Identification and Characterization of Seed Longevity Genes in Barley (*Hordeum vulgare*)

vorgelegt von

Dorothee Charlotte Wozny

aus Heidelberg

Köln, 2017

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln (MPIPZ) in der Abteilung für Pflanzenzüchtung und Genetik (Prof. Dr. M. Koornneef) angefertigt.

Table of Contents

Summary

Plant diversity is being lost at an unprecedented rate. This makes the conservation of plant species to a global task of high priority. Seed conservation, both *in situ* and *ex situ*, is one of the best strategies for the conservation of plant diversity. For germplasm preservation in seed banks, seed longevity is of particular importance. Seed longevity constitutes a measure for the period a seed remains viable when stored under the optimum conditions for that species and/or cultivar. Seed longevity is induced during seed maturation and is involved in retaining germination ability that is gradually lost as a result of ageing. Hence, seed longevity is a major parameter controlling seed quality and an important survival trait in the soil seed bank helping the seed to adapt to changing environmental conditions.

Until now, little is known about the genetic basis of differences in seed longevity because this trait is affected by numerous environmental effects during seed formation, harvest, and storage, and genetic variation between and within species is probably controlled by several genes. The work presented in this thesis aims to identify and characterize seed longevity loci/genes in barley (*Hordeum vulgare*) by combining a quantitative genetics approach with 'omics' technologies. This work is based on a previous Quantitative Trait Loci (QTL) study in recombinant inbred lines (RILs), derived from the short-lived Ethiopian spring barley landrace L94 and the long-lived Argentinian spring barley landrace Cebada Capa (CC). In total, four putative QTLs for seed longevity were identified on chromosomes 1 and 2, and near isogenic lines (NILs) were generated (Adimargono et al., in preparation). Using these near isogenic lines, the four putative QTLs could be confirmed and further mapping of Cebada Capa introgressions was achieved in the so-called L94 NILs using RNA-seq. For candidate and downstream target gene identification, a combined transcriptome and proteome analysis was performed. The analysis of mature, non-aged seeds of the two parental lines and the L94 NILs by RNA-seq and total seed proteomic profiling identified two putative candidate genes and one possible downstream target gene. The (NADP)-dependent Malic Enzyme (NADP-ME) AK248526.1 and the UDP-glycosyltransferase MLOC_11661.1 were chosen as possible candidate genes for two of the four seed longevity QTLs, while the second NADP-ME MLOC_35785.1 was chosen as a possible downstream target. To validate these putative candidate and downstream target genes, a T-DNA knock-out line of a homologous *Arabidopsis* gene, shown to have a seed longevity phenotype by Nguyen (2014), was complemented with the respective Cebada Capa and L94 alleles under the expression of different promoters. Both the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 were able to rescue the *nadp-me1* seed longevity phenotype when being expressed under the control of the 35S- or the Ubi-promoter. In the latter case, only the CC-specific coding sequence had an effect. Since the identified candidate NADP-ME MLOC_35785.1 may affect the redox status in normal and deteriorating seeds, the corresponding *Arabidopsis* mutant *nadp-me1* was combined with the redox reporter roGFP2 to visualize the redox potential in seeds using confocal microscopy. First transformant generation (T_1) lines, showing strong fluorescence in the cytosol, could be generated in this study and allow the measurement of the $NADP^+$/NADPH redox potential in future studies.

Zusammenfassung

Die Pflanzenvielfalt nimmt in einem bislang noch nicht da gewesenen Ausmaß ab. Aufgrund dessen stellt der Erhalt von Pflanzenarten eine globale Aufgabe von hoher Priorität dar. Samenkonservierung, sowohl *in situ* als auch *ex situ*, stellt eine der besten Strategien zum Erhalt der Pflanzenvielfalt dar. Für den Erhalt von Keimgewebe ist die Samenlanglebigkeit besonders wichtig. Die Samenlanglebigkeit stellt ein Maß für diejenige Zeitspanne dar, in der der Samen, unter den optimalen Bedingungen für die Art oder Sorte gelagert, lebensfähig ist. Die Samenlanglebigkeit wird während der Samenreife induziert und ist am Erhalt der Keimungsfähigkeit beteiligt, die graduell aufgrund des Alterungsprozesses verloren geht. Somit ist die Samenlanglebigkeit eine Hauptkomponente der Samenqualität und auch ein wichtiges Überlebensmerkmal der Samenbank in der Erde, da sie dem Samen hilft, sich an wechselnde Umweltbedingungen anzupassen.

Bis jetzt ist wenig über die genetische Grundlage der beobachteten Unterschiede in der Samenlanglebigkeit bekannt, da dieses Merkmal von einer Reihe von Umwelteinwirkungen während der Samenbildung, der Ernte und der Lagerung beinflusst wird. Die genetische Variation zwischen und innerhalb von Pflanzenarten wird zudem von mehreren Genen kontrolliert. Die vorliegende Doktorarbeit hatte das Ziel Samenlanglebigkeits-Loci und Gene in Gerste (*Hordeum vulgare*) anhand der Kombination von quantitativer Genetik und 'omics'-Technologien zu identifizieren und zu charakterisieren. Diese Arbeit basiert auf einer frühreren *Quantitative Trait Locus* (QTL)-Studie in rekombinanten Inzuchtlinien (RILs), die von der Kreuzung der kurzlebigen äthiopischen Gerstenrasse L94 mit der langlebigen argentinischen Gerstenrasse Cebada Capa (CC) stammen. Insgesamt wurden vier mutmaßliche Samenlanglebigkeits-QTLs auf den Chromsomen 1 und 2 identifiziert und nahezu isogene Linien (NILs) generiert (Adimargono et al., in Vorbereitung). Mithilfe dieser nahezu isogenen Linien konnten die vier mutmaßlichen QTLs bestätigt werden und die Kartierung von CC-Introgressionen in den so genannten L94 NILs mithilfe von RNA-seq erreicht werden. Zur Identifizierung von Kandidatengenen und nachgeschalteten Genen wurde eine kombinierte Transkriptom- und Proteomanalyse durchgeführt. Die Analyse von reifen, nicht gealterten Samen der zwei Elternlinien und der L94 NILs mittels RNA-seq und Proteomanalyse führte zur Identifizierung von zwei möglichen Kandidatengenen und einem möglichen nachgeschalteten Gen. Das NADP-abhängige Malatenzym (NADP-ME) AK248526.1 und die UDP-Glykosyltransferase MLOC_11661.1 wurden als mögliche Kandidatengene für zwei der vier Samenlanglebigkeits-QTLs ausgewählt. Das zweite NADP-ME MLOC_35785.1 wurde als mögliches nachgeschaltetes Gen ausgewählt. Für die Validierung dieser Kandidatengene und nachgeschalteten Gene wurde eine T-DNA *knock-out*-Linie eines homologen *Arabidopsis*-Gens, die einen Samenlanglebigkeitsphänotyp zeigte (Nguyen, 2014), mit den entsprechenden Cebada Capa und L94 Allelen unter der Kontrollen von verschiedenen Promotoren komplementiert. Sowohl das NADP-ME MLOC_35785.1 als auch die UDP-Glykosyltransferase MLOC_11661.1 konnten den Samenlanglebigkeitsphänotyp von *nadp-me1* aufheben/retten, wenn sie unter dem 35S- und Ubi-Promotor exprimiert wurden. Im letzten Fall hatte nur die CC-spezifische Kodierungssequenz einen Effekt. Da das identifizierte NADP-ME MLOC_35785.1 den Redoxstatus in

normalen und alternden Samen beeinflussen könnte, wurde die entsprechende *Arabidopsis*-Mutante *nadp-me1* mit dem Redoxreporter roGFP2 kombiniert, um das Redoxpotental in Samen mithilfe von Konfokalmikroskopie visualisieren zu können. Transformanten der ersten Generation (T_1), die ein starkes Fluoreszenzsignal im Zytosol zeigen, konnten in dieser Arbeit generiert werden und ermöglichen die Messung des $NADP^+$/NADPH-Redoxpotentials in zukünftigen Studien.

Chapter 1

General Introduction

Introduction

Practical importance of seed longevity

Plant diversity is being lost at an unprecedented rate. This makes the conservation of plant species to a global task of high priority. In the early 20th century, the realization of the danger of extinction of plant genetic resources led to the establishment of gene banks around the world (Linington and Pritchard, 2001). Seed conservation, both *in situ* and *ex situ*, is one of the best strategies for the conservation of plant diversity. It is estimated that 90% of the six million accessions of plant genetic resources globally held reside in seed banks (Li and Pritchard, 2009). In higher plants, the seed is the main dispersal unit allowing the plant to survive the period between seed maturation and the establishment of the next generation. For that purpose, the seed, mainly in a dry state, is well equipped to sustain extended periods of unfavorable conditions. According to their storage behavior, seeds were classified as orthodox and recalcitrant (Roberts, 1973). Orthodox seeds survive drying and/or freezing during *ex situ* conservation. Recalcitrant seeds undergo little, or no maturation drying and remain desiccation-sensitive both during development and after they are shed. Many monocotyledonous species, which comprise the majority of agricultural plants and include the major grains such as rice, wheat, maize and barley, have orthodox seeds. For germplasm preservation in seed banks, seed longevity is of particular importance.

**Seed longevity and seed quality characteristics**

Seed longevity constitutes a measure for the period a seed remains viable when stored under the optimum environmental conditions for that species and/or cultivar (Barton, 1961). This period is determined by pre-storage and storage conditions, the genetic and physiological storage potential of the seed and is affected by adverse events such as too high or too low temperatures or damage prior to or during storage (Roberts, 1961). The long term storage of seeds, especially under unfavorable conditions, leads to loss of viability. Viability stands for the ability of the seed to germinate and the ability of the seedling to establish itself in the environment in which the seed finds itself (Bradbeer, 1988). Seed longevity has been related to various seed properties such as color, weight and membrane composition. The correlation between these traits and seed longevity is often species- and in some cases even variety-specific (McDonald, 1999). Seeds represent a stage of the life cycle in which plants experience particularly high levels of genotoxic stress leading to genome instability (Waterworth et al., 2011). Environmental stresses including ultraviolet B, ozone, desiccation and rehydration, and air and soil pollutants cause a range of DNA damage products including single-strand and double-strand DNA breaks (SSBs and DSBs). The DNA checkpoint kinases ATM (Ataxia telangiectasia mutated) and ATR (Ataxia telangiectasia and Rad3-related) play crucial roles in the maintenance of genome stability, safeguarding cellular survival and the faithful transmission of genetic information (Waterworth et al., 2016). Seed vigor is defined as the total sum of those properties of the seed which determine the level of activity and performance of the seed or seed lot during germination and seedling emergence (International Seed Testing Association, ISTA). The progressive loss of vigor due to the deterioration of the structures and functions of the seed over time

is termed seed ageing (Mohamed-Yasseen, 1991). The rate of ageing is strongly influenced by environmental and genetic factors such as storage temperature, seed moisture content, and seed quality (Walters, 1998; Walters et al., 2005). Seed quality can be reduced on the mother plant due to adverse environmental conditions, premature germination (Coolbear, 1995) and pathogens (McGee, 2000). Maternal effects are a particular kind of phenotypic plasticity that encompasses the developmental contingency of later life stages upon earlier stages (Donohue, 2009). Maternal environmental effects refer to the particular phenomenon in which the external ecological environment of the maternal parent influences the phenotype of its progeny. They influence natural selection on germination, the expression of genetic variation for germination and even the genes involved in germination. Maternal effects on seed traits can even influence generation time and projected population growth rates. Seed performance after dispersal is highly dependent on parental environmental cues, especially during seed formation and maturation (He et al., 2014). Comparative analysis clearly indicated that temperature plays the most dominant role in both plant and seed performance, whereas light has a prominent impact on plant traits. In comparison to temperature and light, nitrate mildly affected some of the plant and seed traits while phosphate had even less influence on those traits (He et al., 2014).

Deterioration-related damages and defense mechanisms

As shown in figure 1, seeds deteriorate during storage, lose vigor, become more sensitive to stress conditions and ultimately become unable to germinate (Rajjou and Debeaujon, 2008). The long term deterioration which occurs during storage includes membrane and genetic damage (Waterworth et al., 2011), changes in respiratory activity and enzymes and protein damage (Coolbear, 1995; McDonald, 1999). Deterioration also involves physical and chemical changes including the disruption of intracellular integrity, decreased activities of enzymes, lipid peroxidation and non-enzymatic reactions (Wettlaufer and Leopold, 1991; Priestley, 1986). It has been proposed that deterioration partly originates from oxidative reactions during storage (Roqueiro et al., 2010). Hence, the generation of oxidative stress and the defense against oxidative stress or its consequences may play an important role in seed longevity. Dry mature seeds are equipped with antioxidant systems that consist of enzymes and non-enzymatic antioxidants. Superoxide radicals are converted to hydrogen peroxide (H_2O_2) by superoxide dismutase (SOD) (Grene, 2002). H_2O_2 is then neutralized by catalase (CAT) and ascorbate peroxidase (APX). CAT turns H_2O_2 directly into water and oxygen, while APX catalyzes the reaction between ascorbic acid and H_2O_2 to form dehydroascorbate and water (Blokhina et al., 2003). Removal of H_2O_2 by APX requires the involvement of glutathione (GSH) as a part of the ascorbate-glutathione cycle (Nocter and Foyer, 1998). In order to optimize seed life span and to limit the rate of decay, nature has evolved complex systems of protection, detoxification and repair. These defense mechanisms are innate to the seed and define seed quality. A summary of such mechanisms is shown in figure 1.
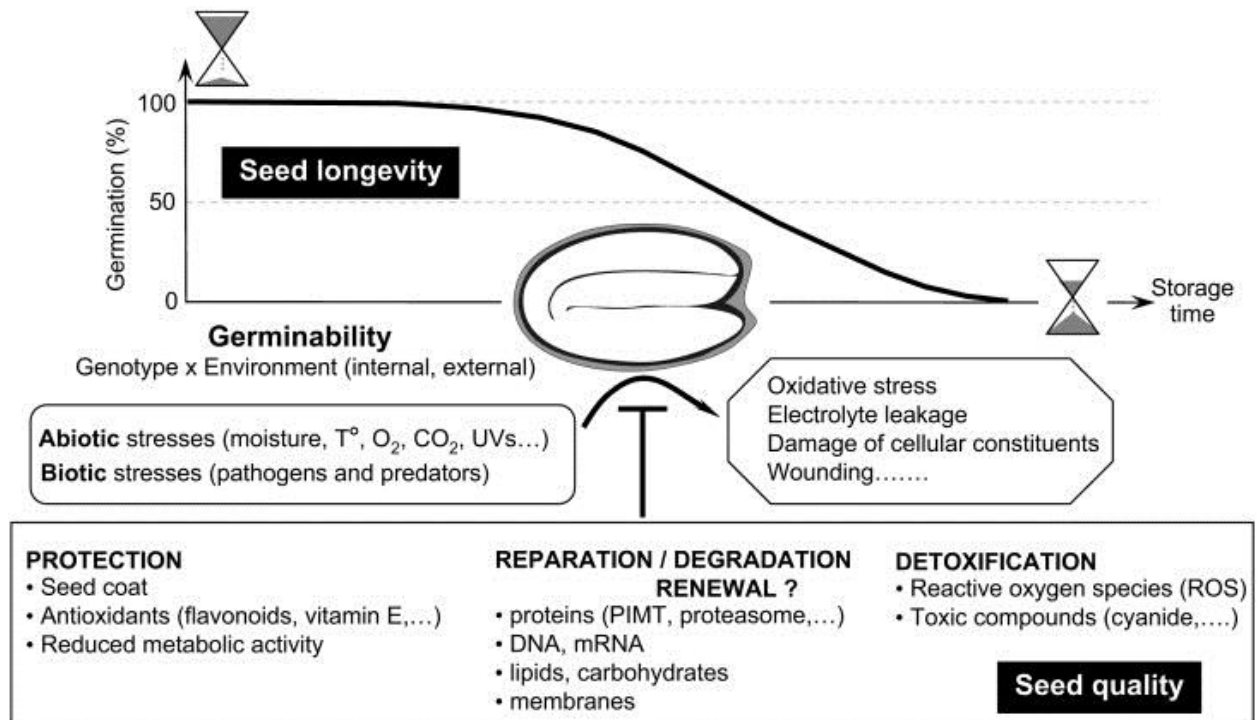
Figure 1. Schematic presentation of the main parameters determining seed longevity (Rajjou and Debeaujon, 2008)

Seed deterioration during storage in soil or in gene banks is generally appreciated as germinability in function of storage time. It results from the interaction between endogenous parameters defining seed quality and environmental parameters such as biotic and abiotic stresses during storage. The arrow indicates induction and the blunt end stands for repression.

Seed longevity measurement

In the laboratory, natural ageing occurs when seeds are stored at ambient conditions for extended periods of time. In order to overcome the long waiting times due to slow natural ageing, artificial ageing methods have been developed and applied to study seed longevity in various species. High relative humidity and high temperature are used in accelerated ageing (AA) and controlled deterioration (CD) tests to mimic seed ageing (Tesnier et al., 2002). While in AA tests, seeds are treated with high humidity and temperature at the same time, seeds are equilibrated in humidity prior to temperature application in CD tests (ISTA). After artificial ageing, seed viability is evaluated by a germination assay. Proteome analysis of *Arabidopsis thaliana* seeds revealed common features between seed ageing with CD (85% relative humidity (RH) at 40°C for up to 7 days) and conventional (up to 11 years at 5°C) ageing (Rajjou et al., 2008).

CHAPTER 1

Genetic variation of seed longevity mutants

The role of important factors influencing seed longevity has been analyzed by using *Arabidopsis* mutants and transgenic lines. The majority of mutants with known effects on seed longevity are seed developmental mutants. Seed maturation is genetically controlled by four major regulators, *ABSCISIC ACID INSENSITIVE3* (*ABI3*), *LEAFY COTYLEDON1* and 2 (*LEC1, LEC2*) and *FUSCA3* (*FUS3*) (Raz et al., 2001). These four factors interact in a network in which *LEC1* and *LEC2* positively regulate *ABI3* and *FUS3*. *ABI3* and *FUS3* positively regulate themselves and each other and form feedback loops essential for their sustained and uniform expression in the embryo (Kroj et al., 2003; Kagaya et al., 2005; To et al., 2006). Mutations in these key regulators lead to rapid loss of viability upon storage. *Abi3*, *lec1* and *fus3* mutants show reduced seed longevity (Ooms et al., 1993; Clerkx et al., 2004a; Sugliani et al., 2009). Seed coat mutants also display a seed longevity phenotype. The seed coat, or testa, acts as a structural barrier to protect both the embryo and seed reserves from biotic and abiotic stresses. Seed coat mutants consist of two major groups. One group, affected in flavonoid pigmentation, is represented by the *transparent testa* (*tt*) and *transparent testa glabra* (*ttg*) mutants (Bürger, 1971; Koornneef, 1990; Shirley et al., 1995). The second group is represented by mutants affected in testa structure. The *aberrant testa shape* (*ats*) mutant ovules lack two cell layers of the integuments and as a result produce heart-shaped mature seeds (Léon-Kloosterziel et al., 1994). Both the *tt* and the *ats* mutants display considerably reduced seed longevity (Debeaujon et al., 2000). In *Arabidopsis*, mutants deficient in flavonoid biosynthesis exhibit a 60% higher level of lipid peroxidation than wild-type plants when exposed to short wave ultraviolet B (Landry et al., 1995). Therefore, it is very likely that seed flavonoids play a protective role against solute leakage, imbibition damage, and oxidative stress. The results showing that pigmentation mutants exhibit more deterioration than their wild-types, are in agreement with this hypothesis. Moreover, the poor storability of *ats* demonstrates that a drastic structural defect can also be very detrimental for seed viability (Debeaujon et al., 2000). Mutations in protection and repair systems that prevent seed vigor loss also lead to decreased seed longevity. *Arabidopsis* mutants affected in vitamin E (lipophilic antioxidant) biosynthesis, *vte1* and *vte2*, exhibited significantly reduced seed longevity (Sattler et al., 2004). The *vte2* mutant is defective in homogentisate phytyl transferase (HPT) and lacks all tocopherols and pathway intermediates which leads to the massive and uncontrolled peroxidation of storage lipids (Sattler et al., 2004). The *vte1* mutant is defective in tocopherol cyclase activity and deficient in all tocopherols, but unlike *vte2*, it accumulates the redox-active biosynthetic intermediate 2,3-dimethyl-6-phytyl-1,4-benzoquinol (DMPBQ) (Sattler et al., 2003). Studies with *vte1* and *vte2* mutants suggest that a primary function of tocopherols is to control non-enzymatic lipid oxidation, especially during seed storage and early germination, and also probably in photosynthetic tissues (Sattler et al., 2004). Proteins and enzymes are also described as factors that may determine seed longevity. Heat stress transcription factor (HSF) over-accumulating seeds of transgenic *Arabidopsis* display enhanced accumulation of heat stress protein and improved tolerance to ageing (Prieto-Dapena et al., 2006). Protein repair appears to play a key role in long-term survival of seeds in the dry state. Important factors might also be structural, influencing the effectiveness by which membranes and other macromolecules are protected (Wolkers et al. 1998). A role for sugars has been suggested and a low ratio of sucrose to

oligosaccharides was found to correlate with long-term longevity of seeds (Obendorf, 1997). Other mechanisms important in dry seeds are the accumulation of amphiphilic molecules such as late embryogenic abundant (LEA) proteins, fructans and the successful formation of a biological glass protecting macromolecules and structural components (Hoekstra et al., 2001; Oliver et al., 2001).

Identification of quantitative trait loci for seed longevity in different species

Seed longevity variation within species behaves as a quantitative trait and genetic studies in rice (*Oryza sativa*) (Miura et al., 2002), *Arabidopsis thaliana* (Clerkx et al., 2004b), soybean (Singh et al., 2008), barley (Nagel et al., 2009), *Aegilops tauschii* (Landjeva et al., 2010) and oilseed rape (Nagel et al., 2011) showed that variation in seed longevity is controlled by several genetic factors, allowing the detection of Quantitative Trait Loci (QTL). QTLs for seed longevity have been identified following both natural ageing and artificial ageing. Using natural ageing, QTLs were identified in *Arabidopsis thaliana* (Nguyen et al., 2012), lettuce (Schwember and Bradford, 2010) and rice (Sasaki et al., 2005). After artificial ageing imposed by a CD test, seed ageing QTLs have been identified in *Arabidopsis thaliana* (Bentsink et al., 2000; Clerkx et al., 2004b; Nguyen et al., 2012), rice (Miura et al., 2002), and wheat (Landjeva et al., 2010). Identification of seed longevity QTLs in barley and possible candidate genes responsible for increased seed life using AA and/or CD tests have been reported by Nagel et al. in 2009. For the genetic studies three different doubled haploid mapping populations were investigated. (1) The ˈSteptoeˈ × ˈMorexˈ (S × M) population, consisting of 150 doubled haploid (DH) lines developed by pollinating the F1 hybrid of the cultivar (cv.) ˈSteptoeˈ × cv. ˈMorexˈ cross with *Hordeum bulbosum* (Kleinhofs et al., 1993). ˈSteptoeˈ is a high-yielding six-rowed feed-type barley (Muir and Nilan, 1973) and ˈMorexˈ, a six-rowed cultivar used as the American malting industry standard (Rasmusson and Wilcoxson, 1979). (2) The OWB mapping population, a set of 94 spring barley DH lines developed again by the *Hordeum bulbosum* method (Costa et al. 2001). (3) The winter barley population W766, resulting from a cross between the two-rowed cultivar ˈAngoraˈ and the accession ˈW704/137ˈ, a two-rowed, short-stemmed, dense-eared winter barley of Japanese origin. Applying both AA and CD tests, a major QTL for longevity was identified on chromosome 5H in the S x M population. For OWB, three QTLs on chromosomes 2H, 5H and 7H were found, whereas W766 analysis yielded one QTL on chromosome 7H. The gene determining naked caryopsis (*nud*) was identified as the candidate gene for the QTL on chromosome 7H in the W766 and OWB populations. Analyzing the OWB population using an AA test, one significant QTL was detected in the distal region of chromosome 2H associated with the gene locus *Zeocriton 1 (Zeo1)*. *Zeo1* is responsible for very compact spikes with long awns and reduced fertility. The S × M population is segregating for the gene *Aleurain* (*Ale*), related to the QTL detected in the proximal region of the long arm of chromosome 5H. *Aleurain* is a barley vacuolar thiol protease. The aleurain gene was synthesized from gibberellic acid-stimulated aleurone cell mRNA (Rogers et al., 1985). The expression is regulated by the plant hormones gibberellic acid and abscisic acid, well-known to be included in germination process. Using natural variation and fine mapping in the crop *Brassica oleracea*, Morris et al. (2016) showed that allelic variation at three loci influences the key vigor trait of rapid germination. Two candidate genes were identified at the principal *SPEED OF GERMINATION* QTL

(*SOG1*). One gene, *BoLCVIG2,* is a homologue of the *Arabidopsis* alternative-splicing regulator PTB1. The other gene, *BoLCVIG1,* was unknown, but different alleles had different splice forms that were coincident with altered abscisic acid (ABA) sensitivity. Furthermore, another QTL, *REDUCED ABSCISIC ACID 1* (*RABA1*), was found to influence ABA content and Morris et al. (2016) provide evidence that this results from the activity of a homologue of the *Arabidopsis* ABA catabolic gene *CYP707A2* at this locus. Lines containing beneficial alleles of these three genes had greater seed vigor suggesting a mechanism in which both seed ABA content and sensitivity to it determines the speed of germination (Morris et al., 2016).

The relationship between seed dormancy and seed longevity

To optimize germination and sustain extended periods of unfavorable conditions, the seed, mainly in a dry state, enters a dormant state. Physiological dormancy can be most simply described as a programmed state that restricts the set of environmental conditions under which a seed will germinate (Bewley, 1997). In plants, the association of seed longevity with seed dormancy has not been studied in detail. In *Arabidopsis*, Clerkx et al. (2004b) studied various mutants with mutations in defined developmental or biochemical pathways. Reduced longevity was shown by mutants with *abscisic acid insensitive3* (*abi3*) and *abscisic acid deficient1* (*aba1*) mutations that are also related to the seed dormancy phenotype of those mutants. Therefore, a positive association between longevity and dormancy can be hypothesized. Studies in rice by Miura et al. (2002) led to the conclusion that these two traits are controlled by different genetic factors. A QTL mapping was performed for dormancy on the same population for which seed longevity traits had been investigated. Whereas the loci for seed longevity were located on chromosomes 2, 4 and 9, QTLs for dormancy were identified on chromosomes 1, 3, 5, 7 and 11. In *Arabidopsis thaliana*, a negative correlation was observed (Nguyen et al., 2012). Integrated QTL analyses for seed longevity in six recombinant inbred line (RIL) populations revealed five *GERMINATION ABILITY AFTER STORAGE* (*GAAS1* to *GAAS5*). GAAS loci were found to be co-located with seed dormancy loci, i.e. *DELAY OF GERMINATION* (*DOG*) loci. Interestingly, lower storability levels correlated with higher seed dormancy levels, and conversely, better storability correlated with lower seed dormancy. Detailed analysis on the co-located *GAAS5* and *DOG1* QTLs revealed that the *DOG1*-Cape Verde Islands allele reduces seed longevity and increases seed dormancy at the same time. The observation of a negative correlation between dormancy and longevity strongly suggests that seeds are able to extend their life span either by dormancy or by an active longevity mechanism. Selection of the different mechanisms could be based on the natural environments in which the seeds are dispersed. Dry environments could result in active longevity mechanisms and humid environments could result in dormancy cycling during which ageing damage may be prevented or repaired. The presence of loci that either improve longevity or increase seed dormancy within one accession would allow adaptive plasticity, resulting in the expression of the optimal phenotype over a range of environments.

Thesis objective

The main aim of this study was to identify and characterize seed longevity loci/genes in barley, which were identified in a previous QTL study using recombinant inbred lines derived from a cross between the Ethiopian spring barley landrace L94 and the line 116-5 (Adimargono et al., in preparation). The line 116-5 was derived from a cross between the short-lived Ethiopian spring barley landrace L94 and the long-lived Argentinian spring barley landrace Cebada Capa. In total, four putative QTLs for seed longevity were identified on chromosomes 1 and 2, and near isogenic lines (NILs) were generated (Adimargono et al., in preparation). Chapter 2 describes the confirmation of these four putative QTLs and further mapping of Cebada Capa introgressions using RNA-sequencing (RNA-seq) in the so-called L94 NIL mapping population. To better understand the complex biochemical and molecular processes underlying seed longevity, transcriptome, proteome and metabolome analyses were performed on the parental lines L94 and Cebada Capa, and the L94 NILs. As described in chapter 3, a total seed proteome profiling was performed to identify proteins that may affect the observed differences in seed longevity between L94 and Cebada Capa. Making use of the recent improvement of the barley genomics infrastructure, an RNA-seq analysis was conducted to identify differentially expressed candidate genes and possible downstream targets. Lastly, mass spectrometry (MS)-based metabolomics was applied to gain insight into biochemical processes affected in seed longevity. Chapter 4 describes the validation of the in chapter 3 identified candidate and downstream targets. For this purpose, the Cebada Capa and L94 alleles were cloned in *Arabidopsis thaliana* to investigate whether they complement the seed longevity phenotype in corresponding *Arabidopsis* mutants. Since one of the identified candidates may affect the redox status in normal and deteriorating seeds, the corresponding *Arabidopsis* mutant was combined with a redox reporter to visualize the redox potential in seeds using confocal microscopy.

References

Barton, L.V. (1961), Seed preservation and longevity. Leonhard Hill Ltd, London

Bentsink, L. et al. (2000), Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. Plant Physiol. 124: 1595-1604

Bewley, J. D. (1997). Seed germination and dormancy. Plant Cell 9: 1055-1066

Blokhina, O. et al. (2003), Antioxidants, oxidative damage and oxygen deprivation stress: a review. Ann. Bot. 91: 179-194

Bradbeer, J. W. (1988), Seed Dormancy and Germination, Blackie and Son Ltd

Bürger, D. (1971), Die morphologischen Mutanten des Göttinger *Arabidopsis*-Sortiment, einschliesslich der Mutanten mit abweichender Samenfarbe. Arabidopsis Inf. Serv. 8: 36-42

Clerkx, E.J.M. et al. (2004a), Genetic differences in seed longevity of various *Arabidopsis* mutants. Physiologia Plantarum 121: 448-461

Clerkx, E.J.M. et al. (2004b), 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 Physiol. 135: 432-443

Coolbear, P. (1995), Mechanisms of seed deterioration. In: Basra, A.S. (Ed.) Seed Quality: Basic mechanisms and agricultural implications. Food Product Press, NewYork, 223-277

Costa, J.M. et al. (2001), Molecular mapping of the Oregon Wolfe Barleys: a phenotypically polymorphic doubled-haploid population. Theor. Appl. Genet. 103: 415-424

Debeaujon, I. et al. (2000), Influence of the Testa on Seed Dormancy, Germination, and Longevity in *Arabidopsis*, Plant Physiol. 122: 403-414

Donohue, K. (2009), Completing the cycle: maternal effects as the missing link in plant life histories. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364: 1059-1074

Grene, R. (2002), Oxidative stress and acclimation mechanisms in plants. *Arabidopsis* Book 1: e0036

He, H. et al. (2014), Interaction between parental environment and genotype affects plant and seed performance in *Arabidopsis*. J. Exp. Bot. 65: 6603-6615

Hoekstra, F.A. et al. (2001), Mechanisms of plant desiccation tolerance. Trends Plant Sci. 6: 431-438

ISTA. International rules for seed testing (1999), Seed Sci. Technol. 27

Kagaya, Y. et al. (2005), LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiol. 46: 399-406

Kleinhofs, A. et al. (1993), A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. 86: 705-712

Koornneef, M. (1990), Mutations affecting the testa colour in *Arabidopsis*. Arabidopsis Inf. Serv.27: 1-4

Kroj, T. et al. (2003), Regulation of storage protein gene expression in *Arabidopsis*. Development 130: 6065-6073

Landjeva, S. et al. (2010), Genetic mapping within the wheat D genome reveals QTLs for germination, seed vigour and longevity, and early seedling growth. Euphytica 171: 129-143

Landry, L.G. et al. (1995), *Arabidopsis* mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. Plant Physiol. 109: 1159-1166

Léon-Kloosterziel, K.M. et al. (1994), A seed shape mutant of *Arabidopsis* that is affected in integument development. Plant Cell. 6: 385-392

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

Linington, S.H. and Pritchard, H.W. (2001), Gene banks. In Levin, S.A. (Ed.) Encyclopaedia of Biodiversity. Vol 3. Academic Press, San Diego, 164-181

McDonald, M.B. (1999), Seed deterioration: physiology, repair and assessment. Seed Science and Technology 27: 177-237

McGee, D.C. (2000), Pathology of seed deterioration. In: Moore, S.H., Yaklich, R.W. (Ed.) Genetic Improvement of seed quality. Crop Science Society of America, Anaheim, Madison, 21-37

Miura, K. et al. (2002), Mapping quantitative trait loci controlling seed longevity in rice (*Oryza sativa* L.), Theor. Appl. Genet. 104: 981-986

Mohamed-Yasseen, Y. (1991), Onion seed aging and plant regeneration *in vitro*, Ph.D. diss., University of Illinois, Urbana

Morris, K. et al. (2016), Trait to gene analysis reveals that allelic variation in three genes determines seed vigour. New Phytol. 212: 964-976

Muir, C.E. and Nilan, R.A. (1973), Registration of Steptoe barley. Crop Sci. 13: 770

Nagel, M. et al. (2009), Seed conservation in *ex situ* genebanks - genetic studies on longevity in barley. Euphytica 170: 5-14

Nagel, M. et al. (2011), Seed longevity in oilseed rape (*Brassica napus* L.) - genetic variation and QTL mapping. Plant Genetic Resources 9: 260-263

Nguyen, T.P. et al. (2012), Natural variation for seed longevity and seed dormancy are negatively correlated in *Arabidopsis*. Plant Physiol. 160: 2083-2092

Nocter, G. and Foyer C.H. (1998), Ascorbate and glutathione: keeping active oxygen under control. Ann. Rev. Plant Physiol. Plant Mol. Biol. 49: 249-279

Obendorf, R.L. (1997), Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. Seed Sci. Res. 7: 63-74

Oliver, A.E. et al. (2001), Non-disaccharide-based mechanisms of protection during drying. Cryobiology 43: 151-167

Ooms, J.J.J et al. (1993), Acquisition of desiccation tolerance and longevity in seeds of *Arabidopsis thaliana*: a comparative study using abscisic acid-insensitive *abi3* mutants. Plant Physiol. 102: 1185-1191

Priestley, D.A. (1986), Seed ageing: implications for seed storage and persistence in the soil, Cornell University press. Ithaca/USA

Prieto-Dapena, P. et al. (2006), Improved resistance to controlled deterioration in transgenic seeds. Plant Physiol. 142: 1102-1112

Rajjou, L. et al. (2008), Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. Plant Physiology 148: 620-641

Rajjou, L. and Debeaujon, I. (2008), Seed longevity: Survival and maintenance of high germination ability of dry seeds, C. R. Biologies 331: 796-805

Raz, V. et al. (2001), Sequential steps for developmental arrest in *Arabidopsis* seeds. Development 128: 243-252

Rasmusson, D.C. and Wilcoxson, R.D. (1979), Registration of ˈMorexˈ barley. Crop Sci. 19: 293

Roberts, E.H. (1961), Viability of cereal seed for brief and extended periods. Annals of Botany 25: 373-380

Roberts, E.H. (1973), Predicting the storage life of seeds. Seed Sd. Technol. 1: 499-514

Rogers, J.C. et al. (1985), Aleurain: a barley thiol protease closely related to mammalian cathepsin H. Proc. Natl. Acad. Sci. USA 82: 6512-6516

Roqueiro, G. et al. (2010), Effects of photooxidation on membrane integrity in *Salix nigra* seeds. Annals of Botany 105: 1027-1034

Sasaki, K. et al. (2005), Mapping of quantitative trait loci controlling seed longevity of rice (*Oryza sativa* L.) after various periods of seed storage. Plant Breed. 124: 361-366

Sattler, S.E. et al. (2003), Characterization of tocopherol cyclases from higher plants and cyanobacteria. Evolutionary implications for tocopherol synthesis and function. Plant Physiol. 132: 2184-2195

Sattler, S.E. et al. (2004). Vitamin E is essential for seed longevity, and for preventing lipid peroxidation during germination. Plant Cell 16: 1419-1432

Schwember, A.R. and Bradford K.J. (2010), Quantitative trait loci associated with longevity of lettuce seeds under conventional and controlled deterioration storage conditions. J. Exp. Bot. 61: 4423-4436

Shirley, B.W. et al. (1995), Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J. 8: 659-671

Singh, R.K. et al. (2008), SSR markers associated with seed longevity in soybean. Seed Science Technology 36: 162-167

Sugliani, M. et al. (2009), Natural modifiers of seed longevity in the *Arabidopsis* mutants abscisic acid insensitive3-5 (*abi3-5*) and leafy cotyledon1-3 (*lec1-3*). New Phytol. 184: 898-908

Tesnier, K. et al. (2002), A controlled deterioration test for *Arabidopsis thaliana* reveals genetic variation in seed quality. Seed Sci. Technol. 30: 149-165

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

Walters, C. (1998), Understanding the mechanisms and kinetics of seed aging. Seed Sci. Res 8: 223-244

Walters, C. et al. (2005), Dying while dry: kinetics and mechanisms of deterioration in desiccated organisms. Integr. Comp. Biol. 45: 751-758

Waterworth, W.M. et al. (2011), Repairing breaks in the plant genome: The importance of keeping it together. New Phytol. 192: 805-822

Waterworth, W.M. et al. (2016), DNA damage checkpoint kinase ATM regulates germination and maintains genome stability in seeds. Proc. Natl. Acad. Sci. USA. 13: 9647-9652

Wettlaufer, S.H. and Leopold, A.C. (1991), Relevance of Amadori and Mailard products to seed deterioration. Plant Physiology 97: 165-169

Wolkers, W.F. et al. (1998), Properties of proteins and the glassy matrix in maturation defective mutant seeds of *Arabidopsis thaliana*. Plant J. 16: 133-143

Chapter 2

Mapping and Confirmation of Quantitative Trait Loci for Seed Longevity in the L94 NIL Mapping Population

Introduction

Natural variation in species with a broad distribution range is expected to permit flexibility and survival in the face of changing environmental circumstances. Cultivated barley, *Hordeum vulgare* L., is grown in a range of diverse environments that vary from sub-arctic to sub-tropical, with greater concentration in temperate areas and high altitudes of the tropics and subtropics (Von Bothmer, 1992). Barley was first domesticated about 10.000 years ago from its wild relative, *Hordeum vulgare* ssp. *spontaneum*, in the area of the Middle East, known as the Fertile Crescent (Badr et al., 2000). *Hordeum vulgare* ssp. *spontaneum* still grows in the Middle East and adjacent regions of North Africa, in both natural and disturbed habitats, such as abandoned fields and roadsides. Growing in widely varying environments, suggest that barley is adaptive to a wide range of soils and it is more tolerant to salinity and alkalinity conditions than other cereals (Epstein et al., 1980; Chen et al., 2007). Most of the responses of growth-related traits to different environmental scenarios are genotype-dependent. The genetic difference among individuals of the same species is often the consequence of allelic differences at multiple loci with a quantitative effect and modification by environmental factors. The identification of genes underlying phenotypic variation has an enormous practical implication by providing a means to improve crop yield and quality. To dissect the genetic architecture of complex traits, quantitative genetics is often performed.

The principle of QTL mapping in segregating populations is based on the genotyping of progenies derived from a cross of distinct genotypes for the trait under study. Phenotypic values for the quantitative traits are then compared with the molecular marker genotypes of the progeny to search for particular genomic regions showing statistically significant associations between polymorphisms and the trait variation. The respective loci that control trait variation are then called Quantitative Trait Loci (QTL). While early QTL mapping studies used restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) markers (Collard et al., 2005), the most common marker types used today are single nucleotide polymorphism (SNP), simple sequence repeat (SSR) or insertion-deletion (indel) markers (Mackay et al., 2009). SNPs between different genotypes of the same species are commonly identified during genome sequencing and the alleles can easily be distinguished from one another using the Cleaved Amplified Polymorphic Sequence (CAPS) method. This method detects the loss or gain of a restriction enzyme recognition site due to a SNP (or a small indel) by amplifying a DNA fragment containing this site and digesting it with a suitable enzyme. The genotypes can then be distinguished by agarose gel electrophoresis (Konieczny and Ausubel, 1993; Neff et al., 1998).

QTL mapping can be performed in different types of populations. In plants, so called 'immortal' mapping populations, consisting of homozygous individuals, are preferred. The use of these populations allows the performance of replications and multiple analyses of the same population in different environments without the need for further genotyping. The effects of each QTL in different environments can be precisely estimated and tested for QTL x environment interactions (Koornneef et al., 2004). Figure 1 shows crossing schemes of the most frequently used types of biparental mapping populations. Recombinant inbred lines (RILs) are the homozygous selfed or sib-mated progeny of the

individuals of an F_2 population using single seed descent (SSD). F_2 populations originate from heterozygous F_1 plants resulting from the cross of two parental lines differing in a trait of interest. In RILs, along each chromosome, blocks of alleles derived from either parent alternate. Because recombination can no longer change the genetic constitution of RILs, further segregation in the progeny of such lines is absent. Therefore these lines constitute a permanent or immortal genetic resource that can be replicated indefinitely by further selfing.
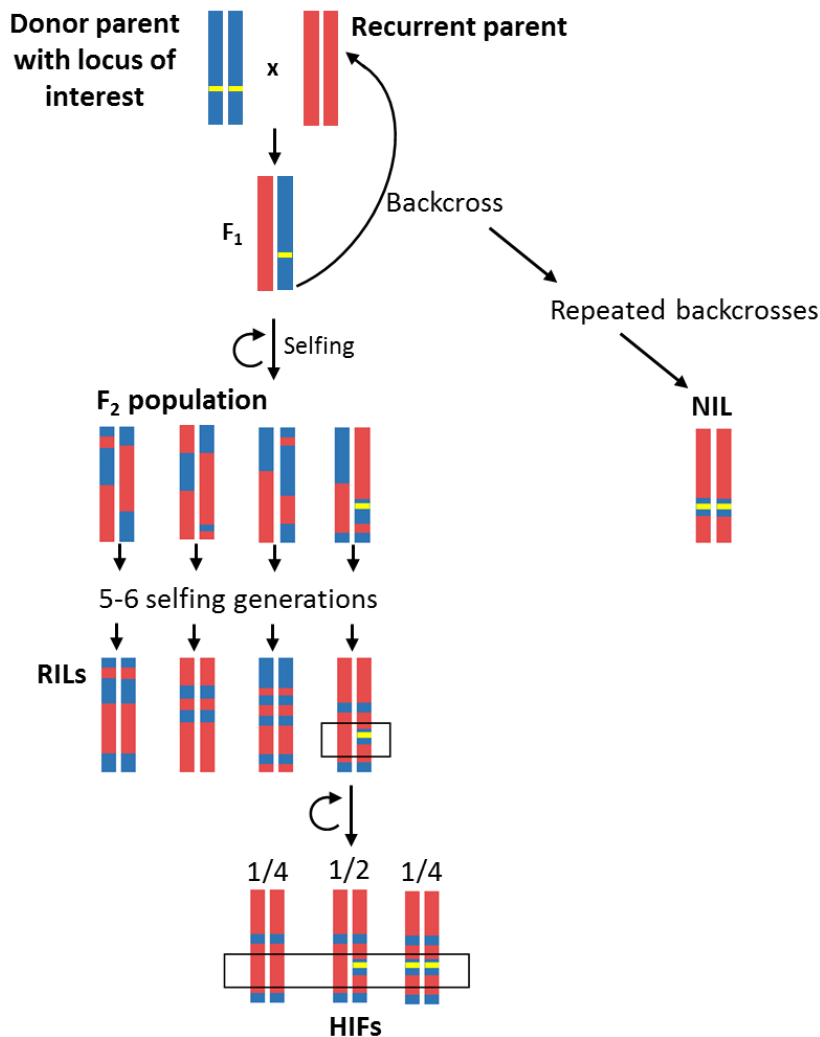

Figure 1. **Crossing scheme for common biparental mapping populations**
One chromosome pair is shown as representative of the diploid parental genomes. The chromosomes derived from the recurrent parent are shown in red, and those of the donor parent in blue. The donor parent carries the locus of interest which is indicated as yellow line. The two parental lines are crossed to produce a heterozygous F_1. Selfing in the F_1 plants gives rise to F_2 plants, largely varying in their genetic constitution. For the generation of recombinant inbred lines (RILs), plants of an F_2 population are continuously selfed until fixed recombinants of the parental lines are obtained. Heterozygous inbred family (HIFs) individuals are derived by selfing RILs still heterozygous around the locus of interest (shown in black rectangle). By selfing such RILs, fixed individuals are obtained for donor or recurrent parent's allele. Near isogenic lines (NILs) originate from an F_1 plant that is repeatedly backcrossed to the recurrent parent. With each round of backcrossing, the number and size of genomic fragments of the donor parent are reduced and a single introgression in the region of the locus of interest can be achieved.

Since the repeated selfing for many generations is very time-consuming, the generation of doubled haploids (DH) can be a time-saving alternative to obtain homozygous lines (Seymour et al., 2012). In this approach, haploid offspring of the F_1 is recovered (e.g. from pollen grains) and converted into the diploid state by spontaneous or induced chromosome-doubling. This produces fully homozygous offspring within one generation, but it requires tissue culture and recombination is limited to events occurring during a single meiosis (Collard et al., 2005; Rakshit et al., 2012). QTL analysis requires the confirmation and fine mapping of QTLs. This can be achieved by employing near isogenic lines (NILs) (Keurentjes et al., 2007) that can be obtained by backcrossing specific genotypes with one or both parents. NILs are introgressions of the QTLs of interest in a contrasting parental background. One QTL can even been introgressed in different backgrounds (Bentsink et al., 2010). The parent used for the backcrossing is called recurrent parent. Molecular markers are used to help monitoring the reduction of the number and size of fragments originated from the donor parent and to speed up the process. Containing only a single introgression per line, NILs increase the power to detect small-effect QTL. The generation of NILs is particularly valuable for those species for which no transformation protocol is established to produce transgenics for the alleles of interest. A further advantage is that in NILs genomic rearrangements, which may happen during transformation, are avoided. For QTL validation within a RIL population, the residual heterozygosity present in RILs can be utilized. Heterozygosity is reduced on average by 50% in every selfed generation, so even after several generations of selfing, a considerable degree of heterozygosity remains within the population. Progeny of a heterozygous RIL at the (QTL) locus of interest, termed heterogeneous inbred family (HIF), will segregate for the locus of interest in a 1: 2: 1 ratio and the homozygous individuals can be considered NILs towards each other (Tuinstra et al., 1997).

To map QTLs for seed longevity, the so called L94 mapping population was used. This population was originally developed to map QTLs for partial resistance to barley leaf rust, which is caused by *Puccinia hordei* (Qi et al., 2000). As parental lines, the spring barley landraces L94 from Ethiopia and Cebada Capa (CC) from Argentina were used. In Ethiopia, barley is one of the most important staple food crops in the highlands (2.000 - 3.000 meters above sea level) and ranks fifth in area and production among the cereals (Woldeab et al., 2015). In Argentina, barley is grown in mountainous regions and is almost exclusively produced for malt, used in the brewing of beer (Ullrich, 2011). For the generation of a mapping population for partial resistance to leaf rust, L94 was selected due to its extreme susceptibility. Cebada Capa, on the other hand, has a high level of partial resistance and carries the gene *Rph7* for hypersensitive resistance (Niks and Kuiper, 1983). One partially resistant line, 116-5 (F_5), was derived from a cross between L94 and Cebada Capa by applying the single seed descent method (SSD). This line was thereafter backcrossed to L94 and after SSD, a RIL population (F_8) containing 117 lines was obtained. Using this RIL population, an AFLP molecular map was constructed to identify QTLs for partial resistance to barley leaf rust (Qi et al., 2000).

As shown in figure 2, L94 and CC differ in their row type and seed phenotype. L94 is two-rowed and produces dark, naked grains, while CC is six-rowed and produces light, hulled grains. Apart from these phenotypes and the before mentioned resistance phenotype to leaf rust, these two landraces also differ widely in their seed longevity phenotype. L94 displays poor seed longevity, while Cebada Capa shows a high level of seed longevity (Adimargono et al., in preparation).
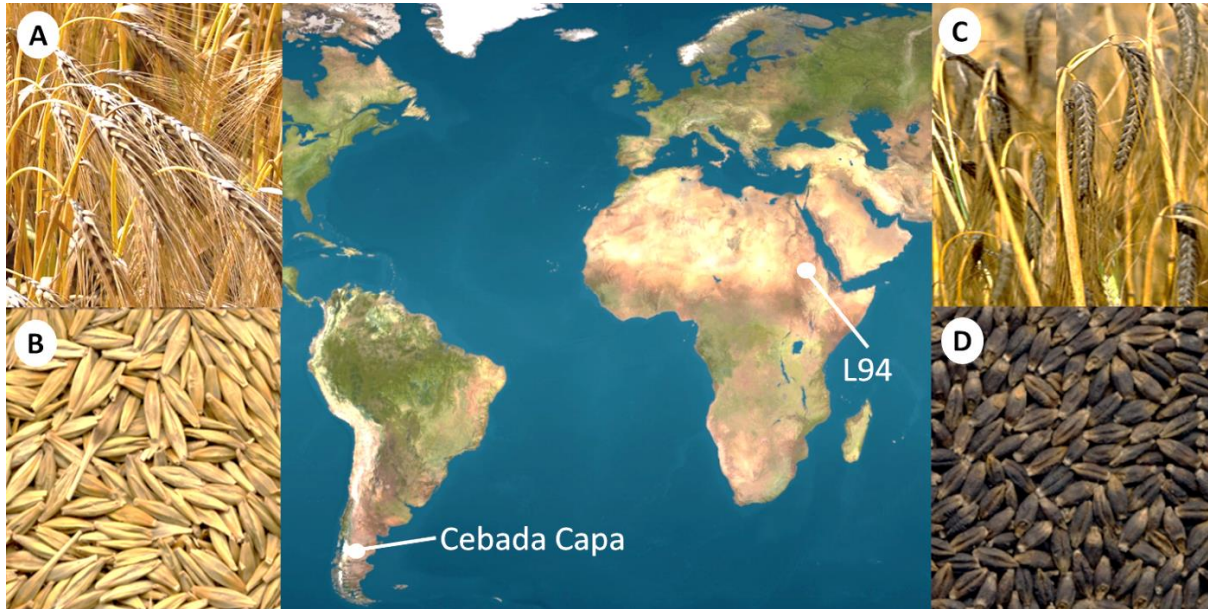
Figure 2. **Geographic location and row type and seed phenotype of the L94 mapping population parentals**
The long-lived parental Cebada Capa is an Argentinian six-rowed (A) spring barley landrace with light, hulled grains (B). The short-lived parental L94 is grown in Ethiopia and is a two-rowed (C) spring barley landrace with dark, naked grains (D). The world map was taken from http://gryllus.net/Blender/PDFTutorials/.

The before mentioned 117 RILs, derived from the cross between L94 and 116-5, were used to map QTLs for seed longevity (Adimargono et al., in preparation). To genotype the RILs, a set of 535 SNP markers from the Barley Oligonucleotide Pool Assay 1 (BOPA1) (Close et al., 2009) was utilized. As it can be seen in figure 3a, the markers are evenly distributed over the seven barley chromosomes. Using controlled deterioration tests, the 117 RILs were phenotyped for seed longevity. As indicated in figure 3a, QTL-mapping resulted in the identification of four putative QTLs (Adimargono et al., in preparation). Three seed longevity QTLs (SLQ1.1-1.3) were found on chromosome 1 and one at the bottom of chromosome 2 (SLQ2). In order to fine-map those four putative QTLs, near isogenic lines were generated. The recombinant inbred line with the strongest seed longevity phenotype, RIL114, was selected for backcrossing to L94. As shown in figure 3b, two F_3 lines (20008-20 and 20001-1) from this second backcross (BC_2) were selected after single seed decent. To obtain NILs for SQL1.2, SLQ1.3 and SLQ2, the line 20008-20 was once again backcrossed to L94 to further reduce introgressions from the long-lived Cebada Capa parent. F_3 lines from this third backcross (BC_3) were selected. As shown in figure 3b, the line 20001-1 was only selfed, but not further backcrossed to generate NILs for the putative seed longevity QTL on the top of chromosome 1 (SLQ1.1).
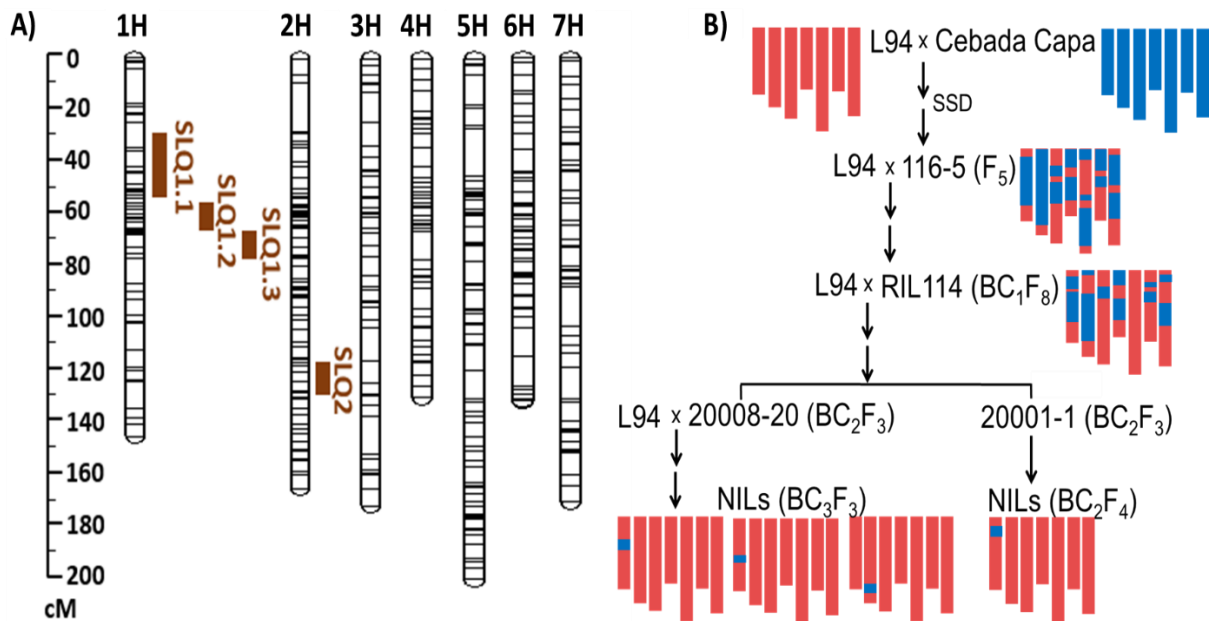
Figure 3. **Location of putative seed longevity QTLs identified in the L94 x 116-5 RIL mapping population and schematic presentation of the generation of the L94 NIL mapping population**

The seven barley chromosomes are schematically presented and abbreviated as 1H to 7H. Chromosomes are colored in red and blue representing the genomic regions of L94 and introgression segments from Cebada Capa, respectively. A) Location of putative Quantitative Trait Loci (QTL) for seed longevity (SLQ) in the L94 x 116-5 recombinant inbred line (RIL) mapping population. Three SLQs (1.1-1.3) are located on chromosome 1 and the SLQ2 is located at the bottom of chromosome 2. The 535 single nucleotide polymorphism (SNP) markers from the Barley Oligonucleotide Pool Assay 1 (BOPA1) (Close et al., 2009), used to genotype the RILs, are presented as black stripes. Positions of the markers are given in centimorgans (cM) (data from Adimargono et al., in preparation) B) Schematic presentation of the generation of the L94 NIL mapping population. The application of the single-seed descent (SSD) method is indicated. The number of filial generations (F) and backcrosses (BC) is indicated in brackets behind each line.

Objectives

The above-mentioned studies described the identification of four putative QTLs for seed longevity on chromosomes 1 and 2, and the generation of the so called L94 NIL mapping population. One main goal was to confirm these putative QTLs in the newly generated L94 NILs employing controlled deterioration tests. These tests were conducted in the same way as for the phenotyping of the 117 RILs, derived from the cross between L94 and 116-5. Making use of the recent improvement of the barley genomics infrastructure, an RNA-seq analysis was employed to further map Cebada Capa introgressions around the QTL regions in the L94 NILs and to exclude the presence of additional introgressions on other chromosomes. In this manner, RNA-seq was used to confirm the ʹcleannessʹ of the L94 NILs. Furthermore, the RNA-seq analysis was used to obtain new SNP markers to be used for the possible future fine mapping of the QTLs.

Results

RNA-seq analysis reveals additional Cebada Capa introgressions in the L94 NILs

RNA-seq was employed to further map Cebada Capa introgressions around the QTL regions in the L94 NILs and to exclude the presence of additional introgressions on other chromosomes, thereby confirming the ˈcleannessˈ of these lines. The further use of RNA-seq for the identification of differentially expressed candidate genes and possible downstream targets is discussed in detail in chapter 3. RNA was extracted from embryos of mature, non-aged seeds harvested at the end of the year 2013. Four replicates of 20 embryos each were used per genotype. Single end sequencing was performed on the HiSeq 2000 (Illumina®) platform of the Genome Center of the Max Planck Institute in Cologne. For each library a minimum of 15 million reads was generated by multiplexing eight libraries. Sequence alignments were performed using a merged dataset of high-confidence (HC) and low-confidence (LC) predicted barley genes (The International Barley Genome Sequencing Consortium, 2012) as reference. Sequence variants between L94 and Cebada Capa were obtained using the GATK UnifiedGenotyper platform and the resulting data was analyzed using R. Figure 4 exemplarily shows the genotype of the NIL selected for SLQ1.2, sequenced by RNA-seq in 2014 (SLQ-1.2(2014)), and its pedigree.
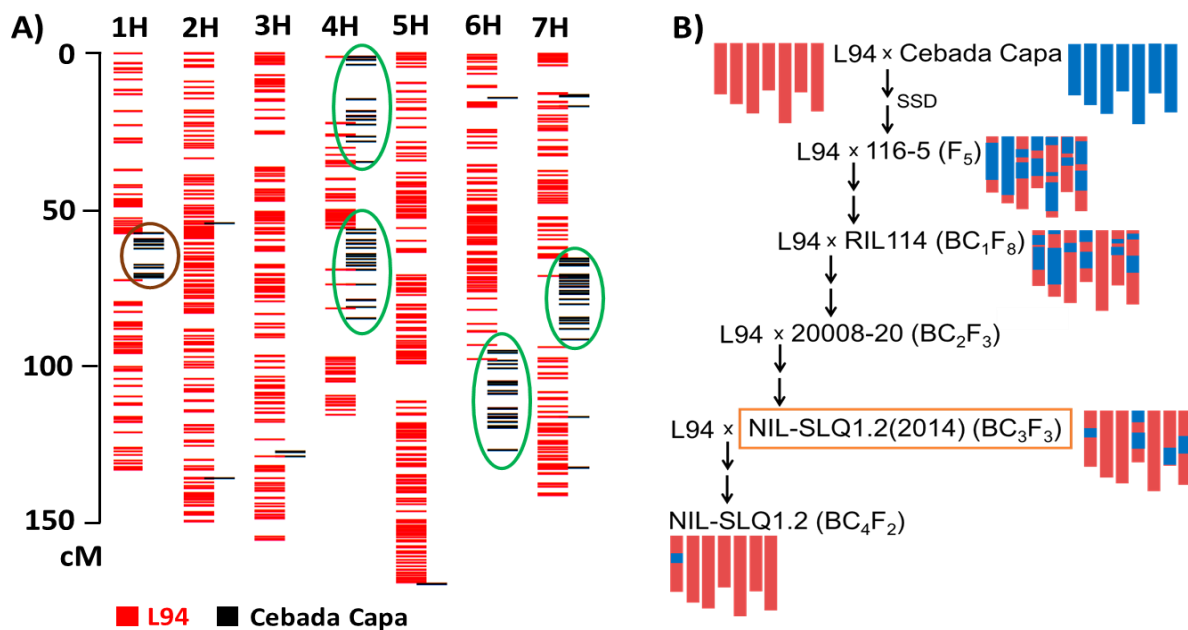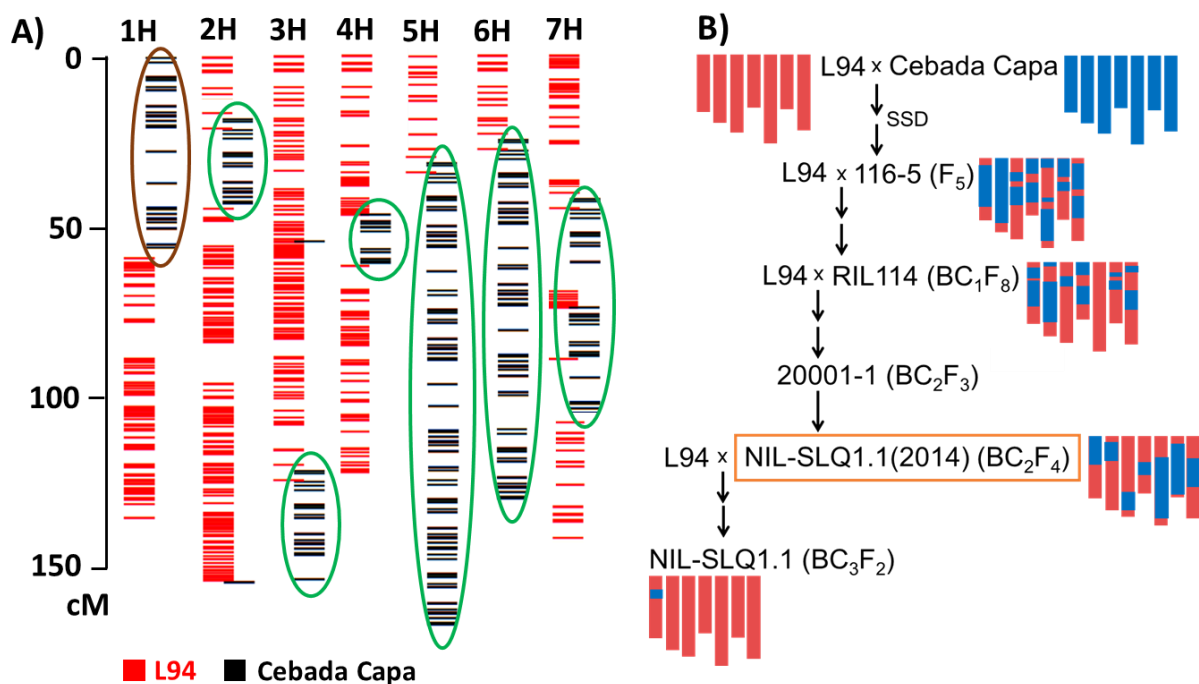

Figure 4. **RNA-seq derived** i**ntrogression pattern and pedigree of NIL-SLQ1.2 sequenced in 2014**

A) Mapping of Cebada Capa introgressions using RNA-seq in a near isogenic line selected for the seed longevity QTL 1.2 (SLQ1.2). The seven barley chromosomes are schematically presented and abbreviated as 1H to 7H. Polymorphic variants between L94 and Cebada Capa are presented as horizontal lines. Variants originated from L94 and Cebada Capa are colored in red and black, respectively. Positions of loci containing the variants are given in centimorgans (cM). Cebada Capa introgressions in the region of SLQ1.2 are highlighted in brown and introgressions on other chromosomes are highlighted in green. B) Pedigree of the in A) presented NIL. Chromosomes are colored in red and blue representing the genomic regions of L94 and introgression segments from Cebada Capa, respectively. The NIL shown in A) is highlighted by an orange rectangle. The application of the single-seed descent (SSD) method is indicated. The number of filial generations (F) and backcrosses (BC) is indicated in brackets behind each line.

As shown in figure 4a, the NIL selected for SLQ1.2 clearly shows an introgression in the respective QTL region and no further introgressions in other SLQ regions on chromosomes 1 and 2. Surprisingly, the NIL exhibits four clear introgressions on chromosomes 4, 6 and 7. These introgressions should no longer be present after the backcrossing of the line 20008-20 to L94 and various steps of selfing as shown figure 3b. Figure 4b shows the pedigree of the sequenced NIL, highlighted by an orange rectangle. Comparing the introgressions of this NIL to RIL114, it can be seen that the introgressions detected on chromosomes 4 and 7 are also present in RIL114. Interestingly, the introgression on the bottom of chromosome 6 is absent in RIL114. A possible explanation for the occurrence of this introgression might be that the L94 plant used for backcrossing of line 20008-20 was not clean. As shown in figure 4b, a further backcross of the sequenced NIL to L94 had to be performed to reduce all additional introgressions and to obtain a clean NIL-SLQ1.2.

All NILs grown in the field in 2013 were sequenced and Cebada Capa introgressions were mapped using RNA-seq and presented in the same way as shown in figure 4. Additional Cebada Capa introgressions outside the SLQ regions were detected in all sequenced L94 NILs. A lot of additional introgressions were detected for the NIL selected for SLQ1.1 as shown in figure 5a.


Figure 5. **RNA-seq derived introgression pattern and pedigree of NIL-SLQ1.1 sequenced in 2014**

A) Mapping of Cebada Capa introgressions using RNA-seq in a near isogenic line selected for the seed longevity QTL 1.1 (SLQ1.1). The seven barley chromosomes are schematically presented and abbreviated as 1H to 7H. Polymorphic variants between L94 and Cebada Capa are presented as horizontal lines. Variants originated from L94 and Cebada Capa are colored in red and black, respectively. Positions of loci containing the variants are given in centimorgans (cM). Cebada Capa introgressions in the region of SLQ1.1 are highlighted in brown and introgressions on other chromosomes are highlighted in green. B) Pedigree of the in A) presented NIL. Chromosomes are colored in red and blue representing the genomic regions of L94 and introgression segments from Cebada Capa, respectively. The NIL shown in A) is highlighted by an orange rectangle. The application of the single seed descent (SSD) is indicated. The number of filial generations (F) and backcrosses (BC) is indicated in brackets behind each line.

The pedigree of the NIL-SLQ1.1 sequenced by RNA-seq in 2014 (NIL-SLQ1.1(2014)), shown in figure 5b, reveals that the majority of these additional introgressions are not present in RIL114. Especially the huge introgression on chromosome 5 is totally absent in RIL114. NIL-SLQ1.1(2014) has almost as many introgressions as 116-5 and even in comparison to this line the introgression on the bottom of chromosome 3 cannot be explained. The larger amount of introgressions in comparison to the NIL-SLQ1.2(2014), shown in figure 4, can be explained by the fact that only two backcrosses had been performed instead of three. In order to be able to reduce the additional introgressions, a further backcross was performed, as shown in the pedigree of figure 5b, to obtain a clean NIL-SLQ1.1.

A schematic representation of all Cebada Capa introgressions detected in the L94 NILs is shown in figure 6. This representation also shows a NIL which was selected due to the fact that it carried both the SLQ1.2 and SLQ1.3. This NIL was also sequenced and named NIL-SLQ1.2+3(2014). NIL-SLQ1.1(2014) is by far the line with the highest number of introgressions. In order to obtain clean NILs, one additional backcross to L94 was performed for all lines. To select clean NILs, SNP markers were utilized to speed up the selection of such lines. These markers were either markers from the Barley Oligonucleotide Pool Assay 1 (BOPA1) (Close et al., 2009), or were designed on the basis of the RNA-seq data and sequences were derived from the barley high- and low-confidence gene set (The International Barley Sequencing Consortium, 2012). Markers were tested on the whole population by the company LGC (https://www.lgcgroup.com/) using KASP™ genotyping assays.
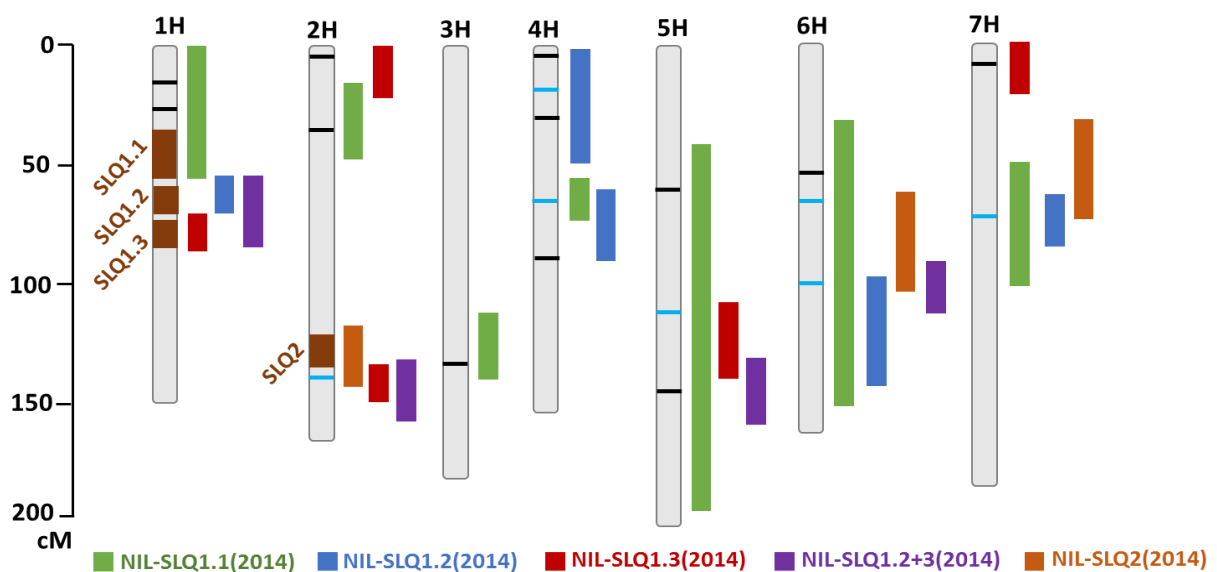
Figure 6. **Schematic representation of detected Cebada Capa introgressions in NILs sequenced in 2014**
The seven barley chromosomes are schematically presented and abbreviated as 1H to 7H. Colored bars indicate the position of Cebada Capa introgressions detected in 2014 sequenced NILs selected for putative Quantitative Trait Loci (QTL) for seed longevity (SLQ). Three SLQs (1.1-1.3) are located on chromosome 1 and the SLQ2 is located at the bottom of chromosome 2, SLQs are indicated by dark brown bars. The positions of the QTL regions and introgressions are given in centimorgans (cM). SNP-markers used for the reduction of introgressions outside the QTL regions are presented as horizontal lines. SNPs colored in black are derived from the barley consensus map (Close et al., 2009) and markers in light blue are derived from the barley high- and low confidence gene set (The International Barley Sequencing Consortium, 2012).

The final result of the ˈcleaningˈ of the NILs sequenced in 2014 is shown in figure 7. Figures 7a, 7b and 7d show the clean NILs obtained for SLQ1.1, SLQ1.2 and SLQ.2. In the case of NIL-SLQ1.3, only the additional introgression on chromosome 5 could be removed, but not the introgressions on chromosomes 2 and 7.
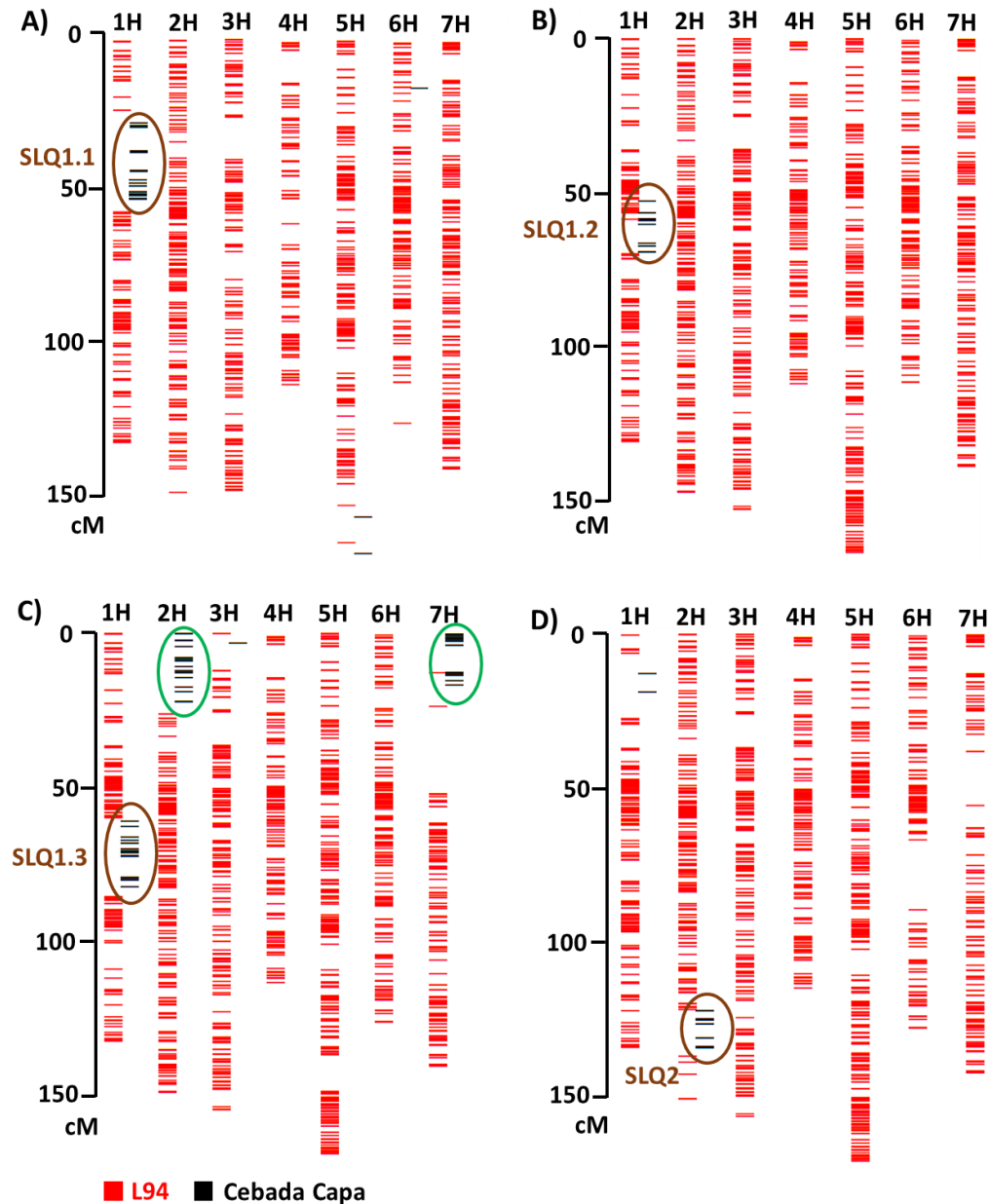

Figure 7. **Mapping of Cebada Capa introgressions in ˈcleanˈ NILs selected for SLQ1.1-1.3 and SLQ2**

The seven barley chromosomes are schematically presented and abbreviated as 1H to 7H. Presented NILs were selected for putative Quantitative Trait Loci (QTL) for seed longevity (SLQ). Three SLQs (1.1-1.3) are located on chromosome 1 and the SLQ2 is located at the bottom of chromosome 2. Cebada Capa introgressions were mapped using R. Polymorphic variants between the two parental lines L94 and Cebada Capa are presented as horizontal lines. Variants originated from L94 and Cebada Capa are colored in red and black, respectively. Positions of loci containing the variants are given in centimorgans (cM). Cebada Capa introgressions in the SLQ regions are highlighted in brown and introgressions on other chromosomes are highlighted in green. A) Schematic representation of the genotype of NIL-SLQ1.1. B) Schematic representation of the genotype of NIL-SLQ1.2. C) Schematic representation of the genotype of NIL-SLQ1.3. D) Schematic representation of the genotype of NIL-SLQ2.

## Confirmation of the four putative QTLs for seed longevity in the L94 NILs

The four putative SLQs, identified in the L94 x 116-5 RIL population, were tested using controlled deterioration tests. These tests were performed at 60°C in an incubator maintaining a relative moisture content of 10%. Seed lots were removed at weekly intervals and tested for their germination ability. Seedlings were scored as normal or abnormal according to the international rules for seed testing (ISTA). Seed survival curves were obtained by plotting the percentage of viable seeds against the storage time. Seed survival curves of L94, Cebada Capa, 116-5 and RIL114 are shown in the left panel of figure 8. In contrast to the other three lines, L94 rapidly loses its viability, so that after 28 days the majority of the seeds is unable to germinate. In contrast to that, it takes around 63 days for Cebada Capa seeds to lose their ability to germinate under the same conditions. The seed survival curves of 116-5 and RIL114 resemble the one of Cebada Capa and in both lines seeds become unable to germinate after around 56 days of storage. To compare seed longevity between different seed lots and lines, seed survival curves are usually transformed according to the so-called probit analysis (Ellis and Roberts, 1980). This analysis is basically a regression analysis expressing percent viability as normal equivalent deviate (NED) viability. Before the widespread use of calculators and computers, it was difficult to make calculations with negative numbers, and so NEDs were converted to probits by adding + 5 to the NED value. By transforming percent viability to NED or probit viability, as shown exemplarily for L94, Cebada Capa, 116-5 and RIL114 in figure 8, a linear relationship between viability and storage period is apparent.

The equation shown between the two panels:

$$v = Ki - \left(\frac{1}{\sigma}\right)p$$
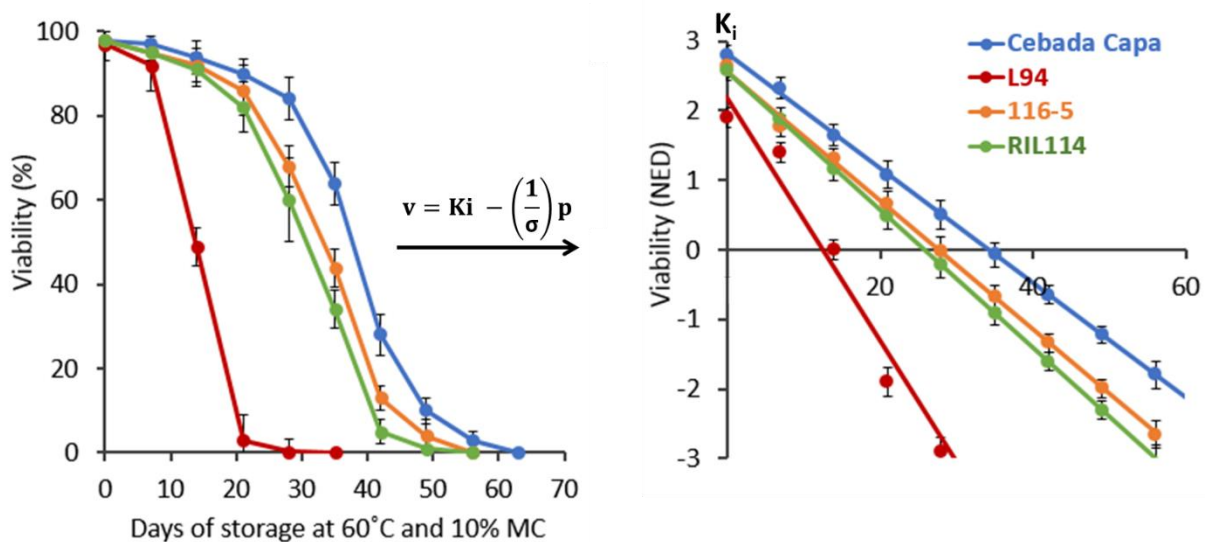
Figure 8. **Seed survival curves of L94, Cebada Capa, 116-5 and RIL114 before and after transformation using the probit analysis**

Seed survival curves obtained by plotting the percentage of viable seeds of L94, CC, 116-5 and RIL114 after controlled deterioration tests against the storage time, are shown on the left. Controlled deterioration tests were performed at 60˚C and a relative moisture content (MC) of 10%. For each genotype at least 10 independent lines were used, and 4 replicates were performed per each time point. Seed survival curves were transformed by the probit analysis (Ellis and Roberts, 1980) into straight lines (shown on the right) according to the shown equation. V is the viability after p days of storage, σ is the time for viability to fall by one NED (normal equivalent deviate) and Ki, the y-intercept, is the initial viability of a seed lot. Using the probit analysis, the viability in percentages is converted into NEDs.

The equation according to which the seed survival curves are transformed is given by $v = K_i - (1/\sigma) p$. V is the viability after p days of storage and K_i, the y-intercept, is the initial viability. σ represents the time for viability to fall by one NED or probit and is found in the slope of the transformed curves given by $-1/\sigma$. As shown in figure 8, the transformed seed survival curve of L94 is steeper and has a larger slope than the ones of the other three analyzed genotypes. The values of K_i also differ. L94 has a K_i value of 2.2, whereas for Cebada Capa a value of 2.8 was obtained. The K_i values of 116-5 and RIL114 are both around 2.6. In order to confirm the four putative SLQs, deterioration tests were performed with the L94 NILs sequenced in 2014 and the cleanest NILs shown in figure 7. For each genotype at least 30 independent lines were used and four replicates were performed per each time point. Figure 9 shows the untransformed and transformed seed survival curves of the in 2014 sequenced L94 NILs and the parental and precursor lines already shown in figure 8. As shown in figure 9a, the seed survival curves of the NILs are more familiar to Cebada Capa, 116-5 and RIL114 than to L94. However, the NILs lose their viability faster than the other three genotypes, but are longer lived than L94. As shown in figure 9b, the transformed seed survival curves of the L94 NILs have a larger slope than the one of L94 but are steeper than the one of RIL114. NIL-SLQ1.1(2014) has the smallest slope and NIL-SLQ2(2014) the largest slope of the NILs. The K_i values of the NILs are higher than the one of L94. NIL-SLQ2(2014) has the largest K_i value of the NILs.
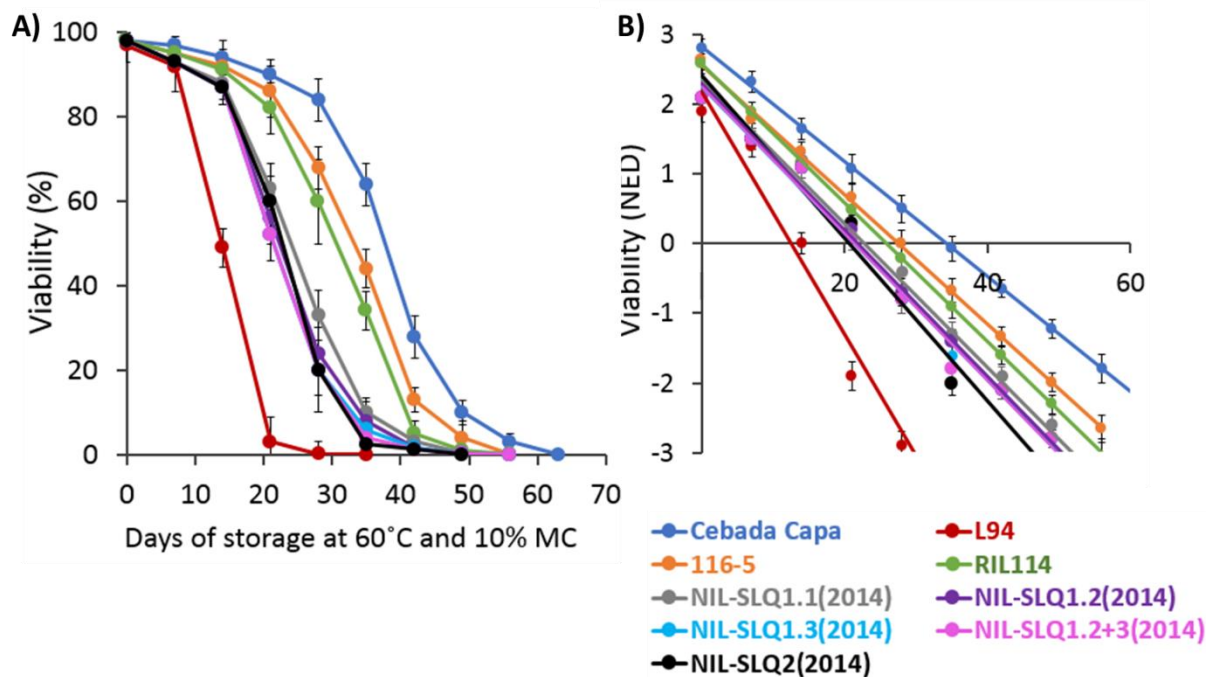
Figure 9. Untransformed and transformed seed survival curves of L94, Cebada Capa, 116-5, RIL114 and the L94 NILs sequenced in 2014

The shown near isogenic lines (NILs) have Cebada Capa (CC) introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2) and were sequenced by RNA-seq in the year 2014. NIL-SLQ1.2+3(2014) harbors CC introgressions in SLQ1.2 and SLQ1.3. A) Seed survival curves were obtained by plotting germination ability of seed lots of L94, CC, 116-5, RIL114 and NILs after controlled deterioration tests against the storage time. Controlled deterioration tests were performed at 60°C and a relative moisture content (MC) of 10%. B) After the probit analysis (Ellis and Roberts, 1980) obtained straight lines. Using the probit analysis, the viability in percentages is converted into normal equivalent deviates (NEDs).

The mean seed longevity (P_{50}) classifies seed collections as short- to long-lived for *ex situ* (Probert et al., 2009) and *in situ* purposes (Long et al., 2008). P_{50} is defined as the time taken in storage for viability to fall to 50%. P_{50} can be calculated by setting v = 0 (equal to 50% germination) and transforming p to P_{50} in the before mentioned equation v = Ki - p/σ. P_{50} is therefore the product of Ki and σ. P_{50} values represent the x-intercepts after the transformation of seed survival curves. A summary of the probit parameters, the intercept Ki and the slope (-1/σ) of the transformed seed survival curves, is shown in table 1 for L94, Cebada Capa, 116-5, RIL114, the in 2014 sequenced L94 NILs and the cleanest NILs.

Table 1. Probit parameters of L94, Cebada Capa, 116-5, RIL114, the L94 NILs sequenced in 2014 and the cleanest NILs tested in controlled deterioration tests.

Controlled deterioration tests were performed at 60˚C and a relative moisture content (MC) of 10%. For each of the shown genotypes at least 30 independent lines were used and 4 replicates were performed per each time point. Seed survival curves were transformed by the probit analysis described by Ellis and Roberts (1980). Probit parameters were calculated in accordance with the longevity equation v = Ki – (1/σ)p. v represents the viability after p days of storage, Ki the initial viability and σ represents the time for viability to fall by one normal equivalent deviate (NED). The standard error (SE) of the slope (-1/σ) is also given. The listed near isogenic lines (NILs) have Cebada Capa (CC) introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). Lines sequenced in 2014 are indicated by brackets behind their name. NIL-SLQ1.2+3(2014) harbors CC introgressions in SLQ1.2 and SLQ1.3.

Line	Probit parameters		
	Intercept (Ki)	Slope (-1/σ)	SE slope
L94	2.1905	-0.1747	0.004
Cebada Capa	2.7964	-0.0862	0.006
116-5	2.6356	-0.0945	0.003
RIL114	2.5822	-0-0998	0.002
NIL-SLQ1.1(2014)	2.3533	-0.1031	0.003
NIL-SLQ1.1	2.3339	-0.1035	0.006
NIL-SLQ1.2(2014)	2.2482	-0.1028	0.003
NIL-SLQ1.2	2.2480	-0.1066	0.005
NIL-SLQ1.3(2014)	2.2479	-0.1056	0.006
NIL-SLQ1.3	2.2482	-0.1052	0.004
NIL-SLQ1.2+3(2014)	2.2489	-0.1057	0.003
NIL-SLQ2(2014)	2.4321	-0.1139	0.001
NIL-SLQ2	2.4318	-0.1137	0.002

P_{50} values were calculated with the probit parameters given in table 1. Figure 10 shows the P_{50} values in days of L94, CC, 116-5, RIL114, the NILs sequenced in 2014 and the cleanest NILs. It can be observed that L94 and Cebada Capa differ a lot in their P_{50} value. Cebada Capa has an almost three times higher value than L94. The NILs sequenced in 2014 and the cleanest NILs have P_{50} values that are significantly higher than the one obtained for L94. The values obtained for the cleanest NILs are all in the same range as the ones for the respective NILs sequenced in 2014. Interestingly, the NIL sequenced in 2014 with both the SLQ1.2 and SLQ1.3 (NIL-SLQ1.2+3(2014)) does not show a higher P_{50} value than the other lines with introgressions in only one SLQ region. Therefore, the two QTLs seem not to be additive. The values of all NILs are significantly lower than the one observed for RIL114. This could be explained by the fact that RIL114 has introgressions in all four SLQ regions. The P_{50} value of RIL114 itself is lower than the one of 116-5. Since the QTLs with the highest LOD (logarithm of odds) score were chosen for further investigation by Adimargono et al. (in preparation), 116-5 still contains other QTL regions for seed longevity which could lead to a higher P_{50} value. Taken all results into consideration, the four putative QTLs for seed longevity could be confirmed in the cleanest NILs.
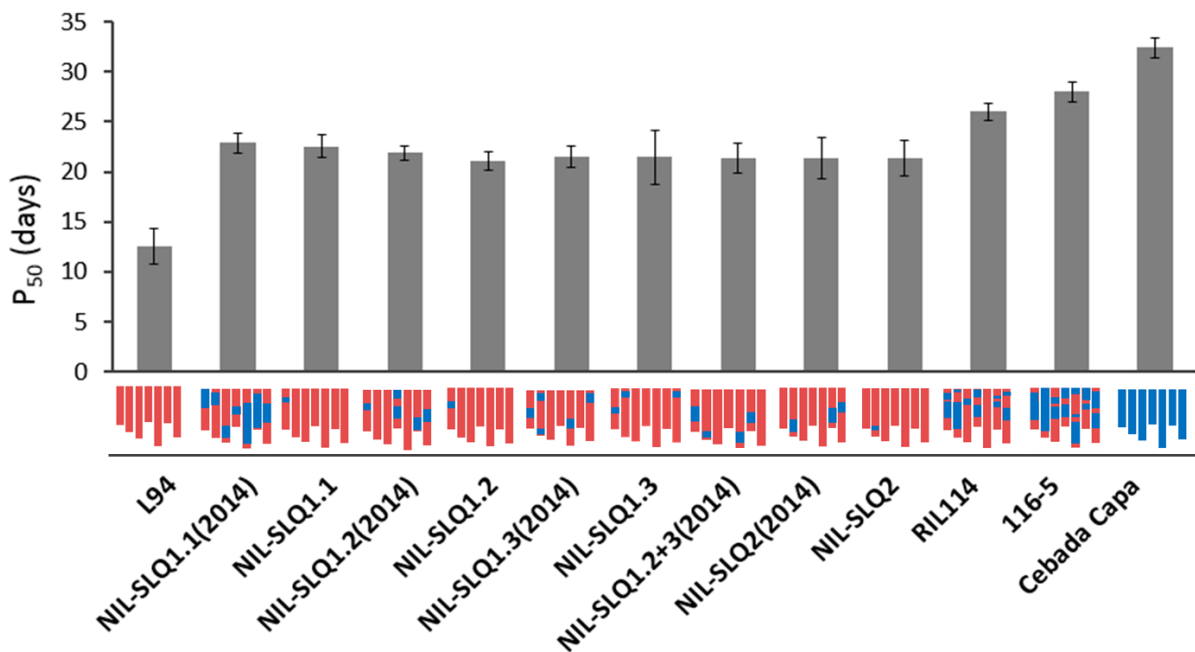
Figure 10. **P_{50} values of L94, Cebada Capa, 116-5, RIL114 and the L94 NILs**

P_{50} values of the presented genotypes were calculated by multiplying the Ki and σ values obtained after the transformation of seed survival curves by the probit analysis (Ellis and Roberts, 1980). Seed survival curves were obtained after analyzing at least 10 independent lines with four replicates per time point in controlled deterioration tests. Controlled deterioration tests were performed at 60˚C and a relative moisture content (MC) of 10%. Schematic representations of the analyzed genotypes show the seven barley chromosomes, the red and blue bars represent the genomic regions of L94 and introgression segments from Cebada Capa (CC), respectively. Near isogenic lines (NILs) show CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). The cleanest NILs, showing only CC introgressions in one seed longevity QTLs, were derived from NILs sequenced by RNA-seq in the year 2014. Lines sequenced in 2014 are indicated by brackets behind their name. NIL-SLQ1.2+3(2014) harbors CC introgressions in SLQ1.2 and SLQ1.3.

Discussion

The results of the before mentioned analyses show that the original L94 NILs, planted in the field in 2013, harbor additional and unexpected Cebada Capa introgressions. These introgressions were detected by an RNA-seq analysis in 2014. The size and amount of these introgressions was unexpected, since a number of backcrosses had already been performed, as shown exemplarily in pedigrees of NILs in figures 4 and 5. With this amount of introgressions, these lines could rather be described as RILs than NILs. Especially the NIL-SQL1.1(2014) harbors large introgressions on several chromosomes and has almost the same amount of Cebada Capa introgressions as 116-5. Two possible explanations for this finding could be given. One possible explanation might be that contaminations or cross-pollinations with Cebada Capa, 116-5 or RILs could have occurred in the field. Usually cross-pollination is prevented by bagging some spikes per plant shortly before flowering with perforated plastic bags, which are not removed until harvest. Seeds derived in such a manner are used as so-called breeder seeds for single seed descent to obtain the next generation. For the contaminations to occur, spikes selected for breeder seeds must have been bagged too late or some of the bags were damaged. A second possible explanation might be that some L94 plants, used for the backcrosses, were not clean and already had some contaminations with Cebada Capa. In this case, several backcrosses must have been performed with unclean L94 lines.

By performing backcrosses with the in 2014 sequenced lines, it was possible to obtain clean NILs with introgressions only in the respective SLQ regions. Only in the case of NIL-SLQ1.3, a completely clean line could not be obtained. As shown in figure 7, this NIL still harbors introgressions on top of chromosome 2 and 7. These introgressions are not located in regions of the four putative SLQs, but in regions in which both RIL114 and 116-5 harbor Cebada Capa introgressions. Therefore, these introgressions need to be reduced in order to exclude the presence of other QTLs in these regions, which could possibly affect the seed longevity phenotype of NIL-SLQ1.3. A further backcross of NIL-SLQ1.3 to L94 could provide an entirely clean line in the future. As shown in figure 10, the four putative SLQs could be confirmed in the clean L94 NILs. Fine mapping of these four SLQs is difficult since the introgressions in the respective SLQs are still of around 20cM. For future QTL fine mapping and characterization of individual loci, HIFs could be generated by backcrossing the obtained clean L94 NILs to L94. Each obtained segregating HIF is independent and contains unique recombination events in genomic regions flanking the QTL. An example of fine mapping using HIFs comes from the investigation of the four loci identified in the *Arabidopsis* population Bay-0 x Shahdara in low and high nitrogen environments by Loudet et al. (2003).

The P_{50} values obtained for L94 and Cebada Capa in the controlled deterioration tests were compared to before published P_{50} estimates. In a study by Probert et al. (2009), P_{50} estimates are given for seed lots of 195 taxonomically diverse species collected worldwide and placed into the same conditions of experimental storage. According to Probert et al. (2009), a P_{50} value ≤ 1 day indicates ˈvery shortˈ seed longevity, P_{50} > 1 to ≤ 10 days ˈshortˈ seed longevity, P_{50} > 10 to ≤ 100 days ˈmediumˈ seed longevity, P_{50} > 100 to ≤ 1000 days ˈlongˈ seed longevity and P_{50} > 1000 days ˈvery longˈ seed longevity. With P_{50} values of 12.54 ± 1.78 and 32.44 ± 1.02 days, L94 and Cebada Capa would be considered as

short- and medium-lived, respectively. In any case, Cebada Capa would not be considered as long-lived barley landrace, since its P_{50} value is far below 100 days. However, the in this study used ageing treatments differ from the majority of the in literature mentioned treatments. In the study of Probert et al. (2009), controlled deterioration (CD) tests were performed at 60% relative humidity (RH) and 45°C. In a recent study on barley seed longevity by Nagel et al. (2014), the same conditions were used. These conditions differ from those used in this study, since the controlled deterioration tests were performed at 10% RH and 60°C. A temperature difference of 15°C could have a strong effect on the ageing process since the reaction rate doubles for every 10°C increase in temperature (Pauling, 1988). Higher temperature could therefore lead to faster loss of membrane integrity, more rapidly reduced energy metabolism, faster impairment of RNA and protein synthesis, and DNA degradation. All these alterations lead to faster ageing, and consequently lower P_{50} values would be observed. To compare the in this studies obtained P_{50} values to those of other studies, the two parental lines and the L94 NILs would be needed to be subjected to CD tests at 60% RH and 45°C. Another important parameter which needs to be considered, is the Ki value of a seed lot. In the probit analysis, a seed lot with a high Ki value and a fast loss of seed longevity has a steeper line than one with the same tendency in viability and a lower Ki value. Therefore, the in other viability studies obtained Ki values need to be compared to those obtained for the L94 NILs and parental lines.

The fact that clean L94 NILs have been obtained in this study enables to investigate how the genetic background in conjunction with environmental conditions during (1) seed development and (2) seed storage conditions, such as temperature and seed moisture content, affects barley seed longevity. To investigate the influence of environmental conditions during seed development, it would be interesting to grow the L94 NILs outside in the field in consecutive years and to compare the effect of the different weather conditions on the seed longevity phenotype of these lines. This material could be subjected not only to CD tests, but also to accelerated ageing (AA) tests. Both of these rapid ageing methods are used to mimic seed behavior in storage (Roberts, 1973). In the before mentioned study by Nagel et al. (2014), AA tests were performed at 43°C and 100% RH for three days according to Hampton and TeKrony (1995). The effect of temperature and moisture content on the seed longevity phenotype of the L94 NILs could be investigated by comparing the effect of the two rapid ageing methods on these lines.

Material and methods

Plant material and growth conditions

For the RNA-seq analysis, conducted at the beginning of 2014, seeds of the barley landraces L94, Cebada Capa, the L94 NILs and the lines 116-5 and RIL114 were sown in 96-well trays. Seeds were germinated in the greenhouse and then transferred outside of the Max Planck Institute for Plant Breeding (MPIPZ Cologne, Germany) site. After six weeks, the plants were transferred to 12 L pots with one plant per pot, each filled with a custom-made peat and clay soil mixture (EinheitsErde® ED73 Osmocote, Einheitserdewerke Werkverband e.V., Sinntal Altengronau, Germany) containing a long term fertilizer. The pots were arranged in 22 rows with an inter-row distance of 1 m, where each row contained 54 pots with a distance of 10 cm. The entire experiment was divided into 4 blocks with at least 4 replicate plants per genotype within each block. To avoid edge effects, the plot was surrounded by border pots containing wild-type barley plants. The plot was irrigated by a sprinkling robot and treated with additional fertilizer or pesticides when necessary. Plants were harvested at maturity and 15 spikes per plant were threshed using the Wintersteiger LD 350 laboratory thresher (Wintersteiger, Ried im Innkreis, Austria).

For the final confirmation of the putative four seed longevity QTLs, using controlled deterioration tests, seeds of the barley landraces L94, Cebada Capa, L94 NILs sequenced in 2014, the cleanest L94 NILs and the lines 116-5 and RIL114 were sown in 96-well trays filled with a custom-made peat and clay soil mixture (EinheitsErde® ED73 Osmocote, Einheitserdewerke Werkverband e.V., Sinntal Altengronau, Germany). Seeds were germinated in the greenhouse under long day (LD) conditions (16h, 22°C day; 8h, 18°C night). After six weeks, plants were transferred to 1.3 L pots and maintained under LD conditions. The pots were arranged in 6 rows with a distance of 20 cm between rows, where each row contained 20 pots with a distance of 10 cm. The entire experiment was divided into 6 blocks with at least 5 replicate plants per genotype within each block. To avoid edge effects, the plot was surrounded by border pots containing wild-type barley plants. The plot was irrigated by hand and treated with additional fertilizer or pesticides when necessary. Plants reached maturity around the some date and were immediately harvested. Spikes were threshed individually by hand and seeds were dried in a drying conditioned at circa 10% relative humidity and 20°C for two months before starting the experiments.

DNA extraction from leaf material

For DNA extraction, approximately 100 mg of leaf tissue of young barley plants was collected into Qiagen™ 96-well collection tubes. The collection tubes were placed into 96-well blocks, sealed with Qiagen™ AirPore tape sheets, stored at -80°C for at least two hours and subsequently freeze-dried with a Christ Alpha 1-4 LSC freeze drier (Martin Christ GmbH, Osterode am Harz, Germany). After the freeze-drying, the AirPore tape sheets were removed and one 5 mm Qiagen™ stainless steel bead was added into each tube and the tubes were sealed with 8-well strips. The 96-well blocks were placed into a Retsch mixer mill (Retsch®, Haan, Germany) and the material was ground two times for 30 s at 25-30 Hz. DNA was extracted of the ground material with the Qiagen™ BioSprint

(Qiagen, Hilden, Germany) robot and the corresponding kit, following the supplier's recommendations. DNA was eluted in 200 μL of the supplied AE buffer and concentrations were measured using the NanoDrop™ spectrophotometer (PEQLAB, Erlangen, Germany). For further analyses, dilutions of 5-10 ng/μL DNA in AE buffer were prepared and stored at 4°C. Undiluted eluates were stored at -20°C.

SNP marker design and analysis by gel electrophoresis

Markers were designed based on the consensus map published by Close et al. (2009) or on the barley high- and low-confidence gene set (International Barley Sequencing Consortium, 2012). SNP positions are included in the supplemented information of Close et al. (2009). In the case of the high- and low-confidence gene set, information about the SNP location were obtained by the performed RNA-seq analysis. SNPs were analyzed for presence of CAPS or dCAPS sites (Neff et al., 1998) using dCAPS finder (http://helix.wustl.edu/dcaps/dcaps.html) and primers for amplification were designed with primer 3 (http://primer3.ut.ee/). SNP markers were assayed by gel electrophoresis. PCR was performed with the Taq DNA polymerase purchased from Ampliqon (Genaxxon bioscience GmbH, Ulm, Germany) and the supplied buffer solutions from the same company. Primers were supplied by Invitrogen (Life Technologies GmbH, Darmstadt, Germany) and diluted to 100 mM stock solutions with sterile dH$_2$O and stored at -20°C. 10 mM dilutions in sterile dH$_2$O were used as working solutions. dNTP sets (4 x 100 mM solutions) were purchased from Biotechrabbit GmbH (Hennigsdorf, Germany) and 10 mM solutions were prepared with sterile dH$_2$O. 22.5 μl volume reactions were set up in 0.2 ml PCR SingleCAP 8er-SoftStrips (Biozym, Hessisch Oldendorf, Germany) according to table 2.

Table 2. Reaction setup for 22.5 μl volume PCR using Taq Ampliqon® DNA polymerase

Component	Volume for 22.5 μl PCR Reaction
Nuclease-free water	to 22.5 μl
10x Buffer	1.5 μl
25 mM MgCl$_2$	1.5 μl
10 mM dNTPs	2.25 μl
BSA	2.25 μl
10 mM Forward Primer	0.45 μl
10 mM Reverse Primer	0.45 μl
Template DNA	4.5 μl (~10-25 ng)
Taq DNA Polymerase	0.07 μl

All reaction components were mixed and centrifuged prior to use and Taq Amplicon® DNA Polymerase was added last in order. PCR tubes were transferred to an Eppendorf® Mastercycler® (ep. 384), preheated to 98°C and thermocycling was performed according to the settings shown in table 3.

Table 3. Touchdown PCR conditions for PCR using Taq Ampliqon® DNA polymerase

Step		Temperature (°C)	Time
1)	Reaction start	95	3 minutes
2)	Denaturation	95	30 seconds
3)	Primer annealing	68 (-1°C per cycle to 52°C)	30 seconds
4)	Primer elongation	72	30 to 90 seconds
Step 2-4 repeated for 16 cycles (touchdown cycles)			
5)	Denaturation	95	30 seconds
6)	Primer annealing	52	30 seconds
7)	Primer elongation	72	30 to 90 seconds
Step 5-7 repeated for 20 cycles (amplification cycles)			
8)	Final extension and hold	72	5 minutes
9)	Hold	4	hold

6x DNA loading dye (ThermoFisher, Oberhausen, Germany) was added to 6 µl of PCR products and the polymorphism was detected by agarose gel electrophoresis at concentrations from minimum 1.5 % (w/v) to 4% (w/v), depending on PCR product sizes. The running buffer stock solution and the agarose gels were prepared as shown in tables 4 and 5. Chemicals and solutions were purchased from Sigma-Aldrich (Taufkirchen, Germany).

Table 4. Recipe for preparation of 5 liter 50x TAE running buffer stock solution

Solution component	Component volumes
Tris base	1211 g
Na_2EDTA	146.2 g
Glacial acetic acid	290 mL
dH_2O	to 5 L
adjust pH to 8.0 with acetic acid	

Table 5. Recipe for preparation of 1-4% agarose gel

Solution component	Component volumes
Agarose	1.6 to 6.4 g (1 to 4 %)
1x TAE buffer	160 mL
Ethidium bromide (10 mg/mL)	3.2 µL

If a sufficient amount of PCR product was obtained, 6 µL of the PCR reaction solution were digested with the respective restriction enzyme and the buffer recommended by the supplier (New England Biolabs GmbH, Frankfurt am Main, Germany) at the enzyme-specific temperature for 1 hour to overnight. Reactions were set up according to table 6. 6x DNA loading dye (ThermoFisher, Oberhausen, Germany) was added to each sample, and samples were run on 3 to 4 % agarose gels (depending on restriction fragment size differences).

Table 6. Reaction setup for standard restriction digest mix

Component	10 µl Reaction
PCR product	6 µl
NEB buffer (10x)	1 µl
BSA	200 µg/mL (optional)
Restriction enzyme	½ Unit
dH₂O	to 10 µl

Detailed information about SNP markers used for QTL confirmation and/ or ˈcleaningˈ of the L94 NILs are given in table 7, which is continued over several pages. Markers were either derived from the Barley Oligonucleotide Pool Assay 1 (BOPA1) (Close et al., 2009) or were designed on basis of the RNA-seq data and sequences were derived from the barley high- and low confidence gene set (The International Barley Sequencing Consortium, 2012). Markers were tested for the whole population by the company LGC (https://www.lgcgroup.com/) using KASP™ genotyping assays.

Table 7. Names and sequences of SNP markers used for QTL confirmation and ˈcleaningˈ of the L94 NILs

Markers derived from the barley consensus map (Close et al., 2009) are named after the corresponding U35 unigene, BOPA names are additionally indicated in brackets. Markers derived from the barley high- and low-confidence gene set (The International Barley Sequencing Consortium, 2012) are named after the corresponding locus and are marked in bold. Chromosome positions in centimorgans (cM) are given according to the two gene sets. The seven barley chromosomes are abbreviated as 1H to 7H. Forward and reverse primer sequences are indicated by F or R, respectively. The expected size of the PCR products is given in base pairs (bp). PCR products were digested with the indicated restriction enzymes according to the manufacturer (New England Biolabs GmbH, Frankfurt am Main, Germany).

Marker name	Position [Chr, cM]	Sequence	Size PCR product	Restriction enzyme
1285 (11_0346)	1H, 18.05 cM	F: GGAGCAAGAGGAACTTCAAAA R: GACGGTGCTTCCAAAAGTGT	181 bp	EarI
17898 (11_1060)	1H, 20.89 cM	F: TCACTCGCCCTTCTGCTTTT R: CTGAACGAGTGGCTCCTG	135 bp	ClaI
17361 (11_0875)	1H, 23.86 cM	F: GGGCAAGGAGATGGATCGGCTCA R: AGGGCGTCGTCGACTACCTCGGGAT	122 bp	FokI
2055 (11_0585)	1H, 33.16 cM	F: GGCAGCCCAAGTCCAAGAGG R: ATATGTCACAGCCACATGTAGCGTC	139 bp	TaqI
7003 (12_0773)	1H, 41.76 cM	F: TGTTGGAGAAGTGGTCGCGAAT R: CACATATTACACACGGAACAGAGTA	81 bp	RsaI

Continuation of Table 7

17863 (11_1052)	1H, 43.27 cM	F: CCAACAATCCTGCTGCCGTT R: AGAGTTCTGATCAGGTGAAGCGCAC	145 bp	BsrI
19855 (11_1239)	1H, 51.70 cM	F: ATCAAAACGAGGAAGCTCCA R: CTGTTCGCAGGACTGTTTCA	150 bp	AccI
14623 (11_0071)	1H, 60.77 cM	F: TCAACGAGGAACCTGCGAAC R: ATTCTTGATATAGAAGAACACGGCG	147 bp	HhaI
16844 (11_0672)	1H, 61.52 cM	F: CAGGGAGACACCAGTGGATGCACTC R: TTGTCGGAGATCCGCAAGG	125 bp	DdeI
15515 (11_0566)	1H, 65.52 cM	F: TCTGGGTTGAGATGAACGAGT R: TCGAATGTCTGAGGCACATG	111 bp	Fnu4III
18187 (11_0059)	1H, 65.96 cM	F: CGTCCCTCCTCGAACACACCGG R: GCGAGCCGCCGGAATCTATCGACCG	130 bp	HpaII
17380 (11_0004)	1H, 66.70 cM	F: TTTTGATGTCACGATGTGCTC R: GTCTGTTTTTGATCGATCAGTCGTA	134 bp	RsaI
15416 (11_1334)	1H, 71.42 cM	F: TCAAGAAAAGAAAACAAAAAACGAA R: TAGCCACGTTGCAATTTTGT	140 bp	MboII
3916 (11_1046)	1H, 73.94 cM	F: GAAAGCAGCGAGCGGTGTAACAACA R: GTTGAGATCTGAGCGGCATGTGCCG	66 bp	AciI
13884 (11_0312)	1H, 75.44 cM	F: GTACTCTTTTGCGACGCTGAGTCGG R: AGCAATCCAAACCACGCACA	150 bp	Hpy188I
MLOC_36339.1	**1H, 80.24 cM**	**F: GTTCCAGGGTGTTCTCAAGCGT** **R: CTTCTGCAGCGGTTTCAGAGTCCGC**	**159 bp**	**HhaI**
AK363419	**1H, 83.56 cM**	**F: AGGCGGAAAGGACCTCTCTGGCCTT** **R: CGACGACGCTGGAGAAGGACC**	**168 bp**	**DdeI**
AK248239.1	**1H, 84.47 cM**	**F: CAGGGCTTCCTTTTACAGAGATTCC** **R: TCCTGTACTGAAGTGCTTTGT**	**159 bp**	**AciI**
17561 (11_1313)	1H, 84.68 cM	F: AGTTGGTGGTCTTTGGGACA R: TGATGAGAATGGACCTGACAA	175 bp	HhaI
17780 (11_1094)	1H, 88.22 cM	F: GCATTAATCAATGTGTTCATTTCTT R: AATTTTACACCCACAGATCC	78 bp	MseI
16783 (11_0771)	1H, 90.97 cM	F: ACAGTTGCCAGCTTCCACAG R: TGCTACTGCATAATCAGGTGATGCG	99 bp	HhaI
MLOC_6767.1	**2H, 3.82 cM**	**F: CCGTCAAAGACAACAACGTTGTGAT** **R: GTGAACTTGTGCCCGGCGATCTT**	**118 bp**	**MboI**
1246 (11_0363)	2H, 6.44 cM	F: CTGGCGGTCTGGTATGCAC R: GTTCCAAGGTGCAGCAAAC	229 bp	HphI
465 (11_0110)	2H, 54.27 cM	F: TTTCCAAGTGCAGACTCCTT R: CAAGGGTCATTGTTGCGATT	237 bp	Cac8I
21251 (11_0064)	2H, 98.59 cM	F: GCAGTTTCCTGGATCAAATCAATCG R: CGGTGGAGGAGATGAAGGGC	123 bp	TaqI
18000 (11_1008)	2H, 128.26 cM	F: ATGGAGCATATGGATGCAGCTCGAC R: TCACCCATGGAAACACCACCA	78 bp	Hpy99I
MLOC_53038.2	**2H, 129.74 cM**	**F: AAGGAGGGTTCTTTCTGCGGCGAC** **R: GCAGCCTCCTGAACTGACTGGCGAC**	**150 bp**	**Hpy99I**

Continuation of Table 7

14800 (11_0403)	2H, 131.76 cM	F: CCTCGGAGTAAAACAGGTTCA R: GGAGCCTCTGGGATGGAG	237 bp	HpyCH4IV
21251 (11_0064)	2H, 98.59 cM	F: GCAGTTTCCTGGATCAAATCAATCG R: CGGTGGAGGAGATGAAGGGC	123 bp	TaqI
18000 (11_1008)	2H, 128.26 cM	F: ATGGAGCATATGGATGCAGCTCGAC R: TCACCCATGGAAACACCACCA	78 bp	Hpy99I
MLOC_53038.2	**2H, 129.74 cM**	**F: AAGGAGGGTTCTTTCTGCGGCGAC** **R: GCAGCCTCCTGAACTGACTGGCGAC**	**150 bp**	**Hpy99I**
14800 (11_0403)	2H, 131.76 cM	F: CCTCGGAGTAAAACAGGTTCA R: GGAGCCTCTGGGATGGAG	237 bp	HpyCH4IV
MLOC_71778.2	**2H, 137.44 cM**	**F: TGCGGATAGTTCTCTCCAAGCTAAAG** **R: AGCTCTGAGTCCAATGGTAGGT**	**169 bp**	**AccI**
4384 (11_1168)	2H, 137.51 cM	F: CACACCGATACCAACATCTTTGATT R: GCTCTTCTGATTCTGTATTCGG	135 bp	HinfI
MLOC_37267.1	**3H, 9.24 cM**	**F: AAACATTGCACTTCAACAACCTGCG** **R: TGATGGTGTAACTAACTTTGGTTGTG**	**112 bp**	**HhaI**
MLOC_15155.2	**3H, 20.75 cM**	**F: ACTGGACCTACTGCCCCAAGATCCT** **R: TCCGTGATGGAGTTGTTGACCTT**	**149 bp**	**MnlI**
2585 (11_1128)	4H, 3.74 cM	F: TGTCACAATTGGTGCCTGAATCTG R: ATCATCTGGTGTTGGTGTTATTGG	242 bp	TaqI
MLOC_10935.1	**4H, 18.48 cM**	**F: TCAAATTTGGTGTCATCTTCAACGA** **R: TCATGGACACCCTGCAGAAGC**	**136 bp**	**Hpy99I**
19453 (11_1254)	4H, 28.40 cM	F: TTTTGGATACATTGAGCCCCCGATGG R: GAGTCAGCCGCACCAGAGAT	232 bp	MseI
MLOC_4376.2	**4H, 61.54 cM**	**F: TCTGGATGGTTGAGACCATGAATCG** **R: CGCTGGCAGGGACGGCATAG**	**300 bp**	**TaqI**
18206 (11_0915)	4H, 93.97 cM	F: GACCAAATGACTGTGGGATGACAA R: CTGACCCGTATATTCATAGGCGTAT	143 bp	TaqI
14974 (11_0664)	4H, 119.08 cM	F: GGAATGAACGACTAAAGAAGATATC R: CTTCCAGTGCGCTTAAATCC	115 bp	TaqI
MLOC_4807.1	**5H, 115.27 cM**	**F: GTGTATTACCGGAAAAGGCAAGCAG** **R: AGGGTTCGTCGGGATTCTTG**	**67 bp**	**MboI**
4175 (11_1071)	6H, 53.94 cM	F: TGGACACGTATAGGGACCAT R: TTTACCAAGCGGCCTGAG	235 bp	SacI
MLOC_5565.1	**6H, 63.45 cM**	**F: CTTCCATTTCCACTGTAAGAAGC** **R: GAATTGCTCCCTCTGGTT**	**109 bp**	**BccI**
2296 (11_0722)	6H, 71.87 cM	F: GTATCCGCAAACCCTCACTG R: CGGAAGGAAAGAGTTTGTCG	225 bp	TaqI
MLOC_3574.2	**6H, 98.05 cM**	**F: GTACCAAAGCTAGAAATCGGGAATCC** **R: AGAAGCTGCTCTTGCTGCTG**	**105 bp**	**Hpy188I**
MLOC_2119.2	**7H, 75.80 cM**	**F: AATCTAGATCGTCCGATCATCCG** **R: GGTGTATGTCCAGTGCTCCAAG**	**132 bp**	**AciI**
859 (11_0212)	7H, 116.33 cM	F: TCACTCCATTTCAAGGTTTTCCGTA R: TTTGCTCTGTTGAGGATCATTTACT	147 bp	AccI

RNA extraction from barley embryos

Approximately 20 embryos were excised from mature barley seeds by carefully slicing open the endosperm with a sterile scalpel, avoiding cutting the embryo, and removing the embryo with a sterile forceps. Embryos were collected in sterile 1.5 ml Eppendorf tubes and flash frozen in liquid nitrogen. The samples may be stored at -80°C before continuing with the RNA extraction procedure. The embryo material was ground using a sterile mortar and pestle, which had been chilled with liquid nitrogen. RNA was extracted from the ground material using the RNAEasy® Mini Kit (50) from Qiagen (Qiagen, Hilden, Germany) following the manufacturer´s protocol and stored at -80°C after DNase treatment (Ambion, Carlsbad, USA).

RNA-sequencing and data analysis

RNA-sequencing was performed by the Genome Center of the Max Planck Institute of Plant Breeding research (MPIPZ, Cologne, Germany). The quality of the RNA was evaluated before library preparation using a bioanalyzer (Agilent). The Illumina TruSeq libraries were prepared using the manufacturer's protocol (version 2, Illumina). Single end sequencing was performed on the HiSeq 2000 (Illumina®) platform. For each library, a minimum of 15 million reads was generated by multiplexing eight libraries. The initial quality control of the raw reads was performed using the FastQC software. Reads were trimmed using the Trimmomatic platform, embedded within the trinity pipeline (Grabherr et al. 2011; https://github.com/trinityrnaseq/trinityrnaseq/wiki), using the following default criteria: phred 33, leading and trailing 3, sliding window 4:15 and a minimum read length of 36. Sequence alignments were performed by Bowtie2 (version 2.1.0; http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) using a merged dataset of High Confidence (HC) and Low Confidence (LC) predicted barley genes (The International Barley Genome Sequencing Consortium, 2012) as reference. SAMtools (Li et al. 2009, version 1.2; http://www.htslib.org/) was used for BAM format conversion, sorting and indexing and read duplicate removal was conducted using the Picard command-line tool MarkDuplicate (version 1.110; https://broadinstitute.github.io/picard/command-line-overview.html). To correct misalignments, the Genome Analysis ToolKit (GATK, version 3.1) realigner was used with recommended settings. Variants were obtained using the GATK UnifiedGenotyper platform (minimum phred score of 30). Refinement of the variants was performed with the GATK Variant Filtration tool (Fisher Strand values FS >30.0; Qual By Depth values QD <2.0) to reduce false positive SNPs. Resulting SNP calls were kept for further analysis if they passed the filtration step and their read coverage exceeded four reads. Transcripts containing filtered homozygous SNPs were mapped to their respective position along the POPSEQ map of barley with R (The International Barley Genome Sequencing Consortium, 2012).

Controlled deterioration tests

Seed longevity was determined as germination ability after an artificial ageing test. Mature, freshly harvested barley seeds were dried in a drying room conditioned at circa 10% relative humidity and 20°C for two months before starting the controlled deterioration tests. Samples of 30 seeds were packed in 2 ml Eppendorf tubes in the drying room. For each genotype at least 10 independent lines

were used, and 4 replicates were performed per each time point. In order to maintain the equilibrium relative humidity, the Eppendorf tubes were sealed with parafilm. Thereafter, they were hermetically sealed in laminated aluminum foil packets. These packages were placed into an Binder BFF115 incubator (Binder, Tuttlingen, Germany) kept at 60°C and were removed at different intervals and tested for germination at 20°C (12 hours light and 12 hours dark) between pleated papers. Seedlings were scored as normal or abnormal according to the international rules for seed testing (ISTA) after 14 days. The number of normal seedlings was plotted against the storage time. The resulting seed survival curves were transformed according to the so-called probit analysis (Ellis and Roberts, 1980) and the values for Ki and σ were determined.

References

Badr, A. et al. (2000), On the origin and domestication history of barley (*Hordeum vulgare*). Molecular Biology and Evolution 17: 400-510

Bentsink, L. et al. (2010), Natural variation for seed dormancy *in Arabidopsis* is regulated by additive genetic and molecular pathways. Proc. Natl. Acad. Sci. 107: 4264-4269

Chen, Z. et al. (2007), Potassium and sodium relations in salinized barley tissues as a basis of differential salt tolerance. Functional Plant Biology 34: 150-162

Close, T.J.et al. (2009), Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10: 582

Collard, B.C.Y. et al. (2005), An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169-196

Ellis, R.H. and Roberts, E.H. (1980), Improved equations for the prediction of seed longevity, Ann. Bot. 45: 13-30

Epstein, E. et al. (1980), Saline culture of crops: a genetic approach. Science 210: 399-404

Grabherr, M. et al. (2011), Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology 29: 644-652

Hampton, J.G. and TeKrony, D.M. (1995), *Handbook of Vigour Test Methods*, 3rd edn. International Seed Testing Association, Zürich, Switzerland

ISTA. International rules for seed testing. (1999), Seed Sci. Technol. 27

Keurentjes, J.J.B. et al. (2007), Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics, 175: 891-905

Konieczny, A., and Ausubel, F.M. (1993), A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4: 403-410

Koornneef, M. et al. (2004), Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu. Rev. Plant Biol. 55: 141-172

Li, H. et al. (2009), Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25: 2078-2079

Long, R.L. et al. (2008), Seed persistence in the field may be predicted by laboratory-controlled aging. Weed Science 56: 523-528

Loudet et al. (2003), Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis*. Plant Physiol. 131: 345-358

Mackay, T.F.C. et al. (2009), The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10: 565-577

Nagel, M. et al. (2014), Genome-wide association mapping and biochemical markers reveal that seed ageing and longevity are intricately affected by genetic background and developmental and environmental conditions in barley. Plant Cell and Environment 38: 1011-1022

Neff, M.M. et al. (1998), dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J. 14: 387-392

Niks, R.E. and Kuiper, H.J. (1983), Histology of the relation between minor and major genes for resistance of barley to leaf rust. Phytopathology 73: 55-69

Pauling, L.C. (1988), General Chemistry, Dover Publications

Probert, R.J. (2009), Ecological correlates of *ex situ* seed longevity: a comparative study on 195 species. Annals of Botany 104: 57-69

Qi, X. et al. (2000), The evidence for abundance of QTLs for partial resistance to *Puccinia hordei* on the barley genome, Molecular Breeding 6: 1-9

Rakshit, S. et al. (2012), Multiparent intercross populations in analysis of quantitative traits. J. Genet. 91: 111-117

Roberts, E.H. (1973), Predicting the storage life of seeds. Seed Science and Technology 1: 499-514

Seymour, D.K. et al. (2012) Rapid creation of *Arabidopsis* doubled haploid lines for quantitative trait locus mapping. Proc. Natl. Acad. Sci. 109: 4227-4232

The International Barley Genome Sequencing Consortium (2012), A physical, genetic and functional sequence assembly of the barley genome. Nature 491: 711-716

Tuinstra, M.R. et al (1997), Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theoretical and Applied Genetics 95: 1005-1011

Ullrich, S.E. (2011), Barley: Production, Improvement, and Uses. Wiley-Blackwell

Von Bothmer, R. (1992), The wild species of *Hordeum*: Relationships and potential use for improvement of cultivated barley. Chapter 1. In: P.R. Shery, ed. Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology. C.A.B. International, Wallingford, Oxon. 3-18

Woldeab, G. et al. (2015), Virulence Spectrum *of Puccinia hordei* of Barley in Western and Central Highlands of Ethiopia. Adv. Crop Sci. Tech. 3: 164

Chapter 3

Analytical Approaches to Understand the Complex
Biochemical and Molecular Processes
Underlying Seed Longevity

Introduction

In the first decade of the 21st century, genomic studies in model plants promoted gene discovery and provided a breadth of gene function knowledge. Driven by recent technological advances, in the second decade, 'omics'-based approaches have made it possible to address the complex global biological systems that underlie various plant functions also in genetically more complex species, including many crop plants. The term 'omics' is collectively used to describe the comprehensive analysis of all biological components of a given system, and often consists of systems biology approaches to assimilate tremendous amounts of data generated by 'omics'-technologies (Yuan et al., 2008). These technologies span an increasingly wide range of fields from genomics (the quantitative study of protein-coding genes, regulatory elements and non-coding sequences), transcriptomics (RNA and gene expression), proteomics (e.g. focusing on protein abundance) to metabolomics (metabolites and metabolic networks) (Mayer, 2011). Advances in DNA-sequencing technology have revolutionized the field of biological sciences, featuring unprecedented innovations in sequencing scale and throughput, and implementations of various novel applications beyond genome sequencing. In particular, next-generation sequencing (NGS) technology provides feasible applications such as whole-genome re-sequencing for variation analysis, RNA-sequencing (RNA-seq) for transcriptome and non-coding RNAome analysis, quantitative detection of epigenomic dynamics, and Chip-seq analysis for DNA-protein interactions (Lister et al., 2009). Additionally, *de novo* assembly from RNA-seq analysis together with the published genomes of related species enable preliminary characterization of novel genomes (Fu et al., 2011; Su et al., 2011).

Mass spectrometry is the most commonly used method for the investigation/identification of analytes in both proteomics and metabolomics. A mass spectrometer measures the mass-to-charge ratio (m/z) of gas-phase ions. In a prototypical 'omics'-experiment, the sample is delivered to the mass spectrometer via a chromatographic device, e.g. liquid chromatography (LC) column, ionized and vaporized in the ion-source. The resulting ions are then separated according to their m/z values and detected to create a mass spectrum, which is characteristic of the molecular mass and/or structure analyzed.

The rise of new sequencing technologies, advances in a wide range of analytical technologies and the emergence of bioinformatics procedures, have contributed to improvements in the quality of research not only in model species, but also in crop plants. Several essential features of barley (*Hordeum vulgare* L.) contribute to the broad utilization of this crop in genetic studies. These features include (1) the crop's diploid nature with a high degree of inbreeding; (2) the low chromosome number (2n = 14), where individual chromosomes are very large in size; (3) the ease of cross-breeding; and (4) the ease of cultivation in a wide range of climatic conditions (Saisho and Takeda, 2011). With a genome size of 5.1 gigabases (Gb), barley has one of the largest and most complex genomes of all economically important food crops. In 2012, a genome-wide physical map of 4.98 Gb of the barley cultivar (cv.) Morex was constructed by high-information-content fingerprinting and contig assembly of 571.000 bacterial artificial chromosome (BAC) clones (The International Barley Genome Sequencing Consortium, 2012). The transcribed complement of the barley gene space was annotated

by mapping 1.67 billion RNA-seq reads (167 Gb) obtained from eight stages of barley development, as well as 28.592 barley full-length cDNAs (Matsumoto et al., 2011) to the whole-genome shotgun assembly. Exon detection and consensus gene modelling revealed 79.379 transcript clusters. Based on a gene-family-directed comparison with the genomes of *Sorghum*, rice, *Brachypodium* and *Arabidopsis*, 26.159 of these transcribed loci fall into clusters and have homology support to at least one other reference genome; they were defined as high-confidence (HC) genes. Due to lack of homology and missing support from gene family clustering, 53.220 transcript loci were considered low-confidence (LC). A characteristic of the barley genome is the abundance of repetitive DNA (Wicker et al., 2009). Approximately 84% of the genome is comprised of mobile elements or other repeat structures. The majority of these consists of retrotransposons, 99.6% of which are long terminal repeat (LTR) retrotransposons (The International Barley Genome Sequencing Consortium, 2012).

Access to 'omics'-tools, including the inventory of genes deduced from the analysis of the *Arabidopsis thaliana* genome, has revolutionized the way research is being performed in different fields of seed biology. Transcriptome analyses were performed to study various seed developmental processes, such as the transition from cell division to cell enlargement in the embryo (Le et al., 2010; Belmonte et al., 2013). A global transcriptome analysis in whole *Arabidopsis* seeds was performed by Nakabayashi et al. (2005). Full-genome microarrays were used for a global transcript analysis of *Arabidopsis thaliana* Cape Verde Island accession (Cvi) seeds in a range of dormant and dry after-ripened states during cycling, indicating a common underlying dormancy control mechanism (Cadman et al., 2006). Proteomic approaches have provided valuable information about the accumulation of various classes of proteins during seed germination and priming (Gallardo et al., 2001). Proteomics approaches have also been proposed as a powerful tool for determining the biological roles and functions of individual proteins and identifying the molecular mechanisms of seed germination, vigor and viability during ageing. Rajjou et al. (2004; 2007) reported that *de novo* mRNA synthesis is not required for seed germination, and that *Arabidopsis* seeds are able to germinate in the presence of the transcription inhibitor α-amanitin. These studies emphasize the importance of RNAs present in the mature seed for germination. Comparison of transcriptome and proteome data has revealed the extent of post-transcriptional regulation during various phases of seed development, as illustrated for biotin metabolism (Chen et al., 2009). Metabolome studies targeted to the seed maturation and desiccation phases (Ruuska et al., 2002; Fait et al., 2006) have been used to build models of metabolic networks operating during seed development and maturation. Currently, 'omics'-approaches are being used to characterize seed vigor, which encompasses seed longevity, tolerance of germination to environmental stresses, and the uniformity and speed of seed germination and seedling establishment. These studies are expected to deliver new markers of seed quality that can be used in breeding programs and/or in biotechnological approaches to improve crop yields.

Using transcriptomic and metabolomic profiling together with a conditional-dependent network of global transcription interactions, Verdier et al. (2013) dissected late maturation events related to the acquisition of desiccation tolerance and longevity in *Medicago truncatula* seeds. The network revealed distinct co-expression modules related to the acquisition of desiccation tolerance, longevity, and pod abscission. The longevity module was enriched in genes involved in RNA processing and translation.

Seed longevity genes were highly connected to two *APETALA2*/ethylene-responsive element binding protein (*AP2*/EREBP) and two basic leucine zipper transcription factors. A heat shock factor was found at the transition of desiccation tolerance and longevity modules, connecting to both gene sets. Metabolome profiles indicated that glucose and stachyose levels are increased and correlate with longevity during seed maturation. The seed longevity proteome has been investigated in dry and imbibed *Arabidopsis* seeds after natural and artificial ageing (Rajjou et al., 2008). The translational potential of dry seeds is reduced during ageing, therefore the maintenance and repair of the functional translational machinery is important for seed longevity (Rajjou et al., 2008). Post-translational modification (PTM) regulation in dry seeds has been shown to play central role in seed dormancy release, metabolism resumption, and ageing processes (Arc et al., 2011). In this study, it was also demonstrated that the elevated accumulation of oxidation (carbonylation) in dry seeds is associated with ageing, and might induce loss-of-function protein and enzyme properties. Therefore, detoxification of reactive oxygen species (ROS), that result in oxidation stress and maintenance of redox homeostasis, are crucial for seed vigor (Rajjou et al., 2008). The detoxification of toxic compounds, e.g. cyanine, accumulating during seed storage and germination, by β-mercaptopyruvate sulfurtransferase (MST) seemed to be related to seed longevity since the abundance of this enzyme was reduced during seed ageing (Rajjou et al., 2008). Proteome analyses in maize (*Zea mays*) seeds upon ageing showed that proteins involved in energy metabolism (glycolysis, tricarboxylic acid cycle, the electron transport chain and oxidative phosphorylation) formed the largest group of down-regulated proteins (Wu et al., 2011; Xin et al., 2011). These findings suggest that the translational capacity and efficiency in seed reserve mobilization are essential mechanisms for seed longevity. Recently, the dry seed proteome of four *Arabidopsis thaliana* genotypes, that carry introgression fragments at the position of seed longevity quantitative trait loci, was investigated (Nguyen et al., 2015). Seeds at two physiological states, after-ripened seeds and aged (stored) seeds, were compared. Aged dry seed proteomes were markedly different from the after-ripened, and reflected the seed longevity level of the four genotypes. Results confirmed the important role of antioxidant systems, notably vitamin E, and indicated that protection and maintenance of the translation machinery and energy pathways are essential for seed longevity. Moreover, a new role for seed storage proteins (SSPs) was identified in dry seeds during ageing. SSPs are suggested to buffer the seed from oxidative stress, thus protecting important proteins required for seed germination and seedling formation.

Objectives

The above mentioned studies revealed that seed longevity is a complex trait, which requires a better understanding of the underlying molecular mechanisms and complex biochemical processes. These processes can be studied by comparing seeds that differ in the times that they were stored, but also by comparing genotypes that differ in their seed longevity properties. For this purpose, a transcriptome, proteome and metabolome analysis was performed on the parental lines of the L94 NIL mapping population, L94 and Cebada Capa (CC), the recombinant inbred line 114 (RIL114) and

L94 near isogenic lines (L94 NILs). A total seed proteome profiling was performed to identify proteins that may affect the observed differences in seed longevity between L94 and CC. Making use of the recent improvement of the barley genomics infrastructure, an RNA-seq analysis was conducted to identify differentially expressed candidate genes and possible downstream targets. Lastly, mass spectrometry (MS)-based metabolomics was applied to gain insight into biochemical processes affected in seed longevity. For all analyses, mature, non-aged seeds were utilized.

Results

RNA-seq analysis of L94, Cebada Capa, RIL114 and the L94 NILs

RNA-seq was employed to identify differentially expressed candidate genes and possible downstream targets for seed longevity, the trait the L94 NILs differ from the recurrent parent L94 (chapter 2). These L94 NILs harbor CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ). Three SLQs are found on chromosome 1, SLQ1.1-1.3, and one SLQ on the bottom of chromosome 2 (SLQ2). The ˈcleanestˈ NILs, showing only CC introgressions in the region of one seed longevity QTL, were derived from NILs sequenced in the year 2014. Those NILs still showed introgressions on other chromosomes. The generation of the ˈcleanˈ NILs is discussed in detail in chapter 2. As shown in chapter 2, the line from which NIL-SLQ1.1 was derived, NIL-SLQ1.1(2014), showed more introgressions than the other NILs sequenced in 2014 and the RIL114. Therefore, the NIL-SLQ1.1(2014) was included in the RNA-seq analysis. The RIL114, which showed the strongest seed longevity phenotype in the original L94 RIL mapping population (Adimargono et al., in preparation), and which was selected for backcrossing to L94 to generate the L94 NILs, was also included in the analysis. A schematic representation of all analyzed genotypes is shown in figure 1a.

As shown in figure 1a, the RNA-seq libraries of the analyzed genotypes have between 6.5 and 9.5 million sequenced reads. The library of the NIL-SLQ1.2 has a very broad size range and the RNA-set obtained for the NIL-SLQ1.1 comprises one clear outlier. For further analyses, this outlier of the NIL-SLQ1.1 was removed. To determine how closely the expression profiles of the samples cluster together, a multi-dimensional scaling (MDS) plot was performed. As it can be observed in figure 1b, the replicates of each genotype cluster together. Dimension 1 of the MDS plot separates the samples according to their amount of CC introgressions. CC is clearly separated from the rest of the samples followed by NIL-SLQ1.1(2014) and RIL114. This observation can be explained by the fact that the NIL-SLQ1.1(2014) harbors more CC introgressions than the RIL114, as shown in the schematic presentation of the analyzed genotypes in figure 1a. Interestingly, the replicates of L94, NIL-SLQ1.2, NIL-SLQ1.3 and NIL-SLQ2 cluster together, while the replicates of the NIL-SLQ1.1 are slightly separated. This might be explained by the fact that the introgression in this NIL on chromosome 1 is bigger than the ones of the other NILs on the chromosomes 1 and 2, respectively. Furthermore, this NIL was derived from NIL-SLQ1.1(2014), which harbors additional CC introgressions. Comparing the samples of RIL114 or NIL-SLQ1.1(2014) to the L94 NILs, the process of the ˈcleaningˈ of the L94 NILs, the reduction of introgressions by backcrossing the RIL114 and NIL-SLQ1.1(2014) to L94, is clearly visible.
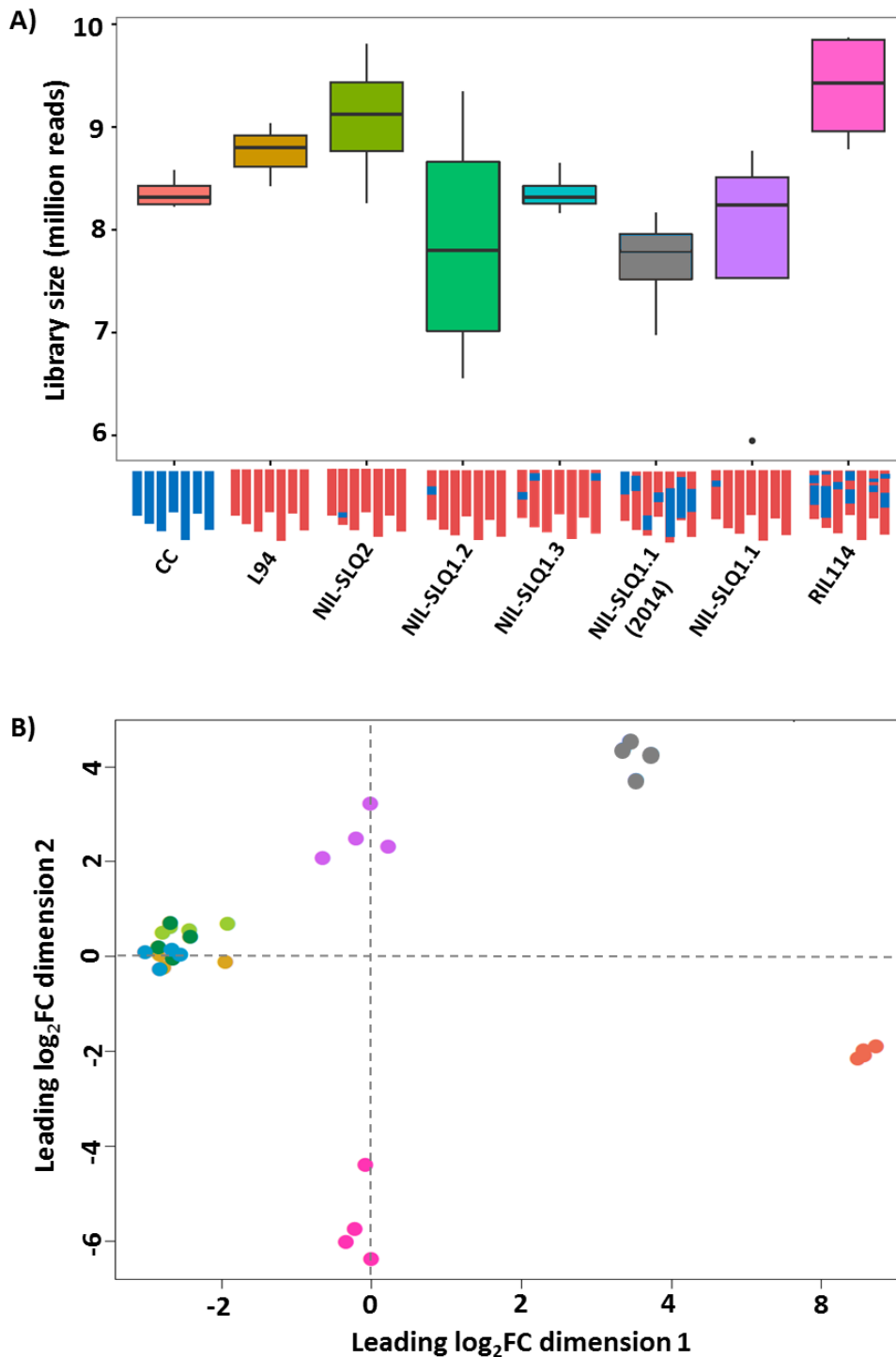
Figure 1. **Box plot of library sizes and multi-dimensional scale (MDS) plot of RNA-data sets derived from the sequencing of embryos from mature, non-aged seeds of L94, Cebada Capa, RIL114 and the L94 NILs**

A) Library sizes are presented as number of short-mapped reads in sequencing a given library. The in the study used genotypes are schematically presented below the box plot. The seven barley chromosomes are colored in red and blue, representing the genomic regions of L94 and introgression segments from Cebada Capa (CC), respectively. Near isogenic lines (NILs) show CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or on chromosome 2 (SLQ2). The in 2014 sequenced NIL is indicated by a bracket behind its name. The recombinant inbred line 114 (RIL114) was also included in the analysis. B) Multi-dimensional scaling (MDS) plot summarizing all RNA-Seq datasets used in this study, leading $\log_2$-fold changes ($\log_2$FC) in dimension 1 and 2 are indicated. The color code, used for the samples, is indicated in A).

Identification of possible candidate genes and downstream targets

RNA-sequencing can be a useful tool to narrow down the number of candidate genes, which comprise differentially regulated genes, as well as transcripts with non-synonymous SNPs, located within the QTL regions. Differentially regulated genes not located within the QTL regions, would be possible downstream targets. Using the Bioconducter package edgeR, pairwise comparisons in gene expression were performed between the samples and the short-lived parental line L94. Differentially regulated reads were identified using a false discovery rate (FDR) of 0.05. Genes with a $\log_2$-fold change ($\log_2$FC) above 1 or below -1 were considered as up- or down-regulated genes, respectively. Table 1 gives an overview over the amount of up- and down-regulated and not-differentially regulated genes in the different samples in comparison to L94. Not surprisingly, the highest amount of up- and down-regulated genes is found in Cebada Capa. The second largest number of differentially regulated genes is found in the NIL-SLQ1.1(2014), which is in accordance with its high amount of CC introgressions. Of the L94 NILs, the NIL-SLQ1.1 and NIL-SLQ1.3 have the largest amount of differentially expressed genes, both having larger CC introgressions than the rest of the L94 NILs. Interestingly, the number of up-regulated genes is only higher than the number of down-regulated genes in the RIL114 and NIL-SLQ2 samples.
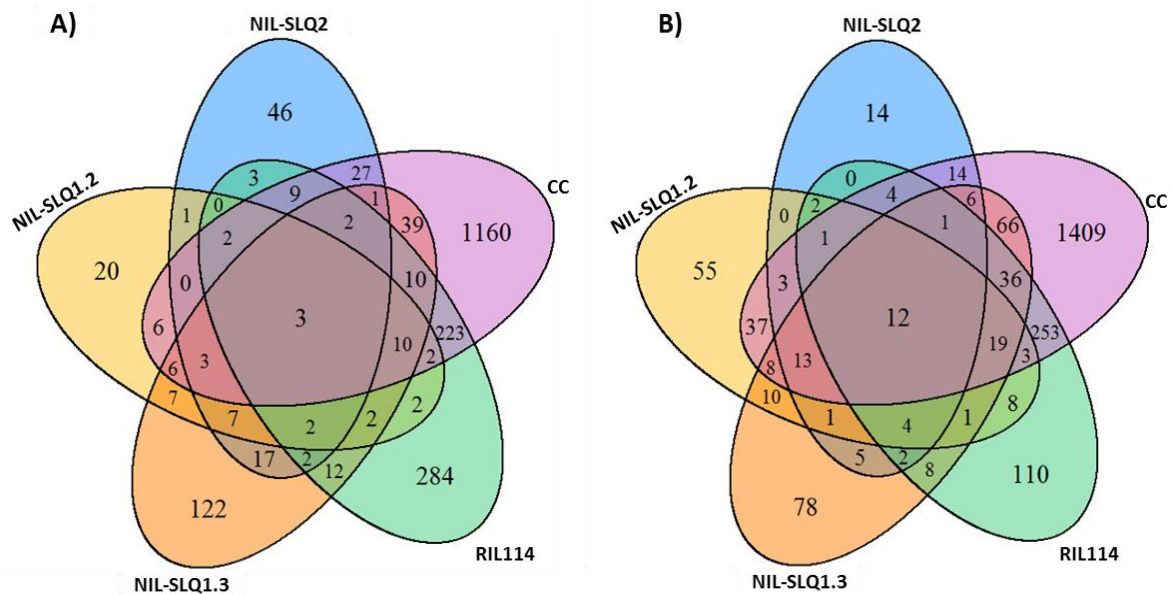
Table 1. Overview over the amount of up- and down-regulated and not-differentially expressed genes in RNA-seq data sets of Cebada Capa, RIL114 and L94 NILs in comparison to L94

The to the parental line L94 compared genotypes comprise Cebada Capa, the recombinant inbred line 114 (RIL114) and near isogenic lines (NILs) harboring CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). The in 2014 sequenced NIL is indicated by a bracket behind its name. Differentially regulated genes were called with a false discovery rate of 0.05. Up-regulated genes have a $\log_2$-fold change ($\log_2$FC) above 1 (>1), and down-regulated genes a $\log_2$FC below -1 (< -1) in comparison to L94.

Sample compared to L94	Down-regulated genes ($\log_2$FC < -1)	Not-differentially expressed genes	Up-regulated genes ($\log_2$FC > 1)
Cebada Capa	1889	31925	1505
RIL114	465	34286	568
NIL-SLQ1.1(2014)	1170	33325	824
NIL-SLQ1.1	335	34804	180
NIL-SLQ1.2	177	35069	73
NIL-SLQ1.3	270	34804	245
NIL-SLQ2	82	35112	125

To identify genes showing a similar gene expression pattern both in the long-lived parental line Cebada Capa, the RIL114 and the L94 NILs, overlaps between differentially expressed genes in these genotypes were performed. Since the L94 NILs cluster closely together apart from the NIL-SLQ1.1, as shown in figure 1b, an overlap was performed for those three NILs with the RIL114 and Cebada Capa. A second overlap was performed for the differentially expressed genes in the NIL-SLQ1.1, the NIL-SLQ1.1(2014) and Cebada Capa. The results of these overlaps, presented in the form of Venn diagrams, are shown in figure 2 and 4.
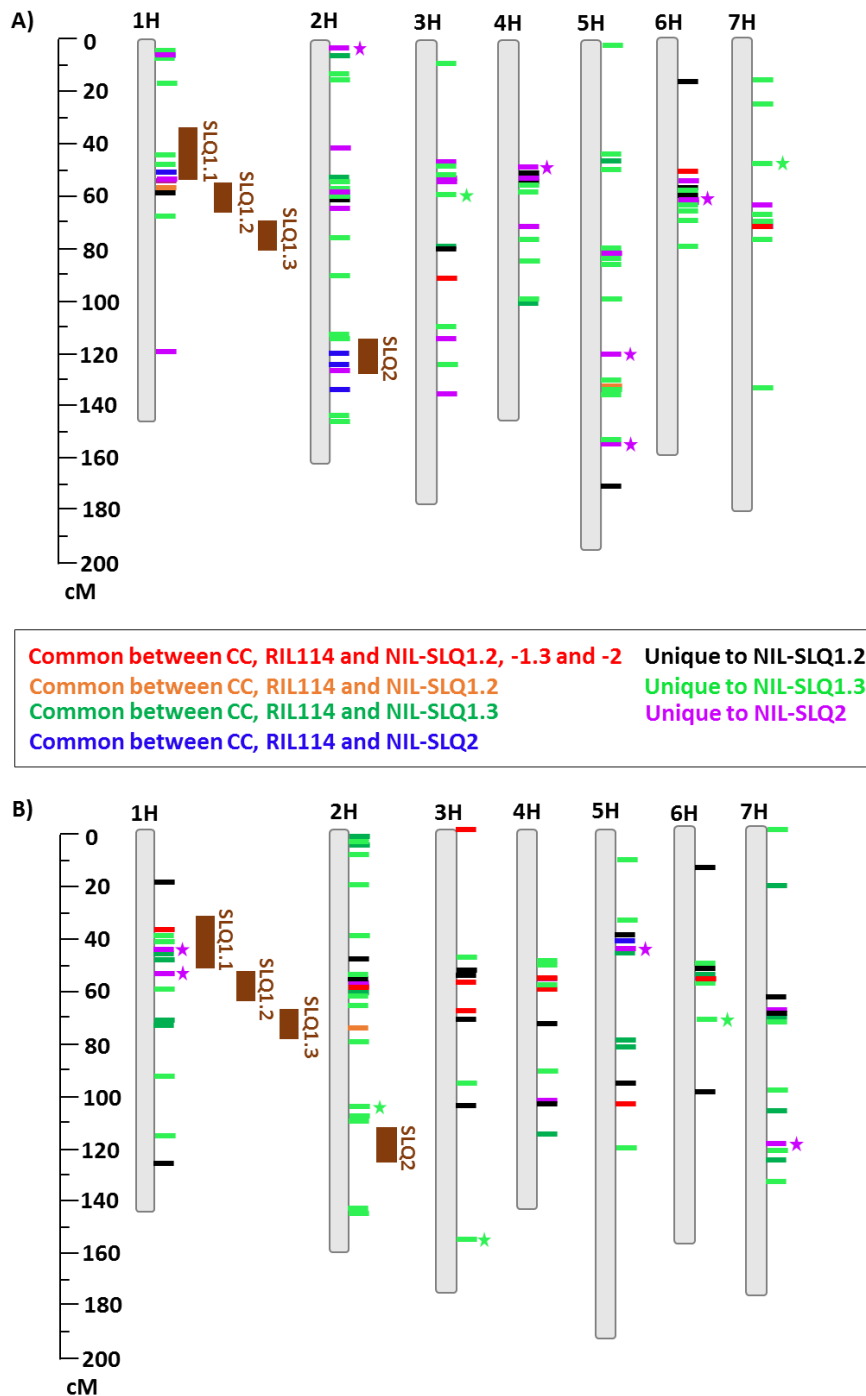
Figure 2 shows the overlaps of up- or down-regulated genes in RNA-seq data sets of Cebada Capa, the RIL114, NIL-SLQ1.2, NIL-SLQ1.3 and NIL-SLQ2. It can be observed that the number of commonly shared up-regulated genes (2a) is four times lower than the number of commonly shared down-regulated genes (2b). In both overlaps, there are a number of genes, which are uniquely found in one of the NILs. For the NIL-SLQ1.3 and NIL-SLQ2, the majority of these unique genes is up-regulated, while for the NIL-SLQ1.2 the majority of these genes is down-regulated.

Figure 2. **Venn diagrams showing overlaps of up- or down-regulated genes in RNA-seq data sets of Cebada Capa, RIL114 and L94 NILs in comparison to L94**

Differentially regulated genes were called with a false discovery rate of 0.05. Up-regulated genes have a log$_2$-fold change (log$_2$FC) above 1 and down-regulated genes a log$_2$FC below -1 in comparison to L94. Overlaps were performed between RNA-seq data sets from Cebada Capa (CC), the recombinant inbred line 114 (RIL114) and near isogenic lines (NILs) showing CC introgressions in Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.2 and 1.3) or chromosome 2 (SLQ2). A) Overlaps between up-regulated genes. B) Overlaps between down-regulated genes.

To assign the genes present in the overlaps of figure 2 as possible candidate genes or downstream targets, the genomic location of these genes was taken into consideration. Furthermore it was checked whether single nucleotide polymorphisms (SNPs) had been detected in the coding sequences of the specific genes in the RNA-seq data analysis. A schematic representation of the location of up- or down-regulated genes, present in the overlaps of figure 2, is shown in figure 3. The presence of SNPs, detected within the coding sequences of genes between the parental lines, is indicated by stars. As it can be seen in figure 3a, the three commonly shared up-regulated genes are located on chromosomes 3, 6 and 7 and do not contain SNPs. Interestingly, SNPs were only detected in genes uniquely up-regulated in NIL-SLQ1.3 and NIL-SLQ2. Those genes are all located outside of SLQ-regions. In the case of the down-regulated genes, shown in figure 3b, one commonly shared gene is located in the region of SLQ1.1, while the rest is located on chromosomes 3, 4 and 5. As observed for the up-regulated genes, SNPs were only detected in genes uniquely down-regulated in the NIL-SLQ1.3 and NIL-SLQ2.

Figure 3. Schematic representation of the location of up- or down-regulated genes present in the overlaps of the Venn diagrams in figure 2.

The seven barley chromosomes are represented as grey bars and abbreviated as 1H to 7H. Genes are represented as colored stripes and the color code, explaining which genes are shared between which RNA-data sets, is shown in the figure itself. Stars indicate genes with single nucleotide polymorphisms (SNPs) between the coding sequences of Cebada Capa and L94. Sample abbreviations are as follows: Cebada Capa (CC), recombinant inbred line 114 (RIL14) and near isogenic lines (NILs) showing CC introgressions in Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). SLQs regions are indicated as vertical brown bars. Gene positions, derived from the barley gene set (The International Barley Sequencing Consortium, 2012), are given in centimorgans (cM). A) Location of up-regulated genes present in the overlaps of figure 2a, B) Location of down-regulated genes present in the overlaps of figure 2b.

A number of differentially expressed genes are located within the SLQ regions. Table 2 gives an overview over the exact location and annotation of these genes. Genes located within the region of SLQ1.1 are highlighted by light grey shading within the table, genes located in the region of SLQ1.2 are unshaded within the table and genes located within SLQ2 are highlighted by dark grey shading.

Table 2. Annotation and position of up- or down-regulated genes present in the overlaps of the Venn diagrams in figure 2 and located within seed longevity QTLs

Gene annotations and positions in centimorgans (cM) are according to the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The seven barley chromosomes are abbreviated as 1H to 7H. Genes located within the region of the seed longevity QTL (SLQ) 1.1 are highlighted by light grey shading, genes located in the region of SLQ1.2 are unshaded and genes located within SLQ2 are highlighted by dark grey shading within the table. The presence or absence of single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in coding sequences is indicated. Possible candidate genes are marked in bold.

Present in	Locus	Annotation	Position	SNPs	Up-/down-regulated
CC, RIL114, NIL-SLQ1.2, NIL-SLQ1.3, NIL-SLQ2	MLOC_77492.1	Unknown protein	1H, 37.61 cM	no	down
CC, RIL114, NIL-SLQ1.3	MLOC_78169.1	Unknown protein	1H, 48.09 cM	no	down
	MLOC_1467.1	Unknown protein	1H, 48.67 cM	no	down
CC, RIL114, NIL-SLQ2	MLOC_61246.1	Unknown protein	1H, 48.09 cM	no	up
NIL-SLQ1.3	MLOC_71710.1	Unknown protein	1H, 48.22 cM	no	up
	MLOC_77043.1	RmlC-like cupins superfamily protein	1H, 43.05 cM	no	up
	MLOC_66307.1	Major facilitator superfamily protein	1H, 41.93 cM	no	down
	MLOC_57601.1	Alpha/beta-Hydrolases superfamily protein	1H, 38.03 cM	no	down
NIL-SLQ2	MLOC_4500.2	LRR and NB-ARC domains-containing disease resistance protein	1H, 48.09 cM	no	up
	MLOC_39058.1	Unknown protein	1H, 48.37 cM	no	up
	MLOC_52077.2	Homeobox-7	1H, 48.09 cM	yes	down
CC, RIL114, NIL-SLQ1.2	**MLOC_50879.1**	**Unknown protein**	**1H, 54.53 cM**	**no**	**up**
NIL-SLQ1.2	**MLOC_20718.1**	**Ribonuclease H-like superfamily protein**	**1H, 56.87 cM**	**no**	**up**
	AK361269	**Leucine-rich repeat (LRR) family protein**	**1H, 58.22 cM**	**no**	**down**
NIL-SLQ2	MLOC_77012.1	Unknown protein	1H, 54.53 cM	yes	down
CC, RIL114, NIL-SLQ2	**MLOC_60558.1**	**Unknown protein**	**2H, 125.35 cM**	**no**	**up**
	MLOC_73587.1	**Bifunctional nuclease I**	**2H, 132.15 cM**	**no**	**up**
NIL-SLQ2	**MLOC_76120.3**	**Unknown protein**	**2H, 125.21 cM**	**no**	**up**

As it can be observed in table 2, the majority of the genes are located in the region of SLQ1.1. Out of these genes, only one gene shows a SNP in its coding sequence and is found to be uniquely down-regulated in the NIL-SLQ2. The genes differentially expressed in NIL-SLQ2 but located in the region of NIL-SLQ1.1 could be downstream targets of NIL-SLQ2. Interestingly, three of the genes located within the SLQ1.2-region are also found to be differentially expressed in the respective NIL. The same occurs for three genes located within the region of NIL-SLQ2. Those six genes, marked in bold in table 2, would be possible candidate genes for SLQ1.2 or SLQ2, respectively.

The three genes commonly up-regulated in all compared samples, as shown in the commonly shared overlap of figure 2a, could be possible downstream targets. The same would occur for the 12 commonly down-regulated genes, which are shown in the overlap of figure 2b. Table 3 gives and overview over the exact location and annotation of these genes.

Table 3. Annotation and position of up- or down-regulated genes commonly shared between CC, RIL114 and L94 NILs, and not located in the QTL regions

Summary of genes commonly up- or down-regulated between Cebada Capa (CC), the recombinant inbred line 114 (RIL114) and L94 near isogenic lines (NILs), harboring CC introgressions in Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.2 and SLQ1.3) or chromosome 2 (SLQ2). Gene annotations and positions in centimorgans (cM) are according to the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The seven barley chromosomes are abbreviated as 1H to 7H. The presence or absence of single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in coding sequences is indicated. Up-regulated genes are highlighted by light grey shading.

| Locus	Annotation	Position	SNPs	Up-/down-regulated
MLOC_77603.1	Prenylcysteine methylesterase	6H, 55.03 cM	no	up
AK251611.1	ABC transporter family protein	7H, 70.61 cM	no	up
AK368005	G-box binding factor 4	3H, 89.13 cM	no	up
MLOC_79997.1	Unknown protein	6H, 54.75 cM	no	down
MLOC_69135.1	Domain of unknown function (DUF966)	3H, 69.90 cM	no	down
MLOC_79373.2	Unknown protein	2H, 59.81 cM	no	down
MLOC_17361.1	Unknown protein	5H, 112.64 cM	no	down
MLOC_35603.1	Retrotransposon protein	No position	no	down
MLOC_46943.1	Unknown protein	No position	no	down
MLOC_9292.1	RNA-directed DNA polymerase (reverse transcriptase)-related family protein	3H, 0.07 cM	no	down
MLOC_32292.1	Unknown protein	4H, 56.41 cM	no	down
MLOC_35813.1	Unknown protein	No position	no	down
MLOC_52762.1	Indeterminate(ID)-domain 14	4H, 59.49 cM	no	down
MLOC_73258.1	Unknown protein	3H, 57.61 cM	no	down

As mentioned before, a second overlap was performed for the differentially expressed genes in the NIL-SLQ1.1, the NIL-SLQ1.1(2014) and Cebada Capa. Figure 4 summarizes the overlaps of up- or down-regulated genes in these samples in the form of Venn diagrams. It can be observed, that the majority of the differentially genes are down-regulated and that a large number of genes is shared between the three compared samples. Interestingly, a relative large number of genes is uniquely up- or down-regulated in the NIL-SLQ1.1(2014). The number of these specific NIL-SLQ1.1(2014) genes is larger than the one observed for RIL114, as shown in figure 2. As shown in chapter 2, NIL-SLQ1.1(2014) harbors large amounts of CC introgressions of which the majority is not found in the RIL114. The larger amount of differentially expressed genes could relate to these additional CC introgressions. Additionally, some genes might be affected in their expression by unique combinations of L94 and CC alleles.
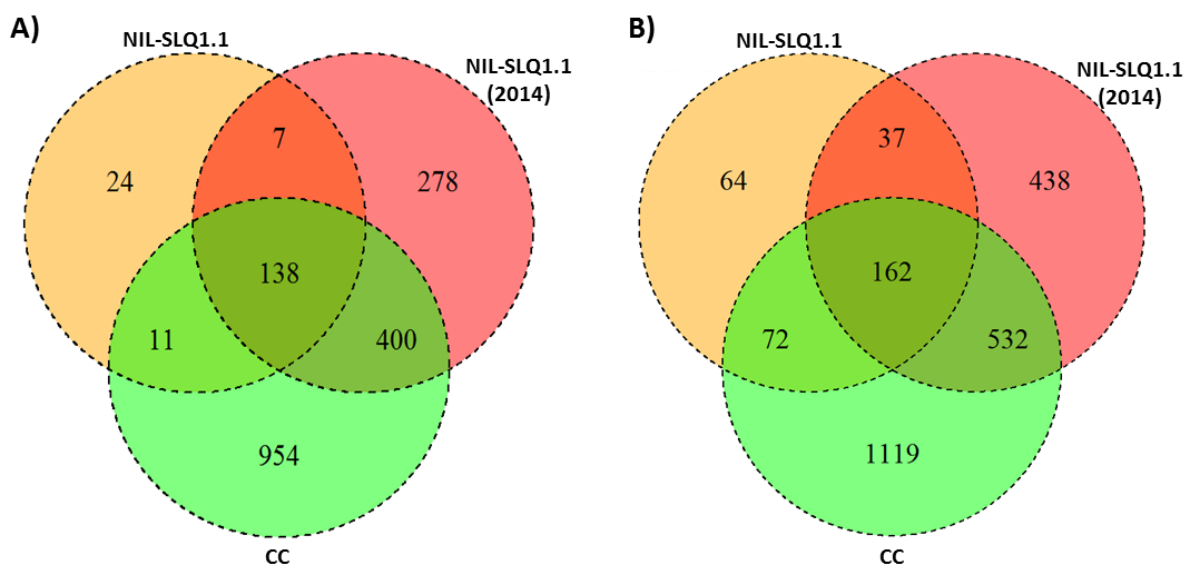
Figure 4. **Venn diagrams showing overlaps of up- or down-regulated genes in RNA-seq data sets of Cebada Capa, NIL-SLQ1.1 and NIL-SLQ1.1(2014) in comparison to L94**

Differentially regulated genes were called with a false discovery rate of 0.05. Up-regulated genes have a $\log_2$-fold change ($\log_2$FC) above 1 and down-regulated genes a $\log_2$FC below -1 in comparison to L94. Overlaps were performed between RNA-seq data sets from Cebada Capa (CC), a near isogenic line (NIL) showing CC introgressions in a Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1), and its precursor NIL sequenced in 2014. A) Overlaps between up-regulated genes. B) Overlaps between down-regulated genes.

As performed for the first overlaps, shown in figure 2, the location and the presence of SNPs in the coding sequence of the in the overlaps present genes were again schematically visualized and are shown in figure 5. Interestingly, SNPs are found within genes of all three overlaps and within genes uniquely up- or down-regulated in the NIL-SLQ1.1. In the down-regulated genes, a number of genes cluster in the regions of the SLQ1.1, SLQ1.2 and SLQ2. These genes could be possible candidate genes. Possible downstream targets could be found in the clusters formed by the up-regulated genes on chromosomes 5, 6 and 7. In the case of the down-regulated genes, cluster of possible downstream targets can be found on chromosomes 2 to 7.
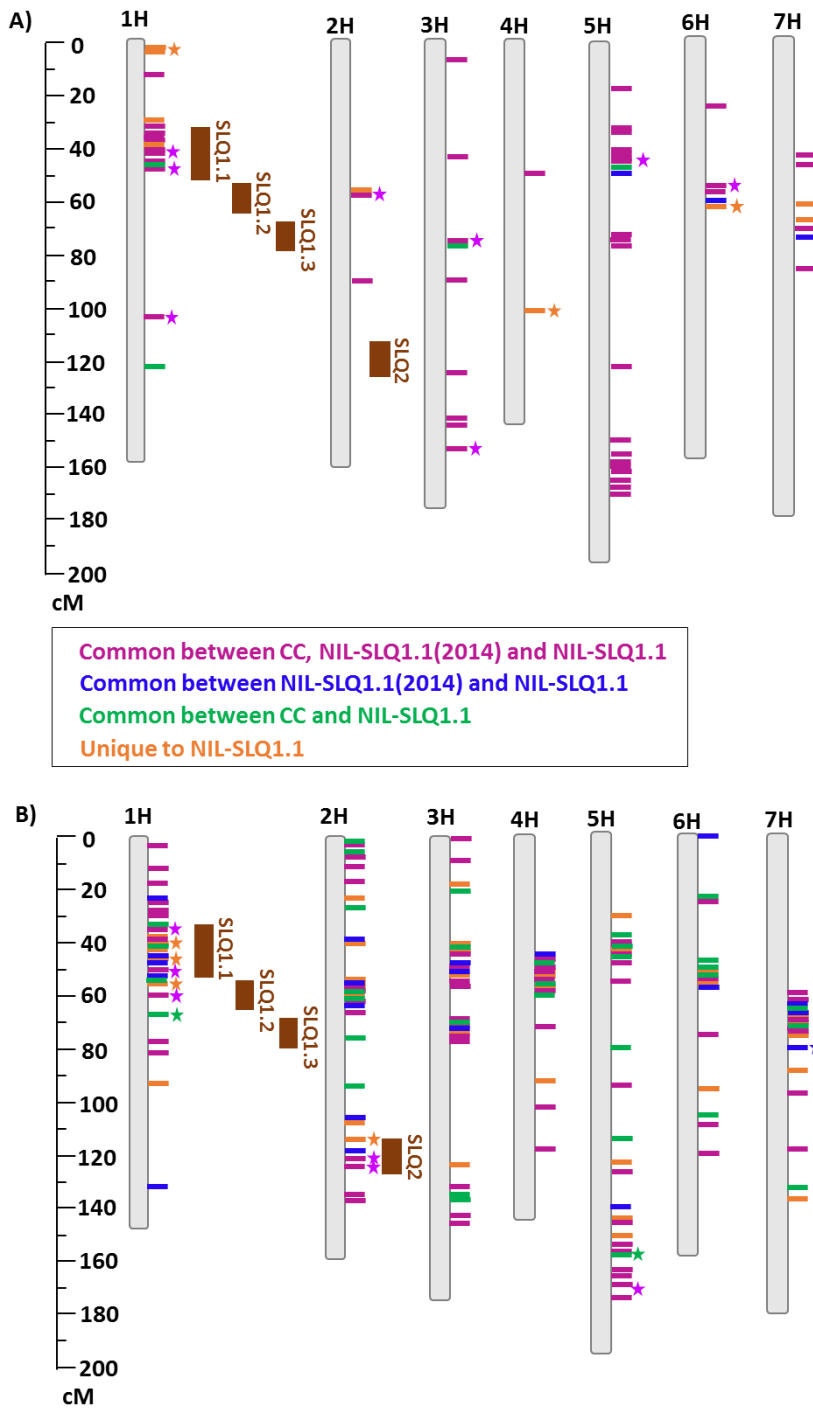
Figure 5. Schematic representation of the location of up- or down-regulated genes present in the overlaps of the Venn diagrams in figure 4.

The seven barley chromosomes are represented as grey bars and abbreviated as 1H to 7H. Genes are represented as colored stripes and the color code, explaining which genes are shared between which RNA-data sets, is shown in the figure itself. Stars indicate genes with single nucleotide polymorphisms (SNPs) between the coding sequences of Cebada Capa and L94. Sample abbreviations are as follows: Cebada Capa (CC), recombinant inbred line 114 (RIL14) and a near isogenic line (NIL) showing CC introgressions in a Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1), and its precursor NIL sequenced in 2014. SLQs regions are indicated as vertical brown bars. Gene positions, derived from the barley gene set (The International Barley Sequencing Consortium, 2012), are given in centimorgans (cM). A) Location of up-regulated genes present in the overlaps of figure 4a, B) Location of down-regulated genes present in the overlaps of figure 4b.

A number of differentially regulated genes are located in the region of SLQ1.1 and form clusters in this region, both in the up- and down-regulated genes. In the up-regulated genes, three of the genes shared between NIL-SLQ1.1, NIL-SLQ1.1(2014) and Cebada Capa have SNPs in their coding sequences. In the case of the down-regulated genes, two genes, shared between the same genotypes, show SNPs. Interestingly, some of the uniquely down-regulated genes in NIL-SLQ1.1 show also SNPs. However, SNPs in the promoters of the rest of the genes cannot be excluded, and therefore all genes located within the region of SLQ1.1 need to be considered possible candidate genes. Further information about possible targets genes, either up- or down-regulated, is summarized in tables 4 and 5.

Table 4. Annotation and position of up-regulated genes present in the overlaps of the Venn diagrams in figure 4 and located within the seed longevity QTL 1.1

Overlaps were performed between RNA-seq data sets from Cebada Capa (CC), a near isogenic line (NIL) showing CC introgressions in a Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1), and its precursor NIL sequenced in 2014. Differentially regulated genes were called with a false discovery rate of 0.05. Up-regulated genes have a $\log_2$-fold change ($\log_2$FC) above 1 and down-regulated genes a $\log_2$FC below -1 in comparison to L94. Gene annotations and positions in centimorgans (cM) are according to the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The presence or absence of single nucleotide polymorphisms (SNPs) in coding sequences between Cebada Capa and L94 is indicated. Genes showing SNPs in their coding sequences are marked in bold.

Present in	Locus	Annotation	Position	SNPs
CC, NIL-SLQ1.1, NIL-SLQ1.1(2014)	**MLOC_57318.2**	**Replication factor-A protein 1-related**	**1H, 48.09 cM**	**yes**
	MLOC_10766.1	Defensin	1H, 36.97 cM	no
	MLOC_48495.2	**CsAtPR5**	**1H, 47.18 cM**	**yes**
	MLOC_61611.1	**XH/XS domain-containing protein**	**1H, 47.95 cM**	**yes**
	MLOC_76988.1	Unknown protein	1H, 48.09 cM	no
	MLOC_58649.1	Serine protease inhibitor, potato inhibitor I-type family protein	1H, 38.03 cM	no
	AK362641	Unknown protein	1H, 48.23 cM	no
	AK252205.1	Unknown protein	1H, 48.23 cM	no
	AK375986	Unknown protein	1H, 48.51 cM	no
	MLOC_19024.1	Unknown protein	1H, 35.69 cM	no
	MLOC_39216.1	Actin binding protein family	1H, 48.09 cM	no
	MLOC_47378.1	Unknown protein	1H, 47.73 cM	no
	MLOC_47866.1	Unknown protein	1H, 48.09 cM	no
	MLOC_60967.1	Unknown protein	1H, 35.69 cM	no
	AK250805.1	Unknown protein	1H, 47.95 cM	no
CC, NIL-SLQ1.1	MLOC_54916.3	Alpha/beta-Hydrolases superfamily protein	1H, 48.09 cM	no
NIL-SLQ1.1	MLOC_45395.1	Serine protease inhibitor, potato inhibitor I-type family protein	1H, 38.03 cM	no

Table 5. Annotation and position of down-regulated genes present in the overlaps of the Venn diagrams in figure 4 and located within the seed longevity QTL 1.1

Overlaps were performed between RNA-seq data sets from Cebada Capa (CC), a near isogenic line (NIL) showing CC introgressions in a Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1), and its precursor NIL sequenced in 2014. Differentially regulated genes were called with a false discovery rate of 0.05. Up-regulated genes have a $\log_2$-fold change ($\log_2$FC) above 1 and down-regulated genes a $\log_2$FC below -1 in comparison to L94. Gene annotations and positions in centimorgans (cM) are according to the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The presence or absence of single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in coding sequences is indicated. Genes showing SNPs in their coding sequences are marked in bold.

Present in	Locus	Annotation	Position	SNPs
CC, NIL-SLQ1.1, NIL-SLQ1.1(2014)	MLOC_12107.3	Xylulose kinase-1	1H, 38.09 cM	no
	MLOC_61374.1	**Unknown protein**	**1H, 41.50 cM**	**yes**
	MLOC_64918.1	**NOD26-like intrinsic protein 5**	**1H, 48.58 cM**	**yes**
	AK356868	NDR1/HIN1-like 25	1H, 48.09 cM	no
	MLOC_1467.1	Unknown protein	1H, 48.76 cM	no
	MLOC_15910.1	Unknown protein	1H, 48.30 cM	no
	MLOC_77492.1	Unknown protein	1H, 37.61 cM	no
CC, NIL-SLQ1.1	MLOC_2662.1	Salt-inducible zinc finger 1	1H, 43.02 cM	no
NIL-SLQ1.1, NIL-SLQ1.1(2014)	MLOC_69613.1	Plastid-lipid associated protein PAP / fibrillin family protein	1H, 47.52 cM	no
	MLOC_71035.6	RECQ helicase SIM	1H, 48.09 cM	no
NIL-SLQ1.1	**MLOC_44161.3**	**Shikimate kinase like 2**	**1H, 47.34 cM**	**yes**
	MLOC_66307.1	**Major facilitator superfamily protein**	**1H, 41.93 cM**	**yes**
	AK356025	**HOPW1-1-interacting 1**	**1H, 41.93 cM**	**yes**
	MLOC_57601.1	Alpha/beta-Hydrolases superfamily protein	1H, 37.08 cM	no

Total seed proteome analysis for L94, Cebada Capa, RIL114 and the L94 NILs

For the analysis, mature, non-aged seeds were ground and protein extracts were analyzed by liquid-chromatography tandem-mass spectrometry (LC-MS/MS). Four replicates per genotype were processed. MS/MS spectra were searched against the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). Peptide spectrum matches and proteins were retained, if they were below a false discovery rate of 1%. Protein hits were only retained, if they were quantified in at least three of the four replicates in any line. In total, 984 proteins were differentially accumulated between L94 and Cebada Capa. Out of these 984 proteins, 654 were present in both parental lines. Of the remaining proteins, 228 proteins were much more abundant in L94 and almost totally absent in Cebada Capa, and 102 proteins were much more abundant in Cebada Capa and almost totally absent in L94. Such proteins were termed L94- or Cebada Capa-specific, respectively.

Gene ontology classification for L94- and Cebada Capa-specific proteins

In order to classify the functions of L94- and CC-specific proteins, a Gene ontology (GO) analysis was performed. GO terms were obtained from the PGSB barley genome database and classified using the CateGOrizer tool (Hu et al., 2008). The frequency of GO terms was analyzed using a GO slim assignment. Proteins were classified in several sub-categories for biological process, cellular component and molecular function. Histogram presentations of proteins belonging to these GO sub-categories are shown in figures 6a and 6b. For the category cellular component, shown in figure 6a, proteins unique to Cebada Capa are classified into less GO terms than proteins unique to L94. Three GO terms are common between L94 and Cebada Capa: membrane, ribosome and nucleus. ˈRibosomeˈ is the most represented GO term for CC-specific proteins. Three GO terms, associated with photosystem complexes I and II and an integral component of membrane, are unique to CC proteins. Seven GO terms are specific for L94 proteins and comprise the highest abundant term ˈintracellularˈ.
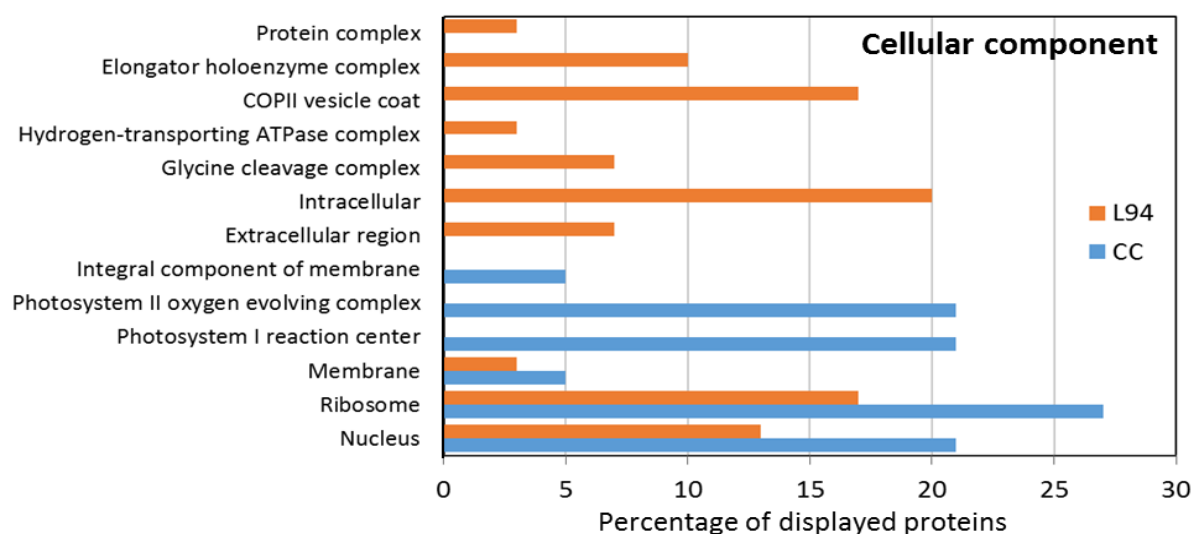
Figure 6a. **Histogram presentation of the Gene ontology (GO) category ˈCellular componentˈ for L94- or Cebada Capa (CC)-specific proteins identified in a total seed proteome analysis of mature seeds by liquid chromatography tandem-mass spectrometry (LC-MS/MS).**

Figure 6b. Histogram presentation of the Gene ontology (GO) category ˈBiological processˈ and ˈMolecular functionˈ for L94- or Cebada Capa (CC)-specific proteins identified in a total seed proteome analysis of mature seeds by liquid chromatography tandem-mass spectrometry (LC-MS/MS).

Comparable to the GO category ˈcellular functionˈ, CC-specific proteins are classified into less GO terms than L94-specific proteins in the categories ˈBiological processˈ and ˈMolecular functionˈ. As shown in figure 6b, five biological processes are shared between CC- and L94-specific proteins. The most abundant GO term for both parental lines in the category ˈBiological processˈ is found in those five common terms. ˈRegulation of transcriptionˈ and ˈMetabolismˈ are the most abundant GO terms for L94 and Cebada Capa, respectively. Only the GO term ˈPhotosynthesis, light harvestingˈ is unique to CC-specific proteins, while L94-specific proteins comprise 12 unique GO terms. Interestingly, the total number of GO terms present in the category ˈMolecular functionˈ is almost three times the number of GO terms in the other two categories. Around one quarter of these GO terms is shared between L94- and CC-specific proteins. Around one fifth is unique to proteins only present in Cebada Capa samples. Almost 60% of the GO terms are unique to L94-specific proteins. The most abundant GO term, ˈHydrolase activityˈ, is shared between L94- and CC-specific proteins.

Cluster analysis of L94- and Cebada Capa-specific proteins in the L94 NILs

For the 228 L94-specific and the 102 CC-specific proteins, a clustering of protein abundance differences was performed in the L94 NILs. The in the LC-MS/MS data analysis obtained label-free quantification (LFQ) intensities were $\log_2$ transformed. $\log_2$ ratios between the samples and L94 were calculated by replacing missing LFQ intensity values with zero. Relative protein abundances are represented as $\log_2$-fold values in the heat map shown in figure 7. In the following, proteins much more abundant in one parent and not per se fully absent in the other parent, are called L94- or CC-specific, respectively. The dendrogram on the y-axis clearly separates CC-specific proteins (increased protein abundance in CC shown in green) and L94-specific proteins (decreased protein abundance in CC shown in red). The dendrogram on the x-axis separates the samples according to their amount of CC introgressions. Cebada Capa is clearly separated from the rest of the analyzed genotypes. NIL-SLQ1.1(2014) and RIL114 are found in close distance to each other since they harbor more CC introgressions than the rest of the samples. Comparably, NIL-SLQ2(2014) and NIL-SLQ1.2+3(2014), used to generate both NIL-SLQ1.2 and NIL-SLQ1.3, also cluster together. While the NILs selected for the seed longevity QTLs on chromosome 1 cluster together, the NIL-SLQ2 is clearly separated. Comparing the protein abundances between Cebada Capa and the rest of the analyzed genotypes, the majority of the CC-specific proteins is also highly accumulated in part of the NILs and the RIL114. In contrast to that, only around one quarter of the L94-specific proteins is lowly accumulated in some of the NILs and/or RIL114. The proteins being either low- or high-accumulated in both CC and the NILs/RIL114 could be proteins affecting the phenotype differences in seed longevity between L94 and Cebada Capa. For this reason, proteins showing the same tendency in protein abundance as CC were further categorized.
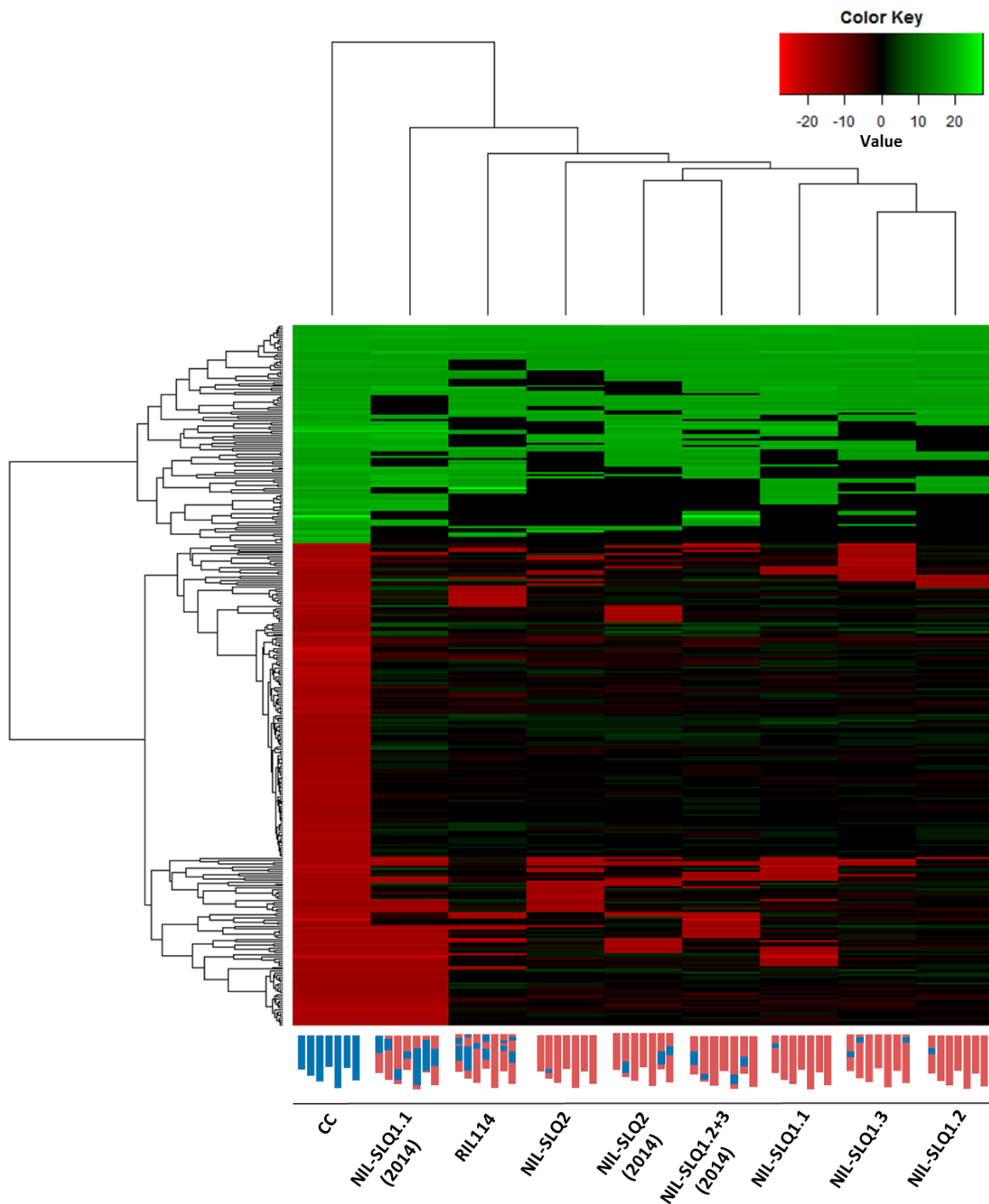
Figure 7. **Heat map representing quantitative differences in protein abundance between Cebada Capa, RIL114, L94 NILs and L94, identified in a total seed proteome analysis of mature seeds by LC-MS/MS**

Heat map depicts normalized $\log_2$-fold changes in protein abundance between genotypes, schematically presented below the heat map, and the parental line L94. Schematic representations of the analyzed genotypes show the seven barley chromosomes, the red and blue bars represent the genomic regions of L94 and introgression segments from Cebada Capa (CC), respectively. Near isogenic lines (NILs) show CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). The cleanest NILs, showing only CC introgressions in one seed longevity QTLs, were derived from NILs sequenced by RNA-seq in the year 2014. Those sequenced lines are indicated by carrying ꞌ2014ꞌ in brackets behind their name. The recombinant inbred line 114 (RIL114) was also included in the analysis. Decreased protein abundance (in $\log_2$-fold change) in comparison to L94 is presented in red and increased abundance in green. Hierarchical clustering analysis of $\log_2$-fold changes in protein abundance for L94- and Cebada Capa-specific proteins is presented as a dendrogram.

Identification of proteins associated with seed longevity

In order to identify those proteins possibly associated with seed longevity, an analysis of the overlaps of proteins, showing a similar level of protein abundance in both CC and the L94 NILs/RIL114, was performed. As shown in figure 7, the log_2-fold values for the CC-specific proteins (increased protein abundance in CC shown in green) in the L94 NILs and the RIL114 were either comparable to Cebada Capa or 0. Log_2-fold values of 0 occur due to missing LFQ intensity values for samples. Figure 8 shows overlaps of proteins highly abundant in both CC and the L94 NILs/RIL114. The mean log2-fold value for proteins highly abundant in the L94 NILs and RIL114 was 21.44 with a standard deviation of 1.28. Overlaps of highly abundant proteins were performed between the 'cleanest' NILs (NIL-SLQ1.1, NIL-SLQ1.2, NIL-SLQ1.3 and NIL-SLQ2), their respective precursor NIL sequenced in 2014, the RIL114 and CC. 25 up to 37 common proteins were shared between the four compared genotypes.
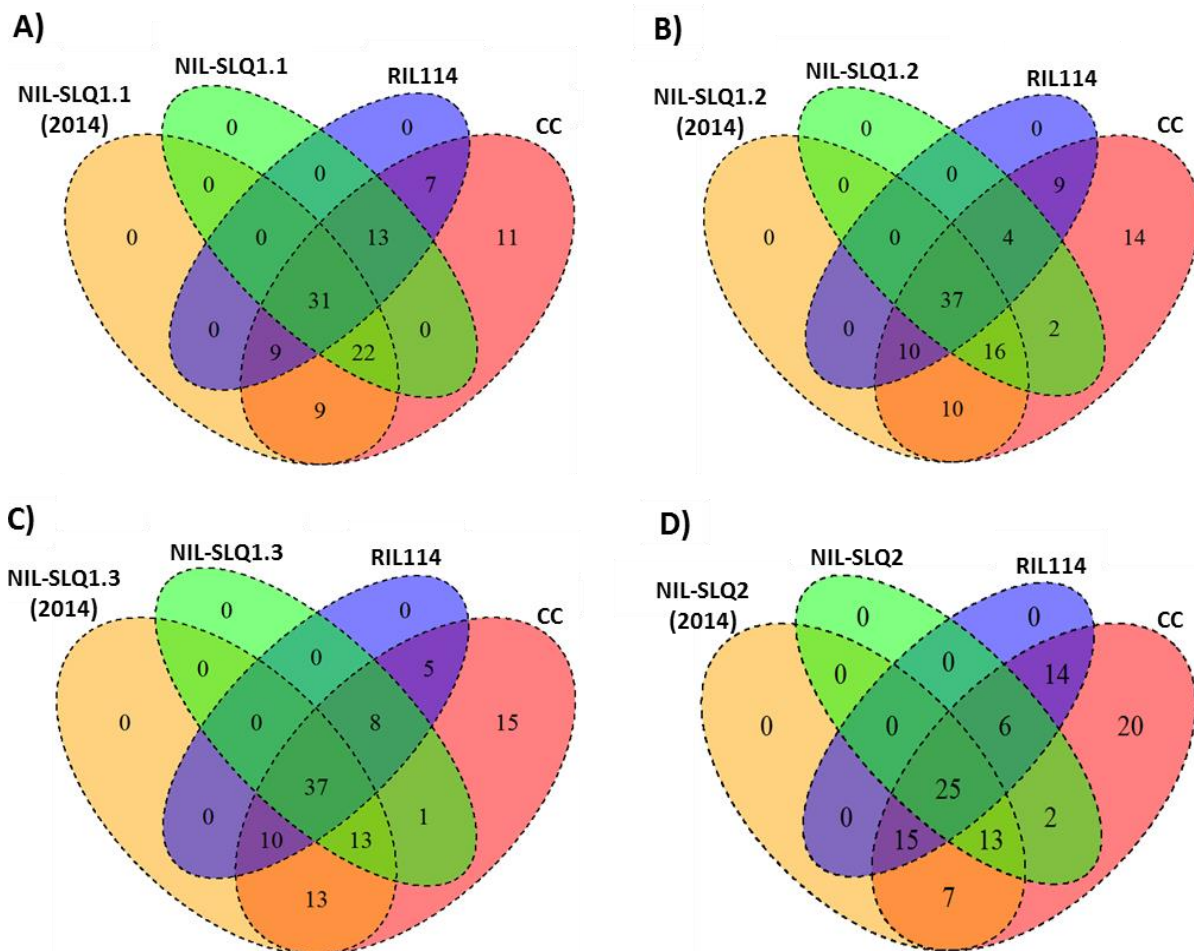
Figure 8. **Venn diagrams showing overlaps of proteins highly abundant in Cebada Capa, RIL114 and the L94 NILs**
Overlaps were performed between near isogenic lines (NILs) harboring Cebada Capa (CC) introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2), their precursor NILs sequenced by RNA-seq in the year 2014, the recombinant inbred line 114 (RIL114) and CC. A) Overlap of proteins highly abundant in NIL-SLQ1.1, NIL-SLQ1.1(2014), RIL114 and CC. B) Overlap of proteins highly abundant in NIL-SLQ1.2, NIL-SLQ1.2(2014), RIL114 and CC. C) Overlap of proteins highly abundant in NIL-SLQ1.3, NIL-SLQ1.3(2014), RIL114 and CC. D) Overlap of proteins highly abundant in NIL-SLQ2, NIL-SLQ2(2014), RIL114 and CC.

Interestingly, 15 proteins were shared between the four performed overlaps (A to D), shown in figure 8. These proteins differ between L94 and the NILs irrespective of their introgression. Since all NILs have an increased longevity, these proteins might be potential common downstream proteins of the different QTLs. Annotations and positions of the genes, encoding the proteins present in the overlaps, are shown in tables 6a, 6b and 6c. Interestingly, two of these genes are found in the region of the SLQ1.1 and two in the region of the SLQ2. These genes are marked in bold in the following three tables.

Table 6a. Annotation and position of genes corresponding to proteins present in overlaps of three Venn diagrams, shown figure 8.

Gene annotations and positions are according to the to PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The seven barley chromosomes are abbreviated as 1H to 7H. Gene positions are given in centimorgans (cM). The abbreviation HC or LC indicate whether the genes form part of the high-confidence (HC) or low-confidence (LC) gene set (The International Barley Sequencing Consortium, 2012). Genes located within the Quantitative Trait Loci (QTL) regions for seed longevity (SLQ) are marked in bold. The presence or absence of single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in coding sequences is indicated.

| Present in	Locus	Annotation	LC/HC set	Position	SNPs
Common in overlaps of A), B), C) and D)	AK250508.1	Germin-like protein 2	HC	No position	yes
	MLOC_72819.2	SNF1 kinase homolog 10	LC	No position	no
	AK373766	BZIP transcription factor	HC	No position	no
	MLOC_65838.1	GRAS family transcription factor	HC	2H, 55.38 cM	yes
	AK374142	DNA-binding storekeeper protein-related transcriptional regulator	HC	7H, 61.47 cM	yes
	AK372353	Transducin/WD40 repeat-like superfamily protein	LC	No position	no
	MLOC_6164.1	Golgin-84	HC	3H, 62.53 cM	yes
	AK375219	2Fe-2S ferredoxin	HC	4H, 57.93 cM	yes
	MLOC_8487.1	Unknown protein	HC	6H, 54.89 cM	yes
	MLOC_56173.1	Syntaxin-61	HC	4H, 51.27 cM	no
	AK251563.1	O-fucosyltransferase family protein	HC	2H, 55.59 cM	yes
	MLOC_65114.2	FASCICLIN-like arabinogalactan-protein	HC	1H, 118.48 cM	no
	MLOC_78630.1	Oxygen-evolving enhancer protein 1	HC	2H, 58.92 cM	no
	MLOC_59602.1	E3 ubiquitin-protein ligase PRT1	HC	6H, 55.02 cM	yes
	MLOC_62389.1	Iron-sulfur cluster insertion protein erpA	HC	5H, 44.93 cM	no
Common in overlaps of A), B) and C)	MLOC_34272.1	Ribulose bisphosphate carboxylase large chain	HC	No position	no
	MLOC_54528.1	Oxygen-evolving enhancer protein 2	HC	2H, 58.64 cM	yes
	MLOC_54108.2	CwfJ domain protein	HC	5H, 48.33 cM	yes
	MLOC_63907.1	Unknown protein	HC	7H, 27.47 cM	no
	AK370622	Cytochrome b561 IPR004877	HC	2H, 76.77 cM	yes
	MLOC_4936.1	50S ribosomal protein L12-2	HC	4H, 43.62 cM	yes
	MLOC_57675.1	**Unknown protein**	**LC**	**2H, 130.31 cM**	**no**

Table 6b. Annotation and position of genes corresponding to proteins present in overlaps of two or three Venn diagrams, shown in figure 8.

Gene annotations and positions are according to the to PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The seven barley chromosomes are abbreviated as 1H to 7H. Gene positions are given in centimorgans (cM). The abbreviation HC or LC indicate whether the genes form part of the high-confidence (HC) or low-confidence (LC) gene set (The International Barley Sequencing Consortium, 2012). Genes located within the Quantitative Trait Loci (QTL) regions for seed longevity (SLQ) are marked in bold. The presence or absence of single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in coding sequences is indicated.

Present in	Locus	Annotation	LC/HC set	Position	SNPs
Common in overlaps of B) and C)	MLOC_52887.1	GDSL esterase/lipase	HC	4H, 1.13 cM	yes
	MLOC_35785.1	NADP-dependent malic enzyme	HC	3H, 46.10 cM	no
	AK250909.1	Pentatricopeptide repeat-containing protein	HC	2H, 73.51 cM	yes
	MLOC_58241.2	Low temperature viability protein	HC	5H, 74.65 cM	yes
	AK357992	Short-chain dehydrogenase/reductase SDR	HC	6H, 55.02 cM	no
	AK354854	Tetratricopeptide repeat protein	HC	No position	yes
Common in overlaps of B) and D)	**MLOC_58032.2**	**Kinase interacting (KIP1-like) family protein**	**HC**	**1H, 48.22 cM**	**no**
Common in overlaps of B), C) and D)	MLOC_38639.1	Unknown protein	HC	5H, 80.34 cM	no
	MLOC_73300.1	Heat shock transcription factor C1	HC	7H, 76.55 cM	no
	MLOC_62505.1	PLAC8 family protein	LC	4H, 110.69 cM	yes
	AK364395	SEC7-like guanine nucleotide exchange family protein	HC	No position	yes
	AK369281	Long chain fatty alcohol oxidase	HC	6H, 54.78 cM	no
	MLOC_39457.1	alpha/beta-Hydrolases superfamily protein	LC	4H, 0.63 cM	yes
Common in overlaps of A), B) and D)	MLOC_76470.1	Cysteine proteinase	HC	No position	no
Common in overlaps of A), C) and D)	MLOC_54015.2	Pre-rRNA-processing protein ESF1	HC	2H, 57.44 cM	yes

65

Table 6c. Annotation and position of genes corresponding to proteins only present in the overlaps of one the Venn diagrams, shown in figure 8.

Gene annotations and positions are according to the to PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). The seven barley chromosomes are abbreviated as 1H to 7H. Gene positions are given in centimorgans (cM). The abbreviation HC or LC indicate whether the genes form part of the high-confidence (HC) or low-confidence (LC) gene set (The International Barley Sequencing Consortium, 2012). Genes located within the Quantitative Trait Loci (QTL) regions for seed longevity (SLQ) are marked in bold. The presence or absence of single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in coding sequences is indicated.

Present in	Locus	Annotation	LC/HC set	Position	SNPs
Only in overlap of A)	AK366918	Stachyose synthase	HC	3H, 39.51 cM	yes
	MLOC_57061.1	Chlorophyll a-b binding protein	HC	4H, 51.55 cM	no
	MLOC_76988.1	**Unknown protein**	**HC**	**1H, 48.08 cM**	**no**
	MLOC_61558.1	Ribulose bisphosphate carboxylase large chain	HC	No position	no
	MLOC_11961.1	Chlorophyll a-b binding protein	HC	2H, 59.35 cM	no
	AK360329	Tubulin beta-1 chain	HC	No position	no
	MLOC_6256.2	Unknown protein	HC	7H, 67.92 cM	no
Only in overlap of C)	MLOC_63731.1	Growth factor	HC	4H, 21.03 cM	no
	AK376359	Cupin family protein	HC	6H, 68.20 cM	yes
Only in overlap of D)	**MLOC_11661.1**	**UDP-glycosyltransferase**	**HC**	**2H, 132.58 cM**	**yes**

Out of the four genes, located within the regions of the seed longevity QTLs SLQ1.1 and SLQ2, two of the corresponding proteins are unique to the overlaps performed for the NIL having an introgression in this specific SLQ. The UDP-glycosyltransferase (MLOC_11661.1) is only found in the overlap of figure 8d, which shows the proteins shared between the NIL-SLQ2, NIL-SLQ2(2014), the RIL114 and CC. Comparably, a protein with unknown function (MLOC_76988.1) is only found in the overlap of figure 8a. Due to their location within the SLQ regions, these two genes are considered possible candidates genes.

Proteins possibly associated with seed longevity, could also be proteins both absent in CC and the L94 NILs/RIL114, but present in the short-lived parental line L94. Such proteins would then be L94-specific. Figure 9, presents the results of the overlaps of proteins present in L94 but absent in CC, the L94 NILs and the RIL114 as Venn diagrams. As shown in figure 9, only one protein is found in the overlaps A and C, respectively. The protein found in overlap A is a glutamate formimino transferase and the protein of overlap C belongs to the protein transducin family. The corresponding genes belong both to the high-confidence (HC) gene set, however, no position is available for them in the PGSB barley genome database.
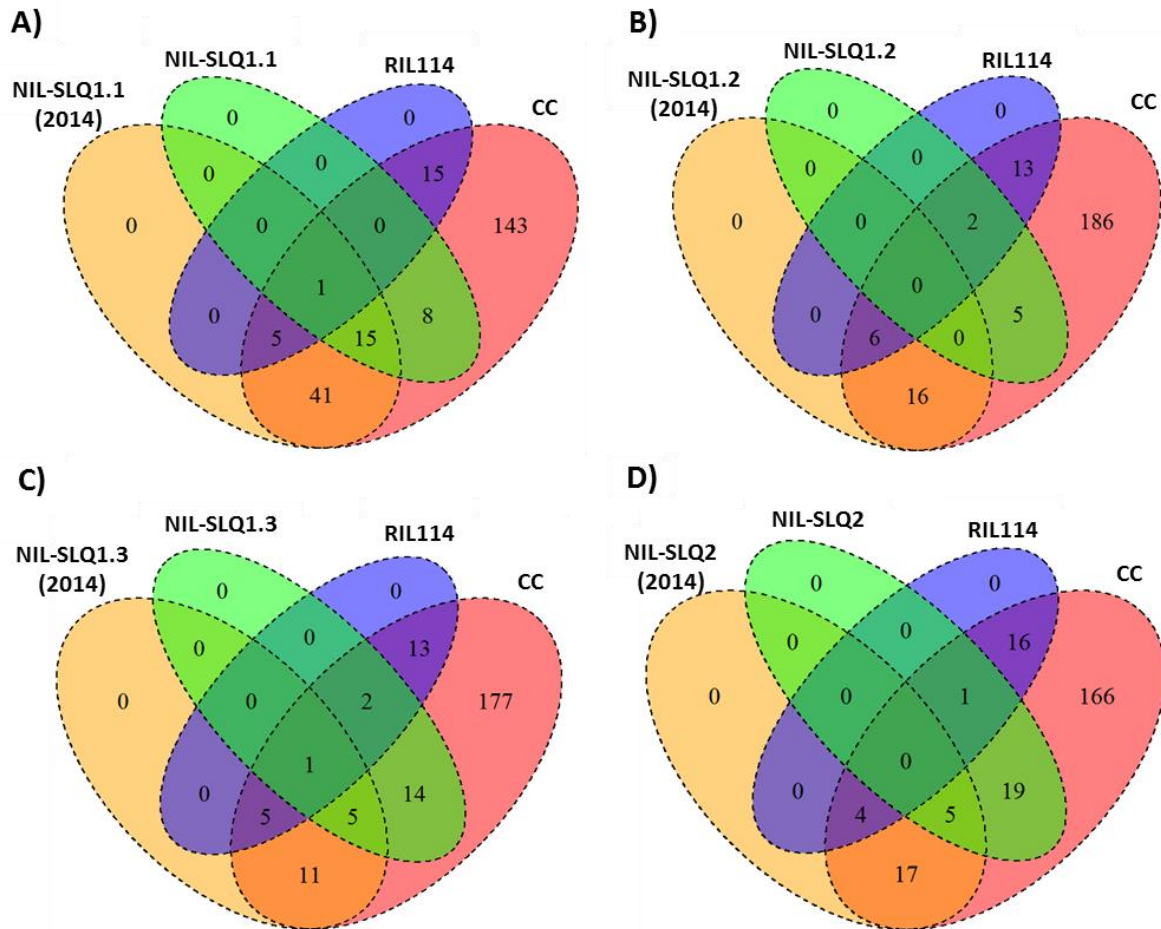
Figure 9. **Venn diagrams showing overlaps of proteins absent in Cebada Capa, RIL114 and the L94 NILs, but present in L94**

Overlaps were performed between near isogenic lines (NILs) harboring Cebada Capa (CC) introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2), their precursor NILs sequenced by RNA-seq in the year 2014, the recombinant inbred line 114 (RIL114) and CC. A) Overlap of proteins absent in NIL-SLQ1.1, NIL-SLQ1.1(2014), RIL114 and CC. B) Overlap of proteins absent in NIL-SLQ1.2, NIL-SLQ1.2(2014), RIL114 and CC. C) Overlap of proteins absent in NIL-SLQ1.3, NIL-SLQ1.3(2014), RIL114 and CC. D) Overlap of proteins absent in NIL-SLQ2, NIL-SLQ2(2014), RIL114 and CC.

Selection of potential candidate and downstream target genes for further analyses

To identify potential candidate and downstream target genes for further analyses, the results of the RNA-seq and total seed proteome analysis were combined. As shown before, the majority of the proteins possibly associated with seed longevity are CC-specific. As shown in tables 6a-c, two of the underlying genes are located within the region of SLQ1.1 and two are located within the region of SLQ2. Due to their location, these genes were considered as possible candidate genes. Figure 10 depicts the relative protein abundances of the four encoded proteins in the form of a heat map. Proteins whose underlying genes are located within the region of SLQ1.1 and SLQ2 are colored in green and blue, respectively. The KIP1-like family protein MLOC_58032.2 and MLOC_76988.1, a protein of unknown function, are possibly associated with SLQ1.1. As shown in figure 10, MLOC_58032.2 is highly accumulated in all presented genotypes apart from NIL-SLQ1.2 and NIL-SLQ1.1. In the RNA-seq analysis, the underlying gene was neither found to be differentially expressed between the

parental lines nor any L94 NIL and L94. In contrast to MLOC_58032.2, MLOC_76988.1 is both highly accumulated in NIL-SLQ1.1 and NIL-SLQ1.1(2014). Additionally, the protein in highly accumulated in Cebada Capa, RIL114 and NIL-SLQ2(2014) in comparison to L94. Comparable to MLOC_58032.2, MLOC_76988.1 was not found to be differentially expressed in samples analyzed by RNA-seq. The UDP-glycosyltransferase MLOC_11661.1 and the unknown protein MLOC_57675.1 were chosen to be possibly associated with SLQ2. As shown in figure 10, this UDP-glycosyltransferase is exclusively highly accumulated in NIL-SLQ2, NIL-SLQ2(2014), RIL114 and Cebada Capa. In contrast to that, the unknown protein MLOC_57675.1 is highly accumulated in all analyzed samples apart from NIL-SLQ2. Neither MLOC_11661.1 nor MLOC_57675.1 were found to be differentially expressed in the conducted RNA-seq analysis.
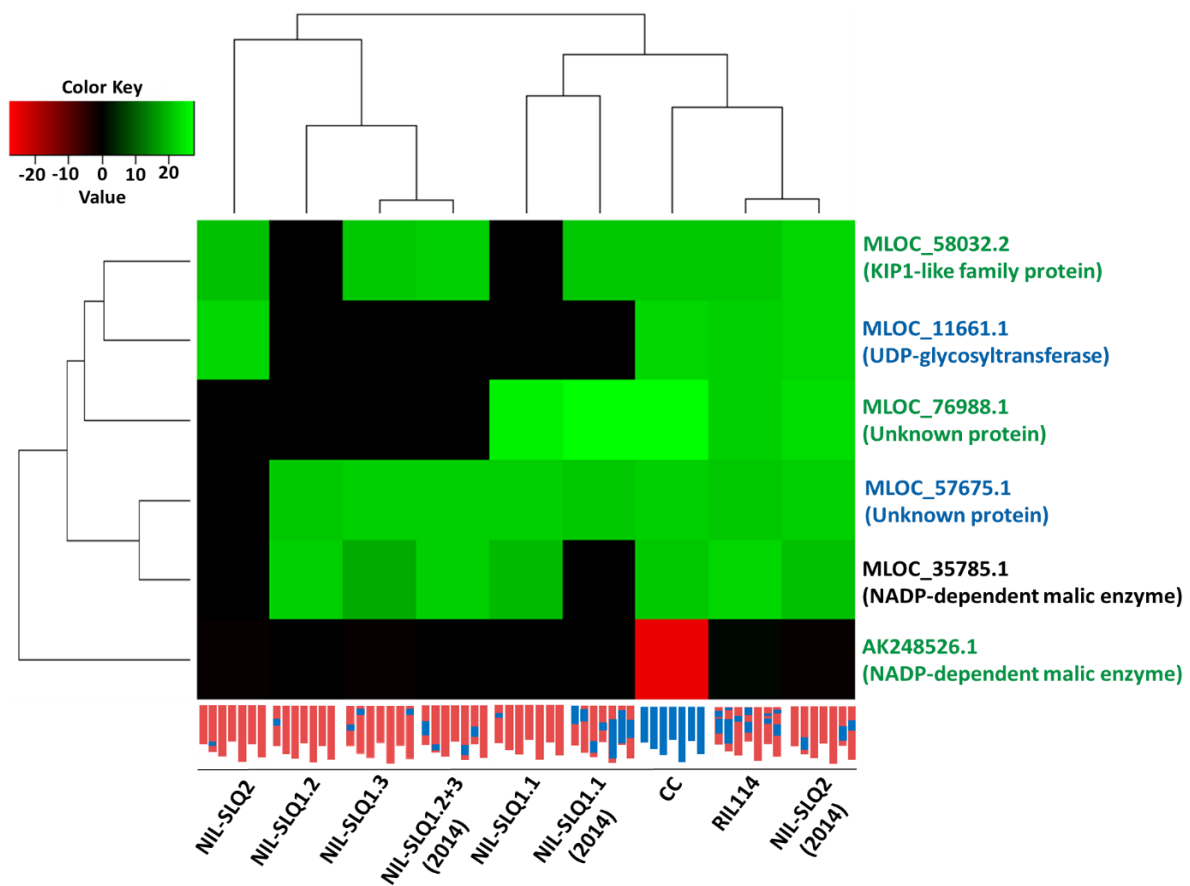
Figure 10. **Heat map of potential candidate genes and downstream targets, representing quantitative differences in protein abundance between Cebada Capa, RIL114, L94 NILs and L94, identified in a total seed proteome analysis of mature seeds by LC-MS/MS**

Heat map depicts normalized log$_2$-fold changes in protein abundance between genotypes, schematically presented below the heat map, and the parental line L94. Schematic representations of the analyzed genotypes show the seven barley chromosomes, the red and blue bars represent the genomic regions of L94 and introgression segments from Cebada Capa (CC), respectively. Near isogenic lines (NILs) show CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). The cleanest NILs, showing only CC introgressions in one seed longevity QTLs, were derived from NILs sequenced by RNA-seq in the year 2014. Those sequenced lines are indicated by carrying ˈ2014ˈ in brackets behind their name. The recombinant inbred line 114 (RIL114) was also included in the analysis. Decreased protein abundance (in log$_2$-fold change) in comparison to L94 is presented in red and increased abundance in green. Proteins associated with SLQ1.1 and SLQ2 are colored in green and blue, respectively. Possible downstream targets are colored in black.

Due to the fact that the UDP-glycosyltransferase MLOC_11661.1 is exclusively highly accumulated in NILs selected for SLQ2, RIL114 and Cebada Capa, the underlying gene was selected for further analyses. The selection was furthermore encouraged by the fact that SNPs were detected in the coding sequence of MLOC_11661.1 by RNA-seq. MLOC_58032.2 and MLOC_57675.1 were not chosen for further analyses, since the respective proteins were only highly accumulated in the NILs sequences in 2014, but not in the clean NILs. Although MLOC_76988.1 was both highly accumulated in NIL-SLQ1.1(2014) and NIL-SLQ1.1, MLOC_76988.1 was not chosen for further analysis due to its unknown function.

Instead, two NADP-dependent malic enzymes were chosen for further analyses due to recent findings in *Arabidopsis thaliana*. The NADP-dependent malic enzyme 1 (NADP-ME1), identified in a combined transcriptomics and proteomics approach, was shown to significantly affect seed longevity (Nguyen, 2014). The relative protein abundances of the two barley malic enzymes are also shown in figure 10. The NADP-dependent malic enzyme MLOC_35785.1 was selected due to its high accumulation in all samples apart from NIL-SLQ1.1(2014) and NIL-SLQ2. Since the underlying gene is located on chromosome 3, as shown in table 6b, MLOC_35785.1 was considered to be a potential downstream target. In the RNA-seq analysis, no differential expression between L94 and the analyzed samples was observed for the underlying gene. Interestingly, another NADP-dependent malic enzyme was identified in the total seed proteome analysis, which showed a contrasting accumulation pattern. As shown in figure 10, the second malic enzyme AK248526.1 is very low abundant in Cebada Capa and not differentially abundant between L94 and RIL114/L94 NILs. Additionally to its contrasting accumulation pattern, the underlying gene is located within the region of SLQ1.1 (Chromosome 1, 49 cM) and differentially expressed between the two parental lines (Log$_2$FC CC/L94: 17.07). The higher expression in Cebada Capa is thereby in contrast to the finding that the respective protein is lower accumulated in Cebada Capa. Due to its location, the malic enzyme AK248526.1 was chosen as possible candidate gene for SLQ1.1. Analyses to elucidate the role of the UDP-glycosyltransferase MLOC_11661.1 and the two malic enzymes are discussed in detail in chapter 4.

Metabolomics analysis of L94, Cebada Capa, 116-5, RIL114 and the cleanest L94 NILs

To gain insight into biochemical processes affected in seed longevity, gas chromatography mass spectrometry-based (GC-MS) metabolite profiling was employed to analyze mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and the cleanest L94 NILs. For each genotype four replicates of 20 seeds each were processed. Figure 11 shows a principal component analysis (PCA) score plot based on the relative metabolic composition of the analyzed samples. The PCA analysis separated all samples into two clusters along the first two axes that together account for around 38% of total variation. A high variation between the replicates of each sample can be observed. Replicates of Cebada Capa, 116-5 and NIL-SLQ1.1 show the highest variation among all analyzed samples. The second principal component seems to separate the samples according to their amount of Cebada Capa introgressions. Cebada Capa and 116-5, the line which harbors by far the most Cebada Capa introgressions, are clearly separated from the other samples. The rest of the samples, including RIL114, the L94 NILs and the parental line L94, cluster together with some separation by PC1. One of the four NIL-SLQ1.1 replicates is distant from this formed cluster.
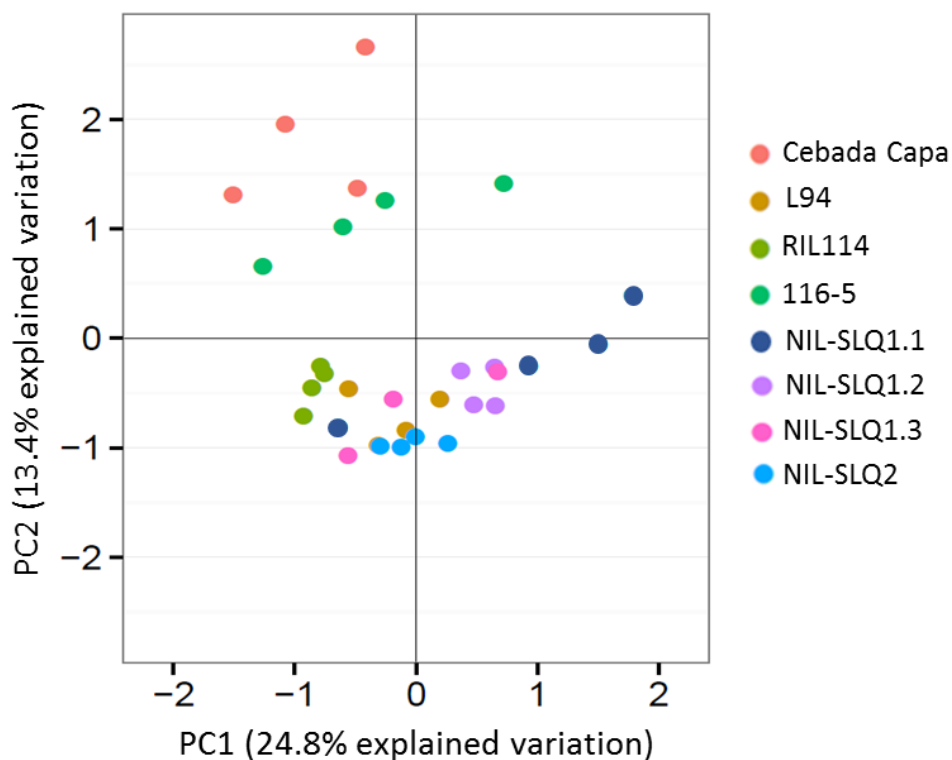
Figure 11. **Principal component analysis (PCA) score plot of GC-MS-based metabolite profiling of mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**

The two first principal components (PC1 and PC2) are plotted on the x- and y-axis, and percentages of the by the components explained variation is given in percentages. Four replicates of 20 mature, non-aged seeds were analyzed for each of the presented genotypes, employing gas chromatography mass spectrometry (GC-MS)-based metabolite profiling. The color code of the analyzed samples is shown. Near isogenic lines (NILs) harboring Cebada Capa (CC) introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2) were analyzed. The recombinant inbred line 114 (RIL114) and the line 116-5, resulting from a cross between L94 and Cebada Capa, were also included in the analysis.

In total, 84 metabolites were identified and quantified. Figure 12 shows a heat map of relative abundances of metabolites detected in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and the L94 NILs. As shown in figure 12, around half of the 84 metabolites are found to be similarly abundant in all samples. Only 17 metabolites are significantly differentially accumulated between L94 and Cebada Capa.
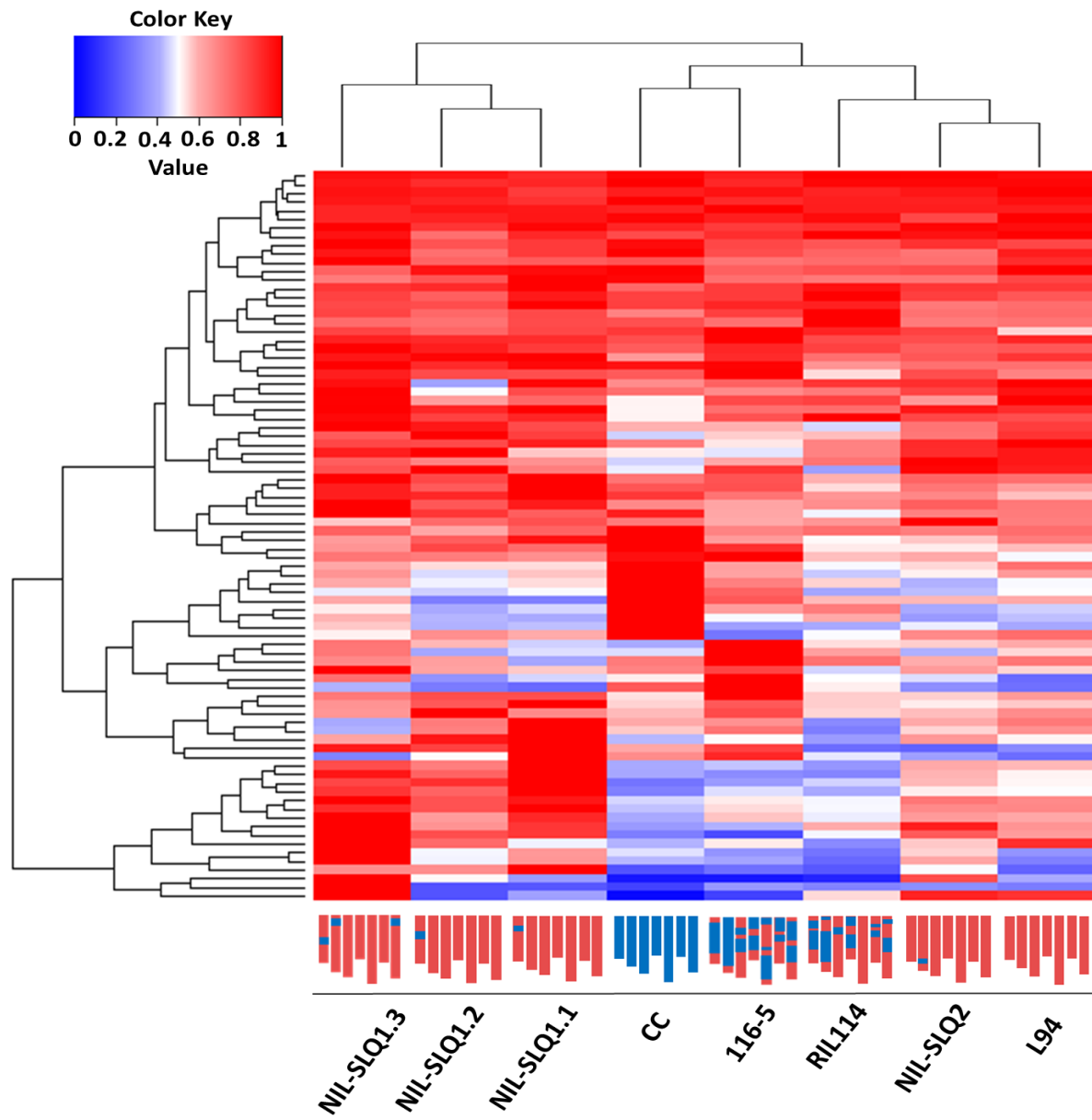
Figure 12. **Heat map of relative abundances of metabolites detected in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs**

Heat map depicts relative abundances, the sample with the highest abundance detected, was set to 1. Relative abundances were calculated by referring metabolite peak intensities of samples to an internal standard. Schematic representations of the analyzed genotypes show the seven barley chromosomes, the red and blue bars represent the genomic regions of L94 and introgression segments from Cebada Capa (CC), respectively. Near isogenic lines (NILs) show CC introgressions in different Quantitative Trait Loci (QTL) for seed longevity (SLQ) on chromosome 1 (SLQ1.1-1.3) or chromosome 2 (SLQ2). The recombinant inbred line 114 (RIL114) and the line 116-5 were also included in the analysis.

Figure 13 shows a pie chart representing the distribution of detected metabolite classes. The majority of the identified metabolites are amino acids, followed by oxalic acids and sugars. Only one flavonoid was detected.
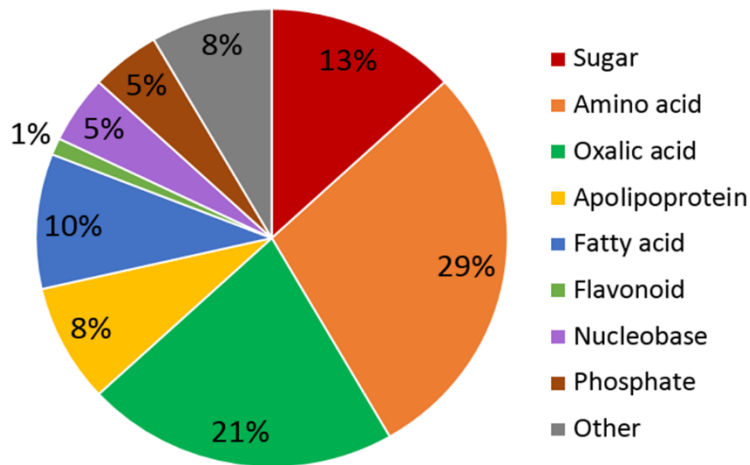
Figure 13. **Pie chart representing the distribution of metabolites detected in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**

The color code of the identified metabolite classes and their respective percentages are indicated.

Out of the 17 metabolites, significantly differentially accumulated between the two parental lines, 11 metabolites are significantly higher abundant in Cebada Capa than in L94. Interestingly, two of the Cebada Capa highly abundant metabolites are low abundant in the rest of the analyzed samples. The relative abundances of these two monosaccharides, arabinose and rhamnose, are shown in figure 14.
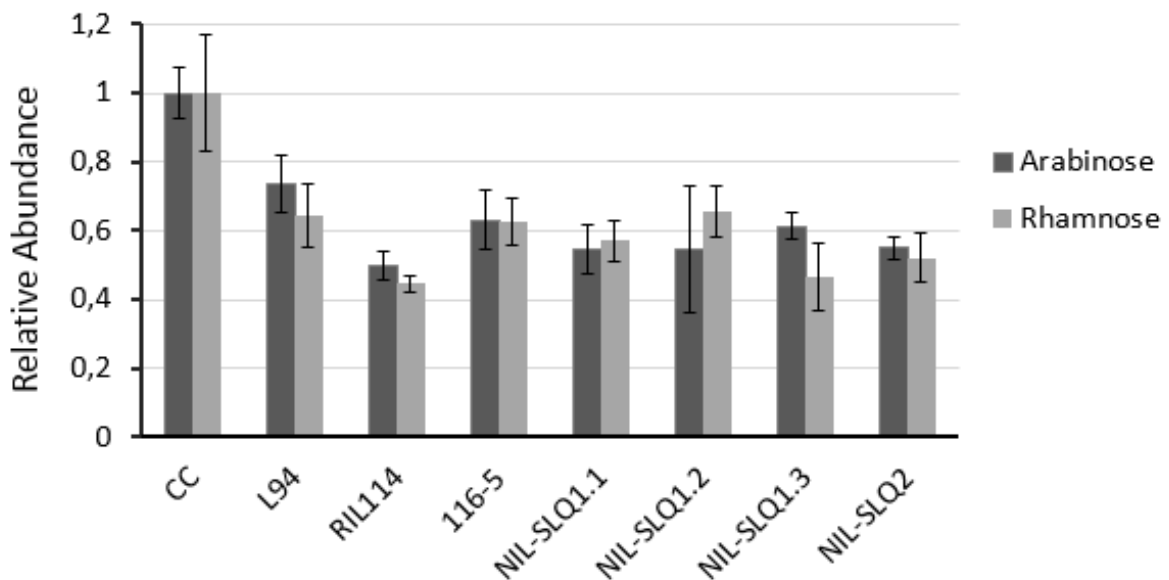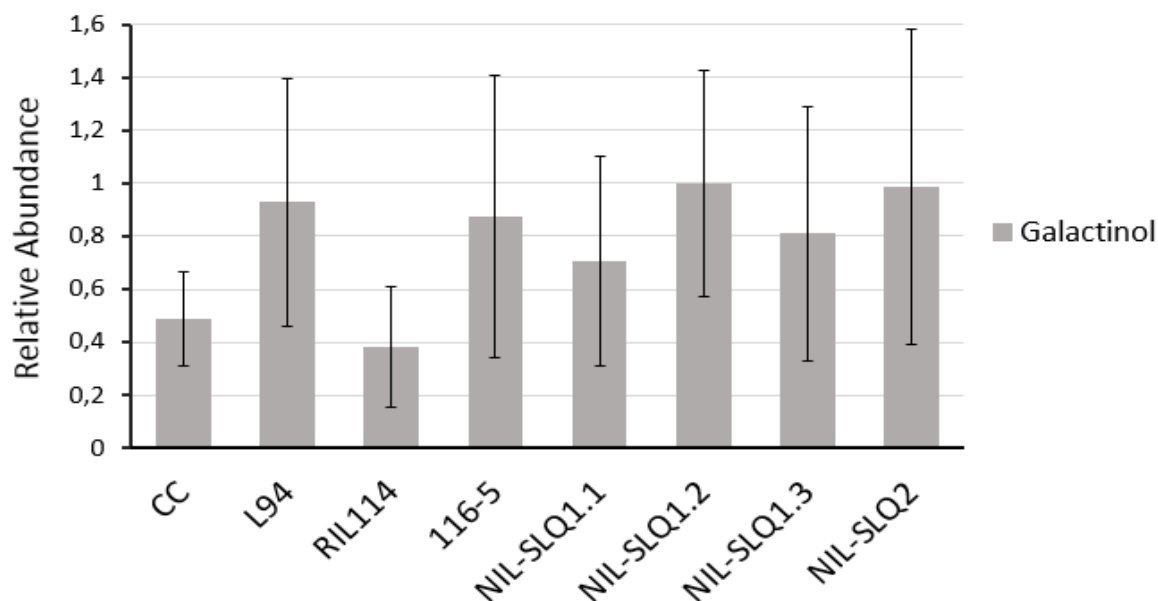

Figure 14. **Bar plot-visualization of arabinose and rhamnose relative abundances in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**

Relative abundances of arabinose (meox1 derivatization) and rhamnose (meox1 derivatization) were calculated by referring metabolite peak intensities of samples to an internal standard. Peak areas were determined using the Xcalibur software (Thermo Scientific). The value of the sample for which the highest abundance was detected, was set to 1. Standard deviations were calculated from four replicates per analyzed sample.

Seed coat mucilages are composed of basic sugar moieties such as arabinose, rhamnose, fructose and mannose (Caffall and Mohnen, 2009). Upon exposure to water, the expanding hydrophilic mucilage ruptures the primary wall and expands to encapsulate the seed, presumably for dispersal, protection, and/or hydration (Arsovski et al., 2010; Western, 2012). Interestingly, Debeaujon et al. (2000) showed that the three *Arabidopsis thaliana* mutants *transparent testa glabra1* (*ttg1*), *aberrant testa shape* (*ats*) and *glabra2* (*gl2*), lacking seed coat mucilage, have reduced seed longevity in comparison to wild-type seed lots after storage for four years at room temperature. Seed coat mucilage has been reported to be produced by seeds or fruits of species in 37 orders, 110 families and at least 230 genera of angiosperms (Yang et al., 2012). From the family *Poaceae*, known as grasses, only *Eragrostis* subspecies (spp.) and *Sporobolus cryptandrus* have been reported to produce seed coat mucilage (Kreitschitz et al., 2009; Grubert, 1974). Since barley has not been reported to produce seed coat mucilage, a possible effect of the higher abundance of arabinose and rhamnose on the seed longevity phenotype of Cebada Capa might be questionable.

In order to identify metabolites possibly influencing the seed longevity phenotype, the metabolomics data set was searched for metabolites showing a similar trend in abundance both in Cebada Capa and the L94 NILs. Two of the Cebada Capa highly abundant metabolites are also highly abundant in some of the NILs. As shown in figure 15, galactose is significantly higher abundant in Cebada Capa, 116-5 and NIL-SLQ1.2, in comparison to L94.
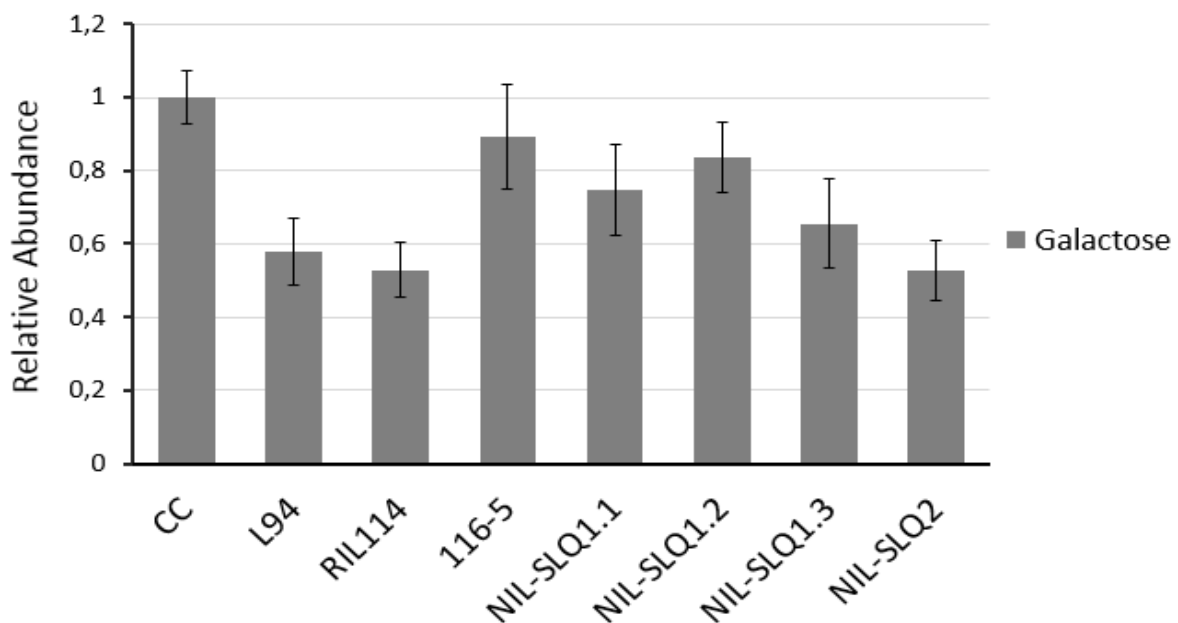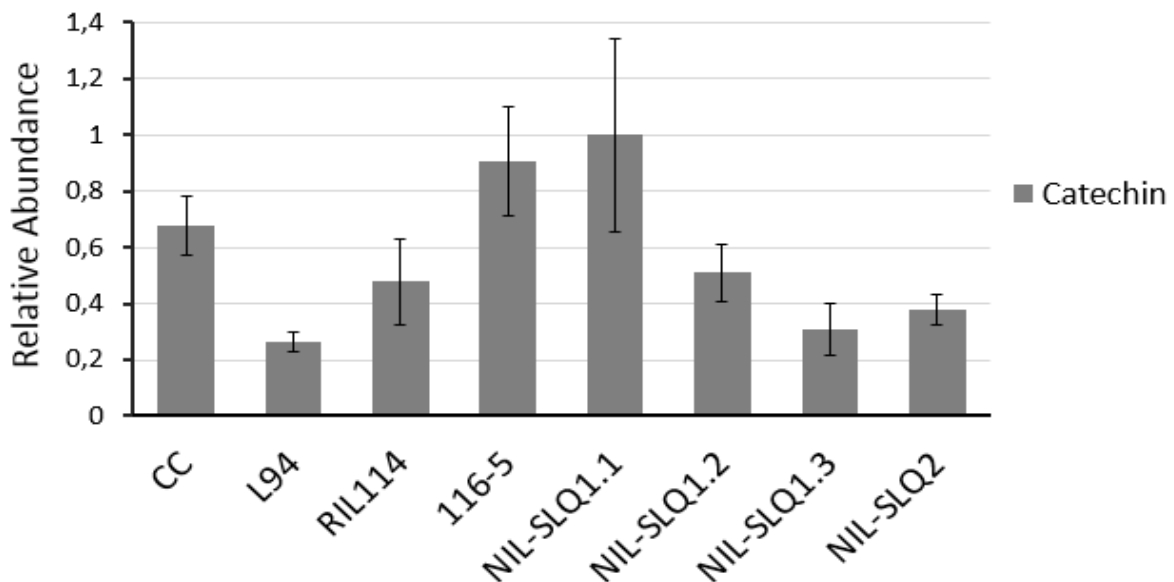

Figure 15. **Bar plot-visualization of galactose relative abundances in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**

Relative abundances of galactose (meox1 derivatization) were calculated by referring metabolite peak intensities of samples to an internal standard. Peak areas were determined using the Xcalibur software (Thermo Scientific). The value of the sample for which the highest abundance was detected, was set to 1. Standard deviations were calculated from four replicates per analyzed sample.

Like arabinose and rhamnose, galactose is a monosaccharide sugar. Sugars in seeds have been proposed to act as signals that regulate and influence seed development (Borisjuk et al., 2004). Several studies have shown that sugars help to maintain the structural integrity of membranes and proteins under dry conditions due to the formation of a glassy state to limit deteriorative reactions (Bernal-Lugo and Leopold, 1995; Nishizawa et al., 2008). Raffinose family oligosaccharides (RFO) are proposed to be involved in desiccation tolerance and seed longevity (Horbowicz and Obendorf, 1994; Obendorf, 1997). RFOs are synthesized from sucrose by the subsequent addition of activated galactose moieties donated by galactinol (Peterbauer and Richter, 2001).

Interestingly, galactinol content has been recently shown to correlate with seed longevity in *Arabidopsis*, cabbage and tomato, and was suggested as marker for seed longevity (Souza Vidigal et al., 2016). Galactinol itself was also detected in the data set. As shown in figure 16, there is no significant difference between the two parental lines and the variation within replicates of the same genotype is very high. Although it is difficult to make a statement about the abundance of galactinol in the analyzed samples, because of the large variation between replicates, galactinol seems to be lower abundant in CC than in L94. This finding is the opposite of what would be expected in the case of galactinol promoting seed longevity.
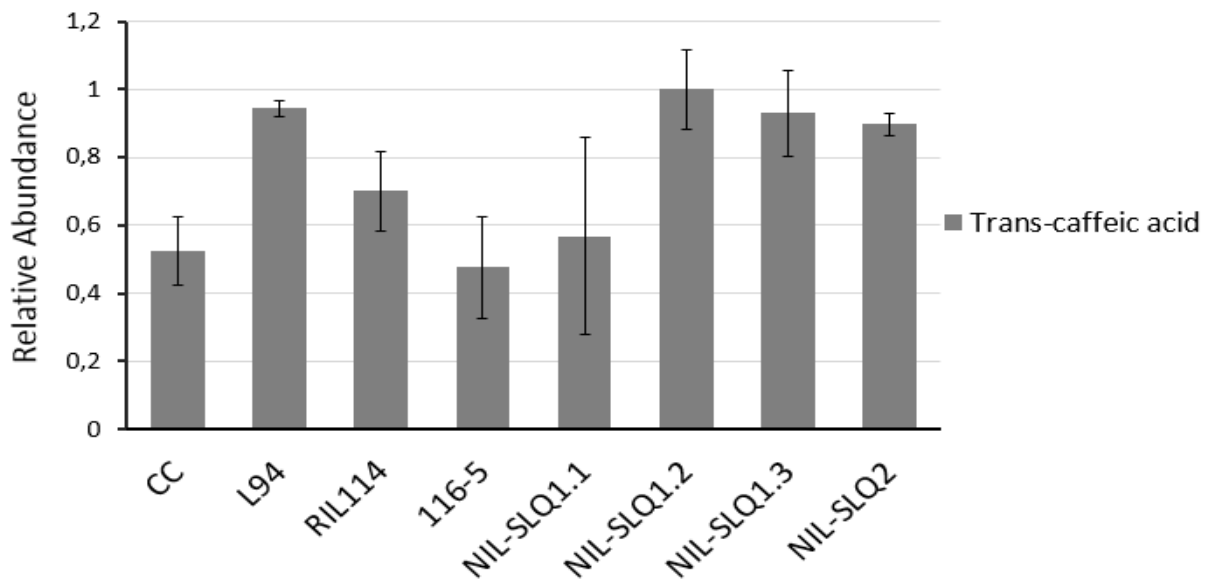

Figure 16. **Bar plot-visualization of galactinol relative abundances in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**
Relative abundances of galactinol (9-times derivatization) were calculated by referring metabolite peak intensities of samples to an internal standard. Peak areas were determined using the Xcalibur software (Thermo Scientific). The value of the sample for which the highest abundance was detected, was set to 1. Standard deviations were calculated from four replicates per analyzed sample.

Apart from galactose, catechin also shows a similar high abundance both in Cebada Capa and in some of the L94 NILs. As shown in figure 17, catechin is significantly higher abundant in Cebada Capa, RIL114, 116-5, NIL-SLQ1.1, NIL-SLQ1.2 and NIL-SLQ2 in comparison to L94. Catechin is a flavan-3-ol which forms part of the chemical family of flavonoids. Flavonoids act as antioxidants in plants due to their ability to

reduce free radical formation and to scavenge free radicals (Pietta, 2000). Examples of oxygen-centered free radicals, known as reactive oxygen species (ROS), include superoxide ($O_2^{\cdot-}$), peroxyl ($ROO^{\cdot}$), alkoxyl ($RO^{\cdot}$), hydroxyl ($HO^{\cdot}$) and nitric oxide ($NO^{\cdot}$). In dry seeds, enzyme activities are extremely reduced and ROS probably originate from non-enzymatic reactions, such as lipid peroxidation that occur even with very low moisture contents (McDonald, 1999) and from Amadori and Maillard reactions (Sun and Leopold, 1995). In hydrated seeds, however, all metabolically activate compartments may become sources of ROS, such as glyoxysomes, peroxisomes, mitochondria, chloroplasts and plasma membranes (Bailly, 2004). It has been proposed that deterioration of seeds during storage partly originates from oxidative reactions (Roqueiro et al., 2010). Hence, the generation of oxidative stress and the defense against oxidative stress or its consequences may play an important role in seed longevity. The high abundance of catechin in Cebada Capa and the before mentioned genotypes, might therefore play an important role for the protection against oxidative stress and could lead to an enhanced seed longevity phenotype.

Figure 17. **Bar plot-visualization of catechin relative abundances in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**

Relative abundances of catechin (5-times derivatization) were calculated by referring metabolite peak intensities of samples to an internal standard. Peak areas were determined using the Xcalibur software (Thermo Scientific). The value of the sample for which the highest abundance was detected, was set to 1. Standard deviations were calculated from four replicates per analyzed sample.

As mentioned before, some metabolites were found to be significantly lower in abundance in Cebada Capa than in L94. As an example, shown in figure 18, the compound trans-caffeic acid was significantly lower in Cebada Capa, RIL114, 116-5 and NIL-SLQ1.1. Caffeic acid is a phenolic acid, and like catechin, promotes antioxidative activity. Phenolic acids act as antioxidants by trapping free radicals (Brewer, 2011). Although the findings of trans- caffeic acid and catechin are contrary, the relative abundance of the two compounds was found to be in the same range as in Cebada Capa. As

shown in figure 17 and 18, the relative abundance of Cebada Capa for catechin is 0.675 ± 0.105, and 0.526 ± 0.099 for trans-caffeic acid.

Figure 18. **Bar plot-visualization of trans-caffeic acid relative abundances in mature, non-aged seeds of L94, Cebada Capa, 116-5, RIL114 and L94 NILs.**

Relative abundances of trans-caffeic acid (3-times derivatization) were calculated by referring metabolite peak intensities of samples to an internal standard. Peak areas were determined using the Xcalibur software (Thermo Scientific). The value of the sample for which the highest abundance was detected, was set to 1. Standard deviations were calculated from four replicates per analyzed sample.

The detection of sugars in the metabolomics analysis, which might possible affect the seed longevity phenotype of Cebada Capa, provides a possible link to the first two described analytical approaches. Combining transcriptomics and proteomics analyses, the UDP-glycosyltransferase MLOC_11661.1 was chosen for further analyses as possible candidate gene for SLQ2. Glycosyltransferases (GTs) are enzymes responsible for glycosylation of plant compounds, which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, for example sugars, lipids, proteins, nucleic acids, antibiotics and other small molecules, with either retention or inversion of configuration at the anomeric center (Lairson et al., 2008). GTs of family 1 characteristically contain a carboxy-terminal consensus sequence termed ʹplant secondary product glycosyltransferase boxʹ (PSPG box) and use uridine 5'-diphospho sugars as sugar donors, and are named UDP-sugar glycosyltransferases (UGTs) (Jones and Vogt, 2001). UGTs might play important roles in maintaining cell homeostasis, regulating plant growth and improving their defense responses to stressful environments (Jones and Vogt, 2001; Bowles et al., 2005). Several UGTs have been shown to glycosylate phytohormones *in vitro* and participate in phytohormone homeostasis *in vivo* (Jackson et al., 2001; Wang et al., 2011). UGTs have also been identified to participate in regulating plant adaptation to abiotic stresses. A recent study showed that overexpression of UGT74E2, encoding an indole-3-butyric acid (IBA) GT, greatly affected indole-3-butyric acid (IBA) homeostasis and the expression of various downstream genes, including transcription factors, glycosidases or genes involved in hormone signaling, leading to increased tolerance to salinity and drought stress in *Arabidopsis thaliana* (Tognetti et al., 2010).

Discussion

Chapter 3 focuses on analytical approaches, which have been performed in this study, to understand the complex biochemical and molecular processes underlying seed longevity. All analyses were performed on mature, non-aged seeds. Conducting an RNA-seq analysis, a number of differentially expressed genes was identified within the SLQ regions, of which some additionally showed SNPs in their coding regions. Clusters of differentially expressed genes, located outside of any SLQ region, could be observed and were considered target genes of the QTLs. To narrow down the number of possible candidate and downstream targets, an overlap between the RNA-seq data and data derived from a total seed proteome analysis on the same genotypes was conducted. Two candidate genes for SLQ1.1 and SLQ2 and one downstream target were selected for further analyses. The selection of these genes was mainly based on the relative abundance patterns of the encoded proteins and recent findings of seed longevity studies in *Arabidopsis thaliana*. However, several SNPs were detected in the coding sequence of the UDP-glycosyltransferase by RNA-Seq, and the malic enzyme AK248526.1 was found to be differentially expressed between the two parental lines. In general, only a small overlap between transcriptomic and proteomic data could be observed. Ongoing improvements of the barley reference genome would certainly facilitate the assignment of differentially expressed genes, with so far unknown location, as candidate genes or downstream targets. The role of selected genes could be tested by ordering T-DNA knock-out lines of *Arabidopsis* and testing them for their seed longevity phenotype, as performed by Nguyen (2014). Further possible analyses, such as the complementation of *Arabidopsis* seed longevity mutants, are described in chapter 4 on the basis of the selected candidate and downstream target genes. Since the SLQs have been identified and confirmed with controlled deterioration tests, as discussed in detail in chapter 2, it might be advisable to perform the same approaches in the future on aged seeds. The expression of specific genes could be monitored over the period of the controlled deterioration tests. A special focus would be on the overlap between data sets generated with mature, non-aged seed and those generated with aged seeds. It might be of special interest to know, whether the three selected possible candidate genes and downstream targets are present in these overlaps.

As shown exemplarily for some of the compounds, the variation within replicates of the same genotype in the metabolomics analysis is relatively high. This variation could be reduced in future analyses by increasing the number of analyzed seeds. Since the space in the Teflon capsules, used for the grinding, is limited, it might also be possible to perform the analysis on embryos instead of whole seeds. It might be interesting to know whether an increase in seed amount, leads to different results for some of the in this study quantified compounds. Special focus would be on compounds such as galactinol, which has been suggested as marker for seed longevity (Souza Vidigal et al., 2016). The identification of mainly sugars and antioxidants e.g. catechin, emphasizes the importance of protection against oxidative stress. Since oxidative stress is considered to play a critical role in seed ageing in various plant species (Bailly, 2004), such compounds could be important for an enhanced seed longevity phenotype. Therefore, it might also be considered to perform future metabolomics analyses not only after controlled deterioration tests, but also after ROS related assays.

Material and methods

Plant material and growth conditions

Seeds of the barley landraces L94, Cebada Capa, L94 NILs sequenced in 2014, the cleanest L94 NILs and the lines 116-5 and RIL114 were sown in 96-well trays filled with a custom-made peat and clay soil mixture (EinheitsErde® ED73 Osmocote, Einheitserdewerke Werkverband e.V., Sinntal Altengronau, Germany). Seeds were germinated in the greenhouse under long day (LD) conditions (16h, 22°C day; 8h, 18°C night). After six weeks, plants were transferred to 1.3 L pots and maintained under LD conditions. The pots were arranged in 6 rows with an inter-row distance of 20 cm, where each row contained 20 pots with a distance of 10 cm. The entire experiment was divided into 6 blocks with at least five replicate plants per genotype within each block. To avoid edge effects, the plot was surrounded by border pots containing wild-type barley plants. The plot was irrigated by hand and treated with additional fertilizer or pesticides when necessary. Plants reached maturity around the some date and were immediately harvested. Spikes were threshed individually by hand and seeds were dried in a drying room, conditioned at circa 10% relative humidity and 20°C, for two months prior to the experiments.

RNA extraction from barley embryos

Approximately 20 embryos were excised from mature barley seeds by carefully slicing open the endosperm with a sterile scalpel, avoiding cutting the embryo, and removing the embryo with a sterile forceps. Embryos were collected in sterile 1.5 ml Eppendorf tubes and flash frozen in liquid nitrogen. The samples may be stored at -80°C before continuing with the RNA extraction procedure. The embryo material was ground using sterile mortar and pestle, which had been chilled with liquid nitrogen. RNA was extracted from the ground material using the RNAEasy® Mini Kit (50) from Qiagen (Qiagen, Hilden, Germany) following the manufacturers protocol and stored at -80 °C after DNase treatment (Ambion, Carlsbad, USA).

RNA-sequencing and data analysis

RNA-sequencing was performed by the Genome Center of the Max Planck Institute of Plant Breeding research (MPIPZ, Cologne, Germany). The quality of the RNA was evaluated before library preparation using a bioanalyzer (Agilent). The Illumina TruSeq libraries were prepared using the manufacture's protocol (version 2, Illumina). Single end sequencing was performed on the HiSeq 2000 (Illumina®) platform. For each library a minimum of 15 million reads was generated by multiplexing eight libraries. The initial quality control of the raw reads was performed using the FastQC software. Reads were trimmed using the Trimmomatic platform, embedded within the trinity pipeline (Grabherr et al., 2011; https://github.com/trinityrnaseq/trinityrnaseq/wiki), using the following default criteria: phred 33, leading and trailing 3, sliding window 4:15 and a minimum read length of 36. Sequence alignments were performed by Bowtie2 (version 2.1.0; http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), using a merged dataset of high-confidence (HC) and low-confidence (LC) predicted barley genes (The International Barley Genome Sequencing Consortium, 2012) as reference. SAMtools

(Li et al., 2009; version 1.2; http://www.htslib.org/) was used for BAM format conversion, sorting and indexing and read duplicate removal was conducted using the Picard command-line tool MarkDuplicate (version 1.110; https://broadinstitute.github.io/picard/command-line-overview.html). To correct for misalignments, the Genome Analysis ToolKit (GATK, version 3.1) re-aligner was used using default settings. The Trimmed Mean of M values (TMM) method was used to normalize libraries sizes. The TMM factor was computed for each lane, with one lane being considered as a reference sample and the others as test samples. Normalized read counts were obtained by dividing raw read counts by these re-scaled normalization factors. Variants were obtained using the GATK UnifiedGenotyper platform (minimum phred score of 30). Refinement of the variants was performed with the GATK Variant Filtration tool (Fisher Strand values FS >30.0; Qual By Depth values QD <2.0) to reduce false positive SNPs. Resulting SNP calls were kept for further analysis, if they passed the filtration step and their read coverage exceeded four reads. Transcripts containing filtered homozygous SNPs were mapped to their respective position along the POPSEQ map of barley with R (The International Barley Genome Sequencing Consortium, 2012). For expression analysis, the reads were aligned to the high-confidence (HC) and low-confidence (LC) gene set as described above and only in this case, the read duplicates were not removed from the BAM file. Raw counts were extracted from the BAM file using Salmon (Patro et al., 2015). Differentially regulated reads were called with a false discovery rate of 0.05 using the R bioconductor package limma-vroom (Ritchie et al., 2015).

Total seed proteome profiling by LC-MS/MS

This analysis was performed by the proteomics unit of the Max Planck Institute for Plant Breeding (MPIPZ Cologne, Germany), with support from Dr. Iris Finkemeier, Dr. Katharina Kramer and Anne Harzen.

Proteolytic digestion and fractionation

Samples of approximately 100 seeds per genotype were ground using a coffee grinder (Siemens, MC23200GB). 200 mg-400 mg of the ground material was transferred to 2 mL Safe-Lock Eppendorf® tubes and whole proteome samples (50 µg) were processed using the FASP method (Wisniewski et al., 2009) as described in detail in Hartl et al. (2015). Cysteines were alkylated by incubation with 55 mM chloroacetamide in urea (UA) buffer for 30 min at room temperature in the dark, followed by three washing steps with UA buffer. Incubation with LysC in UA buffer (1:50 enzyme-to-protein ratio) was performed overnight at room temperature. The samples were diluted with ABC buffer (50 mM ammonium bicarbonate) to a final urea concentration of 2 M. Trypsin digestion (1:100 enzyme-to-protein ratio) was performed for four hours at room temperature. The samples were passed through the filter by centrifugation at 14.000 g for 10 min and residual peptides were eluted with 50 µl ABC buffer. Formic acid was added to a final concentration of 0.5%. Peptides were desalted and pre-fractionated prior LC-MS/MS analysis into three fractions using the Empore Styrene Divenylbenzene Reversed-Phase Sulfonate material (SDB-RPS, 3M), as described in detail by Kulak et al. (2014).

LC-MS/MS data acquisition

Dried peptides were re-dissolved in 2% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) for analysis, and adjusted to a final concentration of 0.1 µg/µl. Samples were analyzed using an EASY-nLC 1000 (Thermo Fisher) coupled to a Q Exactive ™ Plus mass spectrometer (Thermo Fisher). Peptides were separated on 16 cm frit-less silica emitters (New Objective, 0.75 µm inner diameter), packed in-house with reversed-phase ReproSil-Pur C18 AQ 1.9 µm resin (Dr. Maisch). Peptides (0.5 µg) were loaded on the column and eluted for 120 min using a segmented linear gradient of 0% to 95% solvent B (solvent A: 5% ACN, 0.5% formic acid (FA); solvent B: 100% ACN, 0.5% FA) at a flow rate of 250 nL/min. Mass spectra were acquired in data-dependent acquisition mode using a topN (N=15) method. MS spectra were acquired in the Orbitrap analyzer with a mass range of 300-1750 m/z at a FWHM (full width at half maximum) resolution of 70.000 and an AGC target value of $3x10^6$ ions. Precursors were selected with an isolation window of 1.3 m/z. HCD fragmentation was performed at a normalized collision energy of 25. MS/MS spectra were acquired applying an AGC target value of 10^5 ions at an FWHM resolution of 17.500 and a fixed first mass of m/z 100. Singly-positively charged peptides, peptides with a charge state greater than 6, or with unassignable charge state were excluded from fragmentation for MS/MS, dynamic exclusion for 30 seconds prevented repeated selection of precursors.

Data analysis

Raw data were processed using the MaxQuant software (version 1.5.2.8, http://www.maxquant.org/) (Cox and Mann, 2008) with label-free quantification (LFQ) and iBAQ enabled (Cox et al., 2014). MS/MS spectra were searched by the Andromeda search engine against the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/). Sequences of 248 common contaminant proteins and decoy sequences were automatically added during the search. Trypsin was chosen as proteolytic enzyme used for protein digestion and a maximum of two missed cleavages was allowed. The minimal peptide length was set to seven amino acids. Carbamidomethylation of cysteine residues was set as fixed, oxidation of methionine and protein N-terminal acetylation as variable modifications. Peptide-spectrum-matches and proteins were retained, if they were below a false discovery rate of 1%. Subsequent quantitative statistical analyses were performed in Perseus (version 1.5.2.6, http://www.maxquant.org/; Cox and Mann, 2012). Hits were only retained, if they were quantified in at least three of the four replicates in any line. LFQ intensities were log2-transformed. Two-sample t-tests were performed with a p-value of 1% as cut-off. Log2 ratios were calculated by replacing missing intensity values with zero. For the gene ontology (GO) analysis, GO terms were obtained from the PGSB barley genome database (http://pgsb.helmholtz-muenchen.de/plant/barley/) and classified using the CateGOrizer tool (Hu et al., 2008). The heat maps were drawn using the R (https://www.r-project.org) function heatmap.2.

Metabolite profiling by GC-MS

This analysis was performed in the group of Prof. Dr. Ilse Kramer (University Innsbruck, Austria), with support of Dr. Erwann Arc. Chemical derivatization and GC-MS metabolite profiling analysis were performed as previously described (Fiehn, 2006; Fiehn et al., 2008).

Sample preparation

Four replicates of 20 dry seeds per sample were ground in a 5 mL liquid nitrogen-cooled Teflon capsule with an agate ball at 2800 rpm for 30 seconds using a Mikro-Dismembrator S (B. Braun, Biotech International, Germany) and the resulting powder was collected in 2 ml Safe-Lock Eppendorf® tubes. 15 mg of this powder were re-suspended in 1 ml of frozen (-20°C) Water: Acetonitrile: Isopropanol (2:3:3) containing $^{13}C_6$-Sorbitol at 4 µg/ml and extracted for 10 min at 4°C with continuous shaking at 1400 rpm. Insoluble material was removed by centrifugation at 20.000xg for 5 min. 25 µl were then collected and dried for 3 h in a Speed-Vac. One blank tube underwent the same steps as the samples.

Derivatization

Vacuum-dried samples were dissolved in 10 µL of methoxylamine (20 mg/mL pyridine) and incubated at 28°C for 90 min, with continuous shaking, in a thermomixer. 90µl of N-methyl-N-trimethylsilyl-trifluoroacetamide (MSTFA) (Aldrich 394866-10x1ml) were then added and the reaction continued for 30 min at 37°C. After cooling, the content of each tube was transferred to an Agilent vial for injection.

Analysis

Two hours after derivatization, 1 µl of each sample was injected in splitless mode on a Trace 1300 gas chromatograph coupled to a TSQ8000 triple quadrupole mass spectrometer (Thermo Scientific). The column was an Rxi-5SilMS from Restek (30 m with 10 m integra-guard column). The liner (Restek # 20994) was changed before each series of 24 samples analysis. The oven temperature ramp was 70°C for 7 min then 10°C/min to 330°C for 10 min. A constant helium flow rate of 1 mL/min was used. Temperatures were the following: injector: 250°C, transfer line: 300°C and source: 330°C. Sample and blank runs were distributed randomly in the sequencer. An alkane mix (C10 to C20, C22, C24, C28, C32 and C36) was injected in the middle of the queue for external RI calibration. 5 scans per second were acquired.

Data processing

Raw datafiles were analyzed with AMDIS (http://chemdata.nist.gov/mass-spc/amdis/). A home retention indices/ mass spectra library built from the NIST, Golm, and Fiehn databases and standard compounds was used for metabolites identification. Peak areas were then determined using the Xcalibur software (Thermo Scientific).

References

Arc, E. et al. (2011), Reboot the system thanks to protein post-translational modifications and proteome diversity: how quiescent seeds restart their metabolism to prepare seedling establishment. Proteomics 11: 1606-1618

Arsovski, A.A. et al. (2010), Seed coat mucilage cells of Arabidopsis thaliana as a model for plant cell wall research. Plant Signal Behav. 5: 796-801

Belmonte, M.F. et al. (2013), Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. Proc. Natl. Acad. Sci. U S A 110: 435-444

Bailly, C. (2004), Active oxygen species and antioxidants in seed biology. Seed Sci. Res. 14: 93-107

Bernal-Lugo, I. and Leopold, A.C. (1995), Seed stability during storage: raffinose content and seed glassy state, Seed Sci. Res. 5: 75-80

Brewer, M.S. (2011), Natural antioxidants: sources, compounds, mechanisms of action, and potential applications. Comprehensive Reviews in Food Science and Food Safety 10: 221-247

Borisjuk, L. et al. (2004), Seed development and differentiation: a role for metabolic regulation. Plant Biol. 6: 375-386

Bowles, D. et al. (2005), Glycosyltransferases: managers of small molecules. Curr. Opin. Plant Biol. 8: 254-263

Cadman, C.S. et al. (2006), Gene expression profiles of *Arabidopsis* Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. Plant J. 46: 805-822

Caffall, K.H. and Mohnen, D. (2009), The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate Research 344: 1879-1900

Chen, M. et al. (2009), System analysis of an *Arabidopsis* mutant altered in *de novo* fatty acid synthesis reveals diverse changes in seed composition and metabolism. Plant Physiol. 150: 27-41

Cox, J. and Mann, M. (2008), MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26: 1367-72

Cox, J. and Mann, M. (2012), 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics 13: S12

Cox, J. et al. (2014), Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Mol. Cell Proteomics 13: 2513-2526

Debeaujon, I. et al. (2000), Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiol. 122: 403-414

Fait, A. et al. (2006), *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol. 142: 839-854

Fiehn, O. (2006), Metabolite profiling in *Arabidopsis*. Methods Mol. Biol. 323: 439-447

Fiehn, O. et al. (2008), Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J. 53: 691-704

Fu, C.H. et al. (2011), OrchidBase: a collection of sequences of the transcriptome derived from orchids. Plant Cell Physiol. 52: 238-243

Gallardo, K. et al. (2001), Proteomic analysis of *Arabidopsis* seed germination and priming. Plant Physiol. 126: 835-848

Grabherr, M. et al. (2011), Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology 29: 644-652

Grubert, M. (1974). Studies of the distribution of myxospermy among seeds and fruits of *Angiospermae* and its ecological importance. Acta Biol. Venez. 8: 315-551

Hartl, M. et al. (2015), Identification of lysine-acetylated mitochondrial proteins and their acetylation sites. Methods Mol. Biol. 1305: 107-121

Horbowicz, M. and Obendorf, R.L. (1994), Seed desiccation tolerance and storability: dependence on flatulence-producing oligosaccharides and cyclitols-review and survey. Seed Sci. Res. 4: 385-405

Hu, Z.L. et al. (2008), "CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories". Online Journal of Bioinformatics 9: 108-112

Jackson, R.G. et al. (2001), Identification and biochemical characterization of an *Arabidopsis* indole-3-acetic acid glucosyltransferase. J. Biol. Chem. 276: 4350-4356

Jones P. and Vogt T. (2001), Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213: 164-174

Kulak N.A. et al. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat. Methods. 11: 319-24

Kreitschitz, A. et al. (2009), Slime cells on the surface of Eragrostis seeds maintain a level of moisture around the grain to enhance germination. Seed Sci. Res. 19: 27-35

Lairson, L.L. et al. (2008), Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 77: 521–555

Le, B.H. et al. (2010), Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. Proc. Natl. Acad. Sci. U S A 107: 8063-8070

Li, H. et al. (2009). The sequence alignment/map format and samtools. Bioinformatics 25: 2078-2079

Lister, R. et al. (2009), Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr. Opin. Plant Biol. 12: 107-118

Matsumoto, T. et al. (2011), Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. Plant Physiol. 156: 20-28

Mayer, B. (ed.) (2011), Bioinformatics for Omics Data: Methods and Protocols, Methods in Molecular Biology, vol. 719, DOI 10.1007/978-1-61779-027-0_1, © Springer Science + Business Media

McDonald, M.B. (1999), Seed deterioration: physiology, repair and assessment. Seed Sci. Tech. 27: 177-237

Nakabayashi, K. et al. (2005) Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. Plant J. 41: 697–709

Nguyen, T.P. (2014), Seed Dormancy and Seed Longevity: from genetic variation to gene identification, Utrecht University Repository (Dissertation), ISBN: 978-90-393-6094-1

Nguyen, T.P. et al. (2015), A role for seed storage proteins in *Arabidopsis* seed longevity. J. Exp. Bot. 66: 6399-6413

Nishizawa, A. et al. (2008), Galactinol and raffinose constitute a novel function to protect plants from oxidative damage, Plant Physiol. 147: 1251-1263

Obendorf, R.L. (1997), Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance,Seed Sci. Res. 7: 63-74

Patro R. et al. (2015), Salmon: Accurate, versatile and ultrafast quantification from RNA-seq data using Lightweight-Alignment. bioRxiv 021592. doi:10.1101/021592

Peterbauer, T. and Richter, A (2001), Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Sci. Res. 1: 185-197

Pietta P.G. (2000), Flavonoids as antioxidants, J. Nat. Prod. 63: 1035-1042

Rajjou, L. et al. (2004), The effect of α-amanitin on the *Arabidopsis* seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiol. 134: 1598-1613

Rajjou, L. et al. (2007), The use of proteome and transcriptome profiling in the understanding of seed germination and identification of intrinsic markers determining seed quality, germination efficiency and early seedling vigour. Seeds: Biology, Development and Ecology: 149-158

Rajjou, L. et al. (2008), Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. Plant Physiol. 148: 620-641

Ritchie M.E. et al. (2015), Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 20(43): e47

Roqueiro, G. et al. (2010), Effects of photooxidation on membrane integrity in Salix nigra seeds. Annals of Botany 105: 1027-1034

Ruuska, S.A. et al. (2002), Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell, 14: 1191-1206

Saisho, D. and Takeda, K. (2011), Barley: emergence as a new research material of crop science. Plant Cell Physiol. 52:724-727

Souza Vidigal, D. et al. (2016), Galactinol as marker for seed longevity. Plant Sci. 246: 112-118

Su, C.L. et al. (2011), *De novo* assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. Plant Cell Physiol. 52: 1501-1514

Sun, W.K. and Leopold, A.C. (1995), The Maillard reaction and oxidative stress during aging of soybean seeds. Physiol. Plant. 94: 94-104

The International Barley Genome Sequencing Consortium (2012), A physical, genetic and functional sequence assembly of the barley genome. Nature 491: 711-716

Tognetti, V.B. et al. (2010), Perturbation of indole-3-butyric acid homeostasis by the UDP-Glucosyltransferase UGT74E2 modulates Arabidopsis architecture and water stress tolerance. Plant Cell 22: 2660-2679

Verdier, J. et al. (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: 757-774

Wang, J. et al. (2011), N-Glucosyltransferase UGT76C2 is involved in cytokinin homeostasis and cytokinin response in Arabidopsis thaliana. Plant Cell Physiol. 52: 2200-2213

Western, T.L. (2012), The sticky tale of seed coat mucilages: production, genetics, and role in seed germination and dispersal, Seed Science Research 22: 1-25

Wicker, T. et al. (2009), A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J. 59: 712-722

Wisniewski, J.R. et al. (2009), Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J. Proteome Res. 8: 5674-5678

Wu, X.L. et al. (2011), Proteomic analysis of seed viability in maize. Acta Physiol. Plant 33: 181-191

Xin, X. et al. (2011), Proteome analysis of maize seeds: the effect of artificial ageing. Physiol. Plant 143: 126-138

Yang, X. et al. (2012), More than just a coating: Ecological importance, taxonomic occurrence and phylogenetic relationships of seed coat mucilage. Perspectives in Plant Ecology, Evolution and Systematics 14: 434-442

Yuan, J.S. et al. (2008), Plant systems biology comes of age. Trends Plant Sci. 13: 165-171

Chapter 4

Testing of NADP-dependent Malic Enzymes as
Candidates and Downstream Targets for
Seed Longevity QTLs and Their Role in
Redox Regulation

Introduction

Detoxification of reactive oxygen species (ROS) that result in oxidative stress and maintenance of redox homeostasis are crucial for seed vigor (Rajjou et al., 2008). Processes and proteins that could underlie QTLs for seed longevity are related to oxidative stress and among these Nicotinamide Adenine Dinucleotide Phosphate (NADP)-dependent Malic Enzymes (NADP-MEs) could play a role. Levels for some of the latter proteins differed between the two parents and some of the NILs and their recurrent parent, as shown in chapter 3. NADP-MEs are ancient and diverse across bacteria, fungi, animals and plants. As shown in figure 1, they catalyze the reversible oxidative decarboxylation of L-malate to produce pyruvate, CO_2 and NADPH in the presence of a bivalent cation (Drincovich et al., 2001; Fu et al., 2009). Depending on the source of enzyme, either Mg^{2+} and/or Mn^{2+} serve as the metal cofactor (Edwards and Andreo, 1992).
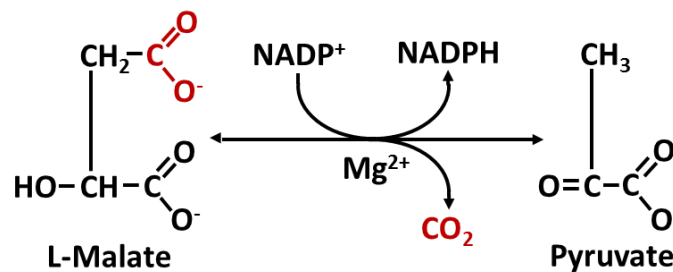
Figure 1. Oxidative decarboxylation of L-malate to pyruvate

In the mammalian liver, NADP-MEs are the major NADPH-generating enzymes for lipogenesis (Bagchi et al., 1987). In plants, NADP-MEs are classified into photosynthetic and non-photosynthetic isoforms based on their participation in carbon fixation metabolism (Drincovich et al., 2001; Maier et al., 2011). The photosynthetic NADP-MEs participate in photosynthesis in the bundle sheath chloroplasts of C_4-plants and in the cytosol of Crassulacean Acid Metabolism (CAM) plants by providing CO_2 to be used in carbon fixation by Ribulose-1,5-Bisphosphate Carboxylase Oxygenase (RuBisCO) (Edwards and Huber, 1981; Honda et al., 2000). In contrast, the non-photosynthetic NADP-MEs are present in the plastids or the cytosol of tissues of C_3-, C_4-, and CAM-plants. These NADP-MEs are involved not only in fatty acid biosynthesis (Shearer et al., 2004) and cytosolic pH control (Martinoia and Rentsch, 1994), but also appear to have a role in fruit ripening (Edwards and Andreo, 1992).

Plant NADP-MEs are divided into four groups (I–IV) based on phylogenetic analysis (Wheeler et al., 2005). As shown in figure 2, groups I and II are from dicots and localized in the cytosol and in the plastids, while group III is specific to monocots. Group IV is found in both monocots and dicots (Wheeler et al., 2005; Christin et al., 2009). Four NADP-ME genes (AtNADP-ME1, AtNADP-ME2, AtNADP-ME3 and AtNADP-ME4) have been identified in the *Arabidopsis thaliana* genome (Wheeler et al., 2005). AtNADP-ME4 is localized in plastids, while the other three isoforms (AtNADPME1-3) are localized in the cytosol (Wheeler et al., 2005; 2009).
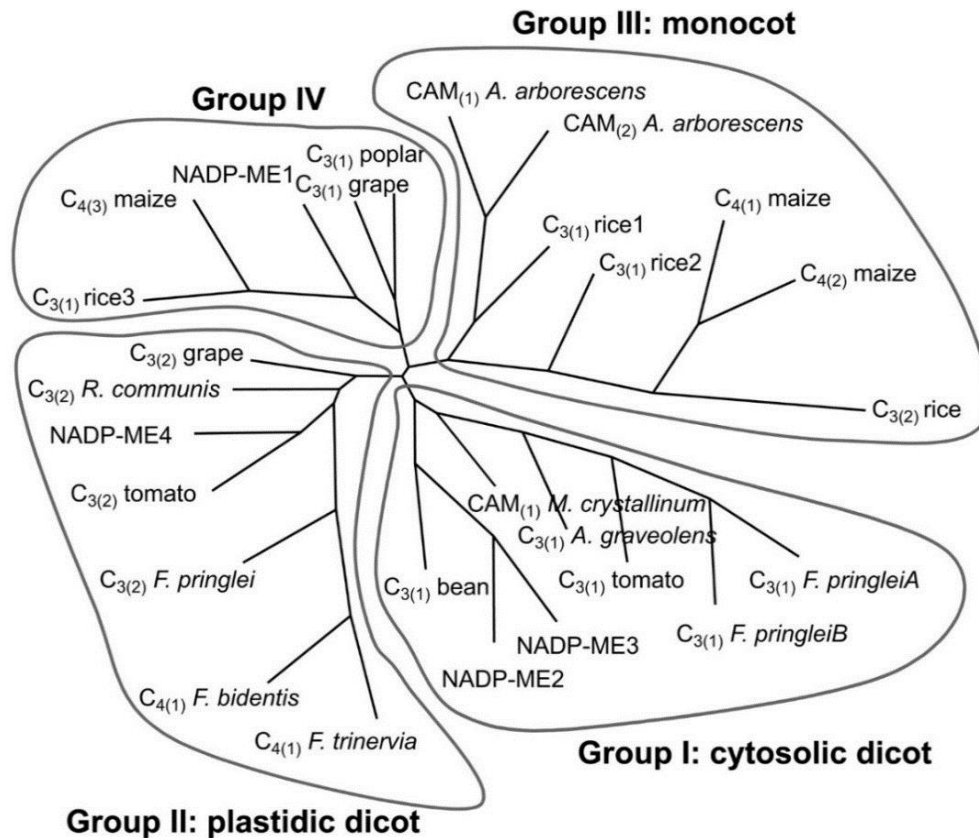
Figure 2. Phylogenetic tree of plant NADP-MEs (Wheeler et al., 2005)

The phylogenetic tree was constructed by Wheeler et al. (2005) using the following mature protein sequences: *Arabidopsis thaliana* NADP-ME1 (At2g19900, Wheeler et al., 2005); *Arabidopsis thaliana* NADP-ME2 (At5g11670, Wheeler et al., 2005); *Arabidopsis thaliana* NADP-ME3 (At5g25880, Wheeler et al., 2005); *Arabidopsis thaliana* NADP-ME4 (At1g79750, Wheeler et al., 2005), $C_{4(1)}$-NADP-ME from maize (Rothermel and Nelson, 1989), *Flaveria trinervia* (Borsch and Westhoff, 1990), and *F. bidentis* (AY863144); $C_{4(2)}$-NADP-ME from maize (Saigo et al., 2004); $C_{4(3)}$-NADP-ME from maize (AY864063); $CAM_{(1)}$-NADP-ME from *Mesembryanthemum crystallinum* (Cushman, 1992) and *Aloe arborescens* (Honda et al., 2000); $CAM_{(2)}$-NADP-ME from *A. arborescens* (Honda et al., 1997); $C_{3(1)}$-NADP-ME from bean (*Phaseolus vulgaris)* (Walter et al., 1990), poplar *(Populus spp.)* (van Doorsseleare et al., 1991), grape berries (Franke and Adams, 1995), tomato *(Lycopersicon esculentum)* (AF001270), *Apium graveolens* (AJ132257), *F. pringlei* (Lai et al., 2002), and rice ($C_{3(1)}$ rice1, rice2, and rice3; Chi et al., 2004); $C_{3(2)}$-NADP-ME from rice (Fushimi et al., 1994), *F. pringlei* (Lipka et al., 1994), tomato (AF001269), grape berries (U67426), and *Ricinus communis* (AF262997). The photosynthetic isoforms were named as $C_{4(1)}$-NADP-ME and $CAM_{(1)}$-NADP-ME; the plastidic nonphotosynthetic NADP-ME isoforms as $C_{4(2)}$-NADP-ME and $C_{3(2)}$-NADP-ME, while the non-photosynthetic cytosolic isoforms as $C_{4(3)}$-NADP-ME, $CAM_{(2)}$-NADP-ME, and $C_{3(1)}$-NADP-ME (Drincovich et al., 2001).

NADPH, the product of the reaction catalyzed by NADP-ME, provides reducing equivalents for biosynthetic reactions and the oxidation reduction involved in protection against the toxicity of reactive oxygen species (ROS), allowing the regeneration of reduced glutathione (GSH) (Rush et al., 1985). GSH and ascorbate (AsA) are crucial antioxidants and form the so-called AsA-GSH cycle, which serves the removal of toxic oxygen species. During the life cycle of the cell, ROS are formed in chloroplasts, mitochondria, plasma membrane, apoplastic space and peroxisomes. Illuminated chloroplasts, for example, produce superoxide radicals ($O_2^{.-}$) and hydrogen peroxide (H_2O_2) from sites on the thylakoids, most commonly the photosystem I (Mehler, 1951). As shown in figure 3, superoxide radicals are converted into hydrogen peroxide by either spontaneous dismutation or by

the antioxidant-enzyme superoxide dismutase (SOD). Hydrogen peroxide is reduced to water by the activity of ascorbate peroxidase (APX) using ascorbate as a cofactor (Asada, 1992). The resulting monodehydroascorbate (MDHA) can be converted into the reduced form of ascorbate (ASC) through the activity of monodehydroascorbate reductase (MDHAR), using NADPH to provide the reducing power. Alternatively, two molecules of MDHA disproportionate to dehydroascorbic acid (DHA) and ascorbate. The latter is reduced using GSH as reductant in a reaction catalyzed by dehydroascorbate reductase (DHAR). Oxidized glutathione (GSSG) is regenerated by glutathione reductase (GR) using NADPH as reductant.
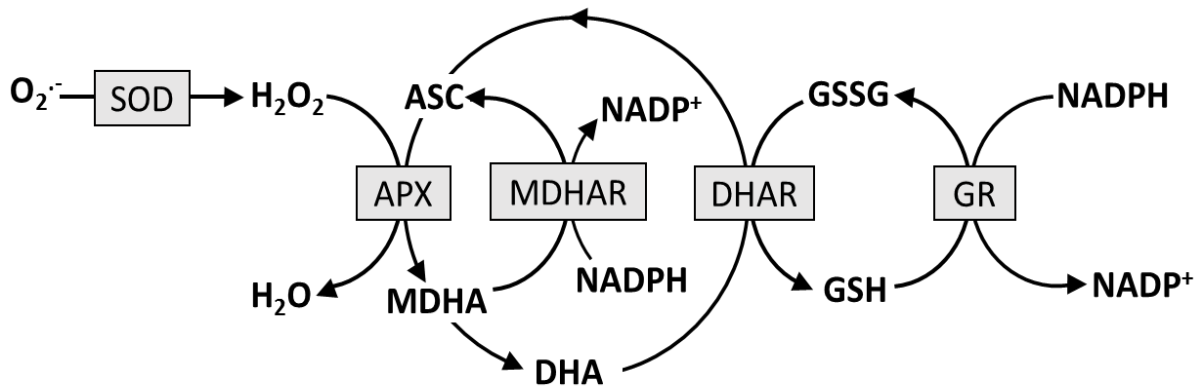
Figure 3. The ascorbate-glutathione (AsA-GSH) cycle

Scheme modified from Foyer and Shigeoka (2011). Abbreviations of enzymes, presented in gray boxes, and intermediates are given. SOD: Superoxide dismutase, APX: Ascorbate peroxidase, MDHAR: Monodehydroascorbate reductase, DHAR: Dehydroascorbate reductase, GR: Glutathione reductase, $O_2^{\cdot-}$: Superoxide radical, H_2O_2: Hydrogen peroxide, ASC: reduced ascorbate, MDHA: Monodehydroascorbate, DHA: Dehydroascorbate, DHA: Dehydroascorbic acid, GSSG: Glutathione disulfide, GSH: Reduced glutathione, $NADP^+$: Nicotinamide adenine dinucleotide phosphate.

Oxidative stress is considered to play a critical role in seed ageing in various plant species (Bailly, 2004). Peroxidation of polyunsaturated fatty acids (PUFAs) and protein oxidation are the biochemical processes most frequently associated with ageing (Stewart and Bewley, 1980; McDonald, 1999), and have frequently been attributed to reactive oxygen species (Bernal-Lugo and Leopold, 1998; McDonald, 1999; Bailly, 2004). Under normal storage conditions, the loss of seed viability is often also associated with the accumulation of DNA breaks and chromosome aberration. ROS induce a variety of lesions in DNA, including oxidized bases, apurinic/apyrimidinic (AP) sites and DNA strand breaks (Klungland et al., 1999). Hence, the defense against oxidative stress or its consequences may play an important role in seed longevity. In order to maintain seed viability, dry mature seeds are equipped with non-enzymatic antioxidants, such as β-carotene and α-tocopherol, and ROS scavenging enzymatic systems, including the before mentioned superoxide dismutase (SOD), catalases (CAT) and peroxidases (POD) (Kibinza et al., 2006). The latter defense mechanisms are only active during seed imbibition at the beginning of germination and are hence inoperative in a dehydrated state (Bailly, 2004).

The generation of ROS is an important mechanism for higher plants to transmit cellular signaling information concerning the changing environmental conditions. Plants have evolved inducible redox-based sensing mechanisms that are activated or amplified in response to abiotic stresses. The antioxidants ascorbate and glutathione are key cellular redox buffers cooperating tightly in the before mentioned AsA-GSH cycle. This ROS detoxification cycle involves the redox pairs ASC/DHA and GSH/GSSG which use the reducing power of the $NADP^+$/NADPH redox pair.

To monitor and measure the glutathione redox potential at a high spatial and temporal resolution in living cells, a redox-sensitive mutant of GFP (roGFP) was generated by the substitution of surface-exposed residues on green fluorescent protein (GFP) with cysteines in appropriate positions to form disulfide bonds (Hanson et al., 2004). Different versions of roGFPs have been generated and tested for their redox dependent fluorescence (Dooley et al., 2004; Hanson et al., 2004). The versions roGFP1 and roGFP2 differ in the S65T mutation, which is present in roGFP2. This mutation causes a distinct shift in the excitation spectrum, whereas the general ratiometric properties are retained. The excitation spectra reported by Dooley et al. (2004) indicate that for excitation at 405 and 488 nm on a confocal laser scanning microscopy (CLSM), roGFP2 offers a much larger dynamic range and is thus more appropriate than roGFP1. RoGFP2 preferentially interacts with glutaredoxins (GRXs) and therefore reports the cellular glutathione redox potential (Meyer et al., 2007). GRX mediates the reaction between roGFP2 and glutathione. The redox-dependent fluorescence of roGFP2 in *Arabidopsis* roots expressing roGFP2 in the cytosol is fully reversible by the addition of the redox reagents dithiothreitol (DTT) and H_2O_2 (Meyer et al., 2007). *In vivo* measurements of glutathione are based on the redox properties of the roGFP2 probe, whose reduced form has an excitation maximum at 488 nm, while the oxidized form has an excitation maximum at 405 nm. As shown in figure 4a, transgenic *Arabidopsis* roots expressing this reporter, show strong fluorescence when excited at 405 and 488 nm, respectively. The ratio of the two images cannot be significantly modified by the perfusion of the root with DTT solution, indicating that roGFP2 is almost fully reduced to about 95% in the undisturbed root. The addition of 1 mM H_2O_2 to the incubation solution results in immediate oxidation of roGFP2, reflected by an increase of the fluorescence ratio from 0.45 to 1.7 (Figure 4a, b and d). This oxidation *in vivo* is much faster than the response of isolated roGFP2 to 10 mM H_2O_2 in vitro (Figure 4e). No further change of the fluorescence ratio can be observed between two and 10 min after the addition of H_2O_2. The addition of 2 mM DTT after the H_2O_2 treatment leads to a decrease of the fluorescence ratio indicating reduction of roGFP2 (Figure 4c, d).
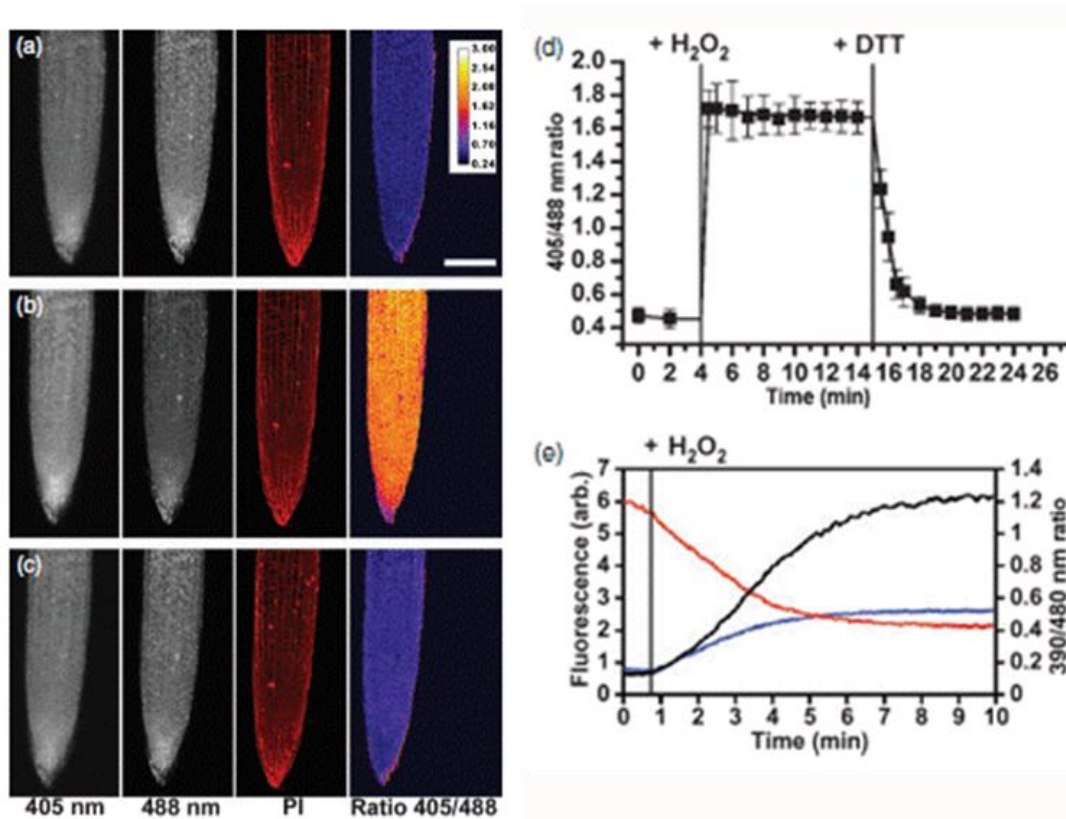
Figure 4. Redox-dependent fluorescence of roGFP2 in the cytosol of *Arabidopsis* roots (Meyer et al., 2007)
Stacks of images taken with confocal laser scanning microscopy with excitation at 405 and 488 nm, the color scale for the ratio values indicates reduced roGFP2 in blue and oxidized roGFP2 in yellow, PI: Propidium iodide staining, scale bar = 100 μm. a) Root tip before the addition of H_2O_2. b) Two minutes after the addition of 1mM H_2O_2. c) Ratio image taken 5 min after the addition of 2 mM DTT to the root previously oxidized with H_2O_2. (d) Time course for the ratio values calculated from CLSM images during the course of successive oxidation and reduction (mean ± SD; n = 4). e) Response of purified reduced roGFP2 to oxidation by 10 mM H_2O_2 *in vitro*. The ratio (black curve) was calculated through division of fluorescence at 390 nm (blue) by the fluorescence at 480 nm (red) (n = 5).

Combining a proteomics and genetics approach, the NADP-dependent malic enzyme1 (NADP-ME1) was shown to significantly affect seed longevity in *Arabidopsis thaliana* (Nguyen, 2014). In a previous study, Nguyen et al. (2012) identified twelve *GAAS* (*GERMINATION ABILITY AFTER STORAGE*) loci controlling seed longevity after natural aging in *Arabidopsis*. In order to investigate seed longevity mechanisms of the three strongest *GAAS* loci (*GAAS1*, *GAAS2* and *GAAS5*), proteome analyses were performed on dry seeds at two physiological states, after-ripened (AR) and 4 year-old (aged) seeds of four genetic lines. The genotypes used were the background line Landsberg *erecta* (L*er*) and three near isogenic lines (NILs) that contain introgression fragments of Cape Verde Islands (Cvi), Antwerp (An-1) and Shakdara (Sha) accessions, respectively, at the position of the earlier identified seed longevity QTL, NIL*GAAS1*-Cvi, NIL*GAAS2*-An-1, and NIL*GAAS5*-Sha. Proteome profiles were compared within the genotypes for the two physiological states (aged vs. AR) and between the genotypes (NILs vs. L*er*) for the AR state. Interestingly, NADP-ME1 was less abundant in aged as compared to AR seeds of NIL*GAAS5*, which makes it a possible marker for seed longevity. A subset of candidate proteins was selected to confirm their role in seed longevity by phenotypic analyses on T-DNA knock-out lines after

artificial ageing (Nguyen, 2014). The T-DNA knock-out line of the NADP-ME1 (*nadp-me1*) showed a seed longevity phenotype, i.e. *nadp-me1* is significantly less storable than the Col-0 wild-type.

In the current study, a similar approach to Nguyen et al. (2012) has been performed. Combining the results of the RNA-seq analysis and the total seed proteome analysis performed on the parental lines and the L94 NILs, two NADP-dependent malic enzymes (AK248526.1 and MLOC_35785.1) and one UDP-glycosyltransferase (MLOC_116611.1) were chosen for further analysis. The results of these analyses are shown and discussed in detail in chapter 3.

Objectives

Validation of the barley NADP-MEs by complementing the *Arabidopsis nadp-me1* mutant

To analyze the role of the two barley NADP-ME candidate proteins, the *Arabidopsis* T-DNA knock-out mutant *nadp-me1*, shown to have a seed longevity phenotype by Nguyen (2014), was complemented with the Cebada Capa and L94 alleles of the two candidates under the *Arabidopsis* ME-specific promoter, the barley ME-specific promoter, the Cauliflower mosaic virus (CaMV) 35S-promoter and Ubiquitin-promoter (Ubi). Col-0, transformed with the same constructs, was utilized as a control. Mature seeds of the generated transgenic lines were tested for their seed longevity phenotype after artificial ageing treatment as described by Nguyen (2014). The role of the UDP-glycosyltransferase was validated in a similar manner.

Generation of *nadp-me1* x roGFP2 double mutants to measure the NADP$^+$/NADPH redox potential

To investigate the redox potential of the NADP$^+$/NADPH couple in mature *nadp-me1* seeds, crossings were performed with transgenic plants expressing roGFP2. Seeds of transgenic *Arabidopsis thaliana* plants either expressing roGFP2 constitutively or GRX1-roGFP2 in the cytosol were kindly provided by Christine Foyer (University of Leeds, England) and Markus Schwarzländer (University of Bonn, Germany), respectively. In addition to the crossings, wild-type Col-0 and *nadp-me1* plants were transformed with the pBinCM binary vector carrying the GRX1::roGFP2 construct, which was also provided by Markus Schwarzländer. To facilitate the visualization of the redox potential state, transparent seeds were generated by crossing the roGFP2 x *nadp-me1* double mutants with the *transparent testa 4* mutant (*tt4*). Due to an EMS-induced mutation in a chalcone synthase gene, encoding for a key enzyme involved in the biosynthesis of flavonoids, the testa of *tt4* is transparent and anthocyanins are absent in leaves, stems and all tissues (Koornneef, 1990).

Results

Identification of NADP-MEs in the L94 NIL mapping population using an analytical approach

In the current study, an approach similar to that of Nguyen et al. (2012) has been performed. An RNA-seq analysis was employed for the identification of differentially expressed candidate genes and possible downstream targets in the L94 NIL mapping population. Additionally to the RNA-seq analysis, a total seed proteome analysis was performed on the parental lines, L94 and Cebada Capa (CC), and the L94 NILs to understand the complex biochemical and molecular processes underlying seed quality and longevity. The results of these analyses are shown and discussed in detail in chapter 3. Combining the results of the RNA-seq and the total seed proteome analysis, two NADP-dependent malic enzymes and one UDP-glycosyltransferase were chosen for further analysis. One malic enzyme, AK248526.1, was identified to be one of the most promising candidate genes because it was located inside the introgression of the L94 NIL chosen for the seed longevity QTL on top of chromosome 1 (SLQ1.1). Interestingly, its protein was found to be present in high abundance in L94 and the L94 NILs, except NIL-SLQ2 and NIL-SLQ1.1(2014), but absent in CC. In contrast, the second malic enzyme MLOC_35785.1 was suggested to be a possible downstream target gene since it was located on chromosome 3, outside of any introgression. In contrast to the first malic enzyme, this malic enzyme was found in high abundance in CC and some of the L94 NILs but absent in L94. As shown in figure 5, AK248526.1 shows the closest phylogenetically relation to the cytosolic rice NADP-ME2, whereas MLOC_35785.1 shows the closest to two plastidic *Zea mays* NADP-ME isoforms.
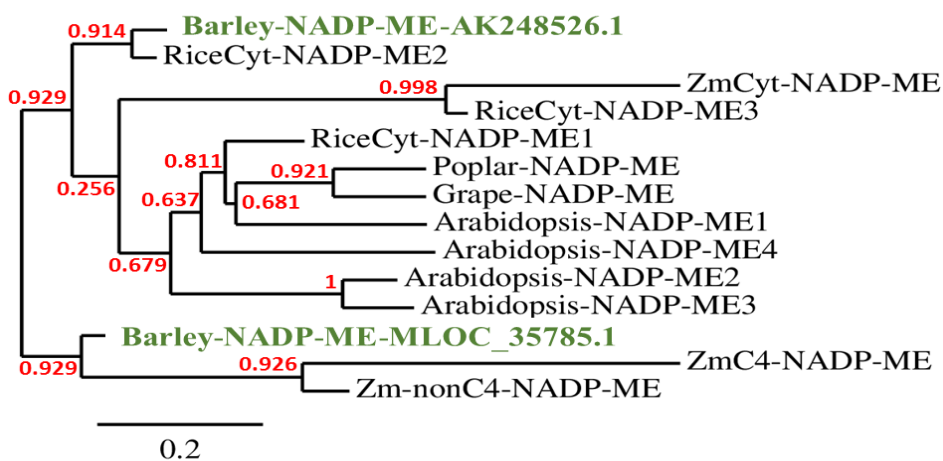

Figure 5. **Phylogenetic tree of barley NADP-MEs and other plant NADP-MEs**

Mature proteins were aligned using ClustalW (1.81) and the phylogenetic tree was visualized using *Phylogeny.fr* (Dereeper et al., 2008). The following sequences, in addition to the two in green highlighted barley NADP-MEs, MLOC_35785.1 and AK248526.1 (The International Barley Genome Sequencing Consortium, 2012), were analyzed: *Arabidopsis thaliana* NADP-ME1 (Arabidopsis-NADP-ME1; At2g19900, Wheeler et al., 2005); *Arabidopsis thaliana* NADP-ME2 (Arabidopsis-NADP-ME2; At5g11670, Wheeler et al., 2005); *Arabidopsis thaliana* NADP-ME3 (Arabidopsis-NADP-ME3; At5g25880, Wheeler et al., 2005); *Arabidopsis thaliana* NADP-ME4 (Arabidopsis-NADP-ME4; At1g79750, Wheeler et al., 2005); *Populus deltoides* NADP-ME (Poplar-NADP-ME, X56233, van Doorsseleare et al., 1991), *Vitis vinifera* NADP-ME (Grape-NADP-ME, L34836, Franke and Adams, 1995) and cytosolic *Oryza sativa* NADP-ME1-3 (Rice-CytME1-3, Chi et al., 2004); *Zea mays* cytosolic NADP-malic enzyme (ZmCyt-NADP-ME, Detarsio et al., 2008); the plastidic *Zea mays* NADP-ME isoforms ZmC4-NADP-ME (Rothermel and Nelson, 1989) and Zm-nonC4-NADP-ME (Saigo et al., 2004). Branch support values are shown in red.

Sanger sequencing of candidate genes and possible downstream targets

To confirm the polymorphisms detected before in the RNA-seq analysis and/or to identify undetected polymorphisms within possible seed longevity candidate or downstream-target genes, the two before mentioned NADP-ME genes were amplified for both CC and L94 and sequenced. In the case of the possible NADP-ME downstream target gene MLOC_35785.1, no single nucleotide polymorphisms (SNPs) were detected between the coding sequences (CDS) of the two parental lines. The same occurred for the possible candidate NADP-ME gene AK248526.1. However, three SNPs were detected within a 2 kb promoter fragment of the latter gene. The position of these SNPs is indicated in figure 6, which shows a schematic representation of the NADP-ME gene AK248526.1.
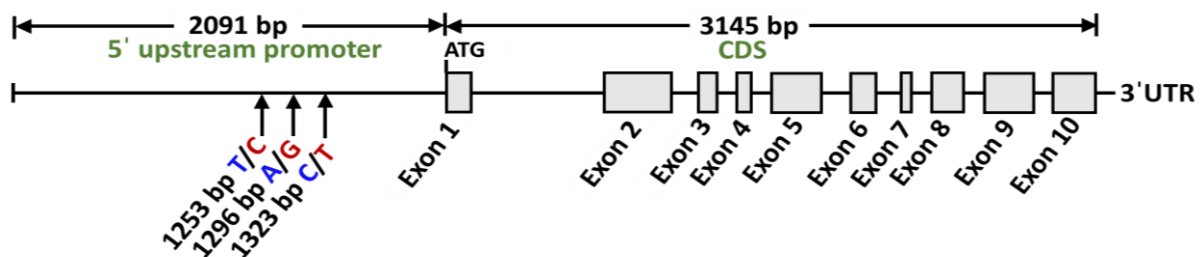
Figure 6. **Schematic presentation of the NADP-ME gene AK248526.1**

Exons are presented as grey boxes and the position of the ATG start codon of the coding sequence (CDS) is given. The three prime untranslated region (3'-UTR) is indicated. The length of the coding sequence and the in the cloning used five prime (5') upstream promoter are given in base pairs (bp). The position of single nucleotide polymorphisms (SNPs) in the promoter sequence are indicated, the respective nucleotides in the Cebada Capa and L94 sequences are given in blue and red, respectively.

In addition to the two before mentioned NADP-MEs, the UDP-glycosyltransferase MLOC_11661.1 was also chosen for further analysis after combining the results of the RNA-seq and the total seed proteome analysis. This UDP-glycosyltransferase was considered as another promising candidate gene due to its location in the seed longevity QTL on the bottom of chromosome 2 (SLQ2) and since it was found in high abundance in CC and some of the L94 NILs, but absent in L94. As shown in table 1, a number of SNPs were detected in the CDS of this gene. The majority of these SNPs are synonymous SNPs, which do not cause a change in the amino acid sequence. However, among the 11 identified SNPs, two are non-synonymous and lead to the in table 1 indicated amino acid change.

Table 1. Nucleotide sequences and amino acid residues of codons containing single nucleotide polymorphisms (SNPs) between Cebada Capa and L94 in the coding sequence of the UDP-glycosyltransferase gene MLOC_11661.1

SNPs in each codon are indicated as bold and underlined nucleotides. The position of the SNPs is given as number of the codon containing these SNPs. Non-synonymous SNPs, causing an amino acid change, are highlighted in grey color. Amino acids names are given in three letter code.

Position of SNP	85	154	179	257	261	272	291	316	317	421	425
Cebada Capa genotype	CG**T**	CC**T**	TC**C**	GG**C**	CG**C**	**C**TA	CA**C**	GG**C**	AA**C**	GT**T**	**G**CT
	Arg	Pro	Ser	Gly	Arg	Leu	His	Gly	Asn	Val	Ala
L94 genotype	CG**A**	CC**A**	TC**T**	GG**A**	CG**T**	**T**TA	CA**G**	GG**T**	AA**T**	GT**C**	**A**CT
	Arg	Pro	Ser	Gly	Arg	Leu	Gln	Gly	Asn	Val	Thr

Testing of homologous *Arabidopsis* T-DNA knock-out lines for their seed longevity phenotype

Similar to the approach of Nguyen (2014), homozygous *Arabidopsis thaliana* T-DNA knock-out lines, homologous to the before mentioned possible barley candidate and downstream target genes, were ordered and tested for their seed longevity phenotype. In the case of the NADP-MEs, the T-DNA knock-out line *nadp-me1* (SALK_036898C) was already shown to have a seed longevity phenotype, as *nadp-me1* was significantly less storable than the Col-0 wild-type (Nguyen, 2014). The barley NADP-ME AK248526.1 shows a 77% protein identity to the *Arabidopsis* NADP-ME1 (At2g19900), while the second NADP-ME MLOC_35785.1 shows 80% protein identity to this gene. Blasting the UDP-glycosyltransferase MLOC_11661.1 against the *Arabidopsis thaliana* genome, leads to the identification of the Galacturonosyltransferase 10 (GAUT10) (At2g20810) as homologue. The barley UDP-glycosyltransferase MLOC_11661.1 shows 70% protein identity to GAUT10. A T-DNA knock-out line of GAUT10 was ordered and tested for its seed longevity phenotype. As shown in figures 7a and 7b, both homologous *Arabidopsis* genes are expressed highly in seeds. Whereas At2g19900 (encoding NADP-ME1) is only highly expressed in late maturation stages of the seed and in dry seeds, At2g20810 (encoding GAUT10) is expressed in different plant organs but shows the highest expression in 24 hours imbibed seeds.
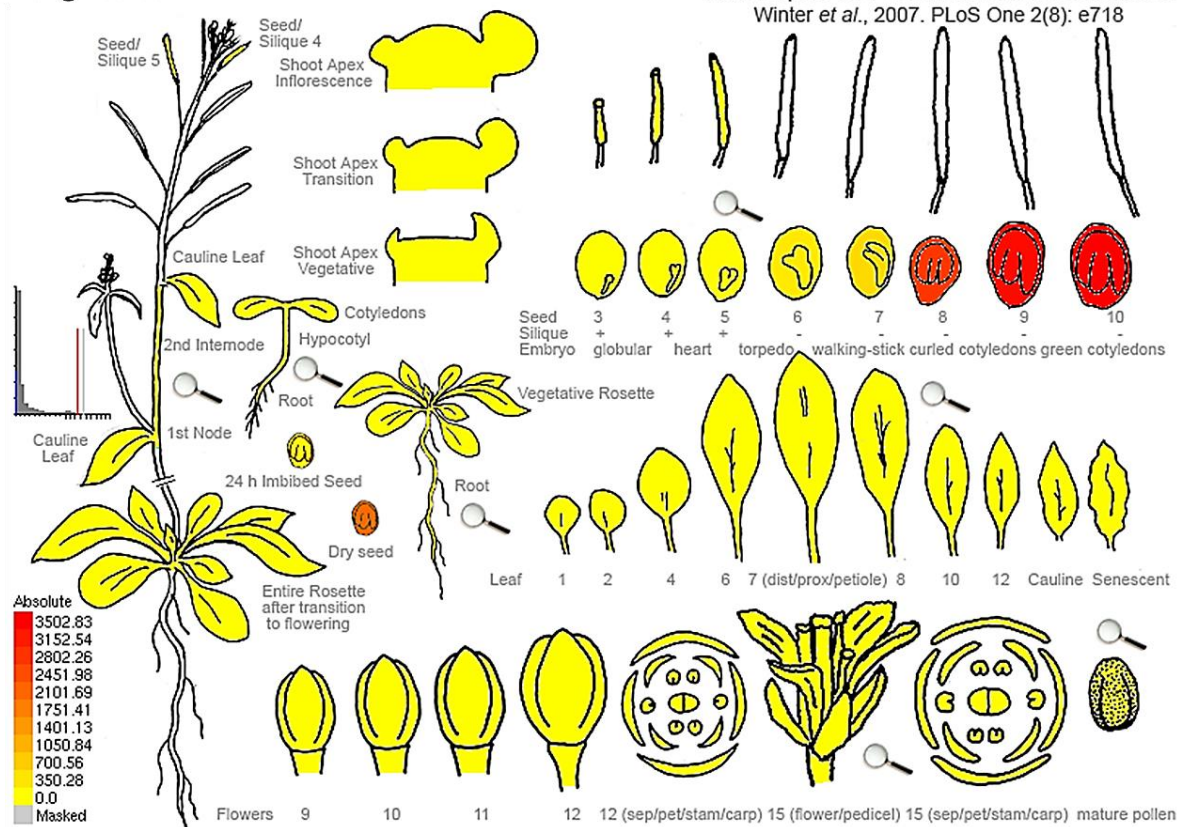
Figure 7a. **Expression profiles of the *Arabidopsis thaliana* gene At2g19900**

The Arabidopsis eFP browser (http://bar.utoronto.ca/) data visualization tool was used to explore in which plant organs the *Arabidopsis thaliana* gene At2g19900, encoding the NADP-dependent malic enzyme 1 (NADP-ME1), is expressed in. Expression values are indicated by a color code ranging from yellow (low expression) to red (high expression).
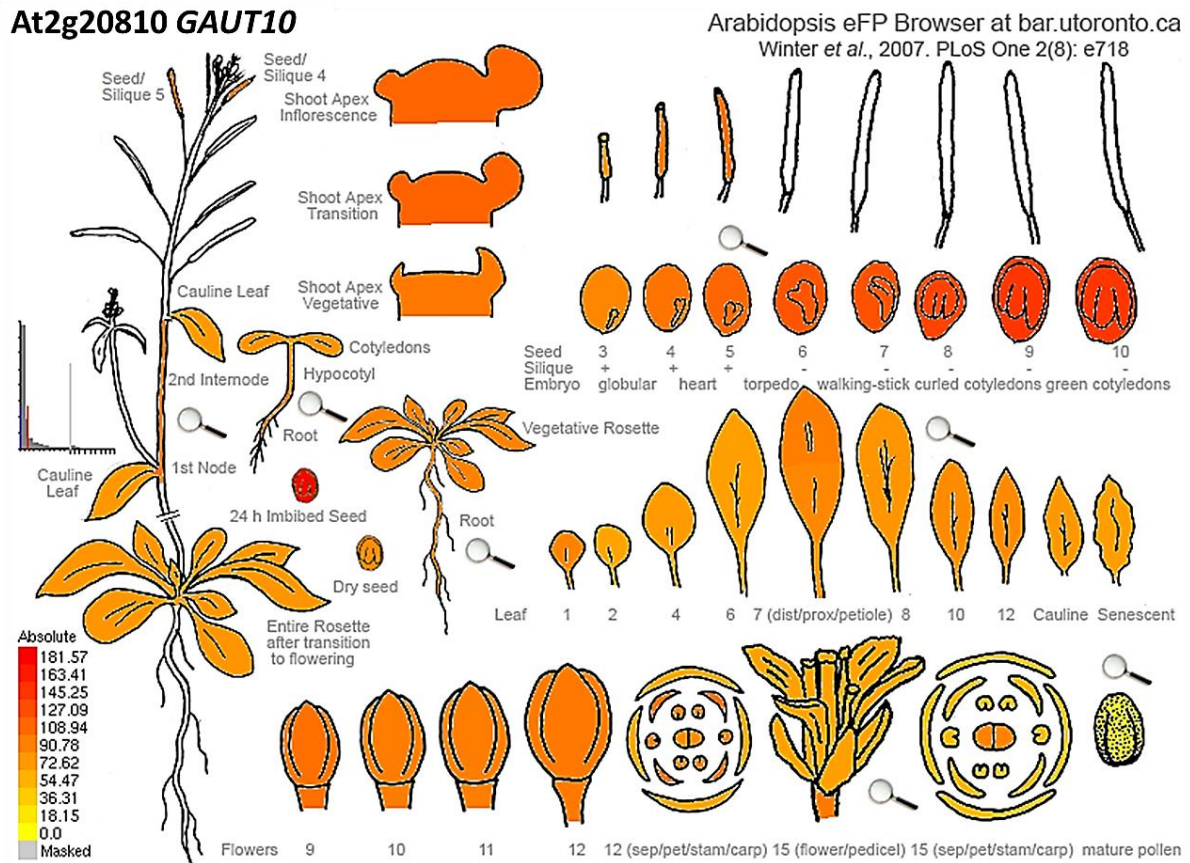
Figure 7b. **Expression profiles of the *Arabidopsis thaliana* gene At2g20810**

The Arabidopsis eFP browser (http://bar.utoronto.ca/) data visualization tool was used to explore in which plant organs the *Arabidopsis thaliana* gene At2g20810, encoding the Galacturonosyltransferase 10 (GAUT10), is expressed in. Expression values are indicated by a color code ranging from yellow (low expression) to red (high expression).

To test the T-DNA knock-out lines for their seed longevity phenotype, an artificial ageing treatment at 37˚C and 80% relative humidity was performed. The T-DNA knock-out line *nadp-me1* (SALK_036898C) and wild-type Col-0 were included in the analysis as controls. The mutant line *dog1* of the *DELAY OF GERMINATION 1* (*DOG1*) gene, identified as a quantitative trait locus involved in the control of seed dormancy, was also included in the analysis since it has been shown to have a seed longevity phenotype (Bentsink et al., 2006). As shown in figure 8, after four days of artificial treatment, *dog1* shows a significant decrease in seed longevity when compared to the wild-type Col-0. The other tested lines, including Col-0, show only a decrease in their germination ability of around 10%. After eight days of treatment, the germination ability of *dog1* decreases further to around 15%. As already reported by Nguyen (2014), *nadp-me1* shows a significant reduction in seed longevity after eight days of ageing. In contrast to *dog1* and *nadp-me1*, *gaut10* does not show a significant reduction in germination ability after 8 days of ageing when compared to the wild-type Col-0. After 12 days of ageing, all tested genotypes exhibit a low germination ability. Taking these observations into consideration, the ordered homozygous T-DNA knock-out line *gaut10* (SALK_029319C) does not show a significant seed longevity phenotype over the whole period of 12 days. However, the seed longevity phenotypes of *nadp-me1*

and *dog1*, reported by Nguyen (2014) and Bentsink et al. (2006), could be confirmed by these experiments.
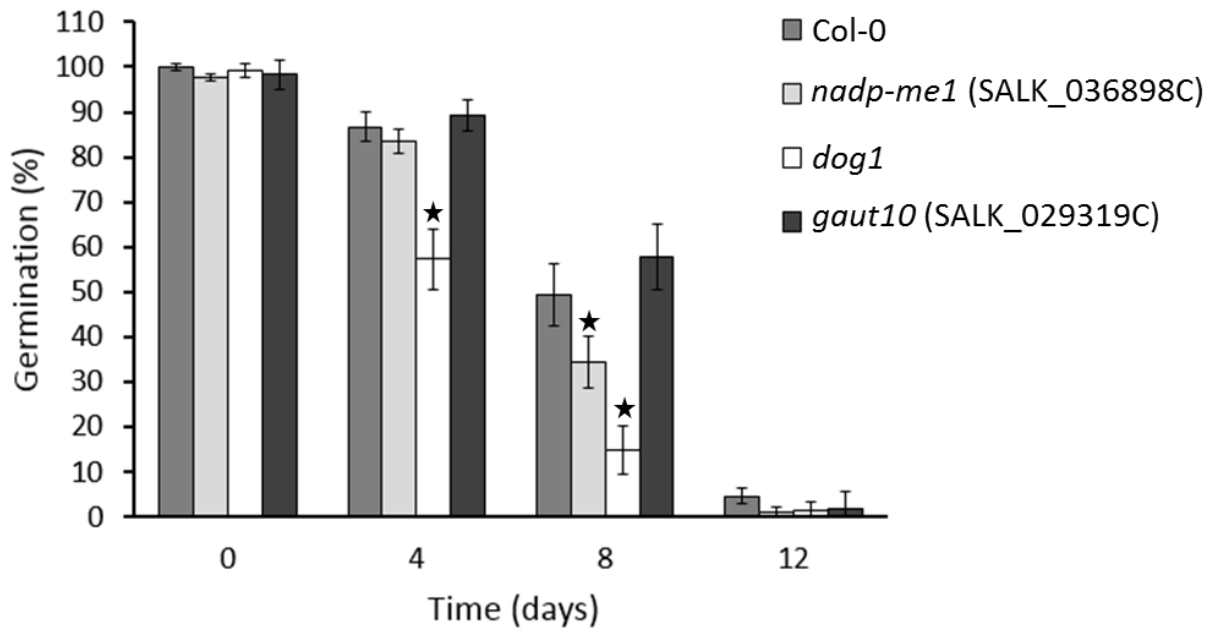
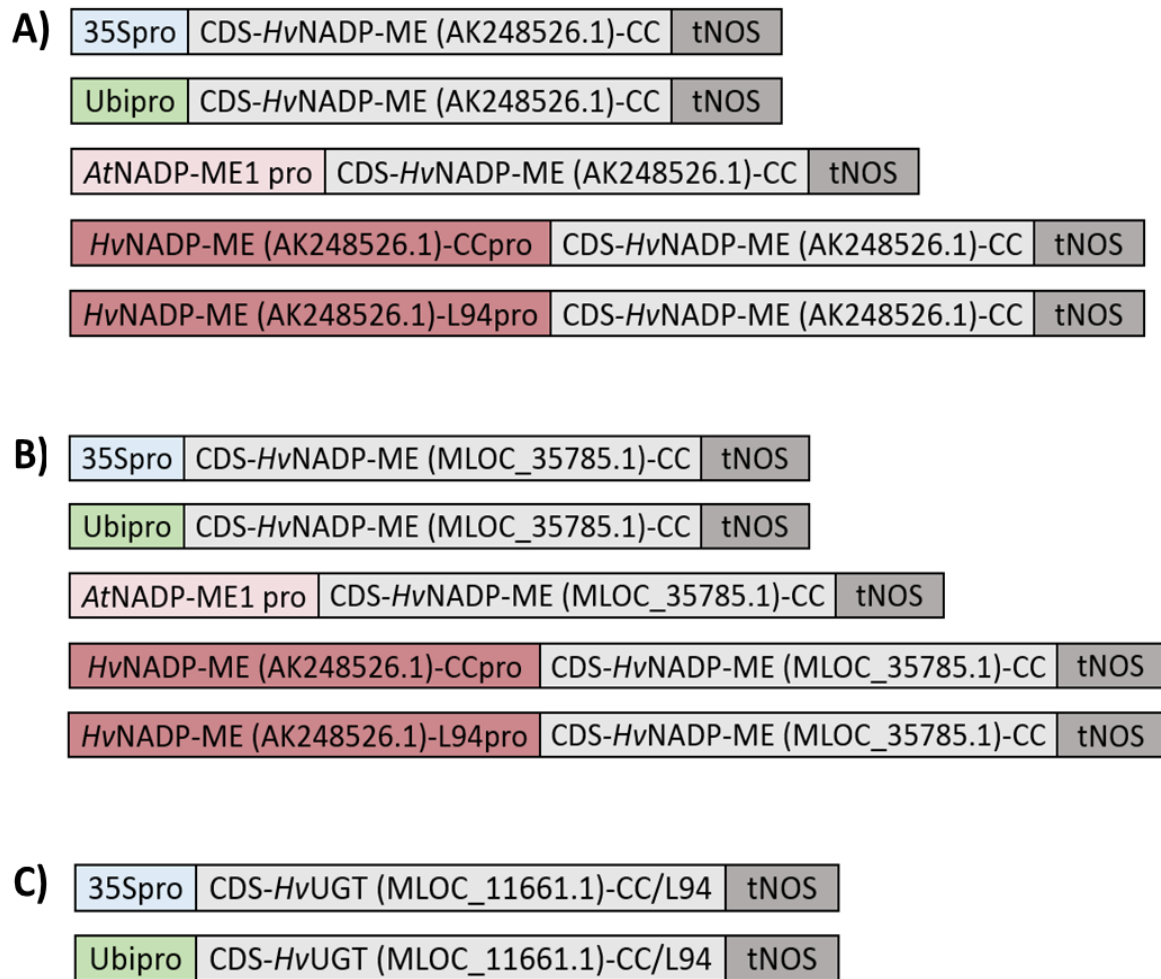**Figure 8**. **Seed longevity of T-DNA knock-out mutants after artificial ageing**

Seed longevity, presented as germination (%) of T-DNA knock-out mutants of candidate genes, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Stock order numbers of the T-DNA knock-out lines are given in brackets. The mutant line *dog1* was included in the tests as control along with the wild-type Col-0. Standard errors are calculated on four biological replicates. The asterisks indicate significant differences between mutants and Col-0 wild-type (P≤0.05).

Complementation of *nadp-me1* with barley alleles of candidate genes and downstream targets

To test the effect of the possible candidate genes AK248526.1 (NADP-ME) and MLOC_11661.1 (UDP-glycosyltransferase), and the possible downstream target MLOC_35785.1 (NADP-ME), the before described seed longevity mutant *nadp-me1* was complemented with the barley alleles and the resulting transgenic lines were tested for their seed longevity phenotype. Figure 9 summarizes all generated constructs. All three barley genes were expressed under the control of the strong constitutive Cauliflower Mosaic Virus (CaMV) 35S-promoter and the Ubiquitin (Ubi)-promoter. The two barley NADP-MEs were additionally expressed under the control of the *Arabidopsis thaliana* NADP-ME1 (At2g19900) promoter, and the Cebada Capa-specific and L94-specific promoters of the NADP-ME candidate gene AK248526.1. Since no SNPs were detected in the coding sequences (CDS) of the two barley NADP-MEs between the parental lines Cebada Capa and L94, only the Cebada Capa CDS was utilized. In the case of the barley UDP-glycosyltransferase MLOC_11661.1, the CDS sequences of both parental lines were cloned to test functional differences between the two allelic variants, which showed SNPs between the parental lines. The final expression constructs were obtained with a double Gateway reaction into the destination vector pEDO097pFR7m24GW containing the *Discosoma* sp. red fluorescent reporter protein DsRed as selection marker. All constructs were introduced into wild-type Col-0 and *nadp-me1* backgrounds by

floral dip *Agrobacterium*-mediated transformation. Transformed seeds expressing DsRed in T_1 and T_2 (first and second transformant generation) lines were selected by fluorescence microscopy in dry state. T_2 lines, showing a segregation ratio of 3:1, indicating the presence of a single functional transgenic locus, and high DsRed-fluorescence, were selected. Artificial ageing experiments were performed with fluorescent seeds from T_2 lines and non-fluorescent sibling seeds were used as negative control for each construct.

A)

| 35Spro | CDS-*Hv*NADP-ME (AK248526.1)-CC | tNOS |

| Ubipro | CDS-*Hv*NADP-ME (AK248526.1)-CC | tNOS |

| *At*NADP-ME1 pro | CDS-*Hv*NADP-ME (AK248526.1)-CC | tNOS |

| *Hv*NADP-ME (AK248526.1)-CCpro | CDS-*Hv*NADP-ME (AK248526.1)-CC | tNOS |

| *Hv*NADP-ME (AK248526.1)-L94pro | CDS-*Hv*NADP-ME (AK248526.1)-CC | tNOS |

B)

| 35Spro | CDS-*Hv*NADP-ME (MLOC_35785.1)-CC | tNOS |

| Ubipro | CDS-*Hv*NADP-ME (MLOC_35785.1)-CC | tNOS |

| *At*NADP-ME1 pro | CDS-*Hv*NADP-ME (MLOC_35785.1)-CC | tNOS |

| *Hv*NADP-ME (AK248526.1)-CCpro | CDS-*Hv*NADP-ME (MLOC_35785.1)-CC | tNOS |

| *Hv*NADP-ME (AK248526.1)-L94pro | CDS-*Hv*NADP-ME (MLOC_35785.1)-CC | tNOS |

C)

| 35Spro | CDS-*Hv*UGT (MLOC_11661.1)-CC/L94 | tNOS |

| Ubipro | CDS-*Hv*UGT (MLOC_11661.1)-CC/L94 | tNOS |

Figure 9. **Schematic representation of constructs used for the transformation of *nadp-me1* and Col-0 wild-type plants**

Promoters (pro), coding sequences (CDS) of Cebada Capa (CC) and L94 and the *Agrobacterium* Nopaline Synthase (NOS) terminators (tNOS) are schematically presented as boxes. Species specific abbreviations are as following: *Hv* =*Hordeum vulgare*, *At* = *Arabidopsis thaliana*. A) Constructs generated for the barley NADP-dependent malic enzyme (NADP-ME) candidate gene AK248526.1. B) Constructs generated for the barley NADP-ME downstream target gene MLOC_35785.1. C) Constructs generated for the barley UDP-glycosyltransferase (UGT) candidate gene MLOC_11661.1.

As shown in figure 10, the NADP-ME candidate gene AK248526.1 is neither able to rescue the seed longevity phenotype of *nadp-me1* under the control of the 35S-promoter, nor the Ubi-promoter. The germination abilities of fluorescent seeds and non-fluorescent sibling seeds after 4 days of artificial

ageing treatment are comparable to those shown in figure 8. After 8 days, the germination ability of fluorescent and non-fluorescent seeds from transformed *nadp-me1* plants is significantly lower than the one of the respective Col-0 wild-type seeds. After 12 days, the germination abilities of all in figure 10 presented transgenic lines is below 10% and no significant difference between *nadp-me1* and Col-0 transgenic lines can be observed.
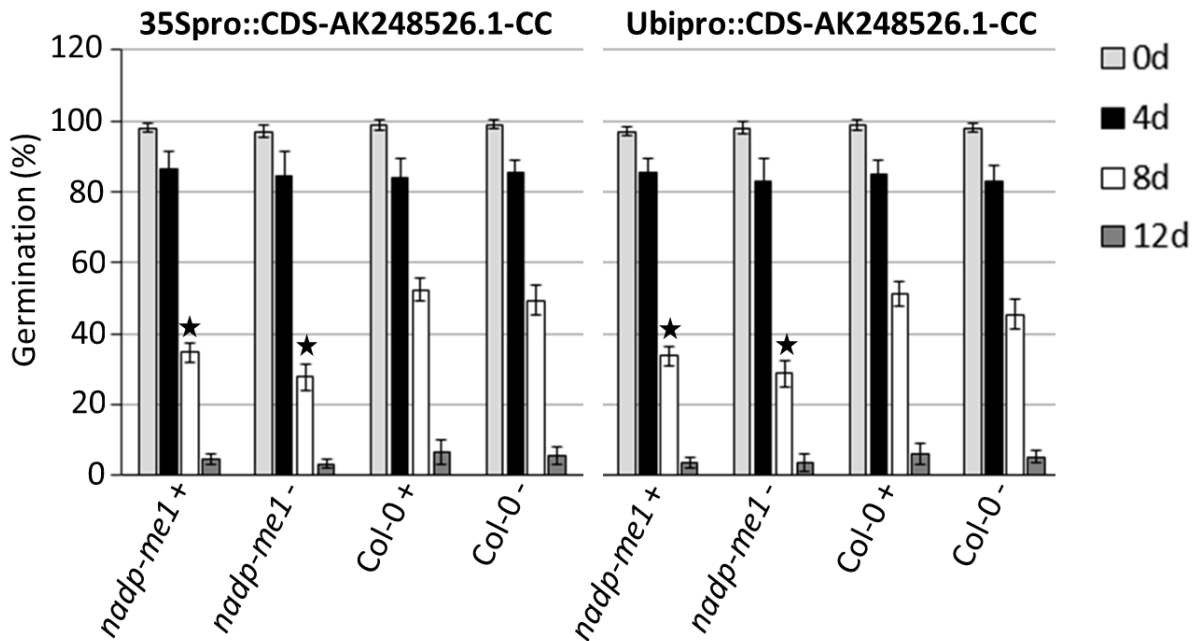
Figure 10. **Seed longevity of transgenic lines, expressing the barley gene AK248526.1 under the control of the 35S- and Ubi-promoter, after artificial ageing**

Col-0 wild-type and *nadp-me1* (SALK_036898C) were transformed with constructs expressing the Cebada Capa (CC) coding sequence (CDS) of the NADP-dependent malic enzyme AK248526.1 under the control of the Cauliflower mosaic virus (CaMV) 35S-promoter (35Spro) and the Ubiquitin-promoter (Ubipro). For the selection of positive transformants, the red fluorescent reporter protein DsRed, integrated in the final Gateway® destination vector pEDO097pFR7m24GW, was used. Seed longevity, presented as germination (%) of T_2 (second generation of transformation) lines, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Non-fluorescent sibling seeds were used as controls. Lines with fluorescent seeds are indicated with plus and those with non-fluorescent seeds with a minus. Standard errors are calculated on four biological replicates. Three lines per construct were tested. The asterisks indicate significant differences between transformed *nadp-me1* and Col-0 wild-type plants (P≤0.05).

As mentioned before, the NADP-ME candidate gene AK248526.1 was additionally expressed under the control of the *Arabidopsis thaliana* NADP-ME1 (At2g19900) promoter, and the Cebada Capa-specific and L94-specific promoters of the candidate gene itself. As shown in figure 11, the germination abilities are comparable to the in figure 10 presented transgenic lines. Again after 8 days of artificial ageing treatment, the germination ability of transformed *nadp-me1* lines was significantly lower than the one of Col-0 plants transformed with the same constructs.
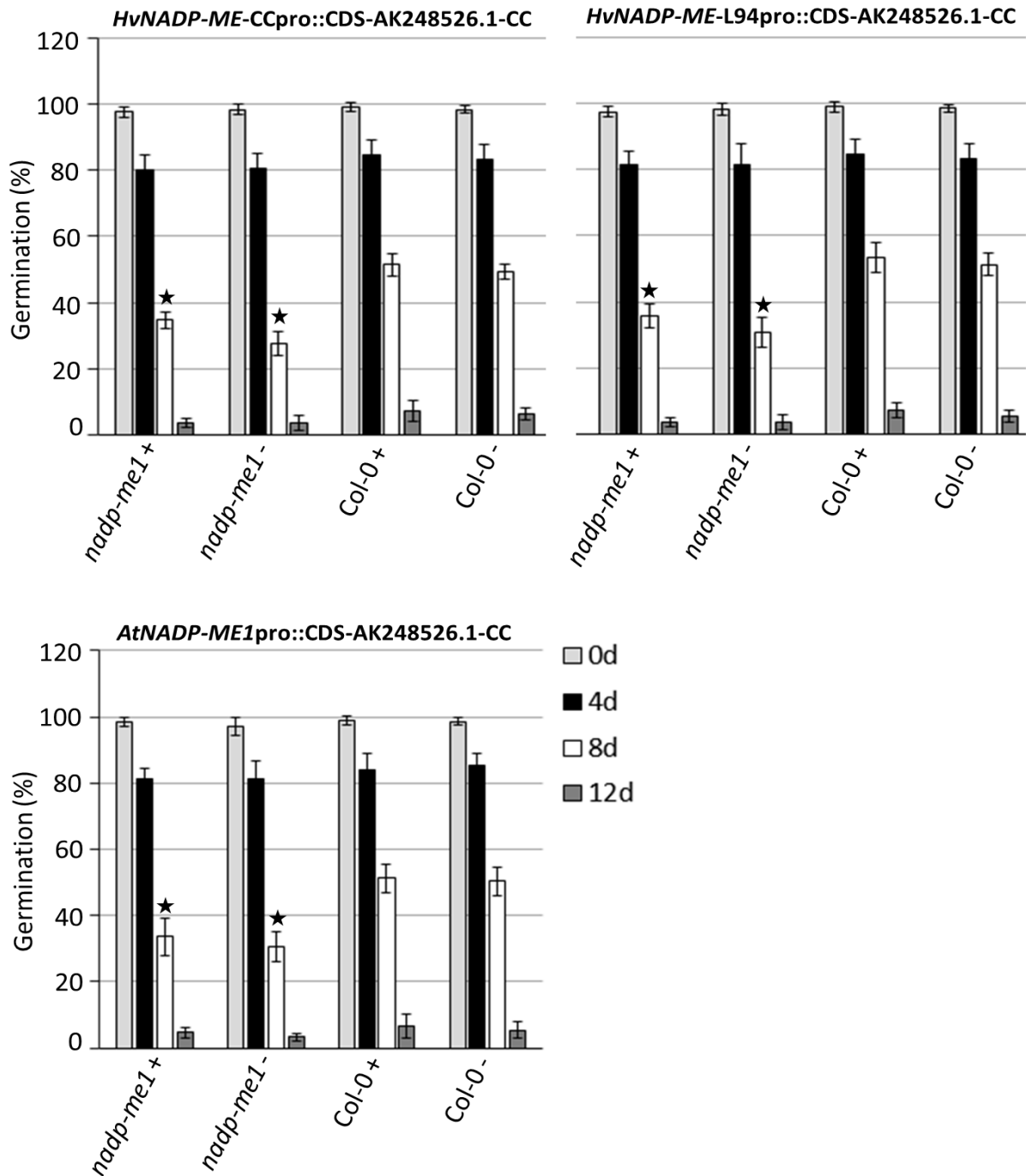
Figure 11. **Seed longevity of transgenic lines, expressing the barley gene AK248526.1 under the control of the *Arabidopsis thaliana* NADP-ME1 promoter and the Cebada Capa- and L94-specific AK248526.1 promoters, after artificial ageing**

Col-0 wild-type and *nadp-me1* (SALK_036898C) were transformed with constructs expressing the Cebada Capa (CC) coding sequence (CDS) of the NADP-dependent malic enzyme AK248526.1 under the control of *Arabidopsis thaliana* NADP-ME1 (At2g19900) promoter and the CC- and L94-specific AK248526.1 promoters. Species specific abbreviations are as follows: *Hv* = *Hordeum vulgare*, *At* = *Arabidopsis thaliana*. For the selection of positive transformants, the red fluorescent reporter protein DsRed, integrated in the final Gateway® destination vector pEDO097pFR7m24GW, was used. Seed longevity, presented as germination (%) of T_2 (second generation of transformation) lines, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Non-fluorescent sibling seeds were used as controls. Lines with fluorescent seeds are indicated with plus and those with non-fluorescent seeds with a minus. Standard errors are calculated on four biological replicates. Three lines per construct were tested. The asterisks indicate significant differences between transformed *nadp-me1* and Col-0 wild-type lines (P≤0.05).

In contrast to the NADP-ME AK248526.1, selected as a possible candidate gene, the second NADP-ME MLOC_35785.1, identified as possible downstream-target, rescues the seed longevity phenotype of *nadp-me1*, both under the control of the 35S- and the Ubi-promoter. As shown in figure 12, significant differences can be observed between fluorescent and non-fluorescent sibling seeds of transformed *nadp-me1* plants after 8 and 12 days of artificial ageing treatment. Interestingly, *nadp-me1* lines expressing the MLOC_35785.1 gene under control of the Ubi-promoter show a germination ability after 12 days, which is even higher than the one of Col-0 lines expressing the same construct.
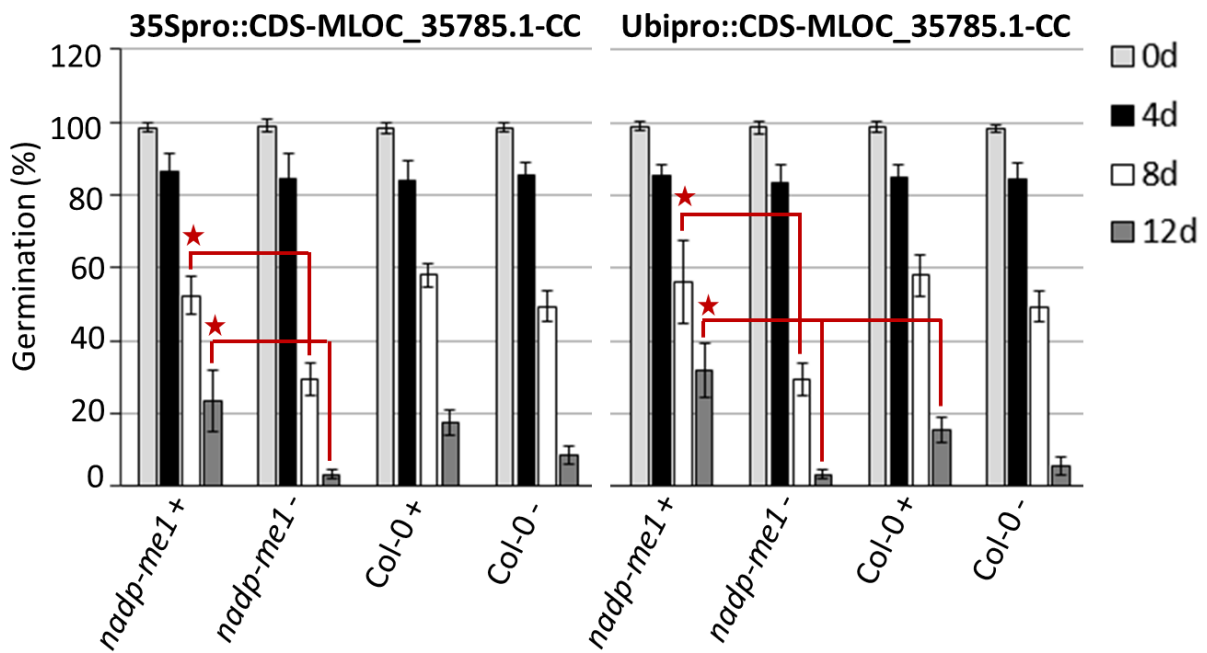
Figure 12. **Seed longevity of transgenic lines, expressing the barley gene MLOC_35785.1 under the control of the 35S- and Ubi-promoter, after artificial ageing**

Col-0 wild-type and *nadp-me1* (SALK_036898C) were transformed with constructs expressing the Cebada Capa (CC) coding sequence (CDS) of the NADP-dependent malic enzyme MLOC_35785.1 under the control of the Cauliflower mosaic virus (CaMV) 35S-promoter (35Spro) and the Ubiquitin-promoter (Ubipro). For the selection of positive transformants, the red fluorescent reporter protein DsRed, integrated in the final Gateway® destination vector pEDO097pFR7m24GW, was used. Seed longevity, presented as germination (%) of T_2 (second generation of transformation) lines, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Non-fluorescent sibling seeds were used as controls. Lines with fluorescent seeds are indicated with plus and those with non-fluorescent seeds with a minus. Standard errors are calculated on four biological replicates. Three lines per construct were tested. Significant differences between transgenic *nadp-me1* and Col-0 lines (P≤0.05) are indicated by red asterisks.

Comparable to the results observed for the other barley NADP-ME, the seed longevity phenotype of *nadp-me1* can neither be rescued when expressing the possible downstream-target gene MLOC_35785.1 under the control of the *Arabidopsis thaliana* NADP-ME1 (At2g19900) promoter, nor the Cebada Capa-specific and L94-specific promoters of AK248526.1. The results shown in figure 13, are similar to the ones shown in figure 11 for the NADP-ME AK248526.1.
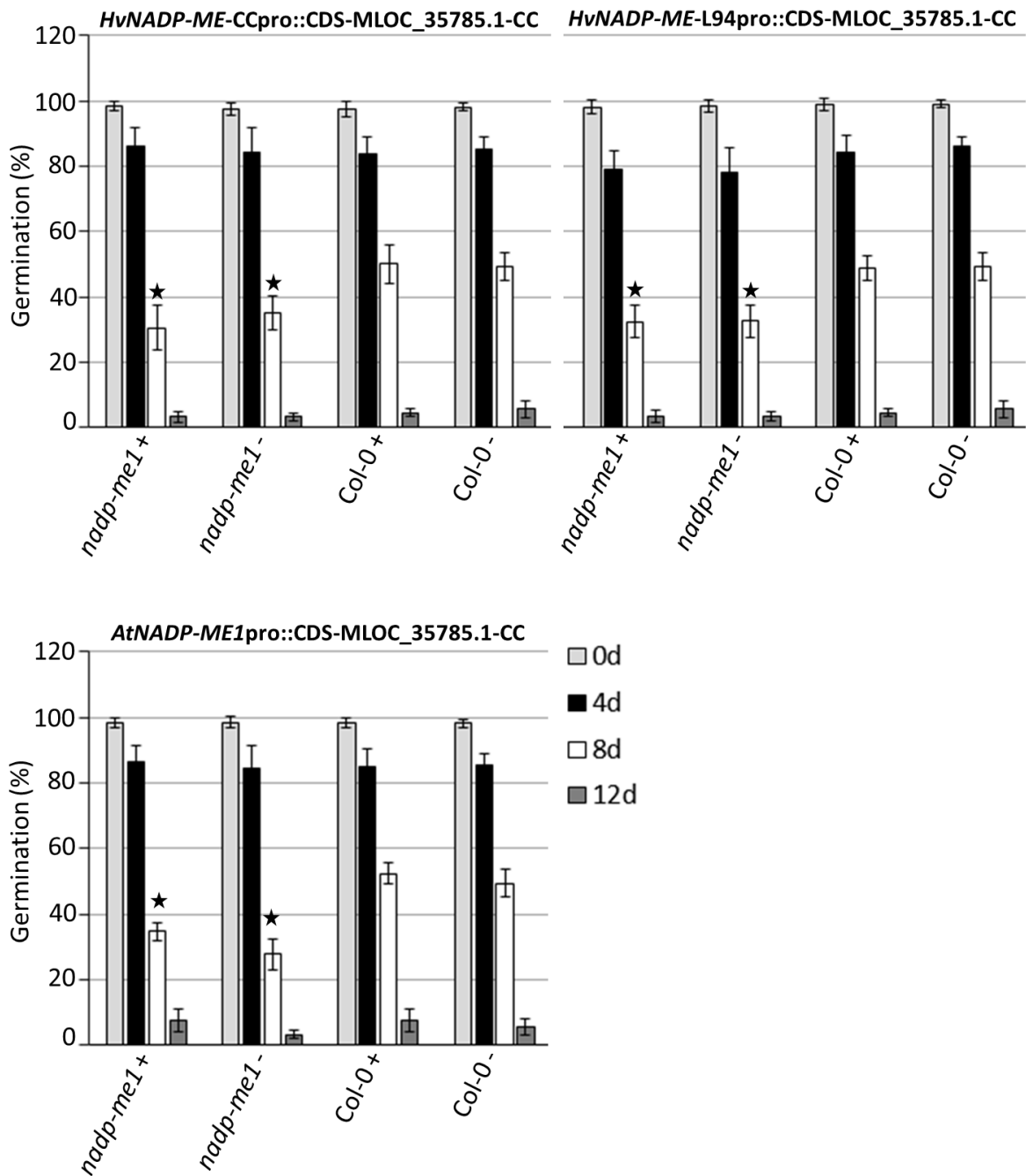
Figure 13. **Seed longevity of transgenic lines, expressing the barley gene MLOC_35785.1 under the control of the *Arabidopsis thaliana* NADP-ME1 promoter and the Cebada Capa- and L94-specific AK248526.1 promoters, after artificial ageing**

Col-0 wild-type and *nadp-me1* (SALK_036898C) were transformed with constructs expressing the Cebada Capa (CC) coding sequence (CDS) of the NADP-dependent malic enzyme MLOC_35785.1 under the control of *Arabidopsis thaliana* NADP-ME1 (At2g19900) promoter and the CC- and L94-specific AK248526.1 promoters. Species specific abbreviations are as follows: *Hv* =*Hordeum vulgare*, *At* = *Arabidopsis thaliana*. For the selection of positive transformants, the red fluorescent reporter protein DsRed, integrated in the final Gateway® destination vector pEDO097pFR7m24GW, was used. Seed longevity, presented as germination (%) of T_2 (second generation of transformation) lines, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Non-fluorescent sibling seeds were used as controls. Lines with fluorescent seeds are indicated with plus and those with non-fluorescent seeds with a minus. Standard errors are calculated on four biological replicates. Three lines per construct were tested. The asterisks indicate significant differences between transformed *nadp-me1* and Col-0 wild-type lines (P≤0.05).

In addition to the two barley NADP-MEs, the role of the UDP-glycosyltransferase MLOC_11661.1 in seed longevity was tested. As shown in figure 14, the seed longevity phenotype of *nadp-me1* is rescued when expressing the CC-specific coding sequence of this gene under the control of the 35S- and the Ubi-promoter. In contrast to the NADP-ME MLOC_35785.1, the germination ability of *nadp-me1* lines expressing the MLOC_11661.1 gene under control of the Ubi-promoter is not higher than the one of Col-0 lines expressing the same construct after 12 days of artificial ageing treatment.
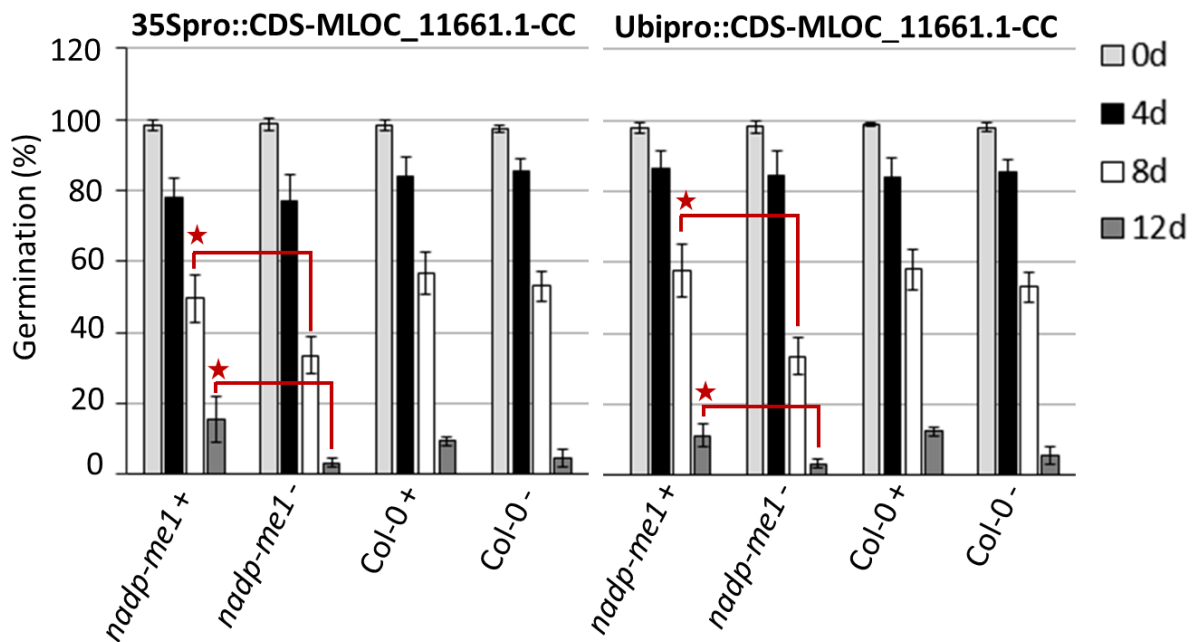
Figure 14. **Seed longevity of transgenic lines, expressing the barley gene MLOC_11661.1 under the control of the 35S- and Ubi-promoter, after artificial ageing**

Col-0 wild-type and *nadp-me1* (SALK_036898C) were transformed with constructs expressing the Cebada Capa (CC) coding sequence (CDS) of the NADP-dependent malic enzyme MLOC_11661.1 under the control of the Cauliflower mosaic virus (CaMV) 35S-promoter (35Spro) and the Ubiquitin-promoter (Ubipro). For the selection of positive transformants, the red fluorescent reporter protein DsRed, integrated in the final Gateway® destination vector pEDO097pFR7m24GW, was used. Seed longevity, presented as germination (%) of T_2 (second generation of transformation) lines, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Non-fluorescent sibling seeds were used as controls. Lines with fluorescent seeds are indicated with plus and those with non-fluorescent seeds with a minus. Standard errors are calculated on four biological replicates. Three lines per construct were tested. Significant differences between transgenic lines ($P \leq 0.05$) are indicated by red asterisks.

Interestingly, the expression of the L94-specific coding sequence under the control of the same promoters does not lead to a complementation of the *nadp-me1* seed longevity phenotype. As shown in figure 15, the germination ability of both fluorescent and non-fluorescent transgenic seeds in *nadp-me1* background exhibit a significantly lower germination rate than the respective seeds of in Col-0 background after 8 days of artificial ageing. This finding emphasizes the effect of the allele of the long-lived parental line Cebada Capa in seed longevity. Summarizing the results of the complementation tests, it can be seen that both the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 are able to rescue the *nadp-me1* seed longevity phenotype

when being expressed under the control of the 35S- or the Ubi-promoter. In the latter case only the CC-specific coding sequence has an effect, while the L94-specific coding sequence is not able to complement the *nadp-me1*, even under the control of strong promoters like the 35S- or Ubi-promoter.
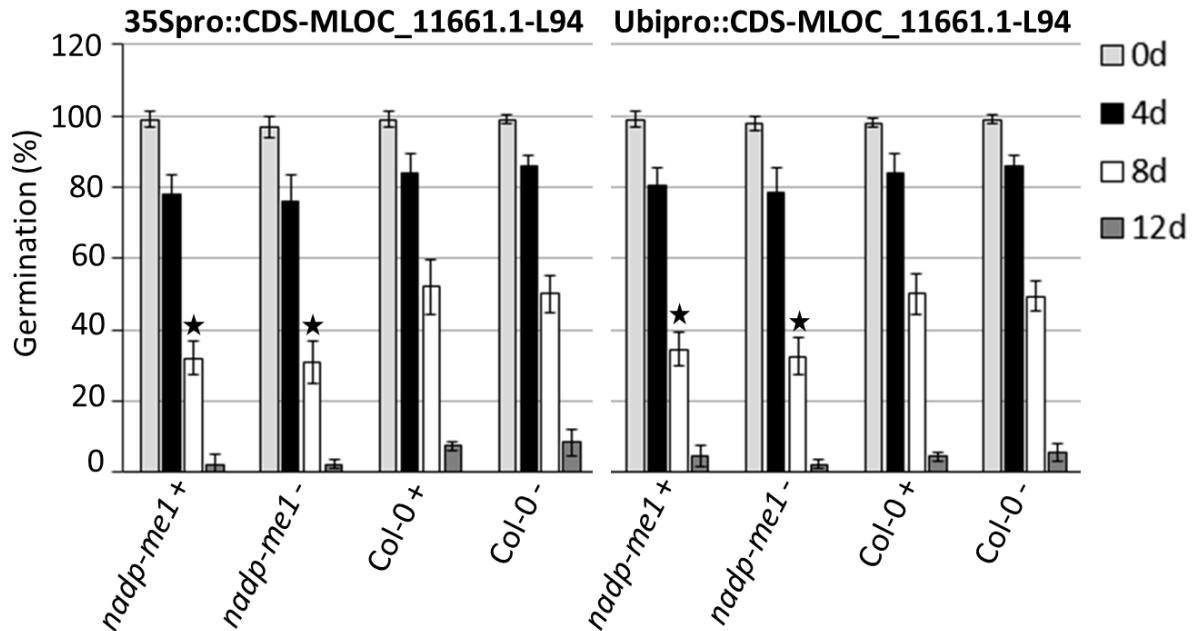

Figure 15. **Seed longevity of transgenic lines, expressing the barley gene MLOC_11661.1 under the control of the 35S- and Ubi-promoter, after artificial ageing**

Col-0 wild-type and *nadp-me1* (SALK_036898C) were transformed with constructs expressing the L94 coding sequence (CDS) of the NADP-dependent malic enzyme MLOC_11661.1 under the control of the Cauliflower mosaic virus (CaMV) 35S-promoter and the Ubiquitin-promoter (Ubi). For the selection of positive transformants, the red fluorescent reporter protein DsRed, integrated in the final Gateway® destination vector pEDO097pFR7m24GW, was used. Seed longevity, presented as germination (%) of T_2 (second generation of transformation) lines, was measured after 0, 4, 8 and 12 days of artificial ageing treatment. Non-fluorescent sibling seeds were used as controls. Lines with fluorescent seeds are indicated with plus and those with non-fluorescent seeds with a minus. Standard errors are calculated on four biological replicates. Three lines per construct were tested. The asterisks indicate significant differences between transformed *nadp-me1* and Col-0 wild-type lines (P≤0.05).

Apart from the validation of the barley NADP-MEs and the UDP-glycosyltransferase candidate and/or downstream target genes, another goal was to generate *nadp-me1* x roGFP2 double mutants to measure the NADP$^+$/NADPH redox potential in mature *nadp-me1* seeds. By this, it could be investigated whether the NADP-ME1 affects the redox potential in seeds. For this purpose, the *nadp-me1* line was either crossed with transgenic lines expressing roGFP2 constitutively (35S::roGFP2) or GRX1-roGFP2 in the cytosol (c-GRX1-roGFP2). Additionally, the *nadp-me1* line was crossed with double mutants between the transparent testa mutant *tt4* and roGFP2 expressing lines.

Seeds of the *tt4* mutant exhibit an altered fluorescence spectrum and high auto-fluorescence

In order to gain knowledge about the excitation behavior of these transgenic lines, excitation spectra of mature seeds of the wild-type Col-0, *tt4*, c-GRX1-roGFP2 and tt4 x 35S::roGFP2 were measured with a scanning fluorescent plate reader. The excitation spectra obtained after three different treatments are shown in figure 16. Seeds in solutions of either 100 µl sterile dH$_2$O, 10 mM of the reduction reagent 1,4-dithiothreitol (DTT) or 5 mM of the oxidation reagent 2,2'-dipyridyldisulfide (DPS) were transferred to the wells of a 96-well microtiter plate. Fluorescence was measured in a scanning fluorescent plate reader in a wavelength range from 360 to 496 nm. As shown in figure 16, seeds of *tt4* and *tt4* x 35S::roGFP2 showed an altered fluorescence spectrum and high auto-fluorescence in all three treatments. No difference in behavior between *tt4* and the *tt4* double mutant with the redox sensor roGFP2 is noticeable. In contrast, the fluorescence spectra of Col-0 and c-GRX1-roGFP2, in Col-0 background, are significantly different. Col-0 shows very low fluorescence in all three treatments while c-GRX1-roGFP2 shows a clear, for roGFP2 characteristic, peak at 488 nm. The absence of this peak in the spectra of *tt4* x 35S::roGFP2 suggests that the sensor signal of the 35S::roGFP2 line might not be strong enough.
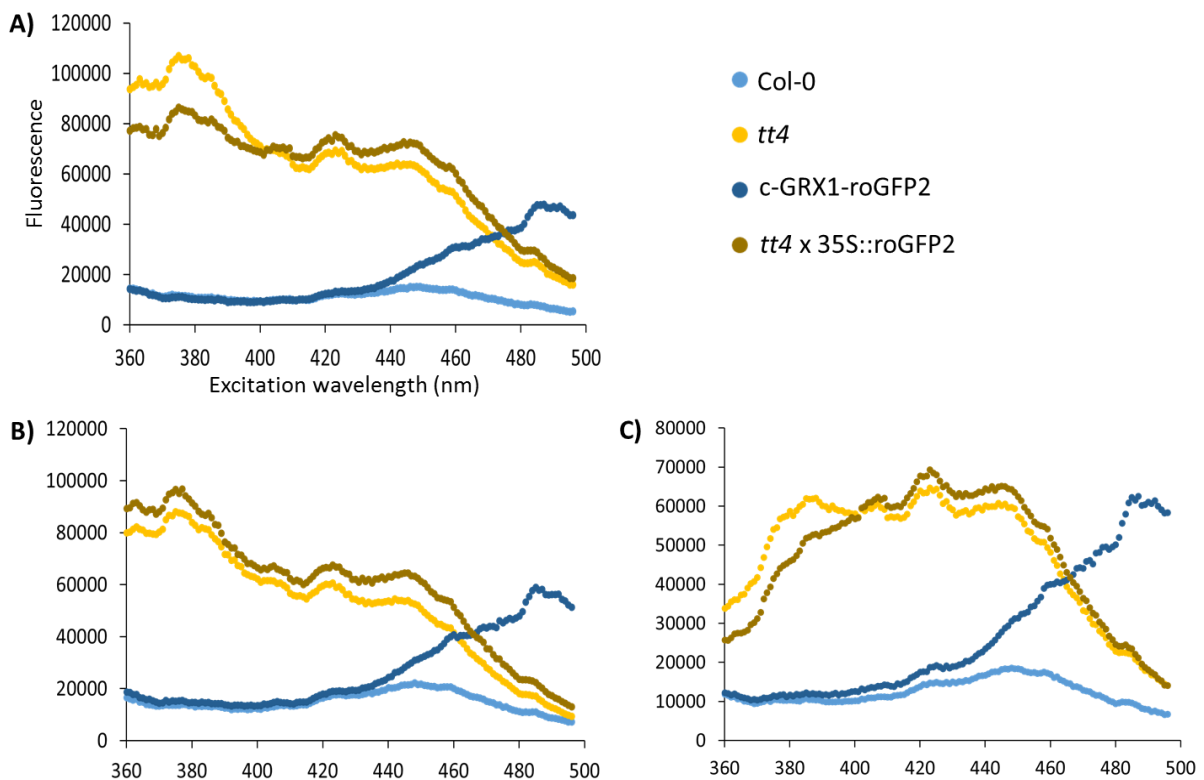

Figure 16. **Excitation spectra of mature seeds of Col-0, *tt4*, c-GRX1-roGFP2 and *tt4* x 35S::roGFP2 measured with a scanning fluorescent plate reader**

Excitation spectra of mature seeds of the transparent test mutant *tt4*, transgenic *Arabidopsis* plants overexpressing roGFP2 in the cytosol, c-GRX1-roGFP2, and a double mutant between *tt4* and an *Arabidopsis* line overexpressing roGFP2 under the control of the 35S-promoter (*tt4* x 35S::roGFP2) were measured with a scanning fluorescent plate reader. Fluorescence was measured in a wavelength range from 360 nm to 496 nm after the addition of 100 µl of A) sterile dH$_2$O, B) 10 mM 1,4-dithiothreitol (DTT) and C) 5 mM 2,2'-dipyridyldisulfide (DPS) to samples of approximately 50 seeds. Three replicates per genotype and treatment were conducted.

Since c-GRX1-roGFP2 exhibits a strong fluorescence signal, the *nadp-me1* mutant line was crossed to this line in order to obtain lines for the measurement of the redox potential of the NADP$^+$/NADPH couple in mature *nadp-me1* seeds. The original idea of producing a *tt4* x roGFP2 x *nadp-me1* triple mutant to facilitate the visualization by measuring the redox potential in transparent seeds was discarded due to the high auto-fluorescence of *tt4* seeds. The high auto-fluorescence of *tt4* and *tt4* x 35S::roGFP2 seeds was also clearly visible in horizontal section images of mature seeds obtained with a confocal laser scanning microscope, as shown in figure 17. In contrast, the auto-fluorescence level of mature Col-0 and c-GRX1-roGFP2 seeds was very low. As shown in figure 17, the fluorescence of c-GRX1-roGFP2 seeds at 488 nm is significantly stronger than the one of the other genotypes. The fluorescence signal is still strong when merging the images obtained at a wavelength 405 nm and 488 nm.
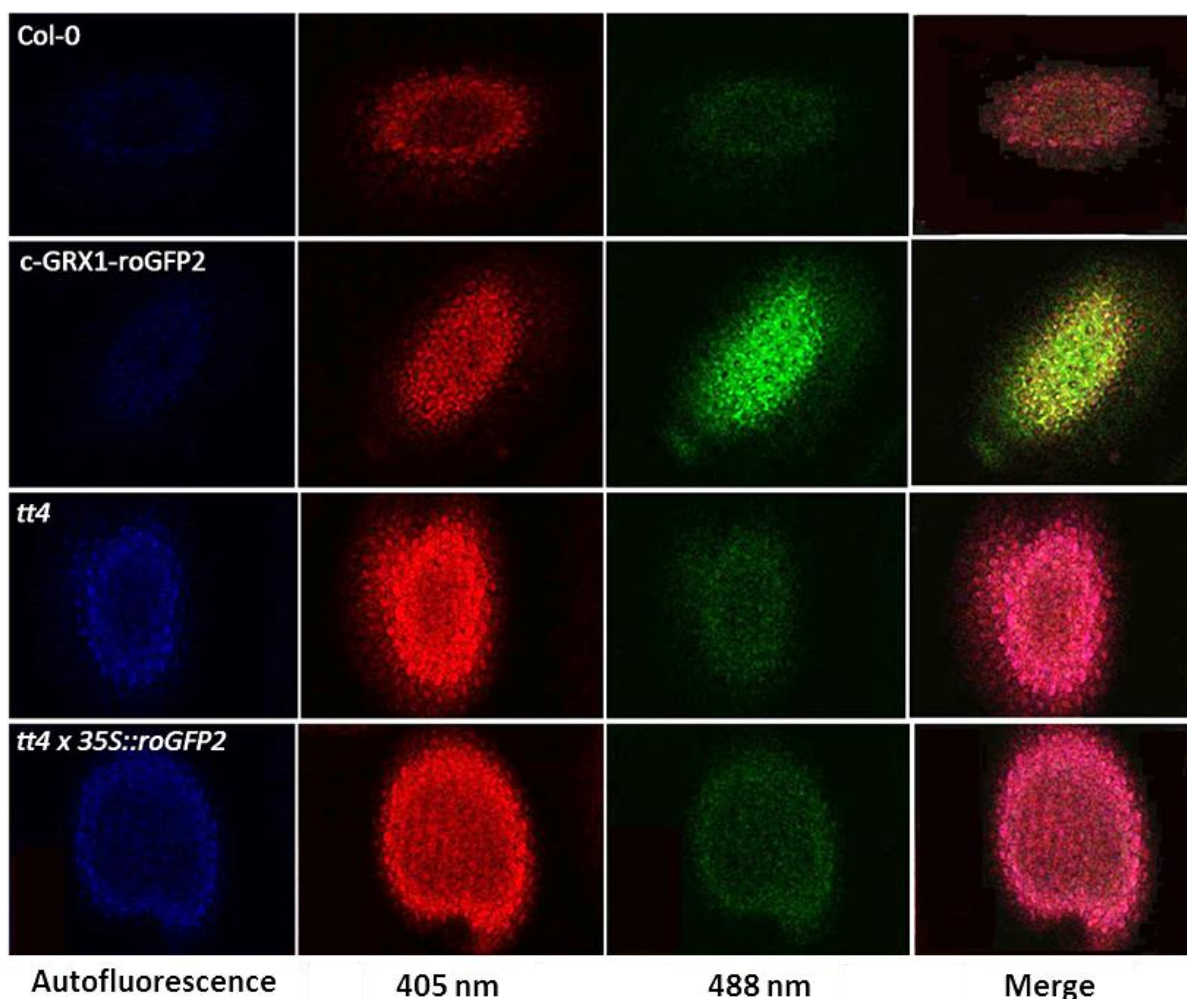
Figure 17. **Confocal images of mature seeds of Col-0, *tt4*, c-GRX1-roGFP2 and *tt4* x 35S::roGFP2 excited at 405 nm and 488 nm**

Horizontal section images of mature seeds of Col-0, *tt4*, c-GRX1-roGFP2 and *tt4* x 35S::roGFP2, placed between microscope slides, were acquired with a confocal laser scanning microscope after excitation at 405 nm and 488 nm. Images obtained at a wavelength of 405 nm and 488 nm were merged.

Taken these observations into consideration, the line c-GRX1-roGFP2 would be suitable to visualize the redox potential in mature seeds. For the selection of fluorescent seeds in the dry state, which could later be used for the measurement of redox potentials by confocal laser scanning microscopy as described by Schwarzländer et al. (2008), the DsRed reporter protein could be utilized. For this purpose, c-GRX1-roGFP2 transgenic lines were generated which express the DsRed reporter protein. The GRX1-roGFP2 gene fragment was amplified from the pBinCAM binary vector carrying the GRX1-roGFP2 construct and transferred with a double Gateway® reaction into the destination vector pEDO097pFR7m24GW containing the reporter protein DsRed. The GRX1-roGFP2 gene fragment was expressed under the control of an Ubi-promoter, as before in the pBinCAM binary vector. Such transgenic lines were generated both in wild-type Col-0 and *nadp-me1* background. The transgenic lines in wild-type Col-0 background could be used as controls. Transformed seeds expressing DsRed in T_1 lines were selected by fluorescence microscopy in dry state and were germinated for three days. The strength and the localization of the GRX1-roGFP2 reporter signal was tested in the hypocotyls of 3-day-old T_1 Col-0 and *nadp-me1* seedlings, expressing Ubi::GRX1-roGFP2, with a confocal laser scanning microscope. As shown in figure 18, the GRX1-roGFP2 reporter exhibits a strong signal in the cytosol and nuclei when excited at 488 nm.

| Bright field | 405 nm | 488 nm | Bright field | 405 nm | 488 nm |
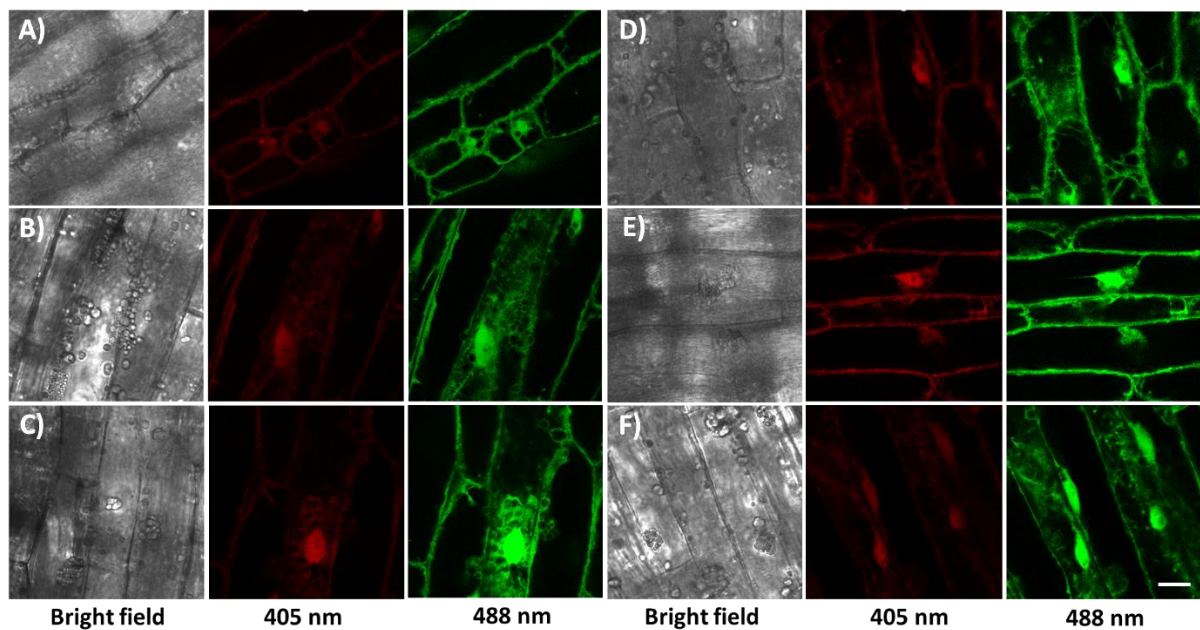
Figure 18. **Confocal images of hypocotyls of 3-day-old Col-0 and *nadp-me1* seedlings expressing Ubi::GRX1-roGFP2**

Horizontal section images of hypocotyls of 3-day-old seedlings, placed between microscope slides, were acquired with a confocal laser scanning microscope after excitation at 405 nm and 488 nm. Bright field images are displayed in the left panels. Scale bar = 0.1mm. A) - C): Images of hypocotyls of 3-day-old Col-0 seedlings expressing Ubi::GRX1-roGFP2; D) - F): Images of hypocotyls of 3-day-old *nadp-me1* seedling expressing Ubi::GRX1-roGFP2.

Fluorescent and non-fluorescent sibling seeds, selected in T_1 lines using a DsRed filter, of the before mentioned transgenic lines were additionally checked under a fluorescence stereo microscope. As shown in figure 19, the fluorescent seeds can clearly be separated from their non-fluorescent sibling seeds and wild-type Col-0 seeds using a DsRed filter. Slight differences in fluorescence among the fluorescent seeds can be observed. Using a GFP filter, the difference in fluorescence between fluorescent and non-fluorescent sibling seeds, selected using a DsRed filter, is less clear. However, some seeds in the selected fluorescent seeds exhibit a high fluorescence. Some of these seeds are highlighted exemplarily by white arrows in figure 19c.
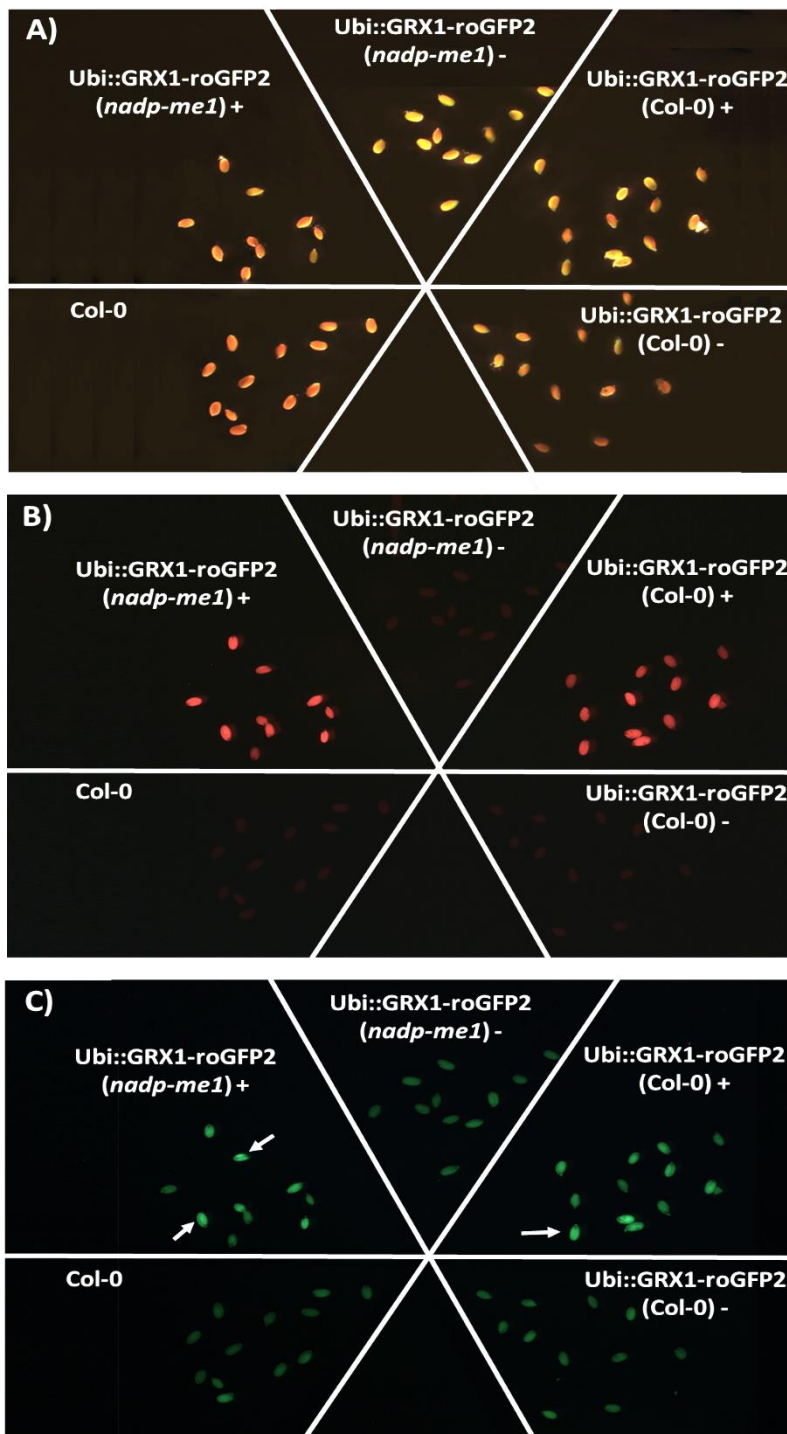
Figure 19. **Fluorescence stereo microscope images of Col-0 seeds and seeds of transgenic lines expressing the Ubi::GRX1-roGFP2 construct**

Images of seeds of Col-0 and of T_1 Col-0 and *nadp-me1* lines expressing the Ubi::GRX1-roGFP2 construct were taken A) with white light illumination, B) with a DsRed filter and C) with a GFP filter. Fluorescent and non-fluorescent sibling seeds of the transgenic lines are indicated by a plus or minus, respectively. Seeds showing a high fluorescence, using the GFP filter, are highlighted by arrows.

Taking the figures 18 and 19 into consideration, it can be said that some T_1 lines exhibit a strong fluorescent signal and could therefore be used for further studies to investigate whether the NADP-ME1 affects the redox potential in seeds. T_1 seeds showing strong fluorescence, both under a DsRed and GFP filter, were selected and grown for future studies in the greenhouse.

Discussion

In this chapter, the role of two NADP-MEs and one UDP-glycosyltransferase in seed longevity was analyzed. The NADP-ME AK248526.1 and the UDP-glycosyltransferase MLOC_116611.1 have been chosen as possible candidate genes for SLQ1.1 and SLQ2, respectively, while the second NADP-ME MLOC_35785.1 has been chosen as a possible downstream target. All three genes have been selected by combining the results of an RNA-seq analysis and total seed proteome analysis on the parental lines and the L94 NILs, as described in detail in chapter 3. The *Arabidopsis thaliana* T-DNA knock-out mutant *nadp-me1*, shown to have a seed longevity phenotype by Nguyen (2014), was complemented with the CC and L94 alleles of the possible candidate and downstream target genes under the expression of different promoters. Interestingly, both the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 are able to rescue the *nadp-me1* seed longevity phenotype when being expressed under the control of the 35S- or the Ubi-promoter. In the latter case, only the CC-specific coding sequence has an effect, while the L94-specific coding sequence is not able to complement the *nadp-me1* even under the control of strong promoters such as the 35S or Ubi-promoter. This finding emphasizes the effect of the allele of the long-lived parental line Cebada Capa in seed longevity. While the NADP-ME MLOC_35785.1 is able to rescue the mutant phenotype, there was no effect observed for the second NADP-ME AK248526.1. This finding could be explained by the observation that the encoded protein was found to be low abundant in Cebada Capa in comparison to L94. Since the T-DNA knock-out line of NADP-ME1 shows a seed longevity phenotype, it would be assumed that the respective protein needs to be highly accumulated in order to protect seeds e.g. against oxidative stress. This hypothesis would be affirmed by the fact that both the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 are higher abundant in Cebada Capa, and in some of the L94 NILs, in comparison to L94.

The in this study observed seed longevity phenotype of transgenic *Arabidopsis* plants expressing the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 under the control of the 35S- or the Ubi-promoter, needs to be confirmed in T_3 generations in the future. Such a confirmation is important to exclude unstable transgene expression. Homozygous T_3 lines can be obtained from T_2 lines, which have been selected due to a 3:1 segregation-ratio in this study. Selection of homozygous T_3 seeds would be performed in dry state using the selection marker DsRed. Alternatively, the selected T_2 lines could be grown and Quantitative Real Time-(qRT)-PCR could be performed using rosette leaf material to select homozygous lines having a single copy of the transformed gene. The time to obtain homozygous T_3 seeds would be the same in both approaches. In the latter approach, qRT-PCR primers would be needed to be developed, tested and possibly optimized.

Interestingly, the expression of neither the NADP-ME AK248526.1 nor the NADP-ME MLOC_35785.1 under the control of the *Arabidopsis thaliana* NADP-ME1 (At2g19900) promoter has an effect on the seed longevity phenotype of *nadp-me1*. This finding could be explained by the fact that both barley NADP-MEs are not that closely related to the *Arabidopsis thaliana* NADP-ME1, but are rather related to rice or maize NADP-ME isoforms, as shown in figure 5. The distinct phylogenetic relationship of the barley and *Arabidopsis* NADP-MEs could lead to non-expressed barley alleles under the control of the *Arabidopsis* promoter. Another explanation would be that the *Arabidopsis* promoter is not strong enough to lead to a detectable change in the *nadp-me1* seed longevity phenotype. Comparable to the results observed for the *Arabidopsis thaliana* NADP-ME1 promoter, the seed longevity phenotype of *nadp-me1* can neither be rescued when expressing the two barley NADP-MEs under the control of the CC-specific, nor the L94-specific promoter of AK248526.1. Few studies are known in which graminaceous promoters are able to induce expression in *Arabidopsis thaliana*. In a study by Kobayashi et al. (2007), promoter fragments of the iron-deficiency-induced gene *IDS3* were fused to the β-glucuronidase (GUS) gene, and this construct was introduced into *Arabidopsis thaliana* and tobacco (*Nicotiana tabacum*) plants. In both *Arabidopsis* and tobacco, GUS activity driven by the *IDS3* promoter showed strongly iron-deficiency-inducible and root-specific expression. The *IDS3* promoter is the first graminaceous promoter to be found to respond to iron deficiency in *Arabidopsis* (Ito et al., 2007), reported that the promoters of the barley and rice *NICOTIANAMINE SYNTHASE 1* (*NAS1*) were not able to induce expression in *Arabidopsis*, despite their ability to respond to iron deficiency in tobacco. The compatibility of the graminaceous iron-deficiency-responsive promoters with trans-acting factors in *Arabidopsis* seems to be highly variable depending on the promoter contexts. Something similar might occur in the case of the NADP-ME barley promoter. Since in both studies expression could be induced in tobacco, it might be interesting to express the constructs with the barley promoters in tobacco and to test the resulting transgenic plants for their seed longevity phenotype. In any case, the NADP-ME MLOC_35785.1 should be additionally expressed under the control of its own promoter in future studies.

To be able to investigate whether the NADP-ME1 affects the redox potential in the future, *nadp-me1* x c-GRX1-roGFP2 double mutants were generated. As shown in this study, seeds of c-GRX1-roGFP2 show a strong GFP signal in wild-type Col-0 background. Apparently the testa proanthocyanin pigments do not hamper the observation of the GFP signal. In contrast to that, seeds of transgenic lines in *tt4* mutant background exhibit an altered fluorescence spectrum and high auto-fluorescence. Due to an EMS induced mutation in a chalcone synthase gene, *tt4* mutants are deficient in flavonoid accumulation, but accumulate an excess of sinapate-esters (Saslowsky et al., 2000). As described by Sheahan (1996), these sinapate-esters have greater UV-B attenuation in *Arabidopsis* leaves than flavonoids. The major sinapate-ester found in seeds is sinapoylcholine, which serves as a reserve of choline and sinapate for the developing seedling (Ruegger et al., 1999). The accumulation of this sinapate ester might lead to the in this study observed high auto-fluorescence. Although the GFP signal is strong enough in Col-0 background, it might still be interesting to obtain transparent *nadp-me1* x c-GRX1-roGFP2 seeds by crossing such lines to a non-fluorescent *tt* mutant. So far, flavonoid fluorescence of *Arabidopsis tt*-mutants has been studied

in the root tip region by Buer et al. (2010). Their fluorescence properties were imaged after the addition of exogenous diphenylboric acid 2-amino ethyl ester (DPBA), which facilitates the visualization of flavonoids *in planta*. According to this study, the *tt* mutants *tt4-tt7* show low flavonoid fluorescence. To exclude auto-fluorescence due to the accumulation of other fluorescent compounds, as it has been observed in the case of *tt4*, *tt5-tt7* should be tested in a similar manner as described in this study.

As mentioned before, T_1 seeds of *nadp-me1* x c-GRX1-roGFP2 transgenic lines showing strong fluorescence, both under a DsRed and GFP filter, were selected and grown for future studies in the greenhouse. The $NADP^+$/NADPH redox potential could be measured in the segregating T_2 population. The in this study described measurement of excitation spectra of seeds with a scanning fluorescent plate reader, after the addition of sterile dH_2O, DTT and DPS, could be used for this purpose.

Since the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 may play a role in seed longevity, as it is suggested by the in this study performed complementation tests, it might be interesting to monitor their activity and abundance during an artificial ageing test. NADP-ME activity could be determined spectrophotometrically as described by Wheeler et al. (2005). In this study, NADP-ME protein concentration was determined by the method of Sedmak and Grossberg (1977) using bovine serum albumin as standard. Additionally, it might be interesting to analyze whether different NADP-ME isoforms can be observed for Cebada Capa and L94. For this purpose, SDS-PAGE could be performed as described by Wheeler et al. (2005). To detect the activity of the UDP-glycosyltransferase MLOC_11661.1, the bioluminescent Promega UDP-Glo™ Glycosyltransferase Assay could be utilized.

As described in this chapter, NADP-MEs could be important for the detoxification of reactive oxygen species. UDP-glycosyltransferases are also implicated in the detoxification of toxins and ROS secondary metabolites (Simon et al., 2014; Krempl et al., 2016). A recent study by Dmitriev et al. (2016) suggests the involvement of UDP-glycosyltransferases (UGTs) in cell wall modification and protection from reactive oxygen species in response to aluminum stress in flax (*Linum usitatissimum*). To test the effect of NADP-MEs and UDP-glycosyltransferases in oxidative stress, it would be interesting to measure ROS content during an artificial ageing test of some of the in this study generated transgenic lines. It would be of special interest to compare the ROS content of transgenic lines expressing the NADP-ME MLOC_35785.1 and the UDP-glycosyltransferase MLOC_11661.1 under the control of the 35S- or the Ubi-promoter, and their respective negative controls in Col-0 and *nadp-me1* background. A lower ROS content in the transgenic lines than in the negative controls would affirm a role of MLOC_35785.1 and MLOC_11661.1 during oxidative stress. The release of reactive oxygen intermediates such as superoxide radicals, hydrogen peroxide, and hydroxyl radicals, could be measured as described by Schöpfer et al. (2001). The by Schöpfer et al. (2001) described measurement, was also used in a study about the role of reactive oxygen species in the regulation of *Arabidopsis* seed dormancy by Leymarie et al. (2012). In the latter study, antioxidant enzyme activities were additionally determined as described by Bailly et al. (1996). Such a measurement might also be interesting for the in this study generated material.

Material and methods

Plant material and growth conditions

Seeds of the barley landraces L94 and Cebada Capa (CC) were sown in 96-well trays filled with a custom-made peat and clay soil mixture (EinheitsErde® ED73 Osmocote, Einheitserdewerke Werkverband e.V., Sinntal Altengronau, Germany). Seeds were germinated in the greenhouse under long day (LD) conditions (16h, 22°C day; 8h, 18°C night). After six weeks, plants were transferred to 1.3 L pots and maintained under LD conditions. Plants were irrigated by hand and treated with additional fertilizer or pesticides when necessary. Leaf material was collected from about seven to eight week old plants and isolated as described in the following.

Arabidopsis plant material used in this study was all derived from the *Arabidopsis thaliana* accession Colombia (Col-0). Seeds were sown on soil and plants were grown in a greenhouse where the temperature was maintained close to 23°C and 16 h of light was provided daily. Ten repetitions per genotype were conducted. Plants were individually harvested at seed maturation, which was reached around the same date, and were not bagged before.

nadp-me1 (SALK_036898C) seeds were kindly provided by Leónie Bentsink (Wageningen University, the Netherlands). Seeds of transgenic *Arabidopsis thaliana* plants constitutively expressing roGFP2 under the 35S-promoter (35S::roGFP2) were kindly provided by Christine Foyer (University of Leeds, England). These transgenic lines originated from Meyer et al. (2007), who transformed *Arabidopsis thaliana* plants using the pBinAR binary vector carrying the 35S::roGFP2 construct (Meyer et al., 2007). Seeds from transgenic *Arabidopsis* plants overexpressing roGFP2 in the cytosol, c-GRX1-roGFP2, were kindly provided by Markus Schwarzländer (University Bonn, Germany). *Arabidopsis thaliana* plants were transformed using the pBinCM binary vector carrying the GRX1::roGFP2 construct via the floral dip method. Seeds of the mutant *dog1-2* were kindly provided by Wim Soppe (MPIPZ Cologne, Germany). Seeds of homozygous T-DNA knock-out lines of seed longevity candidate genes were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). Detailed information about T-DNA knock-out lines used in the longevity tests are listed in table 2. Stock order numbers and sequences of primers used to verify the homozygosity and the presence of the T-DNA insertion are also shown in table 2.

For confocal laser scanning microscopy (CLSM) imaging, seeds from Ubi::GRX1-roGFP2 transgenic plants expressing the *Discosoma* sp. red fluorescent reporter protein DsRed in Col-0 and *nadp-me1* background were surface-sterilized with 70% ethanol twice and re-suspended in sterile deionized water. Seeds were plated on nutrient medium [5 mM KNO_3, 2.5 mM KH_2PO_4, 2 mM $MgSO_4$, 2 mM $Ca(NO_3)_2$, 10 µM Fe-EDTA, 0.1% (v/v) micronutrient mix (Somerville and Ogren, 1982), pH 5.8 solidified with 0.8% phytagel]. Plants were kept at 4°C for 1 day before placing them in vertical orientation in a growth chamber with a diurnal cycle of 16 h light at 22°C and 8 h dark at 18°C. The light intensity was 75 µmol photons m^{-2} s^{-1}. Plants were grown for 3 days before conducting CLSM imaging.

Table 2. Homozygous T-DNA knock-out lines of selected seed longevity candidate genes

Details of the T-DNA knock-out lines tested for their role in seed longevity. For each gene, the gene ID, annotation, T-DNA line code, insertion type and the specific T-DNA left and right primer (LP and RP) sequences for confirming the presence of the insertion in the knock-out line are given. LP and RP primers were designed using the SIGnaL T-DNA verification primer design tool (http://signal.salk.edu/tdnaprimers.2.html).

Gene ID	Annotation	T-DNA line	Insertion	Primer
At2g20810	Galacturonosyltransferase 10	SALK_029319C	exon	RP: TCCTGAATTTGATTGGAGTGC
				LP: GGTTTTACACAGCTCTGCGAG
AT2G19900	NADP-dependent malic enzyme 1	SALK_036898C	intron	RP: CACTTGCAGAGCATACCTTCC
				LP: TGCTTACCAGCAGCTAGAAGC

DNA extraction from leaf material

For DNA extraction, approximately 100 mg of barley leaf tissue and *Arabidopsis thaliana* plants was collected into Qiagen ^{TM} 96-well collection tubes. The collection tubes were placed into 96-well blocks, sealed with Qiagen^{TM} AirPore tape sheets, stored at -80°C for at least two hours and subsequently freeze-dried with a Christ Alpha 1-4 LSC freeze drier (Martin Christ GmbH, Osterode am Harz, Germany). After freeze-drying, the AirPore tape sheets were removed and one 5 mm Qiagen^{TM} stainless steel bead was added into each tube and the tubes were sealed with 8-well strips. The 96-well blocks were placed into a Retsch mixer mill (Retsch®, Haan, Germany) and the material was ground two times for 30 seconds at 25-30 Hz. DNA was extracted of the ground material with the Qiagen ^{TM} BioSprint (Qiagen, Hilden, Germany) robot and the corresponding kit, following the supplier's recommendations. DNA was eluted in 200 µL of the supplied AE buffer and concentrations were measured using the NanoDrop^{TM} spectrophotometer (PEQLAB, Erlangen, Germany). For further analyses, dilutions of 5-10 ng/µL DNA in AE buffer were prepared and stored at 4°C. Undiluted eluates were stored at -20°C.

T-DNA knock-out genotype analyses

Homozygous T-DNA insertion lines were screened using polymerase chain reactions (PCR) with gene specific left (LP) and right (RP) primers (Table 2) and the left T-DNA insertion border primer LBb1.3 (5'-ATTTTGCCGATTTCGGAAC-3'). T-DNA plants that amplified only the insertion product were considered to be homozygous mutants. Primers were supplied by Invitrogen (Invitrogen Life Technologies GmbH, Darmstadt, Germany) and diluted to 100 mM stock solutions with sterile dH$_2$O and stored at -20°C. 10 mM primer dilutions in sterile dH$_2$O were used as working solutions. Approximately 100 mg of leaf tissue was harvested from T-DNA mutants and genomic DNA was extracted according to the section 'DNA extraction from leaf material'. Polymerase chain reactions were performed with Phusion® High-Fidelity DNA Polymerase purchased from New England Biolabs GmbH (Frankfurt am Main, Germany) and the supplied buffer solutions from the same company. dNTP sets (4x100 mM solutions) were purchased from Biotechrabbit GmbH (Hennigsdorf, Germany) and 10 mM solutions were prepared with sterile dH$_2$O. 20 µl-volume reactions were set up in 0.2 ml PCR SingleCAP 8er-SoftStrips (Biozym, Hessisch Oldendorf, Germany) according to Table 3.

Table 3. Reaction setup for 20 µl volume PCR using Phusion® High Fidelity DNA polymerase

Component	20 µl Reaction
Nuclease-free water	to 20 µl
5X Phusion HF Buffer	4 µl
10 mM dNTPs	0.4 µl
10 mM Forward Primer	1 µl
10 mM Reverse Primer	1 µl
Template DNA	~30 ng
Phusion® DNA Polymerase	0.2 µl

All reaction components were mixed and centrifuged prior to use and Phusion® High-Fidelity DNA Polymerase was added last in order. PCR tubes were transferred to an Eppendorf® Mastercycler® (ep. 384) preheated to 98°C and thermocycling was performed according to the settings shown in table 4.

Table 4. Thermocycling conditions for PCR with Phusion® High Fidelity DNA polymerase

Step	Temperature (°C)	Time	Cycles
Initial Denaturation	98	30 seconds	1
Denaturation Annealing Extension	98 60°C 72	5-10 seconds 10 seconds 15-30 seconds per kb	35
Final Extension Hold	72 4	5 minutes hold	1

6x DNA loading dye (ThermoFisher, Oberhausen, Germany) was added to the PCR samples and the polymorphism was detected by agarose gel electrophoresis at concentrations from 1.5 % (w/v) and higher depending on size of differences. The running buffer stock solution and the agarose gels were prepared as shown in tables 5 and 6. Chemicals and solutions were purchased from Sigma-Aldrich (Taufkirchen, Germany).

Table 5. Recipe for preparation of 5 liters 50x TAE running buffer stock solution

Solution component	Component volumes
Tris base	1211 g
Na$_2$EDTA	146.2 g
Glacial acetic acid	290 mL
dH$_2$O	to 5 L

adjust pH to 8.0 with acetic acid

Table 6. Recipe for preparation of 1-4% agarose gel

Solution component	Component volumes
Agarose	1.6 to 6.4 g (1 to 4 %)
1x TAE buffer	160 mL
Ethidium bromide (10 mg/mL)	3.2 µL

Transgenic lines

To express the barley seed longevity candidate genes AK248526.1, MLOC_35785.1 and MLOC_11661.1 under the control of the strong constitutive CaMV 35S-promoter, the Ubiquitin-promoter (Ubi), the promoter of the *Arabidopsis thaliana* NADP-ME1 (At2g19900) and the promoter of the barley NADP-ME candidate gene AK248526.1, the Gateway® system (Invitrogen Life Technology, USA) was used.

The promoters of the *Arabidopsis thaliana* NADP-ME1 (At2g19900) and the barley NADP-ME candidate gene AK248526.1 were amplified from wild-type Columbia (Col-0) genomic DNA and genomic DNA from the two barley parental lines Cebada Capa and L94, respectively, with indicated oligonucleotides (5'-<u>cggggtacc</u>agacctgacctgtgaaccgg-3' and 5'-<u>ttccccccgggg</u>tctaggatcacaaaggaagaaact-3', 1.6 kb; 5'-<u>cggggtacc</u>accccgagtacaacgacact-3' and 5'-<u>ttccccccgggg</u>ctcctcaaactctcggcctcgg-3', 2 kb) and cloned by restriction with XmaI and KpnI (New England Biolabs, Frankfurt am Main, Germany) into the pUC57-L4-KpnI/XmaI-R1 plasmid producing pEN-L4-promoter-R1 clones. Underlined sequences represent XmaI and KpnI sites. The pUC57-L4-KpnI/XmaI-R1 plasmid was generated by Joop Vermeer (University of Lausanne) by introducing L4-KpnI/XmaI-R1 att recombination and restriction sites into pUC57 (Invitrogen). pEN-L4-35S-R1 and pEN-L4-Ubi-R1 plasmids, containing the CaMV 35S and the Ubiquitin promoter as described by Karimi et al. (2007), were kindly provided by Iván Acosta (MPIPZ Cologne, Germany).

Coding DNA sequences (CDS) of the barley seed longevity candidate genes AK248526.1 (5'-<u>ggggacaagtttgtacaaaaaagcaggct</u>ttgaggagatggccggcgg-3' and 5'-<u>ggggaccactttgtacaagaaagctgggtt</u>acctgtagctgcggtagagcg-3') and MLOC_35785.1 (5'-<u>ggggacaagtttgtacaaaaaagcaggct</u>tatgcacaaccttcgccagtacaa-3' and

5'-<u>ggggaccactttgtacaagaaagctgggttt</u>aacgatagctgcggtagacagg-3') and MLOC11661.1 (5'-<u>ggggacaagtttgtacaaaaaagcaggct</u>tatgacggacgaaatgctgagcg-3' and 5'-<u>ggggaccactttgtacaagaaagctgggtt</u>cagcgcatgaagcaccgttg-3') were amplified with oligonucleotides specified in parenthesis containing Gateway attB recombination sites (underlined). cDNA was synthesized from whole embryo total RNA of mature seeds of Cebada Capa and L94. Amplification products were recombined into pDONR221 (Invitrogen) to produce pEN-L1-CDS-L2 clones.

The final expression constructs were obtained with a double Gateway reaction of the respective pEN-L1-CDS-L2 and pEN-L4-promoter-R1 clones into the destination vector pEDO097pFR7m24GW. This vector was generated by inserting the fluorescence-accumulating seed technology (FAST) cassette (Shimada et al., 2010) into pH7m24GW (Invitrogen) by Ester M. N. Dohmann (University of Lausanne). All constructs were introduced into wild-type Columbia (Col-0) and *nadp-me1* backgrounds by floral dip *Agrobacterium*-mediated transformation (Clough and Bent, 1998). Transformed seeds expressing the *Discosoma* sp. red fluorescent reporter protein DsRed in T_1 and T_2 lines were selected by fluorescence microscopy in dry state. T_2 lines, showing a segregation ratio of 3:1, indicating the presence of a single functional transgenic locus, and high DsRed-fluorescence, were selected. Non-fluorescent sibling seeds were selected to be used as negative control for each construct. A minimum of ten independent transgenic lines were used for each construct to perform experiments and verify reproducibility.

To construct Ubi::c-GRX1-roGFP2 transgenic lines expressing the DsRed reporter protein, the GRX1-roGFP2 gene fragment was amplified from pBinCM-GRX1-roGFP2 (Gutscher et al., 2008), which was kindly provided by Markus Schwarzländer (University Bonn, Germany), with indicated oligonucleotides (5'-<u>ggggacaagtttgtacaaaaaagcaggct</u>atggctcaagagtttgtgaactgc-3' and 5'-<u>ggggaccactttgtacaagaaagctgggttt</u>tacttgtacagctcgtccatgcc-3'), containing Gateway attB recombination sites (underlined). Amplification products were recombined into pDONR221 (Invitrogen) to produce pEN-L1-GRX1-roGFP2-L2 clones. The final expression construct was obtained with a double Gateway reaction of the respective pEN-L1-GRX1-roGFP2-L2 and pEN-L4-Ubi-R1 clones into the destination vector pEDO097pFR7m24GW. This construct was introduced into wild-type Columbia (Col-0) and *nadp-me1* backgrounds by floral dip Agrobacterium-mediated transformation (Clough and Bent, 1998). Transformed seeds expressing DsRed in T_1 lines were selected by fluorescence microscopy in dry state.

Sanger sequencing of candidate genes and their promoter sequences

20 µl PCR reactions were performed using Phusion® High-Fidelity DNA Polymerase, purchased from New England Biolabs GmbH (Frankfurt am Main, Germany), to amplify candidate gene coding sequences and the L94- and CC-specific promoter sequence of AK248526.1. For the amplifications, the same primers as for the cloning were used. PCR reactions were set up as shown in table 3. 2 µl of the PCR reactions were loaded on an agarose gel to verify the presence of a clean PCR product. The remaining PCR product was cleaned using the Qiagen QIAquick® PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations and eluted in 30 µL of the supplied AE buffer. DNA concentration was measured with the NanoDrop™ spectrophotometer

(PEQLAB, Erlangen, Germany) and adjusted to the concentration required for Sanger sequencing. Sequencing was performed by the Max Planck Genome Centre Cologne with the primers also used for PCR amplification. DNA chromatogram files in ABI format were aligned and analyzed with the DNASTAR® Lasergene SeqMan Pro assembler (DNASTAR, Madison, USA). Unclean or unsuccessful runs were removed from the analysis.

Artificial ageing

Seed longevity was determined as germination ability after an artificial ageing test (Tesnier et al., 2002). For the artificial ageing test, approximately 50 seeds per replicate were aliquoted for each genotype in 0.2 ml PCR SingleCAP 8er-SoftStrips (Biozym, Hessisch Oldendorf, Germany). For each time point four replicates were conducted, untreated seeds (time point 0) were used as controls. PCR stripes were placed in PCR tube racks and stored above a saturated KCl solution (80% relative humidity) in a closed container at 37°C. After 4, 8 and 12 days, PCR tubes of the corresponding samples were removed from the container and placed in a container with silica gel for three days to reduce relative humidity to 10%.

Germination assays

Immediately after artificial ageing, germination assays were performed. Seeds were sown on one layer of Ø 50 mm filter paper (MN 615, Macherey-Nagel®, Düren, Germany) placed in a Ø 60 mm petri-dish (Greiner dishes with vents, Sigma-Aldrich®, Taufkirchen, Germany). Filter papers were pre-moistened with 580 µl of double-distilled H_2O. Petri-dishes were placed in piles into transparent moisturized containers, and incubated in a germination cabinet (Van den Berg Klimaattechniek) in long-day conditions (12h light at 25°C, followed by 12h darkness at 20°C). After 7 days of incubation, germination counts were done and radicle protrusion was taken to indicate germination. Germinated seeds were removed at time of counting. The collected data was used to plot mean germination percentage curves against time after artificial ageing treatment. Statistic Student's t-tests were performed using a cut-off of $p \leq 0.05$.

Fluorescence microplate assay analysis

Fluorescence microplate assays were performed at the University of Bonn under the supervision of Markus Schwarzländer. Approximately 50 mature seeds of the transgenic lines *tt4*, c-GRX1-roGFP2, the double mutant *tt4* x 35S::roGFP2 and the wild-type Col-0 were aliquoted in 1.5 ml Eppendorf tubes and 100 µl of sterile dH_2O, 10 mM DTT and 5 mM Diphenyl sulfide (DPS) were added. The solutions were directly transferred to the wells of a 96-well microtiter plate. Three replicates per genotype and solution were performed. The 96-well microtiter plate was transferred to a BMG LABTECH CLARIOstar® microplate reader (BMG LABTECH, Ortenberg, Germany) and fluorescence was measured in a wavelength range from 360 nm to 496 nm. The results were represented as means ± standard deviation from three independent measurements.

Confocal laser scanning microscopy (CLSM) imaging

CLSM imaging was performed at the University of Bonn under the supervision of Markus Schwarzländer. 3-day-old seedlings or mature seeds were mounted on a slide in a drop of water and immediately transferred to a Zeiss confocal microscope LSM780 (Carl Zeiss Microscopy, www.zeiss.de/mikro). Images were collected with a 40× lens (Zeiss Objective C-Apochromat 40×/1.2 W Corr M27) in multi-track mode with line switching between 488 nm excitation and 405 nm excitation and taking an average of four readings. The configurations used for the acquisition of images are summarized in table 7.

Table 7. Confocal laser scanning microscopy (CLSM) configurations

Scan Mode	Plane
Scaling	X: 0.09 µm Y: 0.09 µm
Stack size	X: 96.61 µm Y: 96.61 µm
Scan Zoom	2.2
Objective	C-Apochromat 40×/1.2 W Corr M27
Average	Line 4
Pinhole	Ch1: 37 µm ChS1: 37 µm
Filters	ChS1: 508-535
Beam Splitters	MBS: MBS 488/543/633: f-MBS 405/595c DBS1: Mirror FW1: NonelLSM
Wavelength	405 nm T1 2.0% 488 nm T2 1.0%

Stereo fluorescence microscopy imaging

Mature wild-type Col-0 seeds and seeds of transgenic lines were imaged with the Leica M205 FA (Leica Mikrosysteme Vertrieb GmbH, Wetzlar, Germany) stereo fluorescence microscope and fluorescence images were acquired with the Leica sCMOS fluorescence camera. Seeds were maintained in dry state.

References

Asada, K. (1992), Ascorbate peroxidase - hydrogen peroxide-scavenging enzyme in plants. Physiol. Plant. 85: 235-241

Bagchi, S. et al. (1987), Structure and expression of murine malic enzyme mRNA. J. Biol. Chem. 262: 1558-1565

Bailly, C. et al. (1996), Changes in malondialdehyde content and in superoxide dismutase, catalase and glutathione reductase activities in sunflower seeds as related to deterioration during accelerated aging. Physiol. Plant. 104: 646-652

Bailly, C. (2004), Active oxygen species and antioxidants in seed biology. Seed Sci. Res. 14: 93-107

Bentsink, L. et al. (2006), Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. Proc. Natl. Acad. Sci. USA 103: 17042-17047

Bernal-Lugo I. and Leopold A.C. (1998), The dynamics of seed mortality. Journal of Experimental Botany 49: 1455-1461

Borsch, D. and Westhoff, P. (1990), Primary structure of NADP-dependent malic enzyme in the dicotyledonous *Flaveria trinervia*. FEBS Lett. 273: 111-115

Buer, C.S. et al. (2010), Flavonoids: New roles for old molecules, Journal of Integrative Plant Biology 52: 98-111

Chi, W. et al. (2004), Four rice genes encoding NADP-malic enzyme exhibit distinct expression profiles. Biosci. Biotechnol. Biochem. 68: 1865-1874

Christin, P.A. et al. (2009), Evolutionary insights on C4 photosynthetic subtypes in grasses from genomics and phylogenetics. Genome Biol. Evol. 1: 221-230

Clough S.J. Bent A.F. (1998), Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*, Plant J. 16: 735-743

Cushman, J.C. (1992), Characterization and expression of a NADP-malic enzyme cDNA induced by salt stress from the facultative Crassulacean acid metabolism plant, *Mesembryanthemum crystallinum*. Eur. J. Biochem. 208:259-266

Dereeper, A. et al. (2008), Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36: 465-469

Dmitriev, A.A. et al. (2016), Glutathione S-transferases and UDP-glycosyltransferases are involved in response to aluminum stress in flax, Front. Plant Sci. 7, 1-10

Drincovich, M.F. et al. (2001), NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. FEBS Lett. 490: 1-6

Dooley, C.T. et al. (2004), Imaging dynamic redox changes in mammalian cells with green fluorescent protein indicators. J. Biol. Chem. 279: 22284-93

Edwards, G.E. and Huber, S.C. (1981), The C4 pathway. In: The Biochemistry of Plants. A Comprehensive Treatise (Hatch, M.D. and Boardman, N.K.) Academic Press, New York 8: 237-281

Edwards, G.E. and Andreo, C.S. (1992), NADP-malic enzyme from plants. Phytochemistry 31: 1845-1857

Franke, K.E. and Adams, D.O. (1995), Cloning a full-length cDNA for malic enzyme (EC 1.1.1.40) from grape berries. Plant Physiol. 107: 1009-1010

Foyer, C.H. and Shigeoka, S. (2011), Understanding oxidative stress and antioxidant functions to enhance photosynthesis. Plant Physiol. 155: 93-100

Fu, Z.Y. et al. (2009), Cloning, identification, expression analysis and phylogenetic relevance of two NADP-dependent malic enzyme genes from hexaploid wheat. C. R. Biol. 332: 591-602

Fushimi, T. et al. (1994), Nucleotide sequence of rice cDNA similar to a maize NADP-dependent malic enzyme. Plant Mol. Biol. 24: 965-967

Gutscher, M. et al. (2008), Real time imaging of the intracellular glutathione redox potential. Nature Methods 5: 553-559

Hanson, G.T. et al. (2004), Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. J. Biol. Chem. 279: 13044-13053

Honda, H. et al. (1997), Isolation of cDNA for an NADP-malic enzyme from *Aloe arborescens*. DNA Res. 4: 397-400

Honda, H. et al. (2000), An isozyme of the NADPmalic enzyme of a CAM plant*, Aloe arboresens*, with variation on conservative amino acid residues. Gene 243: 85-92

Ito, S. et al. (2007), Interspecies compatibility of NAS1 gene promoters. Plant Physiol Biochem. 45: 270-276

Karimi, M. et al. (2007), Building blocks for plant gene assembly. Plant Physiol. 145: 1183-1191

Kibinza, S. et al. (2006), Sunflower seed deterioration as related to moisture content during ageing, energy metabolism and active oxygen species scavenging. Physiol. Plant. 128: 496-506

Klungland, A. et al. (1999), Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. Proc. Natl. Acad. Sci. USA 96: 13300-13305

Kobayashi, T. et al. (2007), Promoter analysis of iron-deficiency-inducible barley IDS3 gene in Arabidopsis and tobacco plants. Plant Physiol. Biochem. 45: 262-269

Koornneef, M. (1990), Mutations affecting the testa color in *Arabidopsis*. Arabid. Inf. Serv. 27: 94-97

Krempl, C. et al. (2016), Potential detoxification of gossypol by UDP-glycosyltransferases in the two Heliothine moth species *Helicoverpa armigera* and *Heliothis virescens*. Insect Biochem. Mol. Biol. 71, 49-57

Lai, L.B. et al. (2002), Differential regulation of transcripts encoding cytosolic NADP-malic enzymes in C3 and C4 *Flaveria* species. Plant Physiol. 128: 140-149

Leymarie, J. et al. (2012), Role of reactive oxygen species in the regulation of Arabidopsis seed dormancy. Plant Cell Physiol. 53: 96-106

Lipka, B. et al. (1994), The C3 plant *Flaveria pringlei* contains a plastidic NADP-malic enzyme which is orthologous to the C4 isoform of the C4 plant *F. trinervia*. Plant Mol. Biol. 26: 1775-1783

Maier, A. et al. (2011), Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C4 and C3 photosynthesis. J. Exp. Bot. 62: 3061-3069

Martinoia, E. and Rentsch, D. (1994), Malate compartmentation-responses to a complex metabolism. Annual Review of Plant Physiology and Plant Molecular Biology 45: 447-467

McDonald, M.B. (1999), Seed Deterioration: Physiology, Repair and Assessment, Seed Sci. Technol. 27: 177-237

Mehler, A.H. (1951), Studies on reactions of illuminated chloroplasts. I. Mechanism of the reduction of oxygen and other Hill reagents.Arch. Biochem. Biophys. 33: 65-77

Meyer, A.J. et al. (2007), Redox-sensitive GFP in Arabidopsis thaliana is a quantitative biosensor for the redox potential of the cellular glutathione redox buffer. Plant J. 52: 973-986

Nguyen, T.P. et al. (2012), Natural variation for seed longevity and seed dormancy are negatively correlated in *Arabidopsis*. Plant Physiol. 160: 2083-2092

Nguyen, T.P. (2014), Seed Dormancy and Seed Longevity: from genetic variation to gene identification, Utrecht University Repository (Dissertation), ISBN: 978-90-393-6094-1

Rajjou, L. et al. (2008), Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. Plant Physiol. 148, 620-641

Rothermel, B.A. and Nelson T. (1989), Primary structure of the maize NADP-dependent malic enzyme. J. Biol. Chem. 264: 19587-19592

Ruegger, M. et al. (1999), Regulation of ferulate 5 hydroxylase expression in Arabidopsis in the context of sinapate ester biosynthesis, Plant Physiol. 119, 101-110

Rush, G.F. et al. (1985), Organic hydroperoxide-induced lipid peroxidation and cell death in isolated hepatocytes. Toxicology and Applied Pharmacology. 78: 473-483

Saigo, M. et al. (2004), Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. Plant Mol. Biol. 55: 97-107

Saslowsky, D.E. et al. (2000), An allelic series for the chalcone synthase locus in *Arabidopsis*. Gene 255: 127-138

Schwarzländer M. et al. (2008), Confocal imaging of glutathione redox potential in living plant cells. J. Microsc. 231: 299-316

Schöpfer, P. et al. (2001), Release of reactive oxygen intermediates (superoxide radicals, hydrogen peroxide, and hydroxyl radicals) and peroxidase in germinating radish seeds controlled by light, gibberellin, and abscisic acid. Plant Physiol. 125: 1591-1602

Sedmak J.J. and Grossberg, S.E. (1977), A rapid, sensitive, and versatile assay for protein using Coomassie brilliant blue G250. Anal Biochem 79: 544-552

Sheahan, J.J. (1996), Sinapate esters provide greater UV-B attenuation than flavonoids in *Arabidopsis thaliana* (*Brassicaceae*). American Journal of Botany 83: 679-686

Shearer, H.L. et al. (2004), Characterization of NADP-dependent malic enzyme from developing castor oil seed endosperm. Arch. Biochem. Biophys. 429: 134-144

Shimada, T.L. et al. (2010), A rapid and nondestructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. Plant J. 61: 519-528

Simon, C. et al. (2014), The secondary metabolism glycosyltransferases UGT73B3 and UGT73B5 are components of redox status in resistance of *Arabidopsis* to *Pseudomonas syringae* pv. tomato. Plant Cell Environ. 37, 1114-1129

Stewart R.C. and Bewley J.D. (1980), Lipid peroxidation associated with accelerated aging of soybean axes. Plant Physiol. 65: 245-248

Somerville C.R. and Ogren W.L. (1982), Isolation of photorespiration mutants in *Arabidopsis*, in Methods in Chloroplast Molecular Biology, eds Edelman M., Hallick R. B., Chua N. H., editors, Amsterdam: Elsevier Biomedical Press, 129-138

Tesnier K. et al. (2002), A controlled deterioration test for *Arabidopsis thaliana* reveals genetic variation in seed quality. Seed Sci. Technol. 30: 149-165

The International Barley Genome Sequencing Consortium (2012), A physical, genetic and functional sequence assembly of the barley genome. Nature 491, 711-716

Van Doorsseleare, J. et al. (1991) Nucleotide sequence of a cDNA encoding malic enzyme from poplar. Plant Physiol. 96: 1385-1386

Walter, M.H. et al. (1990), Extensive sequence similarity of a bean CAD4 ''cinnamyl-alcohol dehydrogenase'' to a maize malic enzyme. Plant Mol. Biol. 15: 525-526

Wheeler, G. et al. (2005), A comprehensive analysis of the NADP-malic enzyme gene family of *Arabidopsis thaliana*. Plant Physiol. 139: 39-51

Wheeler, G. et al. (2009), Identification of domains involved in the allosteric regulation of cytosolic *Arabidopsis thaliana* NADP-malic enzymes. FEBS J. 276: 5665-5677

Chapter 5

General Discussion

The seed constitutes the main vector of plant propagation and is a critical stage of development with many specificities. Seed longevity and seed dormancy are two major characteristics that control seed quality and are important survival traits in the soil seed bank. While seed dormancy is related to the timing of germination, seed longevity is involved in retaining germination ability that is gradually lost as a result of ageing. Both traits are induced during seed maturation and evolved to adapt to changing environmental conditions. Seed longevity is a major challenge for the conservation of plant biodiversity and for crop success. Knowing and understanding the complex features that govern seed longevity are therefore of major ecological, agronomical and economical importance.

Until now little is known about the genetic basis of differences in seed longevity because this trait is affected by numerous environmental effects during seed formation, harvest, and storage, and genetic variation between and within species is probably controlled by several genes. Seed longevity is a quantitative trait for which variation is often present among naturally occurring accessions of a specific species. The regions within genomes that contain genes associated with a particular quantitative trait are known as Quantitative Trait Loci (QTLs). Only in the past decades, with the advent of molecular marker technologies and the development of QTL-mapping procedures, the identification of genomic regions controlling quantitative traits has become more feasible (Tanksley, 1993; Jansen, 1996). In this respect, the use of homozygous permanent mapping populations such as recombinant inbred lines (RILs) is very efficient due to the possibility to study an indefinite number of traits on the same experimental population and to have replicates (Prioul et al., 1997). To confirm and fine map QTLs, near isogenic lines (NILs), containing only a single introgression fragment from a donor parent into an otherwise homogeneous genetic background, can be generated. QTLs for seed longevity have been identified previously in *Arabidopsis thaliana* populations, using artificial (controlled deterioration tests; Bentsink et al., 2000; Clerkx et al., 2004b; Joosen et al., 2012) and natural (Nguyen et al., 2012) ageing. In crop plants, QTL mapping studies for seed longevity have been described in cabbage (*Brassica oleracea*; Bettey et al., 2000), tomato (*Lycopersicon* sp.; Foolad et al., 1999), rice (*Oryza sativa*; Miura et al., 2002), wheat (Landjeva et al., 2010), *Sorghum bicolor* (Natoli et al., 2002), soybean (Singh et al., 2008) and barley (*Hordeum vulgare*; Nagel et al., 2009).

Recent advances in various 'omics'-technologies provide the possibility to address the complex biological systems that underlie various plant functions, and aim to accelerate candidate gene selection for genes underlying variation at QTLs. Global approaches such as transcriptome and proteome profiling have provided e.g. important new information about mechanisms controlling seed germination and dormancy in *Arabidopsis thaliana* and proved useful for the characterization of potential biomarkers of seed vigor (Holdsworth et al., 2008).

The work presented in this thesis aims to identify and characterize seed longevity loci/genes in barley (*Hordeum vulgare*) by combining a quantitative genetics approach with 'omics'-technologies. This work is based on a previous QTL study in RILs, derived from the short-lived Ethiopian spring barley landrace L94 and the long-lived Argentinian spring barley landrace Cebada Capa, which led to the identification of four putative QTLs for seed longevity on chromosomes 1 and 2 (Adimargono et al., in preparation).

Using near isogenic lines, these four putative QTLs could be confirmed in the present study and further mapping of Cebada Capa introgressions in the so-called L94 NILs using RNA-seq was achieved. For candidate and downstream target gene identification, a combined transcriptome and proteome analysis was performed. A similar approach was recently pursued by Nguyen (2014) in *Arabidopsis thaliana* and led to the identification of genes significantly affecting seed longevity, such as the NADP-dependent malic enzyme 1 (NADP-ME1). The analysis of mature, non-aged seeds of the two parental lines and the L94 NILs by RNA-seq and total seed proteomic profiling identified two putative candidate genes and one possible downstream target gene. Both analyses are only possible due to recent improvements of the barley genomics infrastructure. Due to its relatively large size of 5.1 gigabases and high content of repetitive DNA, sequencing and assembly of the barley genome is a slow process that is still ongoing. The International Barley Genome Sequencing Consortium (IBSC) recently presented a gene-space assembly (The International Barley Genome Sequencing Consortium, 2012) of cultivated barley as an enabling platform for genome-assisted basic research and crop improvement. This assembly represents 86% of the entire barley gene set and, through extensive physical and genetic mapping resources, the sequence contigs have been arranged in a linear order. Moreover, an exome capture assay designed on the basis of the annotated sequence assembly has made approximately 60 Mb of mRNA-coding sequence accessible to cost-efficient high-throughput resequencing (Mascher et al., 2013). However, some regions of the genome might still not be represented in the currently available sequence data and the order of Bacterial Artificial Chromosome (BAC) contigs is not established. Furthermore, a low recombination rate in non-distal chromosome regions impedes mapping of such contigs as well as genes located in these regions. Identification of candidate genes by RNA-seq or exome capture (Mascher et al., 2013) is limited when causal polymorphisms are located in intergenic regions. Moreover, expression analysis is biased by the tissue sampled, the time of sampling and the sequencing depth.

The validation of putative candidate genes in barley itself is still difficult. At the moment, efficient *Agrobacterium tumefaciens*-mediated barley transformation is limited to the spring barley cultivar Golden Promise (Tingay et al., 1997). However, progress has been made in wheat transformation, which is extended to various cultivars (Richardson et al., 2014). Therefore, the currently existing limitation in barley transformation might be overcome in the future and complementation experiments with the mutant of interest might be facilitated. For the production of targeted mutants, which allow the functional characterization of putative candidate genes, the CRISPR/Cas technology looks promising (Xie and Yang, 2013; Shan et al., 2014). Furthermore, targeted gene knock-outs in barley have been achieved using designer transcription activator-like effector nucleases (TALENs). Both methods introduce double strand breaks (DSBs) in the DNA sequence present at a pre-determined genomic site (Gurushidze et al., 2014; Frock et al., 2015). A large number of induced barley mutants had already been created from the 1940s to the 1970s, when mutation breeding programs flourished, and presents excellent resources for forward genetics (Ehrenberg et al., 1966; Gustafsson et al., 1971). These mutant lines with various phenotypes are nowadays available in their original backgrounds and have also been backcrossed to the cultivar Bowman as a recurrent parent to obtain mutant alleles in a nearly isogenic background (Lundqvist et al., 1997; Druka et al., 2011). Mutant collections exist for a

number of morphological traits such as row type, which is considered a key domestication character. The two-rowed spike (inflorescence) is the ancestral state present in wild barley (*Hordeum vulgare* ssp. *spontaneum*). Six-rowed types arose in cultivated barley by loss-of-function mutations in *SIX-ROWED SPIKE 1* (*VRS1*), a gene encoding class I homeodomain-leucine zipper (HD-Zip) transcription factor *HvHOX1,* required to suppress lateral floret development (Komatsuda et al., 2007). To identify the gene underlying the *VRS1* locus, Komatsuda et al. (2007) made use of mutant collections by re-sequencing putative candidate genes from lines harboring different allelic variants of the *VRS1* locus. Such an approach is straightforward since the gene with different non-conservative mutations in different allelic variants is very likely causal to the phenotype of interest. Recently, a similar approach has been pursued for *VRS3*, another recessive locus to modify the extent of lateral floret development (reviewed in Pourkheirandish and Komatsuda, 2007). To identify the underlying gene and causal mutation, van Esse et al. (submitted) analyzed different developmental stages of two allelic mutants of the respective locus by RNA-seq. Using mutant collections, mapping-by-sequencing was recently implemented to rapidly identify a gene underlying the *many-noded dwarf* (*mnd*) phenotype (Mascher et al., 2014). The immediate success of a mapping-by-sequencing experiment, that is, pinpointing a candidate in a single step, can be hindered by many factors. Beyond an intrinsic dependence of genetic mapping on recombination rate and the degree of polymorphism between the parents of the mapping population, sequence-based methods are contingent on genomic resources. In the present study, the identification of a candidate for *MND* was facilitated by the previous characterization of a homolog in rice and the advantageous ratio between physical and genetic distance at the target locus (<1 Mb per cM). Nevertheless, this study illustrates what mapping-by-sequencing can achieve in the context of the current genomic framework of barley despite of its fragmentary structure. However, since the majority of the in collections present barley mutants was mainly selected on morphological characteristics and less on complex physiological characters, a phenotype based approach is still difficult to pursue for seed longevity.

In the present study, the identified putative candidate and downstream target genes were validated by complementing a T-DNA knock-out line of a homologous *Arabidopsis* gene, shown to have a seed longevity phenotype by Nguyen (2014), with the respective Cebada Capa and L94 alleles. Such an approach is time saving due to a multitude of already existing cloning methods and vectors, an efficient *Agrobacterium tumefaciens*-mediated transformation and a fast generation time in *Arabidopsis thaliana*. However, the validation of genes, identified in monocotyledonous cereal crops, in *Arabidopsis* is limited due to the large evolutionary distance between them. Therefore *Arabidopsis* might only be a useful model system for some plant processes.

The current study shows that, despite the limitations in barley research, the combination of recent 'omics' technologies with the already available genetic resources enables the identification of candidate genes. Therefore, these approaches would allow to go beyond QTL mapping and would indeed facilitate the identification of the gene/s underlying a trait QTL, as it is suggested in many studies. It would be interesting to validate the in this study identified NADP-ME MLOC_35785.1 and UDP-glycosyltransferase MLOC_11661.1, which both rescued the seed longevity phenotype of the tested *Arabidopsis* T-DNA knock-out line, in barley itself in the future using the before described

developing technologies such as CRISPR/Cas. Studies on these genes could help to improve already existing methods for the conservation of cereal seeds in gene banks. Given that gene order is largely conserved between the grasses (Moore et al., 1995), results from this research in barley could be transferred to the agronomical important bread wheat (*Triticum aestivum*) and could complement wheat research and breeding. Research on bread wheat itself is complicated due to the very large allohexaploid genome (17 Gbp) (Brenchley et al., 2012). Thanks to its large agronomic importance and relatively small genome (Goff et al., 2002; Yu et al., 2002), rice has been established as a model for domesticated cereals and large advances have been made in this field. However, rice is still evolutionary quite distant from bread wheat. *Brachypodium* is more closely related to the *Triticeae*, but as an undomesticated plant, the investigation of agronomically relevant traits is limited. Barley on the other hand is a good model for bread wheat since it is a domesticated member of the *Triticeae*, has the same basic chromosome number as wheat (n = 7), originates from the same region and is very similar in shoot and inflorescence architecture.

References

Bentsink, L. et al. (2000), Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. Plant Physiol. 124: 1595-1604

Bettey, M. et al. (2000), Quantitative genetic analysis of seed vigor and pre-emergence seedling growth traits in *Brassica oleracea*. New Phytol. 148: 227-286

Brenchley, R. et al. (2012), Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491: 705-710

Clerkx, E.J.M. et al. (2004b), 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 Physiol. 135: 432-443

Druka, A. et al. (2011), Genetic Dissection of Barley Morphology and Development. Plant Physiology 155: 617-627

Ehrenberg L. et al. (1966), On the mutagenic action of alkanesulfonic esters in barley. Hereditas 56: 277-305

Foolad, M.R. et al. (1999), Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato. Plant Breed. 118: 167-173

Frock, R.L. et al. (2015), Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases, Nature Biotechnology 33: 179-186

Goff, S.A. et al. (2002), A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 296: 92-100

Gurushidze, M. et al. (2014). True-breeding targeted gene knock-out in barley using designer TALE-Nuclease in haploid cells. PLoS One 9

Gustafsson Å. et al. (1971), Induced mutations and barley improvement. Theor. Appl. Genet.41: 239-248

Holdsworth, M.J. et al. (2008), Post-genomics dissection of seed dormancy and germination, Trends Plant Sci. 13: 7-13

Jansen, R.C. (1996), Complex plant traits: time for polygenic analysis. Trends Plant Sci. 1: 89-94

Joosen, R.V.L. et al. (2012), Visualizing the genetic landscape of *Arabidopsis* seed performance. Plant Physiol. 158: 570-589

Komatsuda, T. et al. (2007), Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc. Natl. Acad. Sci. U S A 104: 1424-1429

Landjeva, S. et al. (2010), Genetic mapping within the wheat D genome reveals QTLs for germination, seed vigour and longevity, and early seedling growth. Euphytica 171: 129-143

Lundqvist, U. et al. (1997), New and revised description of barley genes. Barley Genetics Newsletter 26: 22-516

Mascher, M. et al. (2013), Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. Plant J. 76: 494-505

Mascher, M. et al. (2014), Mapping-by-sequencing accelerates forward genetics in barley, Genome Biol. 15: R78

Miura, K. et al. (2002), Mapping quantitative trait loci controlling seed longevity in rice (*Oryza sativa* L.), Theor. Appl. Genet. 104: 981-986

Moore, G. et al. (1995), Cereal genome evolution - Grasses, line up and form a circle. Current Biology 5: 737-739

Nagel, M. et al. (2009), Seed conservation in *ex situ* genebanks - genetic studies on longevity in barley. Euphytica 170: 5-14

Natoli, A. et al. (2002), Identification of QTLs associated with sweet sorghum quality. Maydica 47: 311-322

Nguyen, T.P. et al. (2012), Natural variation for seed longevity and seed dormancy are negatively correlated in *Arabidopsis*. Plant Physiol. 160: 2083-2092

Nguyen, T.P. (2014), Seed Dormancy and Seed Longevity: from genetic variation to gene identification, Utrecht University Repository (Dissertation), ISBN: 978-90-393-6094-1

Pourkheirandish, M. and Komatsuda, T. (2007), The importance of barley genetics and domestication in a global perspective, Ann. Bot. 100: 999-1008

Prioul, J.L. et al. (1997), Dissecting complex physiological functions through the use of molecular quantitative genetics. J. Exp. Bot. 48: 1151-1163

Richardson, T. et al. (2014), Efficient *Agrobacterium* transformation of elite wheat germplasm without selection. Plant Cell Tissue Organ Cult. 119: 647-659

Singh, R.K. et al. (2008), SSR markers associated with seed longevity in soybean. Seed Science Technology 36: 162-167

Shan, Q.W. et al. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. Nature Protocols 9: 2395-2410

Tanksley, S.D. (1993) Mapping polygenes. Annu. Rev. Genet. 27: 205-233

The International Barley Genome Sequencing Consortium (2012), A physical, genetic and functional sequence assembly of the barley genome. Nature 491: 711-716

Tingay, S. et al. (1997), *Agrobacterium tumefaciens*-mediated barley transformation. Plant J. 11: 1369-1376

Xie, K.B., and Yang, Y.N. (2013), RNA-Guided Genome Editing in Plants Using a CRISPR/Cas System. Molecular Plant 6: 1975-1983

Yu, J. et al. (2002), A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 296: 79-92

Acknowledgements

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet. Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass sie abgesehen von den unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die RNA-Sequenzierung der in der Arbeit verwendeten Linien erfolgte im Genomzentrum des Max-Planck-Instituts für Pflanzenzüchtungsforschung (MPIPZ) in Köln. Die Proteomanalyse von Samen der Elternlinien und nahezu isogenen Linien (NILs) erfolgte ebenfalls am MPIPZ durch die Arbeitsgruppe von Prof. Dr. Iris Finkemeier. Die Metabolomanalyse von Samen derselben Linien wurde in Zusammenarbeit mit der Arbeitsgruppe von Prof. Dr. Ilse Kranner (Universität Innsbruck, Österreich), unter der Betreuung von Dr. Erwann Arc, durchgeführt. Floureszenzmikroskopische Untersuchungen von roGFP2 exprimierenden transgenen *Arabidopsis*-Linien wurden unter Betreuung von Dr. Markus Schwarzländer (Universität Bonn) durchgeführt.

Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Maarten Koornneef betreut worden.

Ort, Datum   Unterschrift

_____   _____

Lebenslauf

Name: Dorothee Charlotte Wozny

Geburtsdatum/-ort: 10.01.1988 in Heidelberg

Staatsangehörigkeit: deutsch

Familienstand: ledig

Ausbildung

01/2013 – 10/2016 **Promotionsstudium** an der Universität zu Köln, Deutschland
Max-Planck-Institut für Pflanzenzüchtungsforschung Köln, unter der Leitung von Prof. Dr. Maarten Koornneef

10/2010 – 07/2012 **Masterstudium (MSc) in Biologie**
Technische Universität München (TUM)

10/2010 – 09/2011 **Erasmusjahr, Master in Agrarwissenschaftlicher Biotechnologie**
Polytechnische Universität Madrid (UPM)

10/2007 – 07/2010 **Bachelorstudium (BSc) in Biologie**
Technische Universität München (TUM)

06/2007   **Erhalt des Abiturs**
Dom-Gymnasium Freising

2002 – 2007   **Besuch des Gymnasiums**
Dom-Gymnasium Freising

1998 – 2002   **Besuch des Gymnasiums**
Gymnasium Weilheim, Weilheim i. OBB

1994 – 1998   **Besuch der Grundschule**
Grundschule Weilheim, Weilheim i. OBB

Ort, Datum   Unterschrift

_____   _____