.05), as well as 31 molecular function (MF) and 35 cellular component (CC) categories. The down-regulated set showed 98 BP, 20 MF and 17 CC categories (Figure 2). For a complete list of differentially regulated genes, their KEGG and GO enriched categories, see Supplementary file 1.

The down-regulated set contained more genes and was more diverse, as evidenced by the much larger number of GO categories. This likely represents the general growth inhibitory action of the PAC treatment. Apparently, the suppression of GA-biosynthesis has multiple effects across diverse pathways.

Furthermore, a cellular compartmentalization switch was observed in the cellular component GO enrichment analysis. The mitochondrial transcriptome appeared largely suppressed whereas plastid genes were induced. The reduction of mitochondrial gene expression likely reflects the suppression of energy metabolism and preparation for quiescence. The up-regulation of genes located in the plastids is not photosynthesis related as these seeds are not green. Thus, this feature may rather be associated with reserve remobilization, particularly that of lipids. We detected the up-regulation of

plastid located genes such as *FATTY ACID DESATURASE 6* (*FAD6*), *ACETYL CO-ENZYME A CARBOXYLASE BIOTIN CARBOXYLASE SUBUNIT* (*CAC2*), *ASPARTATE AMINOTRANSFERASE 3* (*ASP3*), *CYTIDINEDIPHOSPHATE DIACYLGLYCEROL SYNTHASE 4* (*CDS4*), *LIPOYL SYNTHASE 1* (*LIP1*) and *PLASTID LIPASE1* (*PLIP1*).

Furthermore, the heat shock transcription factors *HSFA2*, *HSFA4A* and *HSFA6B* were up-regulated. Additionally, we found up-regulation of the genes encoding the heat shock proteins *HEAT SHOCK COGNATE PROTEIN 70-1* (*HSC70-1*), *HEAT SHOCK PROTEIN 60* (*HSP60*) and *HEAT SHOCK PROTEIN 101* (*HSP101*). We did not observe an extensive up-regulation of genes related to protective mechanisms against desiccation, besides the antioxidant enzyme *CATALASE 2* (*CAT2*). There were 43 LEAs identified in the transcriptome of both control and PAC treated seeds. However, there was no significant up-regulation of LEA transcripts, commonly associated with the acquisition of desiccation tolerance (Hundertmark and Hincha 2008). We identified only three LEAs, belonging to Group 2 (Hundertmark and Hincha 2008), which were down-regulated upon PAC treatment.

PAC promotes abiotic- and suppresses biotic responses in C. limon seeds

The up-regulated genes were enriched with the GO categories *response to abiotic stress* and *response to radiation* (Figure 2A). These categories contain genes responsible for DNA repair and protection. DNA damage is one of the main problems related to desiccation sensitivity, as was demonstrated in various tolerant and intolerant seeds and other organisms (Boubriak et al. 2000, Boubriak et al. 1997, Faria et al. 2005, Mattimore and Battista 1996).
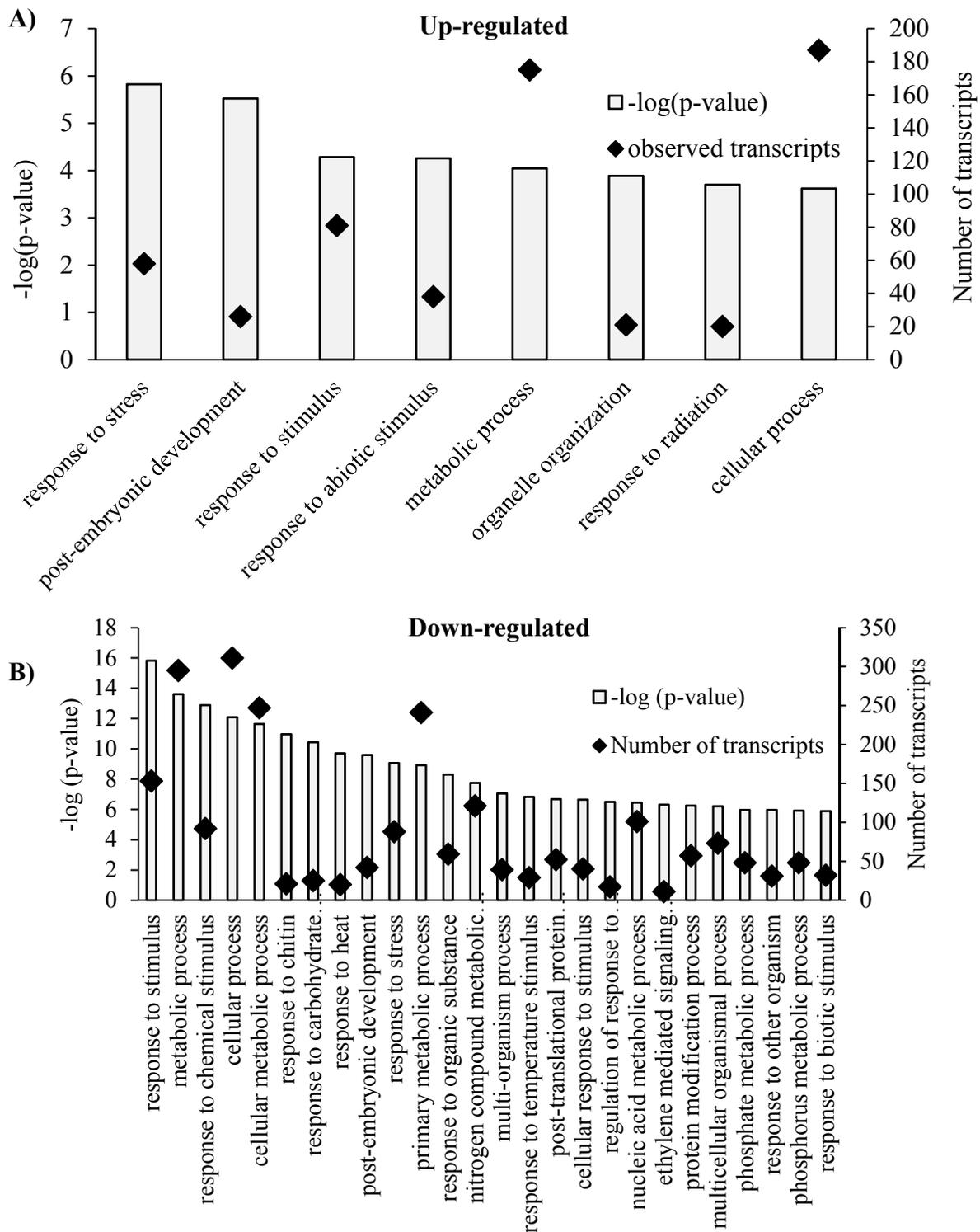
Figure 2. Over-representation analysis of the differentially expressed genes after PAC treatment. The gene set analysed was first filtered through a fold-change and variance cut-off based filter (fold change >2 or <-2 and FDR < 0.05). A) The biological process GOs enriched in the up-regulated set. B) The top 26 biological process GOs in the down-regulated set based on p-value. The total list of GO enrichment analysis can be found in Supplementary file 1.

In addition, in the up-regulated gene set we found the abiotic stress-associated genes *CATALASE 2* (*CAT2*), *EARLY RESPONSIVE TO DEHYDRATION 15* (*ERD15*), *HEAT SHOCK PROTEIN 101* (*HSP101*) and *HEAT SHOCK TRANSCRIPTION FACTOR A2* (*HSFA2*).

Genes related to biotic hormonal pathways, such as jasmonic acid (*ATJAZ1, JAZ2* and *JAZ3*) and salicylic acid (*TGA10*) were down-regulated (Figure 3). Moreover, there was an extensive down-regulation of genes involved in biotic responses, such as those in the categories *response to chitin* and *response to biotic stimulus* (Figure 2B).

Hormonal interactions underlying responses to PAC treatment

PAC inhibits the monooxygenases involved in the oxidation of *ent*-kaurene to *ent*-kaurenoic acid and therefore reduces the plant's ability to synthesize active GAs (Rademacher 1989). GAs are an absolute requirement for germination and inhibition of *ent*-kaurene oxidase by PAC prevents germination (Jacobsen and Olszewski 1993).

Seed germination is regulated by multiple hormones, such as GA, ABA, ethylene, and brassinosteroids (Kucera, Cohn and Leubner-Metzger 2005). The GA-responsive genes include those responsible for synthesis, transport, and signalling of other hormones, suggesting a crosstalk between GA and other hormones (Ogawa 2003). Here multiple hormonal pathways were down-regulated by the PAC treatment of *C. limon* seeds (Figure 3).

We performed a KEGG enrichment analysis of the down-regulated gene set and found a significant enrichment of the category *plant hormone signal transduction* (Supplementary file 1). Several hormone signalling pathways were down regulated. Among them were genes involved in the signalling of auxin, gibberellin, ethylene, brassinosteroids, jasmonic and salicylic acid (Figure 3). The down-regulated genes involved in auxin signalling were *AUXIN RESISTANT 1* (*AUX1*), *INDOLEACETIC ACID-INDUCED PROTEIN 16* (*IAA16*), *AUXIN-RESPONSIVE GH3 FAMILY PROTEIN (GH3.1), SMALL AUXIN UPREGULATED RNA 32* (*SAUR32*), in gibberellin signalling *GIBBERELLIN 2-BETA-DIOXYGENASE 2* (*GA2OX2*), *GA INSENSITIVE DWARF1B* (*GID1B*) and *GA INSENSITIVE DWARF2* (*GID2*), in ethylene signalling *CONSTITUTIVE TRIPLE RESPONSE 1* (*CTR1*), *ETHYLENE RESPONSE 2* (*ETR2*) and *EIN3-BINDING F BOX PROTEIN 1* (*EBF1*), in brassinosteroid signalling *TOUCH 4* (*TCH4*), in jasmonic acid signalling *JASMONATE-ZIM-DOMAIN PROTEIN 1,2* and *3* (*JAZ1, JAZ2* and *JAZ3*) and in salicylic acid signalling *TGACG (TGA) MOTIF-BINDING PROTEIN 10* (*TGA10*).

Conversely, we observed the up-regulation of transcripts involved in ABA signalling, such as *MOTHER OF FT AND TFL1* (*MFT*), *MEDIATOR OF ABA-REGULATED DORMANCY 1* (*MARD1*) and *G-BOX BINDING FACTOR 3* (*GBF3*). Interestingly these genes are also involved in the control of seed maturation and germination (Belmonte et al. 2013, Arana et al. 2014, Zhang et al. 2011).
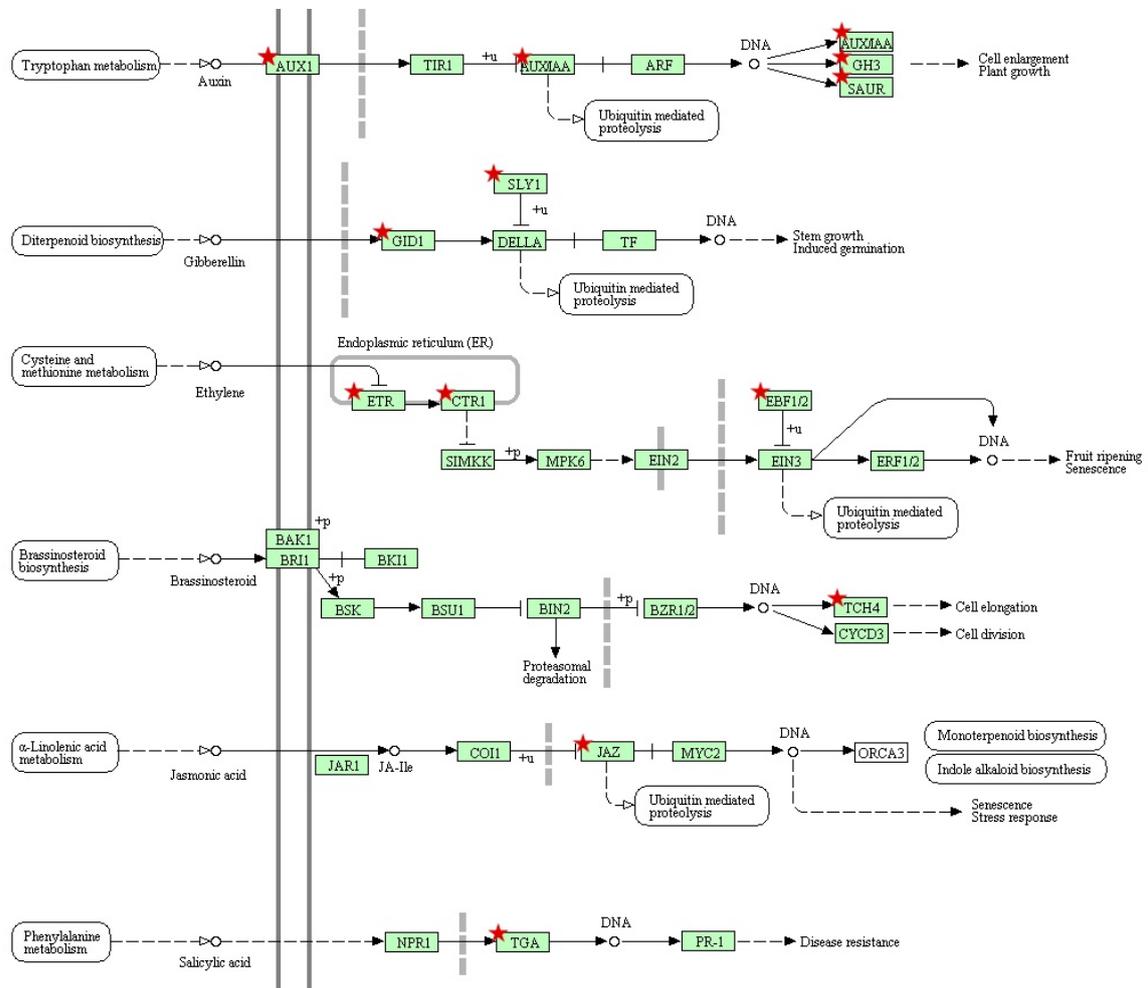
Figure 3. KEGG enrichment analysis (DAVID) of the down-regulated gene set (fold change < -2 and FDR < 0.05). The model depicts the signalling pathway of multiple plant hormones. The red stars indicate down-regulated genes upon PAC treatment of *C. limon* seeds.

Signalling pathways involved in the control of germination and seed desiccation sensitivity in C. limon seeds

The lack of developmental arrest has been associated with seed desiccation sensitivity. Thus, we took a closer look at the differentially regulated genes involved in the promotion or inhibition of germination and, possibly, with acquisition of desiccation tolerance in *C. limon* seeds upon PAC treatment.

In addition to *MFT*, *MARDI1* and *GBF3*, there was a down-regulation of transcripts involved in the stimulation of seed germination through the GA pathway, *PhyE* and *SCL3* (Arana et al. 2014, Zhang et al. 2011). *REVEILLE 1* (*RVE1*) is a promoter of primary seed dormancy in Arabidopsis; its overexpression complements the non-dormant phenotype of the *delay of germination 1* (*dog1*) mutant. It was up-regulated in PAC treated *C. limon* seeds. RVE1 has also been shown to suppress GA-biosynthesis by binding directly to the promoter of *GIBBERELLIN-3-OXIDASE 2* (*GA3ox2*) (Jiang et al. 2016).

Furthermore, we performed an enrichment analysis to find consensus transcription factor (TF) binding sites in the set of differentially regulated genes upon treatment with PAC (Jin et al. 2017). The up-regulated genes showed significant motif enrichment ($p < 0.001$) for 30 TFs (Table 2). Of these, three TFs were indeed identified in the up-regulated gene set, i.e., *RVE1, RVE7* and *GBF3*.

All relevant transcriptional changes identified in the present study involving hormonal signalling and germination control are shown in Figure 4.
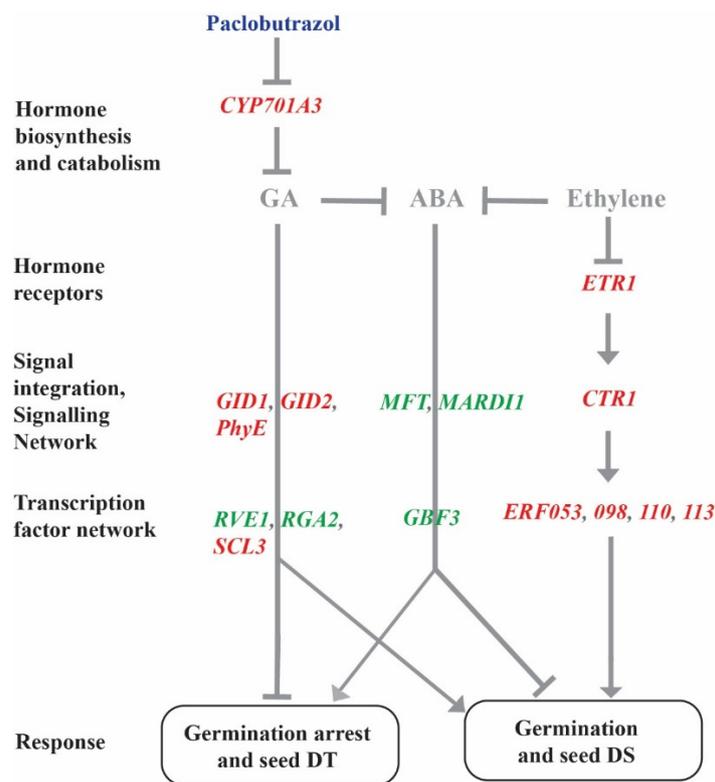
Figure 4. Schematic representation of the interactions between the gibberellin (GA), abscisic acid (ABA) and ethylene signalling pathways in the regulation of the arrest of germination and DS in seeds of *C. limon* treated with paclobutrazol (PAC). Up-regulated genes are marked green and down-regulated red. The model is mainly based on Arabidopsis hormone mutant analyses adapted from Kucera et al. (2005). The positions of some components are speculative and details are discussed in the text.

Table 2. Transcription factors that have over-represented targets on the PAC-treated up-regulated genes. In bold are the transcription factors that were identified in the RNA-seq analysis as up-regulated.

TF	Gene ID	p-value	Query	Background
AT5G17300	**REVEILLE 1 (RVE1)**	**7.45E-07**	**31**	**711**
AT3G09600	REVEILLE 8 (RVE8)	2.77E-06	30	720
AT2G01930	BASIC PENTACYSTEINE1 (BPC1)	3.17E-06	195	8742
AT3G23210	BASIC HELIX-LOOP-HELIX 34 (bHLH34)	3.54E-06	56	1779
AT4G38910	BASIC PENTACYSTEINE 5 (BPC5)	4.23E-06	169	7371
AT1G18330	**REVEILLE 7 (RVE7)**	**1.37E-05**	**28**	**705**
AT3G10113	Homeodomain-like superfamily protein	1.65E-05	28	712
AT5G52660	REVEILLE 6 (RVE6)	3.57E-05	28	743
AT3G10800	(BZIP28)	7.39E-05	54	1888
AT2G36270	ABA INSENSITIVE 5 (ABI5)	9.37E-05	56	1997
AT5G42520	BASIC PENTACYSTEINE 6 (BPC6)	1.41E-04	160	7334
AT2G46270	**G-BOX BINDING FACTOR 3 (GBF3)**	**1.79E-04**	**51**	**1816**
AT3G56850	ABA-RESPONSIVE ELEMENT BINDING PROTEIN 3 (AREB3)	1.82E-04	55	2002
AT1G32150	TRANSCRIPTION FACTOR 68 (bZIP68)	1.93E-04	54	1960
AT1G45249	ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR 2 (ABF2)	1.99E-04	55	2009
AT4G01280	REVEILLE 5 (RVE5)	2.11E-04	23	625
AT3G20840	PLETHORA 1 (PLT1)	2.12E-04	17	402
AT5G08130	(BIM1)	2.62E-04	30	917
AT4G00250	(ATSTKL2)	2.85E-04	17	412
AT2G35530	BASIC REGION/LEUCINE ZIPPER TRANSCRIPTION FACTOR 16 (bZIP16)	2.86E-04	53	1945
AT5G61270	PHYTOCHROME-INTERACTING FACTOR7 (PIF7)	3.39E-04	41	1409
AT5G02840	LHY/CCA1-LIKE 1 (LCL1)	5.31E-04	21	587
AT1G10120	CRY2-INTERACTING BHLH 4 (CIB4)	5.48E-04	35	1175
AT5G67450	ZINC-FINGER PROTEIN 1 (ZF1)	5.95E-04	37	1269
AT5G61620	myb-like transcription factor family protein	7.41E-04	17	447

Discussion

Induction of desiccation tolerance in desiccation sensitive seeds of Citrus limon

The storage of desiccated seeds is important to ensure long term conservation of a broad genetic diversity. In the present study, *C. limon* seeds showed DS phenotype and, hence, cannot be stored for prolonged periods.

It is challenging to determine the cause of the wide variation of seed desiccation sensitivity across the different *C. limon* studies. Here we studied the most sensitive phenotype for *C. limon* seed desiccation sensitivity ever described, with seeds completely losing viability at 10% water content. Possible sources of this variation can be different agronomical practices, adaptation to different environments or varying genetic backgrounds. One of the main breeding targets for citrus species is to lower seed numbers and produce seedless fruits (Varoquaux et al. 2000). This breeding throughout the long *Citrus* domestication history has contributed to the transition to apomixis and reduction of seed numbers (Wang et al. 2017). These domestication events towards loss of seeds possibly contributed to the reduction of their quality, influencing traits such as longevity and/or desiccation tolerance. Citrus species originate in southeast Asia and have been cultivated for thousands of years (Velasco and Licciardello 2014).

Variability in seed dehydration tolerance is also apparent in some other genera, e.g. *Araucaria*, *Dipterocarpus* and *Coffea* (Tompsett 1984, Tompsett 1987, Ellis, Hong and Roberts 1991). This interspecific dehydration tolerance variability may be the result of adaptation to different environments. However, it is also possible that both the adaptation to different environments and breeding contributed to the seed desiccation tolerance variability observed in Citrus.

Here we report, for the first time, induction of desiccation tolerance in a crop with DS seeds. The arrest of germination by inhibiting gibberellin synthesis in recalcitrant seeds was first described in *Hopea odorata* (Garello and Page-Degivry 1995). However, the induction of desiccation tolerance in a DS seed has so far only been described for *Acer saccharinum* seeds (Beardmore and Whittle 2005). Similar to *C. limon*, these seeds are also insensitive to ABA. Application of ABA alone resulted in 0% survival, but ABA combined with tetcyclacis, which inhibited both ABA catabolism and GA synthesis (Rademacher et al. 1987), induced desiccation tolerance in *Acer saccharinum* seeds. We successfully induced desiccation tolerance in *C. limon* seeds with the GA biosynthesis inhibitor PAC.

Acer saccharinum seeds displayed loss of viability when desiccated to water contents below 30%, characterizing recalcitrant behaviour (Becwar, Stanwood and Leonhardt 1983). In this study, *C. limon* seeds lost viability when dried to 10% fresh weight water content. Although we did not evaluate the water content when these seeds

started to lose viability, other studies showed survival of *C. limon* seeds to lower water contents, but always displaying some degree of desiccation sensitivity (Hong et al. 2001, Soetisna, King and Roberts 1985, King et al. 1981). Altogether, it suggests that *C. limon* seeds have an intermediate seed storage behaviour with a variable LSMC. Based on this, we expect that *A. saccharinum* seeds are more sensitive to desiccation than *C. limon* and, thus, induction of desiccation tolerance in *A. saccharinum* is likely to be more complex. The seeds of *A. saccharinum* required a treatment which inhibited both GA-biosynthesis and ABA catabolism while *C. limon* seeds only required the inhibition of GA-biosynthesis to acquire DT.

Overall transcriptomic changes triggered by pacobutrazol

GAs and ABA are important plant growth regulators, acting antagonistically to control many plant developmental processes, including root and stem elongation, rosette expansion, floral induction, anther development and seed germination (Yamaguchi 2008). GA levels are normally decreased and those of ABA increased at later stages of orthodox seed maturation and abnormal high levels of GA can cause precocious germination (White et al. 2000).

Current consensus suggests that seed desiccation sensitivity occurs when seeds fail to undergo the final stages of maturation during seed development (Farrant and Moore 2011). During maturation, a number of key events occur, including the synthesis and deposition of stored reserves (storage proteins, carbohydrates and lipids) and LEA proteins (Verdier et al. 2013).

Beardmore and Whittle (2005) suggested that, by treatment with ABA, desiccation sensitive embryonic axes of *A. saccharinum* probably underwent a process of seed maturation as in orthodox seeds. They associated the induction of desiccation tolerance with the accumulation of storage proteins and dehydrins.

However, upon PAC treatment of *C. limon* seeds, we did not observe significant up-regulation of maturation and desiccation related genes, such as those encoding for seed storage proteins and LEAs, except for HSPs and catalase. Three LEAs, belonging to group 2, were down-regulated upon PAC treatment. Furthermore, this group of LEAs has not been associated with the acquisition of seed desiccation tolerance and its classification as a group of LEA proteins is controversial (Hundertmark and Hincha 2008, Chatelain et al. 2012). It is possible that PAC induced desiccation tolerance in *C. limon* seeds by the regulation of a few essential protective elements whereas most of these were already present in control seeds, e.g. genes encoding 43 LEAs were present in both treatments without significant changes in expression. Furthermore, we observed that the transcriptomes of both control and PAC treated *C. limon* seeds resembled that

of Arabidopsis seeds imbibed for 24h in ABA, which have not germinated yet and are still desiccation tolerant (The Bio-Analytic Resource for Plant Biology 2018).

During the development of orthodox (desiccation tolerant) seeds the reduction of metabolism, particularly the reduction of respiration rate, coincides with their survival after desiccation (Leprince, Buitink and Hoekstra 1999, Pammenter and Berjak 1999). We assume that high GA and low ABA levels in DS seeds are responsible for the maintenance of active metabolism and the skipping of maturation drying. Thus, the inhibition of GA synthesis in *C. limon* seeds counteracted these processes and thus arrested the germination program, suppressed metabolic activity and allowed successful desiccation.

Does PAC treatment rescue C. limon seed desiccation tolerance by up-regulating abiotic and suppressing biotic responsive genes?

We found down-regulation of genes involved in hormonal pathways of biotic stress responses, such as jasmonic acid (*ATJAZ1, JAZ2* and *JAZ3*) and salicylic acid (*TGA10*). We also observed the GO enrichment of biotic response categories among the down-regulated genes set, such as *response to chitin* and *response to biotic stimulus*.

Chitin has been demonstrated to be an elicitor of fungal infection responses in several plants, including Citrus (Gallão et al. 2007, Eckardt 2008, Boller and Felix 2009). Furthermore, the down-regulation of biotic responses, including *response to chitin*, was also demonstrated for the re-establishment of desiccation tolerance in germinated seeds of *Arabidopsis thaliana* upon ABA treatment (Costa et al. 2015).

The protection towards biotic agents is important for imbibed seeds, especially for those seeds shed in ripened fleshy fruits that are more prone to fungal attack. Most recalcitrant seeds in the tropical rainforests are embedded in fleshy fruits (Chin 1977). Chitin triggers lignin deposition as a strategic response to combat fungal infection (El Hadrami et al. 2010). At the same time, lignin hardens the cell walls while desiccation tolerance requires flexible and foldable cells (Moore et al. 2008). It has been demonstrated that recalcitrant seeds show cell wall damage when desiccated (Chandel et al. 1995, Moore et al. 2008). We hypothesize that the down-regulation of biotic responses is not only an isolated consequence of the metabolic arrest caused by PAC but an essential response that prevents the lignification of the cell walls of *C. limon* seeds and thus make these flexible, allowing their folding and preventing damage upon seed desiccation.

The inverse relation between the regulation of biotic versus abiotic stresses and associated hormone pathways, as observed in the present work, has also been found in other studies where, for example, ABA contents are inversely correlated with

susceptibility to pathogens in diverse plant species (Mauch-Mani and Mauch 2005). We assume that the down-regulation of biotic responses and the up-regulation of abiotic responses in PAC treated *C. limon* seeds that leads to desiccation tolerance are interconnected by common regulatory pathways.

We detected the up-regulation of the heat shock transcription factors *HSFA2*, *HSFA4A* and *HSFA6B* that are known to regulate HSPs. Their chaperonin protective role has been shown to provide vegetative abiotic stress tolerance, such as tolerance to heat, salt and oxidative stress in transgenic Arabidopsis studies (Pérez-Salamó et al. 2014, Chauhan et al. 2013). HSPs and heat shock regulated transcription factors are an interesting class of up-regulated genes as they have been associated with diverse abiotic stresses (Wang et al. 2004). It has been suggested that HSPs act as chaperonins to protect other proteins under various stresses (Wang et al. 2004).

Small heat shock proteins (*sHSPs*) have been associated with the acquisition of desiccation tolerance (Wehmeyer and Vierling 2000, Wehmeyer et al. 1996). However, we did not find the up-regulation of sHSPs but only of HSPs: *HEAT SHOCK COGNATE PROTEIN 70-1* (*HSC70-1*), *HEAT SHOCK PROTEIN 60* (*HSP60*) and *HEAT SHOCK PROTEIN 101* (*HSP101*). *HSC70-1* regulates development and abiotic stress responses in *A. thaliana* (Leng et al. 2017), *HSP60* has been implicated in heat shock responses in Arabidopsis (Rikhvanov et al. 2007) and *HSP101* is the most abundant heat stress protein in seeds. It was not detectable in the *A. thaliana* DS-seeded mutant, *abi3-6* (Kotak et al. 2007).

Transcriptional changes associated with hormonal balance and germination control

We expected that the inhibition of gibberellin biosynthesis by PAC, which inhibits the enzyme *ent*-kaurene oxidase (Dalziel and Lawrence 1984), would down-regulate the downstream genes involved in gibberellin biosynthesis and signalling. Indeed, we observed down-regulation of *GA2OX2,* the gene encoding for the rate limiting enzyme GA2 oxidase in gibberellin biosynthesis, as well as members of the GID GA-receptor family of genes *GID1B* and *GID2*. The knock-outs of *GID1* and *GID2* have been shown to supress the germination of Arabidopsis seeds (Hauvermale et al. 2015, Nelson and Steber 2017).

The balance between ABA and GA has been demonstrated to be decisive for the regulation of many physiological processes (Razem, Baron and Hill 2006, Olszewski, Sun and Gubler 2002). Of the ABA biosynthesis genes that responded to the PAC treatment, we found that the transcript coding for *NCED6* was up-regulated but the transcripts coding for the downstream enzyme molybdenum cofactor sulfurase (*ABA3*) were down-regulated. Thus, we expect that the ABA levels are increased as a result of the PAC treatment, although no ABA measurements were performed in this study.

Genes that have been shown to be induced by ABA were up-regulated by PAC, such as *GBF3*, *GPCR-type G protein 1*, *MFT* and *MARD1*.

The up-regulation of *MFT*, *MARD1* and *GBF3* was associated with the arrest of germination and these genes are possibly involved in the acquisition of desiccation tolerance in *C. limon* seeds upon PAC treatment (Figure 4). The transcription factor *G-box binding factor 3* (*GBF3*) is involved in the water stress response during the desiccation phase of *A. thaliana* seed formation (Wohlbach, Quirino and Sussman 2008). Furthermore, *MFT* responds to both ABA and GA signalling to regulate seed germination in *A. thaliana* (Xi et al. 2010) and *MARD1* plays a role in the arrest of germination through ABA-mediated dormancy (He and Gan 2004). The lack of developmental arrest has been associated with seed desiccation sensitivity in the *A. thaliana* mutants *fus3-8*, *lec1-3* and *abi3-5* (Raz, Bergervoet and Koornneef 2001).

GBF3 is involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions (Uno et al. 2000), as well as in the regulation of seed maturation (Belmonte et al. 2013). Together with other bZIP transcription factors, such as *ABI5* and *EEL*, *GBF3* regulates seed maturation and acquisition of desiccation tolerance (González-Morales et al. 2016, Bensmihen et al. 2002). The up-regulated genes are predicted to be targets of *GBF3*. It suggests that *GBF3* is an important regulator of the transcriptional changes occurring in PAC treated *C. limon* seeds and, possibly, with the acquisition of desiccation tolerance in these seeds.

We also observed the down-regulation of other hormonal pathway-associated genes, such as ethylene, which is also a promoter of germination (Linkies and Leubner-Metzger 2012). The Arabidopsis mutant *etr1-2*, for example, is insensitive to ethylene and its seeds are more dormant (Chiwocha et al. 2005). Extensive down-regulation of ethylene biosynthesis and ethylene responsive genes occurred in PAC treated seeds. We assume that the down-regulation of the ethylene pathway is also likely to be associated with arrest of germination of the *C. limon* seeds upon PAC treatment. Mutants insensitive to ethylene, such as *etr1* and *ein2*, are hypersensitive to ABA during germination (Beaudoin et al. 2000, Ghassemian et al. 2000). Endogenous ABA levels are elevated in *etr1* compared with those in the wild type, suggesting that ethylene signalling affects positively ABA levels (Chiwocha et al. 2005).

Taken together, the transcriptomic changes, resulting from the PAC treatment in *C. limon* seeds, are mainly related to the suppression of metabolism as a consequence of the down-regulation of multiple hormonal pathways.

References

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D. J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research,* 25**,** 3389-3402.

Arana, M. V., M. SÁNCHEZ-LAMAS, B. Strasser, S. E. Ibarra, P. D. Cerdan, J. F. Botto & R. A. Sanchez (2014) Functional diversity of phytochrome family in the control of light and gibberellin-mediated germination in Arabidopsis. *Plant, cell & environment,* 37.

Baskin, C. & J. Baskin (2014) Seeds: ecology, biogeography, and evolution of dormancy and germination.

Beardmore, T. & C.-A. Whittle (2005) Induction of tolerance to desiccation and cryopreservation in silver maple (Acer saccharinum) embryonic axes. *Tree physiology,* 25**,** 965-972.

Beaudoin, N., C. Serizet, F. Gosti & J. Giraudat (2000) Interactions between abscisic acid and ethylene signaling cascades. *The Plant Cell,* 12**,** 1103-1115.

Becwar, M., P. Stanwood & K. Leonhardt (1983) Dehydration effects on freezing characteristics and survival in liquid nitrogen of desiccation-tolerant and desiccation-sensitive seeds. *Journal of the American Society for Horticultural Science,* 108**,** 613-618.

Belmonte, M. F., R. C. Kirkbride, S. L. Stone, J. M. Pelletier, A. Q. Bui, E. C. Yeung, M. Hashimoto, J. Fei, C. M. Harada, M. D. Munoz, B. H. Le, G. N. Drews, S. M. Brady, R. B. Goldberg & J. J. Harada (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A,* 110**,** E435-44.

Bensmihen, S., S. Rippa, G. Lambert, D. Jublot, V. Pautot, F. Granier, J. Giraudat & F. Parcy (2002) The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *The Plant Cell,* 14**,** 1391-1403.

Bewley, J. D. (1979) Physiological aspects of desiccation tolerance. *Annual Review of Plant Physiology,* 30**,** 195-238.

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

Bolger, A. M., M. Lohse & B. Usadel (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics***,** btu170.

Boller, T. & G. Felix (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual review of plant biology,* 60**,** 379-406.

Boubriak, I., M. Dini, P. Berjak & D. J. Osborne (2000) Desiccation and survival in the recalcitrant seeds of Avicennia marina: DNA replication, DNA repair and protein synthesis. *Seed Science Research,* 10**,** 307-315.

Boubriak, I., H. Kargiolaki, L. Lyne & D. J. Osborne (1997) The requirement for DNA repair in desiccation tolerance of germinating embryos. *Seed Science Research,* 7**,** 97-106.

Bray, N. L., H. Pimentel, P. Melsted & L. Pachter (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology,* 34**,** 525-527.

Chandel, K., R. Chaudhury, J. Radhamani & S. Malik (1995) Desiccation and freezing sensitivity in recalcitrant seeds of tea, cocoa and jackfruit. *Annals of Botany,* 76**,** 443-450.

Chatelain, E., M. Hundertmark, O. Leprince, S. Le Gall, P. Satour, S. Deligny-Penninck, H. Rogniaux & J. Buitink (2012) Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ,* 35**,** 1440-55.

Chauhan, H., N. Khurana, P. Agarwal, J. P. Khurana & P. Khurana (2013) A seed preferential heat shock transcription factor from wheat provides abiotic stress tolerance and yield enhancement in transgenic Arabidopsis under heat stress environment. *PLoS One,* 8**,** e79577.

Chin, H. 1977. Production and storage of recalcitrant seeds in the tropics. In *Symposium on Seed Problems in Horticulture 83*, 17-22.

Chiwocha, S. D., A. J. Cutler, S. R. Abrams, S. J. Ambrose, J. Yang, A. R. Ross & A. R. Kermode (2005) The etr1-2 mutation in Arabidopsis thaliana affects the abscisic acid, auxin, cytokinin and gibberellin metabolic pathways during maintenance of seed dormancy, moist-chilling and germination. *The Plant Journal,* 42**,** 35-48.

Costa, M. C., K. Righetti, H. Nijveen, F. Yazdanpanah, W. Ligterink, J. Buitink & H. W. Hilhorst (2015) A gene co-expression network predicts functional genes controlling the re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds. *Planta,* 242**,** 435-49.

Dalziel, J. & D. Lawrence (1984) Biochemical and biological effects of kaurene oxidase inhibitors, such as paclobutrazol. *Monograph-British Plant Growth Regulation Group*.

Dennis, G., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane & R. A. Lempicki (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome biology,* 4**,** R60.

Du, Z., X. Zhou, Y. Ling, Z. Zhang & Z. Su (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic acids research*, gkq310.

Eckardt, N. A. (2008) Chitin signaling in plants: insights into the perception of fungal pathogens and rhizobacterial symbionts. *The Plant Cell,* 20, 241-243.

El Hadrami, A., L. R. Adam, I. El Hadrami & F. Daayf (2010) Chitosan in plant protection. *Marine drugs,* 8, 968-987.

Ellis, R., T. Hong & E. Roberts (1991) An intermediate category of seed storage behaviour? II. Effects of provenance, immaturity, and imbibition on desiccation-tolerance in coffee. *Journal of Experimental Botany,* 42, 653-657.

Faria, J. M., J. Buitink, A. A. van Lammeren & H. W. M. Hilhorst (2005) Changes in DNA and microtubules during loss and re-establishment of desiccation tolerance in germinating Medicago truncatula seeds. *J Exp Bot,* 56, 2119-30.

Farnsworth, E. (2000) The ecology and physiology of viviparous and recalcitrant seeds. *Annual Review of Ecology and Systematics*, 107-138.

Farrant, J. M. & J. P. Moore (2011) Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Curr Opin Plant Biol,* 14, 340-5.

Gallão, M. I., A. L. Cortelazzo, M. P. Fevereiro & E. S. d. Brito (2007) Response to chitin in suspension-cultured Citrus aurantium cells. *Brazilian journal of plant physiology,* 19, 69-76.

Garello, G. & M. T. Page-Degivry (1995) Desiccation-sensitive Hopea odorata seeds: Sensitivity to abscisic acid, water potential and inhibitors of gibberellin biosynthesis. *Physiologia Plantarum,* 95, 45-50.

Ghassemian, M., E. Nambara, S. Cutler, H. Kawaide, Y. Kamiya & P. McCourt (2000) Regulation of abscisic acid signaling by the ethylene response pathway in Arabidopsis. *The Plant Cell,* 12, 1117-1126.

González-Morales, S. I., R. A. Chávez-Montes, C. Hayano-Kanashiro, G. Alejo-Jacuinde, T. Y. Rico-Cambron, S. de Folter & L. Herrera-Estrella (2016) Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 201610985.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury & Q. Zeng (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology,* 29, 644.

Gutierrez, L., O. Van Wuytswinkel, M. Castelain & C. Bellini (2007) Combined networks regulating seed maturation. *Trends Plant Sci,* 12, 294-300.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li & M. Lieber (2013) De novo transcript sequence

Du, Z., X. Zhou, Y. Ling, Z. Zhang & Z. Su (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic acids research*, gkq310.

Eckardt, N. A. (2008) Chitin signaling in plants: insights into the perception of fungal pathogens and rhizobacterial symbionts. *The Plant Cell,* 20, 241-243.

El Hadrami, A., L. R. Adam, I. El Hadrami & F. Daayf (2010) Chitosan in plant protection. *Marine drugs,* 8, 968-987.

Ellis, R., T. Hong & E. Roberts (1991) An intermediate category of seed storage behaviour? II. Effects of provenance, immaturity, and imbibition on desiccation-tolerance in coffee. *Journal of Experimental Botany,* 42, 653-657.

Faria, J. M., J. Buitink, A. A. van Lammeren & H. W. M. Hilhorst (2005) Changes in DNA and microtubules during loss and re-establishment of desiccation tolerance in germinating Medicago truncatula seeds. *J Exp Bot,* 56, 2119-30.

Farnsworth, E. (2000) The ecology and physiology of viviparous and recalcitrant seeds. *Annual Review of Ecology and Systematics*, 107-138.

Farrant, J. M. & J. P. Moore (2011) Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Curr Opin Plant Biol,* 14, 340-5.

Gallão, M. I., A. L. Cortelazzo, M. P. Fevereiro & E. S. d. Brito (2007) Response to chitin in suspension-cultured Citrus aurantium cells. *Brazilian journal of plant physiology,* 19, 69-76.

Garello, G. & M. T. Page-Degivry (1995) Desiccation-sensitive Hopea odorata seeds: Sensitivity to abscisic acid, water potential and inhibitors of gibberellin biosynthesis. *Physiologia Plantarum,* 95, 45-50.

Ghassemian, M., E. Nambara, S. Cutler, H. Kawaide, Y. Kamiya & P. McCourt (2000) Regulation of abscisic acid signaling by the ethylene response pathway in Arabidopsis. *The Plant Cell,* 12, 1117-1126.

González-Morales, S. I., R. A. Chávez-Montes, C. Hayano-Kanashiro, G. Alejo-Jacuinde, T. Y. Rico-Cambron, S. de Folter & L. Herrera-Estrella (2016) Regulatory network analysis reveals novel regulators of seed desiccation tolerance in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 201610985.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury & Q. Zeng (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology,* 29, 644.

Gutierrez, L., O. Van Wuytswinkel, M. Castelain & C. Bellini (2007) Combined networks regulating seed maturation. *Trends Plant Sci,* 12, 294-300.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li & M. Lieber (2013) De novo transcript sequence

reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols,* 8**,** 1494-1512.

Hauvermale, A. L., K. M. Tuttle, Y. Takebayashi, M. Seo & C. M. Steber (2015) Loss of Arabidopsis thaliana seed dormancy is associated with increased accumulation of the GID1 GA hormone receptors. *Plant and Cell Physiology***,** pcv084.

He, Y. & S. Gan (2004) A novel zinc-finger protein with a proline-rich domain mediates ABA-regulated seed dormancy in Arabidopsis. *Plant molecular biology,* 54**,** 1-9.

Hilhorst, H. & C. Karssen (1992) Seed dormancy and germination: the role of abscisic acid and gibberellins and the importance of hormone mutants. *Plant growth regulation,* 11**,** 225-238.

Hong, T., N. Ahmad & A. Murdoch (2001) Optimum air-dry storage conditions for sweet orange (Citrus sinensis (L.) Osbeck) and lemon (Citrus limon (L.) Burm. f.) seeds. *Seed science and technology,* 29**,** 183-192.

Hundertmark, M. & D. K. Hincha (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC genomics,* 9**,** 118.

Jacobsen, S. E. & N. E. Olszewski (1993) Mutations at the SPINDLY locus of Arabidopsis alter gibberellin signal transduction. *The Plant Cell,* 5**,** 887-896.

Jiang, Z., G. Xu, Y. Jing, W. Tang & R. Lin (2016) Phytochrome B and REVEILLE1/2-mediated signalling controls seed dormancy and germination in Arabidopsis. *Nature communications,* 7.

Jin, J., F. Tian, D.-C. Yang, Y.-Q. Meng, L. Kong, J. Luo & G. Gao (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research,* 45**,** D1040-D1045.

Khan, M. M., M. A. Alam, M. Abbas & M. J. Iqbal (2003) Studies on seed desiccation tolerance in four citrus species. *Pak. J. Agri. Sci,* 40**,** 55-62.

Khan, M. M., K. Thompson & B. Muhammad Usman Fatima (2002) Role of moisture content and controlled atmosphere in Citrus seed storage. *Int. J. Agric. and Biol,* 4**,** 259-266.

King, M., U. Soetisna & E. Roberts (1981) The dry storage of Citrus seeds. *Annals of Botany,* 48**,** 865-872.

Kotak, S., E. Vierling, H. Bäumlein & P. von Koskull-Döring (2007) A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. *The Plant Cell,* 19**,** 182-195.

Kucera, B., M. A. Cohn & G. Leubner-Metzger (2005) Plant hormone interactions during seed dormancy release and germination. *Seed Science Research,* 15**,** 281-307.

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander & M. Garcia-Hernandez (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research,* 40**,** D1202-D1210.

Leng, L., Q. Liang, J. Jiang, C. Zhang, Y. Hao, X. Wang & W. Su (2017) A subclass of HSP70s regulate development and abiotic stress responses in Arabidopsis thaliana. *Journal of plant research,* 130**,** 349-363.

Leprince, O., J. Buitink & F. A. Hoekstra (1999) Axes and cotyledons of recalcitrant seeds of Castanea sativa Mill. exhibit contrasting responses of respiration to drying in relation to desiccation sensitivity. *Journal of experimental botany,* 50**,** 1515-1524.

Linkies, A. & G. Leubner-Metzger (2012) Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination. *Plant cell reports,* 31**,** 253-270.

Mattimore, V. & J. R. Battista (1996) Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *Journal of bacteriology,* 178**,** 633-637.

Mauch-Mani, B. & F. Mauch (2005) The role of abscisic acid in plant–pathogen interactions. *Current opinion in plant biology,* 8**,** 409-414.

Moore, J. P., M. Vicre-Gibouin, J. M. Farrant & A. Driouich (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiol Plant,* 134**,** 237-45.

Nelson, S. K. & C. M. Steber (2017) Transcriptional mechanisms associated with seed dormancy and dormancy loss in the gibberellin-insensitive sly1-2 mutant of Arabidopsis thaliana. *PLoS One,* 12**,** e0179143.

Ogawa, M. (2003) Gibberellin Biosynthesis and Response during Arabidopsis Seed Germination. *The Plant Cell Online,* 15**,** 1591-1604.

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

Olszewski, N., T.-p. Sun & F. Gubler (2002) Gibberellin signaling biosynthesis, catabolism, and response pathways. *The Plant Cell,* 14**,** S61-S80.

Ooms, J. J., K. M. Leon-Kloosterziel, D. Bartels, M. Koornneef & C. M. Karssen (1993) Acquisition of desiccation tolerance and longevity in seeds of Arabidopsis thaliana (a comparative study using abscisic acid-insensitive abi3 mutants). *Plant Physiology,* 102**,** 1185-1191.

Pammenter, N. & P. Berjak (1999) A review of recalcitrant seed physiology in relation to desiccation-tolerance mechanisms. *Seed Science Research,* 9**,** 13-37.

Pérez-Salamó, I., C. Papdi, G. Rigó, L. Zsigmond, B. Vilela, V. Lumbreras, I. Nagy, B. Horváth, M. Domoki & Z. Darula (2014) The heat shock factor A4A confers salt tolerance and is regulated by oxidative stress and the mitogen-activated protein kinases MPK3 and MPK6. *Plant physiology,* 165**,** 319-334.

Pimentel, H., N. L. Bray, S. Puente, P. Melsted & L. Pachter (2017) Differential analysis of RNA-Seq incorporating quantification uncertainty. *Nature Methods*.

Rademacher, W. (1989) Gibberellins: metabolic pathways and inhibitors of biosynthesis. *Target Sites for Herbicide Action. Edited by Boeger, P. and Sandmann, G***,** 127-145.

Rademacher, W., H. Fritsch, J. E. Graebe, H. Sauter & J. Jung (1987) Tetcyclacis and triazole-type plant growth retardants: Their influence on the biosynthesis of gibberellins and other metabolic processes. *Pest Management Science,* 21**,** 241-252.

Raz, V., J. Bergervoet & M. Koornneef (2001) Sequential steps for developmental arrest in Arabidopsis seeds. *Development,* 128**,** 243-252.

Razem, F. A., K. Baron & R. D. Hill (2006) Turning on gibberellin and abscisic acid signaling. *Current opinion in plant biology,* 9**,** 454-459.

Rikhvanov, E. G., K. Z. Gamburg, N. N. Varakina, T. M. Rusaleva, I. V. Fedoseeva, E. L. Tauson, I. V. Stupnikova, A. V. Stepanov, G. B. Borovskii & V. K. Voinikov (2007) Nuclear–mitochondrial cross-talk during heat shock in Arabidopsis cell culture. *The Plant Journal,* 52**,** 763-778.

Soetisna, U., M. King & E. Roberts (1985) Germination test recommendations for estimating the viability of moist or dry seeds of lemon (Citrus limon) and lime (C. aurantifolia). *Seed science and technology,* 13**,** 87-110.

Supek, F., M. Bošnjak, N. Škunca & T. Šmuc (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one,* 6**,** e21800.

The Bio-Analytic Resource for Plant Biology (2018) Sample Angler.

Tompsett, P. (1984) Desiccation studies in relation to the storage of Araucaria seed. *Annals of Applied Biology,* 105**,** 581-586.

Tompsett, P. (1987) Desiccation and storage studies on Dipterocarpm seeds. *Annals of Applied Biology,* 110**,** 371-379.

Tweddle, J., R. Turner & J. Dickie. 2002. Seed Information Database. Release.

Tweddle, J. C., J. B. Dickie, C. C. Baskin & J. M. Baskin (2003) Ecological aspects of seed desiccation sensitivity. *Journal of Ecology,* 91**,** 294-304.

Uno, Y., T. Furihata, H. Abe, R. Yoshida, K. Shinozaki & K. Yamaguchi-Shinozaki (2000) Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc Natl Acad Sci U S A,* 97**,** 11632-7.

Varoquaux, F., R. Blanvillain, M. Delseny & P. Gallois (2000) Less is better: new approaches for seedless fruit production. *Trends in biotechnology,* 18**,** 233-242.

Velasco, R. & C. Licciardello (2014) A genealogy of the citrus family. *Nature biotechnology,* 32**,** 640-642.

Verdier, J., D. Lalanne, S. Pelletier, I. Torres-Jerez, K. Righetti, K. Bandyopadhyay, O. Leprince, E. Chatelain, B. L. Vu, J. Gouzy, P. Gamas, M. K. Udvardi & J. Buitink (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. *Plant Physiol,* 163**,** 757-74.

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

Wang, W., B. Vinocur, O. Shoseyov & A. Altman (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci,* 9**,** 244-52.

Wang, X., Y. Xu, S. Zhang, L. Cao, Y. Huang, J. Cheng, G. Wu, S. Tian, C. Chen & Y. Liu (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genetics,* 49**,** 765-772.

Wehmeyer, N., L. D. Hernandez, R. R. Finkelstein & E. Vierling (1996) Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Physiology,* 112**,** 747-757.

Wehmeyer, N. & E. Vierling (2000) The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. *Plant Physiology,* 122**,** 1099-1108.

Wen, B., C. Cai, R. Wang, Y. Tan & Q. Lan (2010) Critical moisture content windows differ for the cryopreservation of pomelo (Citrus grandis) seeds and embryonic axes. *CryoLetters,* 31**,** 29-39.

White, C. N., W. M. Proebsting, P. Hedden & C. J. Rivin (2000) Gibberellins and seed development in maize. I. Evidence that gibberellin/abscisic acid balance governs germination versus maturation pathways. *Plant Physiology,* 122**,** 1081-1088.

Wohlbach, D. J., B. F. Quirino & M. R. Sussman (2008) Analysis of the Arabidopsis histidine kinase ATHK1 reveals a connection between vegetative osmotic stress sensing and seed maturation. *The Plant Cell,* 20**,** 1101-1117.

Xi, W., C. Liu, X. Hou & H. Yu (2010) MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in Arabidopsis. *The Plant Cell,* 22**,** 1733-1748.

Yamaguchi, S. (2008) Gibberellin metabolism and its regulation. *Annu. Rev. Plant Biol.,* 59**,** 225-251.

Zhang, Z.-L., M. Ogawa, C. M. Fleet, R. Zentella, J. Hu, J.-O. Heo, J. Lim, Y. Kamiya, S. Yamaguchi & T.-p. Sun (2011) Scarecrow-like 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *Proceedings of the National Academy of Sciences,* 108**,** 2160-2165.

Supplementary material

Supplementary Table 1. Sequencing reads

Sample	Reads mapped	Reads processed	Fraction mapped
Control-1	15,587,103	17,364,908	0.8976
Control-2	19,938,139	22,238,751	0.8965
Control-3	15,329,299	16,657,259	0.9203
Paclobutrazol-1	16,376,057	18,365,376	0.8917
Paclobutrazol-2	16,870,180	18,835,280	0.8957
Paclobutrazol-3	12,865,185	14,398,974	0.8935

Supplementary Table 2. Statistics for RNA-seq mapping

Counts of transcripts, GC content.
Total trinity 'genes': 103,841
Total trinity transcripts: 179,588
Percent GC: 39.66

Statistics based on all transcript contigs
Contig N10: 4,160
Contig N20: 3,277
Contig N30: 2,740
Contig N40: 2,309
Contig N50: 1,933
Median contig length: 612
Average contig: 1,079.95
Total assembled bases: 193,945,868

Statistics based on only longest isoform per 'GENE'
Contig N10: 3,984
Contig N20: 2,988
Contig N30: 2,338
Contig N40: 1,765
Contig N50: 1,213
Median contig length: 375
Average contig: 713.71
Total assembled bases: 74,112,540

6

General Discussion

Desiccation tolerance evolved as the key adaptation to survival on land by the early algal ancestors of terrestrial plants. The first seed plants lost vegetative desiccation tolerance but conserved the seeds with tolerance mechanisms as an important strategy for species survival in unfavourable environments. The mechanisms of desiccation tolerance in seeds are connected to their developmentally regulated maturation programs. We propose that desiccation sensitive (DS) seeds lost tolerance by skipping part of their seed maturation program and advancing towards germination. Moreover, we expect to observe the activation of advanced germination specific genetic elements in transcriptomes of DS seeds.

The transition from quiescent to germinating seed is dependent on developmental genetic programs controlled by the endogenous balance and the signalling pathways of plant hormones and transcription factors. We investigated how these genetic programs were altered throughout the evolution of DS seeds.

In this thesis, I aimed to gain a better understanding of the physiological and molecular determinants of seed desiccation sensitivity. A thorough review of the current state of the art in this field was made and several physiological, genetic and molecular studies were performed. Finally, in the Discussion section, I aim to both connect and synthesize the main findings and insights from the previous chapters, and provide suggestions for potential future research, which have arisen as a result of those acquired insights. Figure 1 summarizes the relevant ecological, physiological, metabolic and molecular aspects of seed desiccation sensitivity regarded in this thesis.

Lack of seed developmental arrest, maturation drying and metabolic quiescence result in desiccation sensitivity.

The hypothesis that developing DS seeds lack the phases of maturation drying and metabolic quiescence, and, thus, proceed directly to the germination phase has long been discussed (Ellis et al., 1991; Farnsworth, 2000; Farrant et al., 1993). This hypothesis is supported by the fact that desiccation tolerant (DT) seeds lose desiccation tolerance upon radicle protrusion (for review see Dekkers et al., 2015).

Desiccation sensitivity is more frequent in seeds that are non-dormant upon shedding (c. 31%), than dormant (c. 9%) (Tweddle et al., 2003). The majority of the non-dormant seeded species are DT; desiccation during maturation causes arrest of development and will cause immediate germination of the quiescent seed once it gets imbibed. Alternatively, DS seeds do not dry and will progress towards germination due to an active metabolism and lack of dormancy.

Similar effects can be observed in desiccation tolerance of e.g. ABA-deficient tomato seeds that germinate inside the moist fruits, probably because their dormancy program is disrupted by a mutation in ABA-biosynthesis (Groot and Karssen, 1992).

Arabidopsis knockout mutants for the transcription factors ABI3, FUS3 and LEC1 also have DS seeds, together with a lack of dormancy and occurrence of premature germination inside the siliques (Raz et al., 2001). Furthermore, the Arabidopsis *abi3-5* knockout mutant has DS seeds with an enlarged apical meristem, confirming the continuation of embryo growth (Rohde et al., 2000). Additionally, the cause of the gradual loss of desiccation tolerance during the progress of germination and, afterwards, of wild type seeds has been associated with the loss of abscisic acid (ABA) sensitivity of the growing embryo (Maia et al., 2014).
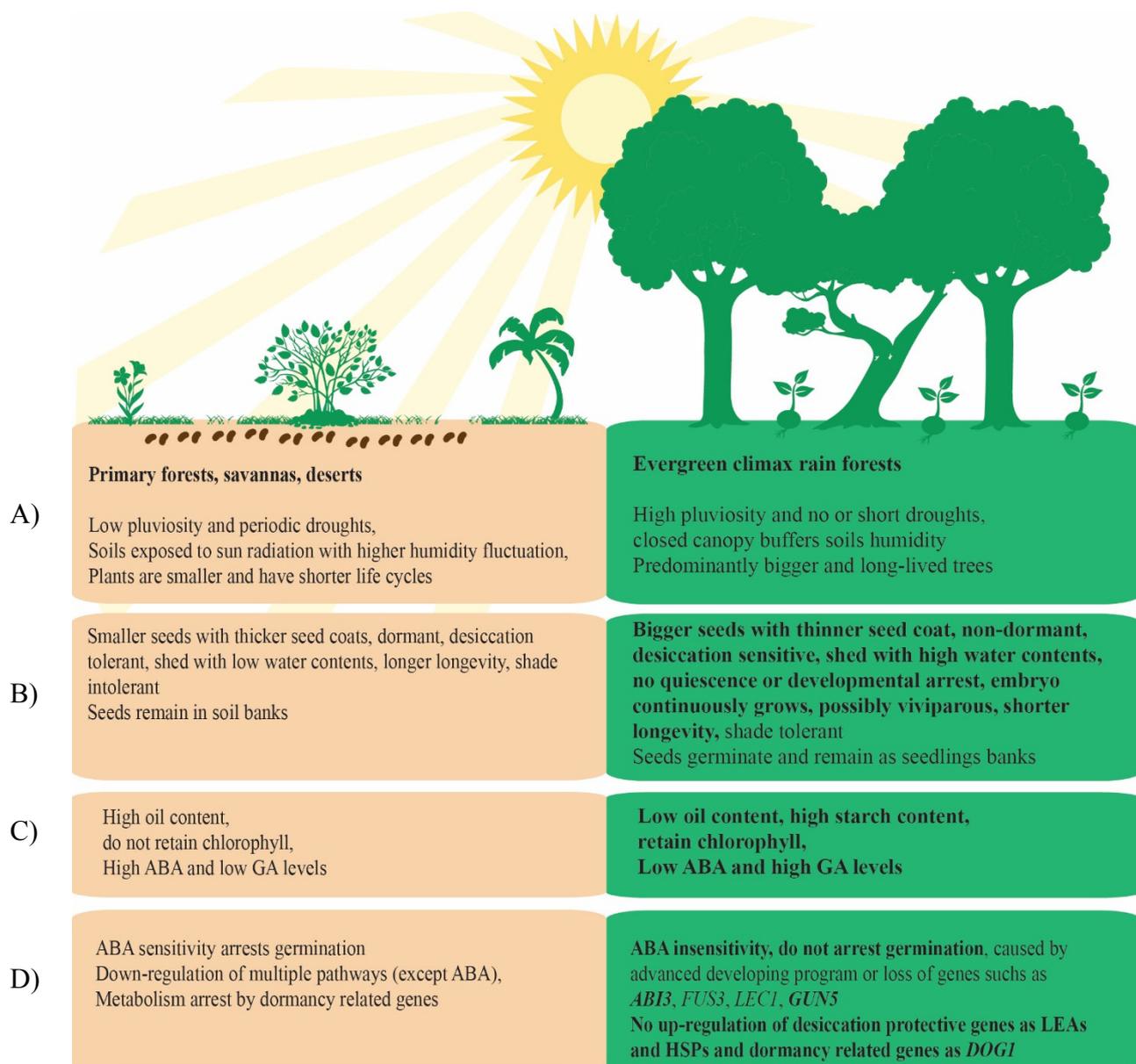
Figure 1. Main aspects related to DT (left panel) and DS seeds (right panel). A) Ecological, B) Seed physiological, C) Metabolic and D) Molecular aspects. Information in bold relates to topics investigated in the present thesis.

Primary embryonic dormancy represents a developmental arrest that prevents fully developed seeds to germinate inside the fruit (Bewley, 1997). The lack of developmental arrest of Arabidopsis *abi3-6* seeds may be associated with a lack of expression of dormancy-associated genes. In Chapter 3 it is shown that the transcriptome of *abi3-6* at late developmental stages is particularly similar to (germinating) Col-0 24h-imbibed seeds. I observed an enrichment of dormancy-associated GO categories only in Col-0 seeds at 10 days after flowering (daf) but not in *abi3-6* at the same stage. Enrichment of the GO category *embryonic development ending in seed dormancy* was also observed upon re-establishment of desiccation tolerance in germinated Col-0 seeds (Maia et al., 2011). Furthermore, the *DELAY OF GERMINATION 1 (DOG1)* gene showed an expression profile that was second best in correlating with desiccation tolerance levels in Col-0 and *abi3-6* (after *CYSTEINE PROTEASE 2*). The genetic interaction of *DOG1* with *ABI3* in the regulation of seed development has been demonstrated by Dekkers et al. (2016). Furthermore, *DOG1* has been identified as a major quantitative trait locus for seed dormancy variability among Arabidopsis ecotypes and *dog1* T-DNA insertional mutants have strongly reduced seed dormancy (Bentsink et al., 2006).

The lack of developmental arrest and suppression of metabolism are also evidenced by the presence of chlorophyll in late stages of *abi3-6* seed development. The *abi3-6* seeds retain chlorophyll after shedding and lack both desiccation tolerance and dormancy. These phenotypes are established during the maturation phase of seeds, when *ABI3* expression peaks (Delmas et al., 2013b; Koornneef et al., 1989; Leprince et al., 1998; Monke et al., 2012).

The majority of green embryos in the eastern Australian rainforest species, including *Castanospermum australe*, are DS (Hamilton et al., 2013). Chlorophyll retention in DT seeds is commonly associated with low seed longevity, arguably due to the deteriorative effects of photo-oxidation (Roqueiro et al., 2010).

In Chapter 4, transcripts related to *root development* and *developmental process* were identified with decreasing abundance in *Medicago truncatula* and increasing in *C. australe* during maturation. The opposite patterns of these developmentally related genes towards the end of seed development indicates that *C. australe* is still developmentally active while *M. truncatula* is arresting its development.

Paclobutrazol (PAC) is an inhibitor of gibberellin biosynthesis. Gibberellins consist of a group of phytohormones responsible for the release of dormancy and stimulation of germination. In Chapter 5, it was demonstrated that PAC arrested both the germination of *Citrus limon* seeds and induced desiccation tolerance. I observed that the major transcriptomic changes caused by PAC in *C. limon* seeds were the down-regulation of multiple hormonal pathways, except that of ABA. The

down-regulation of hormonal pathways involved with promotion of development and metabolism, e.g. gibberellin and auxin, indicates that these seeds are showing development arrest.

The DS seeds of silver maple (*Acer saccharinum*) could be made DT by a treatment with ABA and tetcyclacis, which inhibits both ABA catabolism and GA biosynthesis (Beardmore and Whittle, 2005). Synthesis of putative seed storage proteins and dehydrins were observed upon this treatment, indicating an induction of maturation proteins in embryos that were otherwise advancing towards germination.

Genetic alterations observed among DS-seeded species

Since Roberts (1973) first described the phenomenon of DS seeds, hypotheses have been put forward about their evolutionary and physiological origin(s). One of the early hypotheses attempting to address the cause of desiccation sensitivity of seeds was proposed by Farrant et al. (1993). They suggested that DS seeds lack the genes or competence to produce LEA proteins and, as a result, remain sensitive to desiccation.

Other studies with wild DS-seeded species support the hypothesis that seed desiccation sensitivity is caused by skipping the maturation program and progression towards germination (Ellis et al., 1991; Farnsworth, 2000; Farrant et al., 1993). A similar hypothesis has been proposed for *Arabidopsis thaliana*. Okamoto et al. (2010) stated: "Severe [Arabidopsis] ABA-insensitive mutants, such as *abi3-6* or the *snrk2d/e/i* triple mutant, cannot enter the seed maturation processes". Similarly, Delmas et al. (2013a) stated: "*abi3-6* exhibits pleiotropic effects during seed development that bypass the embryo maturation phase and progress directly into the germination and vegetative phases. This lack of an embryo maturation phase results in seeds that remain green due to failure of embryo degreening in mutants containing severe *abi3* alleles".

Based on the present study and the large variability of DS seed phenotypes observed in nature (Chapter 2), I also support the hypothesis the hypothesis that seed desiccation sensitivity is caused by skipping the maturation program, or part of it, and progression towards germination.

Previous studies have demonstrated intraspecific clinal variation for seed desiccation sensitivity. DS species may show an intraspecific gradient of desiccation tolerance. For example, the seeds of *Acer pseudoplatanus* display a varying tolerance across their native plant distribution ranges (Daws and Pritchard, 2008). The intraspecific variability observed for seed desiccation tolerance reflects the high plasticity of this trait and its possible role in species adaptation to different environments. The variation observed for seed desiccation sensitivity in such

populations is likely a result of their genetic diversity. Studies exploiting genetic mapping may reveal loci associated with desiccation sensitivity.

DS-seeded species are not phylogenetically related (Pammenter and Berjak, 2000). Thus, each DS-seeded species is probably a result of unique unrelated genetic events. However, a genetic bottleneck is possible and thus convergent evolutionary events taking place due to a restricted number of possible genetic alterations that leads to seed desiccation sensitivity could be expected.

Should we expect abrupt genomic changes in the transition from DT to DS seeds? When the seagrass *Zostera marina* moved from a terrestrial to an aquatic lifestyle it was an abrupt behavioural change that led to the loss of massive numbers of genes related to the terrestrial habit, including genes associated with LEAs, ethylene signalling and stomatal development (Olsen et al., 2016). This species does not display an ecological or phenotypical continuum.

However, in our present study of seed desiccation sensitivity our observations are similar to those of *A. pseudoplatanus,* which suggests that the transition from DT to DS seeds is not abrupt and determinant. Possibly, the phenotypical and genomic versatility are important to these species' adaptation to a constantly changing terrestrial environment. Thus, we should not expect a complete degeneration of the desiccation tolerance genomic machinery as a result of the transition from DT to DS seeds, but rather should find key regulatory changes, which are likely reversible.

Initially, I investigated the common transcriptional features that distinguish DS from DT seeds. In DT seeds, tolerance is acquired midway during seed development, concomitantly with seed filling. Previous work has identified transcripts that accumulate during the acquisition of DT in *M. truncatula* seeds (Terrasson et al. 2013; Righetti et al. 2015). Chapter 4 demonstrates that of these transcripts the homologs of 102 genes failed to accumulate during development of DS *C. australe* seeds. These transcripts are related to sugar metabolism, photosynthesis, seed development, protection against abiotic stress and modulation of plant stress responses. Examples of these transcripts are *ABI3*, *ABI5*, LEAs, *TRANSPARENT TESTA PROTEIN*, oleosins, *1-CYSTEINE PEROXIREDOXIN* and α-galactosidases.

C. australe seeds show increased but still low levels of dehydration tolerance towards the late developmental stages. This can be inferred as seed maturation with some sort of acquisition of desiccation tolerance.

Secondly, I performed explorative and unbiased genomic analysis and investigated if the genomic changes associated with desiccation sensitivity are extensive or punctual. Should I find, for example, extensive loss of LEA genes or punctual loss of an essential TF, such as ABI3, in the genome of a DS-seeded species'?

In Chapter 4, no extensive genomic alterations related to desiccation tolerance were found: e.g. *C. australe* has a normal number of LEAs. Additionally, no extensive

gene loss or expansion, loss of synteny, nor increased mutation rates of DT-associated genes were found. Conversely, important regulatory genomic alterations related to hormonal signalling and seed development were observed. It is possible that the genes responsible for the acquisition of desiccation tolerance are pleiotropic and, thus, are not vastly lost along the evolution of DS seeds. Instead, punctual regulatory genomic alterations, possibly hormone related, could have occurred during the evolution of DS-seeds.

The reversion to DT by hormonal manipulations in some DS-seeded species reveals that they still possess the genomic repertoire to do so. As said, *C. australe* has a normal number of LEA genes in its genome compared to DT-seeded species; however, they are not expressed during seed development (Chapter 4). Further studies are necessary to explore the inducibility of desiccation tolerance in *C. australe* and other DS-seeded species and to correlate this with their transcriptomic and genomic features.

When comparing genomes of DS- with DT-seeded species, there appeared to be no proteins retained in all DT species and lost in all DS species. However, several were lost in all DS- and retained in at least half of the DT-seeded species. Among these were a lipid transferase protein and the transcription factors *VERDANDI* and *MYB44-like*. *VERDANDI* is a transcription factor that participates in the ovule identity complex and, when mutated, affects embryo sac differentiation in *Arabidopsis thaliana* (Matias-Hernandez et al., 2010; Mendes et al., 2016). Interestingly, the transcription factors *VERDANDI* and *MYB44-like* were also retained in DT- and intermediate-seeded species. Similarly, *SHP2* was lost in the genome of DS-seeded *Castanea mollissima*. It is involved in ovule integument specification, endosperm- and seed coat development (Ehlers et al., 2016).

One gene (*PLAC8*) was lost without retention of paralogs in three out of four DS-seeded species, namely *C. australe*, *C. mollissima* and *Elaeis guineensis*. The knock-out of this gene caused increased seed and fruit size in maize (Libault and Stacey 2010). In addition, the fw2.2 locus containing the *PLAC8* gene has been suggested to be the key to the evolution of tomato fruit size (Frary 2000). Large seeds and fruits are common features of DS-seeded species and presumably reduce the rate of seed drying and, hence, the risk of desiccation-induced mortality (Daws et al. 2006; Guo et al. 2010).

The gene *GUN5*, a magnesium chelatase involved in retrograde signalling that has been demonstrated to be a hub for ABA signalling to the nucleus (Jiang et al. 2014), was lost in *C. australe* without paralogs. This pathway is strongly affected in *abi5* mutants of the legume *M. truncatula,* which produce seeds with strongly reduced longevity (Zinsmeister et al. 2016) and cannot reacquire desiccation tolerance after germination (Terrasson et al. 2013).

There were 76 genes that had lost their ortholog but kept a paralog in *C. australe.* Among them, 2 genes involved in ABA- and brassinosteroid (BR) signalling were identified:

RETARDED ROOT GROWTH-LIKE (*RRL)* and *MOTHER OF FT* (*MFT*). RRL mediates ABA signal transduction through ABI4 (Park et al. 2015) and MFT regulates seed germination and fertility involving ABA- and BR signalling pathways (Sun et al. 2010). The gene *GPCR-type G* was lost in the genome of the DS-seeded *Castanea mollissima* and when knocked-out in *A. thaliana* it causes ABA hyposensitivity at germination (Pandey et al., 2009).

MFT and *ACTIVATION-TAGGED BRI1 (BRASSINOSTEROID-INSENSITIVE1)-SUPPRESSOR1* (*ATBS1*) showed very contrasting expression patterns during seed maturation in Col-0, as compared to the *abi3-6* mutant (Chapter 3). Moreover, *ATBS1* also showed a contrasting expression pattern between *C. australe* and *M. truncatula* seed maturation (Chapter 4). Interestingly, it is a gene involved in gibberellin- and BR signalling (Schlereth et al. 2010).

The average size of DS seeds is significantly greater than those of DT seeds (Hamilton et al., 2013). Heterochronic evolutionary mechanisms can be defined as changes in the rate and/or timing (onset-offset) of developmental events between ancestral and descendant ontogenies (Gould, 1977; Rice, 1997). Possibly, DS seeds are evolved by acceleration or hypermorphosis, which means a faster, longer and/or further development in relation to its ancestors, contrary to the earlier completion of development, neoteny. This hypermorphosis development might also lead to bigger size of the DS seeds and a more advanced developmental program towards germination program upon shedding.

Seeds of silver maple and *C. limon* are DS but they are still of a relatively small size as opposed to *C. australe* which are considerably bigger. Possibly, *C. australe* seeds underwent a greater heterochronic evolution, getting bigger, developed further towards germination and became more insensitive to ABA. This is possibly the reason why the treatment of induction of desiccation tolerance with hormones and hormone-inhibitors was successful in silver maple[2] (Beardmore and Whittle, 2005) and *C. limon* (Chapter 5) but not in *C. australe* seeds (Chapter 4). Similarly, germinated Arabidopsis seeds are also DS but, at early stages of radicle protrusion, ABA treatment is still able to re-establish desiccation tolerance. Further embryo growth reduces both sensitivity to ABA and the capacity to re-induce desiccation tolerance (Maia et al., 2011).

Altogether, a number of significant genetic alterations among DS-seeded species were observed. These alterations indicate that seed desiccation sensitivity is not a consequence of extensive dismantling of the desiccation tolerance machinery. Instead, punctual regulatory alterations seem to take place during the evolution of DS-seeds. This

---

[2] The first successful induction of desiccation tolerance was a significant achievement towards the conservation of DS-seeded species. However, this initiative has not been followed up until this thesis in which this concept was addressed. Research focused on conservation of DS-seeded species in relation with biodiversity has been neglected. I advocate attention for the urgency of developing conservation protocols for these in face of imminent risk of massive loss of biodiversity due to deforestation and climate change.

suggests that these species could adapt relatively fast to the new DS-seeded life style without extensive genetic modifications. Possibly the re-adaptation back to the DT-seeded life style is also not genetically costly and this would mean that seed desiccation tolerance/sensitivity is a versatile trait that permits populations to adapt to new environments. Such exploratory genetic observations require further study to understand the mechanistic role of these genes in the acquisition of seed desiccation tolerance/sensitivity.

Is hormonal biosynthesis or -signalling altered in desiccation sensitive seeds?

The DS-seeded species studied in this thesis share some striking hormone-related genomic alterations. The Arabidopsis *abi3-6* mutant (Chapter 3) is ABA-insensitive and was discovered by a mutagenesis screening for seeds that germinated in ABA solutions (Nambara et al., 1994). *C. australe* seeds are also insensitive to ABA and germinate even on high concentrations of ABA (Chapter 4). *C. limon* seeds are also insensitive to ABA and required the inhibition of GA-biosynthesis to induce desiccation tolerance (Chapter 5).

A treatment with ABA may re-establish desiccation tolerance in germinated seeds of many species (Buitink et al., 2003; Costa et al., 2017; Maia et al., 2011). Similarly, desiccation tolerance was induced in a DS seed by the treatment of silver maple seeds with a combination of tetcyclacis and ABA (Beardmore and Whittle, 2005).

In Chapter 5, I expected that the inhibition of gibberellin biosynthesis by paclobutrazol (PAC) in *C. limon* seeds would down-regulate the downstream genes involved in the gibberellin-biosynthesis pathway and, thus, arrest germination. The down-regulation of diverse gibberellin-responsive genes including *GA2OX2*, *GID1B* and *SLY1* was indeed observed. However, the extent of the PAC treatment was much broader and it suppressed several other hormone-signalling pathways, probably due to the cross talk between gibberellin signalling and that of other hormones (Weiss and Ori, 2007). Among these are genes involved in the signalling of auxin, abscisic acid, ethylene, brassinosteroids, jasmonic- and salicylic acid. Altogether, I infer that the PAC treatment induced desiccation tolerance in *C. limon* seeds by triggering a metabolic quiescence that suppressed multiple hormonal pathways and prevented them from proceeding towards a germination program.

In Chapter 3, Arabidopsis Col-0 seed development also showed extensive downregulation of multiple hormonal pathways. During the late maturation phase, GO enrichment in the downregulated gene set included multiple hormone-associated categories, such as: *brassinosteroid metabolic process, response to cytokinin, salicylic acid metabolic process, cellular response to ethylene stimulus* and *response to auxin*. Conversely, *abi3-6* GO enrichment in the up-regulated gene set included *jasmonic acid*

biosynthetic process, *salicylic acid biosynthetic process*, *gibberellic acid mediated signalling pathway*, *response to auxin* and *abscisic acid metabolic process*.

The ABA content of *C. australe* embryos was high in the early developmental stages, but declined considerably in both axes and cotyledons in the transition from the yellow-green to the yellow pod stage (Chapter 4). However, ABA content increased to very high values in the seed coat upon the transition from the yellow to the brown pod stage. This high ABA content in the seed coat could play a role in delaying germination of the embryo until ideal conditions for germination are met or to spread the risk of premature seedling death. Interestingly, the seed coat starts to peel away from the embryo in mature seeds and this increases markedly once seeds are shed.

Although *C. australe* seeds possess high levels of ABA, this seems to be not enough to induce desiccation tolerance. Possibly other hormones[3], such as gibberellins, are maintaining the active metabolism of the seeds and resist water loss. The genomes of *C. australe* and other DS-seeded species displayed striking genomic alterations associated with hormonal signaling and biosynthesis. I hypothesize that seed desiccation sensitivity is caused, to some extent, by insensitivity to ABA due to these genomic alterations. Among these, we found that *C. australe* lost the gene *GUN5* that has been demonstrated to be a hub for ABA signalling to the nucleus (Jiang et al., 2014).

Altogether, we confirm the opposite roles of ABA and gibberellins in the acquisition of seed desiccation tolerance. However, also genomic and transcriptomic changes of other hormones were observed, e.g. brassinosteroids. These hormones need to be further studied in order to obtain a broader view of the regulation of seed desiccation tolerance/sensitivity.

Perspectives

Understanding seed physiology allows the investigation of climatic events occurring at the plant level, in a given environment. For example, when plant species are found in a location with fire-stimulated seed germination (Keeley, 1987; Keeley et al., 2011), it indicates that this place undergoes periodic fires or at least these occurred in the past. The presence of seeds with physical dormancy indicates an environment subject to frequent stochastic events such as drought (Hudson et al., 2015). Conversely, finding a plant species with DS seeds in a given location, probably means that this is a long-term stable environment with periodic and abundant rainy seasons.

Seed size has been linked to the speed of plant evolution (Igea et al., 2017). Seed size is a very easily measured and comparable trait, which allowed Igea and co-authors

---

[3]Treatment with or quantification of other hormones, besides ABA, have not been performed in this study.

to estimate the rate of phenotypic changes in seed size and their positive correlation with speciation rates. Nonetheless, seed desiccation tolerance and -sensitivity are adaptive traits that correlate significantly with seed size (Daws et al., 2006), but they were not evaluated by Igea et al. (2017), possibly due to the difficulties to measure and standardize them. Thus, seed desiccation sensitivity could have been the major causal driver of plant evolution speed. However, conclusions were made based only on phenotypical correlation analysis without considering the seed physiology.

Correlations unaccompanied by physiological data are a common practice among ecological, evolutionary and other studies. However, interaction with physiological disciplines may have the power to unravel causative effects of the observed correlations. The physiological data regarding seed behaviour associated with botanical and ecological studies may provide information about the fragility of ecosystems. For example, if a hypothetical forest is represented by 100% DS-seeded species, and if this forest would be cut or burned, these species would potentially become extinct in such a location because these seeds do not form seed banks but are maintained as seedlings which are much more sensitive to fire, leaving no propagules left to regenerate (Lamont et al., 1993).

The regeneration of Amazonian floodplain forests has shown to be more sensitive to fire than the non-flooded ones (Flores et al., 2014). The possible cause of this sensitivity could be explained by the desiccation sensitivity of the seeds from many species in this habitat. The authors did not investigate this but the periodic exposure to the semi-aquatic environment could have led the species to evolve DS seeds.

Genetic markers for seed desiccation sensitivity can support ecological and evolutionary studies. With the development of genomic analysis and profound understanding of genomic features associated with adaptive traits, evolution and ecological studies will be substantiated. The genomes bear not only the genetic information necessary for the development of an organism, but also tell the evolutionary history of the species. Pseudogenes, duplicated genes and whole genome duplications, the number of synonymous and non-synonymous mutations and synteny can reveal the evolutionary history of genes and the selective pressure they have suffered (Chapter 4). The anatomical-physiological adaptations associated with these phenotypes should be traceable into the genome and the dN/dS of the genes associated with these phenotypes may reveal when these phenotypes were acquired or how much selective pressure they suffered. Together with other studies, such as paleontology and geology, genomic studies can also provide valuable information to understand climatic events.

References

Beardmore, T. and Whittle, C.-A. (2005) Induction of tolerance to desiccation and cryopreservation in silver maple (Acer saccharinum) embryonic axes. Tree physiology, **25**(8), 965-972.

Bentsink, L., Jowett, J., Hanhart, C.J., Koornneef, M. (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proc Natl Acad Sci U S A, **103**(45), 17042-17047.

Bewley, J.D. (1997) Seed germination and dormancy. The Plant Cell, **9**(7), 1055.

Buitink, J., Vu, B.L., Satour, P., Leprince, O. (2003) The re-establishment of desiccation tolerance in germinated radicles of Medigago truncatula Gaertn. In Seed Science Research: pp 273-286.

Costa, M., Artur, M., Maia, J., Jonkheer, E., Derks, M., Nijveen, H., Williams, B., Mundree, S.G., Jiménez-Gómez, J.M., Hesselink, T. (2017) A footprint of desiccation tolerance in the genome of Xerophyta viscosa. Nature plants, **3**, 17038.

Daws, M.I., Garwood, N.C., Pritchard, H.W. (2006) Prediction of desiccation sensitivity in seeds of woody species: a probabilistic model based on two seed traits and 104 species. Ann Bot, **97**(4), 667-674.

Daws, M.I. and Pritchard, H.W. (2008) The development and limits of freezing tolerance in Acer pseudoplatanus fruits across Europe is dependent on provenance. CryoLetters, **29**(3), 189-198.

Dekkers, B.J., Costa, M.C.D., Maia, J., Bentsink, L., Ligterink, W., Hilhorst, H.W. (2015) Acquisition and loss of desiccation tolerance in seeds: from experimental model to biological relevance. Planta, **241**(3), 563-577.

Dekkers, B.J., He, H., Hanson, J., Willems, L.A., Jamar, D.C., Cueff, G., Rajjou, L., Hilhorst, H.W., Bentsink, L. (2016) The Arabidopsis DELAY OF GERMINATION 1 gene affects ABSCISIC ACID INSENSITIVE 5 (ABI5) expression and genetically interacts with ABI3 during Arabidopsis seed development. The Plant Journal, **85**(4), 451-465.

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

Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J.G., McCourt, P., Samuel, M.A. (2013b) ABI3 controls embryo degreening through Mendel's I locus. Proc Natl Acad Sci U S A, **110**(40), E3888-3894.

Ehlers, K., Bhide, A.S., Tekleyohans, D.G., Wittkop, B., Snowdon, R.J., Becker, A. (2016) The MADS Box Genes ABS, SHP1, and SHP2 Are Essential for the Coordination of Cell Divisions in Ovule and Seed Coat Development and for Endosperm Formation in Arabidopsis thaliana. PloS one, **11**(10), e0165075.

Ellis, R., Hong, T., Roberts, E. (1991) An intermediate category of seed storage behaviour? II. Effects of provenance, immaturity, and imbibition on desiccation-tolerance in coffee. Journal of experimental botany, **42**(5), 653-657.

Farnsworth, E. (2000) The ecology and physiology of viviparous and recalcitrant seeds. Annual Review of Ecology and Systematics, 107-138.

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

Flores, B.M., Piedade, M.-T.F., Nelson, B.W. (2014) Fire disturbance in Amazonian blackwater floodplain forests. Plant Ecology & Diversity, **7**(1-2), 319-327.

Gould, S.J. (1977) *Ontogeny and phylogeny* Harvard University Press.

Groot, S.P. and Karssen, C.M. (1992) Dormancy and germination of abscisic acid-deficient tomato seeds. Plant physiology, **99**(3), 952-958.

Hamilton, K.N., Offord, C.A., Cuneo, P., Deseo, M.A. (2013) A comparative study of seed morphology in relation to desiccation tolerance and other physiological responses in 71 Eastern Australian rainforest species. Plant Species Biology, **28**(1), 51-62.

Hudson, A.R., Ayre, D.J., Ooi, M.K. (2015) Physical dormancy in a changing climate. Seed Science Research, **25**(2), 66-81.

Igea, J., Miller, E.F., Papadopulos, A.S.T., Tanentzap, A.J. (2017) Seed size and its rate of evolution correlate with species diversification across angiosperms. PLOS Biology, **15**(7), e2002792.

Jiang, S.-C., Mei, C., Wang, X.-F., Zhang, D.-P. (2014) A hub for ABA signaling to the nucleus: significance of a cytosolic and nuclear dual-localized PPR protein SOAR1 acting downstream of Mg-chelatase H subunit. Plant signaling & behavior, **9**(11), e972899.

Keeley, J.E. (1987) Role of fire in seed germination of woody taxa in California chaparral. Ecology, **68**(2), 434-443.

Keeley, J.E., Pausas, J.G., Rundel, P.W., Bond, W.J., Bradstock, R.A. (2011) Fire as an evolutionary pressure shaping plant traits. Trends in plant science, **16**(8), 406-411.

Koornneef, M., Hanhart, C.J., Hilhorst, H.W.M., Karssen, C.M. (1989) In Vivo Inhibition of Seed Development and Reserve Protein Accumulation in

Recombinants of Abscisic Acid Biosynthesis and Responsiveness Mutants in Arabidopsis thaliana. Plant Physiol, **90**(2), 463-469.

Lamont, B.B., Witkowski, E., Enright, N. (1993) Post-fire litter microsites: safe for seeds, unsafe for seedlings. Ecology, **74**(2), 501-512.

Leprince, O., van Aelst, A.C., Pritchard, H.W., Murphy, D.J. (1998) Oleosins prevent oil-body coalescence during seed imbibition as suggested by a low-temperature scanning electron microscope study of desiccation-tolerant and -sensitive oilseeds. Planta, **204**(1), 109-119.

Maia, J., Dekkers, B.J., Dolle, M.J., Ligterink, W., Hilhorst, H.W. (2014) Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated Arabidopsis seeds. New Phytol, **203**(1), 81-93.

Maia, J., Dekkers, B.J., Provart, N.J., Ligterink, W., Hilhorst, H.W. (2011) The re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds and its associated transcriptome. PloS one, **6**(12), e29123.

Matias-Hernandez, L., Battaglia, R., Galbiati, F., Rubes, M., Eichenberger, C., Grossniklaus, U., Kater, M.M., Colombo, L. (2010) VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. The Plant Cell, **22**(6), 1702-1715.

Mendes, M.A., Guerra, R.F., Castelnovo, B., Silva-Velazquez, Y., Morandini, P., Manrique, S., Baumann, N., Groß-Hardt, R., Dickinson, H., Colombo, L. (2016) Live and let die: a REM complex promotes fertilization through synergid cell death in Arabidopsis. Development, **143**(15), 2780-2790.

Monke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hahnel, U., Junker, A., Weisshaar, B., Conrad, U., Baumlein, H., Altschmied, L. (2012) Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. Nucleic Acids Res, **40**(17), 8240-8254.

Nambara, E., Keith, K., McCourt, P., Naito, S. (1994) Isolation of an internal deletion mutant of the Arabidopsis thaliana ABI3 gene. Plant and Cell Physiology, **35**(3), 509-513.

Okamoto, M., Tatematsu, K., Matsui, A., Morosawa, T., Ishida, J., Tanaka, M., Endo, T.A., Mochizuki, Y., Toyoda, T., Kamiya, Y., Shinozaki, K., Nambara, E., Seki, M. (2010) Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of Arabidopsis using tiling arrays. Plant J, **62**(1), 39-51.

Olsen, J.L., Rouzé, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., Dattolo, E., De Paoli, E., Dittami, S., Maumus, F. (2016) The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. Nature, **530**(7590), 331-335.

Pammenter, N. and Berjak, P. (2000) Some thoughts on the evolution and ecology of recalcitrant seeds. Plant Species Biology, **15**(2), 153-156.

Pandey, S., Nelson, D.C., Assmann, S.M. (2009) Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. Cell, **136**(1), 136-148.

Raz, V., Bergervoet, J., Koornneef, M. (2001) Sequential steps for developmental arrest in Arabidopsis seeds. Development, **128**(2), 243-252.

Rice, S.H. (1997) The analysis of ontogenetic trajectories: when a change in size or shape is not heterochrony. Proceedings of the National Academy of Sciences, **94**(3), 907-912.

Roberts, E.H. (1973) Predicting the storage life of seeds. In Proceedings.

Rohde, A., De Rycke, R., Beeckman, T., Engler, G., Van Montagu, M., Boerjan, W. (2000) ABI3 affects plastid differentiation in dark-grown Arabidopsis seedlings. The Plant Cell, **12**(1), 35-52.

Roqueiro, G., Facorro, G.B., Huarte, M.G., de Celis, E.R., García, F., Maldonado, S., Maroder, H. (2010) Effects of photooxidation on membrane integrity in Salix nigra seeds. Annals of Botany, mcq067.

Tweddle, J.C., Dickie, J.B., Baskin, C.C., Baskin, J.M. (2003) Ecological aspects of seed desiccation sensitivity. Journal of Ecology, **91**(2), 294-304.

Weiss, D. and Ori, N. (2007) Mechanisms of cross talk between gibberellin and other hormones. Plant physiology, **144**(3), 1240-1246.

Summary

Desiccation tolerance is a mechanism that allows plant species to avoid extinction by spreading their genetic information through time in a robust and safe vault: their seeds. However, species inhabiting long-stable environments in which the conditions are conducive for immediate germination may produce desiccation sensitive (DS) seeds.

In this thesis I was committed to study the relatively unexplored physiology and genetics of these species. First, in Chapter 2 the literature was reviewed concerning the physiology, evolution and molecular aspects of seed desiccation sensitivity.

Based on meta-analysis of available data on the seed's lowest safe water content for diverse species, I found two remarkable limits for dehydration survival. The first limit is ~0.3 g H_2O.(g dry weight)$^{-1}$ water content with no bulk cytoplasmic water present and DS seeds cannot be dried below it. The removal of water from the hydration shell of the molecules starts below ~0.3 g H_2O.(g dry weight)$^{-1}$ and intermediate seeds can safely be dried until ~0.09 g H_2O.(g dry weight)$^{-1}$. Below this limit, desiccation takes place and only desiccation tolerant (DT) seeds are able to dehydrate safely.

In this chapter, also the available knowledge was surveyed to draw a generalized picture of the evolution of DS-seeded species: seed desiccation tolerance is gradually lost in environments where the conditions are conducive to immediate germination. The lack of selective pressure for desiccation tolerance in these environments causes gradual loss of protective and/or regulatory factors related to seed longevity which will ultimately result in desiccation sensitivity. Concomitantly, the selective pressure towards immediate germination suppresses seed developmental arrest and results in suppression of quiescence.

I decided to look further into the molecular mechanisms underlying this evolutionary hypothesis. Noticing a gap of knowledge on transcriptomes of seed development and maturation in relation to desiccation tolerance, I decided to use the Arabidopsis *abi3-6* mutant as a model for seed desiccation sensitivity. In Chapter 3, it was shown that *abi3-6* seeds had increased transcriptional activity that does not fit with a suppression of metabolic activity, generally observed in DT seeds. A clear association was found of its desiccation sensitivity with the absence of expression of genes responsible for dormancy and germination arrest.

In the next step, I asked whether the gene expression patterns during seed development of the Arabidopsis *abi3-6* mutant and other model systems could also be observed in wild DS-seeded species. For this, I explored the genome and transcriptome of *C. australe* DS-seed development in comparison with other DS- and DT-seeded species (Chapter 4). I observed a lack of gene expression related to protection from desiccation damage and to dormancy during late seed development. In

their genomes, specific relevant alterations were found in *C. australe* and other DS-seeded species associated with the regulation of seed development and control of germination that can lead to seed desiccation sensitivity.

Finally, in Chapter 5 I describe a protocol for suppression of gibberellin signalling by paclobutrazol for induction of desiccation tolerance in DS seeds of *Citrus limon* and studied the associated transcriptional changes. The acquisition of desiccation tolerance appeared to be mainly caused by the suppression of metabolism as a consequence of down-regulation of genes related to multiple hormonal pathways, except ABA.

Despite the large variation for seed desiccation sensitivity in nature, common physiological and molecular alterations have been revealed. With this we set the basis for further studies of the underlying mechanisms of seed desiccation sensitivity that may help to develop effective conservation strategies.

Acknowledgement

In 1983, a strong and perseverant woman gave birth to a baby alongside her equally hard-working husband. Both came from very poor families, lost their fathers very early in life, and had to work as children to help to support their mothers and numerous siblings. They both trusted education to be a path for life change, so they motivated me, insisted I study hard, and invested in me and in my education. For that I am immensely grateful. I dedicate this thesis to the two of them (my parents, Agda Silva and Manoel Marques).

Thanks to my education I can see the world with different eyes. It feels as if I have received a very special gift. I can look at nature and see much more than just plants and animals. I can see how complex, special and fragile nature is. In its midst I see evolution, ecology and genetics. When one discovers how complex human evolution is, or how diverse human cultures are, one can't be a racist. When one discovers the beauty, complexity, and fragility of nature, one can't simply destroy it. Thus, thanks to my education, I have learned to respect human diversity and to respect nature. Obviously, one can be respectful without an education but I believe education to be an important tool to deconstruct intolerance and promote empathy among people and nature.

I also want to thank all the family members that took care of me and send me lots of love. Thank you, brother, sister, step-mom, uncles, aunts and cousins. This thesis was also possible because of you. My godfather and godmother were also special in my life. My godmother was/is super relevant in several educational and cultural aspects of my life.

I also want to thank my friends back in Brazil that were very caring and offered me good companionship and entertainment. I had very interesting cultural experiences together with you.

I thoroughly enjoyed the lively scientific environment provided by Wageningen University, the Laboratory of Plant Physiology and the Wageningen Seed Lab. For this, I thank all of the staff members, students, and technicians that bring this great scientific environment to life. I especially thank friends who spent time with me discussing science and life during coffee breaks; those colleagues who gave us Dutch lessons and conveyed a very sweet-welcoming feeling towards the international students on campus.

Off campus, I found an awesome way to detox and de-stress, diminish frustrations and sporadic bad feelings related to work: Playing music. It worked as a spiritual refuge. For this, I am super happy to thank all the friends/musicians that spent time with me jamming, rehearsing and/or performing. I express my gratitude to all the amazing Droevendaal musicians and friends. I have met and played with many interesting, fun, and caring persons at the house 43, including murga, capoeira, choir, jam sessions, living room concerts, Droefes! Wow! I am immensely grateful to the

Curriculum vitae

About the author

Alexandre Correia Silva de Santana Marques was born in Salvador, Bahia, Brazil on the 13th of January 1983. Alexandre was a child with a strong drive towards discoveries. His family tells stories of him opening up several expensive electronic devices to see what was inside. Unfortunately, at that time the discovery process invariably resulted in the destruction of 99,99% of the devices. His father, a master in the domain of inventions and improvisation, was a major motivator of creativity for Alexandre. Although his great project ideas are yet to appear on the market, they were very much a catalyst for Alexandre's creativity. Maybe in the near future you will see: Houses powered by personal nuclear stations; Airplanes with huge parachutes and ships with big buoys in case of failures. Alexandre's father had great passion for nature. Feeding and identifying birds along with gardening were some of his hobbies. These triggered in Alexandre a similar passion for nature.

Alexandre started his bachelor in Biological Sciences from the Federal University of Bahia in 2002. After his scientific initiation in immunology, he felt much more drawn to plant physiology: How do plants work? After his graduation in 2008, Alexandre started to teach biology for high school students. In 2010, he felt the urge to go back to academia and started the master in Plant Genetic Resources at the State University of Feira de Santana. The master's research project aimed at developing priming methods for *Physalis angulata* seeds and to investigate its associated molecular changes.

During his master's program, Alexandre learned that a major Achilles' heel in plant biodiversity conservation was the current impossibility of keeping *ex situ* collections of recalcitrant seeds. Recalcitrance is estimated to be present in almost half of the tropical rainforest species. After receiving his master's degree, Alexandre decided to investigate and possibly develop methods to store these challenging seeds. This present thesis is the result of this challenge that he took on as his research focus to earn his PhD at Wageningen University on the 22nd of May 2018.

List of Publications

Vidigal, D. S., **Marques, A. C. S. S.**, Willems, L. A., Buijs, G., Méndez-Vigo, B., Hilhorst, H. W., Leónie Bentsink F. Xavier Picó, Alonso-Blanco, C. (2016). Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in Arabidopsis thaliana. Plant, cell & environment, 39(8), 1737-1748.

Marques, A. C. S. S., Buijs, G., Ligterink, W., Hilhorst, H. W. (2018). Evolutionary ecophysiology of seed desiccation sensitivity. *Submitted to Functional Plant Biology*

Marques, A. C. S. S., Costa, M. D., Chathuri, U., Jonkheer, E., Zhao, T., Schijlen, E., Derks, M., Nijveen, H., Marcet-Houben, M., Julca, I., Delahaie, J., Schranz, M. E., Gabaldon, T., Pelleties, S., Leprince, O., Ligterink, W., Buitink, J., Hilhorst, H. W., Farrant, J. M. (2018) The recalcitrant genome of *Castanospermum australe*. (*in preparation for submission*).

Marques, A. C. S. S., Kuil, A., Nogueira, R., Nijveen, H., Somi, C., Immink, R. G. H., Ligterink, W., Hilhorst, H. W. (2018). Understanding seed desiccation tolerance through the desiccation sensitive *Arabidopsis thaliana abi3-6* mutant. (*in preparation for submission*).

Marques, A. C. S. S., Nijveen, H., Somi, C., Ligterink, W., Hilhorst, H. W. (2018). Induction of desiccation tolerance in recalcitrant *Citrus limon* seeds. (*in preparation for submission*).

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: **Alexandre Correia Silva de Santana Marques**
Date: **22 May 2018**
Group: **Laboratory of Plant Physiology**
University: **Wageningen University & Research**

1) Start-up phase	*date*
► **First presentation of your project**	
Title: Clinal variation in fitness of seed germination and seedling establishment in Arabidopsis thaliana	16 Aug 2013
► **Writing or rewriting a project proposal**	
Title: Clinal variation in fitness of seed germination and seedling establishment in Arabidopsis thaliana	03 Jan 2014

Subtotal Start-up Phase   *7.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
8th European Plant Science Retreat for PhD Students , Barcelona, Spain	20-23 Jun 2016
EPS PhD Student Days 'Get2gether 2017', Soest, NL	09-10 Feb 2017
► **EPS theme symposia**	
EPS theme 3 Symposium 'Metabolism and adaptation', Wageningen, NL	11 Mar 2014
EPS theme 3 Symposium 'Metabolism and adaptation', Amsterdam, NL	22 Feb 2016
EPS theme 3 Symposium 'Metabolism and adaptation', Wageningen, NL	14 Mar 2017
EPS Theme 1 Symposium 'Developmental Biology of Plants', Leiden, NL	28 Feb 2017
► **National meetings (e.g. Lunteren days) and other National Platforms**	
Annual meeting 'Experimental Plan Sciences', Lunteren, NL	14-15 Apr 2014
Annual meeting 'Experimental Plan Sciences', Lunteren, NL	13-14 Apr 2015
Annual meeting 'Experimental Plan Sciences', Lunteren, NL	11-12 Apr 2016
► **Seminars (series), workshops and symposia**	
2nd Dutch Seed Symposium, Wageningen, NL	01 Oct 2013
3rd Dutch Seed Symposium, Wageningen, NL	07 Oct 2014
4th Dutch Seed Symposium 6 October 2015, Wageningen	06 Oct 2015
5th Dutch Seed Symposium, Wageningen, NL	04-05 Oct 2016
► **International symposia and congresses**	
4th ISSS workshop on Molecular Aspects of Seed Dormancy and Germination - Paris, France	09-12 Jul 2013
7th International Workshop on Desiccation Sensitivity and Tolerance across Life forms, Aquila Private Game Reserve, South Africa	11-15 Jan 2016
► **Presentations**	
Poster: ALW meeting 'Experimental Plan Sciences', Lunteren	14-15 Apr 2014
Talk and Poster: PPH PhD Trip	22 Apr- 01 May 2015
Talk: 7th International Workshop on Desiccation Sensitivity and Tolerance across Life forms, Aquila Talk : Private Game Reserve, South Africa	11-15 Jan 2016
Poster: ALW meeting 'Experimental Plan Sciences', Lunteren	11-12 Apr 2016

Talk: 8th European Plant Science Retreat for PhD Students, Barcelona, Spain 2016	20-23 Jun 2016
Talk: Genetics Seminar (GSS): 'The Castanospermum australe genome', Wageningen, NL	14 Sep 2016
► **Excursions**	
PPH PhD Trip to NL-Germany-Switzerland (Rijkzwaan, MPIPZ, Biocenter, MPIDB, ETH, UZ, Syngenta)	22 Apr- 01 May 2015

Subtotal Scientific Exposure   *17.8 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
The Power of RNA-seq	16-18 Dec 2013
Bioinformatics - A User's Approach	24-28 Aug 2014
System biology: Statistical analysis for -omics data analysis	15-19 Dec 2015
► **Journal club**	
Literature discussion in Plant Physiology	2013-2017

Subtotal In-Depth Studies   *6.9 credits**

4) Personal development	*date*
► **Skill training courses**	
Information Literacy including EndNote Introduction	15-16 Apr 2014
Social Dutch 1	Sep-Nov 2015
Scientific Writing	Oct-Dec 2015
Mobilising your - scientific - network	17 & 23 Oct 2014

Subtotal Personal Development   *5.4 credits**

TOTAL NUMBER OF CREDIT POINTS*	**37.6**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

* A credit represents a normative study load of 28 hours of study.