



**HAL**  
open science

# Physiological and molecular studies during the acquisition of longevity in soybean (*Glycine max* L. Merrill) seeds

Juliana Lima

► **To cite this version:**

Juliana Lima. Physiological and molecular studies during the acquisition of longevity in soybean (*Glycine max* L. Merrill) seeds. Biochemistry, Molecular Biology. Agrocampus Ouest, 2016. Breton. NNT : 2016NSARI076 . tel-01651580

**HAL Id: tel-01651580**

**<https://theses.hal.science/tel-01651580>**

Submitted on 29 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# AGRO CAMPUS

---

## OUEST

**Juliana Pereira LIMA • 29 avril 2016**

Thèse AGROCAMPUS OUEST  
sous le label de l'Université Bretagne Loire  
pour obtenir le grade de

**DOCTEUR D'AGROCAMPUS OUEST**

Spécialité Biochimie - Biochimie moléculaire et cellulaire

ÉCOLE DOCTORALE • Végétal, environnement, nutrition,  
agroalimentaire, mer (VENAM)

LABORATOIRE D'ACCUEIL • UMR 1345 Institut de Recherche  
en Horticulture et Semences (IRHS)

THÈSE EN COTUTELLE : Universidade Estadual Paulista  
Júlio de Mesquita Filho (Brésil)

**Étude physiologique et  
moléculaire de l'acquisition  
de la longévité des graines  
de soja (*Glycine max* L. Merill)  
pendant leur développement**

**Julia BUITINK**

Directrice de Recherche, INRA Angers, France /  
*présidente de jury*

**José Marcio Rocha FARIA**

Professeur, Universidade Federal de Lavras, Brésil /  
*rapporteur*

**Henk W.M. HILHORST**

Professeur associé, University of Wageningen, Pays-Bas /  
*rapporteur*

**Olivier LEPRINCE**

Professeur, Agrocampus Ouest Angers, France /  
*directeur de thèse*

**Edvaldo Aparecido Amaral DA SILVA**

Professeur, Universidade Estadual Paulista Julio de Mesquita  
Filho, Brésil /  
*directeur de thèse*

N° ordre :  
N° série :

69 p.

## **Thèse AGROCAMPUS OUEST**

sous le label de l'Université de Bretagne Loire pour obtenir le grade de DOCTEUR

D'AGROCAMPUS OUEST

Spécialité : Biochimie-biologie moléculaire et cellulaire

Ecole Doctorale : Végétal, Environnement, Nutrition, Agroalimentaire, Mer (VENAM)

Laboratoire d'accueil : UMR 1345 Institut de Recherche en Horticulture et Semences (IRHS)

présentée par

**Juliana Pereira LIMA**

Etude physiologique et moléculaire de l'acquisition de la longévité des graines de soja (*Glycine max* L. Merrill) pendant leur développement.

**Soutenu à Botucatu (Brésil) le : 29 avril 2016 devant le jury composé de**

M José Marcio Rocha FARIA	Professeur, Universidade Federal de Lavras, Brésil	Rapporteur
M Henk W.M. HILHORST	Professeur associé, University of Wageningen, Pays-Bas	Rapporteur
Mme Julia BUITINK	Directrice de Recherche, INRA Angers, France	Examinatrice
M Olivier LEPRINCE	Professeur, Agrocampus Ouest Angers, France	Directeur de thèse
M Edvaldo Aparecido Amaral DA SILVA	Professeur, Universidade Estadual Paulista "JULIO DE MESQUITA FILHO", Brésil	Directeur de thèse

UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”  
FACULDADE DE CIÊNCIAS AGRONÔMICAS  
CAMPUS DE BOTUCATU

**JULIANA JOICE PEREIRA LIMA**

Supervisor : Prof. PhD. Edvaldo Aparecido Amaral da Silva

Co-supervisor : Prof. PhD. Olivier Leprince

Thesis submitted to the Faculty of  
Agricultural Sciences, UNESP – Botucatu,  
to obtain the title of Doctor of Agronomy  
(Agriculture).

BOTUCATU – SP

April – 2016



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu



**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA TESE: ESTUDOS FISIOLÓGICOS E MOLECULARES DURANTE A AQUISIÇÃO DA LONGEVIDADE EM SEMENTES DE SOJA (*Glycine max* (L.) Merrill)

**AUTORA: JULIANA JOICE PEREIRA LIMA**

**ORIENTADOR: EDVALDO APARECIDO AMARAL DA SILVA**

Aprovada como parte das exigências para obtenção do Título de Doutora em AGRONOMIA (AGRICULTURA), pela Comissão Examinadora:

Prof. Dr. EDVALDO APARECIDO AMARAL DA SILVA  
Dep de Produção e Melhoramento Vegetal / Faculdade de Ciências Agrônomicas de Botucatu

Prof. Dr. HENK W. M. HILHORST  
Wageningen University

Prof. Dr. JOSÉ MÁRCIO ROCHA FARIA  
Dep de Ciências Florestais / Universidade Federal de Lavras

Dra. JULIA BUITINK  
INRA

Dr. OLIVIER H. L. LEPRINCE  
Agrocampus Ouest

Botucatu, 29 de abril de 2016

**Yesterday is an illusion, the future does not exist, only the present is real**

*Byron Katie*

*To God, for supporting me*

## ACKNOWLEDGEMENTS

I ask permission to break the protocol and writing what is in my heart.

I would like at first thank God for giving me the opportunity to do a PhD.

To all my family, especially my brothers Joana and Felipe for their support and unconditional love throughout this trajectory. Your love was essential in every single moment of this journey! *In memoriam*, I thank my parents, I'm sure they are with me all the time.

Claudia and Leonel, you are like real parents and you deserve a special place in this moment. You supported me and welcomed me when I needed angels. Because of you the doctorate became lighter and cool. I'll take you for all my life! Thank you for the friendship, love and support.

To friends that even at a distance were so present; Marcella Nunes, Lauana Muniz, Erica Oliveira, Carol Peixoto, Juliana Stracieri, Pedro Maranhã, Hugo Haggam, Thati Menezes. To my new friends who have recently arrived but it seems we met long ago: Renan Cruz, Luis Eduardo, Mirella! You guys make me laugh a lot! When I was almost freaking out you brought me joy and peace. Juliana Iassia and Rita Camila you are awesome girls! Thank for all talking, support and 'prosa'. Paulete (Paula Caroline) thank you for supporting and advising me when we were at lab working until late of night and weekends! Magnus, you is awesome, my scientific partner, Thanks for joining me!

Dear Deise, what would it be of me without the "*divan*"? I could not have finished this final stage of the doctorate without you. Your support, patience and friendship were essential. I would not have physical and spiritual conditions without you. You are light in my life and in the lives of who pass on your way! To Hamid by the huge support the translation of the thesis into French.

Thank you also to all of the "Seed Analysis Laboratory" in Botucatu, for being present to assist in the performance of the experiments, scientific and personal conversations. Val, without you the lab is never the same! Thanks all from lab, Daiani, Victória, Letícia, Andrea, Gilberto, Barbara, Pedro, Tiago, Alexandre, Renake, Maria Rita. Ju Bravo and Julio, a big thank you both for all scientific discussion and patience! Special thanks to Rubiana, with you I learned how to become a better person. Thanks for all the support and care no matter the circumstance.

I thank my supervisors, Prof. Edvaldo Aparecido Amaral da Silva (Brazil) and Prof. Olivier Leprince (France). You were more than supervisors, you were friends and great teachers who inspire me daily. Gratitude for all teaching, patience and support and for being an example of good professional and person. I know, I'm not easy. You were my masters and Motivation at this stage. Thank you very much !

I enjoy this opportunity to thank also to Dr. Julia Buitink, INRA scientist and head of the Coserto team. Thank you for your teachings, professionalism and learning. It will be forever in my memory and in my heart. You were always available to help me to solve my doubts (which were not few). You are a leader like few others, you can inspire and motivate those who work with you. Your passion for research is contagious. Thank you !

I thank with affection to all the people from 'Institut de Recherche en Horticulture et Semences' laboratory for welcome me so well and for helping me to work during my stay in Angers. A big thank to IRHS team, especially Aurelia, Daniel, Pascale and Coserto team Benoit, David L., Joseph, Julia pétite. You have always been very kinds, cheerfult, patientes and in addition, you understood my difficulties in adapting to the new country and with French language. It was a joy to work with you. Lucien, my office mate! Thank you for every laugh, advice, French grammar correction and all patience. Shivan, you were very kind and understanding. When I was discouraged you could make me have faith again, thank you!

My final thanks, but no less important is for you Chance. You came into my life at the perfect time. Thank you very much for all conversation, support, laughs, encouragement, patience, kindness, friendship, love and also for keeping the faith on us. You are unbelievable ! I'm glad for having you in my life.

I am immensely grateful for the gift that was to meet each of you on this journey! I have no words to express how much I learned from each one. I wish you receive all the good things that you all brought to my life.

Thank you so much !

## AGRADECIMENTOS

Eu gostaria de agradecer primeiramente à Deus por ter me concedido à oportunidade de fazer o doutorado.

À toda minha família, principalmente meus irmãos Joana e Felipe, pelo apoio e amor incondicional e constante durante toda essa trajetória. O amor de vocês foi essencial em cada momento dessa jornada ! *In memoriam*, Agradeço a meus pais, tenho a certeza que estão sempre comigo.

Claudia e Leonel, vocês são como verdadeiros pais de coração para mim e vocês merecem um lugar especial nesse agradecimento. Vocês me apoiaram e acolheram quando eu mais precisei de anjos. Por causa de vocês o doutorado se tornou mais leve e legal. Obrigada por todo aprendizado, amor, apoio e convivência.

Aos amigos que, mesmo à distância estavam tão presentes; Marcella Nunes, Lauana Muniz, Erica Oliveira, Carol Peixoto, Juliana Stracieri, Pedro Maranhã, Hugo Hagmam, Thati Menezes. Aos meus novos amigos que chegaram recentemente, mas parece que nos conhecemos há muito tempo: Renan Cruz, Luis Eduardo, Mirella! Vocês me fazem rir muito! Quando eu estava quase enlouquecendo você me trouxeram alegria e paz. Juliana Iassia e Rita Camila você são impressionantes meninas! Obrigado por toda conversa, apoio e "prosa". Paulete (Paula Caroline) obrigado pelo apoio e conselhos quando estávamos no laboratório até tarde da noite e fins de semana! Magnum, você é incrível, meu parceiro científico, Obrigado por se juntar a mim!

Deise, o que seria de mim sem o divã? Eu não poderia ter terminado essa fase final do doutorado sem você. Seu apoio, paciência e amizade foram essenciais. Eu não teria condições físicas e espirituais sem você. Você é um exemplo de fé e humanidade ! Ao Hamid pelo enorme suporte na tradução de parte da tese em francês.

Agradeço também a todos do "Laboratório de Análise de Sementes" em Botucatu, para estar presente para ajudar no desempenho dos experimentos, nas conversas científicas e pessoais. Val, sem você o laboratório nunca é o mesmo! Obrigado a todos do laboratório, Daiani, Victória, Letícia, Andrea, Gilberto, Barbara, Pedro, Tiago, Alexandre, Renake, Maria Rita. Ju Bravo e Julio, um grande obrigado por toda a discussão científica e paciência! Um obrigado especial à Rubiana, com você eu aprendi a me tornar uma pessoa melhor. Obrigada por todo apoio e cuidado não importasse a circunstância!

Agradeço aos meus supervisores, Prof. Edvaldo Aparecido Amaral da Silva (Brasil) e prof. Olivier Leprince (França). Vocês foram mais que supervisores, foram amigos e grandes mestres que me inspiram diariamente. Gratidão por todo o ensino, paciência e apoio e por ser um exemplo de bons profissionais e pessoas. Eu sei, eu não sou fácil. Vocês foram meus mestres e inspiradores nesta fase. Muito obrigada !

Aproveito esta oportunidade para agradecer também a Julia Buitink, INRA cientista e chefe da equipe de Conserto. Obrigado por seus ensinamentos, o profissionalismo, a aprendizagem. Ele estará para sempre em minha memória e no meu coração. Você estava sempre disponível para ajudar a resolver as minhas dúvidas (que não eram poucos). Você é um líder como poucos, você pode inspirar e motivar as pessoas que trabalham com você, com reponsabilidade. Sua paixão pela pesquisa é contagiosa. Obrigado !

Agradeço com afeto a todos do laboratório 'Institut de Recherche en Horticulture et Semences' por me receber tão bem e por me ajudar a trabalhar durante a minha estadia em Angers. Um grande obrigado à todos do IRHS, especialmente Aurelia, Daniel, Pascalle e à equipe Conserto Benoit, David Lallane, Joseph, Julia petite. Vocês sempre foram bondosos, alegres e pacientes, além disso, vocês entenderam a minha dificuldade em se adaptar ao novo país e com a língua francesa. Foi uma alegria trabalhar com vocês. Lucien, meu companheiro de escritório! Obrigado por cada riso, aconselhamento, correção de gramática francesa e toda a paciência. Shivan, você foi muito gentil e compreensivo. Quando eu estava desanimada você conseguia me fazer ter fé outra vez, obrigado!

Meu agradecimento final, mas não menos importante é para você Chance. Você entrou na minha vida no momento perfeito ! Obrigada por toda conversa, apoio, risadas, compreensão, incentivo, paciência, amizade, amor, e também por manter a fé em nós. Você é inacreditável !!

Estou imensamente grata pelo presente que foi conhecer a cada um de vocês nesta jornada! Não tenho palavras para expressar o quanto eu aprendi com cada um. Desejo que vocês recebam todas as coisas boas que trouxeram para a minha vida.

Muito obrigado !

## SOMMAIRE

Version abrégée en français .....	1
Introduction .....	10
<b>Chapter I. Literature review .....</b>	<b>12</b>
I.1. Importance of soybean in Brazil .....	12
I.2. Seed development and maturation.....	14
I.2.1. Seed filling and formation.....	14
I.2.2. Acquisition of desiccation tolerance .....	15
I.2.3. Acquisition of longevity .....	16
I.3. Molecules and mechanisms involved in dry stage.....	18
I.3.1. Non-reducing sugars .....	19
I.3.2. LEA proteins .....	20
I.3.3. Heat Shock Proteins .....	21
I.3.4. Protection against oxidative stress .....	22
<b>Chapter II. Materials and methods.....</b>	<b>25</b>
II.1. Plant Material .....	25
II.2. Field experiment .....	25
II.3. Sampling points .....	25
II.4. Seed quality assessment .....	26
II.5. Determination of soluble sugars .....	26
II.6. RNA extraction and sequencing .....	27
II.7. Sequences alignment .....	28
II.8. Transcriptome data analysis .....	28
II.9. Gene Ontology analysis .....	29
II.10. Quantitative real time RT-qPCR .....	29
<b>Chapter III. Results .....</b>	<b>31</b>
III.1. Physiological characterization of soybean seeds maturation .....	31
III.1.1. Seed filling and drying .....	32
III.1.2. Acquisition of desiccation tolerance and longevity.....	32
III.2. Changes in soluble sugars during soybean seed maturation .....	36

III.3. Dynamic changes in the transcriptome during late maturation phase of soybean seed.....	39
III.4. Differential expression analysis of genes associated with longevity....	44
<b>Chapter IV. Discussion</b> .....	49
IV.1. Acquisition of longevity and non-reducing soluble sugars accumulation are related in soybean seeds.....	49
IV.2. Gene profiling during maturation of soybean seeds.....	51
IV.3. Transcription factors and longevity in soybean seeds.....	53
<b>Chapter V. Conclusions</b> .....	57
References bibliographiques .....	58

## TABLES

**Table 1.** Characterization of the phenological stages of soybean plants according to Fehr and Caviness (1977) with adaptations

**Table 2.** Days after flowering at different phenological stages of soybean seeds cultivated in the year 2013/2014

**Table 3.** Primers sequences used for the target gene study

**Table 4.** Number of reads per line obtained from total RNA extracted from seeds harvested at the indicated stages during maturation. A and B correspond to biological replicates

**Table 5.** Mapping of single-end reads to the soybean genome. # reads, number of reads following trimming of the libraries for quality; # mapped reads, number of reads that unambiguously mapped to the soybean genome; # unmapped, number of reads that could not be mapped to a single location in the soybean genome. Percentages of unmapped and mapped sequences are also indicated (relative to number of reads)

**Table 6.** Representation of families of transcription factors whose gene expression were clustered using K-means

**Table 7.** Number of differentially expressed genes in seeds during late seed maturation by comparing stage R7.2 and R9

**Table 8.** GO enrichment analysis of DEG that increased with longevity. Analysis was performed with transcripts exhibited a FC R9/R7.2 ratio > 1 using AgriGO with the Arabidopsis TAIR9 background

**Table 9.** GO enrichment analysis of DEG that decrease with longevity. Analysis was performed with transcripts exhibited a FC R9/R7.2 ratio < -1 using AgriGO with the Arabidopsis TAIR background

**Table 10.** GO enrichment analysis of DEG positively correlated with longevity at  $p < 0.001$  (1925 transcripts) using AgriGO with the Arabidopsis TAIR background

**Table 11.** GO enrichment analysis of DEG negatively correlated with longevity at  $p < 0.001$  (1889 transcripts) using AgriGO with the Arabidopsis TAIR background

**Table 12.** Pearson correlation coefficient (r value) of gene expression from RNAseq data and RT-qPCR

## FIGURES

**Figure 1.** Three different phases of development of orthodox seeds (phases I, II and III) and physiological maturity (PM). A – Evolution of the fresh weight (FW), dry weight (DW), water content (WC) during development. B – Acquisition of various aspects of physiological quality during development of orthodox seeds (according to Bewley et al., 2013)

**Figure 2.** Diagram illustrating the three types of profile of acquisition of longevity in post abscission of fruit and seed dispersion (Probert et al., 2007)

**Figure 3.** Maximum temperature (A) and relative humidity (B) at experimental field from sowing to harvest of soybean seeds in crop years 2012/2013 (open circles) and 2013/2014 (black triangle). The bars indicate sowing, beginning of flowering, start of seed filling (stage R5.1), period of acquisition of desiccation tolerance and longevity (from stage R7.1 to R9)

**Figure 4.** RNA integrity number (RIN) values from seed maturation of soybean from stage R7.1 to R9

**Figure 5.** Soybean pods and seeds during maturation at different phenological stages

**Figure 6.** Changes in dry weight (DW) and fresh weight (FW) (mg per seed) and water content (WC) (g H<sub>2</sub>O per g DW). Data are the means of four replications. Bars represent the standard error

**Figure 7.** Changes in water content (WC), fresh weight (FW) and dry weight (DW) from crop years 2012/2013 (open symbols) and 2013/2014 (black symbols) based on phenological stages. Data are the means of four replications of 20 seeds. Bars represent the standard error

**Figure 8.** Rate of water loss during fast drying at 40% RH and 30 °C (A) and slow drying at 45% RH and 25 °C (B). Seeds were harvested in 2012/2013. Data represent the mean of 3 replicates of 20 seeds. Dashed line represents the natural drying in soybean seed maturation. Full line represents the artificial drying time

**Figure 9.** Desiccation tolerance in developing soybean seeds after slow and fast drying at different phenological stages. Data represent the mean of 4 replicates of 25 seed

**Figure 10.** Germination (G, circles) and desiccation tolerance (DT, square) measured after slow drying at 45% RH and 25°C. Seeds were harvested in 2012/2013 (white symbols) and 2013/2014 (black symbols). Data represent the mean of 4 replicates of 25 seeds.

Bars represent the standard error Data are presented as a function phenological stages (A) and in days after flowering (B)

**Figure 11.** Changes in germination percentage of dried seeds harvested at indicated stages of maturation and stored at 35 °C and 75% relative humidity (RH) from crop years 2012/2013 (A) and 2013/2014 (B). Third-order sigmoid curves were fitted to the data to obtain the P50 values. Data are the mean of 3 replicates of 25 seeds. Bars represent the standard error

**Figure 12.** Desiccation tolerance (square) and Longevity (diamond) were expressed as the P50 corresponding to the time of storage (days) at which the dried seed lot lost 50% germination, for both crop years 2012/2013 (black symbols) and 2013/2014 (open symbols)

**Figure 13.** Relationship between the storage time and % of viability expressed in probit for seeds harvested at stages R7.2 and R7.3. Data were fitted with a linear regression according to Ellis & Roberts (1980). The slope  $\sigma$  corresponds to storage time to reduce the viability for 1 value of probit).  $K_i$  corresponds to the viability at storage time 0

**Figure 14.** Evolution of  $K_i$  (A) and Sigma (B) parameters determined by Ellis & Roberts (1980) model during maturation of soybean seeds from crop years 2012/2013 (open circles) and 2013/2014 (black triangle)

**Figure 15.** Changes in non-reducing sugars content during maturation of soybean seeds, from stage R6 to R9. Content of fructose (A), glucose (B), sucrose (C), raffinose (D), stachyose (E) and verbascose (F) in axes (black square) and cotyledons (open diamond) of fresh soybean seeds during development

**Figure 16.** Relation between longevity (p50) and monosaccharides (fructose + glucose) (A), sucrose content (B), ratio sucrose/stachyose (C) and sucrose/RFO (D) in axes (black square) and cotyledons (open diamond) of soybean seeds during maturation. Polynomial model and a linear equation were used

**Figure 17.** Principal component analysis of the transcriptomes during seed maturation. PCA was performed using median centering. Gray circles are replicates of stages R7.2 and R9

**Figure 18.** Over-representation analysis of functional classes during soybean seed development using the PageMan software. Data were subjected to a bin-wise Wilcoxon test, and resulting P-values were adjusted according to Benjamin Hochberg. Over-represented and under-represented classes are indicated in red and green, respectively.

Functional classes and subclasses statistically affected are indicated according to MapMan ontology

**Figure 19.** Clustering displaying the transcription factors during maturation of soybean seeds. The color scale ranges from saturated green for log ratios  $-3.0$  and below to saturated red for log ratios  $3.0$  and above. Each gene is represented by a single row of colored boxes; each phenological stage is represented by a single column. Eleven separate clusters are indicated by colored bars and by identical coloring of the corresponding region of the dendrogram

**Figure 20.** Relationship between P50 taken obtained from seeds collected between stages 7.1 and 9 and log ratios of up-regulated (A) and down-regulated genes (B). Data shown corresponds to the top 5 probes exhibiting both the highest correlation coefficient and expression levels

**Figure 21.** Relative gene expression for RNA-Seq data validation at respective seed developmental stages for genes encoding (A) and (B) 17.6 kDa class II Heat shock protein (HSP 17.6: Glyma17G224900 and Glyma14G000009, respectively); (C) Heat shock protein 101 (HSP101: Glyma17G077500); (D) Heat shock protein 21 (HSP21: Glyma08G318900). Bars and line indicate expression values from qPCR and RNA-Seq experiments, respectively

**Figure 22.** Relative gene expression for RNA-Seq data validation at respective seed developmental stages for genes encoding (A) Heat shock transcription factor A3 (HSFA3: Glyma03G191100); (B) Heat shock transcription factor A6B (HSFA6: Glyma03G157300); (C) bZIP transcription factor family protein (bZIP65: Glyma03G142400); (D) histone acetyltransferase of the GNAT family 1 (GNC5: Glyma.19G160700); (E) NF-X-like 1 (NFXL1: Glyma.09G173000). Bars and line indicate expression values from RT-qPCR and RNA-Seq experiments, respectively

## **ABSTRACT**

Soybean is one of the most important oil crop species for food and feed as well as a range of industrial applications. Producing highly vigorous seeds is a key lever to increase crop production. In soybean, low physiological seed quality is more likely to occur under tropical environment and leads to poor stand establishment and decreased yields. Longevity (i.e. the ability to survive in the dry state for prolonged periods of time) represents an important trait characterizing seed vigor. However, how it is acquired during seed maturation is not known. Here, the objective was to obtain insights into the mechanisms regulating the progressive acquisition of longevity in soybean. Longevity was acquired shortly after seed filling and acquisition of desiccation tolerance (stage R7.2) and continued to increase thereafter. Differences in longevity were found between the 2013 and 2014 harvests. They were related to differences in high temperatures during seed formation, a known factor influencing seed quality in soybean. Using Illumina sequencing, transcriptomes were obtained from 7 stages during the acquisition of longevity. Complex molecular changes occurring at the transcriptional level were observed after R7.2, despite that it is considered as fully mature. GO enrichment analyses revealed a positive association between longevity and RNA processing and modification processes and a negative association with photosynthesis biological process. Also, acquisition of longevity was associated with an increased amount of transcripts encoding HSF and HSP as well as transcription factors related to biotic defense. Our work highlights putative markers that could be used to monitor seed longevity in soybean seed production.

## **RESUMO**

Soja é uma das mais culturas oleaginosas usadas para alimentação animal e humana bem como para uma larga aplicação industrial. Produção de sementes altamente vigorosos é uma alavanca fundamental para aumentar a produção agrícola. Na soja, a baixa qualidade fisiológica das sementes é mais provável de ocorrer em ambiente tropical e leva ao mal estabelecimento do estande e diminuição da produtividade. Longevidade (isto é a capacidade de sobreviver no estado seco durante períodos prolongados de tempo) representa uma importante característica distintiva no vigor das sementes. No entanto, como é adquirida durante a maturação das sementes não é conhecido. Aqui, o objetivo foi obter insights sobre os mecanismos que regulam a aquisição progressiva de longevidade em soja. Longevidade

foi adquirida pouco depois de enchimento de grãos e aquisição de tolerância à dessecação (estádio R7.2) e continuou a aumentar em seguida. Diferenças na longevidade foram encontradas entre os anos agrícolas 2013 e 2014. Eles foram relacionadas com diferenças de temperaturas elevadas durante a formação das sementes, um factor conhecido que influencia a qualidade da semente em soja. Usando sequenciamento da Illumina, transcriptomas foram obtidos a partir de 7 estádios durante a aquisição de longevidade. Foram observadas alterações moleculares complexas que ocorrem em nível transcricional após R7.2, apesar de que ele é considerado como totalmente maduro. Análises de enriquecimento de GO revelou uma associação positiva entre longevidade e processamento do RNA e processos de modificação e uma associação negativa com processo biológico fotossíntese. Além disso, a aquisição de longevidade foi associado com um aumento da quantidade de transcritos que codificam HSF e HSP, bem como fatores de transcrição relacionados com a defesa biótica. O nosso trabalho destaca marcadores putativos que podem ser utilizados para monitorar a longevidade na produção de sementes de soja.

## **RÉSUMÉ**

La production de semences très vigoureuses est un levier clé pour augmenter les rendements agricoles. Chez le soja, les risques de production de semences de faible vigueur germinative sont élevés sous les climats tropicaux, ce qui conduit à une mauvaise installation de la culture, voire un déclassement des lots en grain. La longévité (c-a-d la capacité de survivre à l'état sec pendant de longues périodes) représente une caractéristique importante de la vigueur germinative. Cependant, la façon dont elle est acquise au cours de la maturation de la graine n'est pas connue. L'objectif de la thèse était d'obtenir un aperçu des mécanismes de régulation de l'acquisition de la longévité chez le soja. Celle-ci est acquise peu après la phase de remplissage et l'acquisition de la tolérance à la dessiccation (stade R7.2) et continue d'augmenter jusqu'en fin de maturation. Des différences de longévité ont été observées entre les récoltes 2013 et 2014. Elles seraient liées aux températures élevées au cours de la formation des graines, un facteur connu pour impacter la qualité des semences. Le séquençage Illumina de graines récoltées à 7 stades de développement a révélé des changements complexes se produisant au niveau transcriptionnel dès le stade R7.2, pourtant considéré comme correspondant à la maturité physiologique. Les analyses d'enrichissement GO ont révélé une association positive entre longévité et les processus de maturation des

ARN et régulateurs de la post-transcription et une association négative avec les processus liés à la photosynthèse. L'acquisition de la longévité est aussi associée à une augmentation des transcrits codant pour des HSF et HSP, ainsi que des facteurs de transcription liés à la défense biotique. Notre travail met en évidence des marqueurs putatifs qui pourraient être utilisés pour évaluer la longévité en production de semences de soja.

# Version abrégée en français

## Introduction

La production de graines de légumineuse est stratégiquement et économiquement importante pour la sécurité alimentaire afin d'assurer la croissance de la consommation mondiale de produits carnés. Le soja (*Glycine max*) est l'une des plus importantes espèces cultivées dans le monde avec le riz, le blé, l'orge et le maïs. La graine de soja est riche en protéines et en huiles (LEE et al., 2015), environ 38% et 20% respectivement. Compte tenu de sa capacité à utiliser l'azote atmosphérique, le soja est primordial pour le développement d'une agriculture durable. Au Brésil, la surface des cultures du soja cultivée était de 33,3 millions d'hectares en 2015/2016, représentant une production de graines de plus de 100 millions de tonnes (USDA, 2016), soit 46% du total des graines produites (CONAB, 2016).

L'agriculture moderne exige l'usage de technologies pour optimiser le rendement de la plante. Un premier levier d'action est la production de semences vigoureuses. Les caractéristiques souhaitables en termes de vigueur sont sa capacité à rester vivante pendant le stockage à sec (longévité), une germination élevée, rapide et synchrone même dans des conditions pénalisantes. La vigueur germinative est une caractéristique complexe déterminée par le génome, l'environnement et les interactions génotype x environnement, qui demeurent encore très méconnues. Pour le soja qui pousse dans un environnement tropical, la faible qualité germinative des graines reste un obstacle récurrent qui mène à un pauvre établissement des plantes et donc à une diminution du rendement. Pour atteindre une production stable de graines à haute vigueur, il est nécessaire d'acquérir de nouvelles connaissances sur les mécanismes qui régissent l'acquisition des caractéristiques qui sous-tendent la qualité germinative.

L'objectif de cette étude est de décrire les étapes du développement de la graine pendant lesquelles la longévité est acquise et d'acquérir des connaissances sur les mécanismes moléculaires qui régulent cette acquisition. Pour cela, nous avons caractérisé le transcriptome après séquençage haut-débit RNA-Seq associé à l'acquisition de la longévité de la graine pendant son développement. Nous avons aussi étudié l'évolution de sucres solubles non-réducteurs qui sont impliqués dans le mécanisme de survie à l'état sec.

**Table 1.** Caractérisation des stades phénologiques des graines de soja selon la classification originale de Fehr et Caviness (1977) et adaptée dans ce travail

<b>Stades</b>	<b>Dénomination</b>	<b>Description</b>	<b>Jours après la floraison</b>
R <sub>5,1</sub>	Remplissage du grain	Graines perceptibles au toucher au travers de la gousse, jusqu'à 10% de remplissage des grains	25
R <sub>5,2</sub>	Remplissage du grain	Taux remplissage de 10% à 25%	25
R <sub>5,3</sub>	Remplissage du grain	Taux de remplissage entre 25% et 50%	27
R <sub>5,4</sub>	Remplissage du grain	La Plupart des gousses presentent des graines avec un taux de remplissage entre 50% et 75% de remplissage des grains	27
R <sub>5,5</sub>	Remplissage du grain	Plupart des gousses entre 75% et 100% de remplissage des grains	36
R <sub>6</sub>	Grains développés	Gousses avec 100% de remplissage des grains et des feuilles vertes	47
R <sub>7,1</sub>	Maturation précoce	Jusqu'à 50% des feuilles et les gousses sont jaunes	57
R <sub>7,2</sub>	Maturation précoce	Entre 51% et 75% des feuilles et gousses sont jaunes	63
R <sub>7,3</sub>	Moitié de la maturation	Plus de 76% des feuilles et des gousses sont jaunes	69
R <sub>8,1</sub>	Maturation	Jusqu'à 50% de défoliation	71
R <sub>8,2</sub>	Maturation	Plus de 50% de défoliation avant la récolte	73
R <sub>9</sub>	Maturité de récolte	Le point de recolte	77

## Chapitre 1. Synthèse bibliographique

Les graines acquièrent leur longévité pendant la maturation, après la phase de remplissage (HAY & PROBERT, 1995; SANHEWE & ELLIS, 1996; SINNIAH et al., 1998b; CHATELAIN et al., 2012; RIGHETTI et al., 2015). Durant la maturation de la graine de *Medicago truncatula*, la longévité débute une fois que l'accumulation de réserve est terminée et elle augmente progressivement jusqu'à l'abscission du fruit (CHATELAIN et al., 2012; VERDIER et al., 2013; RIGHETTI et al. 2015). La longévité est une caractéristique complexe affectée par des facteurs environnementaux durant la formation, la récolte et le stockage de la graine, et est généralement contrôlée par de nombreux gènes (CLERKX et al. 2004; DARGAHI et al. 2014). Les mécanismes qui confèrent la longévité des graines ont reçu une attention majeure. Ils incluent l'accumulation de sucres non réducteurs, des protéines LEA et HSP, qui offrent une protection contre la dénaturation et des changements délétères dans la conformation de macromolécules (CHATELAIN et al., 2012; BUITINK & LEPRINCE, 2008). La longévité est aussi conférée par une gamme d'antioxydants qui protègent les graines contre les dommages oxydatifs durant le stockage. Toutefois, les mécanismes régulateurs qui contrôlent l'acquisition progressive de la longévité de la graine demeurent largement inconnus.

Le séquençage haut-débit (RNA-Seq), sont de puissantes techniques pour mettre en évidence les genes exprimés de manière différentielle pendant le développement. (PONJAVIC et al., 2007; YANG et al., 2014), le soja inclus (JONES & VODKIN, 2013). Cependant, ces études se concentrent principalement sur l'embryogenèse et le remplissage de la graine, ne prenant pas ainsi en compte les phases tardives et de la maturation de la graine et l'induction de la longévité.

## Chapitre 2. Matériel et méthodes

L'expérience a été réalisée pendant deux campagnes consécutives en 2012/13 et 2013/14. La zone expérimentale appartient à l'Université d'Etat de Sao Paulo - Université des Sciences Agricoles, UNESP -FCA (Botucatu, SP – Brésil, 22.8842° S, 48.4442° W, climat subtropical). Les graines ont été semées les 4 Décembre 2012 et 5 Décembre 2013 respectivement. Les plantes de soja ont été cultivées à maturation complète et des bourgeons de fleurs ont été marqués juste au début de la floraison. Ensuite, les fruits



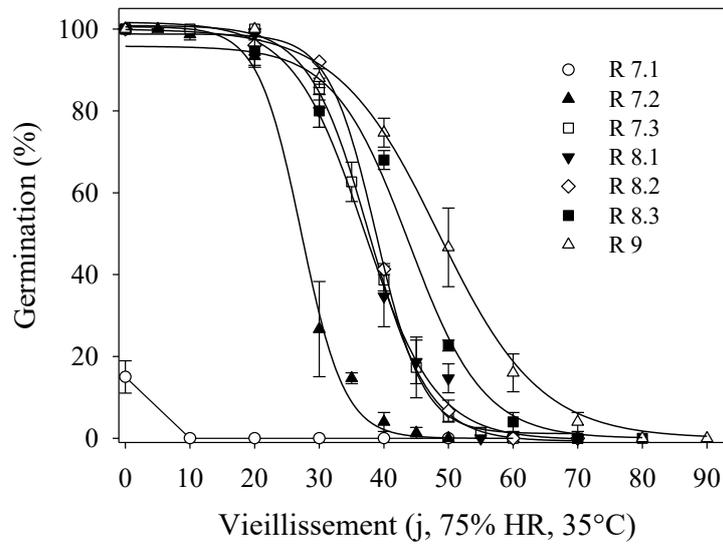
**Figure 1.** gousses de soja et de graines à différents stades phénologiques. Les barres blanches représentent 1 cm.

étaient retirés manuellement des plantes et les graines ont été extraites. Les gousses et les graines ont été récoltées à intervalles différents, basés sur des stades phénologiques suggérés par Fehr & Caviness (1977) avec des adaptations (Table 1) et sur le nombre de jours après floraison (Table 2).

Pour évaluer la qualité physiologique durant la maturation de la graine, la germination, la tolérance à la dessiccation et la longévité ont été mesurées. Pour étudier la longévité, des graines immatures ont été séchées rapidement puis stockées à 75% HR and 35 ° C. Ces conditions représentent un bon compromis entre le vieillissement naturel et le vieillissement accéléré où les hautes températures et l'humidité relative provoquent des effets délétères qui ne sont pas identiques au vieillissement naturel. La longévité est déterminée par le nombre de jours de vieillissement pour diminuer la germination en 50%, la période de demi-viabilité, P50 (TANG et al., 1999, CHATELAIN et al., 2012).

La teneur en sucres solubles (raffinose, stachyose, verbascose, saccharose, fructose et glucose) a été mesurée dans les cotylédons et axes des graines. Les échantillons sont analysés par chromatographie ionique haute performance en phase liquide, en utilisant le système DIONEX. Les sucres sont détectés par ampérométrie après séparation sur une colonne CarboPac PA1 avec un débit constant à 0.25 mL/min de NaOH 0,1 M.

L'étude du transcriptome a été réalisée par séquençage de l'ARN extrait de graines des stades R7.1 à R9 en utilisant la plate-forme HiSeq2500 (Illumina) à l'Universidade de Campinas. A partir des données générées dans le séquençage, les lectures ont été alignées sur le génome de référence de soja 'Williams 82' avec le programme d'alignement Bowtie2. L'estimation de l'expression génique différentielle a été réalisée à l'aide DESeq2. Les données d'expression ont été normalisées en divisant la valeur de l'expression normalisée obtenue dans l'étape n + 1 par la valeur obtenue à l'étape n (ex. R7.2 / R7.1, R7.3 / R7.2, R8.1 / R7.3; ...; R9 / R8.3). La normalisation a été réalisée en utilisant la méthode "*normalize.quantiles.use*" sous R (R Core Team, 2015). La plupart des gènes importants ont été identifiés à 0,5% le taux de fausses découvertes (FDR) calculée en utilisant P-valeur ajustée pour les tests multiples en utilisant la méthode Benjamini-Hochberg (NELSON et al., 1998).



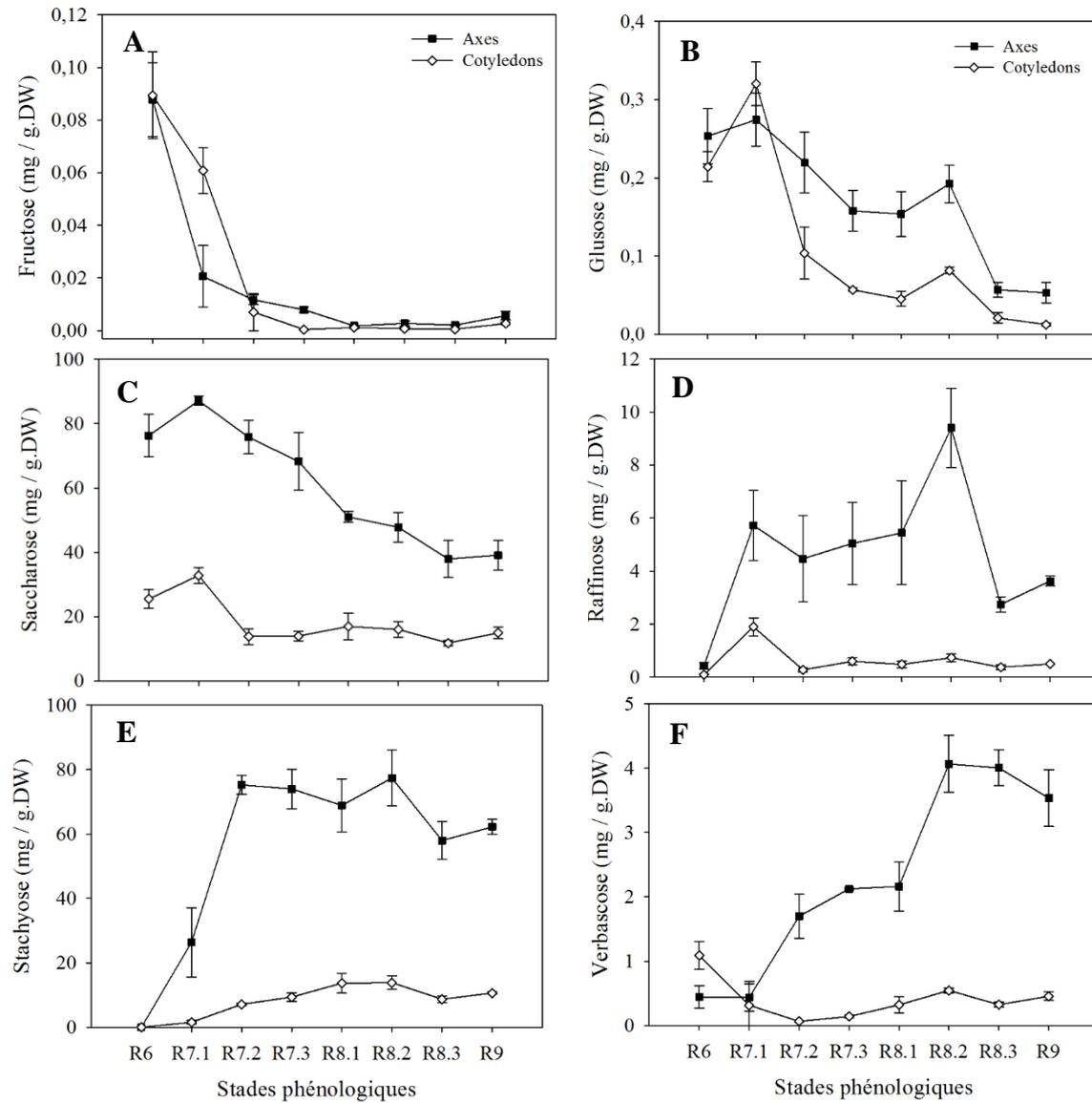
**Figure 2.** Evolution du pourcentage de germination des graines de stades de maturation pendant le stockage à 35 ° C et une humidité relative de 75% de l'année 2013/2014. Les courbes sigmoïdes troisième ordre ont été ajustées aux données pour obtenir les valeurs de P50. Les données sont la moyenne de 3 répétitions de 25 graines. Les barres représentent l'erreur standard.

### Chapitre 3. Résultats physiologiques et biochimiques

Le développement de la graine a été suivi depuis la fin du remplissage (stade 7.1) jusque l'obtention de graines matures (stade 9). Les gousses et les graines du stade R7.1 sont vertes puis la maturité de la masse qui correspond à l'accumulation maximum de poids sec se produit à environ R7.2. La teneur en eau diminue progressivement pendant le développement de la graine allant de 6.0 g H<sub>2</sub>O /g PS au stade R7.2 à 0.15 g H<sub>2</sub>O / g PS au stade 9. Lorsqu'elles sont isolées des gousses, les graines immatures commencent à germer à partir du stade R5.5 dans les deux années de récolte (Figure 1), ce qui équivaut à 36 jours après la floraison (Table 2). La capacité germinative est acquise au stade R7.1 tandis que la tolérance à la dessiccation est acquise au stade R7.2 (Figure 10 A).

Nous avons obtenu les courbes de perte de viabilité au cours du stockage à 35°C, 75 % HR pour les deux années 2012/2013 et 2013/2014 (Figure 2). Ces courbes servent à calculer la longévité des graines, défini comme le P50, c-a-d le temps nécessaire pour obtenir 50% de perte de viabilité du lot de semences. Plus les graines sont matures, plus leur durée de vie est importante durant le stockage. Nous avons découvert que la longévité des graines est déjà très importante au stade R7.2 (26-28 jours) puis continue d'augmenter 1,8 fois jusqu'au stade R9. Il convient de noter que le temps écoulé entre les étapes 7.1 et 7.2 sont approximativement de 10 jours. Plusieurs mécanismes peuvent être impliqués dans l'acquisition de la longévité de la graine, incluant l'accumulation des sucres non-réducteurs, qui sont aussi connus comme des molécules protectrices impliqués dans la tolérance de dessiccation. Cependant, leur rôle dans la longévité n'est pas encore complètement établi. Dans ce travail, nous avons étudié l'évolution de ces sucres pendant la phase de maturation des graines de soja et nous avons examiné une possible corrélation entre leur quantité et l'acquisition de la longévité de la graine. Cette étude a été réalisée sur des graines de soja récoltées entre les stades R6 à R9 pour l'année de récolte 2013/2014.

La teneur en glucose et fructose diminue de manière simultanée avec l'acquisition de tolérance de dessiccation (Figure 3 A et B). Ensuite, les niveaux étaient vraiment faibles au stade R9. Le glucose était plus abondant dans les axes par rapport aux cotylédons. Le saccharose était aussi plus abondant dans les axes que dans les cotylédons durant la période de maturation. Dans les axes, la teneur en saccharose diminue lentement à partir du stade R7.1. La teneur en fructose, glucose et saccharose est plus élevée quand la longévité est faible (étape R7.2). Comme pour le saccharose, les oligosaccharides de la



**Figure 3.** Evolution des sucres non-réducteurs pendant la maturation des graines de soja. Contenu en fructose (A), glucose (B), saccharose (C), raffinose (D), stachyose (E) et verbascose (F) sont indiqués pour les axes (carré noir) et des cotylédons (diamant blanc). Les données sont la moyenne de 3 répétitions. Les barres représentent la déviation standart.

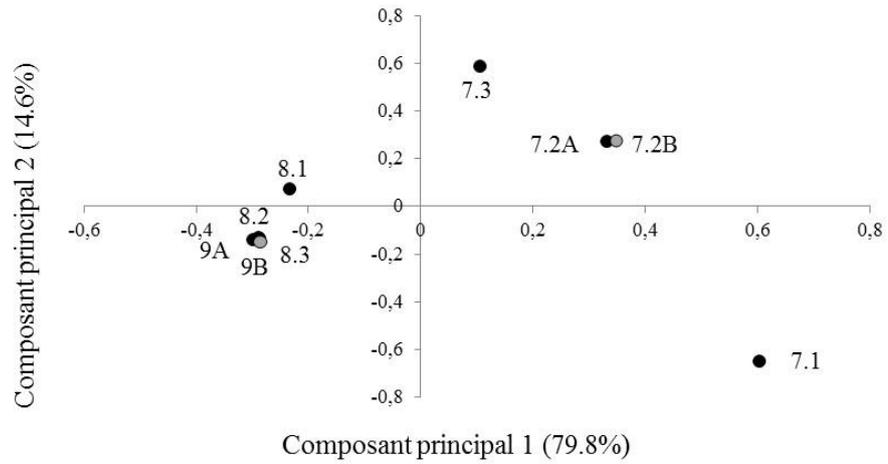
famille du raffinose (RFO – raffinose, stachyose et verbascose) étaient aussi plus abondants dans les axes que dans les cotylédons.

Le stachyose est le RFO le plus abondant. Dans l'axe, les concentrations de raffinose, précurseur du stachyose et du verbascose, augmentent du stade R6 à R7.1. Ils demeurent ensuite stables avant d'atteindre leur maximum au stade R8.1 pour diminuer ensuite au stade R9 (Figure 3 D). Le maximum de raffinose apparaît lors du stade R7.1 dans les cotylédons. Toutefois, son faible taux suggère une possible métabolisation rapide en stachyose durant la maturation. Dans les axes, le stachyose est détecté au stade R7.1. Sa teneur augmente ensuite considérablement dès le stade R7.2 et atteint une valeur constante d'environ 73 mg/g PS au stade R9, c-a-d lorsque la longévité a augmenté de près de deux fois. Dans les cotylédons, la teneur en stachyose augmente régulièrement jusqu'au stade R8.1 (Fig. 3 E) puis demeurent stables. Dans les axes, la teneur en verbascose augmente régulièrement d'environ 8 fois du stade R7.2 à R8.1, ce qui suggère que le métabolisme de RFO reste actif durant la première phase de séchage de maturation.

Nous avons ensuite étudié si la teneur en sucre est liée à l'acquisition de longévité. Dans ce but, les valeurs de P50 ont été confrontés à des contenus de différents sucres. Il y a une relation linéaire négative significative entre la longévité et les monosaccharides à la fois dans les axes et les cotylédons. Une relation similaire a également été trouvée avec la teneur en saccharose dans les axes mais pas dans les cotylédons. Il y a une relation linéaire négative entre la longévité et le rapport saccharose/stachyose dans les axes ( $r = -0.842$ ,  $p = 0.05$ ), ce qui suggère que la régulation du métabolisme de ces sucres et l'acquisition de la longévité sont peut-être liés. Cependant il n'y avait aucune relation significative dans les cotylédons.

Chez les légumineuses, le rapport entre saccharose et RFO représente un bon indicateur de la vigueur de la graine (VANDECASTEELE et al., 2011). Des conclusions antérieures ont suggéré une relation entre longévité et le rapport saccharose/ RFO (HORBOWICZ & OBENDORF, 1994). Plus récemment, en examinant 276 espèces de graines, WALTERS et al. (2005) n'ont trouvé aucune relation entre longévité évalué par P50 et les niveaux de sucres totaux. En revanche, dans les graines en développement de *Medicago truncatula*, il y avait une relation entre l'acquisition de longévité de la graine et l'accumulation du contenu de stachyose (ROSNOBLET et al., 2007; VERDIER et al., 2013).

A partir de ces résultats, nous supposons que les sucres non-réducteurs ne peuvent pas expliquer l'augmentation de la longévité. Selon toute probabilité, il y a d'autres



**Figure 4.** Analyse en composantes principales des transcriptomes pendant la maturation des graines. PCA a été réalisée à l'aide de centrage médiane. Des cercles gris sont répliqués des stades R7.2 et R9.

mécanismes en plus des sucres solubles non-réducteurs qui se produisent en phase de fin de maturation et contribuent sans doute à l'acquisition de la longévité des graines de soja.

#### **Chapitre 4. Analyses transcriptomiques.**

Les graines récoltées au cours de la maturation ont été utilisées pour l'analyse transcriptomique après séquençage des ARN totaux. Selon les stades, entre 14 et 38 millions de reads ont été obtenus. Après alignement sur le génome Wm82.a2.v1, 56,044 transcrits ont été estimés dont 13.477 ont été considérés comme différentiellement exprimé à la fin de la maturation après analyse statistique et normalisation. Une analyse globale de l'expression de ces transcrits a été réalisée par l'analyse en composantes principales (PCA) (Figure 4). La composante 1 explique 80% de la variation et le graphique indique de grandes différences entre les premiers stades et les stades tardifs de la maturation des graines. Ce dernier indique également que le stade R7.3 représente un changement important survenu au niveau du transcriptome. Les raisons de ce changement restent à étudier. Nous suggérons qu'il reflète une reprogrammation pour préparer les graines pour l'état sec et la germination en synthétisant l'ARNm qui seront stockés et utilisés pour la germination. À l'appui de cette hypothèse, les profils d'expression des facteurs de transcription évoluent très fortement au stade 7.3.

Une analyse par enrichissement des ontologies a permis de mettre en évidence que les catégories liées à la chlorophylle et photosystèmes sont surreprésentées au cours de la maturation. Les gènes impliqués dans ces catégories voient leur expression différentielle diminuer au cours du séchage. Cette observation est conforme à de nombreux rapports de la littérature indiquant que la dégradation de la chlorophylle et la différenciation des chloroplastes est un processus actif et nécessaire pendant la maturation des graines (TEIXEIRA et al., 2016). En effet; l'altération de ce processus par des stress semences (TEIXEIRA et al., 2016), par mutation des gènes impliqués (NAKAJIMA et al., 2012) ou traitements chimiques qui portent atteinte à l'activité photosynthétique des graines en développement (ALLORENT et al., 2014) entraîne une perte de longévité ou de vigueur des graines. Les raisons pour lesquelles la rétention de la chlorophylle est préjudiciable à la qualité des graines reste inconnue.

Une autre catégorie sur-représentée est celle de "réponse à la chaleur" (GO: 0009408), «réponse à un stimulus abiotique» (GO: 0009628) et «développement post-



embryonnaire» (GO: 0009791). Une inspection plus approfondie des gènes appartenant à ces processus biologiques a révélé principalement la présence d'homologues de protéines de choc thermique et HSF. Les gènes codant pour des protéines LEA sont déjà été fortement exprimés au stade R7.2. Aucun transcrit n'a été retrouvé dans nos analyses, suggérant que leur accumulation est maximale avant l'acquisition de la longévité. Un constat similaire a également été fait chez *Medicago truncatula*, où les transcrits codant pour les LEA augmentent à mi-course au cours du développement des graines, en corrélation avec la tolérance à la dessiccation et sont restés élevés par la suite (VERDIER et al., 2013). Par conséquent, il est probable que l'accumulation de produits de transcription LEA dans notre matériel a également eu lieu beaucoup plus tôt que le stade R7.1. Une étude du protéome est donc nécessaire afin d'identifier les protéines LEA dont l'abondance serait corrélée avec la longévité.

Nous avons examiné plus en détail les profils d'expression de tous les gènes codant pour les facteurs de transcription différentiellement exprimés au cours de la maturation du soja, ce qui représente 1086 gènes. Ils présentent trois profils d'expression avec un premier profil présentant un pic d'expression dans les premiers stades de maturation; un second profil où le pic est plus tardif au stade 7.2 et 7.3 et représente potentiellement des régulateurs de la longévité et un troisième profil montrant une expression nettement plus tardive. Au sein du second profil, deux familles ont été trouvées significativement enrichies, celles des AP2/EREBP et WRKY. Les facteurs de transcription de la famille AP2/EREBP jouent un rôle important dans le contrôle des processus de développement et la signalisation redox dans le contexte de stress abiotique (DIETZ et al., 2010). Leur surreprésentation dans le développement de graines de soja a également été rapporté par Jones & Vodkin (2013). Plusieurs homologues ont été trouvés pour être co-régulé avec l'induction de la longévité dans la légumineuse modèle *Medicago truncatula* (VERDIER et al., 2013). Les gènes de la famille des facteurs de transcription WRKY régulent l'expression d'autres gènes liés à la défense. Righetti et al. (2015) décrit un mutant *wrky3* chez *Medicago truncatula*, qui présente une longévité inférieure à celle du type sauvage. Corroborant cette observation, le niveau d'expression de *WRKY3* chez le soja augmente de manière concomitante avec la longévité.

Les facteurs de transcription surexprimés en fin de maturation appartiennent aux familles bHLH, AP2/EREBP et HSF. PHYTOCHROME-INTERACTING FACTORS (PIFs) sont une petite sous-famille des bHLH. Dans notre



étude, deux membres de cette sous-famille étaient positivement corrélées avec la longévité, PIF1 et PIF7. PIF1 régule négativement la biosynthèse de la chlorophylle (MOON et al, 2008; OH et al, 2004), et l'expression du gène augmente de manière concomitante à la perte de chlorophylle pendant la phase de maturation. Les gènes codant pour les facteurs de transcription HSF6B, HSFA3, HSFC1, et HSFAD1 ont une expression élevée dans la phase finale de maturation et ont également été positivement corrélé avec la longévité, ce qui suggère un rôle important dans la régulation de ce processus. En effet, Verdier et al. (2013) ont également trouvé un HSF orthologue putatif de HSFA9 chez *Arabidopsis* à l'interface entre les modules de la tolérance à la dessiccation et de la longévité. Un homologue de HSFA9 de tournesol semble réguler la longévité des graines. En effet, la surexpression de ce gène chez la graine de tabac conduit à une stabilité accrue contre le vieillissement accéléré (PRIETO-DAPENA et al. 2006; PERSONAT et al., 2014). Ce HSF interagit également avec le facteur DREB2 d'une manière spécifique à la graine pour améliorer la stabilité contre le vieillissement accéléré (ALMOGUERA et al., 2009).

Un autre facteur de transcription corrélé positivement avec la longévité est NFXL1, une NF-X1-type zinc finger protein. Il est impliqué dans les réactions de défense en réponse à des agents pathogènes (MÜSSIG et al., 2010). Des plantes surexprimant NFXL1 présentent un taux de survie plus élevé lorsqu'elles sont soumises à des stress abiotiques tels que le sel, la sécheresse, ou une intensité lumineuse élevée (LISSO et al., 2006). Chez *Medicago truncatula*, la caractérisation de mutant *nfxl1* a démontré que ce gène régule certains des noeuds du réseau de maturation et présente une longévité diminuée (RIGHETTI et al., 2015). Dès lors, notre travail confirme l'intérêt pour ce gène qui fait le lien entre longévité et défenses contre les stress biotiques

## **Chapitre 5. Conclusions**

Ce travail montre que la longévité chez le soja est acquise dès la fin du remplissage, immédiatement après la tolérance à la dessiccation.

1. Soybean seeds acquire longevity progressively during maturation after seed filling and mass maturity, however their function on longevity is more complex;
2. The accumulation of non-reducing soluble sugars content are related to acquisition of longevity;



3. Throughout the late seed maturation, there are complex changes in the transcriptome that are likely associated with the acquisition of seed longevity and the preparation of the quiescent state. This suggests that physiological maturity is acquired much later than originally described in the literature;



## INTRODUCTION

Production of grain legumes is strategically and economically important for food security to sustain the increase in meat consumption world-wide. Soybean (*Glycine max*) is one of the most important species cultivated in the world. It is considered one of the five sacred grains which are rice, wheat, barley and millet and it was essential for the existence of the Chinese civilization. Soybean seed is rich in protein and oil (LEE et al., 2015) around 38% and 20%, respectively. In Brazil, the area cultivated with soybean in 2015/2016 was 33.3 million of hectares accounting for over 100 million tons of soybean grains (USDA, 2016), which represents 46% of total grain produced in Brazil (CONAB, 2016).

Modern agriculture demands the use of technologies to allow successful crop development and maximal plant yield. In this respect, the production of highly vigorous seeds is crucial to achieve crop production efficiency. The desirable characteristics of high seed vigor are the seed storability in the dry state (longevity), high and synchronous germination and seedling establishment, even under suboptimal conditions. Seed longevity is a complex trait determined by gene by environment interactions that remain poorly understood. For soybean grown under tropical environment, low physiological seed quality remains a recurrent obstacle, leading to poor stand establishment and decreased in yields. To achieve a stable production of high vigor seeds, it is necessary to gain further knowledge into the mechanisms governing seed vigor during seed development.

During maturation, seeds progressively acquire longevity (HAY & PROBERT, 1995; CHATELAIN et al., 2012, RIGHETTI et al., 2015). Seed storability is a complex trait affected by environmental factors during seed formation, harvest and storage, and is usually controlled by several genes (CLERKX et al., 2004; DARGAHI et al., 2014). For example, during seed maturation of *Medicago truncatula*, a model plant closely related to soybean, longevity starts once reserve accumulation is terminated and increases progressively until pod abscission (CHATELAIN et al., 2012; VERDIER et al., 2013; RIGHETTI et al., 2015). Mechanisms that confer seed longevity at dry stage have received major attention. They include the accumulation of non-reducing sugars, LEA and HSP proteins, which confer protection against denaturation and deleterious changes in the conformation of macromolecules (CHATELAIN et al., 2012; BUITINK & LEPRINCE, 2008). Longevity is also conferred by a range of antioxidants that protect the seeds against



oxidative damage during storage. However, the mechanisms controlling the progressive acquisition of seed longevity remains largely unknown.

High-throughput RNA sequencing (RNA-Seq) and unsuing gene annotation and transcriptome analysis is a powerful technique to discover differentially expressed genes regulating development (PONJAVIC et al., 2007; YANG et al., 2014), including in soybean (JONES & VODKIN, 2013). The recent sequencing of the soybean genome (SCHMUTZ et al., 2010), coupled with the next-generation of RNA sequencing technology allow to study changes in gene expression throughout multiple stages of soybean seed development (JONES et al., 2010), but these studies did not focus on the late phase of seed maturation, including the induction of seed longevity and maturation drying.

The aim of this study was to characterize the seed development stages during which longevity is acquired and gain insights into the molecular mechanisms regulating this acquisition. For this purpose, we used the power of high-throughput RNA sequencing technology to elucidate expression profile across a period of seed maturation revealing the transcriptome during acquisition of seed longevity. We also monitored the evolution of non-reducing soluble sugars, which are involved in the mechanism of survival in the dry state.



# I. LITERATURE REVIEW

## I.1. Importance of soybean in Brazil

Soybean, *Glycine max* (L.) Merrill is an autogamous and annual plant belonging to the *Fabaceae* family (Leguminosae), *Faboideae* subfamily (Papilionoidae). Soybean has its center of origin in China and, from there, expanded to other parts of Asia, around the eleventh century b.C (SEDIYAMA et al., 2009). The ancient Chinese literature reveals that soybean may have been cultivated extensively in China at least 2,500 years b.C (HYMOWITZ, 1970).

Soybean is one of the most important cultivated species in the world, being considered one of the five sacred grains such as rice, wheat, barley and millet, considered essential for the existence of the Chinese civilization. Soybean seed is rich in protein and oil (LEE et al., 2015). The average levels of protein and oil are 38% and 20%, respectively (BUBUJIA et al., 2015). The plant can be also used as green manure, forage, silage and pasture. The grain can provide oil for human consumption as well as for the production of biodiesel and lubricants (CAMPANELLA et al., 2010). The soybean bran is used in food and feed and also for the production of processed or semi-processed products (SEDIYAMA et al., 2009).

Currently, world production of soybean is concentrated in the United States, Brazil and Argentina. Brazil is the second largest producer and second largest exporter (USDA, 2016). In Brazil, soybean is cultivated in large areas, advanced technological level, as well as with high productivity. The area cultivated with soybean in the crop years 2015/2016 was 33.3 million of hectares accounting for over 100 million tons (USDA, 2016), which represents 46% of total grains produced in Brazil (CONAB, 2016).

The first reference of introduction of soybean in Brazil was made in the late nineteenth century in the year 1882 in the State of Bahia (SEDIYAMA et al., 2005). From the 1960s, the rapid development of soybean in the country led to a new and strong sector, characterized by a high demand for technologies. In the late 60s, the research on soybean was small and concentrated in the southern regions of the country, where lines and cultivars from USA were first introduced (EMBRAPA, 2004). In the following decades, soybean was established as the main Brazilian agribusiness crop, going from 1.5 million tons in 1970 to 15 million tons in 1979 (EMBRAPA, 2004).

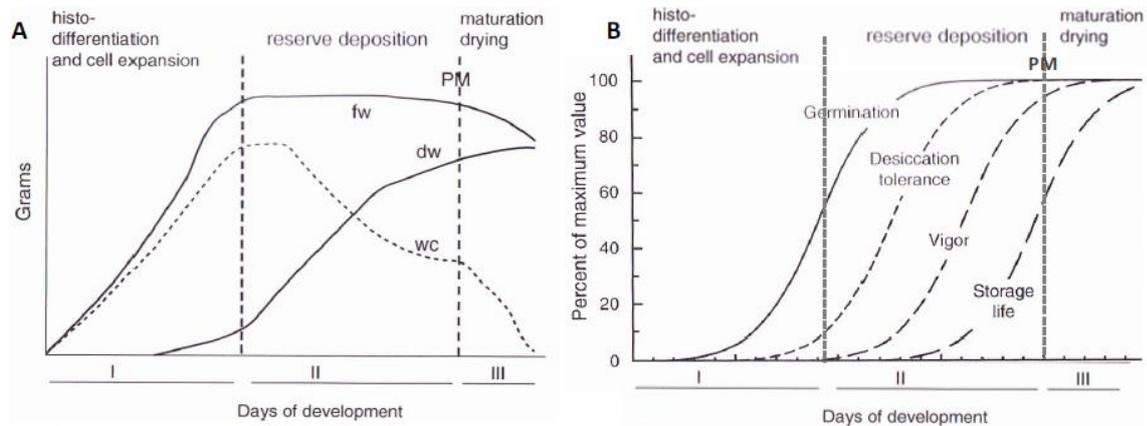


The history of the Brazilian agriculture has changed significantly due to the growth of soybean production in the country at the following years. Soybean was largely responsible for the development of the Brazilian commercial agriculture (EMBRAPA, 2006). Even with an annual growth of cultivated area with the crop, the productivity continues to grow at higher rates than in other countries (FAO, 2010). This significant increase in productivity is also due to the fact that the country has made significant investments in the generation and dissemination of high-level and specific technologies for each region from Brazil over the years.

The seed is the basis of all this progress for soybean cultivation, it is the essential ingredient for the production process, and its quality is considered an indispensable element. The increasing in soybean production brought demands for seeds with higher physiological quality. To obtain soybean with high physiological quality is necessary to know the phenology of plant and seed, and when the attributes of seed quality are acquired during development.

The characterization by phenological stages on the plant and seeds allow detailing the description of the plant cycle with respect to the utilities in each stage. It is possible to use phenology for more specific purposes, such as fertilizers coverage in phytosanitary treatments or even in observing of the acquisition of some traits of seed physiological quality, associated with well-defined stages (PASCALE & DAMARIO, 2004). In soybean, development stages or phenological stages are divided into vegetative stage (V1, V2 ... Vn) and reproductive stage (R1, R2, R3, ..., R8) (FEHR & CAVINESS, 1977). The characterization of soybean most used and accepted in the world is proposed by Fehr and Caviness (1977), which is based on characteristics of the plant and pods. For example, stage R1 is characterized by the beginning of flowering. The seed filling starts at stage R5 stage. Stage R7 is the beginning of maturation and it is observed one typical pod on the main stem that has reached its mature pod color. Howell et al. (1959) reports that the phenological stage R7 marks the mass maturity in soybean seed, which is characterized by the maximum accumulation of dry weight, with water content ranging from 50% to 60% and this usually occurs when the pods and seeds become yellow (RITCHIE et al., 1994). At this stage the seed off the parent plant. After ceasing the supply of nutrients, it begins a slow process of water loss, up to stage R8, with around 13% of water content. According to Fehr

and Caviness (1977), stage R8 is the full maturity, which is described as ninety-five percent



**Figure 1.** Three different phases of development of orthodox seeds (phases I, II and III) and physiological maturity (PM). A – Evolution of the fresh weight (FW), dry weight (DW), water content (WC) during development. B – Acquisition of various attributed of physiological quality during development of orthodox seeds (according to Bewley et al., 2013).

of the pods with mature pod color. In the transition from stage R7 to R8, the metabolism of desiccation tolerance mechanisms is enhanced (Oliver & Bewley, 1997), however the exact moment that it happens is unknown.

Ritchie et al (1994) suggest a scale based on Ferh & Caviness (1977), from which R5, R7 and R8 stages could be divided to increase the precision of identification of the phenological stages. However, they do not suggest how it could be that subdivision and its characterization of the pods and seeds. A subdivision of the phenological stages would allow us to obtain more information at the level of acquisition of seed quality, including longevity, which is acquired in the final phase of maturation, and so far, it is unknown when this trait is higher during soybean seed maturation.

## **I.2. Seed development and maturation**

### **I.2.1. Seed filling and formation**

The seed development is a critical stage in the life cycle of plants which is controlled genetically and involves an ordered sequence of morphological, physical, physiological and biochemical changes that occur from the ovule fertilization to disconnection of a mature seed from mother plant (Beltrati & Paoli, 2003).

In the majority of seeds, development can be conventionally divided into three phases (Figure 1A). Phase I (embryogenesis) corresponds to embryogenesis and is characterized by cell division and elongation followed by increase in fresh weight of the whole seed (GRUWEZ et al., 2013). Phase II represents the phase of seed filling during which cells are expanding as they synthesize storage protein and oil. The reserve accumulation results in a significant increase in dry weight during this phase and a slight decrease of water content. In addition, during this period the seed physiological quality attributes are acquired sequentially such as the capacity to germinate, acquisition of desiccation tolerance, vigor and longevity (Figure 1B) (BEWLEY et al., 2013).

The ability to germinate is first acquired. In some species occurs two peak of germination, the first one at the beginning of development, however seed has not reserves enough to form normal seedlings, thus some regulators as ABA prevents germination and then the ability to germinate is progressively acquired during seed maturation. Desiccation

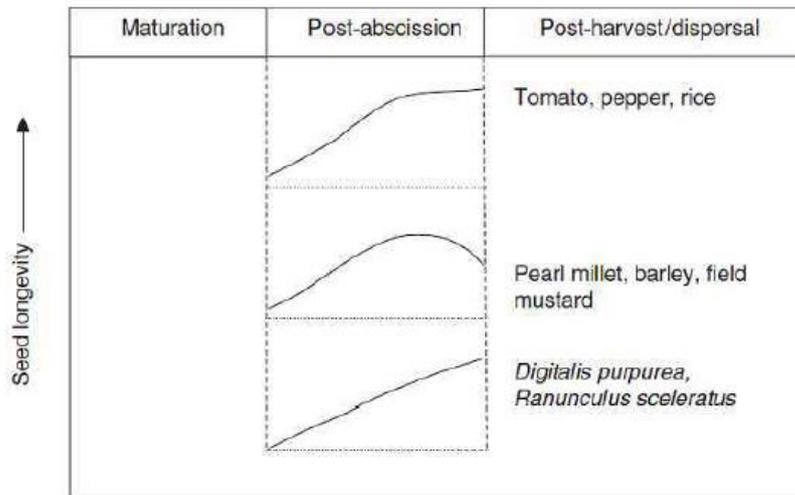


tolerance is defined as the ability to survive a rapid drying that bring the seed water content to values found in mature seeds (i.e. 5-12% according to the species) and it is acquired after germinability. Then vigor is acquired which is the ability of a seed to germinate fast and uniform and also forming seedlings under adverse environmental conditions.

During phase III, seed development ends with a drying phase of maturation or late desiccation. During this phase the accumulation of reserves ceases. This last phase is situated between physiological maturity and the maturity of harvest (abscission of fruit). According to Bewley et al., (2013) physiological maturity (also termed mass maturity) is the point of development when seed dry weight has reached its maximum value. In some cases, several components of seed quality, including desiccation tolerance, vigor and longevity, develop after physiological maturity (Fig 1B), a period in which seed moisture content is declining (BEWLEY et al., 2013). This natural drying phase is important to bring the seeds into a quiescent dry state, characterized by a arrest of metabolism and formation of a glassy state that prevent seed deterioration or even germination of a seed when in the fruit.

### **I.2.2 Acquisition of desiccation tolerance**

Desiccation tolerance (DT) is defined as the ability to survive the removal of almost all the cellular water without irreversible damage and resume normal activity upon rehydration (LEPRINCE & BUITINK, 2010). It is widely found in reproductive structures (pollen, spores, orthodox seeds) as well as in the vegetative parts of a few species (so-called resurrection plants) (ALPERT & OLIVER, 2002). DT constitutes the exception of the established principle that life consists of chemical interactions in liquid water. Desiccation results in reduced cellular volumes and causes the compaction of cytoplasmic components. Such compaction increases molecular interaction, which results in protein denaturation and membrane fusion in desiccation-sensitive tissues (HOEKSTRA et al., 2001). During seed late maturation phase, the cytoplasm of seed cells transforms from a liquid state to a solid-like state, the so-called glassy. In the glassy state, cellular components are stabilized and their mobility is severely restricted. The glassy state can be compared to a solid, thermodynamically unstable without regular structure and a high viscosity (WALTERS, 1998; BUITINK & LEPRINCE, 2004). It is formed by increasing the concentration of solute in the cell during the drying phase of the maturation. The formation and the presence of the glassy state in seeds does not explain the desiccation tolerance, because the critical water



**Figure 2.** Diagram illustrating the three types of profile of acquisition of longevity in post abscission of fruit and seed dispersal (Probert e al., 2007).

content below which the sensitive tissues die is greater than that in the form of glassy state (BUTINK & LEPRINCE, 2004). However, the formation of glassy state has been observed both in desiccation tolerant and recalcitrant seeds, suggesting that it is not associated with desiccation tolerance but is essential for survival during storage (BUTINK & LEPRINCE, 2008). Desiccation tolerance of seeds is a pre-requisite to seed longevity, as shown in recalcitrant seeds, which are sensitive to desiccation and cannot be stored for long periods (ANGELOVICI et al., 2010; OOMS et al., 1993).

DT is established during maturation considerably earlier than the maturation drying itself and the acquisition at different organs does not occur simultaneously (LEPRINCE et al., 1993). DT may be acquired before, during or after maximum dry weight of seed varying according to the species (PROBERT et al., 2007). In soybean, it is acquired at the end of seed filling (BLACKMAN et al., 1991). Seeds that have the ability to tolerate desiccation and with ability to store in dry state are called orthodox because they undergo some degree of desiccation (BEWLEY et al., 2013; CASTRO et al., 2004).

### **I.2.3. Acquisition of longevity**

Seed longevity is defined as the total time-span during which seeds remain viable. Seed longevity is an important trait for ecology, agronomy and economy. Remarkably seed longevity of for more than 1000 years have been reported for some species such as *Nelumbo nucifera* (SHEN-MILLER, 2002) and *Phoenix dactylifera* (SALLON et al., 2008), whereas other species such as onion and pepper show a relatively short longevity (DEARMAN et al., 1986). It has been also reported that seed longevity varies within a plant species such as *Arabidopsis* (CLERKX et al., 2004) and wheat (LANDJEVA et al. 2010).

The acquisition of longevity occurs after the acquisition of DT during the late drying phase. It is suggested that the mechanisms involved to protect seeds during desiccation are also involved in maintaining the viability during storage (SANO et al., 2015).

According to Probert et al. (2007), three types of acquisition of longevity profile can be observed in relation to abscission of the fruit or seed from the mother plant (Figure 2) 1) longevity increases progressively during post-abscission phase then remains constant. This profile is found in tomato, pepper and rice seeds 2) in maize, spring barley,



and wild mustard seeds, longevity increases and then gradually decreases prior to the dispersion of seeds (Figure 2). This profile also occurs in *Medicago truncatula* seeds (CHATELAIN et al., 2012). 3) Seeds of *Digitalis purpurea* (HAY & PROBERT, 1995) and *Anemone nemorosa* are characterized by a steady increase in longevity throughout development and after abscission period (Figure 2).

It is well established that longevity continues to increase after mass maturity (VERDIER et al., 2013; CHATELAIN et al., 2012; PROBERT et al., 2007) However, we do not know in which profile soybean seeds should be ranked..

Evaluation of the seed longevity during the natural aging under low temperature and humidity conditions requires a long time because the natural aging occurs when the seeds are stored under controlled environmental conditions for long periods of time. These stored seeds can be taken at certain intervals to test their ability to germinate (seed longevity). Reduction in seed germination is an indicator of loss of seed vigor during the aging process. Under laboratory conditions, orthodox seeds can remain viable for many years. To overcome the waiting time of natural aging, artificial aging methods were developed and used to study the longevity of the seeds of several species. Aging rate of the seeds depend on the seed moisture content, temperature, and seed initial quality (WALTERS, 1998; WALTERS et al., 2005). High relative humidity and/or high temperature are used to accelerate aging and deterioration of seeds (TESSIER et al., 2002). However, it is not clear whether these artificial ageing conditions mimic the biochemical and molecular events that occur during the natural aging of the seeds. According to Rajjo et al. (2008) and Tessier et al. (2002), markers of deterioration are found both in artificially and natural aged *Arabidopsis* seeds, suggesting the mechanisms of seed deterioration do not depend on storage conditions. However, most of the data in the literature suggest otherwise. For example, genetic studies in *Arabidopsis* show that QTLs describing loss of viability under natural and artificial aging are not located on the same loci (BENTSINK et al., 2000). In rye and wheat, longevity measured in natural storage conditions did not correlate with longevity measured under more deleterious conditions (Niedzielski et al., 2009). In lettuce (MIRA et al., 2010) production of volatiles during storage leading to seed deterioration was different according to the storage conditions.



To predict loss of seed viability over time, Ellis & Roberts (1980) developed a model according to the equation:

$$v = K_i - p/\sigma$$

Where  $v$  is the seed viability after  $p$  days of aging,  $K_i$  is the initial quality of seed lot expressed in probit and  $\sigma$  refers to the time taken for viability to fall by 1 probit during storage.  $\sigma$  is the function of the storage conditions according to the equation:

$$\log \sigma = K_E - C_w \log m - C_H t - C_Q t^2$$

$K_E$ ,  $C_w$ ,  $C_H$  e  $C_Q$  are specific variables from studied cases,  $m$  is the seed water content (expressed as % of fresh material) and  $t$  is temperature at which the seeds were stored. Longevity may be expressed by the  $K_i$  considering that  $\sigma$  is constant (Ellis et al, 1993; Hay & Probert 1995; and Sinniah et al., 1998a). Another way to quantify the longevity is to determine the number of days of aging to decrease the germination in 50%, half-viability period, P50 (TANG et al., 1999, CHATELAIN et al., 2012).

### **I.3. Molecules and mechanisms involved in dry state**

To survive in the dry state, the seed must prevent damage to its cellular components, both during the loss of water and subsequent rehydration. Cell membranes and large molecules such as proteins and nucleic acids are particularly vulnerable because the hydrophilic and hydrophobic interactions of its components with water are essential to maintain their functional 3D structures. Synthesis of protective mechanisms must be initiated during seed maturation to preserve the integrity of those cellular components as the water is removed.

Further dehydration leads to the formation of a glassy state, which can be understood as a liquid solution with viscosity properties of a solid, which does not form crystals even at very low temperatures and with stability in a wide temperature range (KOSTER, 1991). The glassy state prevents denaturation of molecules and formation of molecular aggregates, occupies space in tissues. During dehydration, it serves to prevent excessive increase in collapse of tissue, solute concentration and pH changes (BURKE,



1986; BUITINK & LEPRINCE, 2008). The glassy state is thought to confer longevity because the restricted movement of molecules within the cytoplasm reduces the rate of deteriorative reactions during storage.

Several mechanisms are present in the cell to survive under desiccation. Non-reducing sugars, LEA and HSP proteins and protection against oxidative damage contribute to the survival of the cell in dry stage. These different mechanisms are described in more detail below.

### **I.3.1. Non-reducing sugars**

During and after seed filling phase, non-reducing soluble sugars accumulate during the maturation in the form of sucrose and raffinose family oligosaccharides (RFO), among them, raffinose, stachyose and verbascose (KUO et al., 1997; PETERBAUER et al, 2001). These sugars are formed by successive addition of galactose to sucrose to form first raffinose through raffinose synthase enzyme then stachyose and verbascose by the action of stachyose synthase and raffinose synthase.

These non-reducing sugars are able to protect the protein and phospholipid bilayers from denaturation induced by water loss and form a glassy state. Two mechanisms have been proposed to explain the protective role of sugars. They are preferential exclusion (or preferential hydration) and water replacement (HOEKSTRA et al., 2001). According to the first one mechanism, sugars maintain a layer of water around macromolecules, thus avoiding the denaturation. Preferential exclusion is likely to be the main mechanism of protecting macromolecules in organisms against moderate water loss (HOEKSTRA et al., 2001).

At the second one mechanism, these sugars can prevent the harmful effects of desiccation on the cellular membranes, as the hydroxyl groups of sugars can gradually replace the OH groups from water molecules (CROWE et al., 1996), replacing the water normally associated with the surfaces of the membranes, thus maintaining the spacing of lipid groups. Sugars such as oligosaccharides and sucrose, replace the water molecules, while keeping the membranes structure and functionality (CROWE et al., 1996; HOEKSTRA et al., 2001) and proteins such as the phosphofructokinase. The hypothesis of water replacement may explain the stabilization of membranes or other macromolecules, and



is supported by *in vitro* studies on liposomes (CROWE et al, 1997; HOEKSTRA et al., 1997) and protein (CARPENTER et al 1987; ALLISON et al., 1999).

Sugars also can interact with proteins, avoiding changes in its conformation and hence, loss of its function (LEPRINCE et al., 1993). Sucrose, maltose and trehalose stabilize and prevent aggregation of enzyme in a dry state in an in “vitro” system (CARPENTER et al, 1987; CARPENTER et al., 1990).

Non-reducing sugars also participate in the formation of the glassy state. The glassy state can be compared to a solid, thermodynamically unstable without regular structure and a high viscosity (WALTERS, 1998; BUITINK & LEPRINCE, 2004). Oligosaccharides, such as raffinose and stachyose, are more effective in forming glasses than sucrose and monosaccharides which have more plasticizing effects (HOEKSTRA et al., 2001). The solid-like glassy matrix in dry seeds is supported by the absence of monosaccharides, however Ooms et al. (1993) suggested that the ratio between oligo and mono-saccharides is more important for desiccation tolerance than its absolute abundance.

The formation of the glassy state may explain the longevity of the seeds (BUITINK et al, 1998; BUITINK & LEPRINCE, 2004). The composition of sugars and particularly sucrose / oligosaccharide ratio is also used as a marker of seed longevity. According to Horbowicz and Obendorf (1994) a ratio <1 is characterized as important to seed longevity (greater than 10 years). However, by biophysical methods, Buitink et al. (2000) were able to show, by manipulating the seed longevity through priming, that there was no connection between the molecular properties of glassy state, oligosaccharide content and the survival in the dry state. Therefore, the question of whether soluble sugars play a role in longevity and what role they would play is still unclear.

### **I.3.2. LEA proteins**

Late embryogenesis abundant proteins (LEA) are small stress proteins that accumulate during maturation. They are thought to play a protective role in the dry state, contributing to the DT. However, their mode of action is not yet clear (BERJAK & PAMMENTER, 2000). The accumulation pattern, the abundant occurrence and the physical characteristics LEA proteins indicate a role in the formation of glass cytoplasm. Commonly associated with desiccation and abiotic stress tolerance, LEA proteins are among the intrinsically disordered proteins in aqueous solution. They undergo desiccation-induced



folding during cell drying suggesting that these proteins could carry out distinct functions under different water states (SANO et al., 2015).

In solution, the LEA group proteins are mostly unstructured however during water loss they assume an alpha-helical conformation (WISE & TUNACLIFFE, 2004; BERJAK et al., 2007).

Because of their amphipathic nature, they are able to prevent the denaturation of macromolecules and stabilize intracellular structures. Several studies show that these proteins are not specifically associated with the DT. Some of these proteins have also been detected in recalcitrant seeds (BRADFORD & CHANDLER, 1992; FINCH-SAVAGE et al, 1994) or vegetative tissues sensitive to drying but tolerant to osmotic stress (TUNNACLIFFE & WISE, 2007; LEPRINCE & BUITINK, 2010).

The role of LEA proteins in seed longevity remains enigmatic. Weichberg et al. (1994) observed that there was no correlation between the content of dehydrins and longevity in seeds of *Ranunculus sceleratus*. In contrast, the accumulation of heat stable protein 57 kDa and unknown nature -but assumed to be an LEA protein- correlated with increased longevity during seed development in *Brassica campestris* (SINNIAH et al., 1998b). In maize, the abundance of the group 1 LEA protein named EMB564 was suggested to be related to seed viability (WU et al. 2011). In recent proteomic study, it was revealed that half of the LEA proteome accumulated during the acquisition of seed longevity in *Medicago truncatula*, rather than DT. This suggests that LEA proteins play an important role in the long-term survival in the dry state (CHATELAIN et al. 2012). Biophysical studies show that LEA proteins could participate in the formation and properties of the glassy state (BUITINK & LEPRINCE, 2004). More research is needed to understand how LEA proteins contribute to seed longevity.

### **I.3.3. Heat Shock Proteins (HSP)**

Heat shock proteins (HSP) are usually found in plants subjected to drought and are associated with tolerance to desiccation (JOSHI & NGUYEN, 1996). HSPs are encoded by multigene families. They are divided into several classes according to their molecular weight as HSP110, HSP90, HSP70, HSP60 and those of low molecular weight (15-30 kDa) termed small HSP (sHSP) (VIERLING, 1991). They were found in different cellular compartments, including mitochondria and chloroplasts. HSP proteins act as



molecular chaperones in case of stress. They bind transiently and non-covalent to the protein and allow this take on a conformation that make them functional and prevent its aggregation and degradation (CRAIG et al, 1993; HENDRICK & HARTL, 1993).

Small HSPs (sHSP) have recently been associated with tolerance to desiccation (ALMOGUERA & JORDANO, 1992; WEHMEYER et al., 1996). These proteins are induced by the same stresses as LEA and their synthesis coincides with the acquisition of DT. The mechanism of action of small HSPs is different from those of the LEA proteins: they do not form molecular shields but interact with the substrate proteins to prevent their aggregation. The sHSPs are part of the ATP-independent chaperone proteins (equally with the LEA proteins), although some studies have shown that the presence of ATP can stimulate (MUCHOWSKI & CLARK, 1998) or otherwise inhibit (SMYKAL et al., 2000) its chaperone activity. Also they work in collaboration with other families of ATP-dependent molecular chaperones, such as HSP70, DnaK or DnaJ (NAKOMOTO & VIGH, 2007).

There is a correlation between the level of the sHSPs and seed longevity. Bettey & Finch-Savage (1998) showed that lots of *Brassica oleracea* seeds with higher content of HSP17.6 were less sensitive to accelerated aging, which suggests a positive effect of sHSPs on longevity. In transgenic tobacco plants, overexpression of sunflower Heat Shock Factor A9 (HSFA9), a heat stress transcription factor, led to enhanced accumulation of HSPs and to improved seed longevity (PRIETO-DAPENA et al., 2006).

#### **I.3.4. Protection against oxidative stress**

Reactive oxygen species (ROS) include radicals derivatives from oxygen with unpaired electrons as singlet-oxygen ( $^1\text{O}_2$ ), superoxide anion ( $\text{O}_2^-$ ), hydroxyl radical ( $\cdot\text{OH}$ ) and nitric oxide ( $\text{NO}^\cdot$ ). They also include  $\text{H}_2\text{O}_2$  which is not a radical species but can degrade into  $\text{O}_2$  in the presence of Fe as catalysts (KRANNER & BIRTIC, 2005). These species are unstable (with the exception of the  $\text{H}_2\text{O}_2$ ) and highly reactive.

The accumulation of ROS leads to the dysfunction of mitochondria, enzyme inactivation, the membrane disruption, and oxidation of lipids, proteins and genetic material (DNA and RNA) (MOLLER et al., 2007).

The loss of seed viability during storage is largely attributed to oxidation of macromolecules. Rajjou et al. (2008) showed that the controlled deterioration in



*Arabidopsis* seed, promoted a significant protein carbonylation. The carbonylation is an irreversible oxidation of proteins leading to denaturation and degradation (DALLE-DONNE et al., 2006). In wheat, the viability loss caused by accelerated aging at 45° C and 100% relative humidity is associated with the accumulation of H<sub>2</sub>O<sub>2</sub> (LEHNER et al., 2006). In sunflower, the loss of viability during storage at 35° C is proportional to the accumulation of H<sub>2</sub>O<sub>2</sub> and lipid peroxidation, estimated by the amount of malondialdehyde (MDA) (KIBINZA et al., 2006). However, this tendency seems to be highly dependent on the moisture content at which the seeds were deteriorated. Also changes in the glutathione redox potential (E<sub>GSSG/2GSH</sub>) towards more oxidizing conditions has been correlated with the loss of viability during storage (KRANNER & BIRTIC, 2005).

Seeds require efficient antioxidant systems to prevent excessive oxidation of macromolecules. It is important to distinguish between passive mechanisms including non-enzymatic ROS scavenging systems and active mechanisms including enzymatic ROS detoxification (SANO et al., 2015).

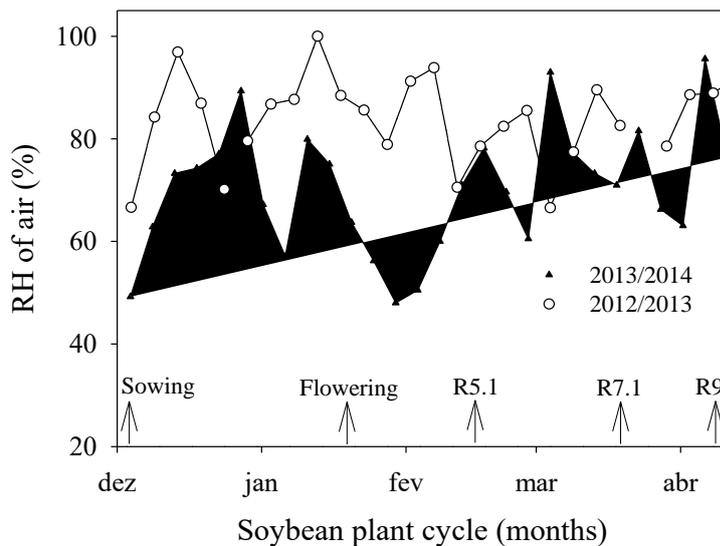
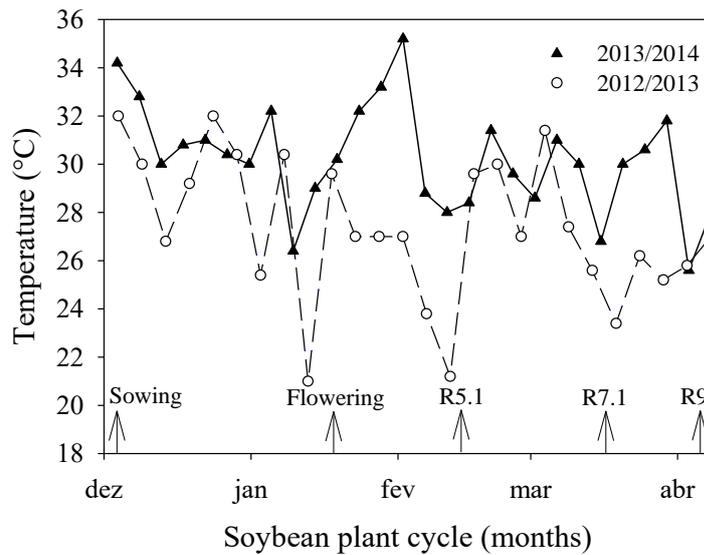
These passive mechanisms include low molecular weight antioxidants including tocopherols, ascorbate and glutathione (SANO et al., 2015). Tocopherol (Vitamin E) acts against phospholipid radicals and prevents lipid peroxidation during storage of the seed, germination and stand establishment (GRENE, 2002; SATTLER et al., 2004). *Arabidopsis vte1:vte2* mutants that are deficient in tocopherols such as tocopherols exhibited a significant reduction in seed longevity (SATTLER et al., 2004). However, seed viability loss is not always associated with a loss in tocopherols (SEAL et al., 2010).

In order to remove excess of ROS accumulation during seed storage and control free radical production generated by the reinitiation of metabolism upon imbibition, seeds use a set of antioxidant enzymes such as superoxide dismutases, catalases, glutathione and ascorbate peroxidases, monodehydroascorbate, dehydroascorbate and glutathione reductases (BAILLY, 2004; KUMAR et al. 2015). Superoxide radicals are converted to hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) by the enzyme superoxide dismutase (SOD) (Grene, 2002; Bailly et al., 2004). H<sub>2</sub>O<sub>2</sub> is oxidized by catalase (CAT) and ascorbate peroxidase (APX). CAT directly converts H<sub>2</sub>O<sub>2</sub> into water and oxygen while APX catalyzes the reaction between ascorbic acid (vitamin C) and H<sub>2</sub>O<sub>2</sub> to form dehydroascorbate and water (Blokhina et al., 2003). Removal of H<sub>2</sub>O<sub>2</sub> by APX requires the involvement of glutathione (GSH) as a part of ascorbate, glutathione cycle (NOCTER & FOYER, 1998). Metallothionein (MTs) play a role in cleaning the ROS.



Some studies have focused on the impact of accelerated aging on enzymatic systems (KIBINZA et al., 2006; LEHNER et al., 2006). In pepper seeds aged at 60 °C, seed viability levels were negatively correlated with lipid peroxidation, and positively associated with antioxidant enzyme activities (e.g. superoxide dismutase, catalase, peroxidase) (DEMIRKAYA, 2013). Similarly, high-temperature treatment (40 °C) of soybean seeds also resulted in greater membrane damage (XIN et al. 2014), as well as decreased antioxidant enzyme activities (superoxide dismutase, and enzymes of the ascorbate-glutathione cycle). Similar changes were also observed in tomato seeds aged at high humidity (20 °C/75% RH (relative humidity) and 30 °C/45% RH (DE VOS et al. 1994). These investigations indicate that the reduction of antioxidant capacity has a significant contribution to the loss of seed viability during ageing.

Although, several other proteins that act as antioxidants or in related ROS signaling, such as thioredoxins, peroxiredoxins and glutaredoxins, have been identified in seeds, their roles in seed longevity remain to be described.



**Figure 3.** Maximum temperature (A) and relative humidity (B) at experimental field from sowing to harvest of soybean seeds in the crop years 2012/2013 (open circles) and 2013/2014 (black triangle). The bars indicate sowing, beginning of flowering, start of seed filling (stage R5.1), period of acquisition of DT and longevity (from stage R7.1 to R9).

## **Chapter II. MATERIAL AND METHODS**

### **II.1 Plant Material**

The soybean cultivar used in this study was BRS 284 which were stored for 8 months before sowing, belonging to the early maturing group with cycle of 120-126 days.

### **II.2 Field experiment**

The experiment was performed in two consecutive years, 2012/13 and 2013/14. The experimental area belongs to the São Paulo State University - College of Agricultural Science, UNESP - FCA (Botucatu, SP – Brazil). The seeds were sown on December 4<sup>th</sup>, 2012 and December 5<sup>th</sup>, 2013, respectively. All the phytosanitary controls were performed according to recommendations of Raij et al., (1997). The field ambient temperatures were measured with an electric thermograph during the experiment and local relative humidity data were obtained from the Department of Engineer at the College of Agricultural Science, UNESP-Botucatu-SP (Figure 3). Figure 3 shows temperature and relative humidity during plant cycle for both years. The irrigation system was implemented in the field for water supply during absence of rain; it avoided the hydric stress on soybean plants.

### **II.3. Sampling points**

Soybean plants were grown until full maturation and flowers buds were tagged just at the beginning of flowering, prior to anthesis. Following, the fruits were manually removed from the plants and the seeds were extracted. The pods and seeds were harvested at different time intervals based on phenological stages proposed by Fehr and Caviness (1977) with adaptations (Table 1) and on the number of days after flowering (Table 2). The subdivisions of the phenological stages R7 (R7.1, R7.2 and R7.3) and R8 (R8.1, R8.2, and R9) were made in order to obtain higher precision on the acquisition of quality attributes of seeds, such as germination, desiccation tolerance and longevity.

**Table 1.** Characterization of the phenological stages of soybean plants according to Fehr and Caviness (1977) with adaptations.

<b>Stages</b>	<b>Denomination</b>	<b>Description</b>
R <sub>5,1</sub>	Grain filling	Grains perceptible to touch up to 10% of grain filling
R <sub>5,2</sub>	Grain filling	Most pods grain filling 10% to 25%
R <sub>5,3</sub>	Grain filling	Most pods between 25% and 50% grain filling
R <sub>5,4</sub>	Grain filling	Most pods from 50% to 75% grain filling.
R <sub>5,5</sub>	Grain filling	Most pods from 75% to 100% grain filling.
R <sub>6</sub>	Grains developed	Pods with 100% grain filling and green leaves
R <sub>7,1</sub>	Early maturation	Start to 50% of yellowing leaf and pods
R <sub>7,2</sub>	Early maturation	Between 51% and 75% yellow leaves and pods
R <sub>7,3</sub>	Half of maturation	More than 76% of leaves and pods yellow
R <sub>8,1</sub>	Full maturation	Start to 50% defoliation
R <sub>8,2</sub>	Full maturation	More than 50% defoliation before harvest
R <sub>9</sub>	Harvest maturation	Point of harvest

## II.4 Seed quality assessment

To evaluate the physical and physiological quality during seed maturation the followed experiments were performed:

**Fresh weight (FW) and dry weight (DW)** – The fresh and dry weight of the seeds were evaluated by using precision analytical balance. Fresh weight and subsequent dry weight were determined in an oven at 60° C until the dry weight stabilizes. Four replications of 20 seed each were used and the results were expressed in mg per seed.

**Determination of water content** – Water content was gravimetrically by comparing fresh weight and dry weight after an incubation in the oven at  $105 \pm 3^\circ$  C for 24 hours, with four replicates of 20 seeds. The water content was expressed as gram water per gram of dry weight (g H<sub>2</sub>O per g DW).

**Germination test** - Germination test (G) was evaluated with four replicates of 25 seeds each using paper roll moistened with water and at 25°C. The germination percentage was scored daily by counting seeds with radicle with length  $\geq 1$  mm.

**Desiccation tolerance** - Immature seeds were submitted to enforced drying using an airflow at 43% RH at 25°C (slow drying) generated by saturated salt solution of K<sub>2</sub>CO<sub>3</sub> and at 40% RH at 30 ° C (fast drying). the seeds were kept over drying until their water content reach 0.10 gH<sub>2</sub>O.gDW. After drying, seeds were imbibed as described above. They were considered as desiccation-tolerant when they germinated.

**Longevity** – Artificially dried or mature seeds were placed over a saturated solution of NaCl (75% RH) at 35°C in hermetically sealed boxes for different time spans. Thereafter, they were imbibed as described above. Viability was assessed using the germination assay described above. Longevity was expressed as P<sub>50</sub>, defined as the time (days) at which the stored seed lot lost 50% of viability during storage.

## II.5. Determination of soluble sugars

Sugar contents (raffinose, stachyose, verbascose, sucrose, fructose and glucose) were measured in cotyledons and embryonic axes from the phenological stage R6 onwards. The embryo parts were isolated, lyophilized for 72 h, and then ground to a powder in liquid nitrogen. A solution of 40 ul of melezitose (4 g / L) was added as an internal control

**Table 2.** Days after flowering at different phenological stages of soybean seeds cultivated in the year 2013/2014.

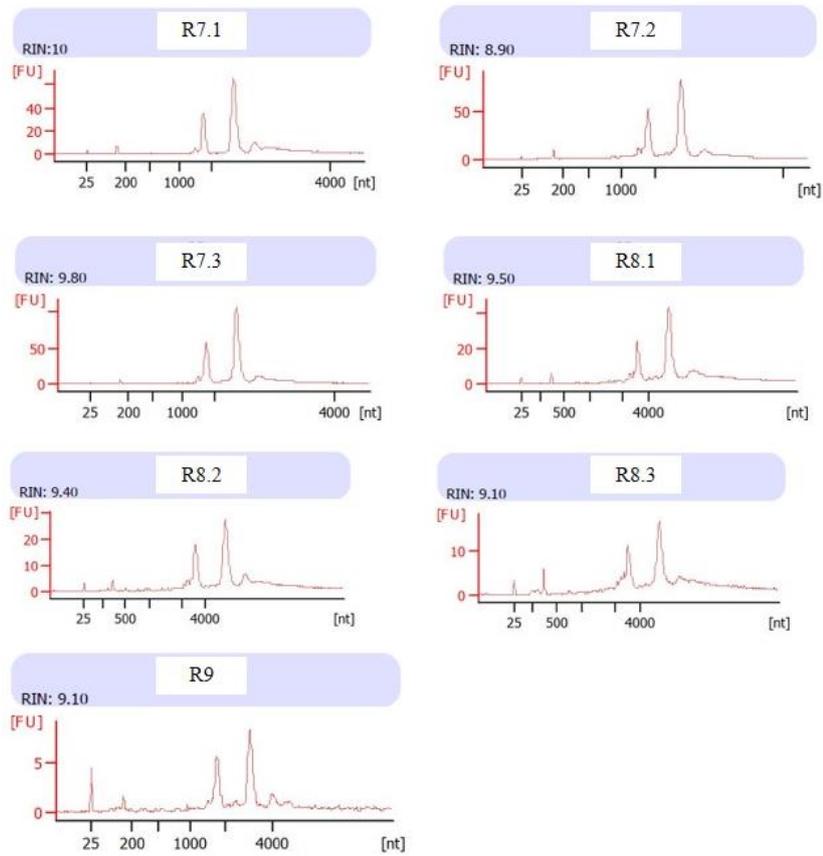
<b>Stages</b>	<b>Days after flowering</b>
R <sub>5,1</sub>	25
R <sub>5,2</sub>	25
R <sub>5,3</sub>	27
R <sub>5,4</sub>	27
R <sub>5,5</sub>	36
R <sub>6</sub>	47
R <sub>7,1</sub>	57
R <sub>7,2</sub>	63
R <sub>7,3</sub>	69
R <sub>8,1</sub>	71
R <sub>8,2</sub>	73
R <sub>8,3</sub>	74
R <sub>9</sub>	77

together with 1 ml of a 80% methanol (v/v) was added to the mixture. The samples were incubated for 15 min at 76° C and then methanol was evaporated in a speed-vac (Mi Four Vac concentrator, Genevac) for 2 h at 40° C. The pellet was resuspended in 1 ml of ultra-pure water and centrifuged for 5 min at 4 °C at 13,500 g (R5417, Eppendorf). The supernatant was diluted from 25 to 80 times. The samples were analyzed by high performance liquid chromatography (Dionex Corp., Sunnyvale, CA) using a pre-column AS11 ion exchange and a Dionex CarboPac PA1 column (2 x 250mm) at a constant flow rate of 0.25 ml/min of 0.1 M NaOH. Sugars were detected by amperometry. Their nature and concentration were determined by comparison with standards of known concentrations. The concentration of verbascose was calculated based on the peak area of stachyose due to the absence of pure standard. Sugar contents were expressed in mg per mg DW.

## **II.6 RNA extraction and sequencing**

Seeds from crop year 2013/2014 were selected to proceed with transcriptome study. We harvested seeds from up to 200 plants at each stage. Then, the seeds were homogenized and frozen. Total RNA was extracted from at least 15 fresh seeds from stages R7.1 to R9 using the NucleoSpin® RNA Plant kit (Macherey-Nagel) according to the manufacturer instruction. RNA integrity was assessed by analysis on a 1% agarose gel, and concentration was additionally determined by UV spectrophotometry (260 nm, NanoDrop 1000, Thermo Fischer Scientific). RNA integrity number (RIN) was measured using a Bioanalyzer (Agilent Technologies). Those samples exhibiting a RIN values ranging between 8.9–10.0 were used for further analysis (Figure 4).

High quality of total RNA samples were used for library preparation and mRNA sequencing at the Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LaCTAD) from the University of Campinas (UNICAMP). cDNA libraries were generated using the TruSeq RNA sample preparation kit (Illumina) to obtain single-end 100-bp long RNA-seq. The libraries were subjected to enzymatic fragmentation by connecting terminal adapters to these fragments and deposited on a flowcell containing 3 lanes via robotic instrument cBOT (Illumina). On the flowcell surface there are oligonucleotides that are complementary to the adapters of libraries. The flowcell was placed in a HiSeq2500 where occurred the incorporation of nucleotides containing fluorescently labeled dideoxy terminators on the fragments linked to the sequencing primer. Upon the



**Figure 4.** RNA integrity number (RIN) values during soybean seed maturation from stages R7.1 to R9.

occurrence of the incorporation of a nucleotide, the fluorescence was excited with a number of lasers and captured by cameras.

## **II.7 Sequences alignment**

Sequencing data quality control was performed prior to data analysis. Using the alignment program Bowtie2 (LANGMEAD et al., 2009), reads were aligned to the ‘Williams 82’ soybean reference genome (assembly Glyma.Wm82.a1.0, annotation v2.0) with 88,647 Glyma 2.0 cDNA soybean gene models determined by the Soybean Genome Project, Department of Energy, Joint Genome Institute (SCHMUTZ et al., 2010), updated on August 13, 2014 available in <http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=PhytozomeV10>. Bowtie2 is an ultrafast and memory-efficient tool to align sequenced reads to long reference sequences.

## **II.8 Transcriptome data analysis**

Estimation of differential gene expression was performed using DESeq2, v1.11.21 (LOVE et al., 2014) available as a R/ Bioconductor package (GENTLEMAN et al., 2004). This package is well adapted to perform differential analysis of count data from small replicate numbers and large dynamic range. Following the procedures described in the Deseq2 documentation, read count tables were loaded into R, normalized using the default method for Deseq2.

Relative expression data were normalized by dividing the mean normalized gene expression value obtained in the stage n+1 by the value obtained in stage n (i. e.,  $R7.2/R7.1$  ;  $R7.3/R7.2$  ;  $R8.1/R7.3$  ;... ;  $R9/R8.3$ ). We performed another normalization method consisting in comparing each stage to the most immature stage (R7.1) i.e.,  $R7.2/R7.1$  ;  $R7.3/R7.1$ ;  $R8.1/R7.1$ ;... ;  $R9/R7.1$  ; and also each stage to the most mature stage (R9) i.e.,  $R7.1/R9$ ,  $R7.2/R9$  ;  $R7.3/R9$  ; ... ;  $R8.3/R9$  . However both attempts failed to obtain expression profiles of marker genes whose expression is known to increase during seed development.

Quantile normalization makes it possible to ensure that all datasets are drawn from the same distribution. Given a reference distribution, a target distribution is normalized by replacing each of its values by the value of the variable with the same rank in the reference distribution.

If the reference distribution contains multiple samples, the target **Table 3**. Primers sequences used for the target gene study.

<b>Gene</b>	<b>Glyma2.0</b>	<b>Forward</b>	<b>Reverse</b>
20S *	Glyma.06G078500	CACCAACACACGATACAAC	TCCCAACCACCAACAATTAACC
60S *	Glyma.15G271300	GGCAGAGAAGGAGGAGAA	ACCTAGCACCCAAGTAAGA
HSP 17.6	Glyma.17G224900	ACGAAAGAGGGACGAAGA	CATTCTCAGGCAGCACAA
HSP 17.6	Glyma.14G100000	TGCGGATGTGAAGGAATATC	AAGCACGTTGTCGTCTTC
HSP 101	Glyma.17G077500	GAGGGTGCAACTTGATAGTC	GCCTTGTCTTCTCTTTCTC
HSP 21	Glyma.08G318900	AACATGCTGGTGGTGAAG	AGGGCTATCCTGTGGTTAT
HSFA3	Glyma.03G191100	CATCAGGTTGGTGGCAATA	GCATTAGCACACTCCTTTCT
HSFA6B	Glyma.03G157300	GAGTGTGAGAGTTGGAAGTG	CCAGCCTCTCTTGTGATTG
BZIP65	Glyma.03G142400	TCGTGAGACAGGCTGATAA	GGCTCGAAGACGATGAAAG
GCN5	Glyma.19G160700	GGAGGACCCGACTAAGATT	CTGCTTCCTCCTCTTTCTTG
NFXL1	Glyma.09G173000	CCGATACCTACCTGCTCTAA	ACGGCACTTCTGTGAAAC
WRKY3	Glyma.07G227200	GGGCTTCAATGGATCCTAAA	GGCAAGTGTGTGGCTATTA

\*Reference gene

and reference distributions will only be identical if the reference distribution is first normalized across all samples (TOMPSON et al., 2016). Quantile normalization was performed using the *normalize.quantiles.use.target* method of the preprocessCore package (BOLSTAD, 2015) in the R statistical environment (R Core Team, 2015). Most significant genes were identified at 0.5 % false discovery rate (FDR) calculated using P-value adjusted for multiple testing using the Benjamini-Hochberg method (NELSON et al., 1998).

## **II.9 Gene Ontology analysis**

Gene ontology enrichment analysis was performed using AgriGO (bioinfo.cau.edu.cn/agriGO/analysis.php) (DU et al, 2010) using Arabidopsis as genome model and the Fisher test (significance level of 0.05 with the Hochberg FDR multi-test adjustment method). Overrepresentation analysis of functional classes during soybean seed development were also analyzed using PageMan (THIMM et al., 2004). Data were subjected to a bin-wise Wilcoxon test, and resulting P values were adjusted according to Benjamin Hochberg

## **II.10 Quantitative real time RT-qPCR**

RNA-Seq output was validated using quantitative RT-qPCR. RNA extraction was performed as previously described of three biological replicates of developing stages from R7.1 to R9. First strand cDNA was synthesized from two µg high quality total RNA using High Capacity RNA-to-cDNA kit (Applied Biosystems) following manufacturer's instructions. Quantitative real time PCR was performed in a 12 µl reaction volume, constituted of 3 µl of cDNA solution at concentration 2µg/ µL, 6 µl of 2× SYBR Green qPCR ReadyMix (Sigma Aldrich), 0.5 µl of 10 mM of both reverse and forward gene-specific primers solution (10mM), 2.5 µl of UltraPuredistilled water.

The amplification was performed in a thermocycler Eco Real-Time (Illumina) with consisted an initial step of incubation at 2 minutes at 50°C, 3 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 30 seconds at 60°C for 30 seconds. At the end of the reactions was performed a melting curve to assess specificity which consisted the following steps: 15 seconds at 95°C, 55°C and 95°C respectively. The data were analyzed in EcoStudy program version 5.0 (Illumina).

Genes were selected based on the relative expression of transcripts after RNA-seq analysis from this study and on literature. The primers used (Table 3) were



preferably designed in the 3' end of the transcript and, if possible, spanning an intron–exon border. The  $T_m$  of the primers was between 58 and 62°C. Routinely a melting curve analysis was performed after the qPCR run (between 55 and 95°C with 0.5°C increments each 10 s). For all primers, a single peak was observed, confirming the synthesis of a single product.

Primer efficiency was calculated through the LinRegPCR program (RUIJTER, 2009). Relative expression levels were calculated using the comparative  $\Delta\Delta(Ct)$  method (LIVAK; SCHMITTGEN, 2001) using both reference genes 60S Ribosomal protein and 20S proteasome (TEIXEIRA, 2016), which also had no variation of expression among stages based on our RNAseq data.



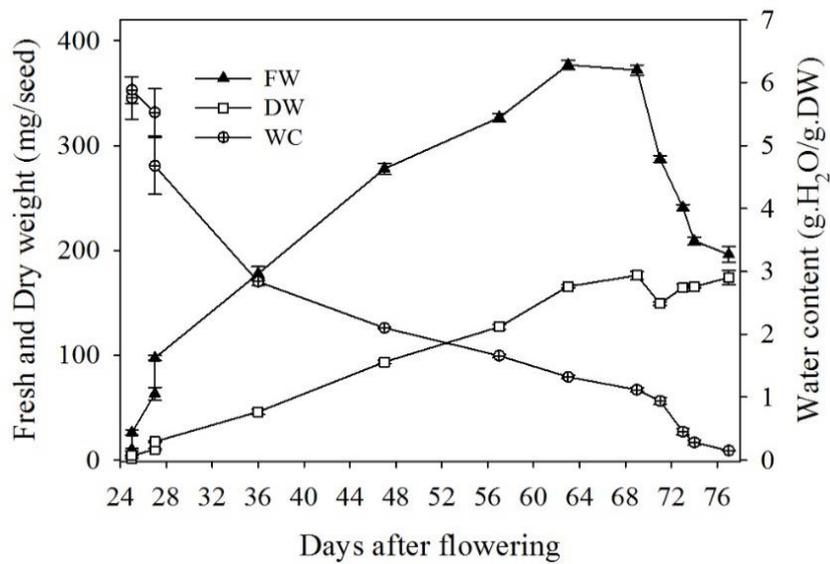
**Figure 5.** Soybean pods and seeds from different phenological stages. White bars represents 1 cm.

## **Chapter III. RESULTS**

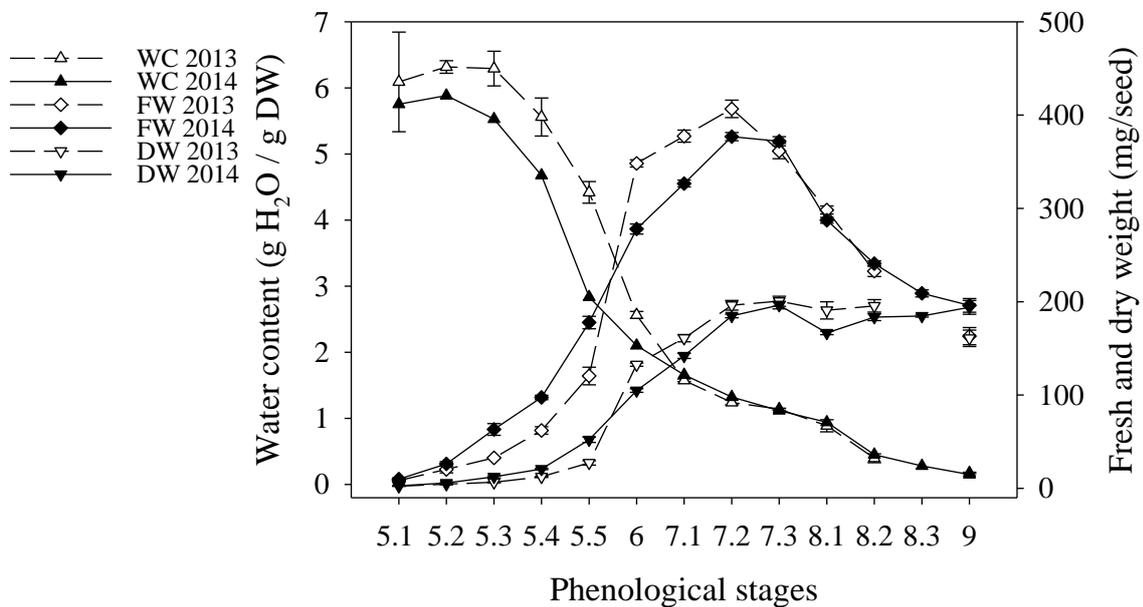
### **III.1. Physiological characterization of soybean seeds during maturation**

The aim of this study was to characterize the soybean seeds during development and observe when the quality attributes are acquired. Therefore, seed development was daily monitored by observing the characteristics of phenological stages, with the subdivisions proposed, based on soybean plant and pods. Pods and seeds were collected from seed filling to mature seeds and were divided into 2 lots. The first lot was directly frozen in liquid nitrogen for further RNA extraction from crop year (2013/14) and the second lot used for the physiological assessment (fresh and dry weight, water content, germination, desiccation tolerance and longevity) and sugar content determination.

Harvests were based on the phenological stages. It is interesting to note that at stage R7 proposed by Fehr and Caviness (1977) is highlighted by chlorophyll degradation, in another word, it has seeds from green to yellow color. According to the researchers previously mentioned, multiple processes related to acquisition of seed quality are already taking place. Working on subdivisions, pods and seeds remain green until stage R7.1 (Figure 5), i.e. 57 DAF (Table 2). Then, the seeds gradually lose their chlorophyll. From stage R7.3, seed size decreased gradually due to natural drying. Stage R8, which is divided into R8.1, R8.2 and R9, is characterized by reducing of water content at late phase of maturation. At the stage R9, the seeds are mature and have brown coloration beyond the reduced water content (0.15 g H<sub>2</sub>O/g .DW). The subdivision of the phenological stages allows higher precision in obtaining information related to the characterization of soybean seeds during maturation.



**Figure 6.** Changes in dry weight (DW), fresh weight (FW) (mg per seed) and water content (WC) (g H<sub>2</sub>O per g DW). Data are the means of four replications of 20 seeds. Bars represent the standard error. Data were from the 2013/2014 harvest.



**Figure 7.** Changes in water content (WC), fresh weight (FW) and dry weight (DW) from crop years 2012/2013 (open symbols) and 2013/2014 (black symbols) based on phenological stages. Data are the means of four replications of 20 seeds. Bars represent the standard error.

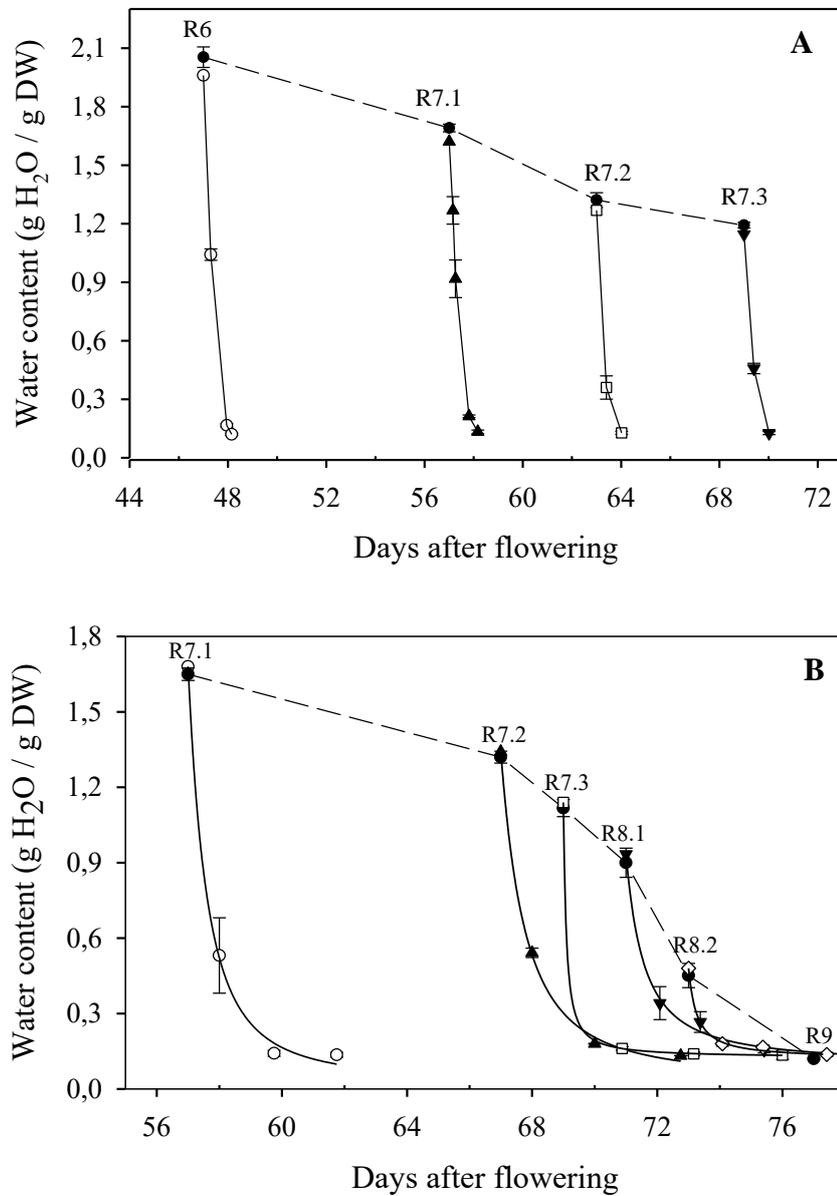
### III.1.1. Seed filling and drying

Figure 6 shows the changes in fresh weight, dry weight and water content in soybean seeds in days after flowering (DAF) for the year 2013/2014. Between 25 and 63 DAF, i.e., stages R5.1 and R7.2 respectively (Table 2), fresh and dry weight of the seed increased indicating that they were in filling phase. The mass maturity that corresponds to the maximum dry weight accumulation occurred at around 63 DAF (stage R7.2). The water content decreased progressively during seed development ranging from up 6.0 g H<sub>2</sub>O / gDW to 0.15 g H<sub>2</sub>O / g DW. The fresh weight decreased significantly from 69 DAF (stage R7.3), indicating that the drying phase started from there, when seeds reached a water content around 1,11 g H<sub>2</sub>O / g DW.

The measurements were repeated in both crop years 2012/2013 and 2013/2014 and compared between them to test the reproducibility of water content, fresh and dry weight data and natural drying phase during development and the results presented as phenological stages at Figure 7. Overall, we found that the changes in fresh and dry weight and water content followed a very similar profile in both crop years.

### III.1.2. Acquisition of desiccation tolerance and longevity

In the crop year 2012/2013 we evaluated desiccation tolerance of the seeds subjected to both fast and slow drying. The seeds that germinate after drying were considered tolerant to desiccation. The slow drying consisted of a drying treatment at 45% RH and 25 ° C and fast drying at 40% RH at 30 ° C. After slow drying, seeds harvested at the indicated stages reached a water content of 0.15 g/g: (i.e. equivalent to that of mature seeds) after 3 days (Figure 8 B); whereas during fast drying this time was reduced to 24 hours (Figure 8 A). Figure 9 compares the acquisition of desiccation tolerance for these two drying regimes. There was no difference in the germination between drying methods (Figure 9): tolerance to desiccation was acquired at stage R7.2. The drying at 45% RH and 25 ° C was selected to continue the study. The study of acquisition of germination, desiccation tolerance and longevity in soybean seeds during development was repeated in the following year 2013/2014 in order to ensure reproducibility of the data. Figure 10 A and B shows the acquisition of germination and desiccation tolerance respectively in both crop years as a function of phenological stages and seed age. When isolated from the pods, immature seeds

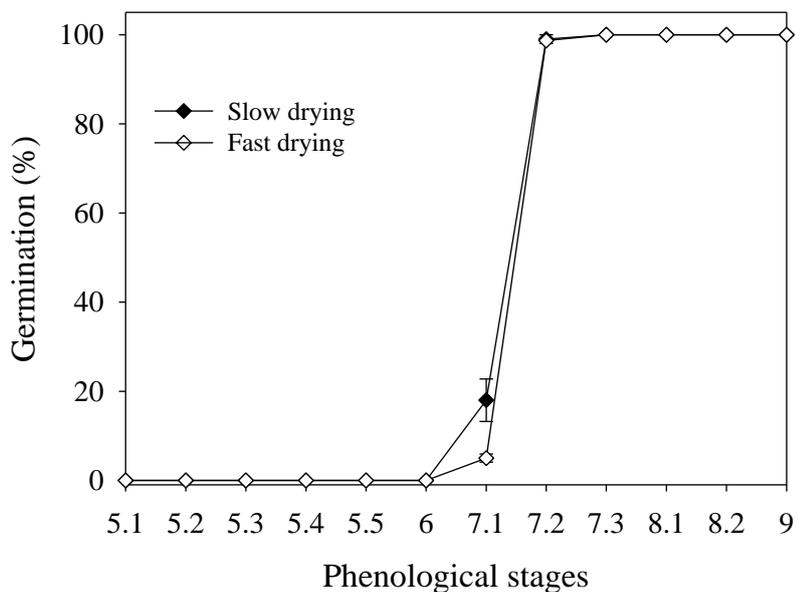


**Figure 8.** Rate of water loss during fast drying at 40% RH and 30 °C (A) and slow drying at 45% RH and 25 °C (B). Seeds were harvested in 2012/2013. Data represent the mean of 3 replicates of 20 seeds. Bars represent the standard error. Dashed line represents the natural drying in soybean seed maturation. Full line represents the artificial drying time.

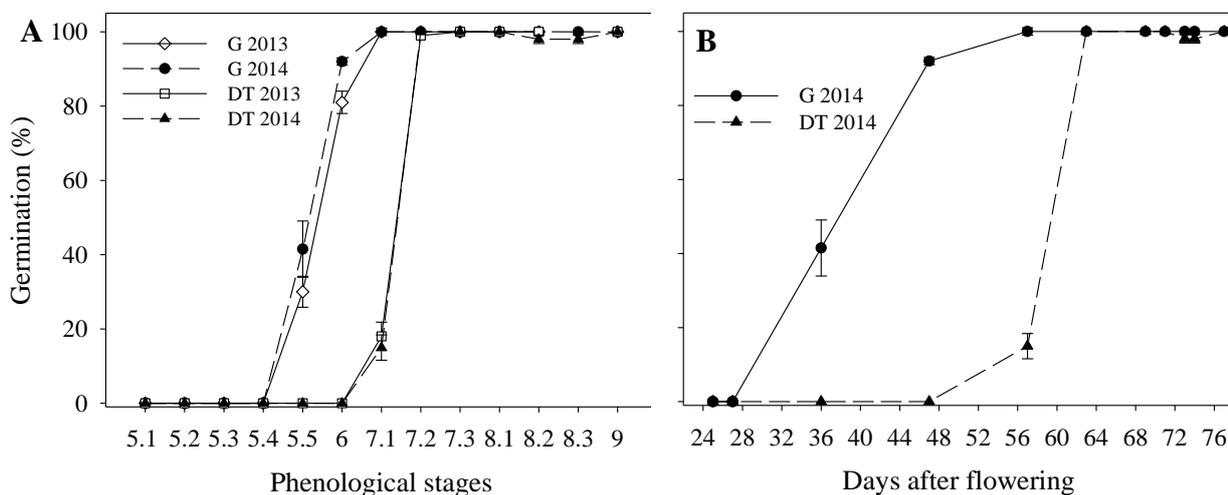
started to germinate from stage R5.5 in both crop years (Figure 10 A), which represents 36 days after flowering (Figure 10 B). Maximum germination (i.e. 100%) was obtained at stage R7.1. Desiccation tolerance in both crop years was acquired between stages R6 and R7.2, (Figure 10 A).

To study the longevity, immature and mature seeds were stored at 75% RH and 35 ° C. Those conditions represent a good compromise between natural aging which would occurs within years and an accelerated ageing where high temperature and relative humidity induces deteriorative effects that should be similar to natural ageing.

At indicated time during storage, seed samples were taken out to assess the germination as an indicator of the viability of the seed lot. We obtained the viability loss curves over time for both years 2012/2013 and 2013/2014 (Figure 11). Seeds from stage R7.1 were partially tolerant to drying with a germination percentage before storage at approximately 20%, then immediately decreased after 5 days of storage. The older the seeds, the longer they remained alive during storage. We found that the storage capacity of the immature seeds gradually increased from the stage R7.2 to R9 in both crop years (Figure 11). As expected from the literature, the loss of viability could be modeled using a reverse sigmoidal fit.



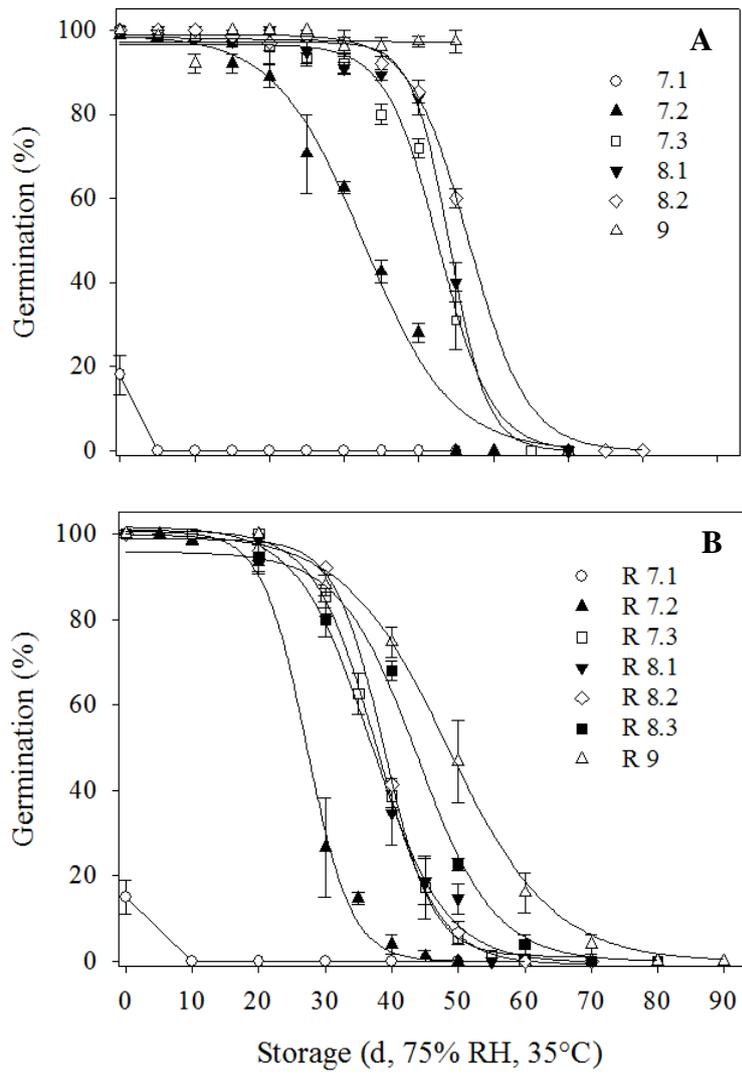
**Figure 9.** Desiccation tolerance in developing soybean seeds after slow and fast drying at different phenological stages in 2012/2013. Data represent the mean of 4 replicates of 25 seeds. Bars represent the standard error.



**Figure 10.** Germination (G, circles) and desiccation tolerance (DT, square) measured after slow drying at 45% RH and 25°C. Seeds were harvested in 2012/2013 (white symbols) and 2013/2014 (black symbols). Data represent the mean of 4 replicates of 25 seeds. Bars represent the standard error. Data are shown as phenological stages (A) and in days after flowering (B).

From these curves, the time to lose 50% of viability during storage (i.e. P50) was used to quantify the longevity. P50 values for the two crop years are shown at Figure 12. For being desiccation sensitive, it was not possible to determine a P50 for seeds harvested at R7.1 stage. The P50 at stage R7.2, was already at 26-28 days, indicating that 50% of the maximal longevity measured in mature seeds was already acquired. It should be noted the time lapse between stage R7.1 and R7.2 is approximately 10 days. From stages R7.2 to R9 P50 values increased 1.8 fold. It was not possible to calculate the P50 for the stage R9. We could estimate based on the crop year 2013/2014 that P50 at R9 would be around 60 days from crop year 2012/2013.

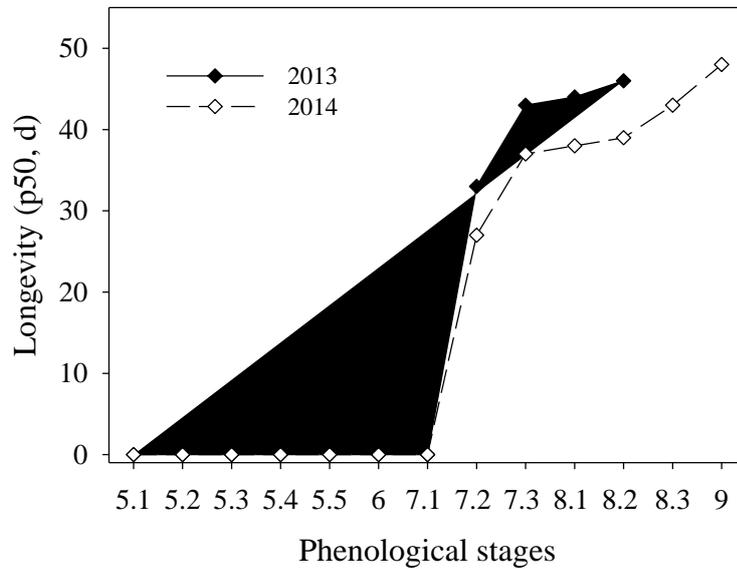
An alternative to describe the acquisition of longevity was to use the viability equation of Ellis and Roberts (1980). At this case, the germination percentages were converted to probit, which makes the relationship between the storage time and probits linear. The linear regression, gives access to of the slope  $\sigma$  (referring to the time taken for viability to fall by 1 probit) and  $K_i$  (initial viability of the seed lot at  $p = 0$  days) parameters (Figure 13).  $\sigma$  reflects the rate of aging while  $K_i$  is the initial physiological quality before aging (ELLIS & ROBERTS, 1980). An example is shown in Fig. 13.



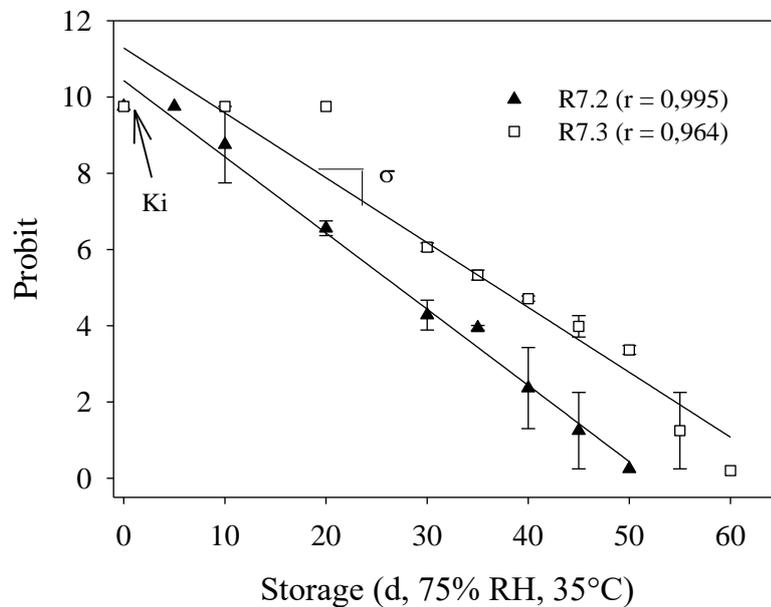
**Figure 11.** Changes in germination percentage of dried seeds harvested at indicated stages of maturation and stored at 35 °C and 75% relative humidity (RH) from crop years 2012/2013 (A) and 2013/2014 (B). Third-order sigmoid curves were fitted to the data to obtain the P50 values. Data are the mean of 3 replicates of 25 seeds. Bars represent the standard error.

Ki and  $\sigma$  values were obtained for developing seeds between stages R7.2 and R9 (Figure 14). There was little variation in Ki values along the soybean seed maturation in both crop years 2012/2013 and 2013/2014 (Fig 14A), suggesting that the initial seed quality is acquired at R7.2. In contrast, the  $\sigma$  values increased progressively during maturation of soybean seeds, indicating that the mature seed deteriorated at a slower speed compared to immature seeds from seed harvested in 2013/2014. In the crop year 2012/2013 was not possible determine Ki and  $\sigma$  at stage R9, as explained previously, storage was until 50 days and the germination was still 100%.

Our data also shows clearly that environmental conditions in the field influenced the acquisition of seed longevity but not germination and desiccation tolerance. The crop year 2012/2013 during seed development was on average colder than 2013/2014. Yet the longevity of immature seeds collected in the latter year was much higher for any stages of 2012/2013. Thus, acquisition of longevity was affected by higher temperatures in 2013/2014.



**Figure 12.** Desiccation tolerance (square) and Longevity (diamond) were expressed as the P50 corresponding to the time of storage (days) at which the dried seed lot lost 50% germination, for both crop years 2012/2013 (black symbols) and 2013/2014 (open symbols).



**Figure 13.** Relationship between the storage time and % of viability expressed in probit for seeds harvested at stages R7.2 and R7.3. Data were fitted with a linear regression according to Ellis & Roberts (1980). The slope  $\sigma$  corresponds to storage time to reduce the viability for 1 value of probit).  $K_i$  corresponds to the viability at storage time 0).

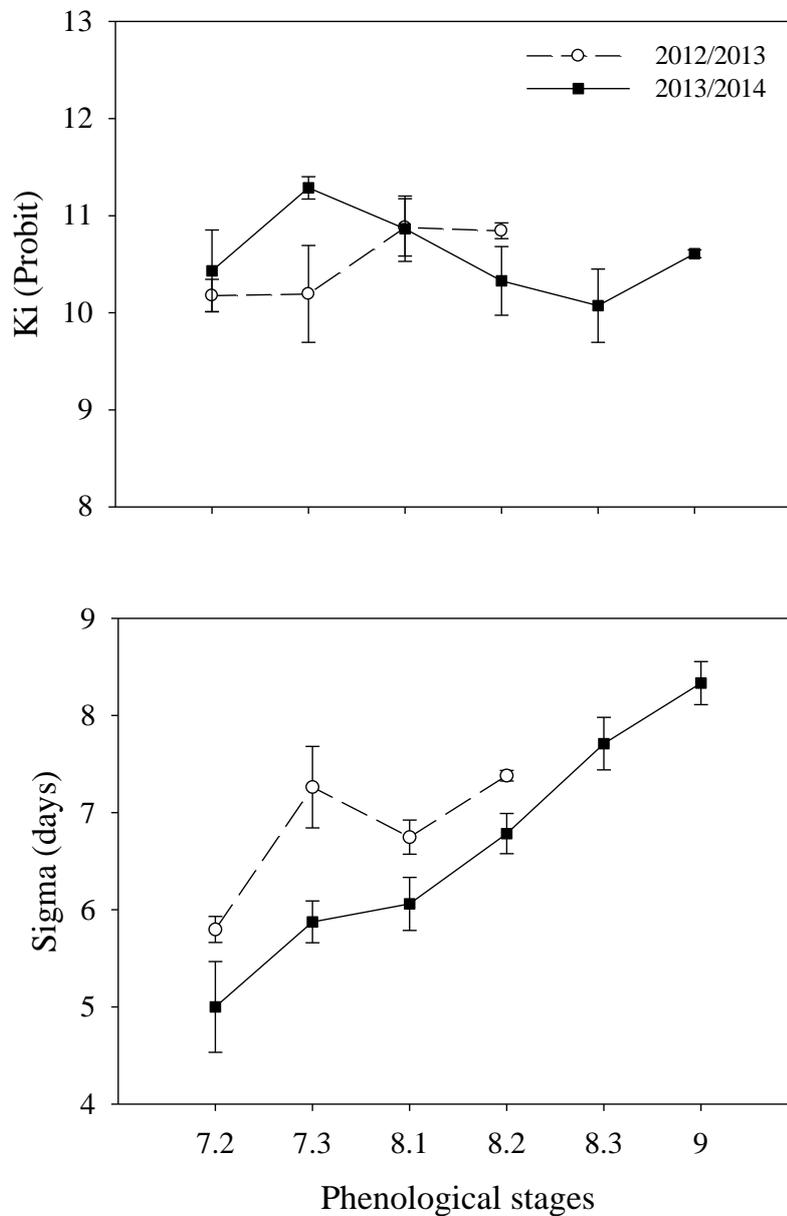
### III.2 Changes in soluble sugars during soybean seed maturation

Several mechanisms may be involved in the acquisition of seed longevity, including the accumulation of non-reducing sugars which are known as protective molecules involved in desiccation tolerance. Although their direct implication in seed longevity is not yet secured, their amount correlate with storability of orthodox seeds (STEADMAN et al., 1996; VANDECASTEELE et al., 2011). More recently, a correlation between amount of galactinol, a precursor of RFO and seed longevity was performed in three species of the Brassicaceae family (de SOUZA VIDIGAL et al., 2016). In this research, we studied the evolution of these sugars during the maturation phase of soybean seeds and investigated a possible correlation between their amount and the acquisition of seed longevity. This study was performed in fresh soybean seeds from stages R6 to R9 for the crop year 2013/2014.

Glucose and fructose content decreased concomitantly with the acquisition of desiccation tolerance (Figure 15 A and B). Fructose content decreased between stages R6 and R7.2 in axes and cotyledons, ranging from 0.09 to 0.02 mg / g DW in axes and 0.09 to 0.01 mg / g DW in cotyledons, respectively. Thereafter, the levels were very low until stage R9. Glucose was more abundant in axes compared to cotyledons. It gradually decreased in axes between R6 and R7.3 ranging from 0.25 to 0.15 mg / g DW, whereas in cotyledons the glucose content increased from R6 to R7.1 then decreased sharply from R7.1 to R7.2 (Figure 15 B). Thereafter, it decreased progressively between stages R7.3 and R9.

Sucrose was more abundant in axes than in cotyledons throughout the maturation period. However, its steady-state levels varied according to the organ studied. In axes, it slowly decreased from stage R7.1 onwards whereas in cotyledons, it decreased from R7.1 to R7.2 then remained steady at around 15 mg/g DW until stage R9 (Figure 15 C). The decrease of monosaccharides is inverse to the acquisition of longevity, which begins to be acquired between stages from R7.2 to R9. Content of fructose, glucose and saccharose is higher when longevity is low (stage R7.2).

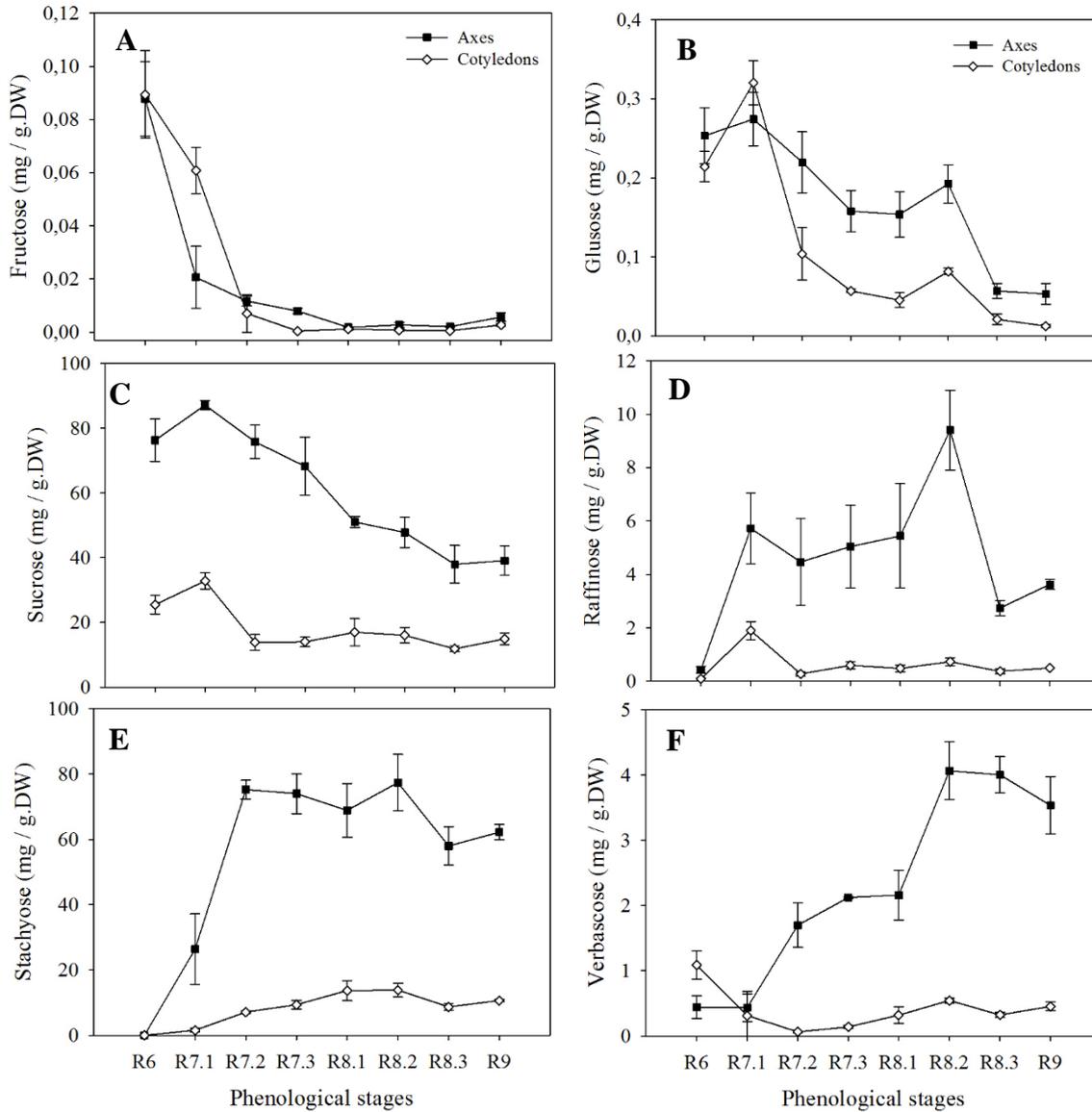
Like for sucrose, RFO (raffinose, stachyose and verbascose) were also more abundant in the axes than in cotyledons. Stachyose represent the most abundant RFO. In the axis, the concentrations of raffinose, precursor of stachyose and verbascose, increased from stage R6 to R7.1, remained stable, then peaked at stage R8.1 and then decreased at stage R9 (Figure 15 D). In contrast, the peak of raffinose occurred at stage R7.1 in the cotyledons. However, its low amount throughout maturation suggests that it may be rapidly metabolized in



**Figure 14.** Evolution of Ki (A) and Sigma (B) parameters determined by Ellis & Roberts (1980) model during maturation of soybean seeds from the crop years 2012/2013 (open circles) and 2013/2014 (black triangle).

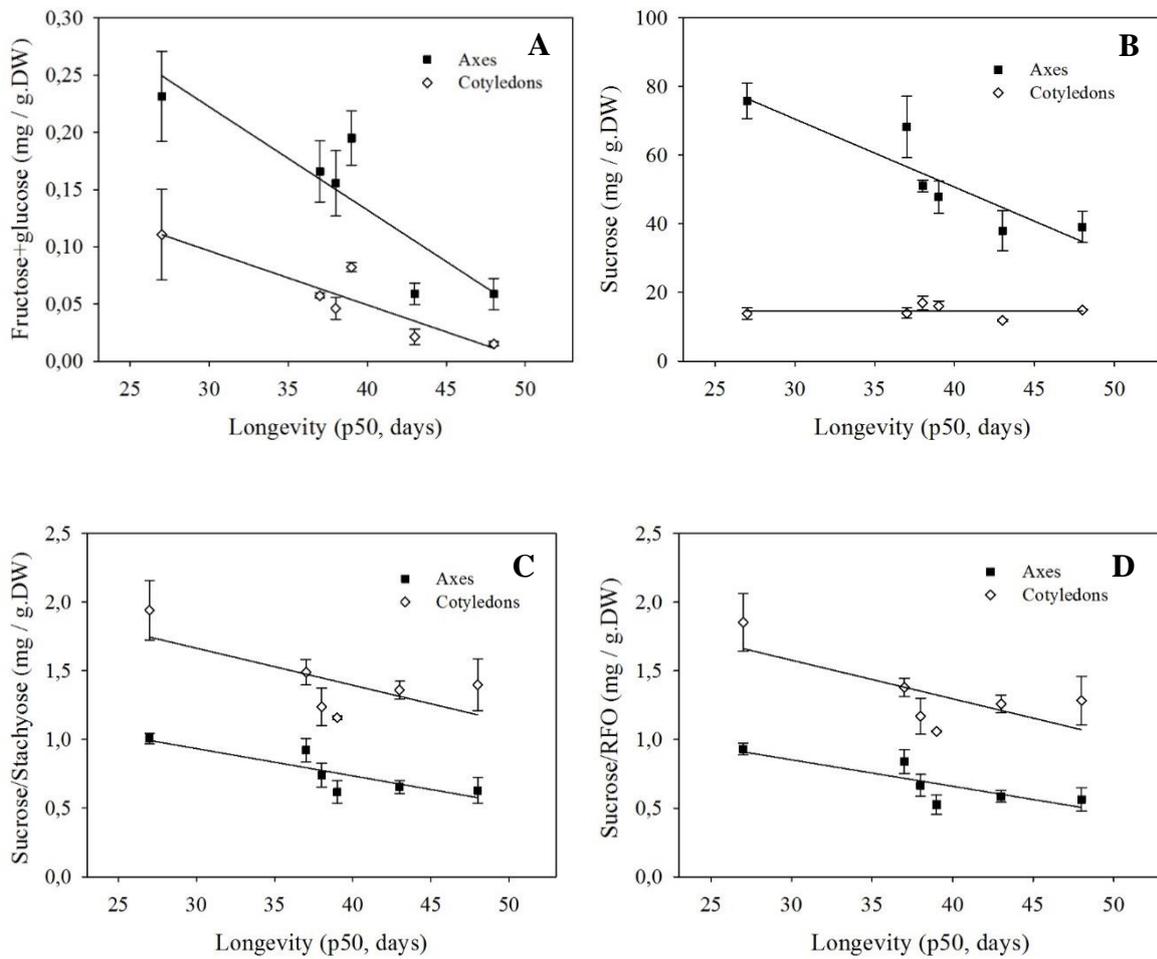
stachyose. In axes, stachyose were detected at stage R7.1 then its content increased dramatically at stage R7.2 and reached a constant value of about 73 mg / g DW until R9, that is when longevity increased by almost two fold. In the cotyledons, stachyose contents increased steadily until stage R8.1 (Fig. 15 E) and remained stable. In the axis, verbascose levels increased steadily of about 8 fold from stage R7.2 to R8.1, suggesting that the RFO metabolism remains active during the first phase of maturation drying.

Next, we investigated whether sugar content were correlated with the acquisition of longevity. For this purpose, P50 values were plotted against contents of the different sugars or different mass ratios (Fig 16) from stage R7.2 to R9. Stage R7.1 was not used to perform the correlation test because it was an outlier. There was a significant negative linear relationship between longevity from stage R7.2 to R9, i.e., from 27 to 48 days (p50) and monosaccharides ( $r = -0.886$ ;  $p = 0.05$  and  $r = -0.906$ ;  $p = 0.01$ ) in axes and cotyledons, respectively (Figure 16 A). A similar relationship was also found with sucrose content in axes ( $r = -0.896$ ;  $p = 0.05$ ). There was no significant correlation between longevity and sucrose content in cotyledons (Figure 16 B). There was a negative linear correlation between longevity and the sucrose/stachyose ratio in axes ( $r = -0.842$ ,  $p = 0.05$ ), suggesting that regulation of the metabolism of these sugars and acquisition of longevity may be coupled (Figure 16 C). In legumes, the ratio between sucrose and RFO represent a good indicator of seed vigor (VANDECASTEELE et al., 2011). In our experiment, the Suc/RFO ratio decreased drastically before the acquisition of seed longevity (Fig. 16 D) then remained more or less constant. Sucrose/RFO ratios in axes was negatively correlated with longevity ( $r = 0.882$  ;  $p = 0.01$ ), however there was no significant correlation in cotyledons.



**Figure 15.** Changes in non-reducing sugars content during maturation of soybean seeds. Content in fructose (A), glucose (B), sucrose (C), raffinose (D), stachyose (E) and verbascose (F) are shown for axes (black square) and cotyledons (open diamond). Data are the mean of 3 replicates. Bars represent the standard deviation.

From these results, we assume that non-reducing sugars can not explain the increase in longevity. More sampling points during maturation are required to obtain a more reliable and conclusive correlation between sugar contents and longevity. Probably, there are other mechanisms in addition to the non-reducing soluble sugars that take place at the late maturation phase and may contribute on acquisition of longevity in soybean seeds.



**Figure 16.** Relation between longevity (p50) and monosaccharides (fructose + glucose) (A), sucrose content (B), ratio sucrose/stachyose (C) and sucrose/RFO (D) in axes (black square) and cotyledons (open diamond) of soybean seeds during acquisition of longevity. Polynomial model and a linear equation were used.

### **III.3. Dynamic changes in the transcriptome during the late maturation phase in soybean seeds**

Whole soybean seeds comprised of cotyledons, axes, and seed coat, were sampled at seven stages of seed development, from stage R7.1 to R9 from crop year 2013/2014. These stages were chosen because they correspond to post embryo development and the beginning of the acquisition of longevity after the acquisition of DT (Figure 5). After RNA extraction, the sequencing of RNA was performed by Illumina HiSeq 2500 platform in triplicate (Lane 1, 2 and 3), generating between 14 and 38 million of reads per phenological stage, totaling more than 176 million of reads from mRNA libraries (Table 4), which were available on file in Fastq format. The amount of reads were aligned to the soybean reference genome, Williams 82 genome assembly 2.0 sequence, annotated by *Joint Genome Institute* - JGI (GOODSTEIN et al., 2012). The number of mapped and unmapped reads and their respective percentages are shown at Table 5. Between 13 and 34 million of reads were mapped (a total of up to 161 million of reads), it represents more than 90% of the reads mapped. Less than 11% of the amount of reads were not mapped, it shows the high sensitivity and specificity of the alignment to the reference genome. Read counts (number of reads mapped to a given gene) were estimated from the sequence alignment data for all the annotated gene models, which allowed the identification of 56044 transcripts presents in 88,647 transcripts from reference genome.

**Table 4.** Number of reads per line obtained from total RNA extracted from seeds harvested at the indicated stages during maturation. A and B correspond to biological replicates.

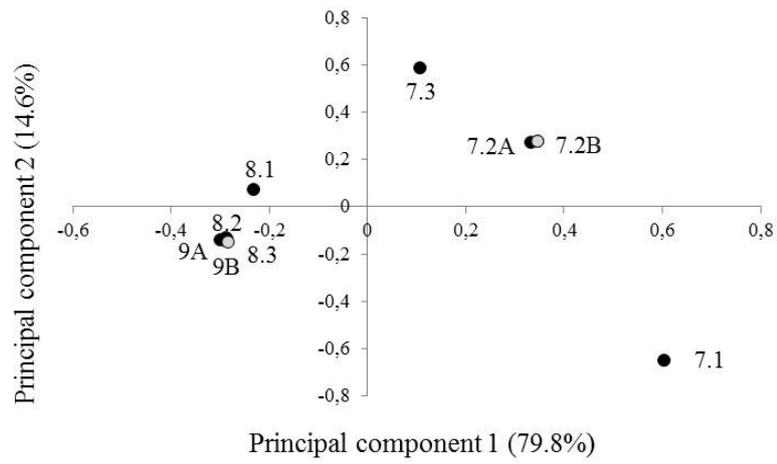
Stage	#Reads			Total
	Lane 1	Lane 2	Lane 3	
7.1	8,508,855	8,579,093	8,585,083	25,673,031
7.2 A	12,617,101	12,724,090	12,747,648	38,088,839
7.2 B	11,378,954	11,494,158	11,507,896	34,381,008
7.3	7,373,672	7,432,068	7,461,631	22,267,371
8.1	4,813,779	4,848,066	4,871,492	14,533,337
8.2	7,311,411	7,365,578	7,380,419	22,057,408
8.3	8,231,449	8,315,885	8,317,657	24,864,991
9A	9,543,605	9,635,004	9,645,974	28,824,583
9B	7,630,951	7,706,218	7,272,435	23,064,604

**Table 5.** Mapping of single-end reads to the soybean genome. # reads, number of reads following trimming of the libraries for quality; # mapped reads, number of reads that unambiguously mapped to the soybean genome; # unmapped, number of reads that could not be mapped to a single location in the soybean genome. Percentages of unmapped and mapped sequences are also indicated (relative to number of reads).

Stages	# reads	# mapped	% mapped	#unmapped	%unmapped
R7.1	25673031	24112015	93,92	1561016	6,08
R7.2	38088839	34971702	91,82	3117137	8,18
R7.3	22267371	20475504	91,95	1791867	8,05
R8.1	14533337	13392050	92,15	1141287	7,85
R8.2	22057408	19621133	89,95	2436275	11,05
R8.3	24864991	22707980	91,33	2157011	8,67
R9	28824583	26457357	91,79	2367226	8,21

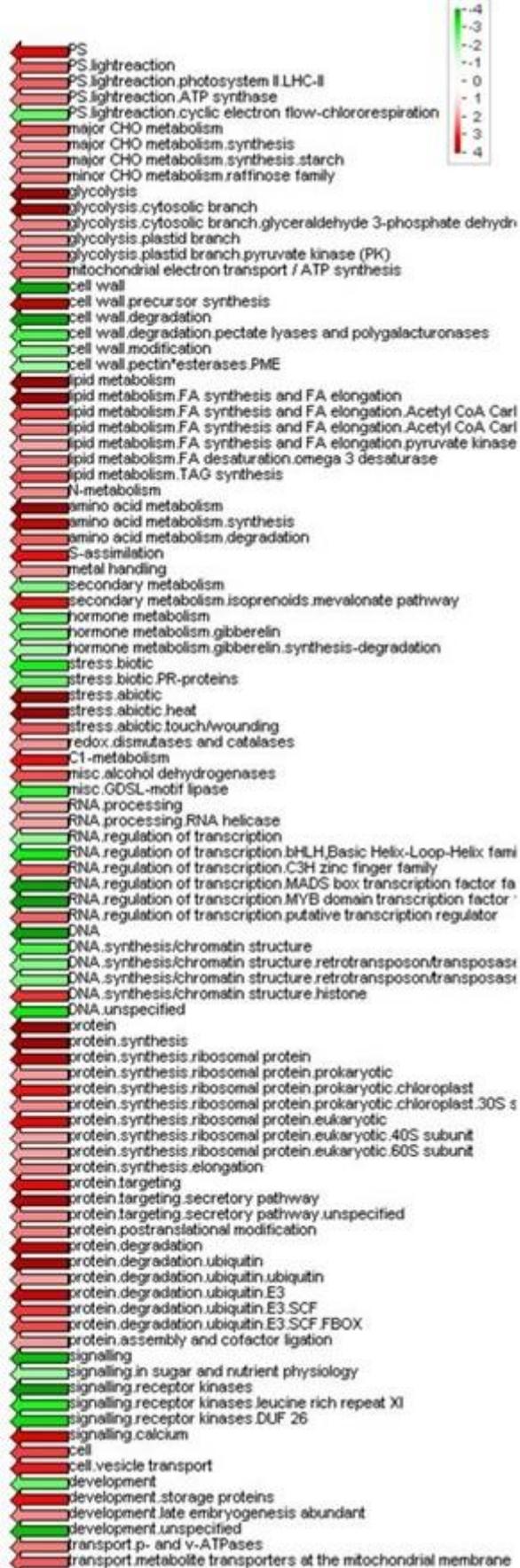
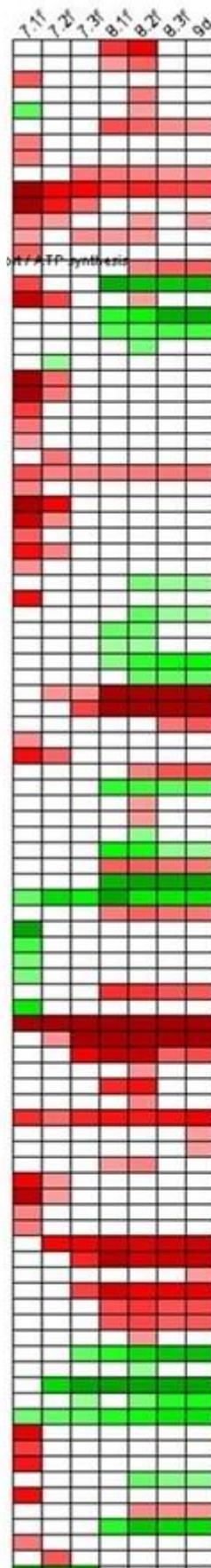
There were 56,044 transcripts detected throughout development. A global transcriptome analysis was carried out to identify changes in gene expression in soybean seeds during maturation. After normalization and background removal, data were filtered to remove transcripts with variance among stages  $< 1$ . From this data set, 13,477 transcripts were selected as differentially expressed throughout the complete seed maturation and retained for further analysis.

Principal component analysis (PCA) was performed to highlight the dynamic changes of transcriptomes from the different phenological stages (Figure 17). The kinetics of seed maturation is visible in the first and second component, representing 80 and 14% of the transcriptome variation, respectively. This variation was mostly attributed to developmental mechanisms that occurred mainly between stage R7.1 to R8.1. Transcriptomes of stages R7.1, R7.2 and R7.3 were very dissimilar, indicating that despite the arrest of seed filling and acquisition of desiccation tolerance, major changes in gene expression can still happen. Interestingly, R7.3 was separated from the rest of the transcriptomes in both components of the PCA plot, suggesting that this stage is pivotal to maturation. Transcriptomes of stages R8.1, R8.2, R8.3 and R9 were clustered together, forming a discrete group, suggesting that putative regulatory processes governing the late phases of maturation did not occur at the transcription level. Interestingly, a largely shift in the transcriptome occurred at R7.3 in both directions. We also obtained a biological duplicate for the transcriptomes at stages R7.2 and R9 (Figure 17). For both stages, the replicates clustered together, indicating a good reproducibility of our sampling method and sequencing.



**Figure 17.** Principal component analysis of the transcriptomes during seed maturation. PCA was performed using median centering. Gray circles are replicates of stages R7.2 and R9.

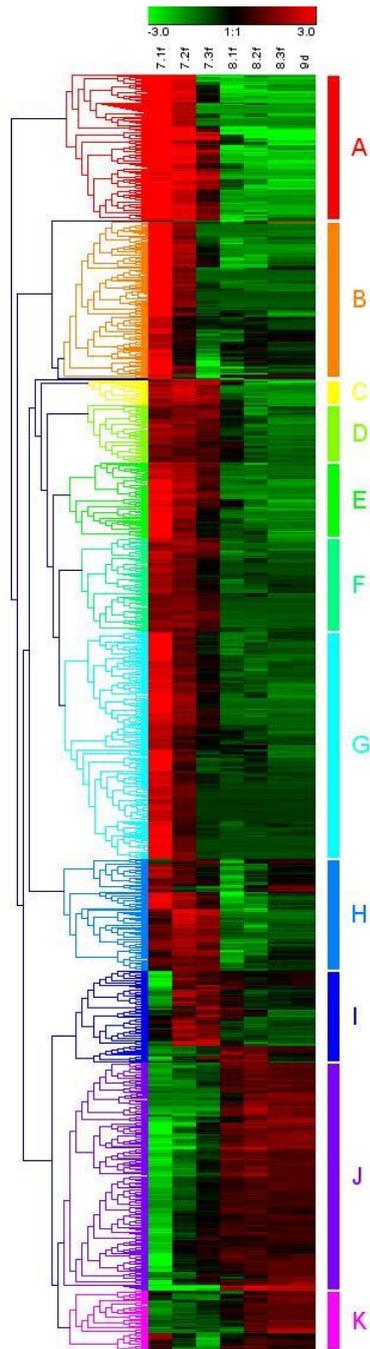
An overrepresentation analysis of functional classes was performed to obtain an overview of the different biological process that occurs at the different stages of maturation. For this purpose, we used Pageman (MAPMAN, GABI Germany), a gene ontology (GO) tool that generate annotated profiles of differentially expressed genes (Figure 19, USADEL et al. (2006)). This analysis compares a gene set of interest to a reference set of a certain functional category as a GO term. As result, categories that are over-represented in the experimental gene set appear in red and the categories that are under-represented appear in green.



**Figure 18.** Over-representation analysis of functional classes during soybean seed development using the PageMan software. Data were subjected to a bin-wise Wilcoxon test, and resulting P-values were adjusted according to Benjamin Hochberg. Over-represented and under-represented classes are indicated in red and green, respectively. Functional classes and subclasses statistically affected are indicated according to Map- Man ontology.

At stage R7.1 when seeds are still green (Figure 5), there is an over-representation for photosystem II, carbohydrates synthesis, fatty acids synthesis and N-metabolism. At the end of seed filling (stages R7.1 and R7.2) there is overrepresentation for photosynthetic activity, glycolysis, amino acid metabolism, lipid synthesis (top of Fig. 18), protein targeting, vesicle transport and development of storage protein (bottom part of Fig. 18). While genes associated with DNA synthesis were under-represented (Fig. 18). This suggests that despite seed filling has ended in terms of accumulation of dry weight. The expression of genes associated with synthesis of storage reserves is not yet turned off. This occurred probably at stage R7.3 considering the absence of over-representation of these categories. The shift in transcriptome observed at R7.3 on the PCA plot was also observed in Figure 18 where many categories started to be over- or under- represented at this stage (when the seeds lose their chlorophyll, Figure 5) or stage R8.1. The over-represented categories were RFO metabolism, response to abiotic stress, ribosomal proteins synthesis and degradation, in particular, genes associated with protein degradation via the SCF family of modular E3 ubiquitin pathway and genes associated with histone and chromatin modification. The under-represented categories were cell wall degradation, GA metabolism and those associated with biotic stress, including stress biotic PR proteins (representative of passive defense against pathogens) and signaling receptor kinases that contain a LRR motif (bottom of Figure 18, representative of the pathogen response pathway). Categories associated with regulation of transcription were also found to be in our transcriptome at stage R8.1 onwards. Some were over-represented such as Zn finger family whereas others were under-represented such as MYB domain transcription factors (middle of Fig. 18).

Since Figure 18 suggests a complex transcriptional regulation of transcription factors during the late phase of seed maturation, we examined in more details the kinetics of all the transcription factors (TF) that were differential expressed in developing soybean seeds. After filtering, a list of 1,086 TFs could be retrieved and clustered based on their expression profiles (Figure 19). A number of 32 families represents the TFs previously



**Figure 19.** Clustering displaying the TFs during maturation of soybean seeds. The color scale ranges from saturated green for log ratios  $-3.0$  and below to saturated red for log ratios  $3.0$  and above. Each gene is represented by a single row of colored boxes; each phenological stage is represented by a single column. Eleven separate clusters are indicated by colored bars and by identical coloring of the corresponding region of the dendrogram.

filtered (Table 7). Eleven clusters were obtained. Clusters A, C, D, E, F and H had a similar expression patterns, characterized by a high level of transcripts during early maturation (stages R7.1 to R7.3) and low abundance during the end of maturation. Because of this, we considered as a single cluster called ACDEFH. For the same reason the cluster B and G and J and K were considered as BG and JK, respectively. BG corresponded to differentially expressed genes that decreased early during maturation, before the induction of longevity whereas JK corresponded to transcripts that appeared during the late phase of maturation drying.

Transcripts from cluster I and JK were studied. Cluster I exhibits a transitory profile characterized by transcript levels reaching a peak at stages R7.2 and R7.3 (Figure 19). This leads us to assume that the TFs from cluster I might be involved in regulating the acquisition of seed longevity and/or the preparation of the dry state.

We calculated which families were significantly enriched in cluster I and JK. AP2/EREBP and WRKY families were most represented at cluster I with 20% and 12.31%, respectively (Table 6). AP2/EREBP is one the largest family of TFs with pleotropic roles involved in stress adaptation and development like flower organ formation. In this respect, the three most highly abundant transcripts whose levels peaked at R7.3 were encoding genes homologue to the DREB family (Dehydration-Responsive Element Binding protein) transcription factor and AIL6 (AINTEGUMENTA-LIKE 6), respectively. ABI3/VP family was not over-represented. Within the WRKY family, the most differentially expressed gene within cluster I was an homologue of WRKY72, which contributes to basal immunity against pathogens in tomato and Arabidopsis.

In the cluster JK the most representative families of TFs were AP2/EREBP, bHLH and putative transcription regulators (with 10.5; 11.60 and 11.05%, respectively). Members of the basic helix-loop-helix family (bHLHs) of constitutive nuclear TFs play a central role in the molecular mechanisms of phytochrome signal transduction (DUEK; FANKHAUSER, 2005). In our results, PIF homologues (PIF7, Photoreceptor interacting factors) was present, which interacts specifically with the far-red light-absorbing Pfr form of phyB.

**Table 6.** Representation of families of transcription factors whose gene expression were clustered using K-means .

Family of transcription fator	ACDEFH (%)	BG (%)	I (%)	JK (%)
ABI3/VP1	0,30	0,40	1,54	0,00
AP2/EREBP	8,04	7,54	<b>20,00</b>	10,50
ARF	1,79	3,17	1,54	2,76
ARR	1,49	1,59	0,00	1,10
AS2/LOB	1,19	1,59	3,08	3,31
B3	0,89	1,98	0,00	0,55
bHLH	13,39	11,11	6,15	11,60
bZIP	5,06	5,95	1,54	4,97
C2H2 family	9,52	6,35	6,15	3,87
C3H	2,38	0,79	0,00	2,21
CO-like	2,08	1,19	1,54	1,66
CPP	0,00	0,40	0,00	0,00
DOF	3,27	1,98	3,08	2,21
EIL	0,30	0,40	0,00	0,55
Ets	0,00	0,40	0,00	0,00
G2-like	1,79	1,98	1,54	5,52
GATA	2,38	3,17	1,54	1,10
GePB	0,00	0,40	0,00	0,55
GRAS	2,98	3,97	0,00	3,87
Homeobox	6,85	9,13	9,23	6,63
HDA	0,30	0,79	0,00	0,00
HSF	1,79	0,00	1,54	3,31
JUMONJI	0,30	0,00	0,00	0,55
MADS box	5,95	3,57	6,15	1,66
MYB	7,74	11,90	7,69	7,18
MYB-related	2,38	1,98	3,08	4,42
Putative transcription regulator	<b>3,87</b>	8,73	9,23	11,05
TCP	2,38	0,40	0,00	0,55
Trihelix	0,89	2,78	3,08	2,76
WRKY	6,55	4,76	12,31	4,42
YABBY	1,79	0,40	0,00	0,00
zf-HD	2,38	1,19	0,00	1,10
Total (%)	100	100	100	100
Total (number of genes)	336	251	64	181

### **III.4 Differential expression analysis of genes associated with longevity.**

To identify genes that might be involved in longevity of soybean seeds, we performed a differential expression study by using RNAseq at stages R7.2 and R9, respectively. Longevity of seeds at stage R9 are almost 2-fold higher than longevity of seeds harvested and dried at stage R7.2 (Figure 12). In addition, RNAseq was carried out in duplicates for each stages, which allowed us to perform statistical analysis. Differential gene expression analyses were performed using the statistical tool, DESeq, v1.12.1 (Anders et al., 2010). Fold-change (FC) was calculated by dividing mean normalized gene expression value in R9 over that in R7.2. Most significant genes were identified at 0.05 false discovery rate (FDR) calculated using P-value adjusted for multiple testing using Benjamini-Hochberg method (Severin et al., 2010).

We identified a total of 3,063 genes with significant differential expression between stages R7.2 and R9. Of these differentially expressed genes (DEGs), 1248 (40.7 %) and 1815 (59.3 %) transcripts were identified as up- and down-regulated, respectively. The number of DEG with  $\log_2$  FC ratio of  $>1$  and  $<-1$  was 742 and 1525, respectively (Table 7).

Functional enrichment of gene ontology (GO) was performed with homologes of Arabidopsis using the AgriGO platform ([bioinfo.cau.edu.cn/agriGO/analysis.php](http://bioinfo.cau.edu.cn/agriGO/analysis.php)) (DU et al, 2010) with the following settings: TAIR9 background Fisher test with the Benjamini-Hochberg method for multiple testing to obtain adjusted P-values. The enriched results were then filtered using adjusted-P-value  $\leq 0.01$  (or 1 % FDR) to obtain significant enriched terms. Table 8 shows highly significant enriched GO terms identified for up regulated with  $\log_2$  FC R9/R7.2 ratio  $> 1$ .

**Table 7.** Number of differentially expressed genes in seeds during late seed maturation by comparing stage R7.2 and R9 .

<i>Log<sub>2</sub> FC R9/R7.2</i>	<b>&gt;1</b>	<b>&gt;2</b>	<b>&gt;3</b>
# transcripts up regulated in R9	742	64	09
	<b>&lt;-1</b>	<b>&lt;-2</b>	<b>&lt;-3</b>
# transcripts down regulated in R9	1525	281	20

**Table 8.** GO enrichment analysis of DEG that increased with longevity. Analysis was performed with transcripts exhibited a FC R9/R7.2 ratio > 1 using AgriGO with the Arabidopsis TAIR9 background.

<b>GO Term</b>	<b>Biological process</b>	<b>DEG</b>	<b>p-value</b>	<b>FDR</b>
GO:0009408	response to heat	25	2,0E-16	3,2E-13
GO:0009628	response to abiotic stimulus	70	9,1E-16	4,9E-13
GO:0009791	post-embryonic development	47	7,3E-16	4,9E-13
GO:0009266	response to temperature stimulus	34	2,2E-12	9,0E-10
GO:0006950	response to stress	82	1,4E-11	4,4E-09
GO:0050896	response to stimulus	116	2,4E-10	6,4E-08
GO:0010035	response to inorganic substance	23	4,7E-10	9,3E-08
GO:0000302	response to reactive oxygen species	14	4,3E-10	9,3E-08
GO:0032501	multicellular organismal process	70	4,7E-09	8,4E-07
GO:0007275	multicellular organismal development	66	3,1E-08	5,0E-06
GO:0006979	response to oxidative stress	21	1,9E-07	2,7E-05
GO:0003006	reproductive developmental process	39	2,5E-07	3,3E-05
GO:0034641	cellular nitrogen compound metabolic process	26	3,3E-07	4,0E-05
GO:0032502	developmental process	69	3,6E-07	4,1E-05
GO:0042542	response to hydrogen peroxide	9	4,6E-07	4,9E-05
GO:0009642	response to light intensity	11	4,9E-07	4,9E-05
GO:0048608	reproductive structure development	38	6,6E-07	6,2E-05
GO:0009314	response to radiation	28	1,1E-06	9,1E-05
GO:0006807	nitrogen compound metabolic process	98	1,3E-06	1,0E-04
GO:0009416	response to light stimulus	27	1,9E-06	1,3E-04
GO:0009644	response to high light intensity	7	5,8E-05	2,1E-03

Most of the DEGs associated with significantly enriched biological processes were associated with response to stress. The most significant biological processes were ‘response to heat’ (GO:0009408), ‘response to abiotic stimulus’ (GO:0009628) and ‘post-embryonic development’ (GO:0009791). A deeper inspection of the DEGs belonging to these biological processes revealed mostly the presence of homologs of heat shock protein and HSF. Interestingly, biological processes associated with response to light were also over-represented. This GO category contained several genes homologous to HSP and PHYA and PHYB.

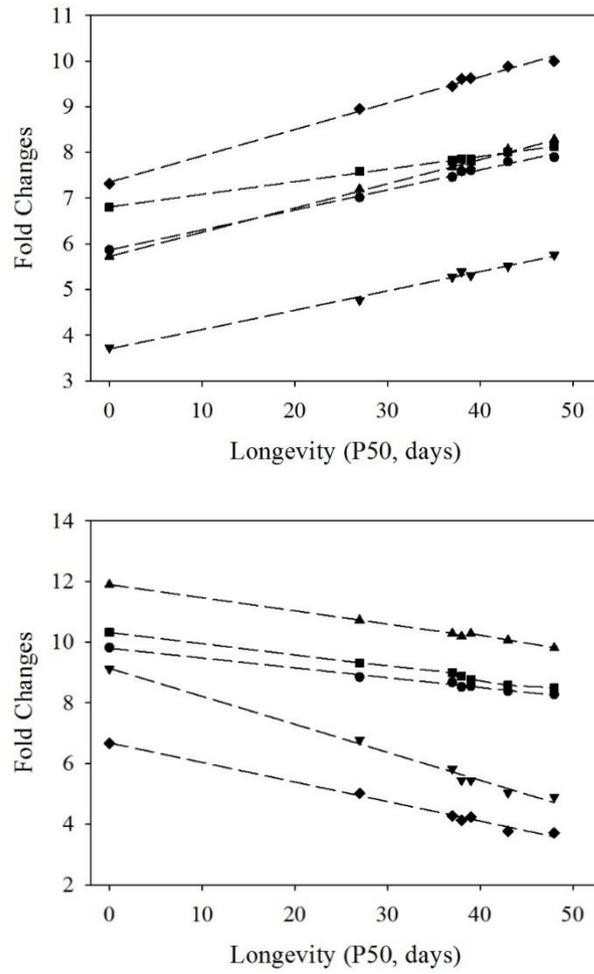
The GO categories that were associated with DEGs that were down-regulated are shown in table 9. ‘Response to stimulus’ (GO:0050896), ‘response to stress’ (GO:0006950) and ‘response to abiotic stimulus’ (GO:0009628) were the most significant. GO associated with photosynthesis was also over-represented, despite the fact that developing seeds at stage R7.2 had already lost their chlorophyll. This category contained mainly genes encoding proteins from the photosynthetic electron transport chain in photosystem II and chloroplast-encoded genes.

Next, we calculated the coefficient correlations between DEGs and acquisition of longevity by plotting P50 values against log<sub>2</sub> values. We selected those DEGs that showed a high correlation (>0.95) as illustrated in Figure 23 A and Figure 23 B for up- and down-regulated genes.

**Table 9.** GO enrichment analysis of DEG that decrease with longevity. Analysis was performed with transcripts exhibited a FC R9/R7.2 ratio < -1 using AgriGO with the Arabidopsis TAIR background.

<b>GO term</b>	<b>Description</b>	<b>DEG</b>	<b>p-value</b>	<b>FDR</b>
GO:0050896	response to stimulus	210	4,8E-18	3,5E-16
GO:0006950	response to stress	132	3,7E-14	1,3E-12
GO:0009628	response to abiotic stimulus	91	8,0E-12	1,9E-10
GO:0015979	Photosynthesis	26	1,1E-11	2,0E-10
GO:0008152	metabolic process	382	7,5E-09	1,1E-07
GO:0051179	Localization	95	1,4E-07	1,5E-06
GO:0006091	generation of precursor metabolites and energy	27	1,5E-07	1,5E-06
GO:0009791	post-embryonic development	45	7,8E-07	6,2E-06
GO:0009719	response to endogenous stimulus	60	7,1E-07	6,2E-06
GO:0044237	cellular metabolic process	305	7,6E-06	5,5E-05
GO:0006810	Transport	85	8,8E-06	5,8E-05
GO:0051234	establishment of localization	85	9,7E-06	5,8E-05
GO:0019748	secondary metabolic process	32	1,9E-05	1,0E-04
GO:0009987	cellular process	388	2,9E-05	1,5E-04
GO:0009058	biosynthetic process	183	3,0E-04	1,5E-03
GO:0005975	carbohydrate metabolic process	43	3,5E-04	1,5E-03
GO:0009605	response to external stimulus	26	3,4E-04	1,5E-03
GO:0044249	cellular biosynthetic process	174	7,5E-04	3,0E-03
GO:0009607	response to biotic stimulus	32	1,6E-03	6,1E-03

These amounted to 1925 for up-regulated transcripts and 1889 for down regulated ones. Using this list, a GO enrichment analysis was performed as described above. The main GO categories for up-regulated transcripts were associated with post-embryonic and seed development (Table 10). They contained a high proportion of genes associated with ribosome and ribonuclease activity such as the homologues of emb2171 (embryo defective 2171; a structural constituent of ribosomes), emb2444 (a RNA binding protein), DOMINO1 (which is thought to have a role in ribosome biogenesis), emb1687 (having a ribonuclease activity) and several other embryo defective genes with unknown function (MEE14 (maternal effect embryo arrest), MEE11, emb2738, emb1789, emb 2737). The GO “organelle organization” contained genes with a variety of functions but we noticed a high proportion of homologs associated with chromatin conformation and chromatin silencing such as GCN5 (a histone acetyltransferase), two histone H1/H5, HDA15 (histone deacetylase), VIM1 (Variant in methylation 1, a protein binding to methyl cytosines of CG, CNG and CNN motifs), TFL2 (Terminal flower 2, a gene involved in chromatin silencing associated with the epigenetic maintenance of the vernalized state in the inflorescence meristem), SUVH5 (SU(VAR)3-9 homolog 5, a histone methyltransferase contributing to maintenance of H3mK9).



**Figure 20.** Relationship between P50 taken obtained from seeds collected between stages 7.1 and 9 and log ratios of up-regulated (A) and down-regulated genes (B). Data shown corresponds to the top 5 probes exhibiting both the highest correlation coefficient and expression levels.

The GO enrichment analysis of down-regulated genes generated a long list including 490 terms that remained very broad and not very instructive. The 13 most significant, based on FDR, are shown at Table 11. When we restricted the analysis with the top 50 genes that showed a strong decrease in transcript level during the acquisition of longevity, the term cellular metabolic process (GO: 0044237) was obtained (P value =2,1 E-4, FDR= 0.032). This finding probably reflects the maturation drying that imposes a progressive decrease in metabolic activities.

**Table 10.** GO enrichment analysis of DEG positively correlated with longevity at  $p < 0.001$  (1925 transcripts) using AgriGO with the Arabidopsis TAIR background.

<b>GO term</b>	<b>Description</b>	<b>DEG</b>	<b>p-value</b>	<b>FDR</b>
GO:0009791	post-embryonic development	75	1,10E-13	2,9E-10
GO:0009790	embryonic development	48	3,50E-07	4,6E-04
GO:0009987	cellular process	539	1,60E-06	1,1E-03
GO:0009793	embryonic development ending in seed dormancy	42	1,60E-06	1,1E-03
GO:0048316	seed development	44	6,70E-06	3,5E-03
GO:0010154	fruit development	45	1,00E-05	4,4E-03
GO:0032501	multicellular organismal process	122	1,20E-05	4,6E-03
GO:0003006	reproductive developmental process	67	1,50E-05	4,9E-03
GO:0007275	multicellular organismal development	117	2,40E-05	6,8E-03
GO:0006996	organelle organization	48	3,10E-05	8,1E-03
GO:0044237	cellular metabolic process	404	5,10E-05	1,1E-02
GO:0006839	mitochondrial transport	12	5,20E-05	1,1E-02
GO:0008152	metabolic process	481	5,60E-05	1,1E-02
GO:0032502	developmental process	128	5,80E-05	1,1E-02
GO:0007005	mitochondrion organization	9	7,30E-05	1,3E-02

**Table 11.** GO enrichment analysis of DEG negatively correlated with longevity at  $p < 0.001$  (1889 transcripts) using AgriGO with the Arabidopsis TAIR background.

<b>GO term</b>	<b>Description</b>	<b>DEG</b>	<b>p-value</b>	<b>FDR</b>
GO:0009987	cellular process	620	5,20E-23	1,70E-19
GO:0008152	metabolic process	564	2,50E-20	4,20E-17
GO:0044237	cellular metabolic process	458	9,10E-15	1,00E-11
GO:0044238	primary metabolic process	467	2,70E-14	2,20E-11
GO:0050896	response to stimulus	248	8,30E-14	5,50E-11
GO:0044262	cellular carbohydrate metabolic process	53	3,80E-13	2,10E-10
GO:0009791	post-embryonic development	72	5,20E-13	2,40E-10
GO:0006796	phosphate metabolic process	100	6,20E-13	2,40E-10
GO:0006793	phosphorus metabolic process	100	6,50E-13	2,40E-10
GO:0034641	cellular nitrogen compound metabolic process	58	1,50E-12	4,90E-10
GO:0016310	Phosphorylation	91	9,80E-12	3,00E-09
GO:0043412	macromolecule modification	118	9,60E-11	2,60E-08
GO:0006464	protein modification process	109	1,40E-10	3,50E-08

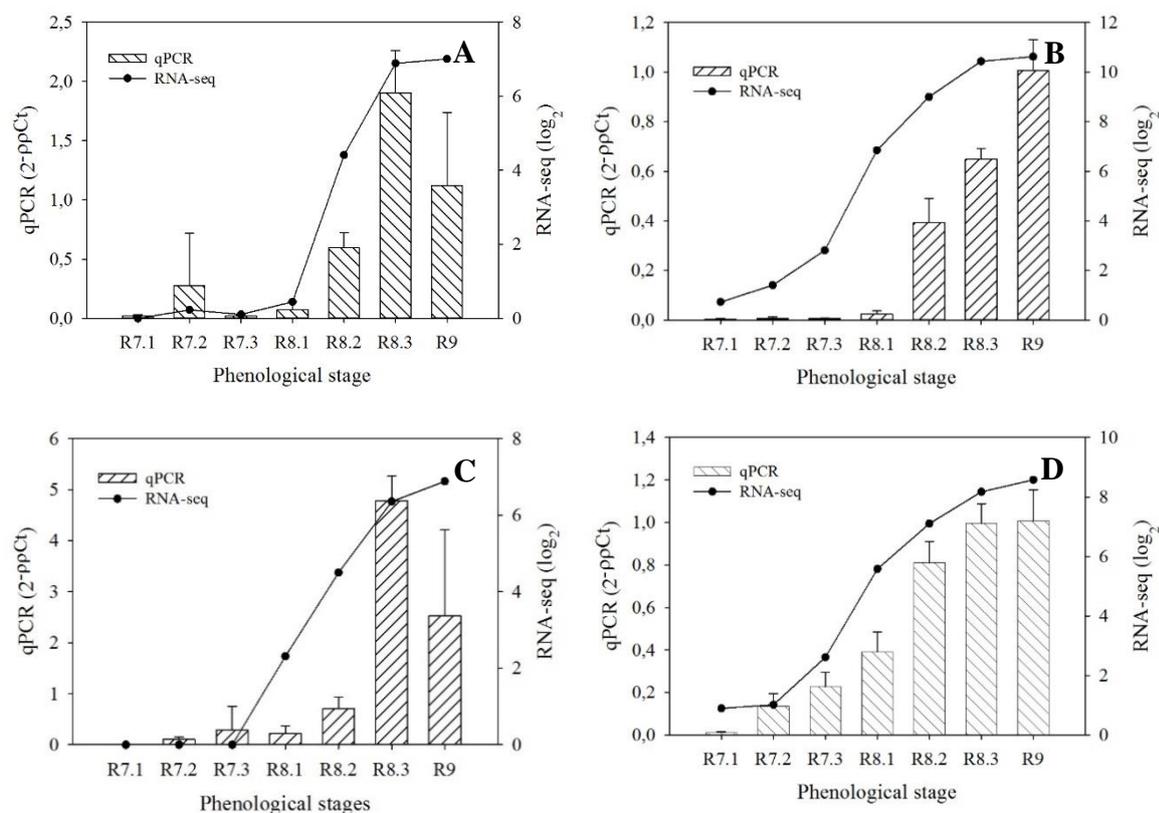
Some genes positively correlated with longevity had their expression confirmed by using RT-qPCR. The objective was to confirm the accuracy of the RNA-Seq data. There were highly significant correlations between relative gene expression profile obtained by RT-qPCR and RNASeq analyses at the 0.05 significance level by Pearson correlation test (Table 12). Figure 21 shows the selected genes coding for heat shock proteins - HPS (sHSP17.6, HSP101, HSP21). These genes were at the list of transcripts that increased during seed maturation.

The genes coding for transcription factors HSFA3 (HEAT SHOCK TRANSCRIPTION FACTOR A3; HSFA6 (heat shock transcription factor a6b); bZIP 65,;GNC5 (HISTONE ACETYLTRANSFERASE OF THE GNAT FAMILY 1) and NF-X-like 1, were selected for expression studies by using RT-qPCR (Figure 22). The ratio of the relative expression profile between RNA-Seq and RT-qPCR analysis was significant and positive for genes encoding both HSFA3 and HSF6B (Figure 22 A and B, respectively), GNC5 (Figure 22 D) and NFXL1 (Figure 22 E). In general, the expression profiles obtained from RNA Seq data were similar to those obtained by RT-qPCR.

**Table 12.** Pearson correlation coefficient (r value) of gene expression from RNAseq data and RT-qPCR .

Glyma2.0	Name of gene	r value
Glyma.17G224900	17.6 kDa class II heat shock protein	0,840*
Glyma.14G099900	17.6 kDa class II heat shock protein	0,859*
Glyma.17G077500	Heat shock protein 101	0,813*
Glyma.08G318900	heat shock protein 21	0,950**
Glyma.03G191100	HSFA3	0,739
Glyma.03G157300	HSFA6B	0,826*
Glyma.03G142400	bZIP65	0,798*
Glyma.19G160700	GCN5	0,712
Glyma.09G173000	NFXL1	0,793*

Significant at \*  $p < 0.05$  and \*\*  $p < 0.01$



**Figure 21.** Relative gene expression for RNA-Seq data validation at respective seed developmental stages for genes encoding (A) and (B) 17.6 kDa class II Heat shock protein (HSP 17.6: Glyma17G224900 and Glyma14G000009, respectively); (C) Heat shock protein 101 (HSP101: Glyma17G077500); (D) Heat shock protein 21 (HSP21: Glyma08G318900). Bars and line indicate expression values from qPCR and RNA-Seq experiments, respectively.

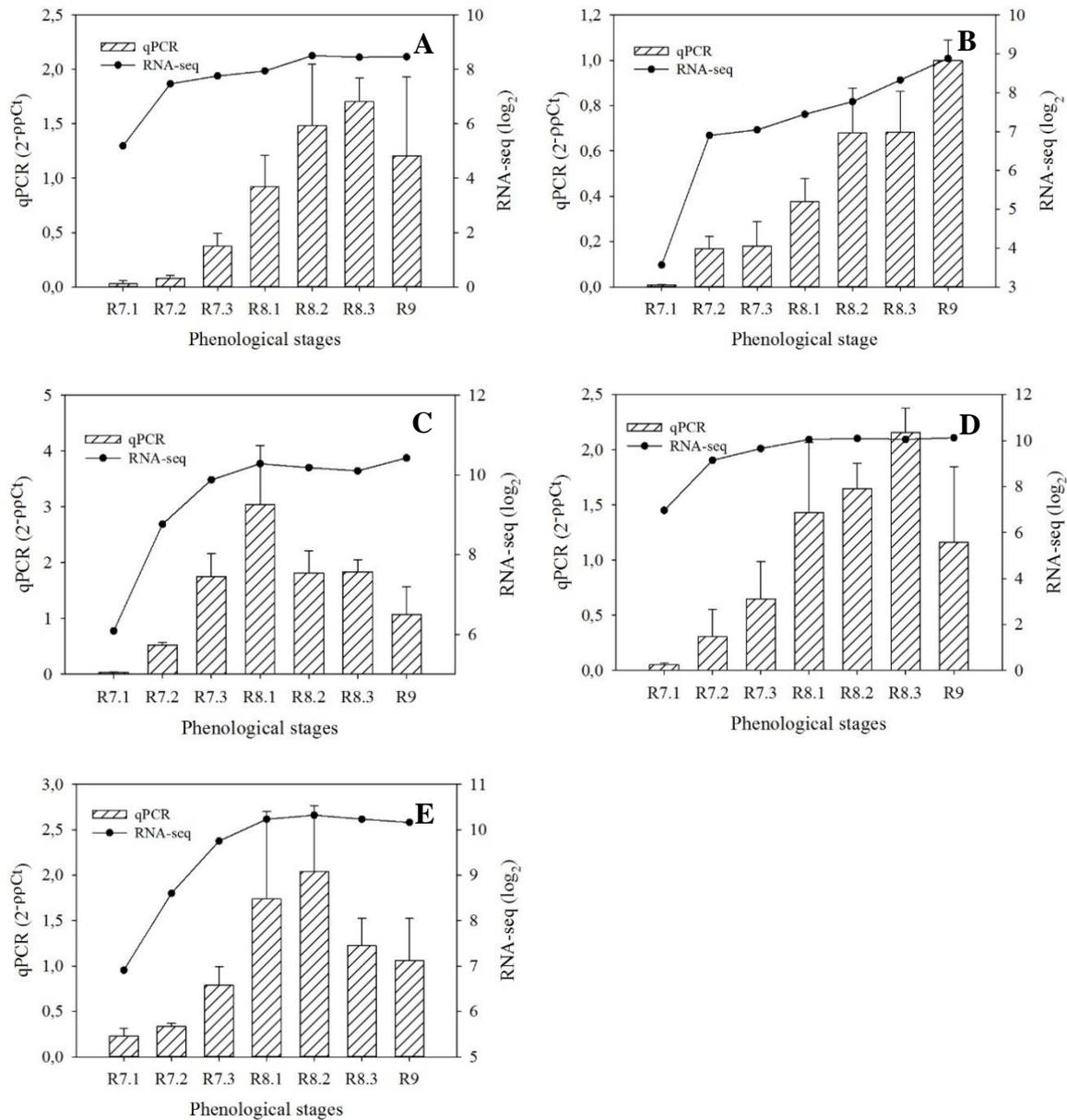
## Chapter IV. DISCUSSION

### IV.1 Acquisition of longevity and non-reducing soluble sugars accumulation are related in soybean seeds.

In this study, we attempted to know how longevity is regulated during seed maturation. Longevity is a key attribute of physiological quality. Seed quality is highly important for crop establishment since one of the major problem in soybean seed production in the tropics is the fast loss of seed viability during storage due to environmental conditions of high temperature and relative humidity (NKANG & UMHO, 1996). This work aimed also to get new insights into the maturation processes that occur after seed filling and prepare the seeds to the dry, quiescent stage.

A number of functional changes take place during the seed maturation, including the ability to germinate and to tolerate desiccation (BLACKMAN et al., 1991; OBENDORF et al., 1980; BLACKMAN et al., 1992). Our results show that at 36 DAF (stage R6) seeds already have ability to germinate in both crop years (Figure 10). Similar observations were reported by Blackman et al. (1992) at 34 DAF for soybean seeds. The ability to germinate after drying is acquired from stage R7.2, i.e. 63 DAF. Immediately after the acquisition of desiccation tolerance (Figure 10), seed longevity is progressively acquired during maturation (Figure 11 and 12). A similar profile on the acquisition of longevity was also reported for wild mustard (SINNIAH et al., 1998a), clover (HAY et al., 2010), barley (PROBERT et al., 2007). In contrast, the initial seed quality (Ki) which is thought to be a good indicator of seed vigor and potential longevity (ELLIS & ROBERTS, 1980) did not change during maturation.

The profile of the acquisition of longevity was studied in two crop years, 2012/2013 and 2013/2014. The profile was identical in both crop years. However, there was difference in longevity evaluated by P50 and sigma between crop years (Figures 11, 12). This could be explained by the higher temperature (around 7°C higher) from stage R7.1 to R9 in crop year 2013/2014 (Figure 3). Indeed, the acquisition of seed longevity is sensitive to environmental effects during seed formation (FILHO & ELLIS, 1991; SINNIAH et al. 1998a; KOCHANNEK et al., 2011, RIGHETTI et al., 2015). It could be argued that a warm environment brings forward to an earlier stage the time at which longevity is acquired. However, this argument can not explain our data for two reasons. Here, warm temperatures



**Figure 22.** Relative gene expression for RNA-Seq data validation at respective seed developmental stages for genes encoding (A) Heat shock transcription factor A3 (HSFA3: Glyma03G191100); (B) Heat shock transcription factor A6B (HSFA6: Glyma03G157300); (C) bZIP transcription factor family protein (bZIP65: Glyma03G142400); (D) histone acetyltransferase of the GNAT family 1 (GNC5: Glyma.19G160700); (E) NF-X-like 1 (NFXL1: Glyma.09G173000). Bars and line indicate expression values from RT-qPCR and RNA-Seq experiments, respectively.

induced a reduction in storability throughout maturation. Also, our sampling strategy based on phenological stages should eliminate putative differences in the progression of maturation processes due to temperature that would be otherwise observed when this progression is based on seed age. The reasons why cold temperatures is more favorable for seed longevity remains unknown. Likewise, the reasons why temperature did not influence initial seed quality (Ki) remains to be investigated. This is in contrast with data from the literature suggesting that the parental environment influences Ki (KOCHANNEK et al., 2011).

Regarding the non-reducing soluble sugars, despite the large amount of data available, their implication in seed longevity remains a contentious and enigmatic issue. Earlier findings suggested a correlation between longevity and Suc/RFO ratio (HORBOWICZ & OBENDORF, 1994). More recently, by examining 276 seed species, Walters et al. (2005) did not find a correlation between longevity assessed by P50 and sucrose or RFO levels. In contrast, in developing seeds of *Medicago truncatula*, there was a correlation between acquisition of seed longevity and accumulation of stachyose contents (ROSNOBLET et al., 2007; VERDIER et al., 2013). Here we established a relationship between longevity, and monosaccharide and sucrose contents and sucrose/stachyose and sucrose/RFO ratios (Figure 16). However, care must be taken in the interpretation of these relationships because there is no longevity data point between 0 and 25 DAF, which represents half of the maximal longevity. This makes the data point of P50= 0 an outlier that could skew the correlation coefficient.

The correlation was negative between longevity and sucrose and monosaccharides (Figure 16 A e B). This allows us to assume that the sucrose does not seem to play a protective role. However, it may be involved in the regulation of longevity. It is known that sucrose is involved in the regulation of the development of legume seeds (WEBER et al, 2005; SHEEN, 2010). Another explanation could be that sucrose is mobilized to form stachyose.

We observed a negative correlation between longevity and the sucrose/stachyose ratio in axes, which is consistent with the hypothesis of the presence of oligosaccharides as protective agents to form glassy state (BUTINK & LEPRINCE, 2004). In *Medicago truncatula*, seeds defective in MtSNF4b, a gene that is part of kinase complex regulating energy homeostasis showed impaired accumulation of raffinose family oligosaccharides compared with control seeds and as well as reduced longevity and vigor. Correlations between longevity and accumulation of oligosaccharides during maturation



have been observed in *snf4b* mutants of *M. truncatula* (ROSNOBLET et al. 2007). Rosnoblet et al. (2007) reported that stachyose may be important for the germination, after storage for being a rapidly metabolized sugar, and therefore, it would be an indicator of high longevity. However, high stachyose content do not necessarily provide nutrients to accelerate the emergence of the radicle. Thus, we hypothesize that stachyose content is more related to the acquisition of desiccation tolerance than longevity, because of its accumulation increased strongly when soybean seeds become desiccation tolerant, i.e., between stages R7.1 and R7.2.

These results do not discard a role of oligosaccharides on the acquisition of longevity in soybean seeds. However, only the content of non-reducing sugars can not explain the increase in longevity. More sampling points during maturation are required to obtain a more reliable and conclusive correlation between sugar contents and longevity. All these data suggest that the function of non-reducing sugars on longevity is more complex.

#### **IV.2. Gene profiling during maturation of soybean seed**

Many RNA sequencing studies on seed development have focused on soybean, largely because of its global economic importance (BECKER et al., 2014). In this study, we focused at the late phase of seed maturation, from 57 and 77 DAF, i.e. from stage R7.1 to R9, when seed longevity is acquired. A total of 7 sampling points were sampled during the late maturation phase and maturation drying. Therefore, this work is complementary to previous transcriptome studies, focusing on embryogenesis and seed filling (JONES & VODKIN, 2013; SEVERIN et al., 2010; JONES et al., 2010).

A list of 56,044 transcripts were estimated from the sequence mapping data for all the annotated gene models of soybean reference genome, from which 13,477 were considered differentially expressed during late maturation. A global analysis of the expression from these transcripts was performed by principal component analysis (PCA) (Figure 18). It showed that component 1 explained 80% of the variation. This shows large differences between the early and late stages of seed maturation. There is a large difference from transcripts from stage R7.1 to R7.2. Between these two stages, desiccation tolerance is acquired and longevity started to increase. Due to the lack of an intermediate point between R7.1 and R7.2, it is not possible to discriminate between processes and mechanisms that are



taking place in relation with desiccation tolerance from those associated with the onset of longevity. Our PCA plot show that the stage R7.3 represents a key stage at the molecular level as highlighted by a major shift occurring at the transcriptome level. The reasons for this shift remain to be investigated. Phenotypically, at that stage, seeds have lost their chlorophyll and longevity has already increased significantly. Perhaps, the transcriptome shift reflects a reprogramming to prepare seeds for the dry state and germination by synthesizing mRNA that will be stored and used for germination. In support of this hypothesis, the TF profiles also change dramatically at R7.3. Transcripts associated with protein degradation via the SCF family of modular E3 ubiquitin pathway increased during maturation drying. This observation is in agreement with results from Jones et al. (2010) and Jones & Vodkin (2013) in soybean seeds. These proteins are known to filter the proteome by degrading key regulatory proteins as main targets (DUPLAN & RIVAS, 2014). This suggests that the drying seeds are setting up machinery for post-translational regulation before entering in the dry state and also to rearrange their proteome upon imbibition.

During seed maturation, GO categories that are related to chlorophyll binding proteins and photosystems were over-represented. These transcripts were found to be down-regulated during maturation drying. This is consistent with many reports showing that chlorophyll breakdown and chloroplast degradation is an active and necessary process during seed maturation (TEIXEIRA et al. 2016). Impairments of these processes by high temperature during seed formation (TEIXEIRA et al., 2016), by genetic modification (NAKAJIMA et al. 2012) or chemical treatments impairing photosynthetic activity in developing seeds (ALLORENT et al., 2014) leads to reduced seed quality and/or seed longevity. The reasons why retention of chlorophyll is detrimental for seed quality remains unknown.

Other biological processes that were over-represented during final stages of development are related to chaperones, and other products associated with protein-protein interactions and protein folding (Figure 19). HSP could assist in creating proper-folding conditions during seed desiccation. The mRNAs for these factors or the proteins encoded may be produced later in seed development and then stored in the seed for use during the early stages of imbibition and germination.

Genes encoding LEA proteins were already highly expressed during maturation at stage R7.2 and none was found to further increase significantly during maturation drying. Jones & Vodkin (2013) observed induction of LEA genes but these



researchers compared a seed filling stage corresponding to the maximum fresh weight with dry mature seeds, thus skipping the entire maturation period. Therefore, it is difficult to compare this data set with ours. In contrast, Asakura et al. (2012) found that LEA gene expression increased during seed filling, which is much earlier than observations made by Jones & Vodkin (2013). A similar finding was also found during development of *Medicago truncatula*, where LEA transcripts increased mid-way through seed development, correlating with desiccation tolerance and remained high thereafter (VERDIER et al., 2013). Furthermore, there was a delay between LEA transcripts accumulation and synthesis of LEA proteins, which occurred mostly during the late phase of seed maturation (CHATELAIN et al., 2012). Therefore, it is likely that accumulation of LEA transcripts in our material also occurred at a much earlier than stage R7.1. A proteome study is therefore necessary to identify those LEA proteins whose abundance is correlated with seed maturation and longevity. Several lines of evidence suggest that LEA proteins have a role in seed longevity. In *Arabidopsis*, the strong down-regulation of three seed specific dehydrins resulted in a decrease in survival during storage (HUNDERTMARK et al., 2011), whereas in *M. truncatula*, the four most abundant seed LEA proteins correlated with the increase in longevity during maturation (CHATELAIN et al., 2012). These proteins may protect cellular structures in a variety of ways, such as by stabilizing cell membranes, chelating ions that have been concentrated by water loss, or binding the remaining available water (CUEVAS-VELAZQUEZ et al. 2016).

### **IV.3. Transcription factors and longevity in soybean seeds**

Several are associated with the regulation of seed development and maturation. Studies to characterize regulators of the maturation phase were performed in *Arabidopsis* (KOORNNEEF et al., 1984; MEINKE, 1992; KEITH et al., 1994; MEINKE et al., 1994; WEST et al., 1994). The central roles of TF in the control of embryo and seed development were established initially through investigations using mutations. Loss-of-function mutations in maturation TF genes caused embryo lethality or separation of embryo parts, because mutant embryos are intolerant to desiccation or defective in the accumulation of storage proteins and lipids (BECKER et al. 2014).



In our study we noticed that 1086 TF were differentially expressed during late maturation. We performed a cluster analysis and obtained three patterns of expression profile (Figure 22) each of them characterized by a peak of expression at stage 7.1-7.2, 7.2-7.3 and from stage 8.1. The second profile contains transcription factors with a peak in expression at stage R7.2 and R7.3, implying an important regulatory role of these genes before seed drying. Two families were significantly enriched, AP2/EREBP and WRKY. AP2/EREBP transcription factors play an important role in controlling developmental processes and in hormone, sugar and redox signaling in relation with abiotic stresses (DIETZ et al., 2010). Their over-representation in developing soybean seeds was also reported by Jones & Vodkin (2013). Several homologues were found to be co-regulated with the induction of longevity in the model legume *Medicago truncatula* (VERDIER et al., 2013). A DREB2 from sunflower enhanced seed longevity of tobacco when ectopically over-expressed with a heat shock factor, HaHSFA9 (ALMOGUERA et al., 2009). Several WRKY genes regulate the expression of defense-related genes. For example, Redekar et al. (2015) reported that 8 WRKY transcription factors were up-regulated in the early stages of soybean seed development in *3mlpa* mutant. *3mlpa* is an advanced generation recombinant inbred line (RIL) with three *lpa*-causing mutations, derived from a bi-parental cross of V99-5089 and CX-1834, and it provides a higher reduction in PA (phytic acid) content of soybean seed relative to those with single mutations. In summary, the *lpa* (low phytic acid) causing mutations in *3mlpa* mutant may play a role in initiating defense responses during seed development. Verdier et al. (2013) described a *wrky3* mutant of *Medicago truncatula*, which showed higher longevity than wild type. They reported the presence of the defense genes in the longevity module, points to a role of passive, developmentally regulated biotic defense in the acquisition of longevity during late seed maturation. Here, the expression level of *WRKY3* increased as longevity was acquired in soybean seeds. It would be interesting to investigate whether *WRK3* also plays a role in seed longevity in soybean

“Response to abiotic stress and light” are GO biological process that are enriched at late stages of maturation. Genes present in these GO categories include members of the bHLH, AP2/EREBP and HSF families. Some of these TF could be involved in regulation of genes in the dry seed by anticipating germination, decreasing the time and energy needed to produce essential proteins once imbibition begins (JONES & VODKIN. 2013). bHLH TF have been implicated in numerous biological processes in plants including responses to light, cold, and hormones, epidermal cell fate determination, developmental



patterning in roots and flowers and anthocyanin biosynthesis (LIU et al., 2008; PAYNE et al., 2000), development and dehiscence of the seed and seed pod (HEISLER et al. , 2001; GROSZMANN et al. 2011) and responses to light quality and photoperiod (CASSON et al., 2009; HUQ & QUAIL, 2002). PHYTOCHROME-INTERACTING FACTORS (PIFs) are a small subfamily of bHLH TFs that directly interact with the photoactivated phytochromes under specific light conditions (CASTILLON et al., 2007). In our study, two members of this subfamily were positively correlated with longevity, namely PIF1 and PIF7. PIF1 negatively regulates chlorophyll biosynthesis (MOON et al., 2008; OH et al., 2004), and its expression increases as soybean seeds lose their chlorophyll during maturation phase. PIF7 acts similarly to PIF3 in prolonged red light as a weak negative regulator of phyB-mediated seedling deetiolation (LEIVAR & MONT, 2014).

Heat shock transcription factors HSFA6B, HSFA3, HSFC1 e HSFAD1 have a high expression at the late maturation phase (cluster JK, table 7) and were positively correlated with longevity, suggesting an important role in regulating this process. Using network analysis in developing *Medicago truncatula* seeds, Verdier et al., (2013) also found a MtHSF (putative ortholog of the Arabidopsis HSFA9) that was placed at the interface between DT and longevity modules. HSF has been found to regulate seed longevity. Overexpression of HaHSFA9 in mutants led to an increased stability against accelerated ageing of tobacco seeds on longevity beyond thermal tolerance of seedlings (PRIETO-DAPENA et al., 2006). This TF also interacts with the drought-responsive factor HaDREB2 in a seed-specific manner to enhance stability against accelerated aging (ALMOGUERA et al., 2009).

Another interesting TF found in our study, among TFs positively correlated with longevity, it was NFXL1, a NF-X1-type zinc finger protein. It has been implicated in the trichothecene phytotoxin-induced response as well as in the general defense response (MÜSSIG et al., 2010). In addition, plants overexpressing NFXL1 display a higher survival when subjected to abiotic stresses such as salt, drought, or high light intensity (LISSO et al., 2006). Additionally, characterization of *nfxl1* TF mutant demonstrated that this gene regulates some of the network nodes and exhibits impaired acquisition of longevity during maturation (RIGHETTI et al. 2015).

Analysis of putative gene regulatory is an excellent way to identify possible regulators of seed development. However, experimental validation and functional characterization of the TFs are required to validate them. At this study we suggest a



characterization study on mutants from HSF, WRKY3, PIF1 and NFXL1 as putative candidates to contribute to longevity in soybean seeds.



## Chapter V. CONCLUSIONS

3. Soybean seeds acquire longevity progressively during maturation after seed filling and mass maturity;
4. The accumulation of non-reducing soluble sugars content are related to acquisition of longevity, however their function on longevity is more complex;
3. Throughout the late seed maturation, there are complex changes in the transcriptome that are likely associated with the acquisition of seed longevity and the preparation of the quiescent state. This suggests that physiological maturity is acquired much later than originally described in the literature;
4. HSF and several TF associated with biotic defense (WRKY3 and NLFX1) are candidate genes whose putative role in seed longevity deserve further characterization;
5. Biological processes enriched for response to abiotic stress and response to light are related to acquisition of longevity.



## REFERENCES BIBLIOGRAPHIQUES

ALMOGUERA C, JORDANO J. 1992. Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular weight heatshock protein and Lea mRNAs. *Plant Mol Biol* 19: 781-792.

ALMOGUERA C, PIETRO-DAPENA, P. DÍAZ-MARTIN, J. ESPINOSA, J.M. CARRANCO, R. JORDANO, J. 2009. The HaDREB2 transcription factor enhances basal thermotolerance and longevity of seeds through functional interaction with HaHSFA9. *BMC Plant Biology*, 9, 12p.

ALLORENT, G. OSORIO S, VU JL, FALCONET D, JOUHET J, KUNTZ M, FERNIE AR, LERBS-MACHE S, MACHEREL D, COURTOIS F, FINAZZI G. 2014. Adjustments of embryonic photosynthetic activity modulate seed fitness in *Arabidopsis thaliana*. *New phytologist*, 205, 707–719.

ALPERT, P.; OLIVER, M. J. 2002. Drying without dying. In: Desiccation and survival in plants: Drying without dying, Black M, Pritchard HW, eds. Wallingford, UK, **CABI Publishing**, p. 3-43.

ANGELOVICI, R, GALILI G, FERNIE A. R, FAIT A. 2010. Seed desiccation: A bridge between maturation and germination. *Trends in Plant Science*, 15, p. 211-218.

ASAKURA T, TAMURA T, TERAUCHI K, NARIKAWA T, YAGASAKI K, ISHIMARU Y, ABE K. 2012. Global gene expression profiles in developing soybean seeds. *Plant Physiology and Biochemistry*, 52, p. 147–53.

BAILLY C. 2004 Active oxygen species and antioxidants in seed biology. *Seed Science Research*, 14, p. 93-107.

BELTRATI CM, PAOLI AAS. 2003 Sementes. Pp. 399-424. In: Appezatoda-Glória, B. & Carmello-Guerreiro, S.M. *Anatomia vegetal*. Viçosa, UFV.

BENTSINK L, ALONSO-BLANCO C, VREUGDENHIL D, TESNIER K, GROOT SP, KOORNNEEF M. 2000. Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiology*, 124 p. 1595–1604.

BECKER MG, HSU SW, HARADA JJ, BELMONTE MF. 2014. Genomic dissection of the seed. *Frontiers in Plant Science*, 5, p. 1 – 16.

BERJAK P, PAMMENTER NW. 2000. What ultrastructure has told us about recalcitrant seeds. *Revista Brasileira de Fisiologia Vegetal*, 12, p. 22–55,

BERJAK P, PROCTOR MCF, OLIVER MJ, WOOD AJ, ALPERT P, STARK LR, CLEAVITT N, MISHLER BD. 2007. Seed desiccation tolerance mechanisms. In: Jenks MA and Wood AJ eds. *Plant Desiccation Tolerance*. Blackwell publishing, Iowa USA, 151-192,



- BETTEY M, FINCH-SAVAGE WE. 1998. Stress protein content of mature Brassica seeds and their germination performance. **Seed Science Research**, 8, p. 347–355.
- BEWLEY JD, BRADFORD, K., HILHORST, H., NONOGAKI, H. 2013. **Seeds: physiology of development, germination and dormancy**. 3 ed. New York: Springer, 392 p.
- BLACKMAN SA, WETTLAUFER SH, OBENDORF RL, LEOPOLD AC. 1991. Maturation proteins associated with desiccation tolerance in soybean. **Plant Physiology**, 96, p. 868-874.
- BLACKMAN SA, OBENDORF RL, LEOPOLD AC. 1992. Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. **Plant Physiology**, 100, p. 225-230.
- BLOKHINA O, VIROLAINEN E, FAGERSTEDT KV. 2003. Antioxidants, oxidative damage and oxygen deprivation stress: a review. **Annals of Botany**, 91 Spec No, p. 179-194.
- BOLSTAD BM. 2015. Preprocesscore: A Collection of Pre-Processing Functions. R package version 1.30.0.
- BUITINK J, WALTERS C, HOEKSTRA FA, CRANE J. 1998. Storage behavior of *Typha latifolia* pollen at low water contents: Interpretation on the basis of water activity and glass concepts. **Physiologia Plantarum**, 103, p. 145-153.
- BUITINK J, HEMMINGA MA, HOEKSTRA FA. 2000. Is there a role for oligosaccharides in seed longevity? An assessment of intracellular glass stability. **Plant Physiology**, 122, p. 1217-1224.
- BUITINK J, LEPRINCE O. 2004. Glass formation in plant anhydrobiotes: survival in the dry state. **Cryobiology**, 48, p. 215-228.
- BUITINK J, LEPRINCE O. 2008. Intracellular glasses and seed survival in the dry state. **Comptes Rendus Biologies**, 331, p. 788– 795.
- BURKE MJ. 1986. The glassy state and survival of anhydrous biological systems. In: LEOPOLD AC. (Ed.). **Membranes, metabolism and dry organisms**. Ithaca: Cornell University Press, p. 358-363.
- CARPENTER JF, MARTIN B, CROWE LM, CROWE JH. 1987. Stabilization of phosphofructokinase during air-drying with sugars and sugar/transition metal mixtures. **Cryobiology**, 24, p. 455-464.
- CARPENTER JF, CROWE JH, ARAKAWA T. 1990. Comparison of solute-induced protein stabilization in aqueous solution and in the frozen and dried states. **Journal of Dairy Science**, 73, p. 3627-3636.



- CASSON SA, FRANKLIN KA, GRAY JE, GRIERSON CS, WHITELAM GC, HETHERINGTON AM. 2009. Phytochrome B and PIF4 regulate stomatal development in response to light quantity. **Current Biology**, 19, p.229–234.
- CASTILLON A, SHEN H, HUQ E. 2007. Phytochrome interacting factors: central players in phytochrome-mediated light signaling networks. **Trends Plant Science**, 12, p. 514–521.
- CASTRO RD, BRADFORD KJ, HILHORST HWM. 2004. Desenvolvimento de sementes e conteúdo de água. In: FERREIRA, A. G.; BORGHETTI, F. **Germinação: do básico ao aplicado**. Porto Alegre, editora Artmed.
- CHATELAIN E, HUNDERTMARK M, LEPRINCE O, LE GALL S, SATOUR P, DELIGNY-PENNINCK S, ROGNIAUX H, BUITINK J. 2012. Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. **Plant Cell Environment**, 35, p. 1440–1455.
- CLERKX EJM, VRIES HB, RUYS GJ, GROOT SPC, KOORNNEEF M. 2004. Genetic differences in seed longevity of various *Arabidopsis* mutants. **Plant Physiology**, 121, p. 448–461.
- CONAB. Companhia Nacional de Abastecimento. 2016. **Acompanhamento da safra brasileira grãos**. v. 03. Safra 2015/16, n. 04 – Quarto levantamento, jan.
- CRAIG EA, GAMBILL BD, NELSON RJ. 1993. Heat shock proteins: Molecular chaperones of protein biogenesis. **Microbiology Molecular Biology Reviews**, 57, p. 402-414.
- CROWE JH, HOEKSTRA FA, NGUYEN KH, CROWE LM. 1996. Is vitrification involved in depression of the phase transition temperature in dry phospholipids? **Biochimica et Biophysica Acta (BBA) – Biomembranes**, 1280, p. 187-196.
- CROWE JH, OLIVER AE, HOEKSTRA FA, CROWE LM. 1997. Stabilization of dry membranes by mixtures of hydroxyethyl starch and glucose: The role of vitrification. **Cryobiology**, 35, p. 20-30.
- CUEVAS-VELAZQUEZ CL, SAAB-RINCÓN G, REYES JL, COVARRUBIAS AA. 2016. The unstructured N-terminal region of *Arabidopsis* group 4 Late Embryogenesis Abundant Proteins (LEA) is required for folding and for chaperone-like activity under water deficit. **The Journal of Biological Chemistry**, 291, 22 p.
- DEARMAN J, BROCKLEHURST PA, DREW RLK. 1986. Effects of osmotic priming and ageing on onion seed germination. **Annals of Applied Biology**, 108, p. 639–648.
- DALLE-DONNE I, ALDINI G, CARINI M, COLOMBO R, ROSSI R, MILZANI A. 2006. Protein carbonylation, cellular dysfunction, and disease progression. **Journal of Cellular and Molecular Medicine**, 10, p.389-406.



- DARGAHI H, TANYA P, SRINIVES P. 2014. Mapping of the genomic regions controlling seed storability in soybean (*Glycine max* L.). **Journal of Genetics**, 93, n. 2, p. 365-370.
- DE SOUZA VIDIGAL D, WILLEMS L, VAN ARKEL J, DEKKERS BJ, HILHORST HW, BENTSINK L. 2016. Galactinol as marker for seed longevity. **Plant Science**, 246, p. 112–118.
- DEMIRKAYA M. 2013. Relationships between antioxidant enzymes and physiological variations occur during ageing of pepper seeds. **Horticulture, Environment and Biotechnology**, 54, 97–102
- DIETZ KJ, VOGEL MO, VIEHHAUSER A. 2010. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. **Protoplasma**, 245, p. 3–14.
- DU Z, ZHOU X, LING Y, ZHANG Z, SU Z. 2010. AgriGO: a GO analysis toolkit for the agricultural community. **Nucleic Acids Research**, 38, p. 64–70.
- DUEK PD, FANKHAUSER C. 2005. bHLH class transcription factors take centre stage in phytochrome signaling. **Trends in Plant Science**, v. 10, p. 51-64.
- DUPLAN V, RIVAS, S. 2014. E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. **Frontier in Plant Science**, 42, 6p.
- ELLIS RH, HONG TD, JACKSON MT. 1993. Seed production environment, time of harvest, and the potential longevity of seeds of three cultivars of rice (*Oryza sativa* L.). **Annals of Botany**, 72, p. 583-590.
- ELLIS RH, ROBERTS EH. 1980. Improved equations for the prediction of seed longevity. **Annals of Botany**, 45, p. 13- 30.
- EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. 2004. **Tecnologias de produção de soja região central do Brasil**. Londrina, Embrapa – Soja. 237p. (Sistemas de Produção 04).
- EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. 2006. **Tecnologias de produção de soja região central do Brasil**. Londrina, Embrapa – Soja. 225p. (Sistemas de Produção 11)
- FAO. Food and Agriculture Organization of the United Nations. **FAOSTAT** – (2010).
- FEHR WR, CAVINESS CE. 1977. **Stages of soybean development**. Ames: Iowa State University of Science and Technology, 11p (Special Report, 80)
- FILHO CP, ELLIS RH. 1991. The development of seed quality in spring barley in four environments. I. Germination and longevity. **Seed Science Research**, 1, p. 163-177.



GENTLEMAN RC, CAREY VJ, BATES DM, BOLSTAD B, DETTLING M, DUDOIT S, ELLIS B, GAUTIER L, GE Y, GENTRY J, HORNIK K, HOTHORN T, HUBER W, IACUS S, IRIZARRY R, LEISCH F, LI C, MAECHLER M, ROSSINI AJ, SAWITZKI G, SMITH C, SMYTH G, TIERNEY L, YANG JY, ZHANG J. 2004. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, 5, R80.

GOODSTEIN DM, SHU S, HOWSON R, NEUPANE R, HAYES RD, FAZO J, MITROS T, DIRKS W, HELLSTEN U, PUTNAM N, ROKHSAR DS. 2012. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, p. 1178 – 1186.

GRENE R. 2002. Oxidative stress and acclimation mechanisms in plants. **Arabidopsis Book**, 1, p: 1-20.

GROSZMANN M, PAICU T, ALVAREZ JP, SWAIN SM, SMYTH DR. 2011. SPATULA and ALCATRAZ, are partially redundant, functionally diverging bHLH genes required for Arabidopsis gynoecium and fruit development. **Plant Journal**, 68, p. 816–829.

GRUWEZ R, LEROUX O, DE FRENNE P, TACK W, VIANE R, VERHEYEN K. 2013. Critical phases in the seed development of common juniper (*Juniperus communis*). **Plant Biology**, 15, p. 210–219.

HAY FR, PROBERT RJ. 1995. Seed maturity and the effects of different drying conditions on desiccation tolerance and seed longevity in foxglove (*Digitalis purpurea* L.). **Annals of Botany**, 76, p. 639–647.

HAY FR, SMITH RD, ELLIS RH, BUTLER LH. 2010. Developmental changes in the germinability, desiccation tolerance, hardseededness, and longevity of individual seeds of trifolium ambiguum. **Annals of Botany**, 105, p. 1035-1052.

HEISLER MGB, ATKINSON A, BYLSTRA YH, WALSH R, SMYTH DR. 2001. SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. **Development**, 128, p.1089–1098,

HENDRICK JP, HARTL F. 1993. Molecular chaperone functions of heat-shock proteins. **Annual Review of Biochemistry**, 62, p. 349-384.

HOEKSTRA FA, WALKERS WF, BUITINK B, GOLOVINA EA, CROWE JH, CROWE LM. 1997. Membrane stabilization in the dry state. **Comparative Biochemistry and Physiology Part A: Physiology**, 117, p. 335-341.

HOEKSTRA FA, GOLOVINA EA, BUITINK B. 2001. Mechanisms of plant desiccation tolerance. **Trends in Plant Science**, 6, p. 431– 438.

HORBOWICZ M, OBENDORF RL. 1994. Seed desiccation tolerance and storability: Dependence on flatulence-producing oligosaccharides and cyclitols? Review and survey. **Seed Science Research**, 4, p. 385-405.



HUQ E, QUAIL P H. 2002. PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in Arabidopsis, **EMBO Journal**, 21, p. 2441–2450.

HYMOWITZ T. 1970. On the domestication of the Soybean. **Economy Botany**, 24, 4, p. 408 – 421.

JONES SI, GONZALEZ DO, VODKIN LO. 2010. Flux of transcript patterns during soybean seed development. **BMC Genomics**, 11, pp. 136-149.

JONES SI, VODKIN LO. 2013. Using RNA-Seq to Profile Soybean Seed Development from Fertilization to Maturity. **PLOS ONE**, 8, p. 1-12.

JOSHI CP, NGUYEN HT. 1996. Differential display-mediated rapid identification of different members of a multigene family, HSP16.9 in wheat. **Plant Molecular Biology**, 31, p. 575-584.

KEITH K, KRAML M, DENGLER NG, MCCOURT P. 1994. Fusca3: a heterochronic mutation affecting late embryo development in Arabidopsis. **Plant Cell**, 6, p. 589-600.

KIBINZA S, VINEL D, COME D, BAILLY C, CORBINEAU F. 2006. Sunflower seed deterioration as related to moisture content during ageing, energy metabolism and active oxygen species scavenging. **Physiologia Plantarum**, 128, 496–506.

KOCHANEK J, STEADMAN KJ, PROBERT RJ, ADKINS SW. 2011. Parental effects modulate seed longevity: exploring parental and offspring phenotypes to elucidate pre-zygotic environmental influences. **New Phytologist**, 191, P. 223–233.

KOORNNEEF M, REULING G, KARSSSEN CM. 1984. The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. **Physiology of Plant.**, 61, p. 377- 383.

KOSTER KL. 1991. Glass formation and desiccation tolerance in seeds. **Plant Physiology**, 96, p. 302–304.

KRANNER I, BIRTIC S. 2005. A modulating role for antioxidants in desiccation tolerance. **Integrative and Comparative Biology**, 45, p. 734–740.

KUMAR SJ, PRASAD SR, BANERJEE R, THAMMINENI C. 2015. Seed birth to death: dual functions of reactive oxygen species in seed physiology. **Annals of Botany**, 116, p. 663–668.

KUO TM, LOWELL C, SMITH PT. 1997. Changes in soluble carbohydrates and enzymatic activities in maturing soybean seed tissues. **Plant Science**, 125, p. 1-11,

LANDJEVA S, LOHWASSER U, BÖRNER A. 2010. Genetic mapping within the wheat D genome reveals QTL for germination, seed vigour and longevity, and early seedling growth. **Euphytica**, 171, p. 129-143.



- LANGMEAD B, TRAPNELL C, POP M, SALZBERG SL. 2009. Ultrafast and memory efficient alignment of short DNA sequences to the human genome. **Genome Biology**, 10, p. 25-32.
- LEE Y Y, PARK HM, HWANG TY, KIM SL, KIM MJ, LEE SK, SEO MJ, KIM KJ, KWON YU, LEE SC, KIM YH. 2015. A correlation between tocopherol content and antioxidant activity in seeds and germinating seeds of soybean cultivars. **Journal of the Science of Food and Agriculture**, 95, p. 819–827.
- LEIVAR P, MONTE E. 2014. PIFs: systems integrators in plant development. **The Plant Cell**, 26, n. 56–78.
- LEHNER A, MAMADOU N, POELS P, CÔME D, BAILLY C, CORBINEAU F. 2008. Changes in soluble carbohydrates, lipid peroxidation and antioxidant enzyme activities in the embryo during ageing in wheat grains. **J. Cereal Sci.** 47, p. 555-565.
- LEPRINCE O, BUITINK J. 2010. Desiccation tolerance: From genomics to the field. **Plant Science**, 179, p. 554–564.
- LEPRINCE O, HENDRY GAF, MCKERSIE BD. 1993. The mechanisms of desiccation tolerance in developing seeds. **Seed Science Research**, 3, p. 231-246.
- LISSO J. 2006. The AtNFXL1 gene encodes a NF-X1 type zinc finger protein required for growth under salt stress. **FEBS Lett.** 580, p. 4851–4856.
- LIU X, YU X, LI K, KLEJNOT J, YANG H, LISIERO D, LIN C. 2008. Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. **Science**, 322, p. 1535–1539.
- LOVE MI, HUBER W, ANDERS S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, 15, pp. 550.
- MEINKE DW. 1992. A homoeotic mutant of *Arabidopsis thaliana* with leafy cotyledons. **Science**, 258, p. 1647-1650.
- MEINKE DW, FRANZMANN LH, NICKLE TC, YEUNG EC. 1994. *Leafy Cotyledon* mutants of Arabidopsis. **Plant Cell**, 6, p. 1049-1064.
- MIRA S, GONZÁLEZ-BENITO ME, HILL LM, WALTERS C. 2010. Characterization of volatile production during storage of lettuce (*Lactuca sativa*) seed. **Journal of Experimental Botany**, 61, p. 3915–3924.
- MOLLER IM, JENSEN PE, HANSSON A. 2007. Oxidative modifications to cellular components in plants. **Annual Review of Plant Biology**, 58, p. 459-481.
- MOON J, ZHU L, SHEN H, HUQ E. 2008. PIF1 directly and indirectly regulates chlorophyll biosynthesis to optimize the greening process in Arabidopsis. **Proceedings of the National Academy of Sciences**, 105, p. 9433–9438.



MÜSSIG C, SCHRÖDER F, USADEL B, LISSO J. 2010. Structure and putative function of NFX1-like proteins in plants. **Plant Biology**, 12, p 381-94.

NAKAJIMA S, ITO H, TANAKA R, TANAKA A. 2012. Chlorophyll b Reductase Plays an Essential Role in Maturation and Storability of Arabidopsis Seeds. **Plant Physiology**, 160, p. 261–273.

NAKOMOTO H, VIGH L. 2007. The small heat shock proteins and their clients. **Cellular and Molecular Life Sciences**, 64, p. 294-306.

NIEDZIELSKI M, WALTERS C, LUCZAK W, HILL LM, WHEELER LJ, PUCHALSKI J. 2009. Assessment of variation of seed longevity within rye, wheat and the intergeneric hybrid triticale. **Seed Science and Research**, 19, p. 213-224.

NKANG A, UMHO EO. 1996. Six month storability of five soybean cultivars as influenced by stage of harvest, storage temperature and relative humidity. **Seed Science and Technology**, 25, p. 93–99.

NOCTER G, FOYER CH. 1998. Ascorbate and glutathione: keeping active oxygen under control. **Annual Review of Plant Physiology and Plant Molecular Biology**, 49, p. 249-279.

OBENDORF RL, ASHWORTH EN, RYTKO GT. 1980. Influence of seed maturation on germinability in soybean. **Crop Science**, 20, p. 483-486.

OH E, KIM J, PARK E, KIM JI, KANG C, CHOI G. 2004. PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in *Arabidopsis thaliana*. **Plant Cell**, 16, p. 3045–3058.

OOMS J, LEON-KLOOSTERZIEL KM, BARTELS D, KOORNNEEF M, KARSSSEN CM. 1993. Acquisition of desiccation tolerance and longevity in seeds of *Arabidopsis thaliana* (a comparative study using abscisic acid-insensitive *abi3* mutants). **Plant Physiology**, 102, p. 1185–1191.

PAYNE CT, ZHANG F, LLOYD AM. 2000. GL3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. **Genetics**, 156, p. 1349–1362.

PETERBAUER T, LAHUTA LB, BLÖCHL A, MUCHA J, JONES DA, HEDLEY CL, GÖRECKI RJ, RICHTER A. 2001. Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. **Plant Physiology**, 127, p. 1764-1772.

PRIETO-DAPENA P, CASTAÑO R, ALMOGUERA C, JORDANO J. 2006. Improved resistance to controlled deterioration in transgenic seeds. **Plant physiology**, 142, p. 1102.

PROBERT R, ADAMS J, CONEYBEER J, CRAWFORD A, HAY F. 2007. Seed quality for conservation is critically affected by pre-storage factors. **Australian Journal of Botany**, 55, p. 326-335.



RAIJ B, CANTARELLA H, QUAGGIO JA, FURLANI AMC. 1997. **Recomendações de Adubação e Calagem para o Estado de São Paulo**. Campinas: IAC, 285p. (Boletim técnico, 100).

RAJJOU L, LOVIGNY Y, GROOT SP, BELGHAZI M, JOB C, JOB D. 2008. Proteome-wide characterization of seed aging in Arabidopsis: a comparison between artificial and natural aging protocols. **Plant Physiology**, 148, p. 620-641.

REDEKAR NR, BIYASHEV RM, JENSEN RV, HELM RF, GRABAU EA, MAROOF MAS. 2015. Genome-wide transcriptome analyses of developing seeds from low and normal phytic acid soybean lines. **BMC Genomics**, 16, 1074

RIGHETTI K, VU JL, PELLETIER S, VU BL, GLAAB E, LALANNE D, PASHA A, PATEL RV, PROVART NJ, VERDIER J, LEPRINCE O, BUITINK J. 2015. Inference of longevity-related genes from a robust co-expression network of seed maturation identifies new regulators linking seed storability to biotic defense-related pathways. **Plant Cell**, 27, p. 2692–2708.

ROSNOBLET C, AUBRY C, LEPRINCE O, VU BL, ROGNIAUX H, BUITINK J. 2007. The regulatory gamma subunit snf4b of the sucrose non-fermenting-related kinase complex is involved in longevity and stachyose accumulation during maturation of *Medicago truncatula* seeds. **Plant Journal**, 51, p. 47-59.

RUIJTER JM. 2009. LinRegPCR (11.0) Analysis of quantitative RT-PCR data (Manual). 29 p.

SALLON S, SOLOWEY E, COHEN Y, KORCHINSKY R, EGLI M, WOODHATCH I, SIMCHONI O, KISLEV M. 2008. Germination, genetics, and growth of an ancient date seed. **Science**, 320, pp. 1464.

SANO N, RAJJOU L, NORTH HM, DEBEAUJON I, MARION-POLL A, SEO M. 2015. Staying Alive: Molecular Aspects of Seed Longevity. **Plant and Cell Physiology**, 0, p. 1–15.

SANHEWE AJ, ELLIS RH. 1996. Seed development and maturation in *Phaseolus vulgaris*. 2. Post-harvest longevity in air-dry storage. **Journal of Experimental Botany**, 47, p. 959–965.

SATTLER SE, GILLILAND LU, MAGALLANES-LUNDBACK M, POLLARD M, DELLAPENNA D. 2004. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. **Plant Cell**, 16, p. 1419-1432.

SCHMUTZ J, CANNON SB, SCHLUETER J, MA J, MITROS T, NELSON W, HYTEN DL, SONG Q, THELEN JJ, CHENG J, XU D, HELLSTEN U, MAY GD, YU Y, SAKURAI T, UMEZAWA T, BHATTACHARYA MK, SANDHU D, VALLIYODAN B, LINDQUIST E, PETO M, GRANT D, SHU S, GOODSTEIN D, BARRY K, FUTRELL-GRIGGS M, ABERNATHY B, DU J, TIAN Z, ZHU C, GILL N, JOSHI T, LIBAULT M, SETHURAMAN A, ZHANG X, SHINOZAKI K, NGUYEN HT, WING RA, CREGAN P, SPECHT J, GRIMWOOD J, ROKHSAR D, STACEY G,



- SHOEMAKER RC, JACKSON SA. 2010. Genome sequence of the palaeopolyploid soybean. **Nature**, 463, p. 178–183.
- SEAL CE, ZAMMIT R, SCOTT P, FLOWERS TJ, KRANNER I. 2010. Glutathione half-cell reduction potential and  $\alpha$ -tocopherol as viability markers during the prolonged storage of *Suaeda maritima* seeds. **Seed Science Research** 20, p. 47–53.
- SEDIYAMA T, TEIXEIRA RC, REIS MS. 2005. Melhoramento da soja. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: UFV, p. 553 – 604.
- SEDIYAMA T, TEIXEIRA RC, BARROS HB. 2009. Origem, evolução e importância econômica. In: SEDIYAMA, T. (Ed). **Tecnologias de produção e usos da soja**. Londrina: Mecenaz, p. 1–15.
- SEVERIN AJ, WOODY JL, BOLON YT, JOSEPH B, DIERS BW, FARMER AD, MUEHLBAUER GJ, NELSON RT, GRANT D, SPECHT JE, GRAHAM MA, CANNON SB, MAY GD, VANCE CP, SHOEMAKER RC. 2010. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. **BMC Plant Biology**. 10, p. 1-16.
- SHEN-MILLER J. 2002. Sacred lotus, the long-living fruits of China Antique. **Seed Science and Research**, 12, p. 131-143.
- SHEEN J. 2010. Discover and connect cellular signaling. **Plant Physiology**, 154, p. 562-566.
- SINNIAH UR, ELLIS RH, JOHN P. 1998a. Irrigation and seed quality development in rapid-cycling brassica: Seed germination and longevity. **Annals of Botany**, 82, p. 309-314.
- SINNIAH UR, ELLIS RH, JOHN P. 1998b. Irrigation and seed quality development in rapid-cycling Brassica: soluble carbohydrates and heat-stable proteins. **Annals of Botany**, 82, p. 647–655.
- STEADMAN K, PRITCHARD HW, DEY PM. 1996. Tissue-specific Soluble Sugars in Seeds as Indicators of Storage Category. **Annals of Botany**, v. 77, p. 667-674.
- SMÝKAL P, MASÍN J, HRDÝ I, KONOPÁSEK I, ZÁRSKÝ V. 2000. Chaperone activity of tobacco HSP18, a small heat-shock protein, is inhibited by ATP. **The Plant Journal**, 23, p. 703–713.
- TANG SD, TEKRONY DM, EGLI DB, RUCKER M. 1999. Survival characteristics of corn seed during storage: I. Normal distribution of seed survival. **Crop Science**, 39, p. 1394-1400.
- TEIXEIRA RN, LIGTERINK W, FRANÇA-NETO JDEB, HILHORST HW, DA SILVA EA. 2016. Gene expression profiling of the green seed problem in Soybean. **BMC Plant Biology**, 1, p. 16 – 37.
- THIMM O, BLÄSING O, GIBON Y, NAGEL A, MEYER S, KRÜGER P, SELBIG J, MÜLLER LA, RHEE SY, STITT M. 2004. MAPMAN: a user-driven tool to display



genomics data sets onto diagrams of metabolic pathways and other biological processes. **Plant Journal**, 37, p. 914 – 939.

TOMPSON JA, TAN J, GREENE CS. 2016. Cross-platform normalization of microarray and RNA-seq data for machine learning applications. **PPerrJ**, 4, p. 1-19.

TUNNAcliffe A, WISE MJ. 2007. The continuing conundrum of the LEA proteins. **Naturwissenschaften**, 94, p. 791-812.

USADEL B, NAGEL A, STEINHAUSER D, GIBON Y, BLÄSING OE, REDESTIG H, SREENIVASULU N, KRALL L, HANNAH MA, POREE F, FERNIE AR, STITT M. et al. 2006. PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. **BMC Bioinformatics**, 7, 8 p.

USDA. **United States Department of Agriculture**. 2016. Foreign Agricultural Service Circular Series WAP 1-16, January.

VANDECASTEELE C, TEULAT-MERAH B, MORERE-LE PAVEN MC, LEPRINCE O, LY VU B, VIAU L, LEDROIT L, PELLETIER S, PAYET N, SATOUR P, LEBRAS C, GALLARDO K, HUGUET T, LIMAMI AM, PROSPERI JM, BUITINK J. 2011. QTL analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigor in *Medicago truncatula*. **Plant, Cell and Environment**, 34, p. 1473-1487.

VERDIER J, LALANNE D, PELLETIER S, TORRES-JEREZ I, RIGHETTI K, BANDYOPADHYAY K, LEPRINCE O, CHATELAIN E, VU BL, GOUZY J, GAMAS P, UDVARDI MK, BUITINK J.. 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 Physiology**, 163, p. 757–774.

VIERLING E. 1991. The roles of heat shock proteins in plants. **Annual Review of Plant Physiology and Plant Molecular Biology**, 42, p. 579–620.

WALTERS C. 1998. Understanding the mechanisms and kinetics of seed aging. **Seed Science and Research**, 8, p. 223-244.

WALTERS C, WHEELER LM, GROTENHUIS JM. 2005. Longevity of seeds stored in a genebank: Species characteristics. **Seed Science Research**, 15, p 1-20.

WEBER H, BORISJUK L, WOBUS U. 2005. Molecular physiology of legume seed development. **Annual Review of Plant Biology**, 56, p. 253-279,

WECHSBERG GE, PROBERT RJ, BRAY CM. 1994. The relationship between 'dehydrin-like' proteins and seed longevity in *Ranunculus sceleratus* L. **Journal of Experimental Botany**, 45, p. 1027-1030.

WEHMEYER N, HERNANDEZ LD, FINKELSTEIN RR , VIERLING E. 1996. Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. **Plant Physiology**, 112, p. 747-757.

WEST, M, YEE KM, DANA O J, ZIMMERMAN JL, FISCHER RL, GOLDBERG RB, HARADA JJ. 1994. *LEAFY COTYLEDON1* is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. **Plant Cell**, 6, p. 1731-1745.

WISE MJ, TUNNACLIFFE A. 2004. POPP the question: what do LEA proteins do? **Trends in Plant Science**, 9, p. 13-17.

WU XL, LIU H, WANG W, CHEN S, HU X, LI C. 2011. Proteomic analysis of seed viability in maize. **Acta Physiology Plant**, 33, p. 181-191.

XIN X, TIAN Q, YIN G, CHEN X, ZHANG J, NG S, LU X. 2013. Reduced mitochondrial and ascorbate-glutathione activity after artificial ageing in soybean seed. **Journal of Plant Physiology**, 171, 140-147.

## RÉSUMÉ

Étude physiologique et moléculaire de l'acquisition de la longévité des graines de soja (*Glycine max* L. Merrill) pendant leur développement.

La production de semences très vigoureuses est un levier clé pour augmenter les rendements agricoles. Chez le soja, les risques de production de semences de faible vigueur germinative sont élevés sous les climats tropicaux, ce qui conduit à une mauvaise installation de la culture, voire un déclassement des lots en grain. La longévité (c-à-d la capacité de survivre à l'état sec pendant de longues périodes) représente une caractéristique importante de la vigueur germinative. Cependant, la façon dont elle est acquise au cours de la maturation de la graine n'est pas connue. L'objectif de la thèse était d'obtenir un aperçu des mécanismes de régulation de l'acquisition de la longévité chez le soja. Celle-ci est acquise peu après la phase de remplissage et l'acquisition de la tolérance à la dessiccation (stade R7.2) et continue d'augmenter jusqu'en fin de maturation. Des différences de longévité ont été observées entre les récoltes 2013 et 2014. Elles seraient liées aux températures élevées au cours de la formation des graines, un facteur connu pour impacter la qualité des semences. Le séquençage Illumina de graines récoltées à 7 stades de développement a révélé des changements complexes se produisant au niveau transcriptionnel dès le stade R7.2, pourtant considéré comme correspondant à la maturité physiologique. Les analyses d'enrichissement GO ont révélé une association positive entre longévité et les processus de maturation des ARN et régulateurs de la post-transcription et une association négative avec les processus liés à la photosynthèse. L'acquisition de la longévité est aussi associée à une augmentation des transcrits codant pour des HSF et HSP, ainsi que des facteurs de transcription liés à la défense biotique. Notre travail met en évidence des marqueurs putatifs qui pourraient être utilisés pour évaluer la longévité en production de semences de soja.

## ABSTRACT

*Physiological and molecular studies during the acquisition of longevity in soybean (*Glycine max* L. Merrill) seeds*

*Soybean is one of the most important oil crop species used for food and feed as well as a range of industrial applications. However, producing highly vigorous seeds is a key lever to increase crop production efficiency. Low physiological seed quality, which is more prone to occur under tropical environment, leads to poor stand establishment and decreased in yields. Seed longevity is the ability to survive the dry state for prolonged periods of time and represents an important trait for seed quality. Here, the objective was to obtain insights into the mechanisms regulating the progressive acquisition of longevity. Using Illumina high-throughput sequencing, RNA was sequenced from seven different stages during the acquisition of longevity, generating between 14 and 38 million of reads. These reads were aligned to the *Glycine max* Wm82.a2.v1 gene model. Differentially expressed transcripts (DET) were correlated with the increase in seed longevity. Transcriptome and GO enrichment analyses of these DET revealed a significant over-representation of terms associated with response to stress and RNA processing and modification. Photosynthesis biological process was related to low seed longevity. HSF and several TF associated with biotic defense (WRKY3 and NLFX1) are candidate genes whose putative role in seed longevity deserve further characterization. We also performed the determination of the content of non-reducing soluble sugars, and we observed that the accumulation of non-reducing soluble sugars are related to acquisition of longevity but only the accumulation of them is not enough to explain the increase in longevity.*



AGROCAMPUS OUEST • Institut supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage  
65 rue de Saint-Brieuc – CS84215 – F-35042 Rennes Cedex  
Tél. : 02 23 48 50 00  
[www.agrocampus-ouest.fr](http://www.agrocampus-ouest.fr)

