UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CAMPUS DE BOTUCATU

TOWARDS UNDERSTANDING THE INFLUENCE OF SEED MATURATION ON PHYSIOLOGICAL SEED QUALITY IN LEGUMES

RUBIANA FALOPA ROSSI

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

BOTUCATU - SP
July – 2016

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CAMPUS DE BOTUCATU

TOWARDS UNDERSTANDING THE INFLUENCE OF SEED MATURATION ON PHYSIOLOGICAL SEED QUALITY IN LEGUMES

RUBIANA FALOPA ROSSI

Supervisor: Prof. Dr. Edvaldo Aparecido Amaral da Silva
Co- Supervisor: Prof. Dr. Olivier Henri Leon Leprince

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

BOTUCATU - SP
July – 2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO – DIRETORIA TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - UNESP – FCA – LAGEADO – BOTUCATU (SP)

Rossi, Rubiana Falopa, 1986-
R833t Towards understanding the influence of seed maturation on physiological seed quality in legumes / Rubiana Falopa Rossi. – Botucatu : [s.n.], 2016
xiv, 157 f. : ils. color., tabs.

Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agronômicas, Botucatu, 2016
Orientador: Edvaldo Aparecido Amaral da Silva
Coorientador: Olivier Henri Leon Leprince
Inclui bibliografia

1. Soja - Sementes - Qualidade. 2. Soja - Sementes – Desenvolvimento. 3. Germinação. 4. Longevidade. I. Silva, Edvaldo Aparecido Amaral. II. Leprince, Olivier Henri Leon. III. Universidade Estadual Paulista "Júlio de Mesquita Filho" (Câmpus de Botucatu). Faculdade de Ciências Agronômicas. IV. Título.

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

FCA

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: TOWARDS UNDERSTANDING THE INFLUENCE OF SEED MATURATION ON PHYSIOLOGICAL SEED QUALITY IN LEGUMES

AUTORA: RUBIANA FALOPA ROSSI
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 Ciencias Agronomicas de Botucatu

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

Prof. Dr. OLIVIER HENRI LEON LEPRINCE
Agrocampus Ouest - França

Dra. JULIA BUITINK
INRA - França

Dra. NATHALIE NESI
UMR - França

Botucatu, 15 de julho de 2016

I dedicate my thesis to my dear parents

João Rossi and Silvia Helena Falopa.

You are everything in my life.

ACKNOWLEDGEMENTS

One day I was told that the acknowledgements section of the thesis or dissertation is reserved to thank those who contributed directly to the accomplishement of the work in question. This looks easy, but for me this it is a little complicated. How restricting in few words something as broad as my gratitude? Throughout my walk there were so many people who contributed to my professional and personal growth. I don't even know where to begin.

Perhaps a good start is to begin to thanks my supervisors. In chronological order: Prof. Dr. Stela Maris Kulczynski was my first supervisor. Thank you to have shared with me your passion: the research. And thank you for teaching me to "look with the eyes of those who want to see". I will never forget that phrase.

Then, I started my activities in Botucatu and I had the honor of being received by Prof. Dr.Claudio Cavariani who guided me during my master studies. It was from him that I had my first official professional conversation and since then I carry with me your advices. Thank you, not only for your guidance, but for sharing with me your sincere friendship.

After that, Prof. Dr. Edvaldo A. A. da Silva became my PhD supervisor. Thank you for having supervised me, for believing in my work and for giving me the opportunity to do part of my doctorate in Angers-France. Thank you very much for contributing to my professional and personal growth.

To my supervisor in Angers Prof. Dr. Olivier Leprince. I am grateful for your supervision, guidance, suggestions and friendship; without your help I could not have finished my thesis successfully. Your commitment in seed science and encouragement made me to feel more confident to achieve my desire.

To Dr. Julia Buitink, I know you made many efforts to guide me. Thank you for your dedication and patience. It was a real pleasure to meet you in Angers and I have learned so much about seed science with you.

Participate in a sandwich doctorate program was great and wonderful challenge to me. To join the Conserto team was a privilege. Benoit Ly Vu, Joseph Ly Vu, David Lalanne, Julia Zinsmeister, Lucien Sahl, Shivan Yousif and friends group MitoStress e SESaN: Thank you very much! I really miss you all. Thank you for having received me and have not measured efforts to help me.

I want to thank a special person my dear Prof. João Nakagawa that was never officially my superviror, but at infinite time gave important suggestions to my work. It is beautiful to see your dedication to academic life. Even, more beautiful, is your lowness. You are admirable. Thank you!

In fact, the whole Department of Agriculture (FCA/Unesp), teachers, staff, technicians and colleagues, Vera Cereda, Amanda Bedetti, Eliane Gonçalves, Dorival Arruda, Thiago Gianeti and Iara Brito thanks for all assistance. A special thanks to Valeria Giandoni. A friend summarized what you mean to us in a simple phrase: "Valeria is the heart of the seed lab!!" And she's right. Thank you very much Val.

Célio Mariano, Casimiro Alves, Aparecido da Silva, Valdemir dos Santos, Milton Mateus, Antônio Camargo e Ciro de Oliveira, your help with the experiments in the field was essential.

The all friends of graduation course, especially partners and aggregates of the Seed Lab: Alexandre Carneiro, Pedro Bento, Thiago Alexandre, Vitor Maia, Leonel Pereira Neto, Girlanio Holanda, Júlio Maia, Andréia Akemi, Daiani Ajala, Letícia Moreno, Karina Ducatti, Victória Pires, Maria Rita Gilli, Bárbara Valario, Juliana Lima, Denise Basso, Bruna Luiza de Souza, Hellen Síglia Demétrio and Juliana Gimenez. To all of you thanks for your help and for sharing your knowledge and friendship above all. I always say that each new friendship, which we make throughout life, enriches us for what it reveals to us.

To Dr.Juliana Bravo, I want to extend my thanks on behalf of the entire seed laboratory. Ju, thank you for always help me. Thanks for your affection.

Lilian Silveira, Renake Teixeira and “Camilas” of my life: Camila Aquino and Camila Abrahão, thank you. It seems an exaggeration, but you were "sunshine on cloudy days". Each one in its particularity, and with special ways to speak, to advise and specially to hear. Anyway, I lack words to thank you all. I already miss you. I am a lucky person, because I have friends like you at my side.

I thank my longtime friend, Leandro Tropaldi, by sincere friendship and assistance from the graduation times to the present day. I hope our friendship lasts forever. Thank you!

Akeo Veerman, Gabriel Escobar, Raquel Lobo and David, thank you for making my days in Angers even more fun.

I would like to thank my godmother Isabel Falopa and my friends Mônica Aranha, Miriam Campos, Silvana Marques and Cristiane Cavalini who understood my absence, praying for my success and always expecting my return. Thank you.

I want share this moment with my eternal little Larissa Piccinin and my godson Arthur Teixeira. Thank you for understand my absence. I love both of you.

It is impossible to summarize what has been a doctorate in my life. I learned a lot, but I also discovered that I have much to learn; I found some answers, nevertheless, I found new questions. I overcame some fears and some obstacles. I met a wonderful person that has been important in my life. You encouraged me during my journey. To Danilo Almeida, my thanks for your affection, patience and constructive criticism. You were very important for the accomplishment of this project.

Lastly, I want to thank those people who are the basis, my safe haven: my parents, Silvia Helena Falopa and João Rossi. I left you for the last because without you nothing, absolutely nothing would be possible. Without your support, I would have retreated at the first obstacle. Nevertheless, you made a strong person. Thank you for investing in my dreams and be part of it. You are everything to me.

Now, a cycle has ended in my life. Cities, places and events, all a day passes.But, you, my friends, will continue with me, in my history and in my heart.

Thank you!

Merci beaucoup!

To the foundation institute,
São Paulo Research Foundation (FAPESP) for funding this
research by the consent of a scholarship
(FAPESP Process Nº: 2013/04598-8).

AGRADECIMENTOS

Um dia me disseram que a seção dos agradecimentos da tese ou da dissertação é o espaço reservado para agradecermos aqueles que contribuíram diretamente com a execução do trabalho em questão. Mas para mim isto está sendo um pouco complicado. Como restringir em breves palavras algo tão amplo quanto a minha gratidão? Ao longo da minha caminhada foram tantas pessoas que contribuíram para meu crescimento profissional e pessoal que eu nem sei por onde começar a agradecer.

Talvez um bom começo seja agradecendo meus orientadores. Em ordem cronológica, Prof. Dra. Stela Maris Kulczynski, minha primeira orientadora. A senhora, o meu muito obrigada. Obrigada por ter dividido comigo a sua grande paixão: a pesquisa e, por ter me ensinado a "olhar com olhos de quem quer ver". Jamais esquecerei essa frase.

Em seguida, iniciei minhas atividades em Botucatu e tive a honra de ser recebida pelo Prof. Dr. Claudio Cavariani quem me orientou durante a execução do meu mestrado. Foi com ele que tive minha primeira conversa oficialmente profissional e, desde então carrego comigo suas colocações e conselhos. Professor meu muito obrigada, não só pela sua orientação, mas por partilhar comigo sua sincera amizade.

No doutorado fui orientada pelo Prof. Dr. Edvaldo Aparecido Amaral. Muito obrigada por ter me orientado, por ter acreditado em meu trabalho e por ter me dado a oportunidade de fazer parte do meu doutorado em Angers-Fr. Obrigada por ter contribuído para o meu crescimento profissional e pessoal.

Para meu supervisor em Angers Prof. Dr. Olivier Leprince. Sou grato pela sua supervisão, orientação, sugestões e amizade; sem a sua ajuda eu não poderia terminar a minha tese com sucesso. O seu compromisso com a ciência e o seu incentivo me fez sentir mais confiante para realizar meu desejo.

Para Dr. Julia Buitink, eu sei que você não mediu esforços para me auxiliar. Obrigado por sua dedicação e paciência. Foi um verdadeiro prazer conhecê-la em Angers e eu tenho aprendido muito sobre sementes com você.

Participar do programa de doutorado sanduíche foi um maravilhoso desafio e fazer parte da equipe ConSerto foi uma honra para mim. Benoit Ly Vu, Joseph Ly Vu, David Lalanne, Julia Zinsmeister, Lucien Sahl, Shivan Yousif e os amigos do

grupo MitoStress e SESaN meu muito obrigada. Sinto muita saudade de vocês. Obrigada por terem me recebido e por não terem medido esforços para me auxiliar.

Quero agradecer aquele que, nunca foi oficialmente meu orientador, mas, por infinitas vezes desempenhou esse papel. Meu querido professor João Nakagawa, muito obrigada. Suas sugestões sempre foram importantes. Eu acho lindo de se ver sua dedicação à vida acadêmica. Ainda mais linda é a sua humildade. O senhor é admirável. Obrigada.

A todos do Departamento de Produção Vegetal – Agricultura, professores, funcionários, técnicos e colegas, meu muito obrigado. Vera Cereda, Amanda Bedetti, Eliane Gonçalves, Dorival Arruda, Thiago Gianeti e Iara Brito obrigada por toda ajuda. Um agradecimento especial a Valéria Giandoni. Uma amiga resumiu o que você significa para nós em uma simples frase: "a Valéria é o coração do laboratório de sementes!", e ela tem razão. Muito Obrigada Val.

Célio Mariano, Casimiro Alves, Aparecido da Silva, Valdemir dos Santos, Milton Mateus, Antônio Camargo e Ciro de Oliveira, pela essencial contribuição nas atividades do experimento.

A todos os amigos do curso de pós-graduação, em especial os companheiros e agregados do Laboratório de Análise de Sementes: Alexandre Carneiro, Pedro Bento, Thiago Alexandre, Vitor Maia, Leonel Pereira Neto, Girlanio Holanda, Júlio Maia, Andréia Akemi, Daiani Ajala, Letícia Moreno, Karina Ducatti, Victória Pires, Maria Rita Gilli, Bárbara Valario, Juliana Lima, Denise Basso, Bruna Luiza de Souza, Hellen Síglia Demétrio e Juliana Gimenez. A todos vocês obrigada pela ajuda nos trabalhos, pelos conhecimentos divididos e pela amizade acima de tudo. Eu gosto de afirmar que cada novo amigo que ganhamos no decorrer da vida nos enriquece, nos aperfeiçoa, não tanto pelo que nos dá, mas pelo que nos revela de nós mesmos.

A Juliana Bravo, eu quero estender meu agradecimento em nome de todo o laboratório de sementes. Ju, muito obrigada por estar sempre disposta a nos socorrer. Obrigada por seu carinho.

As minhas amigas Lilian Silveira, Renake Teixeira e as "Camilas" da minha vida: Camila Aquino e Camila Abrahão. Parece exagero, mas vocês foram "o sol em dias nublados". Cada uma na sua particularidade, no seu jeito especial de falar, de aconselhar e principalmente de ouvir. Enfim, me faltam palavras para agradecê-las. Já

sinto tanta saudade de vocês. Eu sou uma pessoa de sorte, porque eu tenho amigas como vocês ao meu lado.

Agradeço meu amigo de longa data, Leandro Tropaldi, pela sincera amizade e auxilio desde os momentos da graduação até os dias de hoje. Que nossa amizade prolongue por toda vida. Obrigada!

Akeo Veerman, Gabriel Escobar, Raquel Lobo e Dav Id obrigada por tornarem meus dias em Angers ainda mais divertidos.

Quero agradecer a minha madrinha Isabel Falopa e as minhas amigas Mônica Aranha, Miriam Campos, Silvana Marques e Cristiani Cavallini que compreenderam a minha ausência, rezando pelo meu sucesso e aguardando carinhosamente o meu retorno. Muito obrigada.

Eu quero dividir este momento com minha eterna pequena Larissa Piccinin e com meu afilhado Arthur Teixeira. Obrigada por entender minha ausência. Amo vocês.

É impossível resumir o que foi o doutorado na minha vida. Aprendi muito, mas também descobri que tenho muito mais para aprender. Encontrei algumas respostas, no entanto encontrei novas perguntas. Eu superei alguns medos e venci alguns obstáculos. No meio de tantos acontecimentos, eu conheci uma pessoa maravilhosa que tem sido importante na minha vida. Você me incentivou durante a execução do meu trabalho. À você, Danilo Almeida obrigada pelo seu carinho, sua paciência, suas correções e críticas construtivas. Você foi muito importante para a realização desse projeto.

E, por último, quero agradecer aqueles que são minha base, meu porto seguro: meus pais: Silvia Helena Falopa e João Rossi. Eu deixei para agradecer vocês por último, pois sem vocês nada, absolutamente nada disso seria possível. Sem o vosso apoio eu teria recuado no primeiro obstáculo. Mas, vocês me fortaleceram. Obrigada por investir nos meus sonhos e fazerem parte dele. Vocês são tudo para mim!

Agora um ciclo se encerra na minha vida, cidades, casas, acontecimentos, tudo um dia ficará para trás. Mas, vocês, meus amigos, continuarão comigo, na minha história e no meu coração.

Muito Obrigada!

Agradecemos à Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio financeiro e pela concessão da bolsa de doutorado (Processo FAPESP Nº: 2013/04598-8).

CONTENTS

TABLES

FIGURES

ABSTRACT

During seed maturation, germination, desiccation tolerance and longevity are acquired sequentially. Seed maturation is terminated by a desiccation phase that brings the embryo to a quiescent state. In the seed production chain, the stage of maturity at harvest is the first factor that influences seed longevity and crop establishment. After harvest, seeds are usually dried to water content compatible with long term storage and post-harvest treatments. However, there is a lack of understanding of how seed longevity is acquired during seed maturation and how premature drying impacts longevity and resumption of cellular activities during imbibition. This was addressed here by comparing transcriptome changes associated with maturation drying and imbibition of seeds of soybean and *Medicago truncatula*, harvested at an immature stage and mature dry stage. The immature stage corresponded to end of seed filling when longevity was not acquired while other vigor traits were acquired. Transcriptome characterization in soybean revealed that enforced drying was not similar to maturation drying *in planta*, which stimulated degradation of chlorophyll and synthesis of protective chaperones. Eighty-nine % of the differentially expressed genes during a 18h-imbibition period showed a similar pattern between immature and mature seeds, consistent with a comparable germination between stages. An analysis of the 147 transcripts that increased during imbibition of mature seeds but not in immature seeds suggested an activation of processes associated with shoot meristem development and DNA repair. These data were compared with imbibing immature and mature seeds of Medicago and revealed an overrepresentation of genes involved in phototropism, seed coat and innate immunity in mature seeds. This work should provide new tools to optimize harvest at maximum seed quality.

Keywords: seed quality, seed development, germination, longevity, RNAseq.

INFLUÊNCIA DA MATURAÇÃO DE SEMENTES NA QUALIDADE FISIOLÓGICA DE LEGUMINOSAS

Botucatu, 2016. 157p.

Tese (Doutorado em Agronomia/Agricultura) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista.

Autor: RUBIANA FALOPA ROSSI

Orientador: EDVALDO APARECIDO AMARAL DA SILVA

Co – Orientador: OLIVIER HENRI LEON LEPRINCE

RESUMO

Durante a maturação da semente, a germinação, a tolerância à dessecação e a longevidade são adquiridos sequencialmente. A maturação da semente termina com a fase de dessecação que traz o embrião a um estado de repouso. Na cadeia de produção de sementes, o estádio de maturação no momento da colheita é o primeiro fator que influencia a longevidade das sementes e estabelecimento da cultura. Após a colheita, as sementes são normalmente secas para um teor de água compatível com os tratamentos pós-colheita e armazenamento a longo prazo. No entanto, há uma falta de compreensão de como a longevidade das sementes é adquirida durante a maturação da semente e qual o impacto da secagem prematura na longevidade e na retomada das atividades celulares durante a embebição. Esta questão foi abordada aqui, comparando alterações transcriptoma associados com a secagem maturação e embebição de sementes de soja e *Medicago truncatula*, colhidos em um estádio imaturo e estádio seco maturo. A fase imatura correspondeu final de enchimento de grãos, quando a longevidade não foi adquirida enquanto outros traços de vigor foram adquiridos. A caracterização do transcriptoma de soja revelou que a secagem forçada não era semelhante à maturação de secagem na planta, o que estimulou a degradação da clorofila e síntese de chaperones de proteção. Oitenta e nove % dos genes diferencialmente expressos durante um período de 18 horas de embebição mostrou um padrão similar entre as sementes imaturos e maduros, consistente com uma germinação comparáveis entre os estágios. Analisando os 146 transcritos que aumentam durante a embebição de sementes maduras, mas não em sementes imaturas sugeriu uma activação dos processos associados ao desenvolvimento de meristema e reparação do DNA. Esses dados foram comparados com sementes imaturas e maturas de Medicago durante a imbebição e revelou uma sobre-representação de genes envolvidos no

fototropismo, revestimento de sementes e imunidade inata em sementes maturas. Este trabalho deve fornecer novas ferramentas para otimizar a colheita de sementes no ponto máximo de qualidade.

Palavras-chave: qualidade de sementes, desenvolvimento de sementes, germinação, longevidade, RNAseq.

CONTRIBUTION À LA COMPRÉHENSION DE L'EFFET DE LA MATURATION DES GRAINES SUR LEUR QUALITÉ PHYSIOLOGIQUE CHEZ LES LÉGUMINEUSES.

Ecole Doctorale: Végétal, Environnement, Nutrition, Agroalimentaire, Mer (VENAM)
Laboratoire d'accueil : UMR 1345 Institut de Recherche en Horticulture et Semences (IRHS).
Autor: Rubiana Falopa ROSSI
Directeur de thèse: Professeur Edvaldo Aparecido AMARAL DA SILVA
Co directeur de thèse : Professeur Olivier Henri Leon LEPRINCE

RÉSUMÉ

Pendant la maturation des graines, la germination, tolérance à la dessication et longévité sont acquises de manière séquentielle. La maturation s'achève par la dessication qui amène l'embryon à l'état de quiescence. Au cours de leur production, la maturité des graines à la récolte est le premier facteur qui influence la longévité et l'établissement de la culture lors du semis. Les graines récoltées sont ensuite séchées à une teneur en eau permettant leur conservation. On ne comprend pas comment la longévité est installée pendant la maturation et comment un séchage prématuré influence la longévité et la reprise des activités cellulaires pendant l'imbibition. L'objectif de la thèse était de répondre à ces questions en comparant les transcriptomes de graines immatures et matures de soja et *Medicago truncatula* pendant la dessication et l'imbibition. Les graines immatures furent récoltées après le remplissage avant la dessiccation, lorsque la longévité n'est pas encore acquise. Chez le soja, la comparaison des transcriptomes des graines immatures et matures montre que le séchage forcé n'est pas identique à la dessication *in planta* qui se caractérise par la synthèse de protéines chaperones. Plus de 89% des gènes différentiellement exprimés après 18 h d'imbibition présentent des profils d'expression identiques dans les graines immatures et matures, en accord avec la germination comparable de celles-ci. L'analyse des transcrits dont la teneur augmente uniquement pendant l'imbibition des graines mature suggère la mise en place de mécanismes de réparation. La comparaison de ces données avec Medicago montre que l'imbibition des graines matures se caractérise par une sur-représentation des gènes liés au phototropisme, à la testa et réponse immunitaire. Ce travail doit permettre le développement d'outil d'analyse de la maturité des graines lors de leur récolte.

THÉSE EN FRANÇAIS

Le Soja (Glycine max, 2n = 40 chromosomes) est une plante, annuelle de la famille des légumineuses, originaire d'Asie de l'Est. Les premières traces de son utilisation datent de 2.838 av. J.-C. en Chine ce qui indique qu'il peut être l'une des plus anciennes espèces cultivées (MORSE, 1950 cited in BOZATO and BOZATO, 1987). Au Brésil l'introduction du soja a eu lieu en 1882 dans les Etats de Bahia et de São Paulo (D'UTRA, 1882; DAFFERT, 1893, cited in BOZATO and BOZATO, 1987). Actuellement, le Brésil est le deuxième producteur mondial de soja dépassé par les Etats-Unis.Ssa production a atteint 100 millions de tonnes et un rendement moyen de 3,037 kg ha^{-1} pour sur une surface d'environ 33 millions d'hectares (CONAB, 2016). Les graines sont riches en protéines et traditionnellement utilisées pour la production d'huile et l'alimentation humaine et animale (QIU and CHANG, 2010).

A cause de l'importance croissante de soja pour l'économie brésilienne il y a un effort continu d'augmenter sa production, principalement en augmentant son rendement par surface cultivée. Dans ce contexte, l'utilisation de semences de haute qualité est une condition préalable à l'implantation de culture efficace et avec de bon rendement. Pour la plupart des espèces, la qualité des semences est acquise au cours du développement de la graine pendant la phase de maturation tardive (BEWLEY et al., 2013).

Des études suggèrent que les caractéristiques liées à la qualité des semences sont acquises à la maturité physiologique et commence à etre perdu peu de temps après que les graines atteignent la fin de la phase de maturation (REN et al 2009;

FRANÇA NETO et al 2007; OBENDORF et al., 1980; TEKRONY et al., 1979). En revanche, d'autres études ont mis en évidence que la qualité physiologique des graines est acquise de façon séquentielle après la phase de remplissage lorsque la teneur en eau des graines diminue et sont toujours présente chez la plante mère (CHATELAIN et al., 2012; PROBERT, 2007; DEBEAUJON, 2000; ELLIS et al., 1993; ELLIS et al., 1987).

Selon Zanakis et ces collègues (1994) dans les graines de soja la phase finale de maturation et de séchage (dessiccation), après la maturité physiologique, est considéré comme une étape importante du développement de la graine; où plusieurs composants de la qualité, tels que la tolérance à la dessiccation, la vigueur et la longévité sont acquis. Cependant chez le soja il n'est pas encore clair à quel moment la tolérance à la dessiccation, la germination et la longévité sont acquises au cours du développement de la graine, ainsi que l'influence de l'acide abscissique (ABA) sur ces événements. Chez plusieurs espèces comme le haricot, l'Arabidopsis, la tomate, le poivron et le melon, la qualité des semences augment à la phase de maturation tardive (BEWLEY et al. 2013).

Il est généralement admis que la maturation-dessiccation favorise la transition d'un mode de développement à un programme orienté vers la germination (BEWLEY et al., 1997). Ainsi, les graines matures non dormantes, comme les graines de soja, germent après l'imbibition. Il est intéressant de noter qu'une transcription *de novo* des gènes n'est pas nécessaire pour la synthèse des protéines au début de l'imbibition, ce qui suggère que la phase initiale de la germination ne dépend que d'ARNm préexistants « stockés » accumulés au cours du développement des graines (RAJJOU et al., 2004). Cela confirme l'hypothèse que la transition de la phase de développement vers la germination reste contrôlée au cours du développement des graines.

A maturité, les graines sèches d'Arabidopsis contiennent entre 12000 et 17000 transcrits caractérisés (voir revue, WEITBRECHT et al. , 2011). Ces transcrits ne sont pas tous nécessaires pour la germination et l'identité d' ARNm spécifiques et nécessaires pour la germination et quand ils sont synthétisés au cours du développement ne sont toujours pas connus. Une étude récente a montré que, pendant les 3 premières heures d'imbibition des graines d'Arabidopsis il y avait une traduction de certains ARNms du programme de développement tels que les LEA et les protéines de stockage, indiquant qu'un sous-ensemble d'ARNm stockés dans les graines sèchescaractéristiques du programme de maturation et ne disparaît pas au cours du séchage de maturation (au cours de la phase de dessiccation) (GALLAND et al., 2014).

Cependant, il n’est pas connu si une déshydratation artificielle des graines immature induirait un profil des ARN stockés similaire ou totalement différent de celui de la déshydratation associée à la maturation qui dure 14 jours. Cette information permettrait d’avoir une idée si ces ARNs sont nécessaires pour la germination et la longévité. Cela nous informera si la déshydratation artificielle est identique à celle associé à la maturation. Du point de vue de l’agriculture, ces informations peuvent être utiles pour identifier des indicateurs putatifs pour évaluer si la déshydratation suivant la récolte et le traitement des graines est capable de remplacer la déshydratation naturelle.

Cependant, des différences dans les niveaux de transcrits observés durant la maturation peuvent être présentent durant la germination, en d’autre mots, les différences dans les niveaux de transcrits durant les premières étapes d’imbibition peuvent être attribués aux degrés de maturité de la graine. Un autre élément intéressant est que durant l’imbibition plusieurs gènes peuvent être exprimés. Par contre les gènes exprimés ne sont pas seulement en lien avec la germination et les performances de la plantule, rendant difficile de séparer l’expression de gènes liés à la vigueur des graines ce celle des autres gènes. Ainsi, une analyse comparative entre les deux espèces de légumineuses permettra probablement d’éliminer un nombre important de gènes non spécifiques.

Dans ce contexte, l'objectif de cette étude est de comprendre comment la maturation des graines influe sur les mécanismes moléculaires qui régulent et conduisent à la germination des graines de soja.

Pour cette étude, les plantes de soja (BRS 284) ont été cultivées en champ à Botucatu, dans l’Etat de São Paulo en Brazil en 2013 et 2014. La maturation des graines a été suivie au cours du développement après le marquage des fleurs de selon les étapes phénologiques proposées par Fehr et Caviness (1977) avec des modifications. Les graines ont été récoltées à différents stades de reproduction pendant le remplissage et le séchage durant la maturation. La capacité germinative, la croissance de la plantule dans des conditions optimales et stressantes (15°C or NaCl); l’acquisition de la tolérance à la dessiccation; la longueur et la masse sèche de l’hypocotyle avec la teneur en acide abscissique (ABA) ont été suivis au cours du développement des graines. La Germination à la lumière et à l’obscurité, l'acquisition de la longévité ont été suivis pour les graines de soja dans les stades R7.2 et R9.

Pendant le développement, une accumulation plus élevée en eau a été observée dans les stades précoces du développement (R5.1 à R5.5) et une diminution

de la teneur en eau a été observée à partir de l'étape 7.1 jusqu'au point de récolte (stade R9). L' accumulation maximale de masse fraîche, dans les deux années de récolte, a été observée au stade R7.2 dans la récolte de l'année 2013; et aux stades R7.2 et R7.3 pour les graines récoltées dans la récolte de l'année 2014. Il a été noté que l'accumulation maximale de masse sèche se produit aux stades R7.2; R7.3 et R8.2 dans la récole de l'année 2013 et aux stades R7.3 et R9 dans la culture de l'année 2014. En ce qui concerne l'ABA sa teneur est faible au début de l'embryogenèse, mais elle augmente rapidement pour atteignant un pic de 3190,76 ng par graines au stade R6, puis elle diminue lentement jusqu'à la maturité physiologique. La capacité à germer a été progressivement acquis au cours du remplissage précoce des graines, entre les stades R5.5 et R6. La germination après séchage (tolérance à la dessiccation TD) est acquise entre le stade R6 et R7.2, avant le début du séchage de maturation qui a eu lieu entre le stade R8.1 et le stade R9 (mature).

Comme le soja est l'une des cultures mondiales les plus importantes, de nombreuses études ont été réalisées afin d'augmenter son rendement et produire des semences de qualité. Cependant, les opinions diffèrent quant à la qualité physiologique. En effet, il y a une incohérence entre les mécanismes qui conduisent à des graines de très haute qualité physiologique chez le soja et la façon dont ces mécanismes sont progressivement installés, maintenus ou perdus pendant la maturation.

Certaines recherches suggèrent que les caractéristiques à l'origine de la qualité physiologique sont acquises à la maturité physiologique (CARVALHO and NAKAGAWA, 2000) qui correspond au stade R7 du développement du soja (FEHR; Caviness, 1977) et sont perdus par la suite (FRANÇA-NETO et al, 2007; OBENDORF et al. 1980; REN et al. 2009; TEKRONY et al. 1979). Cependant, lorsque les graines achèvent le processus de remplissage, la plupart des espèces ne procèdent pas directement au séchage final, mais présentent ce que l'on appelle une «phase de maturation tardive» ('late maturation phase'), qui jusqu'à présent a reçu peu d'attention (PROBERT et al. 2007, BEWLEY et al., 2013).

Au cours des dernières années, de nombreuse études ont été menées avec les graines de *Medicago truncatula* (GALLARDO et al., 2003; ROSNOBLET et al., 2007; CHANTELAIN et al., 2012; VERDIER et al., 2014; RIGHETI et al., 2015); une plante modèle étroitement proche des légumineuses cultivées d'intérêt économique, telles que la luzerne (*Medicago sativa*), le soja (*Glycine max*), et le pois (*Pisum sativum*).

En raison des hypothèses sur la maturité physiologique et sa relation avec le stade R7.2. Dans la culture 2014, nous avons cherché à comparer la vigueur des graines immatures au stade R7.2 avec les graines matures au stade R9. Chez les graines immatures sèches au stade R7.2, la longueur de l'hypocotyle la plus élevée est mesurée à 48 h après l'imbibition ensuite elle diminue progressivement après ce stade. En outre, il a été observé que les graines sèches immatures au stade R7.2 germent plus vite que les graines matures (au stade R9) suggérant que les graines au stade R7.2 sont à pleine maturité lorsque le potentiel de germination est concerné. Cependant, il n'y a pas de différence en termes de poids sec des plantules ou le pourcentage de germination dans des conditions stressantes.

Bien que la différence de longévité entre les graines aux stades R7.2 et R9 est évidente. Les graines récoltées au stade R7.2 ont pris environ 27 jours pour atteindre le P50. Alors que, les graines récoltées au stade R9 ont atteint le P50 après 48 jours de stockage, soit presque deux fois plus long que dans les graines de stade R7.2.

Comme il y a un manque de compréhension autour de comment la maturation des graines influe sur les mécanismes moléculaires qui régulent et conduisent à la germination des graines de soja. Nous avons essayé de proposer certaines questions dans cette étude afin de trouver s'il y a des changements majeurs dans les profils de transcription au cours des derniers stades de la maturation des graines après remplissage et si le séchage prématuré imposer induit un changement radical dans le transcriptome et d'étudier les mécanismes de germination au niveau du transcriptome entre les deux stades immature et mature.

Dans ce but, l'ARN a été extrait de graines au stade R7.2 lorsque la tolérance à la dessiccation est acquise mais pas la longévité, avant et après séchage et à partir de graines sèches matures récoltés au stade R9. En outre, l'ARN a été extrait de graines imbibées pendant 18 heures dans l'obscurité, avant l'émergence de la radicule.

Le séquençage et les analyses bioinformatiques (i.e. étude transcriptomique) ont été réalisés à l'Institut de Recherche en Horticulture et Semences (IRHS) à Angers en France. Les séquences obtenues ont été filtrées à l'aide des paramètres de pureté fournies par Illumina, pour l'élimination des amorces, des vecteurs, des adaptateurs et des séquences répétées qui ont tout simplement pas d'intérêt ou qui d'une certaine façon peuvent affecter l'analyse.

Ensuite les séquences restantes, avec la qualité souhaitée, ont été alignées sur le génome de référence du soja, *Glycine max* 'Williams 82'(*Wm*82.a2.v1) disponible gratuitement sur le site Web: www.phytozome.net/soybean, en utilisant le programme Bowtie2. La normalisation et l'analyse statistique ont été réalisées à l'aide du package DESeq2 du logiciel R Bioconductor, sous la supervision de l'équipe de bioinformatique de l'INRA. L'analyse de données RNAseq a produit entre 14 et 38 millions lectures (« *reads* »)par bibliothèque (« *library* »). Toutes les bibliothèques sont cartographiées à au moins 90%. La base de données Soybase (http://www.soybase.org/genomeannotation/) a été utilisée pour l'annotation. Pour le soja environ 56044 transcrits ont été identifiés.

Des analyses de différence de niveau d'expression de gènes ont été réalisé à l'aide de l'outil statistique, DESeq, v1.12.1 (ANDERS et al. , 2010). Les variations du niveau d'expression (Fold-change) ont été calculées en divisant la moyenne de l'expression normalisée de R9 par celle de R7.2. Les gènes significatifs ont été identifié en utilisant une valeur seuil de 0.05 et une *p*-value corrigée en utilisant la méthode Benjamini-Hochberg (SEVERIN et al., 2010). Les analyses durant l'imbibition ont été réalisé en suivant la même méthode, *i.e.*, les variations d'expression ont été calculé en divisant les moyennes de l'expression normalisée de R9 imbibé pendant 18h par celle de R9 non imbibé. Il en été fait de même pour le stade R7.2

Une fois l'annotation des données effectuée les données ont été filtrées et seul les transcrit exprimés de façon significative (BH <0,05) ont été retenus. Ils ont ensuite été classés en deux catégories en fonction de la valeur numérique Log_2 (*ratio*): les gènes positifs et négatifs, respectivement les gènes surexprimés et sous-exprimés. L'étape suivante consistait à regrouper les transcrits selon leurs profils d'expression.

Lorsque l'on compare les changements de l'expression des transcrits au cours de la maturation on observe que les transcrits les plus exprimées dans les graines matures (stade R9) sont associés aux protéines de choc thermique (« HeatShockProteins –HSP »).

Le profil des gènes associé à la longévité des semences a été étudié par l'analyse des transcrits exprimés de manière différentielle avant et après l'acquisition de la longévité (i.e. entre les deux stades R7.2 et R9). Une analyse d'enrichissement de GO sur les 500 premiers transcrits différentiellement exprimés qui ont des niveaux de transcription plus élevés sur l'acquisition de la longévité a révélé une sur-représentation des termes

associés à la synthèse de HSP (« response to heat », « proteinfolding », « response to ER stress »).

Un nombre élevé de GO catégories associées à la réponse à la lumière ont également été sur-représentés tel que « response to high light intensity », « far-red light photoreceptoractivity », « response to verylow fluence », « red light stimulus », « photoreceptoractivity ». Une analyse plus approfondie des gènes appartenant à ces catégories a révélé la présence de phytochrome A et des homologues PIF (PIF7, Photoreceptorinteractingfactors), ce qui suggère un rôle pour la lumière rouge comme signal environnemental lors de la maturation des graines. Ceci est un résultat nouveau et surprenant parce que les graines de soja matures ne sont pas soumises à un contrôle de la germination par la lumière.

Peut-être que ces gènes sont nécessaires pour contrôler la photomorphogenèse pendant la levée des semis et leurs transcripts sont déjà synthétisés au cours de la maturation. Cependant, la germination des graines immatures au stade R7.2 ou R9 n'a pas révélé de différences lorsque les graines sont imbibées à la lumière (200 pmol, lumière continue). Une hypothèse alternative est que les graines en développement pourraient réagir à des signaux lumineux tels que la photopériode et le contrôle de la qualité de la lumière du processus de maturation et de l'acquisition de la longévité. Il est connu que la lumière qui pénètre dans les tissus du fruit chez le pois et le soja et qui atteignent l'embryon est principalement la lumière dans la région rouge lointain (ALLEN et al. , 2009). Aussi l'expression des gènes dans les graines en développement chez le soja est contrôlé par l'horloge circadienne avec un phasage spécifique d'organe « the circadianclockwith an organ-specificphasing » (HUDSON et al., 2010). Cependant, dans nos expériences, les graines ont été récolté toujours rigoureusement au même moment de la journée. En outre, parmi les loci génétiques caractérisant les groupes de maturité entre les accessions de soja sont deux gènes paralogues codant photorécepteur PhyA-like impliqué dans la réponse à la photopériode et le contrôle du temps de floraison aux locus de maturité E3 et 4 (LIU et al., 2008). Des travaux supplémentaires sont nécessaires pour comprendre la prévalence des termes *GO* associés à la perception lumineuse pendant la maturation et un rôle putatif de *PhyA* dans la maturation des graines. Les principales catégories de *GO* qui étaient sur-représentés dans les 500 des gènes les plus exprimés de façon différentielle avec les niveaux de transcription et qui ont diminué entre le stade R7.2 et le stade R9 ont

été associées à la photosynthèse. Ceci est expliqué par le fait de la dégradation brusque de la chlorophylle dans ce stade.

Chez le soja, les événements moléculaires conduisant au processus de germination sont encore loin d'être complètement connu. Il a été démontré que durant la dernière étape de la maturation, les graines sont capables de synthétisés des ARNm en préparation à l'état déshydraté. Ces ARNm stockés sont suspectés d'être utilisés durant l'imbibition de la graine afin d'assurer la germination (RAIJOU et al., 2004; GALLAND and RAIJOU, 2015). Les graines immatures déshydratées et les graines matures ne montrent pas de différence sur le plan de la germination. En cohérence avec ces observations nous avons trouvé que la vaste majorité des transcrits des stades R7.2 et R9, 82% des gènes différentiellements, exprimés durant l'imbibition, suivent la même tendance dans ces deux stades durant l'imbibition ce qui suggère que les principaux facteurs régulant la germination sont probablement mis en place au moins durant le stade R7.2. Nos données justifient de revoir l'hypothèse que la germination repose uniquement sur des ARNm produits durant la maturation pour produire de nouvelles protéines durant l'imbibition et assurer la germination. Parce que, conformément à Bewley et al. (2013) et Rajjou et al. (2004), la germination se produit uniquement avec l'ARNm qui sont emmagasinée pendant la maturation. Cependant, certains ARNm ne sont pas exprimés de façon significative au cours de la maturation et, en revanche, ont été exprimés et significativamentes pendant l'imbibition. Dans notre analyse du transcriptome, nous nous trouvons dans les deux étages, un changement dans l'ARNm de comparer les graines sèches et imbibé (18h). Donc, nous proposons que l'hypothèse mentionnée ci-dessus devrait être revue dans les graines de soja.

1. INTRODUCTION

Soybean (*Glycine max*, 2n = 40 chromosomes) is a native legume from East of Asia and the first reports of its use are from 2838 B.C. in China (MORSE, 1950 according to BOZATO and BOZATO, 1987). In Brazil, the introduction of soybean took place in 1882 in the State of Bahia and the State of Sao Paulo (D'UTRA, 1882; DAFFERT, 1893, according to BOZATO and BOZATO, 1987). Currently, Brazil is the second largest soybean producer in the world, preceded only by the United States and according to the latest information released by CONAB, soybeans are cultivated in an estimated area of 33 million of hectares with an estimated production of 100 million tons and an average yield of 3037 kg ha^{-1} (CONAB, 2016). Soybean is a rich source of protein and oil and has been traditionally used for oil production, food and feed (QIU and CHANG, 2010).

With the continued increase in world demand for sources of plant oil and proteins, soybean production has spread rapidly to tropical regions. In Brazil, there is a continuous effort to increase its production, mainly by increasing its yield per area. Therefore, it is imperative to know the physical characteristics of the plant, its growth stages, the nutritional demand, requirements of water, thermal and photoperiodic for proper management practices to reach increasing in soybean yield. The proper establishment of a seed production field requires careful planning, including: the choice of the region, respecting the requirement of culture in relation the availability of water (ranging between 450 and 800 mm per cycle, being higher during germination to emergence and flowering to seed filling) and temperature (ranging between 20 °C and 30 °C); the choice of the area,

considering the history, the crop rotation and the physical properties of the soil, such as fertility, drainage and topography; the choice of the cultivar analyzing the maturity group, always considering the latitude. Other care are also necessary, such as row widths, plant population (200–230 thousand plant per hectares, depending on cultivar), the weed control, pest insects control and diseases are also important in the management of culture. The sowing date is one of the factors that most influence the yield of soybeans. Seed germination and seedling emergence are favored by temperatures between 25 °C and 30 °C. Soil temperature below 10 °C results in delay in seed germination and subject to the action of soil-borne pathogens. For good seedling emergence, the soil should not exceed 85% of available water and not be less than 50%. In addition to the temperature and humidity requirements, it is necessary considered the photoperiod. Once the soybean is a term and photosensitive specie, it is subject to physiological and morphological changes when their demands are not met (SEDIYAMA et al., 1993; BERGAMIN et al., 1999; SANTOS 2008). The theoretical best time of soybean sowing in any area suitable to its cultivation, is between 30 and 45 days before the summer solstice, this time is considered sufficient for the plant meet its growing season and develop with height and size compatible to high productivity and mechanized harvesting. In general, the varieties adapted to Brazilian conditions are cycled between 90 and 150 days (EMBRAPA, 2011).

Allied to the good management practices, the use of high quality seeds are the first critical factors leading to crop yield. Seed quality consists of genetic purity, physical and physiological quality and seed health (POPINIGS, 1985). All this attributes are important to determine the quality of the seeds, but one in particular, has received more attention: the physiological quality. Represented by germination, vigor and longevity, the physiological quality determines the performance in the field, affecting establishment of the seedlings, plants development and crop yield (BEWLEY and BLACK, 1994).

The physiological quality traits are not acquired in the same time. The capacity to germinate is acquired prior to maximum dry weight. This is followed by the development of desiccation tolerance. Concomitantly, seed vigor is acquired, which is represented by greater speed of germination, uniform seedling establishment and tolerance of stressful conditions during germination. Good seedling establishment is essential for crop production to be sustainable and profitable and therefore, a critically important trait for farmers and growers. Finally, longevity increases in the last stages of development

(BEWLEY et al. 2013). Production of seeds with high physiological quality (or vigor) is a paramount to maintain soybean expansion. Ideally, the seed harvesting should occur when all the above characteristics reach their maximum levels. However, there is no consensus as to when this occurs during the maturation. On the one hand, agronomy and seed technology studies consider that physiological quality is maximum when the seed filling has ended (so-called mass maturity) and can decrease thereafter during the end of maturation drying or during seed processing (TEKRONY et al., 1979; OBENDORF et al., 1980; FRANÇA NETO et al., 2007; REN et al., 2009). This occurs because, sometimes, in agronomic crops, such as soybean, the emphasis on seed production is associated with dry weight accumulation and crop yield. On the other hand, for seed physiologist, physiological maturity refers to the developmental stage at which seeds achieve maximum viability and vigor, which is not necessarily correlated with seed filling. In many species, further maturation drying to 45% moisture is necessary to achieve maximum germination speed and absence of abnormal seedlings (ELLIS et al., 1987; ZANAKIS et al., 1994; CHATELAIN et al., 2012). Seed longevity, another key factor implicated in physiological quality increase continuously after seed filling until dispersal or harvest (ELLIS et al., 1987; ELLIS et al., 1993; PROBERT et al., 2007; CHATELAIN et al., 2012). For practical reasons, seeds are harvested during the maturation drying, otherwise they would be crushed during the harvest. Therefore, commercial harvest has to be delayed until the seed moisture decreases to levels that are compatible to harmless mechanical handling. During this period, the seeds that remain on the plant are highly prone to deterioration, particularly when humidity and/or temperature remain high, conditions that typically occur in tropical regions.

The quality of soybean seeds is partly affected by the genetic of the plant. The trend in breeding programs was initially to develop genotypes able for cultivation in tropical regions, at different latitudes. Continuously, has been sought to develop genotypes with desirable traits for resistance to diseases and pests. In the last years have been sought genotypes with increased in oil content, protein and lignin in seed and tolerance to water stress, aiming to be a way of stabilizing the productivity. The genetic variability among genotypes for quality seed is information that should be considered by breeders during the strain selection process. Since the genotypes can express themselves differently in relation to seed quality. An example is the difference of soybean genotypes for resistance to mechanical damage (CARBONELL and KRZYZANOWSKI, 1995)

which has been related to the higher lignin content in soybean seed coat (CAPELETI et al., 2005). Susceptibility to mechanical damage is associated to its lignin content, while longevity and potential deterioration in the field have been related to the degree of permeability of the integument (SOUZA and MARCOS FILHO, 2001). An example is the variability among soybean cultivars varying in color of the integument. The integument black coloring soybeans exhibit slower imbibition, increased resistance to deterioration in field, greater thickness antifungal properties, and higher lignin content compared to light-colored seed coats (CHACHALIS and SMITH 2000; SANTOS et al., 2007; MERTZ et al., 2009; DELLAGOSTIN et al., 2011).

The genotype can influence the intensity of the deterioration process. However, the seed quality is more closely related to environmental factors than genetic factors. Several studies have demonstrated variability in soybean seed composition caused by field intemperism (KEIRSTEAD, 1952; KANE et al., 1997; WATANABE and NAGASAWA, 1990; OBENDORF et al., 1998; WILSON, 2004). Generally, the temperature variation is a considerable factor in the plant growth especially during seed development (DORNBOS JR., 1995; WILSON, 2004, REN et al., 2009). High temperatures linked to excessive rainfall during the maturation can result in seed deterioration, irreversibly affecting seeds germination and vigor (TEKRONY et al, 1980; COSTA et al, 1994).When soybean seeds develop under elevated temperature, it was observed an increase of total oil and oleic acid concentration in seeds, whereas linolenic acid decrease (REN et al., 2009; CARRERA et al., 2011) and there is a negative correlation between oil and protein in soybean seeds (WATANABE and NAGASAWA, 1990).

Difficulties on germination and reduced seed longevity may be associated with response to environmental stress during development. Variations are associated with low stachyose, sucrose, and other nonreducing soluble carbohydrates, and reduction of phosphorus stored in the form of phytic acid (myo-inositol-1-phosphate) (WILSON, 2004). At molecular level, it was observed changes associated with high temperature in FAD2 enzyme (HEPPARD et al., 1996) and heat shock proteins (HSPs) (NAGAO et al., 1995). Sucrose binding protein (SBP) plays a critical role in sucrose uptake in soybean seed (GRIMES et al., 1992). Through the proteomic analysis, Ren et al. (2009) were able to identify 20 proteins whose accumulations were changed due to high

temperature. The authors stressed that high temperature during seed development results in changes in the vigor and longevity and changes in seed protein expression profiles.

The occurrence of green seeds at the end of the maturation process has been reported as another issue for Brazilian soybean growers. The green seed is a problem, because it reduces seed quality and oil quality (GOMES et al., 2003; ZORATTO et al., 2009, PÁDUA et al., 2009; TEIXEIRA et al., 2016). According to Teixeira et al. (2016) the chlorophyll retention is generally associated with pronounced increase in temperature, which leads to a rapid decrease in water content and impairment of the natural degreening. Therefore, the production of high quality seeds requires that the maturation and harvesting phases occur under mild temperatures (COSTA et al., 2003; FRANÇA-NETO et al., 2007). In Brazil there are studies analyzing the appropriate regions for the production of high quality soybean seeds. Agroclimatic zoning for the state of Parana (COSTA et al., 1994) and state of Minas Gerais (PÁDUA et al., 2014), was already performed, and the agroclimatic zoning for other regions has being researched.

While the late phase of seed development appears to be critical in order to harvest at maximum physiological quality and despite the lack of consensus as to when maximum seed vigor is acquired in soybean, we still lack basic knowledge of the molecular processes occurring during the late phase of maturation after seed filling when seed vigor is acquired. In soybean, transcriptome studies have generated a wealth of data describing seed development, mainly during embryogenesis and filling (HAJDUCH et al., 2005; HUDSON, 2010; JONES et al., 2010, LIBAULT, 2010; SEVERIN et al., 2010; ASAKURA et al., 2012; SHA et al., 2012, SHAMIMUZZAMAM and VODKIN, 2012; AGHAMIRZAIE et al., 2013). However all these transcriptome studies never included developmental stages after seed filling, while 20-40 days can pass between mass maturity and dry mature seeds depending on the environmental conditions during cultivation. In order to optimize harvesting processes to ensure high quality seeds, it is therefore essential to revisit the molecular events occurring during seed maturation in association with the acquisition of various characteristics associated with seed quality.

It is generally inferred that desiccation during seed maturation promotes the transition from a developmental mode to a germination-oriented program (BEWLEY and BLACK, 1994). In Arabidopsis, *De novo* transcription is not required for protein synthesis during early imbibition, suggesting that the initial phase of germination depends only on pre-existing “stored” mRNA, which have accumulated during seed

development (RAJJOU et al., 2004). At maturity, dry seeds of Arabidopsis contain approximately 12000 transcripts which have been characterized (reviewed in WEITBRECHT et al., 2011). This implies that these transcripts must be synthesized during seed maturation. However, when this occurs has not been investigated. Also, not all stored transcripts are necessary for germination. The identity of the specific mRNAs required for germination and when they are synthesized during development is still not known. A recent study on arabidopsis showed that during the first 3h of imbibition seeds translate some mRNAs that were part of developmental program such as LEAs and storage proteins, indicating that a subset of stored mRNA in dry seeds are characteristic of the maturation program and did not disappear during maturation drying (GALLAND et al., 2014). It is not known whether an artificial drying treatment in immature seeds would induce a similar or entirely different profile of stored mRNA compared to maturation drying. Such information would be useful to provide putative molecular indicators to assess the maturity status of harvested soybean seeds and whether post-harvest drying during seed processing is sufficient to replace natural maturation drying.

Considering the fact that the soybean genome is not small, that during soybean seed imbibition many genes are expressed, and that are not only genes related to seed(ling) performance, making it difficult to separate gene expression related to germination and seed vigor from genes involved in other functions. The soybean genome size approximately 975Mb is captured in 20 chromosomes, with 56044 protein-coding loci and 88647 transcripts have been predicted (SCHMUTZ et al., 2010). According to Mudge et al. (2005), there are highly syntenic regions in the genomes of soybean and *Medicago truncatula*, that is, regions of gene content conserved between both species. These authors reported that the up to 75% of soybean genes are colinear with *M. truncatula*, therefore, an interesting way towards understand the mechanisms involved in germination and seed vigor would be compare soybean genoma with *M. truncatula.*

Thus, there is a lack of understanding of how seed longevity is acquired during seed maturation and how premature drying impacts longevity and resumption of cellular activities during imbibition. This was addressed here by comparing transcriptome changes associated with maturation drying and imbibition of seeds of soybean and *Medicago truncatula*, harvested at an immature fresh stage and mature dry stage.

After a brief literature review and presentation of material and methods, we will first present and discuss the results obtained on soybean then on *Medicago truncatula,* following the order of the different objectives described above

2. REVIEW

2.1 Seed development

Seed development is a complex process that can be divided into three partially overlapping phases: embryogenesis, seed filling and late maturation phase (BEWLEY and BLACK, 1994). This complex process involves the interplay of a network of many developmental processes and metabolic pathways together with their interactions with the environment (BEWLEY et al., 2013).

The soybean growth stages are illustrates in figure 1. The cotyledons appear above the soil surface and provide nutrients from 7 to 10 days after sown. The time between sowing and the seed emergence (VE) is strongly influenced by the seed vigor, the soil physical conditions and fluctuations in soil temperature. New vegetative stages are developed every 5 days with normal temperatures (McWILLIAMS et al.,1999).

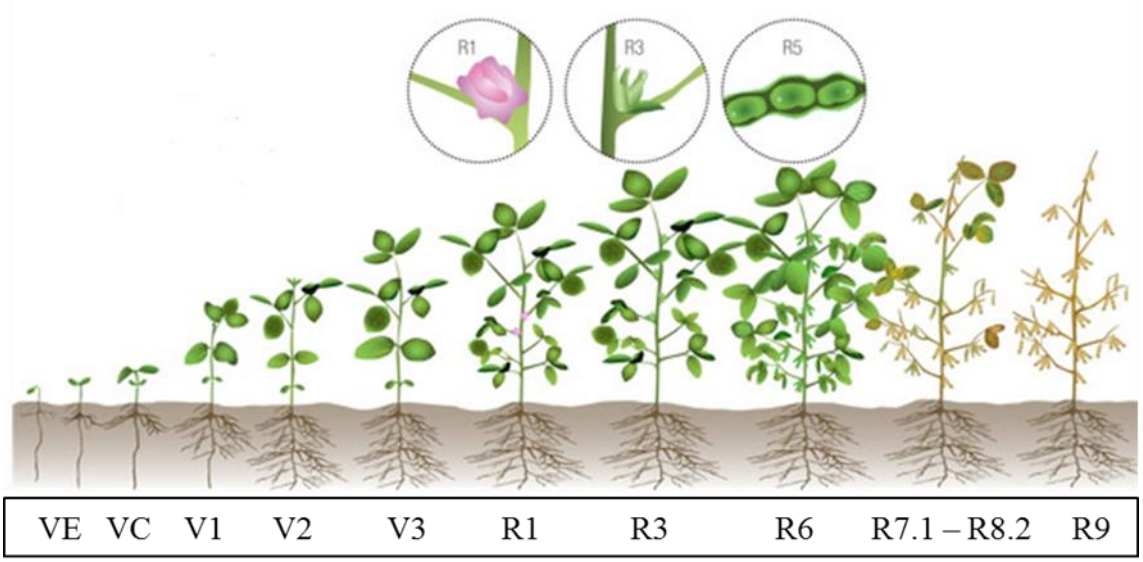

Figure 1. Soybean growth stages (available in McWilliams et al., 1999).

Induction of flowering and differentiation of flower parts are considered the starting points for seed formation. In soybean, the start of the flowering occurs after 40 to 70 days after seedlings emergence, depending on the cultivar, sowing date and weather conditions (FEHR and CAVINESS, 1977). According to Ritchie et al. (1982) in soybean seed development, around 3 to 5 days after flowering (DAF) the embryo reaches the globe or globular stage. Then, 8 to 10 DAF, starts to formation of cotyledons during the heart stage. The development continues, around 22 to 24 DAF the cotyledons are developed by taking the total size of the seed and the primordium of the first trifoliate leaf begins to develop. Figure 2, presents in detail the soybean seed development (available in LE et al., 2007).

Embryogenesis marked as a ‘lag’ phase initiated at fertilization and dominated by histodifferentiation and early cell expansion. It is a period of active cell division characterized by a rapid increase in water content with little dry weight accumulation (BEWLEY and BLACK, 1994). Maturation characterized by the ‘seed filling’, involving a rapid deposition of storage compounds (proteins, lipids and carbohydrates) and cell expansion. These nutritive reserves subsequently support the initial growth during germination and seedling establishment (BAUD et al., 2002). The final phase called late maturation or maturation drying starts when the seed has reached its

maximum dry weight and start to undergo maturation drying in preparation for the quiescent and/or dormant state (BEWLEY and BLACK, 1994).

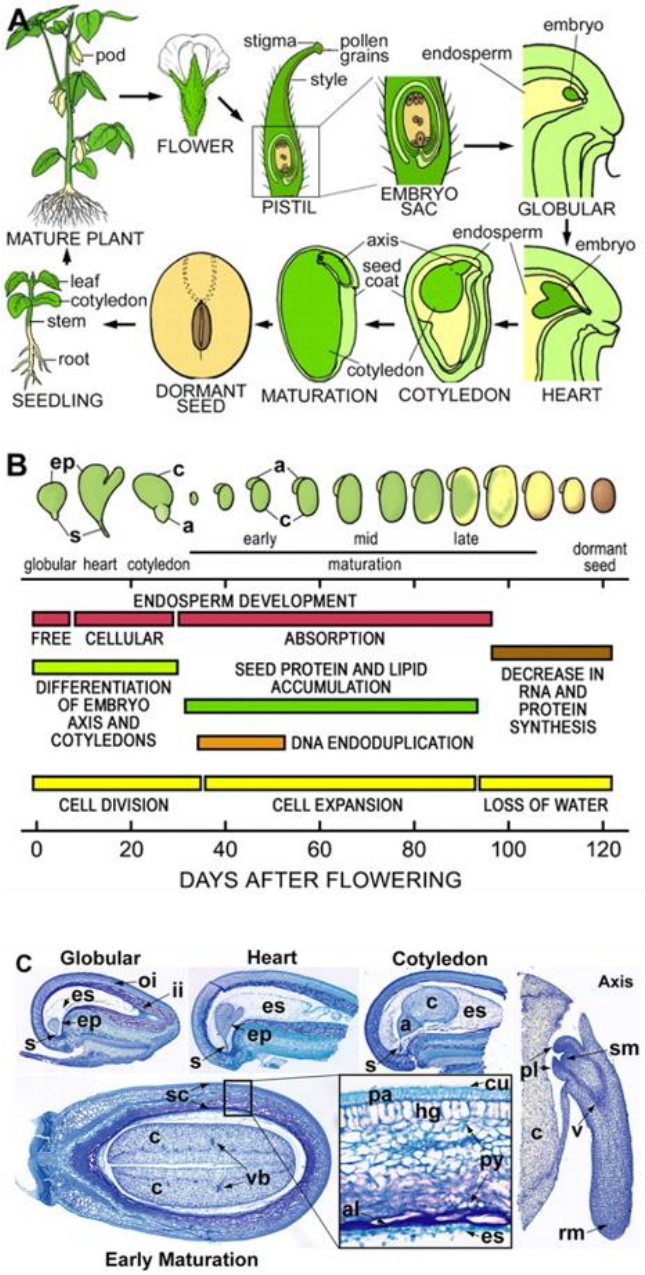

Figure 2. Soybean seed development (available in Le et al., 2007). **A**, Cartoon depicting soybean life cycle. **B**, Schematic representation of soybean seed development. Embryo morphologies and developmental events were adapted and modified from Goldberg et al. (1989). **C**, Paraffin transverse 10-µm sections of soybean globular, heart, cotyledon, and early maturation seeds. Inset contains a magnified view (40x) of the seed coat. Axis longitudinal section was obtained from an early maturation seed. a, Axis; al, aleurone; c,

cotyledon; cu, cuticle; ep, embryo proper; es, endosperm; hg, hourglass cells; ii, inner integument; oi, outer integument; pa, palisade layer; pl, plumule; py, parenchyma; rm, root meristem; s, suspensor; sc, seed coat; sm, shoot meristem; v, vascular tissues; vb, vascular bundle (LE et al., 2007).

2.1.1 Embryogenesis

The embryogenesis begins with fertilization. The fusion of the male gamete and the egg nuclei produces a diploid zygote that will differentiate and form the embryo. Following fertilization, the zygote undergoes an asymmetrical mitotic division on that gives rise to a small apical cell and a large basal cell. Further divisions lead to the formation at the preglobular embryo and the suspensor. During the seed development, the suspensor transports nutrients and growth-regulating hormones to the growing embryo, serving as the connection between maternal and endosperm tissue. In the globular stage, the embryo develops radial patterning through a series of cell divisions, with the outer layer of cells differentiating into the protoderm. In the next phase, designated heart stage, start up the development of the cotyledons, which is accompanied by changes in the internal structure beginning the organization of the tissue systems. Bilateral symmetry is apparent from the heart stage. Following, the embryo changes to a heart-shaped appearance in dicots. In monocots, only a single cotyledon is formed. Continued cell division, growth and differentiation lead to the late heart stage and then to the torpedo stage. The shoot meristem differentiates from the apical-central region of the embryo. The ground meristem of the globular stage develops into the storage parenchyma cells of the cotyledons. The procambium forms the vascular conducting tissue within in the mature embryo. The axis tissues in the mature seed are the root axis (radicle), shoot axis (plumule) and hypocotyl (BEWLEY et al., 2013).

The endosperm (triploid) development also occurs during the embryo development. The endosperm division starts to advance and occur faster than the zygote division. This happens because the embryo can use the endospermatic tissue for its development during maturation and subsequent germination/growth stages (MARCOS FILHO, 2005; BEWLEY et al., 2013). In most dicot seeds, the endosperm is almost completely consumed during development of the embryo. In these seeds, the reserves are stored in the cotyledons (MARCOS FILHO, 2005). In soybean seeds, the embryo is

formed by two cotyledons and embryo axis, which are responsible for 90% of the soybean seed weight.

2.1.2 Seed filing

The synthesis of stored reserves occurs during the seed growth period after histodifferentiation. The early and mid phases of maturation are dominated by the action of ABA, initially synthesized in the maternal tissues and later in the embryo and endosperm, although to a lower level in the embryo and endosperm (NAMBARA and MARION-POLL, 2003). Subsequently, abscisic acid (ABA) levels decline during late maturation. Most seeds contain large quantities of nutrient reserves, mainly carbohydrates, oils (triacylglycerols), proteins, and small amount of phytin, The reserves are located within the embryo itself or in an alternative storage tissue, namely the endosperm (cereals; castor bean) or the perisperm (sugar beet) (BEWLEY et al., 2013).

The storage reserves in the seed are synthesized during seed development. The synthesis occurs in specific cellular compartments for each type of substance: starch and lipid in plastids, and proteins in cytosol and endoplasmic reticulum. The accumulation of carbohydrates precedes the lipids and proteins because the carbohydrates are precursors or participate indirectly in the synthesis of these compounds in the seeds (MARCOS FILHO, 2005). The mature soybean seed is composed on average of 37% protein, 17% oil, 26% carbohydrates and 5% minerals, which together are essential for the growth, development and reproduction (MEDIC et al., 2014). Also is composed by secondary metabolites, such as alkaloid and phenolic compounds, including phenolic acid, lignin, and isoflavones; which are associated with plant defense and survival mechanisms under biotic and abiotic stress environment factors such as drought, heat and diseases (PELTIER et al., 2009). During development of soybean seed, oil accumulation is faster in the early stages, while proteins accumulate in later stages. The content of monosaccharides and disaccharides decreases during the course of development and seed maturation, while raffinose family oligosaccharides accumulate during the last weeks before full maturation (SALDIVAR et al., 2011; MEDIC et al., 2014).

Sucrose is the first source of carbon in the developing of seeds, and amino acids are the first nitrogen source. In some legumes, e.g., soybean and garden pea, the translocated sucrose produced by photosynthesis in the leaves and pods may be stored temporality as starch in the pod prior to remobilization and transfer to the developing seeds

(FERREIRA and BORGHETTI, 2004; MARCOS FILHO, 2005; BEWLEY et al., 2013). Although the starch is present in significant amounts in early development soybean seed, it is degraded to less than 1% at maturity, being considered a transitory (SALDIVAR et al., 2011).

Hexose, sucrose, amino acids and other solutes are transported to the seed via the phloem, through osmotic gradient. The pressure gradient from the font to the drain guides the mass flow in the phloem. Assimilates are translocated from the parental plant via a vascular strand that branches from the vascular tissue running through the pod. This strand passes through the funiculus and into the seed coat. Passage of assimilates through the funiculus, and from the seed coat into the cotyledons is aided by the presence of transfer cells. In soybean seeds, the phloem in the seed coat, may exhibit an extensive reticulate network (MARCOS FILHO, 2005; BEWLEY et al., 2013).There is no direct vascular connection between the embryo within the seed and the mother plant. Nutrients are therefore released into the extracellular space (apoplast) between the parental and filial tissues, and then imported into the liquid endosperme and embryo.

The compounds stored and contents in soybean seed can vary considerably across genotypes, mother plant nutrition, environmental conditions (light intensity, water availability and temperature) during development and incidence of pests and diseases. Although the soybean seed derived from the same mother plant appear to be homogeneous, it has been well known that the seed produced at the canopy may have a higher protein content, lower oil content (COLLINS and CARTTER (1956) cited by HUBER et al., 2016), and reduced concentration of mineral, as Mg, Fe and Cu (HUBER et al., 2016) than the seeds from the pods of the bottom

The increase in oil concentration in the lower seeds may be resulted from the increased duration of seed filling period, because the oil accumulation in the seeds starts before the protein accumulation (SALDIVAR et al., 2011). However, Huber et al. (2016) found that the protein and oil concentration gradients at lower to the top of the plant did not correlate with the difference in the seed filling period. The oil accumulation tends to reach a plateau before the protein accumulation and consequently, the percentage of oil content often decrease with increasing duration of seed filling period rather than increase. Analyzing the concentration of metabolites, according to seed size and node position, they suggest that a greater offer of asparagine, to seeds developing in top of the canopy, can contribute to further accumulation of storage protein.

Sharma et al. (2013) also analyzed the soybean seed according to positions. The higher rate of lipid peroxidation and lower peroxidase activity were found in seed at basal position as compared to apical positions. These results suggests that the higher lipid content in seeds from basal portion of soybean stem axis compared to apical portion might be related to higher deteriorative changes during storage.

In relation to phenolic compounds concentration (lignin and isoflavones) and cell wall boron, Bellaloui (2012), found that concentration is higher in bottom seed than in top seed along the main stem of soybean. This researcher emphasizes that this trend cannot be generalized in soybean genotypes unless enough germplasm is tested.

2.1.3 Late maturation

The third phase of seed development is the least characterized. It is mainly represented by the loss of water leading to the quiescent state after seed filling. However, when the seed filling is terminated, seeds of most species do not necessarily proceed directly to final drying, (SANHEWE and ELLIS, 1996). In arabidopsis seed, this period comprising of 1-2 days (BAUD et al., 2002; VISCENTE-CARBAJOSA and CARBONERO, 2005; LE et al., 2010), for medicago seeds in around 21 days (CHATELAIN et al., 2012) and in soybean seeds from 20 to 40 days (ZANAKIS et al., 1994).

It is generally inferred that desiccation during seed maturation promotes the transition from a developmental mode to a germination-oriented program. Indeed, slow drying at an immature stage generally induces the ability to germinate. In cereals, this can be at a very early during seed development soon after embryogenesis but in most species, slow drying must occur during seed filling. By analyzing germination of castor bean seeds during development, it was observed that immature seeds that were not subjected to artificial drying, failed to complete germination; only 30-40% of the seeds harvested at 50 DAF reached normal post-germination growth, while seeds harvested at 35 DAF and subjected to drying, showed 100% normal post-germination growth (KERMODE and BEWLEY, 1985). However, the maturation drying has received little attention in the literature. For example, it is not clear whether it is a passive phenomenon primarily

controlled by environmental conditions or a function of the physiological activity of the mother plant or seed.

2.1.4 Regulation of seed maturation programs

The transition between the different phase of seed development is coordinated by specific interactions between growth regulators (mainly ABA) and alteration in sucrose/hexose balance that is imported into the developing seeds. (VICENTE-CARBAJOSA et al., 1998; SANTOS-MENDOZA et al., 2005; GUTIERREZ et al., 2007). Seed maturation program in Arabidopsis is dependent on a transcriptional network referred to as LAFL developmental network, according to figure 3 (JIA et al., 2014). This transcriptional network is composed by several master regulators: three members of the B3 family of transcription factors, LEAFY COTYLEDON (LEC2), ABSCISIC ACID-INSENTITIVE 3 (ABI3) and FUSCA 3 (FUS3), and a fourth regulator, (LEC1) a HAP3 subunit of the CCAAT-box binding transcription factor (CBF). The role and importance of these regulators have been discovered based on the severe defects in various maturation programs as in the table 1.

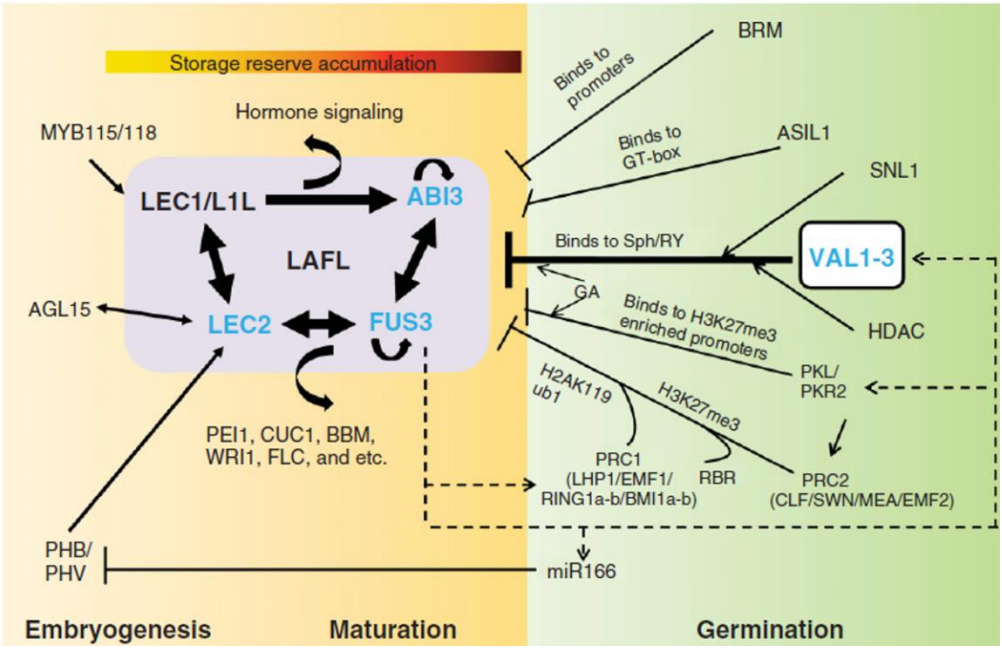

Figure 3. LAFL and VAL networks regulate the seed to seedling phase transition. Spatial and temporal patterns of LAFL gene expression are refined by mutual interactions, according to Jia et al. (2014). Important direct targets of LAFL factors include (1) SSP and

LEA genes, (2) transcription factor genes that control seed specific processes including PEI1, CUC1, BBM, WRI1 and FLC, and (3) genes that function in major hormone metabolism and signaling pathways. AGL15, PHB/PHV and MYB are proposed to act upstream of LAFL network. The LAFL network is repressed by VAL B3 factors and other repressors during seed germination to enable the transition to seedling development. Black lines with arrows indicate activation, and black lines ending with bars indicate repression. Inferred functions with less experimental evidence are indicated by dashed lines.

Table 1. Main phenotypes of mutants in the master regulators of seed maturation. Adapted from To et al. (2006) and Gutierrez et al. (2007).

Phenotype	Gene				
	Wild Type	LEC1	LEC2	FUSCA3	ABI3
Chlorophyl retention in dry seeds	No	Yes	Yes	No	Yes
Expression of storage protein	Normal	Reduced	Reduced	Reduced	Reduced
Oil content	Normal	Reduced	Reduced	Reduced	Reduced
Trichomes	No	Yes	Yes	Yes	No
Desiccation tolerance	Yes	No	Yes	Partial	No

There is a spatial and temporal pattern for expression of *LEC1*, *LEC2*, *ABI3* and *FUS3* during seed development, and their functions overlap in the general control of seed maturation (LE et al., 2010).

Other studies have shown that ABI3, FUS3 and LEC2 can regulate the expression of an oleosin gene (oleosin gene) and the accumulation of lipids (CROWE et al., 2000; SANTOS-MENDOZA et al., 2005). The loss or inactivation of *ABI3* function changes the accumulation of reserve compounds and leads to loss of desiccation tolerance, dormancy, ABA sensitivity to germination and chlorophyll degradation (VICENTE-CARBAJOSA and CARBONERO, 2005). The control and accumulation of oil in seeds is also related to the master regulator LEC2 by its regulating activity on *WRINKLED 1* (*WRI1*) (BAUD et al., 2007), which protein product targets fatty acid synthesis genes (BAUD et al., 2002; MAEO et al., 2009). *FUS3* is also related to flavonoid biosynthesis, as it represses the expression of *TRANSPARENT TESTA GLABRA1* (*TTG1*) in the protoderm (TSUCHIYA et al., 2004, LEPINIEC et al., 2006).

2.2 Acquisition of physiological quality during development

Seed physiological quality or seed vigor is a collection of properties inherent to a seed lot that allow a high and uniform level of germination, which will produce uniform, vigorous seedlings without defects under various environmental conditions, including stressful conditions. The maintenance of physiological quality during handling after harvest and subsequent storage is also critical. Poor storage conditions and/or the absence of protective mechanisms allowing long term storage will contribute to a fast deterioration of the seed lots and negatively impact germination and seedling emergence. Therefore, physiological quality also include the capacity to tolerate desiccation and the capacity to revive without defects after extended periods of storage in the dry state (i.e. longevity). Seed physiological quality can be influenced by several factors, including nutritional deficiencies in the mother plants, occurrence of pests and diseases, stressful conditions during seed development, drying, processing, storage and transport (CARVALHO and NAKAGAWA, 2000).

During seed development, the different characteristics of physiological quality are not acquired synchronously. The capacity to germinate (germinability) is acquired prior to maximum dry weight. This is followed by the development of desiccation tolerance (in orthodox seeds). Concomitantly, seed vigor is acquired, which is represented by greater speed of germination, uniform seedling establishment and greater tolerance of stressful conditions during germination. Finally, longevity increases in the last stages of development, just prior to or associate with the maturation drying phase, as shown in Figure 4 (BEWLEY et al., 2013). The acquisition of these different components will be described in the following sections.

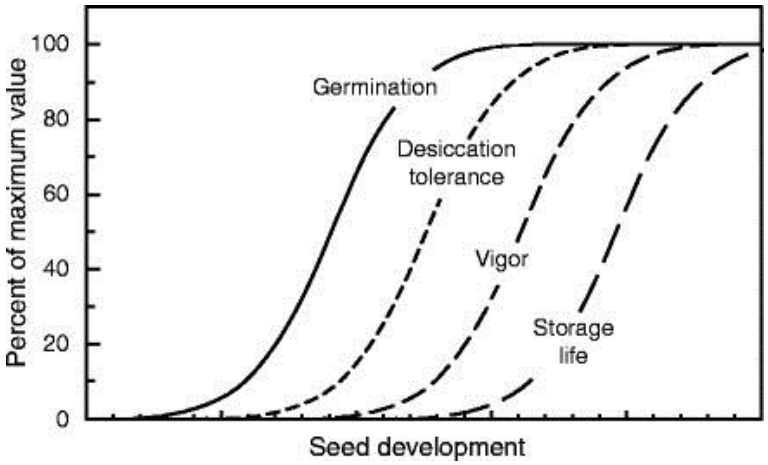

Figure 4. Generalized pattern of development of seed quality components (BEWLEY et al., 2013).

Ideally, seed crops, like soybean, should be harvested when maximum physiological quality is acquired. However, in commercial practices, this is very difficult to achieve for several reasons. Despite text book patterns similar to figure 4, there is no consensus in the literature as to when the maximum quality occurs during seed development. This is particularly true for soybean, where maturation is further complicated by the existence of several maturity groups and determinate and undeterminate cultivars. Also, the environment during seed development will influence the different seed traits. Finally, the mechanical harvest imposes that seeds are dried to a window moisture content to avoid threshing and mechanical damage.

Physical variations (such as weight, shape) or physiological (such as viability, vigour) seed attributes may be related with the pod position on the plant, or yet with the seed position in the pod (ILLIPRONTI et al., 2000), however the relationship between the position and the physiological quality attributes are not well proven.

Ghassemi-Golezani et al. (2012) stated that soybean seeds originated in the pods at the top of the stem exhibits higher quality, compared with those from the middle and lower part of the plant. Whereas for the chickpea, was reported that large and vigorous seeds were obtained from the lower part of the canopy (GHASSEMI-GOLEZANI et al. 2010). In thorough work, Illipronti et al. (2000) analyzed the relationship between the time of pod set and seed position on the mother plant, as stem type (main stem, primary branch or secondary branch), mains stem section, branch section and pod section in which of the seeds were derived. It was found that the seeds from the

main stem were heavier and larger than seeds of branches. However, in relation of seed viability it was not found different for seeds from different main stem sections or different pod sections. PRASAD et al (2000) observed difference in seed weight and also increase in the percentage of seed germination, being higher on the main stem. They emphasize though seed vigour may range between the seed produced on the same mother plant, and even among between the seed produced on the same pod. As mentioned above, the deterioration during the storage might be related to the seed position on the plant (SHARMA et al., 2013).

2.2.1 Germinability

Germinability can be defined as the ability of an immature seed to germinate. In legume seeds, germinability is usually acquired early during the seed filling in the middle of the development phase. According to Zanakis et al. (1994) and Obendorf (1980), germinability in soybean increases between 25 and 40 days after flowering (DAF). Endogenous factors from the mother plant and from the embryo, including ABA, prevent premature germination of immature seeds before the end of development (ACKERSON, 1984; HSING et al., 1995; MISRA and BEWLEY, 1985; ROSENBERG and RINNE, 1986; TEKRONY et al., 1979). In soybean seeds, as in other crops, there is a clear relationship between the embryo ABA content and the ability of immature embryos to germinate. During the normal course of embryogenesis, ABA content remains high and thus, precocious germination also do not occur. The maximum level of ABA in soybean seeds occurs between 18 and 21 days after flowering (DAF) (QUEBEDEAUX et al., 1976; ACKERSON, 1984). In addition, analyzing the germination of seeds and embryos of alfafa during development and ABA content, Xu et al. (1991) noticed that the capacity of embryos and seeds to germinate increased at the same time that content of ABA declined, and reaches maximum values shortly before physiological maturity, fluctuating only between 97% and 100%; however, DT in soybean seeds was not attained until 65 - 70 DAF.

The abscisic acid acts at different stages along the formation and maturation of the seed. As mentioned before, in the early stage of embryogenesis, the ABA produced by the maternal tissue prevents viviparity, promoting growth and development of the embryo (CHEN et al., 2002; FREY et al., 2004). During maturation, this hormone is

produced by the embryo tissue, allowing the synthesis of protein and lipid reserves. The ABA controls the transition from dormancy to germination in seeds; germination is preceded by the decrease of their endogenous content and the dormancy by maintaining its accumulation in the mature seeds (KERMODE, 2005; KUCERA et al., 2005).

In Karssen et al. (1983) and Frey et al. (2004) works with mutant and wild type species, it was shown that seed dormancy development is dependent on ABA, that is synthesized in the embryo and does ABA from maternal sources. The use of ABA synthesis inhibitor during embryogenesis resulted in the formation of non-dormant embryos (HILHORST, 1995) as well as the use of mutants deficient in ABA synthesis or perception produced non-dormant seeds (KARSSEN, 1995). The inhibitory effect in loosening cell walls of the embryo, the accumulation of β- tubulin protein, cross-microtubule organization, and replication of DNA nuclear during imbibition are some of the papers related to ABA during seed germination (HILHORST, 1995; TAIZ and ZEIGER, 2013).

Both the abscisic acid (ABA) as gibberellic acid (GA) are important regulators of germination and dormancy in seeds. Gibberellins are known as growth hormones to act on the extensibility of the cell wall, the enzyme activity and the change in osmotic potential interfering therefore directly on germination, mobilization of sugars and breaking of dormancy (KARSSEN, 1995). During seed germination, GA induces the hydrolytic enzymes synthesis such as proteases and amylases on cereals. These enzymes degrade nutritive reserves stored in the endosperm or embryo, of the mature seeds. This degradation of reserves provides energy to sustain the growth of seedlings (FERREIRA and BORGHETTI, 2004; TAIZ e ZEIGER, 2013). However, this can be inhibited by the ABA action, since the ABA induces the synthesis of proteins that prevent the GA synthesis. Seed germination is often associated with a sharp drop in the ratio of ABA and GA. The ABA inhibitory effect on germination can be reversed by GA_3, when used in concentrations that exceed their content, but from a critical level of ABA their action prevails inhibitory (CUNHA and CASALI, 1989).

In barley, it was noted that the ABA blocks the signal cascade activated by GA, before the formation of GAMyb transcription factor through the kinase protein PKABA1, which is induced by ABA (ZENTELLA et al., 2002). As a consequence this inhibiting, the gene transcription of hydrolytic enzymes does not occur and paralyzes the degradation. The balance between the levels of these hormones and their signaling

pathways are important (FINKELSTEIN et al., 2008). Finch-Savage et al. (2007) observed that genes involved in GA synthesis, as GA3*ox1*, have increased expression to break dormancy and germination, while the genes involved in ABA synthesis have decreased expression as NCED4; NCED6; NCED9.

Seeds germination is traditionally described by the "three-phasic pattern of germination", proposed by Bewley and Black (1978). Phase I: imbibition, characterized by the fast water uptake, resulted from the low matric potential of the dry seed. The water continues to get inside the seeds until all cellular content turns totally hydrated. Phase II: plateau, when the water content is relatively constant and the respiration level is low. Phase II has variable duration according to plant species, tissue composition, and hydric potential of the substrate. Phase II is characterized by preparatory biochemical activities like synthesis or release of more hexoses after initial hydrolysis and synthesis of enzymes and mRNAs. Phase III is characterized by a rapid increase in water uptake that corresponds to embryo growth resumption after visible germination (radicle protrusion). Finally, seedling establishment is initially supported by metabolites produced by hydrolysis and conversion of stored substances in the seed, as proteins, carbohydrates, and oils, until the plant turns photosynthetically active (BEWLEY and BLACK, 1978; BEWLEY et al., 2013).

In dry seeds, the phospholipid bilayer are thought to be in a gel phase. During imbibition, they turn to a crystalline state, resulting in cell damages, with a release of ions and low molecular weight metabolites from the seeds. The repair process of such desiccated membranes and the damages imposed by the quick imbibition visible in phase I starts during the imbibition (HOEKSTRA et al., 1999). Imbibing cotton seeds, and may act to stabilize and enhance cellular compartmentation (SANDOVAL et al., 1995).

Imbibing seeds have a fast resumption of metabolic activity. The essential structures and enzymes to metabolism resumption are inside the dry seeds, having persisted to desiccation and maturation phases. One of the earliest steps to metabolism resumption is respiration, which is detected already a few minutes after imbibition begins. The glycolytic and oxidative pentose phosphate pathways resumes during Phase I, and Krebs' cycle enzymes are activated (BOTHA et al., 1992). These enzymes may be protected in the dry seed by mitochondrion-specific LEA proteins (GRELET et al. 2005).

DNA might suffer from damage during desiccation and rehydration. Therefore, the damaged DNA must be repaired during seed germination. In

arabidopsis seeds the inactivation of DNA LIGASE VI, exhibited by *atlig6* mutants, results in a late germination (WATERWORTH et al., 2010). Macovei et al. (2011) observed that formamidopyrimidine-DNA glycosylase (FPG) and 8-oxoguanine DNA glycosylase/lyase (OGG1) are positively regulated during the initial phases of germination in *Medicago truncatula* seeds, suggesting its involvement in DNA repair. Another DNA repair enzymes are the poly (ADP-ribose) polymerases (PARPs) such as NIC2. Hunt and Holdsworth (2007) reported *nic2-1* mutant seeds has high levels of nicotinamide and retarded germination.

The time course of seed germination is related to both sequential and selective mRNA translation, emphasizing a fine regulation of the translational machinery (GALLAND et al., 2014).

At maturity, dry seeds contain around 10.000 transcripts called "stored mRNAs", which are associated with messenger ribonucleoprotein complexes (mRNPs). Global transcriptome analyses of several species, e.g. Arabidopsis, tomatoes and sunflowers, using high throughput approaches such as microarrays indicate that transcriptional activity is initiated in seeds during the first few hours after the start of imbibition (HAI) (WEITBRECHT et al., 2011; RAJJOU et al., 2012; GALLAND et al., 2014; GALLAND and RAJJOU, 2015). In arabidopsis seeds, the most genes positively regulated around 3 HAI are that one associated to primary metabolism, including pentose phosphate pathway since respiration is one of the earliest cell activities during imbibition. Howell et al. (2011), observing early metabolic and transcriptome changes in rice germination, found that glycolysis and the TCA cycle are activated to support energy demanding processes.

2.2.2 Acquisition of desiccation tolerance

Desiccation tolerance (DT) is the ability of certain organisms to deal with extreme water loss to levels below 0.1g H_2O per gram dry weight and completely recover metabolic function after rehydration (LEPRINCE and BUITINK, 2010; FARRANT et al., 2007). In orthodox seeds, the acquisition of DT is initiated concomitantly with the accumulation of reserves and is established before the end of seed maturation (PAMMENTER and BERJAK, 1999; BEWLEY et al., 2013). In soybean, Sun and Leopold (1993) and Zanakis et al. (1994) found that the embryonic axis acquire

germinability around 34 DAF while the the maximum dry weight was acquired after 48 DAF. However, both studies gave contrasting results regarding the induction of desiccation tolerance. According to Sun and Leopold (1993), it was acquired at physiological maturity (48 DAF) whereas Zanakis et al. (1994) showed it was was acquired at around 66 DAF, after the end of seed filling and after physiological maturity (Figure 5).

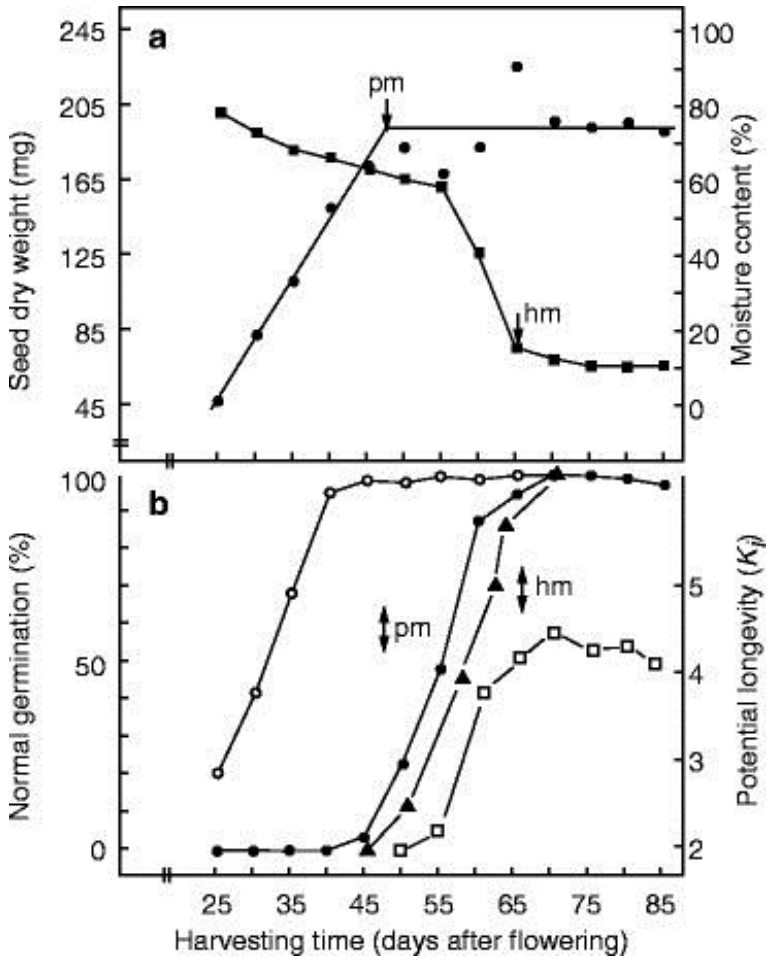

Figure 5. Changes in (a) seed dry weight (●) and moisture content (% wet basis, ■) and (b) the capacity of freshly harvested (o) and dried (to 10–11% moisture content, ●) seeds to germinate normally within 14 days (o, ●) or for dried seeds to germinate within 3 days after imbibition (▲). Also shown in bare changes in the potential longevity (Ki) of seeds harvested and dried at different times of development and then aged at 12% moisture content and 40°C (□). Modified from Zanakis et al. (1994) in Bewley et al. (2013).

Desiccation tolerance is a multifactorial phenomenon in which many components act in synergism (for reviews see HOEKSTRA et al., 2001; LEPRINCE and BUITINK, 2010). They include protective mechanisms that can be grouped in three categories, namely: 1) protection by stabilization of membranes and proteins by non-reducing sugars such as sucrose and raffinose oligosaccharides family, late embryogenesis abundant (LEA) proteins and heat shock protein (HSP). 2) protection against oxidative damage and ROS production by a range of antioxidant compounds including tocopherols, glutathione, together with a coordinated response of metabolism during drying and 3) protection against structural stresses imposed by drying such as cell wall modification, reorganization of endomembranes and cytoskeleton, vacuolization and chromatin condensation.

The carbohydrates, proteins and lipids correspond to the main sort of seed reserves, and their proportions depend on each species and the development stage of the seeds. The soluble sugars correspond to a small part among the carbohydrates in the seeds, highlighting the glycose, fructose, mannose, galactose, sucrose, and the oligosaccharides from the raffinose family. Moreover, the soluble sugars correspond to an important protection to the seeds, limiting the damages due to desiccation of mature seeds (BUCKERIDGE et al. 2000).

The accumulation of raffinose family oligosaccharides may occur due to the conversion of monosaccharides, decreasing the availability of substrate for respiration and, consequently, decreasing the metabolic activity during the stages of desiccation and storage (LEPRINCE et al. 1992, PAMMENTER and BERJAK 1999). According to Hoekstra et al. (1994), soluble sugars can prevent the harmful effects of desiccation over the cell membranes during the dehydration period. Soluble sugars have hydrogen bonds, and they replace the water over the membrane surface, holding the space of lipids, and avoiding the transition of the liquid phase to the gel phase. Hoekstra et al. (1994) also reports that intolerant seeds to desiccation has soluble sugars in their composition, however, the forms that they are deposited during the dehydration do not results in protection. Additionally, some specific proteins are synthetized lately in the seed development, and they has been associated to desiccation tolerance or seed longevity, as late abundant embryogenesis (LEAs) and heat shock proteins (HSPs).

Late abundant embryogenesis proteins are largely common among plant species and can be grouped in diverse homologue families (WISE, 2003). The LEAs

are composed by hydrophilic amino acids, resulting in high solubility and high denaturation resistance (WANG et al., 2003). The LEAs can protect cellular structure, membranes, and other proteins, once they acts as hydration buffer, sequestering ions e rehydration of proteins (reviewed by TUNNACLIFFE and WISE 2007).

According to Blackman et al. (1991), LEAs are responsible for the desiccation tolerance acquisition in soybean seeds. The level of LEAs was correlated to desiccation tolerance, as in the development phase as well as in the germination phase, increasing with the tolerance desiccation and decreasing during imbibition. Although, there are evidences that LEAs acts in some role in the tolerance to desiccation, their biological role in seeds developed in vivo still not clear (BEWLEY et al., 2013).

The heat shock proteins, or stress proteins, also referred to as molecular chaperones, play in an important role in protein stabilization such as assembling of multi-protein complex, folding or unfolding, transport or sorting of proteins into correct compartments at sub-cellular level, control of cell-cycle and signaling, as well as cell protection against stress or apoptosis (LINDQUIST and CRAIG, 1988). In plants, has already been reported more than 30 families of HSPs, that are highly preserved (WALTERS, 2013), and synthetized during the embryo development and fruits maturation (KALEMBA et al., 2012).

According Vertucci and Farrant (1995), the role of HSPs is related to conservation and repair of macromolecular structure during dehydration and hydration. Although the HSPs are synthetized in organisms as a response to heat, the balance of synthetized proteins and the relative importance of the HSPs families for stress tolerance are very different among the species (QUEITSCH et al., 2000). The gene expression for HSPs during the development of seeds generally occur as the same time of gene expression for LEAs, and both has a positive response to ABA.

In studies with *Medicago truncatula* seeds, Larre et al. (2006), reported that the loss of desiccation tolerance during the germination followed the reduction of HSPs amount. In *Arabidopsis thaliana* embryos, it was also reported the accumulation of HSPs in the middle phase of maturation and that HSPs amount is high after the acquisition of desiccation tolerance (WEHWEYER and VIERLING, 2000). Gallardo et al. (2001) observed that the amount of some LEAs and HSPs was decreased at the end of germination process, while some other remained during all the germination

process, showing that some LEA and HSPs acts in protection role, not just in the seed maturation, but also during all the germination process.

The regulatory mechanisms and signaling pathways controlling the induction of these protective mechanisms leading to DT are still poorly understood. However, it is known that ABA plays an important role in desiccation tolerance. ABA controls transcription of the SSP genes, and possibly the genes responsible for accumulation of other reserves (VICENTE-CARBAJOSA and CARBONERO, 2005). High levels of ABA, or appropriate ABA/GA levels, are thought to maintain the seed in a quiescent or dormant state, germination being brought about by the reduction in ABA and/or reversal ABA/GA ratios (WHITE and RIVIN 2000, VICENTE-CABAJOSA and CARBONERO 2005). Thus, it is not surprising that ABA is thought to play an important role in acquisition and loss of desiccation tolerance. Probably the best reported role for ABA in control of desiccation tolerance is the regulation of lea gene transcription (BRAY, 1993, KERMODE 1990, KERMODE and FICH-SAVAGE 2002). In many cases, exogenous application of ABA can stimulate the accumulation of LEAs.

The ABA-deficient and -insensitive mutants of *Arabidopsis thaliana* enable the elucidation of crucial factors involved in desiccation tolerance. Three different loci for ABA insensitivity, designated *abi1*, *abi2*, and *abi3*, have been described. The *abi1* and *abi2* mutants suffer from water stress in the vegetative stage, whereas the *abi3* mutation specifically affects seed development (KOORNNEEF et al., 1984). The *abi3* mutant seeds exhibit reduced LEA contents and reduced chromatin condensation. In addition, the fact that chlorophyll is retained in dry seeds of *abi3* mutants might be interpreted as a symptom of lack of coordination to metabolic reduction during drying. According to review by Sano et al. (2016), the ABA, mediated by ABI3, regulates pathways in seed dormancy, desiccation tolerance and seed longevity. In Arabidopsis seed, a central role of seed longevity signaling has been attributed to downstream ABI3 through the expression of heat shock transcription factor A9 (HSFA9; KOTAK et al., 2007).

2.2.3 Acquisition of longevity

During seed development, seed longevity is the last quality traits to be acquired (BEWLEY et al., 2013). Seed longevity is defined as the capacity to remain viable during storage in the dry state (BUITINK and LEPRINCE 2004; CHATELAIN et

al., 2012). It is acquired progressively during the late phase of seed maturation but this has not been characterized extensively. In soybean and Medicago, longevity increases steadily during maturation drying and reaches a maximum at around seed abscission. This decline is usually attributed to field weathering, corresponding to wet and hot conditions that prevent seed drying or worse rehydrate seeds (TEKRONY et al., 1980; CHATELAIN et al., 2012). In soybean production, high temperatures linked to excessive rainfall during the maturation can result in seed deterioration, irreversibly affecting seeds germination and vigor (TEKRONY et al., 1980; COSTA et al., 1994).

In other species, longevity can increase steadily during maturation until seed abscission after which it remains steady or continues to increase during further drying (PROBERT et al., 2007).

After harvest, seed longevity depends on the moisture content (or ambient relative humidity), oxygen pressure and temperature during storage. High temperature and high humidity induces a rapid deterioration whereas storage in dry and cold conditions (e.g. 5°C, 15% RH) can prolong the shelf life by a factor 10 to 100 according to the species. This increases is due to the formation of a glassy state, which correspond to an amorphous matrix where the cytoplasmic viscosity is so high that it resembles a solid (BUITINK and LEPRINCE, 2004). In the glassy state, there is no active metabolism because the high viscosity prevents the formation of ATP. In addition, deleterious reactions are severely slowed down, thereby stabilizing the seeds. Several authors have used so-called accelerated aging treatments as a proxy to determine seed longevity and thereby seed vigor. Typically, seeds are stored at 42°C and 100% RH which induces a combination of progressive rehydration and heat stress which rapidly deteriorates the seeds (TEKRONY et al., 1980, RAJJOU et al., 2008). There are claims that these types of treatments mimic the storage behavior in optimal conditions or represents a good test to assess physiological maturity. However, both genetic and biochemical evidence show that the mechanisms of deterioration are not identical at different water content/temperature combinations.

Seed longevity is attributed to two main strategies: protection and repair (reviewed in SANO et al., 2015; RAJJOU and DEBEAUJOU, 2008, BUITINK and LEPRINCE, 2008). Protection is offered by the presence of high concentrations of non-reducing sugars, including sucrose and RFO that are good glass formers. Also, in vitro studies have shown that these sugars protect the cellular integrity during desiccation by

stabilizing membranes and proteins on and can also act as antioxidant. Other protective compounds include LEA proteins (HUNDERTMARK et al., 2011; CHATELAIN et al., 2012), heat shock proteins (PRIETO-DAPENA et al., 2006) and a set of antioxidant defenses against oxidative stress such as glutathione, tocopherols and flavonoids present in the testa (DEBEAUJON et al., 2000; SANO et al., 2015). Seed longevity has also been shown to be associated with DNA and protein repair systems such as DNA ligase (WATERWORTH et al., 2010), protein L-isoaspartyl methyltransferase (OGÉ et al., 2008) and methionine sulfoxide reductases (MSRs; CHATELAIN et al., 2013).

2.2.4 Acquisition of physiological maturity

We mentioned above that seed harvesting must occur when the seed potential is the highest, that which corresponds to the so-called physiological maturity (maximum germinatibility, longevity and vigor). Several works have focused on determining the criteria that define physiological maturity using a range of physiological assays (percentages of germination and desiccation tolerance, seedling growth, various parameters of seed longevity), technological assays (accelerated aging, controlled deterioration, percentage of abnormal seedling) and biochemical assays (electrolyte leakage, chlorophyll content). In soybean, seed maturity was first associated with leaf abscission, and leaf and pod colors. Willard (1925) related the maturity point to yellowing of soybean leaves, changes in pod color, and to a leaf abscission rate of approximately 50%. According to Marcos Filho (2005) the seed physiological maturity was proposed for the first time by Shaw and Loomis (1949), using as criterion the maximum dry weight. This stage was also called “relative maturity” by Aldrich (1943) and “morphological maturity” by Anderson (1955), all of them had as criterion the maximum dry weight (MARCOS FILHO, 2005). However, data from other experiments showed that physiological maturity does not always correspond to maximum dry weight. Oat, beans, rice, and soybean seeds acquired high germination and vigor after maximum dry mass was reached, suggesting that biochemical changes promoting metabolic adjustments after the end of reserve deposition are still happening and can influence seed quality (MARCOS FILHO, 2005). Ellis and Pietra Filho (1992) showed that the maximum physiological potential occurs after physiological maturity and proposed the term “mass maturity” to denominate the point of maximum dry mass, discriminating it from the point of maximum

vigor. Zanakis et al. (1994) observed also in soybean seeds that the maximum dry mass was reached 45 DAF (Figure 5a), when the seeds are intolerant to desiccation, and the desiccation tolerance is acquired between 45 and 55 DAF, much later than mass maturity. Also, seed vigor, assessed by the Ki of the viability equation that is used to assess seed longevity during storage increases after mass maturity (Figure 5b).

Increase in seed quality, after the reach maximum dry weight, has been confirmed in other studies (ELLIS et al., 1987; ELLIS et al., 1993; PROBERT, 2007; CHATELAIN et al., 2012).

3. MATERIAL AND METHODS

3.1 Plant materials, growth conditions and sampling points

3.1.1 Soybean

The study was performed at the São Paulo State University - College of Agricultural Science, UNESP - FCA (Botucatu, SP – Brazil). The soybean cultivar used in this study was BRS 284, fitting to early maturity group with cycle of 120-126 days. The experiment was performed in 2 consecutive years, 2012/13 and 2013/14. The experimental area was located at latitude 22°50' S and longitude 48°25' W, and altitude of 750 m. The climate is a humid subtropical with dry winters and a well-defined dry season between the months of May and September. Mean rainfall averages is about of 1500 mm, while the highest monthly mean temperature is over 23°C and the lowest is below 18°C (CEPAGRI). The soil in the experimental area is classified as a Nitossolo Vermelho (EMBRAPA, 2006). It was collected soil samples before soybean sowing in the years of 2012 and 2013, to evaluate fertility according to Raij et al. (1997). The chemical characteristics of the soil in the experimental area are presented in table 2.

Table 2. Chemical characteristic of the soil at 0-0.20 m depth before sowing in 2012 (A) and 2013 (B).

	pH	MO	P_{resin}	K	Ca	Mg	H+Al	SB	CTC	V
	$CaCl_2$	$g.dm^{-3}$	$mg.dm^{-3}$	.....		$mmol_c.dm^{-3}$	.....			(%)
A	5,8	30	41	1,3	38	20	32	59	91	65
B	5,2	20	31	2,5	42	19	37	63	100	63

According to recommendation for Sao Paulo State, Brazil, (RAIJ et al., 1997), it was applied 250 kg ha^{-1} of the formulate 02-20-20 in the furrows to provide 5 kg ha^{-1} of N, 50 kg ha^{-1} de P_2O_5 e 50 kg ha^{-1} de K_2O.

Seeds were treated with fungicide Carboxin and Thiran (Vitavax-Thiran 200 SC) at a dose of 2.5 mL kg^{-1} of commercial product (c.p) per seeds and insecticide Thiametoxan (Cruiser 700 WS), at a dose of 2,0 mL kg^{-1} of c.p. per seeds. At the day of sowing, the seeds were inoculated with the bacteria *Bradyrhizobium*, using the c.p. Biomax-Biosoja, at a dose of 2 mL kg^{-1} of seeds. The seeds were sown on December 4th 2012 and December 6th 2013 respectively. There were used 0.45 m between rows and was maintained a population of 17 plants m^{-1}. All the phytosanitary controls were performed according to EMBRAPA (2011).

The BRS284 genotype used in this experiment was selected from a list of conventional materials, i.e, non- transgenic, and considered recommended for the climatic conditions of the growing region.

The development of the fruits was monitored and soybean flowers in the main stem were tagged. Flowers from the lower part of the plant and from branches were not tagged. During the development the pods were manually removed from more than 200 plants at each stage and the seeds were quickly extracted and homogenized. The harvest points were performed at different time intervals, based on days after flowering and phenological scale of soybean plant development (Table 3). As this scale considers mostly observations about the plant, leaf and pod colors, we observed carefully and added information about the seed characteristics to choose the correct harvest point; detailed changes were observed, such as changes in seed water content, seed weight, changes in the color of the cotyledons and embryonic axis.

3.1.2 Medicago

Plants of *Medicago truncatula* were grown in a sterile mix of vermiculite and soil in a growth chamber at 23°C with a 16 hours photoperiod at 200 µmol m^{-2} s^{-2} light intensity and 60–70% relative humidity. The development was well monitored; flowers were tagged and developing seeds were removed from de pods at 28 days after pollination and at final abscission, corresponding to mature dry seeds. Immature seeds were rapidly dried according to Chatelain et al (2012).

Table 3. Characterization of the phenological stages of soybean plants at the main stem proposed by Fehr and Caviness (1977) and Ritchie et al. (1982).

Reproductive Stages	Nomination	Description
R1	Beginning flowering	One open flower at any node.
R2	Full flowering	Open flower at one of the two uppermost nodes on the main stem with a fully developed leaf
R3	Beginning pod	Pod with 5mm long at one of the four uppermost nodes on the main stem with a fully developed leaf
R4	Full pod	Pod with 2cm long at one of the four uppermost nodes on the main stem with a fully developed leaf
R5	Beginning seed	Seed with 3 mm long in the pod at the for uppermost nodes of the main stem with a fully developed leaf
R5.1	~ 10% filling	Seeds ~ 5mm long in the pod (perceptible to touch)
R5.2	~ 11 - 25% filling	Seeds ~ 7mm long in the pod
R5.3	~ 26 - 50% filling	Seeds ~ 8mm long in the pod
R5.4	~ 51 - 75% filling	Seeds ~ 10mm long in the pod
R5.5	~ 76 - 100% filling	Seeds ~ 11mm long in the pod
R6	Full seed	Pod containing a green seed that fills the pod cavity at one of the four uppermost nodes on the main stem with a fully developed leaf (plant with green leaves)
R7	Begin of maturation	One normal pod on the main stem that has reached its mature pod color
R7.1		~ 50% yellowing of leaves and pods
R7.2		~ 51 and 75 leaves and yellow pods
R7.3		~ 75 pods and yellow leaves
R8	Full maturation	~ 95% of the pods have reached their mature pod color
R8.1		Beginning at 50% defoliation
R8.2		More than 50% defoliation at early harvest
R 9	Harvest maturity	Point of harvesting

Table 4. Days after flowering (DAF) at different reproductive stages of soybean seed cultivated in the 2013/2014.

Reproductive stages (R)	Days after flowering (DAF)
R5.1	25
R5.2	25
R5.3	27
R5.4	27
R5.5	36
R6	47
R7.1	57
R7.2	63
R7.3	69
R8.1	71
R8.2	73
R9	77

3.2. Water content determination

Water content was determined by oven drying at 105 ± 3° C for 24 hours (BRASIL, 2009), with four replicates of 20 seeds. The water content of the seeds was expressed as gram of water per gram of dry weight (g H_2O per g DW).

3.3. Fresh weight (FW) and dry weight (DW)

The fresh and dry weights of the seeds were evaluated by using precision analytical balance. Fresh weight and subsequent dry weight were determined in an oven at 60^o C for 72 hours or more until constant weight. Four replications of 20 seed each were used and the results were expressed in mg per seed.

3.4. Seed quality assessment

Physiological quality of seeds was assessed as described below:

Germination – germination test was evaluated by using four replicates of 25 soybean seeds harvested in different reproductive stages (Table 4), using paper rolls, imbibed in water at 25ºC in the dark. The immature dry seeds at the stage R7.2

and mature seeds at stage R9 were also evaluated by using paper rolls, imbibed in water at 15°C or imbibed in salt solution (75 mM or 100 mM NaCl) at 25°C, both condition in dark. Or imbibed in water but in light condition (200 µmol m-2 s-2). For *Medicago truncatula*, three replicates of 50 seeds of each stage were imbibed in water at 20°C in the dark; in cold condition at 10°C in the dark and in salt condition. For both seeds, soybean and Medicago, the germination percentage was scored daily by counting radicle protrusion, when the radicle has more than 2 mm length (BEWLEY and BLACK, 1994).

Normal seedlings – normal seedling formation was scored after eight days of imbibition. The seedlings were considered normal when presented all the essential structures to ensure the development of a plant. In other words presence of primary and secondary root well developed, hypocotyl, cotyledon, epicotyl and plumule healthy (BRASIL, 2009).

Desiccation tolerance - desiccation tolerance (DT) was evaluated by scoring germination (radicle protrusion and normal and abnormal seedlings) after rapid drying of the seeds over an airflow of 42% relative humidity (RH), generated by a saturated salt solution of K_2CO_3, according to Chatelain et al. (2012). The seeds were dried to approximately 0,10 g H_2O per gram dry weight; independently of the reproductive stage. For this, the drying time varied from 24 to 72 hours. Seeds were considered desiccation tolerant when they germinated.

Hypocotyl length – the length of 10 hypocotyls was measured after 48 hours of imbibition in water, at 25ºC in the dark or in the light.

Hypocotyl dry weight – hypocotyls were placed to dry in an oven with air circulation at 80°C for 36 hours.

Longevity – Immature dried seeds at stage R7.2 and mature seeds at stage R9 were kept over a saturated solution of NaCl (75% RH) at 35°C in hermetically sealed plastic boxes (CHATELAIN et al., 2012). At different times during storage, seeds were imbibed as described above and germination was recorded. Longevity was expressed as P50, defined as the time (days) for which the seeds lose 50% of its viability during storage.

3.5. Abscisic acid (ABA) quantification

Quantification of abscisic acid (ABA) biosynthesis was performed using High Performance Liquid Chromatography (HPLC). Three replications of 15

embryos (cotyledons and embryonic axis) of fresh soybean seeds were lyophilized for 72h. They were frozen in liquid nitrogen and ground to a powder. ABA extraction was done as proposed by Ma et al. (2013), using 150 mg of embryos for each repetition. Extraction solution were composed of methanol, acetonitrile, Milli-Q water and acetic acid (v: v: v: v, 40: 40: 20: 1)

The chromatographic separation was performed on HPLC Shimadzu model Prominence comprised of degasser mobile phase DGU-20A, the quaternary pumping system comprised of pump LC-20AD, sampler model SIL-20ACHT, controller CBM-20A and oven model columns CTO-20AC. The column used was a Sinergi 2.5 Hydro RP-100A 50x4.6mm, maintained at 40°C during quantification. The column effluent was introduced into the mass spectrometer type triple quadrupole (MS/MS) brand AB Sciex model 4500 equipped with ionization source type ESI (eletrospray) at the interface. The results were expressed in nanograms per seeds (ng per seed).

3.6. Statistical analysis for seed quality assessment

An ANOVA test was performed on seed quality and gene expression data at the level of 5% ($p \leq 0.05$). The averages were compared by T-test at 5% probability level.

3.7. Transcriptome analysis

For the transcriptome studies, two reproductive stages of soybean seeds were selected and analyzed before and after 18 h of imbibition 9 (Table 5).

Table 5. Reproductive stages of soybean seeds used in the transcriptome analysis.

Reproductive stages		Hours after imbibition
Immature seeds	R7.2 fresh	0h
	R7.2 dried	0h
	R7.2 dried	18h
Mature seeds	R9	0h
	R9	18h

3.8. RNA extraction and sequencing (RNAseq)

In the crop year 2013/2014, seeds were harvested from more than 200 plants at each stage and stored - 80°C in deep freeze. Total RNA extraction was performed from two independent duplicates using the NucleoSpin® RNA Plant kit (Macherey-Nagel); using immature seeds (stage R7.2) and mature seeds at stage R9. R9. 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.

High quality of total RNA samples were used for library preparation and mRNA sequencing at the Central Laboratory of High Performance Technologies (LaCTAD) in Campinas, SP – Brazil. 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 and sequenced using a Hiseq 2500(Illumina). On the flowcell surface there are oligonucleotides that are complementary to the adapters of libraries. The flowcell was placed in a HiSeq 2500 where occurred the incorporation of nucleotides containing fluorescently labeled dideoxy terminators on the fragments linked to the sequencing primer. Upon the occurrence of the incorporation of a nucleotide, the fluorescence was excited with a number of lasers and captured by cameras.

3.9. Bioinformatics analysis

Transcriptome studies were performed at the Research Institute of Horticulture and Seeds (IRHS, *Institut de Recherche en Horticulture et Semence)* in Angers – France.

The bioinformatics analysis was done in three steps. The first step it was a filtering process, then the mapping by analysis of the candidate genes. The sequences obtained were filtered using parameters of purity provided by Illumina, for the removal of primers, vectors, adapters and long repeated sequences of bases that were not of interest or, in some way, can affect the analysis. Then, the alignment program Bowtie2 (LANGMEAD et al., 2009) was used to align reads to the 'Williams 82' soybean reference

genome, assembly *Glycine max* Wm82.a2.v1 (SCHMUTZ et al., 2010) as a reference (www.phytozome.net/soybean).

The normalization and statistical analysis were performed by using the tool DESeq2, v1.11.21 (LOVE et al., 2014) available as package of the software R Bioconductor (GENTLEMAN et al., 2004). 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).

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 was calculated by dividing the mean gene expression value obtained in the immature stage (R7.2) by the value obtained in mature stage (R9). i.e., the immature stage was considered as the control for the calculation of expression relative to mature (treatment). The ratio of expression between both stages was considered through the logarithmic division (Log_2 [R7.2 : R9]).

3.10. Gene Ontology Analysis

GO enrichment analysis of differentially expressed genes it was performed using Soybase GO-enrichment tool (http://www.soybase.org/goslimgraphic_v2/dashboard.php, MORALES et al., 2013) using GO annotation terms from genome version *Glyma* 2.0. Over- and under-represented GO terms were identified by comparing the GO term enrichment within differentially expressed genes and the entire soybean genome using Fisher's exact test with a Bonferroni correction. Also performed the Gene ontology enrichment analysis using AgriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php, Du et al., 2010), using Arabidopsis thaliana genome annotation (TAIR 10) as genome model and data was corrected for multiple testing error (FDR). For both database the significant level of 0.05 ($p\text{-}value \leq 0.05$) was used for filtering.

3.11. cDNA synthesis and quantitative RT-qPCR

The RNA used for gene expression analysis by RT-qPCR was extracted from the soybean developing stages presented at table 3. Three replicates were used for RNA extraction as previously described. cDNA synthesis was performed by using the iScriptTM Reverse Transcription Supermix for RT-qPCR according to the manufacturer's protocol (Bio-Rad). For a reaction of 20 µL, was used 4 µL of iScript RT Supermix and 1 µg RNA) and completing the volume with Nuclease-free water. The reactions were incubated in a thermocycler at 25°C for five minutes for the initial activation of the enzyme and at 42°C for 30 minutes for cDNA synthesis, followed by inactivation of the enzyme at 85°C for five minutes and concluding the cycle at constant temperature of 12°C.

For the expression of candidate genes (Glyma.14G047000, *Glyma.*03G224700, *Glyma.*02G200500, *Glyma.*10G066100 and *Glyma.*14G217700) was used LuminoCt SYBR Green qPCR ReadyMix (Sigma Life Science), and the volume of the reactions in a total solution of 12.0 µL corresponding to: 6 µL (1x) of SYBRgree; 0,25 µL (10mM) of forward and 0,25 µL (10mM) of reverse; 3 µL (5ng/ µL) of cDNA and 2,5 µL of water.

Primer design was performed by using the Software PrimerQuest Tool (http://www.idtdna.com/Primerquest/Home/Index). The primers used were designed in the 5'-3' end of the transcript. The Tm of the primers was between 54 and 58ºC; amplicon of 75 - 100 base pairs (Table 6). Primer efficiency was calculated, after the RT-qPCR, through the LinRegPCR program (RUIJTER et al., 2013).

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 specify 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 of Illumina. Ct values obtained from the Eco Real-Time software (Illumina) were analyzed in an excel sheet by the ΔΔCt method (LIVAK and SCHMITTGEN, 2001). RT-qPCR data of each gene of interest were normalized against the average of the three reference genes used. The gene expression data obtained was analyzed using the REST® program that performs the comparative quantification method of the "Pair-Wise Fixed Reallocation Randomization Test"

(PFAFFL et al., 2002). Correlation between the expression values detected by RNA-Seq and qRT-PCR for the 5 genes tested in both stages, R7.2 and R9, with 0h and 18h of imbibition were analyzed by Pearson´s correlation.

3.12. *Medicago truncatula*: RNA extraction, affymetrix microarray, data extraction and normalization

For transcriptomic analysis of *Medicago truncatula* three replicates of 50 immature seeds fresh, immature seeds dried and mature seeds, before and after 6h and 24 of imbibition were rapidly frozen in liquid nitrogen and stored at -80°C prior to RNA isolation. Total RNA extraction was performed according to the NucleoSpin® RNA Plant kit (Macherey-Nagel); RNA integrity was assessed by analysis on a 1% agarose gel; quantified and evaluated for purity using the Bioanalyzer 2100 (Agilent).

Nimblegen Microarray, data extraction, normalization and transcriptomic data analysis of seed development during imbibition were performed according to the methodology described in Verdier et al (2013).

For the gene ontology enrichment analysis using PlantGSEA (http://structuralbiology.cau.edu.cn/PlantGSEA; YI et al., 2013), using *Arabidopsis thaliana* genome annotation (TAIR 10) as genome model and data was corrected for multiple testing error (FDR). For both database the significant level of 0.05 (p-value≤0.05) was used for filtering.

Table 6. Primers sequences used as target and reference genes (mRNAs) used in RT-qPCR reactions.

Target genes		**Primers**
Class I KNOX homeobox transcription factor *Glyma.*14G047000	Forward	CAACCTAGATTCTCAAGCTGAA
	Reverse	CTTCTTGAGGCTCCCTAGATA
WRKY transcription factor *Glyma.*03G224700-	Forward	CTTTGGCAGCAGCAATCT
	Reverse	TGCGGTGAACCTGGAATA
ABI3L domain class transcription factor *Glyma.*02G200500	Forward	GCCATACCATCACCAACAA
	Reverse	CGAACTCGAACTAGAACTGC
Heat stress transcription factor A3 *Glyma.*10G066100	Forward	AGCCATGTCTTCTGAAACTC
	Reverse	AAGGTCTTCTGCACCTACTA
Auxin response factor Glyma.14G217700	Forward	CAGGTTCTGTAGGAAGATCAAT
	Reverse	GTGTCATTCAGCAGACCAT
Primer set		**Reference genes**
Proteasome subunit beta type *Glyma.*06G078500	Forward	TCACACAGCTCACTGATAATG
	Reverse	GAAGGAAGTAGCGCACATAG
	Frag. Length	90
40S ribosomal protein S20-2 *Glyma.*03G142300	Forward	CTGTGGTGAAGGTACCAA
	Reverse	CTCCACACCAGGTTCAAT
	Frag.Length	127
Polyadenylate-binding protein *Glyma.*14G084500	Forward	GACGCAGACGGTAAATCAA
	Reverse	TTCCAACATACCACTCCTTATC
	Frag. Length	121
Importin beta-2 subunit family protein *Glyma.*20G106300	Forward	GATAATAAGCGGGTCCAAGAG
	Reverse	GTCATCAGGTGCTTCAGTATAA
	Frag. Length	110

4. RESULTS

4.1. CHARACTERIZATION OF ACQUISITION OF SEED QUALITY AND ABSCISIC ACID QUANTIFICATION DURING SOYBEAN SEED DEVELOPMENT.

Physiological quality seed characteristics include the ability to germinate, tolerate desiccation and remain viable during storage. Ideally, the seed harvest should occur when all the characteristics reach their maximum levels. However, there is no consensus during the maturation process of soybean seeds about when this occurs. The technology for producing soybean seed recommends that the better moment for harvest it is the physiological maturity point (FRANÇA NETO, 1984). Believed that the characteristics of seed quality are acquired in this point and quickly then lost (TEKRONY et al. 1979; OBENDORF et al. 1980; FRANÇA NETO et al. 2007; REN et al. 2009). However, there is increasing evidence, that these characteristics are acquired sequentially after the physiological maturity when the seeds lose water and are still in the mother plant (ELLIS et al. 1987; ELLIS et al. 1993; DEBEAUJON, 2000; PROBERT, 2007; CHATELAIN et al. 2012).

As there are still a lack of consensus in the literature about the acquisition of soybean seed quality during seed development. Thus, we analyzed the development phases considering the phenological scale and days after pollination, observing the morphology characteristics of the pods and seeds. We also analyzed the change in water content, fresh and dry weight during seed development. In addition, we

performed physiological assessment (germination, desiccation tolerance, vigor and longevity) and ABA quantification.

4.1.1 Characterization of seed maturation stage

Soybean seed maturation was characterized using phenological stages presented in Figure 6, and by monitoring changes in seed weight and water content. Seed filling started at stage R5.2. Fresh and dry weights increased steadily onwards until stage R7.2. Thereafter, dry weight was stable, indicating the end of the filling phase.

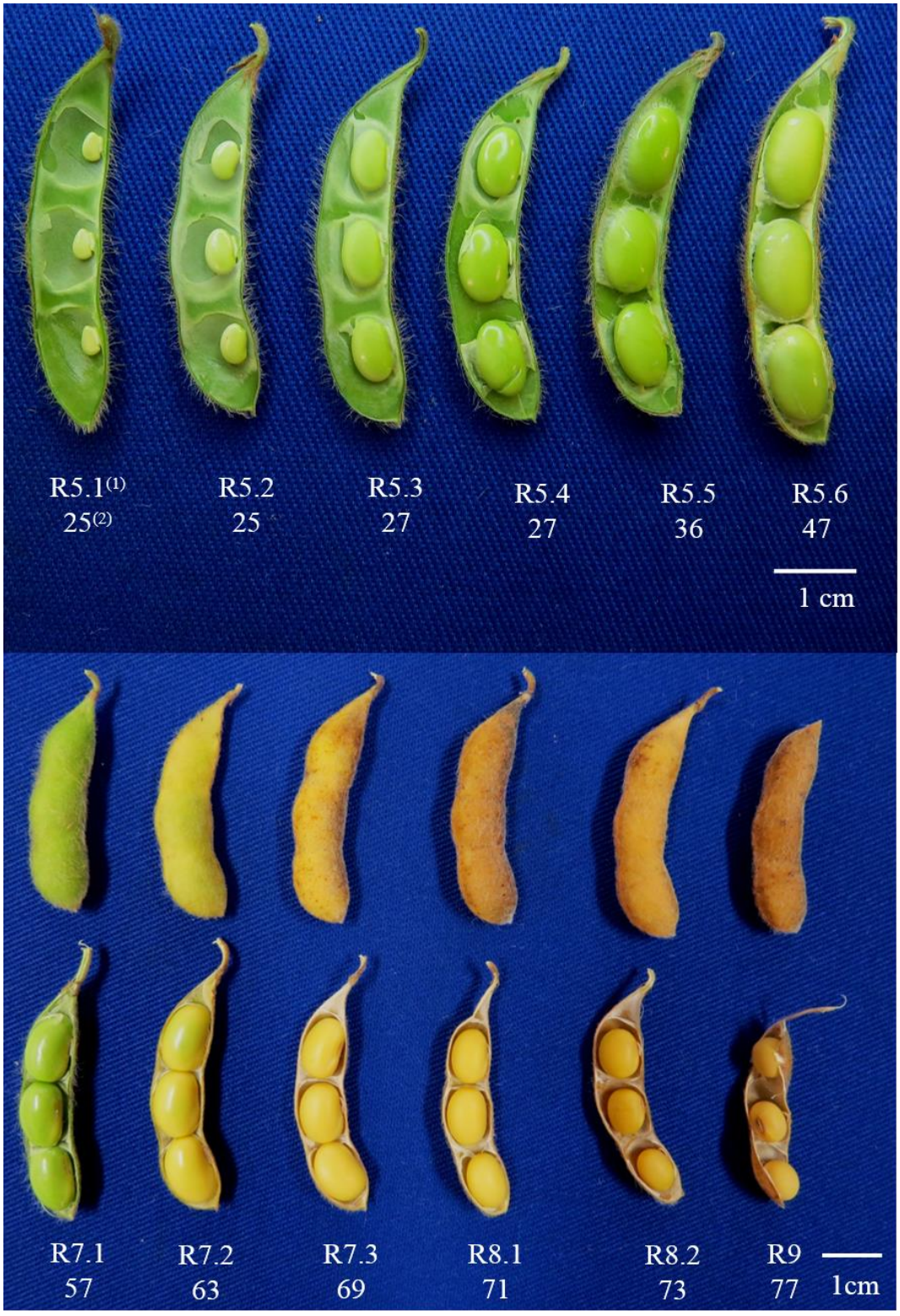

Figure 6. Development of soybean pods and seeds. The characterization of the reproductive stages based on phenological scale (FEHR and CAVINESS, 1977; RITCHIE et al., 1982). (1)Phenological scale; (2)Days after flowering.

Between R7.1 and 7.2, pod and developing seeds turned yellow as a result of chlorophyll degradation. However, a central part of the cotyledons remained green. Water content varied during the reproductive stages (Figure 7). Between stages R5.3 and R7.2, water content decreased steadily due to the replacement of the water by DW. At stage R7.2, the seed water content was 1,32 ± 0,02 g H_2O/g DW. Thereafter, it decreased steadily until stage 9, corresponding to the post-abscission stage where water content was 0,12. Therefore, the maturation drying was considered to be from stage R7.2 to R9 since DW did not change significantly.

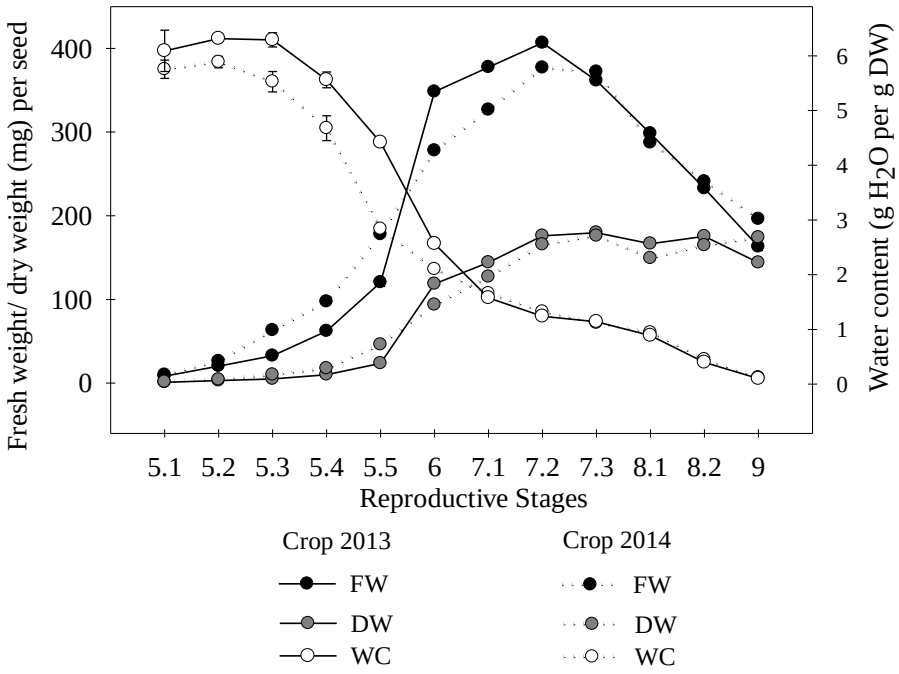

Figure 7. Characterization of the soybean seed development. Changes in water content (WC, g H_2O/ g DW), in fresh weight (mg/seed, FW) and dry weight (DW, mg/seed) and water content (g H_2O per g DW) at indicated reproductive stages. Data are the average of four replications using 20 seeds. Error bars indicate standard errors.

Next, we determined changes in ABA contents during development. ABA levels increased from stage R5.1 until R6, corresponding to a maximum (Figure 8). Thereafter, it declined progressively. At stage 7.2, the amount was still about 372,42 ± 21,85 (ng per seed) whereas it became undetectable at stage 7.3.

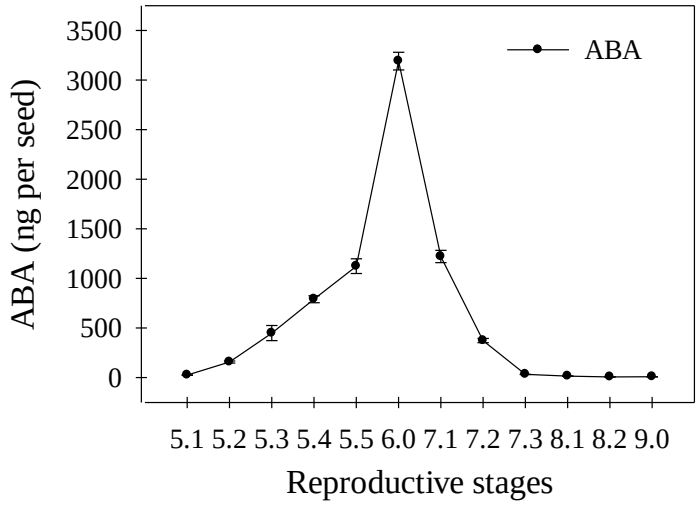

Figure 8. Changes in abscisic acid content during seed maturation at indicated reproductive stages. Data are expressed a seed basis and correspond to the average of three replications using 150 mg of embryos. Error bars indicate standard errors.

4.1.2 Acquisition of physiological quality during seed maturation

The capacity to germinate was progressively acquired during early seed filling, between stages R5.5 and R6 (Figure 9). Desiccation tolerance (DT), defined as the ability to germinate after rapid drying to 0,10 g water/g DW was acquired between stage R6 and stage R7.2. Acquisition of normal seedling establishment occurred also between stages R7.1 and R7.2, concomitantly with DT (Figure 9 b). No effect of drying was observed, indicating that physiological maturity was obtained at R7.2.

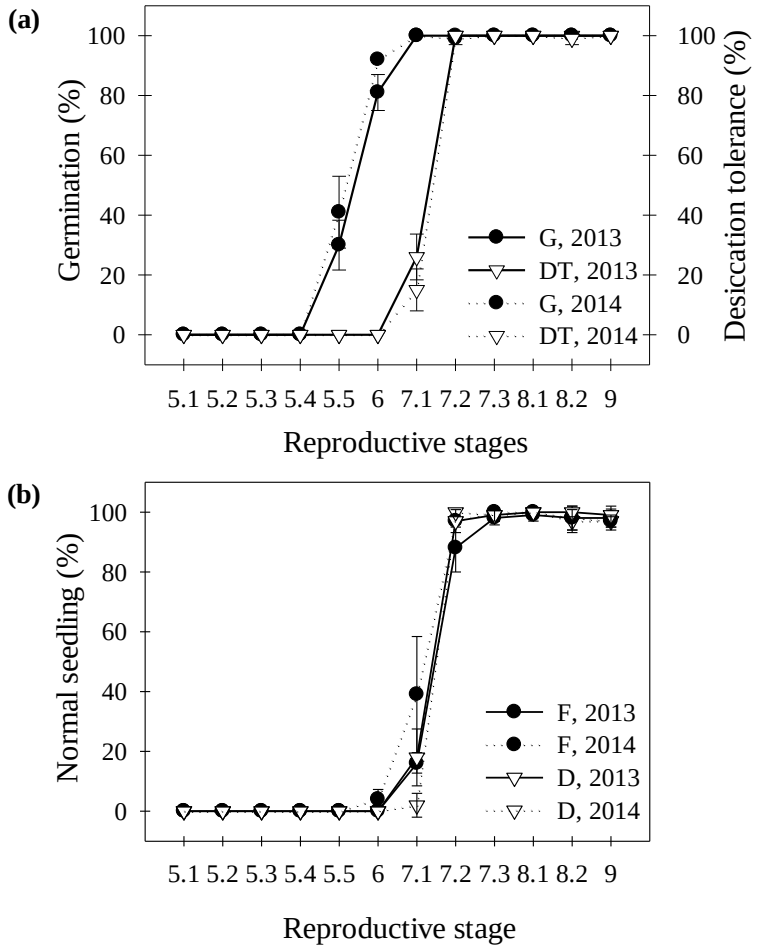

Figure 9. Acquisition of physiological quality during development. **(a)** Changes in percentage of germination (G) and desiccation tolerance (DT) of seeds harvested at different reproductive stages. (**b)** Changes in percentage of normal seedlings of seeds harvested at different reproductive stages before (F) and after an enforced rapid drying. Data points are average of four replications of 25 seeds obtained from two crop year (2013 and 2014).

To further assess physiological maturity, we tested when germination in stressful conditions was acquired as a proxy for a vigor test. Two stressful conditions were applied: salt stress using 100 mM NaCl and cold (15°C). Germination and seedling emergence of seeds from stages R7.2 and R9 were tested, since these represent the extremes of maturity (Figure 10a). There was no difference in the germination speed or final percentage of germination and seedling emergence between stages R7.2 and R9 in these different conditions, suggesting that seed vigor is already acquired at stage R7.2. Next, we tested whether there existed difference in hypocotyl elongation between mature

and immature stages. For this purpose, hypocotyl length and dry weigh were measured after 48h of imbibition in optimal conditions. Figure 10 (b) shows that no significant difference in both parameters was found between the different stages, thereby confirming that the capacity to establish a normal seedling is fully acquired at R7.2 in our laboratory conditions.

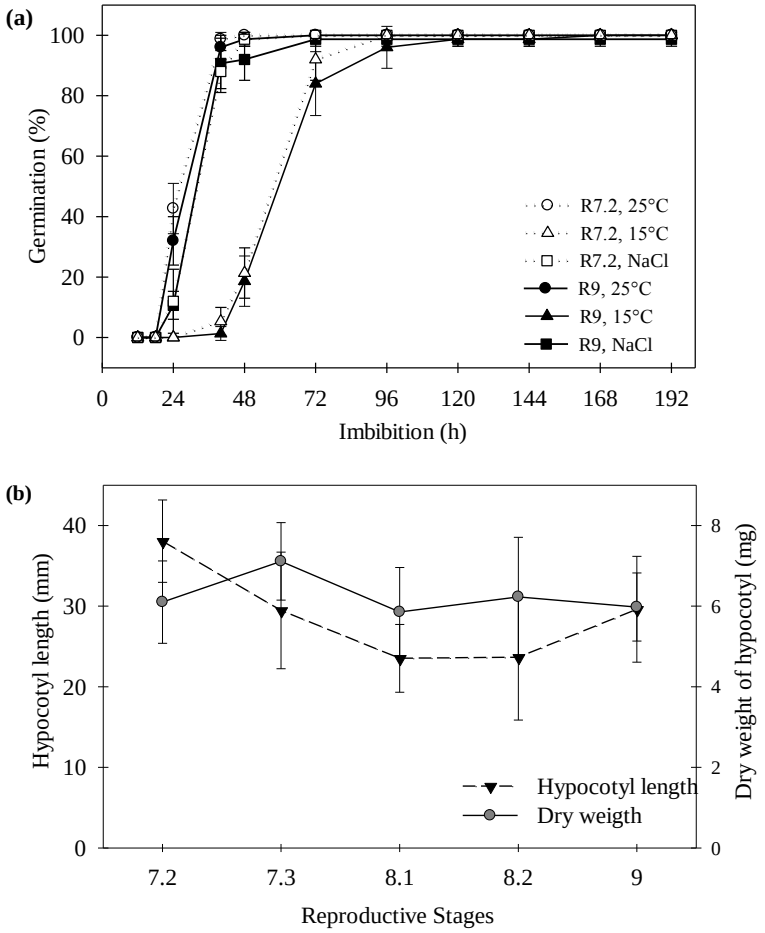

Figure 10. Seed vigor is already acquired at stage R7.2. (a) Germination of immature (stage R7.2) dried and mature (stage R9) seeds in optimal (25°C) and stressful conditions (15°C or 100mM NaCl). Data points are average of four replicates of 25 seeds; error bars indicate standard deviation. (b) Hypocotyl length and dry weight of 48 h imbibed seedlings obtained at indicated reproductive stages. Data are the average of 10 hypocotyls. Error bars indicate standard deviation.

4.1.3 Characterization of longevity in immature (stage R7.2) and mature (stage R9) seeds

During soybean seed maturation no physiological differences regarding germination were observed between immature and mature seeds. However, developing seeds need also to acquire longevity (i.e. the capacity to survive in the dry state for extended periods of time during storage, CHATELAIN et al., 2012). Here longevity was tested for seeds harvested and dried at stage R7.2 and for mature seeds R9. Storage conditions were 75% RH and 35°C which is supposed to induce a moderate ageing (CHATELAIN et al., 2012). During storage, seeds of stage 7.2 started to lose their viability after 21 days of storage (Figure 11). Half the population was dead after 27 days as calculated by the fitting curve. In contrast, it took 48 days to kill 50% of mature seeds.

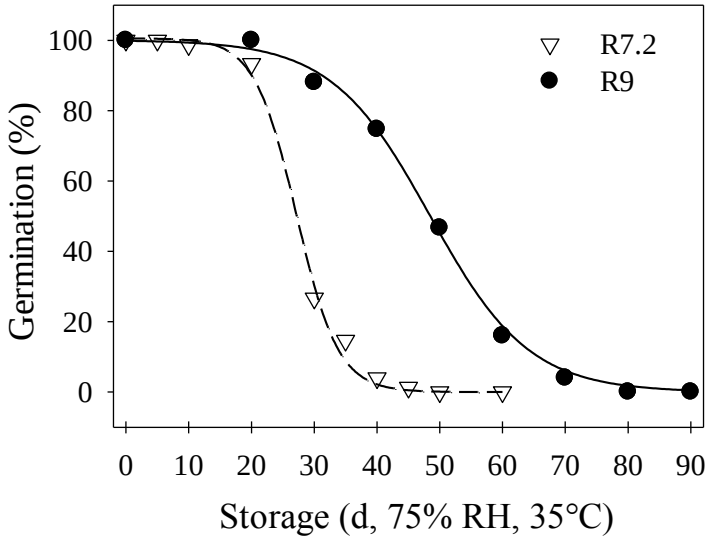

Figure 11. Changes in germination percentages of seeds at stages R7.2 and R9 during storage at 35°C and 75% relative humidity (RH). Third-order sigmoid curves were fitted to the data to obtain the P_{50} values, corresponding to the time of storage (days) at which the dried seed lot reached 50% germination.

4.2. TRANSCRIPTOME ANALYSIS DURING SOYBEAN SEED MATURATION

Our physiological characterization shows that there is no phenotypic difference in germination, seedling emergence and seed vigor between stages R7.2 and R9, suggesting that physiological maturity in terms of seed emergence was reached at R7.2. Nevertheless, the developing seeds need to mature further to acquire full longevity and therefore its full agronomic potential. Seeds from stage R7.2 contains 1,32 ± 0,02 g H_2O/g DW. Thus the progress in seed maturation occurs during maturation drying. It has been inferred that desiccation promotes the transition from developmental mode to a germination-oriented program (BEWLEY, 1997). All of the physiological experiments were performed using immature seeds that were submitted to an artificial drying treatment and compared with R9. Thus we can hypothesize that during maturation drying, the seeds will accumulate all mRNAs that will be stored in the dry quiescent state and used during imbibition to ensure germination (RAJJOU, 2004). However, it is not known whether an artificial drying treatment of immature seeds at R7.2 that lasts for 2-3 days would induce a similar or entirely different profile of stored RNA compared to maturation drying that lasts around 14 days. Such information would be useful to have an idea of those mRNA that are necessary for germination and longevity. It will also give some insights into whether artificial drying is identical to a maturation drying. For an agricultural point of view, such information could be useful to provide putative molecular indicators to assess whether post-harvest drying during seed processing was sufficient to replace natural maturation drying.

To assess changes in transcriptome, a high throughput sequencing on total RNA was performed on two reproductive stages of soybean seeds, immature R7.2, fresh and dried (enforced fast drying), and stage R9 (in planta drying), with two biological replicates per stage. Illumina RNAseq produced between 14 and 66 million reads per library, all libraries having close to or above 90% mapped reads (Table 7). Mapped reads were aligned to 88.647 soybean gene models determined by the Soybean Genome Project available on Phytozome, using the program Bowtie 2 and R packages.

Table 7. Mapping of single-end reads of developing seeds 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. Data of two replicates (rep1 and rep2), in freshly harvested seeds (F) or rapidly dried seeds (d), at 0h and 18h of imbibition, for seeds harvested at the stage R7.2 and R9.

Stage/Replicates	# reads	# Mapped reads	%Mapped reads	#Unmapped reads	%Unmapped reads
R7.2 F rep 1	38 088 839	34 971 702	91,8	3 117 137	8,2
R7.2 F rep 2	34 381 008	31 799 644	92,5	2 581 364	7,5
R7.2 d 0h rep 1	14 259 082	12 881 662	90,3	1 377 420	9,7
R7.2 d 0h rep 2	66 764 127	60 626 937	90,8	6 137 190	9,2
R7.2 d 18h rep 1	23 345 963	21 399 094	91,7	1 946 869	8,3
R7.2 d 18h rep 2	28 353 184	26 028 156	91,8	2 325 028	8,2
R9 0h rep 1	28 824 583	26 457 357	91,8	2 367 226	8,2
R9 0h rep 2	23 064 604	21 097 336	91,5	1 967 268	8,5
R9 18h rep 1	25 774 186	23 183 699	89,9	2 590 487	10,1
R9 18h rep 2	33 725 757	30 287 347	89,8	3 438 410	10,2

Annotation was performed using the Soybase database (http://www.soybase.org/genomeannotation/) and lead to the identification of 56.044 transcripts. After annotation, differentially expressed genes (DEG) were considered significant when the BH value was < 0.05.

4.2.1 Transcripts associated with seed longevity

Considering that seed longevity doubled in dried seeds between R7.2 and R9 stages, we first considered DEG genes associated with this process. Table 8 shows there were only 742 DEG transcript being up-regulated at least 2-fold whereas 1525 were down-regulated.

Table 8. Number of differential transcripts between immature dried seeds ranked according to the level of expression calculated on a Log_2 ratio.

	# Log_2(ratio) transcripts							
Comparison	> 1	> 2	> 3	> 4	< -1	< - 2	< - 3	< - 4
R9 /R7.2d	742	60	9	1	1525	272	20	0

Next, a GO enrichment analysis of the top 500 DEG was performed. Results are shown in Table 9 and revealed a significant over-representation of terms associated with the synthesis of HSP (response to heat, protein folding, response to ER stress; Table 10).

Table 9. Gene Ontology (GO) enrichment analysis of transcripts corresponding to DEG up-regulated between R7.2d and R9. Analysis was performed on the top 500 genes using SoyBase, with annotations for *Wm*82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	#Expressed	*p*-value	GO_ description
GO:0009408	659	51	2,57E-18	response to heat
GO:0006457	779	49	6,79E-14	protein folding
GO:0042542	511	42	8,43E-16	response to hydrogen peroxide
GO:0009644	582	41	4,61E-13	response to high light intensity
GO:0034976	483	30	7,19E-08	response to endoplasmic reticulum stress
GO:0009220	315	22	2,49E-06	pyrimidine ribonucleotide biosynthetic process
GO:0006334	128	14	6,22E-06	nucleosome assembly
GO:0006626	235	14	9,85E-03	protein targeting to mitochondrion
GO:0031516	6	5	5,70E-07	far-red light photoreceptor activity
GO:0009883	6	5	5,70E-07	red or far-red light photoreceptor activity
GO:0008020	8	5	5,20E-06	G-protein coupled photoreceptor activity
GO:0010161	23	5	1,28E-02	red light signaling pathway
GO:0004673	41	5	4,90E-02	protein histidine kinase activity
GO:0010203	4	3	8,42E-03	response to very low fluence red light stimulus
GO:0009881	6	3	1,14E-02	photoreceptor activity

A high number of categories associated with response to light were also over-represented such as “response to high light intensity”, “far-red light photoreceptor activity”, “response to very low fluence”, “red light stimulus”, “photoreceptor activity” (Table 10). A closer inspection of genes belonging to these categories revealed the presence of phytochrome A and PIF homologues (PIF7, Photoreceptor interacting factors), suggesting a role for red light as an environmental cue during seed maturation. Thus we hypothesized that light might play a role in germination.

To test this, we performed imbibed seeds in complete darkness and in the light at 30°C. The speed and percentage of germination were monitored (Figure 12) together with the root and hypocotyl length after during imbibition (Table 10). Regardless of the light conditions, both immature and mature seeds germinated at the same time and to the same level. No significant differences in the root and hypocotyl lengths were observed between R7.2 and R9, suggesting that the skotomorphogenesis and de-etiolation processes were similar between stage R7.2 and R9 (Table 10).

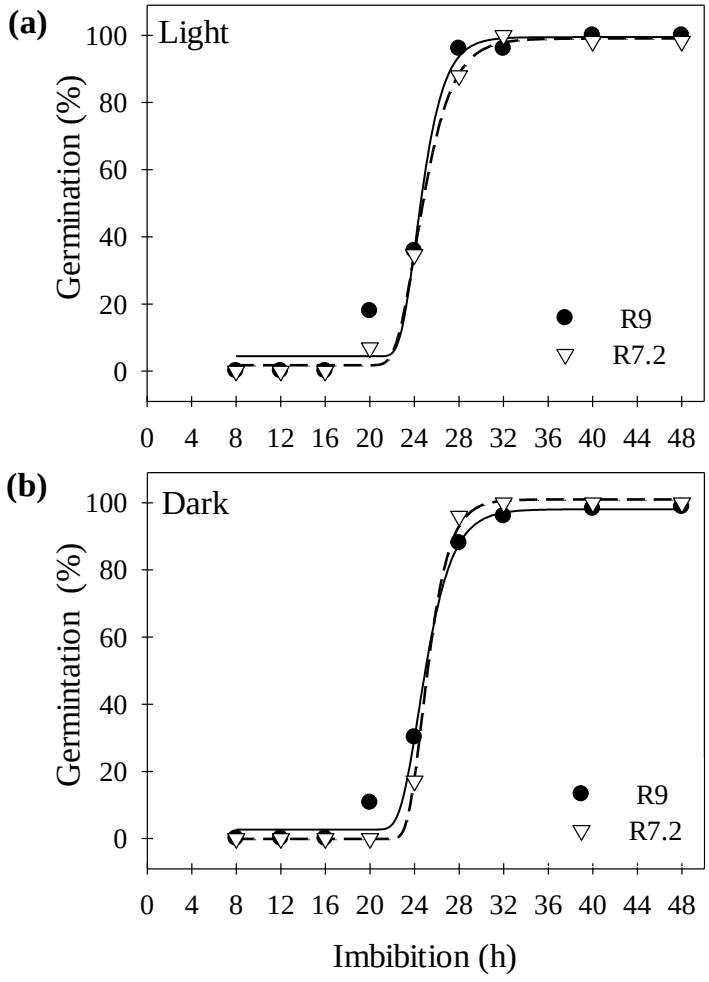

Figure 12. Germination of seeds harvested at the indicated reproductive stages (R7.2 and R9) exposed to be continued light (200 μmol m^{-2} s^{-2}) or dark conditions. Data points are average of two replications of 28 seeds.

Table 10. Root and hypocotyl length measured after 120 h of imbibition of immature dried seeds (stage R7.2) and mature (stage R9) in to be continued light (200μmol m^{-2} s^{-2}) or dark conditions. Data are the average ± SE of 25 seedlings.

Stages of maturation		
Light		
Organ	R7.2	R9
Root	15,78 ± 0,54	15,90 ± 0,67
Hypocotyl	5,90 ± 0,25	6,13 ± 0,22
Dark		
	R7.2	R9
Root	15,06 ± 0,45	15,30 ± 0,77
Hypocotyl	9,37 ± 0,49	9,35 ± 0,63

The main GO categories that were over-represented in the top 500 DEG with transcript levels that decreased between stage R7.2 and stage R9 were associated with photosynthesis (Table 11, Supplementaty Table 2) suggesting that photosynthesis would be detrimental for longevity. Some transcripts corresponded to homologues encoding chloroplastic genes such as proteins D1 (psbA), D2 (psbD), A1 (psaA) and A2 (psaB), subunits of photosystem II and I (TEIXEIRA et al., 2016).

Table 11. Gene Ontology (GO) enrichment analysis of transcripts corresponding to DEG down-regulated between R7.2d and R9. Analysis was performed on the top 500 genes using SoyBase, with annotations for *Wm*82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
GO:0015979	452	62	1,93E-39	photosynthesis
GO:0006091	116	51	1,10E-60	generation of precursor metabolites and energy
GO:0006354	225	51	1,29E-43	DNA-dependent transcription, elongation
GO:0019684	320	35	2,09E-18	photosynthesis, light reaction
GO:0010207	372	24	2,46E-07	photosystem II assembly
GO:0045333	53	12	2,94E-09	cellular respiration
GO:0009772	28	11	2,39E-11	photosynthetic electron transport in photosystem II
GO:0015986	58	7	7,09E-03	ATP synthesis coupled proton transport
GO:0009767	39	6	7,15E-03	photosynthetic electron transport chain
GO:0009769	12	4	8,11E-03	photosynthesis, light harvesting in photosystem II
GO:0006176	5	3	1,31E-02	dATP biosynthetic process from ADP

4.2.2 Impact of enforced premature drying on differentially expressed genes

To investigate whether premature drying induced changes in the transcriptome, we first compared the transcriptome response of immature dried seeds (stage R7.2d) with that of mature seeds (stage R9), so-called “natural drying” with that of the fresh immature seeds fresh (stage R7.2f) vs artificially dried seeds (R 7.2d), so called “enforced drying” (stage R9; Table 12). First, we compared the number of DEG ranked according to the level of expression. Here, a large number of DEG was found to be up and down-regulated with approximatively the same intensity (Table 12). Using these values, we calculated from each intensity categories the ratio between the numbers of DEG being up-regulated and that of DEG being down-regulated for both artificial and enforced drying. Results are presented in Figure 13.

Table 12. Number of differential transcripts between artificial drying (R7.2 d/R7.2 f) and natural drying (R9/R7.2f) in soybean seeds. Only transcripts differently expressed with BH-value < 0.05 were considered.

	# Log_2(ratio)							
Comparison	> 1	> 2	> 3	> 4	< -1	< - 2	< - 3	< - 4
Enforced drying (R7.2d/ R7.2f)	4554	1698	569	199	4218	1882	639	198
Natural drying (R9/ R7.2f)	5475	1814	614	188	6919	3644	1459	521

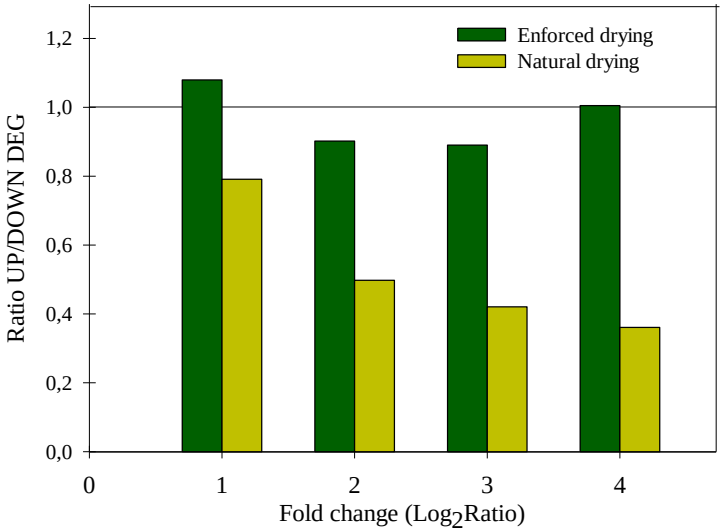

Figure 13. Ratio between the numbers of UP and DOWN differentially expressed genes (DEG) for each intensity categories. Value of 1 means an equal amount of DEG being up or down regulated.

While the ratio were around 1 for immature seeds that were dried artificially, the values were much lower for naturally dried seeds, particularly for those transcripts that were highly abundant (Figure 13). This strongly suggests that RNA metabolism is not identical during drying, with the natural drying favoring a degradation of RNA.

Venn diagrams were constructed for genes that were differentially up-regulated or down-regulated (Figure 14) between enforced drying and natural drying. For the up-regulated DEG, there was more or less an equal distribution between transcripts that were common to natural and artificial drying (39% of both transcriptomes) and those

specific to each drying treatment. A total of 42% of DEG were specific to artificial drying of immature seeds whereas 46% were specific to natural drying. However, for the down-regulated genes, there was a strong overlap between both transcriptome, with only 16 % being specific to artificial drying. Also, there was a large number of DEG that specifically decreased during natural drying representing 57% of the transcriptome. This confirms the idea that there is an increased degradation of transcripts during natural drying.

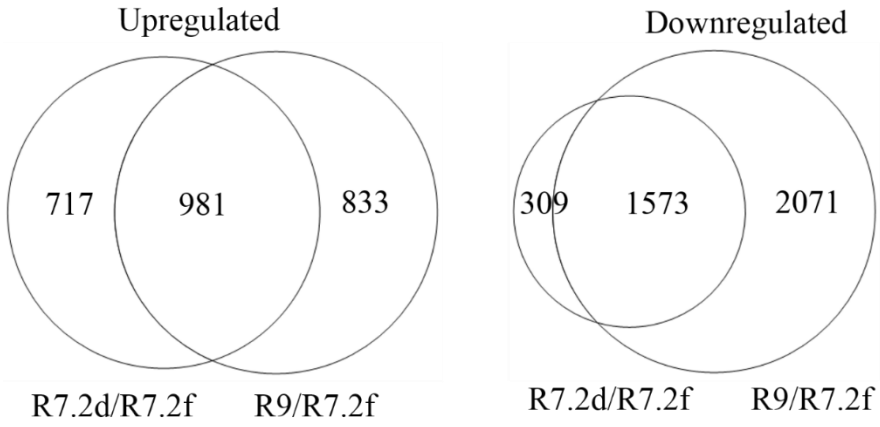

Figure 14. Venn diagrams of DEGs that are up-regulated and down-regulated during enforced drying (R7.2d/R7.2f) and natural drying (R9/R7.2f). Only transcripts with $2<Log_2$ ratio>2 were taken into account.

To gain further knowledge of genes that were specifically up- and down-regulated by natural and artificial drying treatments, a GO enrichment analysis was performed on each category. The transcriptome containing the 717 transcripts whose level was specifically increased during enforced drying of immature seeds were associated a significant over-representation of terms associated with photosynthesis, such as “photosynthesis”, “photosynthesis, light reaction”, “photosynthesis, light harvesting in photosystem II”, “photosynthetic electron transport in photosystem II”, “photosynthetic electron transport chain”, “chlorophyll binding”, “photosystem II assembly”(Table 13).

Table 13. Gene Ontology (GO) enrichment analysis of transcripts DEG up-regulated to artificial drying. Analysis was performed on the top 717 genes using SoyBase, with annotations for Wm82.a2.v1 (Glyma 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
Overrepresented				
GO:0015979	452	80	1,75E-55	photosynthesis
GO:0010207	372	22	7,13E-05	photosystem II assembly
GO:0019684	320	45	3,49E-26	photosynthesis, light reaction
GO:0006354	225	66	5,24E-61	DNA-dependent transcription, elongation
GO:0006091	116	66	6,97E-85	generation of precursor metabolites and energy
GO:0008137	58	15	1,86E-12	NADH dehydrogenase (ubiquinone) activity
GO:0016168	58	11	2,66E-07	chlorophyll binding
GO:0045333	53	13	7,21E-10	cellular respiration
GO:0080044	50	6	2,57E-02	quercetin 7-O-glucosyltransferase activity
GO:0016651	40	11	3,47E-09	oxidoreductase activity, acting on NAD(P)H
GO:0009767	39	10	1,76E-07	photosynthetic electron transport chain
GO:0080046	29	5	1,66E-02	quercetin 4'-O-glucosyltransferase activity
GO:0009772	28	9	1,43E-07	photosynthetic electron transport in photosystem II
GO:0042773	24	8	1,03E-06	ATP synthesis coupled electron transport
GO:0045156	19	7	1,86E-06	electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity
GO:0080045	16	4	1,97E-02	quercetin 3'-O-glucosyltransferase activity
GO:0003959	14	8	1,81E-09	NADPH dehydrogenase activity
GO:0009769	12	7	8,38E-08	photosynthesis, light harvesting in photosystem II
GO:0045157	7	3	2,76E-02	electron transporter, transferring electrons within the noncyclic electron transport pathway of photosynthesis activity
GO:0006176	5	3	1,94E-02	dATP biosynthetic process from ADP
Underrepresented				
GO:0006952	1116	3	2,54E-02	defense response

A closer look at the most abundant transcripts (Supplementary Table 3) in immature dried seeds (R7.2d) revealed the presence of “Pectinesterase”, “Hydroxyproline-rich glycoprotein-like protein”, “Protein PsbN”, “Ring finger protein”, “E3 ubiquitin-protein ligase”, among proteins related to abiotic stress responses. Also, we found transcripts encoding a glutathione ɣ-glutamylcysteinyltransferase-3 and glutatione peroxidase, two enzymes involved in glutathione mediated protection against oxidative damage and maintaining redox homeostasis (MAY et al., 1998; POTTERS et al. 2002).

The transcriptome containing the 833 transcripts whose level was specifically increased during natural maturation drying of immature seeds were associated a significant over-representation of genes involved in protein folding, including DNAj family, chaperones and heat shock proteins (Table 14). Among the mostly up-regulated genes, we noticed a “Heat shock transcription factor A6b” (HSFA6b), “Thioredoxin-like 4A” and “Asparagine synthetase”. Also, it was noted the presence of genes related with proteins degradation and recyclying (Ubiquitin-conjugating enzyme), detoxification of secondary metabolites (MATE efflux family protein) and biotic stress.

Table 14. Gene Ontology (GO) enrichment analysis of transcripts DEG up-regulated to natural drying. Analysis was performed on the top 833 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
GO:0006457	779	31	4,64E-02	protein folding
GO:0009644	582	29	1,05E-03	response to high light intensity
GO:0042542	511	27	7,94E-04	response to hydrogen peroxide
GO:0006334	128	13	4,90E-04	nucleosome assembly

For the 309 genes that were down-regulated in artificially dried immature seeds, the enrichment analysis revealed only 1 GO containing only 3 genes, suggesting that the down-regulation does not affect a particular biological process (Table 15). However, when we looked at the top down-regulated genes, we noticed several transcripts involved in ethylene signaling, including a homologue of ESR1 (ethylene response sensor 1), 2 homologues of EIN3-binding F box protein 1. The presence of several transcription factors (AP2 transcription factor, NAC, MYBR domain class, TIFY) was also noticed.

Table 15. Gene Ontology (GO) enrichment analysis of transcripts DEG down-regulated to artificial drying. Analysis was performed on the 309 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
GO:0030643	7	3	1,26E-02	cellular phosphate ion homeostasis

The 2071 down-regulated genes associated with the natural drying process were associated with cell wall categories ("cellulose synthase activity", "primary cell wall biogenesis" and "cellulose synthase (UDP-forming) activity"), and secondary metabolism ("acetyl-CoA metabolic process"; Table 16).

Table 16. Gene Ontology (GO) enrichment analysis of transcripts DEG down-regulated to natural drying. Analysis was performed on the 2071 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
GO:0016126	363	39	2,09E-03	sterol biosynthetic process
GO:0006084	178	22	3,31E-02	acetyl-CoA metabolic process
GO:0016759	74	13	1,62E-02	cellulose synthase activity
GO:0016760	56	11	2,30E-02	cellulose synthase (UDP-forming) activity
GO:0004579	10	6	9,78E-04	dolichyl-diphosphooligosaccharide-protein glycotransferase activity
GO:0009833	13	6	1,42E-02	primary cell wall biogenesis

4.3. TRANSCRIPTOME COMPARISONS OF IMMATURE SEEDS DRIED (STAGE R7.2d) AND MATURE SEEDS (STAGE R9) DURING IMBIBITION

In the previous section 4.2 our transcriptome data suggest that during the late stage of maturation, seeds are able to synthesize mRNA in preparation for the dry state. It is estimated that 12.000 and 13.500 transcripts are stored in the dry seeds of Arabidopsis (NAKABAYASHI et al., 2005) and our soybean DEG are in the same range. These stored mRNAs are thought to be used during seed imbibition to ensure germination. The proof of this hypothesis comes from the observation that blocking transcription with α-amanitin does not prevent radicle emergence during imbibition in arabidopsis (RAJJOU et al., 2004); indicating that newly synthesized proteins are translated from stored mRNA. However, complex changes in the stored mRNA have been reported during imbibition prior to radicle emergence. Some are rapidly degraded whereas others are maintained to high levels (GALLAND and RAJJOU, 2015). Using ribosome profiling techniques that monitor translation, it was possible to describe which transcripts were engaged into translation in dormant seeds where germination is blocked compared to non dormant seeds (BASBOUS-SERHAL et al., 2015). In non-dormant seeds like soybean, it is not clear which mRNAs are important for germination and which one can be degraded because they are no longer necessary to ensure germination while they were necessary for maturation and remained highly abundant in dry seeds. Molecular biology experiments to address this issue are difficult to perform because measuring the rate of transcript degradation is very complex (BAZIN et al. 2011). Also, the use of α-amanitin should be avoided because it introduces artefacts at the transcriptome levels when used in cell culture. An additional layer of complexity comes from the observation that maturation-specific genes can be expressed after germination upon osmotic treatment that blocks radicle growth and re-induces desiccation tolerance (BUITINK et al., 2006). Therefore, the molecular comparison between stages R7.2 and R9, where no germination difference is observed, represents is an interesting physiological system to study changes in transcript levels that occur during early imbibition before germination of seeds that stored different sets of mRNAs as shown previously.

As explained below, major differences occurred during seed imbibition of R7.2 compared to R9 stages. Therefore, the next question was to assess how

these differences could be attributed to the level of maturity of the seed lots and highlight those biological processes that are (un)necessary for germination. Also, the only phenotypic difference between harvested seeds at stage R7.2 and mature seeds is related to an increased longevity. Thus, we hypothesize that a transcriptome approach comparing changes during early imbibition of two seeds lots differing in longevity could give access to repair mechanisms that would be synthesized during imbibition of seeds from stage R9 or, alternatively to deteriorative mechanisms taking place during imbibition of immature seeds**.** To assess changes in transcriptome, we obtained RNASeq data sets from entire dry seeds from stage R7.2 (enforced fast drying) and stage R9 (in planta drying) before at imbibition (0h) and after 18 hours of imbibition, a time point when seeds were fully hydrated but had not germinated. Below we discuss the results found.

4.3.1 Profiles of differentially expressed genes (DEG) during imbibition

After normalization and background removal, data was filtered to remove transcripts with variance < 1 among stages. From this extended data set, 13.477 transcripts were selected as differentially expressed genes (DEG) between the different stages. Next, a clustering analysis was performed to describe the different profiles of differentially expressed genes during imbibition. Transcripts clustered into 13 patterns of expression that are shown in Figure 15 These patterns suggest three categories of expression: genes associated with germination (letter G), maturation (M) or longevity (L).

1. The pattern G corresponds to genes that up or down regulated during imbibition (profiles A, B) regardless of the maturity of seeds. It also include genes that up- or down regulated after 18 h of imbibition in R7.2 seeds to reach similar levels of expression as in stage R9 (profiles K-N).

2. Maturation associated genes (M) correspond to transcript that are always significantly different between stages R7.2 and R9 regardless of the imbibition time or their profile during imbibition. Transcripts of the M category could be putative markers of the level of maturity reached by the seeds since their level of expression is always higher or lower in stage R9 compared to stage R7.2 during the first 18h of imbibition.

3. Genes were declared to be associated with longevity based on a putative repair function. Thus expression levels L genes could correspond to repair

mechanisms where the genes could act as an activator (profiles G, I) or repressor (H-J). The amount of genes varies greatly between the different profiles. The vast majority of them behaved the same way, with transcripts increasing (profiles A, C) or decreasing (profiles B, D) during imbibition and were associated with germination.

The fact that different profiles can be found during seed imbibition implies a complex mRNA metabolism where transcripts are selectively recognized from immature seeds and being degraded. In the following sections, we will describe the following patterns according the relevance of the objectives of this work and to the amount of transcript. Therefore, we will focus on profiles A, B, G, N, and L.

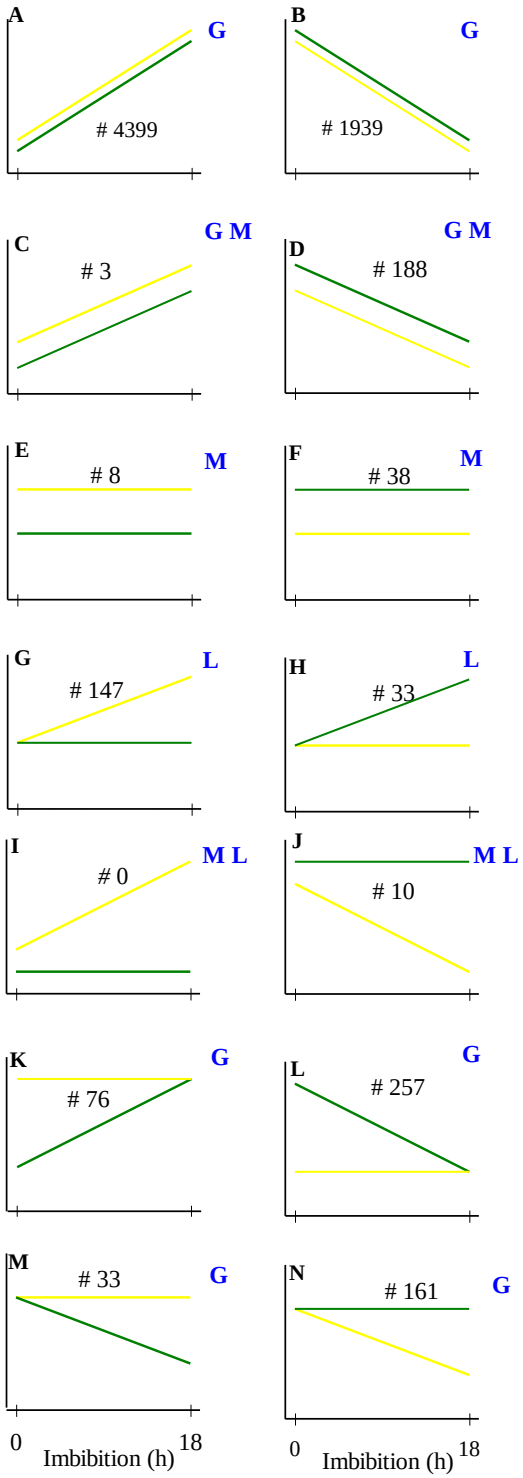

Figure 15. Profiles of expression of differentially expressed genes in dried seeds from R7.2 (green) and R9 stages (yellow), before (0h) and after 18 h of imbibition in optimal conditions. Numbers on the right indicate the amount of transcripts exhibiting the profile and letters (blue) indicate the type of pattern: G, germination associated genes, M maturity genes, L longevity associated genes.

4.3.2 Transcriptome changes during seed imbibition (Profile A and Profile B)

Firstly, we verified whether the pattern of differentially expressed genes (DEG) were similar during imbibition between stage R7.2 and R9 using the expression profile that were linked to germination (Figure 15, A-B). The number of up- and -down DEG after 0 and 18h were counted and ranked in four categories based on their level of expression (log_2 ratio from 1 to 4 for lowly and highly differentially expressed transcripts).

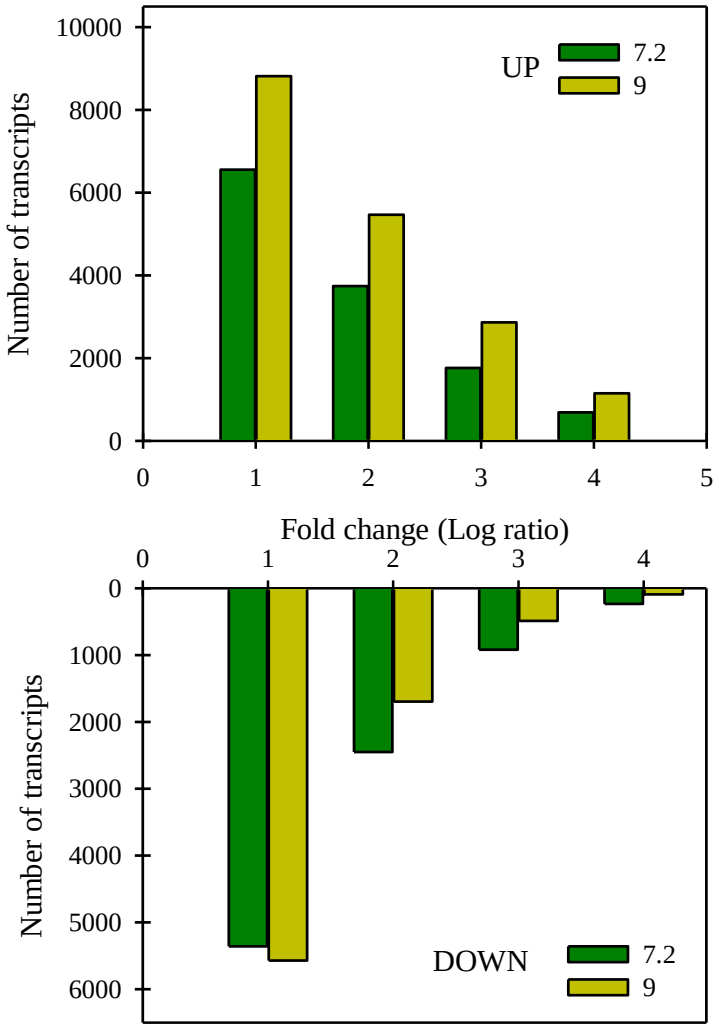

Figure 16. Number of differential transcripts between 0h and 18h of immature (stage R7.2) and mature (stage R9) seeds. Significant differences between transcript levels were assessed using a BH-value < 0.05.

Figure 16 shows that the overall pattern of expressions was similar for both up- and down- genes in both stages, suggesting no gross difference in RNA metabolism between both stages. However, slight differences were observed. We observed a higher amount of differentially up-regulated genes in stage R9, particularly for those that are highly expressed during imbibition. Indeed, compared to R7.2, they were 50% more transcripts that increased 16 fold ($\log_2$ 4) during imbibition of seeds from stage R9. An inverse trend was found for transcripts that decreased during imbibition. Compared to R7.2, they were 40% less transcripts found at 18h of imbibition of stage R9 seeds that had 16 fold-decreased level. This suggests that RNA degradation is likely to be more important in mature seeds. It is noteworthy that these differences do not mean that the expression profile of those genes were radically different during imbibition. Indeed, when we counted the amount of DEG at 18h between stage R7.2 and R9, none were found to be more than two fold ($\log_2$ ratio) differentially up-regulated.

To unravel the molecular processes occurring during early imbibition, we obtained the expression profiles of DEG that increased and decreased during maturation, regardless of their maturity stage. This represented respectively 4399 and 1939 being up and down-regulated transcripts (Figure 16). Next, we performed a GO enrichment analysis on both profiles using the expression data from soybase.org as a background (Table 17). To increase the significance of the analysis, we limited the number of input genes to 1081 transcripts, corresponding to a minimum of 8-fold increase during imbibition. Results are shown in Table 17.

Table 17. Gene Ontology (GO) enrichment analysis of transcripts DEG increased during imbibition. Analysis was performed on the 1081 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	#Expressed	*p*-value	GO description
Over-represented				
GO:0000271	274	35	1,37E-11	polysaccharide biosynthetic process
GO:0010075	450	41	7,04E-09	regulation of meristem growth
GO:0006084	178	24	5,31E-08	acetyl-CoA metabolic process
GO:0010817	212	26	7,46E-08	regulation of hormone levels
GO:0048767	504	41	2,28E-07	root hair elongation
GO:0016126	363	33	7,51E-07	sterol biosynthetic process
GO:0004553	546	42	2,26E-06	hydrolase activity, hydrolyzing O-glycosyl compounds
GO:0009932	248	26	2,26E-06	cell tip growth
GO:0009825	265	26	9,06E-06	multidimensional cell growth
GO:0043481	310	27	5,79E-05	anthocyanin accumulation in tissues in response to UV light
GO:0042335	131	15	1,58E-03	cuticle development
GO:0005975	911	49	2,62E-03	carbohydrate metabolic process
GO:0000038	104	13	2,81E-03	very long-chain fatty acid metabolic process
GO:0071555	448	30	3,01E-03	cell wall organization
GO:0051740	14	5	1,14E-02	ethylene binding
Under-represented				
GO:0003735	733	3	1,73E-03	structural constituent of ribosome
GO:0006412	763	3	4,83E-03	Translation

Biological processes that were more over-represented during imbibition were associated with growth of different tissues (root tip, root hair, meristem, cell wall elongation). Also, genes involved in lipid reserve mobilization appears to be activated, as highlighted by the categories “Very long chain fatty acid metabolic process” and “acetyl-CoA metabolic process”. The GO “regulation of hormone levels” did not contain genes whose function are related to a particular hormone but exhibited genes that had overlapping GO annotation with other categories such as “sterol biosynthesis process”, “anthocyanin accumulation”. Interestingly, the 5 genes representing the GO “ethylene binding” corresponded to probes encoding five different ethylene receptors, including an Arabidopsis homologue of ESR1.

GO enrichment analysis of DEGs that were down-regulated reveals a high number of categories associated with myo-inositol and inositol-phosphate (Table 18). In seeds, inositol serves a matrix to store phosphate in the form of phytate. Many genes have been characterized and were shown to be expressed during seed development. The DEG found here represents biosynthetic genes. Thus, during imbibition they represent sets of transcripts that are no longer necessary and being degraded. The GO "protein folding" and "response to ABA" was also highly significant. These categories contained genes involved in ABA metabolism and signaling pathway such as homologues of ATAF1 involved in attenuating ABA signaling, OST1 (OPEN STOMATA 1), a member of ABA activated SNF1-related protein kinases (SnRK2) involved in ABA signaling, ABI5 (ABA INSENSITIVE 5), APF2, a bZIP factor interacting with ABI5, KEG (KEEP ON GOING) a, RING E3 ligase involved in abscisic acid signaling by degrading ABI5, PP2CA (PROTEIN PHOSPHATASE 2CA), a negative regulator of ABA, ABA1 involved in ABA synthesis,. CYP707A2; (+)-abscisic acid 8'-hydroxylase involved in ABA degradation. Interestingly PMIT (PROTEIN-L-ISOASPARTATE METHYLTRANSFERASE 1), an ABA induced protein involved in repairing protein in aged seeds was also present in this category.

A close inspection of the transcriptome also highlighted many homolog genes responsible for a cross talk between light and ABA/GAI signaling such FRS11 (FAR1-related sequence 11), LZF1 (LIGHT-REGULATED ZINC FINGER PROTEIN 1, FRS6 (FAR1-related sequence 6); PIF4 (phytochrome interacting factor 4), PAT1 (phytochrome A signal transduction 1).

The GO category "chlorophyll oxygenase" represents genes involved in chlorophyll degradation. Chlorophyl breakdown occurs during the late phase of seed development at around stage R7.2. Since the transcripts disappeared both in R7.2 and R9 imbibing seeds, this suggests that during maturation, they were not degraded during maturation. Chlorophyll retention in dry mature soybean seeds represents a major agronomic problem that leads to economic losses as green seeds have poor nutritional and physiological quality. They were also transcripts involved in photosynthesis.

Table 18. Gene Ontology (GO) enrichment analysis of transcripts DEG decreased during imbibition. Analysis was performed on the 1939 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
Overrepreseted				
GO:0009737	1254	82	4,17E-03	response to abscisic acid stimulus
GO:0006457	779	60	5,09E-04	protein folding
GO:0009414	888	60	3,25E-02	response to water deprivation
GO:0010264	140	20	4,57E-04	myo-inositol hexakisphosphate biosynthetic process
GO:0003954	40	11	1,02E-04	NADH dehydrogenase activity
GO:0045333	53	10	2,90E-02	cellular respiration
GO:0016857	14	6	3,60E-03	racemase and epimerase activity, acting on carbohydrates and derivatives
GO:0052726	10	5	8,83E-03	inositol-1,3,4-trisphosphate 5-kinase activity
GO:0047325	10	5	8,83E-03	inositol tetrakisphosphate 1-kinase activity
GO:0052725	11	5	1,57E-02	inositol-1,3,4-trisphosphate 6-kinase activity
GO:0010277	12	5	2,60E-02	chlorophyllide a oxygenase [overall] activity
GO:0009268	6	4	2,98E-02	response to pH
Underrepresented				
GO:0016788	372	1	8,25E-03	hydrolase activity, acting on ester bonds

We also looked at the top 30 transcripts that were mostly down-regulated during imbibition to have an idea about those transcripts that were unstable during imbibition. Many of them corresponded to genes having protective functions during stress such as the LEA gene Seed Maturation protein PM39, HSPs, DNAj and other less characterized genes with chaperone functions (Bcl2-associated anthanogen-like protein (Table 19).

Table 19. List of the top 30 transcripts whose level decrease the most during imbibition.

Gmax 2.0 ID	Tax Hit	*p*-value	Name
Glyma.12G237200	Q1SKZ9_MEDTR	4,00E-43	Zinc finger, CCHC-type
Glyma.12G085000	E6ZE62_BETVM	7,00E-13	ATP synthase subunit alpha
Glyma.03G193400	G7L603_MEDTR	0	Xylem serine proteinase
Glyma.01G116600	G7KS96_MEDTR	7,00E-67	Alcohol dehydrogenase-like protein
Glyma.16G184100	D7LKY7_ARALL	1,00E-24	Protease inhibitor/seed storage/lipid transfer protein family protein
Glyma.17G034400	Q2HSK9_MEDTR	7,00E-100	Ubiquitin-conjugating enzyme, E2
Glyma.08G324100	O24101_MEDTR	3,00E-27	MtN5 protein
Glyma.15G242700	B9I8W9_POPTR	1,00E-96	Purine permease family protein
Glyma.08G156600	B4YYB2_THEHA	5,00E-09	ST63-2
Glyma.16G192200	Q9SWB3_SOYBN	1,00E-93	Seed maturation protein PM39 (Fragment)
Glyma.16G192000	Q9SWB3_SOYBN	6,00E-126	Seed maturation protein PM39 (Fragment)
Glyma.13G173200	Q9LK41_ARATH	4,00E-172	Beta-1,3-glucanase
Glyma.16G189900	Q9SWB3_SOYBN	0	Seed maturation protein PM39 (Fragment)
Glyma.16G185900	Q9SWB3_SOYBN	2,00E-147	Seed maturation protein PM39 (Fragment)
Glyma.07G030800	Q9SSQ8-2	7,00E-33	-2 Isoform 2 of 26.5 kDa heat shock protein, mitochondrial
Glyma.16G192700	Q9SWB3_SOYBN	3,00E-81	Seed maturation protein PM39 (Fragment)
Glyma.16G193200	Q9SWB3_SOYBN	4,00E-144	Seed maturation protein PM39 (Fragment)
Glyma.07G043600	G7L932_MEDTR	7,00E-72	17.4 kDa class III heat shock protein
Glyma.16G190100	Q9SWB3_SOYBN	3,00E-96	Seed maturation protein PM39 (Fragment)
Glyma.12G130000	G7JDS4_MEDTR	1,00E-177	Chaperone protein dnaJ
Glyma.18G284900	G7KUT0_MEDTR	3,00E-59	Bcl-2-associated athanogene-like protein
Glyma.09G139600	Q9SWB3_SOYBN	0	Seed maturation protein PM39 (Fragment)
Glyma.U039700	Q9SWB3_SOYBN	4,00E-98	Seed maturation protein PM39 (Fragment)
Glyma.08G071000	G7JCQ8_MEDTR	0	White-brown-complex ABC transporter family
Glyma.08G177500	W0TQL8_ACAMN	0	Dihydropyrimidinase
Glyma.05G092300	ATPAM_PHAVU	3,00E-169	ATP synthase subunit alpha, mitochondrial
Glyma.15G083900	G7ZXJ0_MEDTR	2,00E-90	Transcription factor TGA5
Glyma.17G074500	FH13_ARATH	0	Formin-like protein 13
Glyma.08G056600	U5FRT1_POPTR	2,00E-159	Hydrolase family protein
Glyma.17G074500	FH13_ARATH	0	Formin-like protein 13

4.3.3 Analysis of transcription factors during imbibition

We investigated whether the patterns of transcription factors (TF), whose expression varied during imbibition between immature (stage R7.2) and mature (stage R9). Only transcripts with BH-value < 0.05 and similar profiles for immature and mature seeds were considered. We first calculated the % of enrichment of TF in the different profiles (Table 20). For this, we assumed that there were no specific enrichment in the expression profile A and B, that is DEG up and down regulated during seed imbibition of both stages. TF represented 4.5% of the total amount in both UP and DOWN DEG.

Table 20. Number of transcription factors (TF) presents in different profiles of genes expressed during imbibition, between immature (stage 7.2) and mature (stage 9) and P values after comparing these values with those of profile A and B. NS not significant, NT not tested.

Profiles	# genes	# TF	P Value
C	3	0	NS
D	188	1	NS
E	8	2	NS
F	38	0	NT
G	**147**	**13**	**0.014**
H	33	1	NS
I	0	0	NT
J	10	0	NT
K	76	5	NS
L	257	8	NS
M	33	2	NS
N	161	4	NS

Therefore, we inspected the list of the 13 transcription factors from category G (Table 21) The mostly up-regulated DEG was a Class I KNOX box homeobox factor whose homolog in Arabidopsis is BUM1. This TF is required for shoot apical meristem (SAM) formation during embryogenesis and for SAM function throughout the lifetime of the plant. We also noted the presence of GRAS-type TF, whose homologue is SGR7, (SHOOT GRAVITROPISM 7) and is essential for normal shoot gravitropism. There is also an ABI3-like transcript. In Arabidopsis, this AP2/B3-like transcriptional

factor family is apparently not related to ABA signaling. Expressed in the lateral organ boundary region, it operates in a network of regulatory genes controlling leaf serration.

4.3.4 Down regulated transcripts during seed imbibition of stage R7.2 only (Profile L)

Next, we investigated which genes were down regulated during imbibition of R7.2 genes but were already down regulated in stage R9 before imbibition. With 257 transcripts, this profile (L, Figure 15) represents the second largest one. We reasoned that these represent transcripts are likely to represent these transcripts that were not degraded progressively during enforced drying. Nevertheless, they can decrease in two different cellular contexts: seed drying during maturing and seed imbibition prior to germination. The GO enrichement analysis (Table 22) revealed a few number of significant categories that are closely related. Indeed, both contained mainly chloroplastic genes involved in transcription and elongation with the chloroplasts. These include ribosomal genes, genes encoding proteins functioning in the photosystem1 and electron transport chain. Paradoxically, while chloroplasts will be resynthesized later after germination during the greening process, transcripts are first degraded even if they could be reused later. A closer inspection of the transcriptome did not reveal any transcript that could be conspicuously linked to degradative processes that could occur as a result of damage incurred during enforced drying.

Table 21. Annotation of the transcription factors expressed corresponding to Profile G.

***Gmax* 2.0 ID**	**Tax Hit**	***p*-value**	**Name**	**R9 18h/ R7.2d 18h**	**R9 18h/ R9 0h**
Glyma.14G047000	A5Y4H1_MEDTR	2,00E-161	Class I KNOX homeobox transcription factor	1,09	5,04
Glyma.02G026300	Q0PJI4_SOYBN	0,00E+00	MYB transcription factor MYB138	1,08	4,18
Glyma.08G137900	G7LAC7_MEDTR	0,00E+00	BZIP transcription factor bZIP133	1,32	3,08
Glyma.05G105600	K7NCY5_QUESU	0,00E+00	GRAS family transcription factor	1,05	2,60
Glyma.06G125100	G7J693_MEDTR	2,00E-124	Ethylene responsive transcription factor 1a	2,10	2,46
Glyma.03G224700	G7KUM9_MEDTR	9,00E-155	WRKY transcription factor	1,30	2,44
Glyma.02G200500	D9ZIL2_MALDO	3,00E-71	ABI3L domain class transcription factor	1,08	2,41
Glyma.10G066100	D1M7W9_SOLLC	9,00E-120	HSF Heat stress transcription factor A3	1,97	1,85
Glyma.13G100500	D9ZJ13_MALDO	1,00E-155	HD domain class transcription factor	1,07	1,70
Glyma.05G157000	Q0GPG2_SOYBN	6,00E-115	BZIP transcription factor bZIP111	1,24	1,63
Glyma.13G249800	G7IK44_MEDTR	2,00E-167	Transcription factor bHLH112	1,03	1,59
Glyma.14G217700	I1MC56_SOYBN	0,00E+00	Auxin response factor	1,06	2,92
Glyma.09G052400	Q948W7_PEA	2,00E-123	Zinc-binding protein	1,03	1,43

Table 22. Gene Ontology (GO) enrichment analysis of transcripts DEG that decreased during imbibition of R7.2 genes but were already degraded in stage R9 before imbibition (profile L of Figure 14). Analysis was performed on the 257 genes using SoyBase, with annotations for Wm82.a2.v1 (Glyma 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
GO:0015979	452	14	7,30E-03	photosynthesis
GO:0006091	116	12	6,71E-08	generation of precursor metabolites and energy
GO:0006354	225	12	1,16E-04	DNA-dependent transcription, elongation

4.3.5 Decreased transcript during imbibition of mature seeds only (Profile N)

Next, we examined a third largest category (Figure 15, profile N) representing transcripts that decreased during imbibition of stage R9 seeds but remained high in immature seeds. This category is interesting because shows that degradation mechanisms are distinct between mature and immature seeds. Although this category contained a higher number of transcripts (161), the GO enrichment analysis did not reveal processes that are of significance in our study (Table 23).

Table 23. Gene Ontology (GO) enrichment analysis of transcripts DEG that decreased during imbibition of R9 seeds but remained high in stage R7.2 during imbibition (profile N of figure 14) .Analysis was performed on the 161 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	**#Genome**	**#Expressed**	***p*-value**	**GO_ description**
GO:0006878	17	3	2,67E-02	cellular copper ion homeostasis
GO:0016048	4	2	4,45E-02	detection of temperature stimulus

4.3.6 Increased transcripts during imbibition of mature seeds (Profile G)

Profile G is characterized by transcripts increasing during imbibition of stage R9 seeds but remaining low and steady in stage R7.2 seeds. It could

contain genes that encode repair processes since, according to CHEN et al. (2013) they can be activated even if symptoms of loss of viability are not visible. To check this, we screened the annotation of all 147 probes for a known repair processes and found two probes. First, DNA2, a member of DNA replication helicase family protein (*Glyma*.06G310500) whose Arabidopsis homologue was recently found to be associated with DNA repair The second probe is (*Glyma*.18G194300), the Arabidopsis homolog of a Werner syndrome-like exonuclease whose expression was significantly upregulated in stage 9 during imbibition but not stage R7.2 (Figure 17).

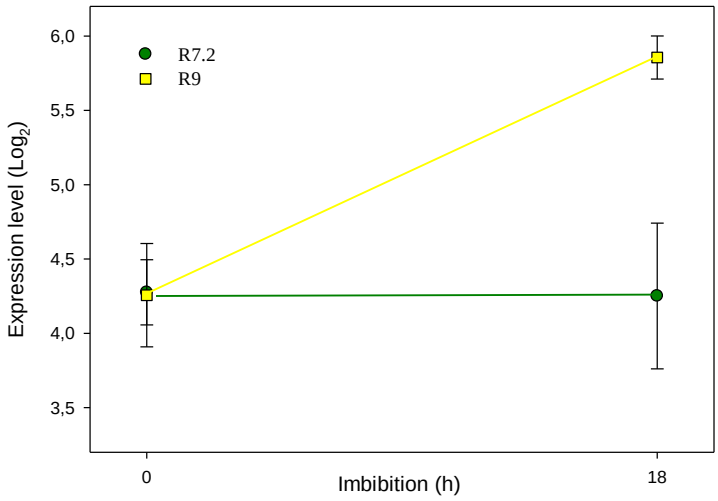

Figure 17. Changes in expression of probe *Glyma*.18G194300 encoding a 3'-5' exonuclease whose closest homologue in Arabidopsis is a Werner syndrome-like exonuclease. Data are expressed as normalized read counts/ millions and are the mean of two replicate samples. Error bars indicate standard errors.

Next, GO enrichment analysis was performed (Table 24). Categories that were over-represented were related to organ formation, in particular in relation to light (red, far-red and blue), and meristem activity, such as flower morphogenesis, organ morphogenesis, cotyledon morphogenesis. Many of these transcripts were related to auxin signaling and auxin.

Table 24. Gene Ontology (GO) enrichment analysis of transcripts DEG that increased during imbibition of R9 seeds but remained low in stage R7.2 during imbibition (profile G of figure 14). Analysis was performed on the 147 genes using SoyBase, with annotations for Wm82.a2.v1 (*Glyma* 2.0). #Genome: Genome Count present in this GO category; # Expressed: Number of submitted genes present in this GO category.

GO_id	#Genome	# Expressed	*p*-value	GO_ description
GO:0048519	139	12	4,59E-07	negative regulation of biological process
GO:0048439	191	13	1,74E-06	flower morphogenesis
GO:0010051	193	12	1,85E-05	xylem and phloem pattern formation
GO:0009887	289	14	3,30E-05	organ morphogenesis
GO:0010358	19	5	1,13E-04	leaf shaping
GO:0009855	326	14	1,42E-04	determination of bilateral symmetry
GO:0009944	203	10	1,96E-03	polarity specification of adaxial/abaxial axis
GO:0009886	79	6	1,41E-02	post-embryonic morphogenesis
GO:0048507	119	7	1,67E-02	meristem development
GO:0048826	12	3	3,72E-02	cotyledon morphogenesis

4.4. VALIDATION OF THE RNAseq DATA

As well discussed in this work, the only physiological difference found between stage R7.2 and R9 was in relation to seed longevity, wherein R9 is increased. Thus, study the changes in transcript levels, during imbibition of the seed these stage, sparked further more our interest The validation of the RNAseq data was performed by using RT-qPCR. Therefore, five interesting genes belonging to the profile G; profile characterized by increasing transcripts during imbibition at stage of R9, but remaining low and steady at stage R7.2 (Figure 18). The expression level (Log_2 fold change) obtained by of RNA-Seq analysis are shows in the Table 25.

Table 25. Relative expression by RNA seq (Log_2-Fold change) transcription factors of selected for validation.

Id *Gmax*	Name (Soybase)	Log_2 R7.2 0h	Log_2 R7.2 18h	Log_2 R9 0h	Log_2 R9 18h
Glyma.14G047000	Class I KNOX homeobox transcription factor	2,41	3,53	1,80	6,84
Glyma.03G224700	WRKY transcription factor	4,24	4,34	4,26	6,70
Glyma.10G066100	HSF Heat stress transcription factor A3	6,06	6,70	7,58	9,44
Glyma.14G217700	Auxin response factor	2,52	3,74	3,16	6,08
Glyma.02G200500	ABI3L domain class transcription factor	5,57	5,93	5,51	7,91

The correlation between expression value by RNAseq and quantitative real-time PCR (RT-qPCR) is present in Figure 18.

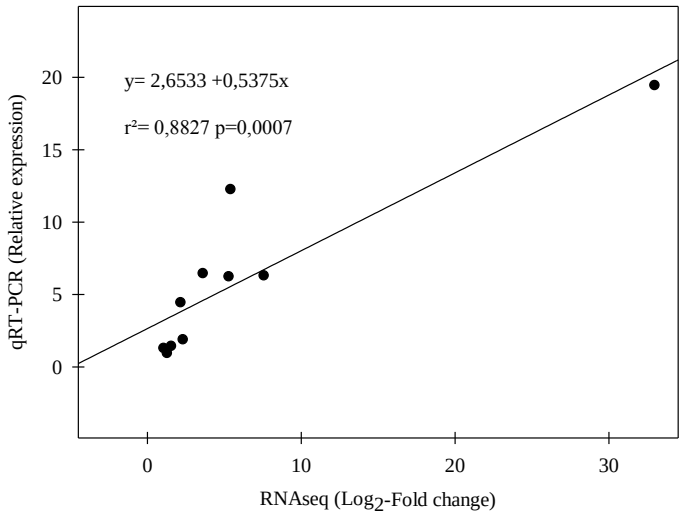

Figure 18. Correlation between RNA*seq* and quantitative real-time PCR (RT-qPCR). Comparison of Log_2 fold change of 5 genes obtained by RNA-seq and RT-qPCR. Real-time PCR was performed using three biological repetitions in duplicate cDNA from each RNA-seq sample. Statistically significant Pearson correlation is shown between the expression levels measured using real-time PCR and RNA-seq.

We validated the expression of Class I KNOX homeobox transcription factor. These transcripts are important regulators in maintaining the shoot apical meristem and patterning adequate of organ initiation (HAKE et al., 2004). KNOX I

was strongly expresso by RT-qPCR for mature seed was observed increased 18.2 fold change during imbibition (Figure 19 a). WRKY transcription factor it was also validated. This family is one of the largest families of transcriptional regulators in plants, some of them have been related to response to biotic and abiotic stress (RUSHTON et al 2010) and responsible for regulated phosphate homeostasis in Arabidopsis seeds (SU et al., 2015). Recently the WRK3 was related to seed longevity of *M. truncatula* (RIGHETTI et al 2015), with then an important transcription factor found in our date.

Longevity genes in *M. truncatula* were found to be enriched in binding sites for auxin-binding factors (RIGHETTI et al. 2015), implying that auxin also may have a role in seed longevity. Consistent with these observations, we found Auxin Responsive Factor family being up-regulate in our data by RNAseq analysis. So, we analysis its expression by RT-qPCR and confirmed its expression, with 5 fold change more for stage R9 during imbibition (Figure 19 d).

Our transcriptome revealed also the presence of HSF, a family of transcription factors that are considered important for longevity seed (PRIETO-DAPENA et al., 2006; ALMOGUERA et al., 2009; PERSONAT et al., 2014).

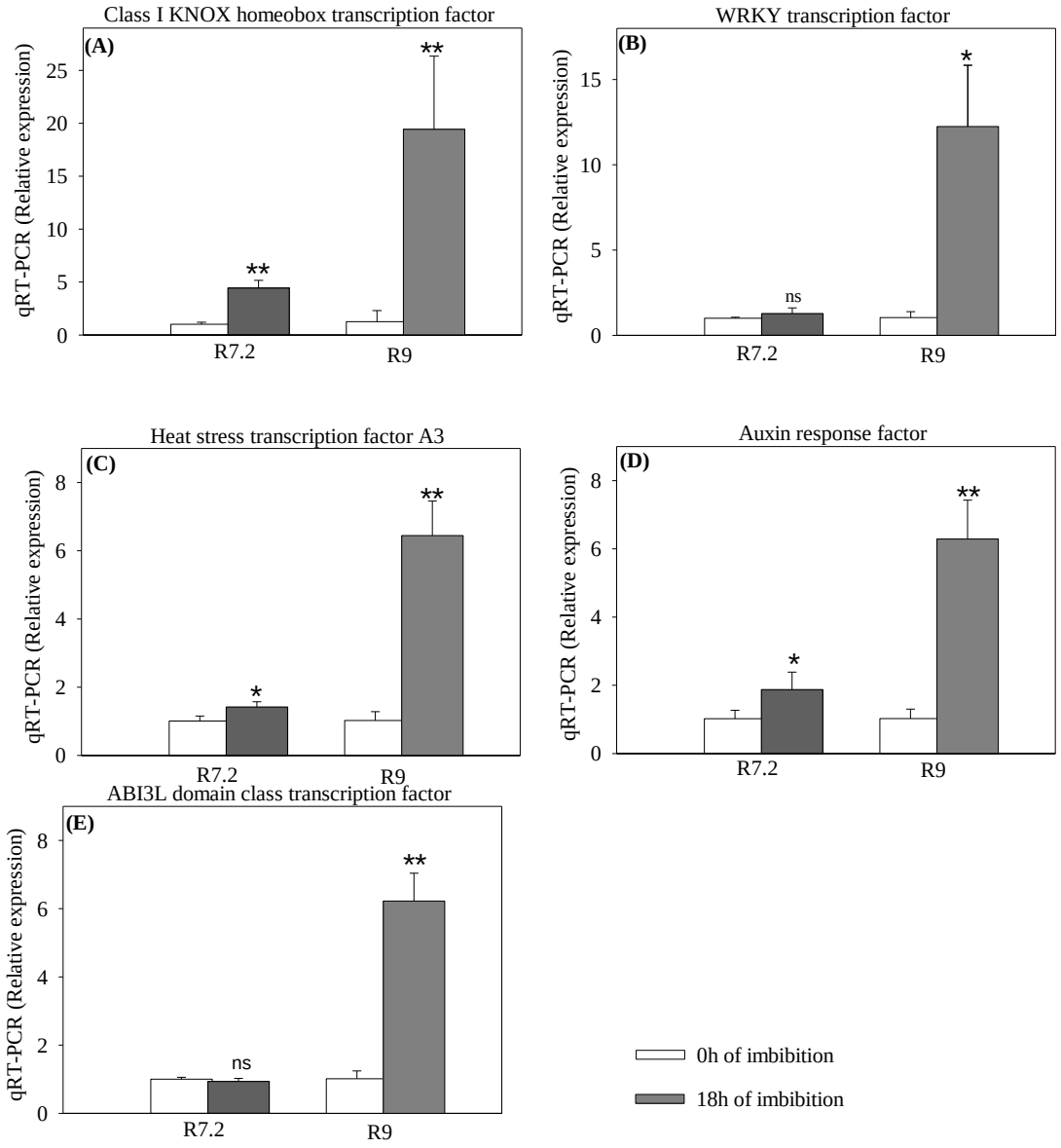

Figure 19. Relative expression levels (RT-qPCR) of indicated transcription factors being upregulated during imbibition. Gene identification number: (A) *Glyma.*14G047000; (B) *Glyma.*03G224700; (C) *Glyma.*10G066100; (D) *Glyma.*14G217700; (E) *Glyma.*02G200500. Samples were collected of stage R7.2 and R9. From control was used seeds not imbibition at each stage. Each column represents the average data with standard errors from three biological repetitions. The data were analyzed using the Relative expression software tool (REST 2009 –Qiagen) and differences were considered: ns, no significant; ** significant p -value < 0.01 and * significant p-value < 0.05.

4.5. COMPARATIVE ANALYSIS BETWEEN SOYBEAN AND *MEDICAGO TRUNCATULA* TO IDENTIFY GENES RELATED TO SEED VIGOR ACQUISITION DURING MATURATION

During imbibition, immature soybean seeds exhibiting a lesser longevity that mature seeds exhibited many changes in the transcriptome that were specific to the maturity level. However, during imbibition, many genes are expressed that are not only related to seed(ling) performance, making it difficult to separate gene expression related to seed vigor from genes involved in other functions. Therefore, to understand which mechanisms that might be related to seed quality, we wished to compare our transcriptome data with those of developing seeds of *Medicago truncatula*. *Medicago truncatula* emerged as a model plant for legume genetics and genomics (COOK, 1999). It is related to economically relevant legumes such as alfalfa (*Medicago sativa*), pea (*Pisum sativum*) and soybean (FRUGOLI and HARRIS, 2001). Thus, a comparative analysis between the two legume species is expected to eliminate a large number of genes that are no specific.

4.5.1 Acquisition of seed vigor during maturation of *Medicago truncatula*

First, we selected a stage during Medicago development that exhibited the same characteristics than those of immature soybean at stage R7.2. This stage corresponded to seeds of 28 days after pollination, corresponding to the end of seed filling (CHATELAIN et al., 2012). The water content of the seeds at 28 DAP was 0,52 g H_2O g DW^{-1}. After harvest, immature seeds were rapidly dried back at 43% RH. Both dried and mature seeds were equilibrated at this same RH prior to the physiological experiments. Thereafter, the water content drying was 0,091 ± 0,002 and 0,11 (±0,002) g H_2O g DW^{-1} for immature and mature seed batches, respectively. Next, we compared the different characteristics of seed physiological quality. Seeds at 28 DAP germinated to 100%, albeit slower than fully mature seeds (Figure 20).

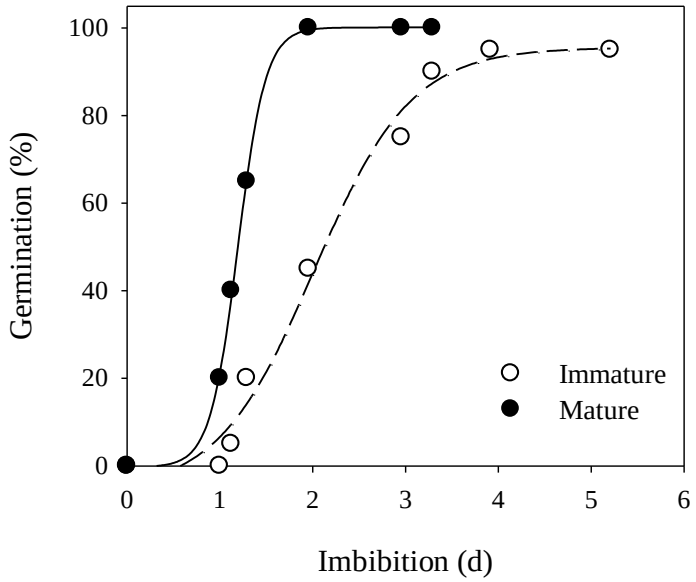

Figure 20. Germination of *Medicago truncatula* dried seeds harvested at 28DAP (immature) and mature. Data are the means of 3 replicates of 50 seeds of each stage.

Cold imbibition at 10°C decreased speed of germination, but no difference was found in the delay or percentage of germination between the two stages (Table 26). However, the percentage of normal seedlings under cold imbibition was significantly lower in the immature seed lot compared to the mature lot 43% of the immature seeds developed in abnormal seedlings (Figure21). No difference was observed germination when seeds were imbibed in salt (data not shown).

Table 26. Speed of germination (T50, days) of immature (28DAP) and mature *Medicago truncatula* seeds during imbibition at 10°C and 20°C, in the dark.

Temperature	**T50 (d)**	
	Immature (28DAP)	**Mature**
10°C	5,0	2,9
20 °C	2,1	1,2
10 °C /20 °C	2,4	2,4

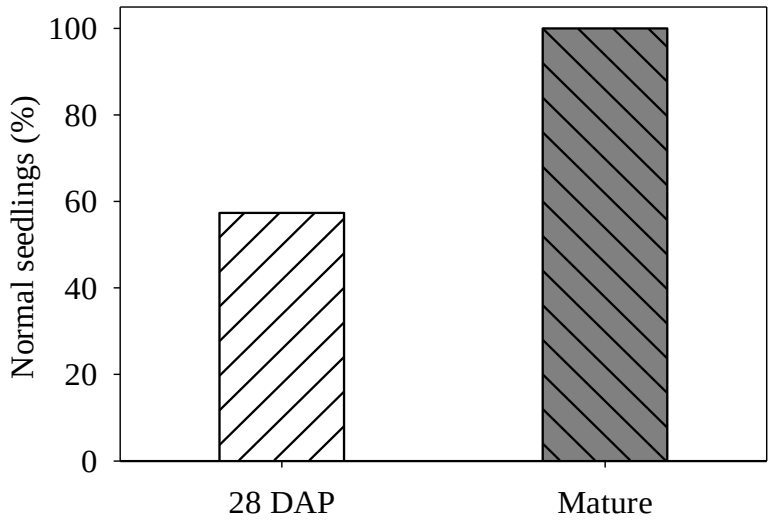

Figure 21. Changes in percentage of normal seedlings of *Medicago truncatula* seeds harvested at 28 DAP (immature) and mature imbibed for 7d at 10°C in the dark. Each column represents the average of three replications of 50 seeds.

Like in soybean, the major difference between 28 DAP seeds and mature seeds was their seed longevity. During storage in conditions similar to that used for soybean (75% RH and 35°C), germination decreased almost directly upon incubation (Figure 22). Half of the seed population had died after 7 days of storage. In contrast, further maturation led to an increase in seed storability. Mature seeds could be stored for two weeks before a decrease in viability and half of the seed population died after only 25 days of storage (Figure 22). Longevity was increased by almost four-fold in mature seeds compared to immature seeds. These findings are similar to those observed for soybean (Figure 11). Immature seeds of stage R7.2 started to lose their viability after 21 days of storage (Figure 11). Half the population was dead after 27 days. In contrast, it took 48 days to kill 50% of mature seeds.

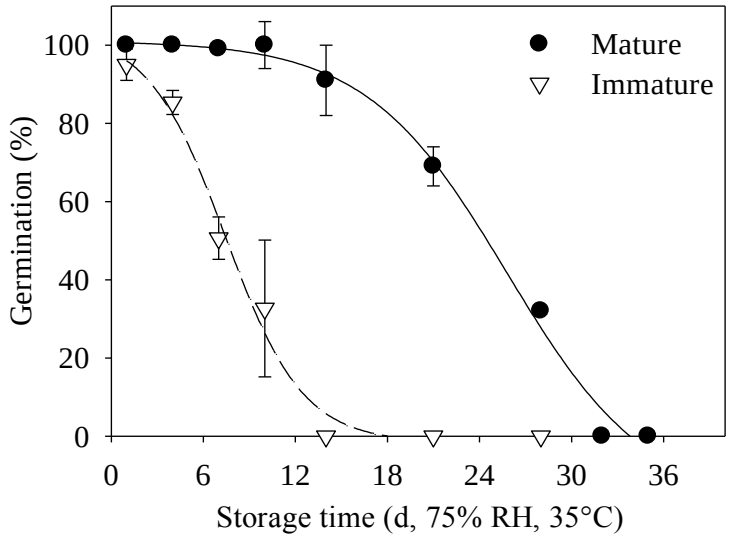

Figure 22. Survival curves of immature (28 DAP) and mature *Medicago truncatula* seeds showing a decrease in percentage of germination during storage at 35°C and 75% relative humidity (RH). Germination was tested at different storage time intervals using three replicates of 50 seeds. Error bars represent standard deviation.

4.5.2 Characterization of genes that are differentially expressed in immature vs mature seeds of legumes

Transcriptomes were obtained from immature (28DAP) dried and mature seeds. Total RNA was isolated in three biological replicates and used to hybridize the Affymetrix *M. truncatula* GeneChip containing 102.121 probe sets. After statistical analysis, differentially expressed genes (DEG) were considered significant when the *p*-value was < 0,01. Genes differential expression being up-regulated corresponding the expression level of Log_2>1; and being down-regulated corresponding the expression level of Log_2 < -1 (Table 27).

Table 27. Number of differential transcripts between immature dried and mature seeds of *M. truncatula* ranked according to the level of expression calculated on a Log_2 ratio. Mat: for mature seeds; Imm d: for immature dried seeds.

	Log_2(ratio)							
	> 1	> 2	> 3	> 4	< -1	< - 2	< - 3	< -4
Mat 0h/ Imm d 0h	640	72	7	1	1788	272	20	0

A comparative analysis was performed between the homologues of Medicago and soybean using the Arabidopsis AGI number, since no database exists that provides a correct and complete comparison of the two legume species. Considering the seed longevity increases in both soybean and Medicago, after filling while the seeds are losing water, we first analyzed DEG related with longevity.

It was observed that in *Medicago truncatula* seeds, there were only 640 DEG transcripts being up-regulated at least 2-fold, whereas 1788 were down-regulated, between immature seed dried (28DAP) and mature. This result was similar numerically when analyzing the transcripts expressed in soybean seeds between immature seed dried (R7.2) and mature (R9); wherein 742 DEG transcript being up-regulated and 1525 were down-regulated.

Among UP-regulated transcripts, only 7%, that corresponding to 36 homologous genes in Arabidopsis, were found to be common in both species. A closer look at those transcripts (Table 28) revealed genes encoding MALATE SYNTHASE, PHY A, PHYB, NFXlike 1, ABI-1-like 1, and SCARECROW-like 14.

When we compare the 1788 DEG transcripts being down-regulated in Medicago during seed maturation with the 1525 were down-regulated for soybean also during maturation, it was found 223 genes are common for both species. This corresponds to 20% DEG transcripts expressed during maturation. The GO enrichment analysis corresponding at these transcripts is show on the table 29. It were related to processes associated with with response to stimulus, metabolism and defense.

Table 28. Genes up-regulated in soybean and medicago during seed maturation.

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*p*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	*p*-value	Name Arabdopsis gene
Medtr_v1_002024	AC233577_31_1	AT5G03860.2	0,00E+00	Glyma.17G128000	I1MUM7_SOYBN	0	AT5G03860.2	0	malate synthase
Medtr_v1_006650	MEDTR1G085160_1	AT1G09570.1	0,00E+00	Glyma.10G141400	B5U9F5_SOYBN	0	AT1G09570.1	0	phytochrome A
Medtr_v1_006650	MEDTR1G085160_1	AT1G09570.1	0,00E+00	Glyma.19G224200	C1PHB8_SOYBN	0	AT1G09570.1	0	phytochrome A
Medtr_v1_006650	MEDTR1G085160_1	AT1G09570.1	0,00E+00	Glyma.20G090000	B4YB07_SOYBN	0	AT1G09570.1	0	phytochrome A
Medtr_v1_077941	IMGA|contig_68627_1.1	AT5G27950.1	0,00E+00	Glyma.20G229600	U7E2V5_POPTR	0	AT5G27950.1	0	P-loop containing nucleoside triphosphate hydrolases superfamily protein
Medtr_v1_004695	MEDTR1G019890_1	AT4G24620.1	3,60E-01	Glyma.02G212600	I1JGY5_SOYBN	0	AT4G24620.1	0	phosphoglucose isomerase 1
Medtr_v1_006290	MEDTR1G079370_1	AT2G02360.1	4,00E-67	Glyma.10G169300	T2DNH1_PHAVU	6,00E-113	AT2G02360.1	2,00E-60	phloem protein 2-B10
Medtr_v1_018521	MEDTR3G085440_1	AT1G27920.1	0,00E+00	Glyma.11G249900	G7JC01_MEDTR	0	AT1G27920.1	0	microtubule-associated protein 65-8
Medtr_v1_017850	MEDTR3G072830_1	AT2G41705.2	1,00E-119	Glyma.04G252000	G7J3J9_MEDTR	0	AT2G41705.2	2,00E-145	camphor resistance CrcB family protein
Medtr_v1_071574	IMGA|contig_8315_1.1	AT1G53210.1	0,00E+00	Glyma.12G228000	E5GBG6_CUCME	1,00E-153	AT1G53210.1	1,00E-124	sodium/calcium exchanger family protein / calcium-binding EF hand family protein
Medtr_v1_071574	IMGA|contig_8315_1.1	AT1G53210.1	0,00E+00	Glyma.13G272000	E5GBG6_CUCME	0	AT1G53210.1	0	sodium/calcium exchanger family protein / calcium-binding EF hand family protein
Medtr_v1_077815	IMGA|contig_58561_1.1	AT2G39970.1	1,00E-110	Glyma.11G227400	G7KBX5_MEDTR	6,00E-169	AT2G39970.1	1,00E-160	Mitochondrial substrate carrier family protein
Medtr_v1_024414	MEDTR4G062100_1	AT1G07710.1	0,00E+00	Glyma.U033200	G7JSN3_MEDTR	0	AT1G07710.1	0	Ankyrin repeat family protein
Medtr_v1_024414	MEDTR4G062100_1	AT1G07710.1	0,00E+00	Glyma.12G074900	G7JSN3_MEDTR	0	AT1G07710.1	0	Ankyrin repeat family protein
Medtr_v1_046651	MEDTR7G111060_1	AT1G02305.1	1,00E-154	Glyma.19G223300	T2DLX7_PHAVU	0	AT1G02305.1	0	Cysteine proteinases superfamily protein
Medtr_v1_072024	IMGA|contig_11361_1.1	AT5G05250.1	1,00E-40	Glyma.06G142900	C7IVU8_9POAL	2,00E-25	AT5G05250.1	3,00E-46	unknown protein
Medtr_v1_072024	IMGA|contig_11361_1.1	AT5G05250.1	1,00E-40	Glyma.04G222500	C7IVU8_9POAL	2,00E-25	AT5G05250.1	3,00E-44	unknown protein
Medtr_v1_073748	IMGA|contig_51360_2.1	AT2G30580.1	1,00E-102	Glyma.10G162800	B9H8H8_POPTR	3,00E-132	AT2G30580.1	1,00E-123	DREB2A-interacting protein 2
Medtr_v1_073371	NA	AT2G46225.1	1,00E-100	Glyma.16G030200	G7L308_MEDTR	9,00E-148	AT2G46225.1	4,00E-131	ABI-1-like 1
Medtr_v1_073371	NA	AT2G46225.1	1,00E-100	Glyma.07G062200	G7L308_MEDTR	2,33E-156	AT2G46225.1	1,00E-127	ABI-1-like 1
Medtr_v1_020457	MEDTR3G114840_1	AT4G35790.2	0,00E+00	Glyma.11G081500	I1LI58_SOYBN	0	AT4G35790.2	0	phospholipase D delta
Medtr_v1_011512	MEDTR2G048720_1	AT1G30220.1	0,00E+00	Glyma.05G142400	G7IGM5_MEDTR	0	AT1G30220.1	0	inositol transporter 2
Medtr_v1_046086	MEDTR7G100110_1	AT1G07670.1	0,00E+00	Glyma.03G175200	F9W2W4_MEDTR	0	AT1G07670.1	0	endomembrane-type CA-ATPase 4

Table 28. Continuation…

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*p*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	*p*-value	Name Arabdopsis gene
Medtr_v1_004109	MEDTR1G009360_1	AT4G32690.1	3,00E-18	Glyma.14G140400	A2TDC3_9ROSI	2,00E-104	AT4G32690.1	3,00E-94	hemoglobin 3
Medtr_v1_089457	IMGA|contig_64154_1.1	AT4G28300.1	3,00E-51	Glyma.13G297200	B9IQU2_POPTR	3,00E-153	AT4G28300.1	3,00E-110	Protein of unknown function (DUF1421)
Medtr_v1_084563	IMGA|contig_74723_1.1	AT1G63850.1	0,00E+00	Glyma.08G016800	Q7X711_ORYSJ	0	AT1G63850.1	0	BTB/POZ domain-containing protein
Medtr_v1_022917	MEDTR4G023030_1	AT2G44500.1	0,00E+00	Glyma.09G003300	G7JVS2_MEDTR	0	AT2G44500.1	0	O-fucosyltransferase family protein
Medtr_v1_051083	MEDTR8G079710_1	AT5G67150.1	1,00E-74	Glyma.17G152600	G8A1G0_MEDTR	2,00E-161	AT5G67150.1	5,00E-109	HXXXD-type acyl-transferase family protein
Medtr_v1_010757	MEDTR2G034040_1	AT2G18790.1	0,00E+00	Glyma.15G140000	I1MGE5_SOYBN	0	AT2G18790.1	0	phytochrome B
Medtr_v1_010757	MEDTR2G034040_1	AT2G18790.1	0,00E+00	Glyma.09G035500	B4YB10_SOYBN	0	AT2G18790.1	0	phytochrome B
Medtr_v1_099183	NA	AT2G35110.1	1,00E-66	Glyma.20G019300	E3UV65_MEDTR	0	AT2G35110.2	0	transcription activators
Medtr_v1_023121	MEDTR4G027040_1	AT1G10170.1	0,00E+00	Glyma.08G055900	G7JH33_MEDTR	0	AT1G10170.1	0	NF-X-like 1
Medtr_v1_018957	MEDTR3G091570_1	AT5G24300.2	0,00E+00	Glyma.06G129400	Q75T80_PHAVU	0	AT5G24300.2	0	Glycogen/starch synthases, ADP-glucose type
Medtr_v1_050784	MEDTR8G075190_1	AT4G35930.1	2,00E-72	Glyma.16G123900	G7LB69_MEDTR	4,00E-128	AT4G35930.1	1,00E-92	F-box family protein
Medtr_v1_077006	NA	AT2G46550.1	1,00E-62	Glyma.03G258600	-	-	AT2G46550.1	9,00E-80	unknown protein
Medtr_v1_024497	MEDTR4G064150_1	AT1G07530.1	1,00E-168	Glyma.11G138400	G7JUT6_MEDTR	0	AT1G07530.1	0	SCARECROW-like 14
Medtr_v1_001224	AC225530_1_1	AT3G21090.1	1,00E-86	Glyma.12G020300	G7JRM5_MEDTR	0	AT3G21090.1	0	ABC-2 type transporter family protein
Medtr_v1_093809	IMGA|contig_163394_1.1	AT3G56640.1	0,00E+00	Glyma.16G084100	B9H8R1_POPTR	0	AT3G56640.1	0	exocyst complex component sec15A
Medtr_v1_066683	NA	AT4G17080.1	1,00E-116	Glyma.05G067200	B8XY96_BRARP	8,00E-49	AT4G17080.1	2,00E-54	Histone H3 K4-specific methyltransferase SET7/9 family protein
Medtr_v1_017942	MEDTR3G076820_1	AT1G23870.1	7,00E-17	Glyma.02G033500	G8XR07_GOSAR	0	AT1G23870.1	0	trehalose-phosphatase/synthase 9
Medtr_v1_071318	IMGA|contig_60598_1.1	AT4G00430.1	1,00E-145	Glyma.01G220600	G7KC21_MEDTR	0	AT4G00430.1	0	plasma membrane intrinsic protein 1;4
Medtr_v1_071318	IMGA|contig_60598_1.1	AT4G00430.1	1,00E-145	Glyma.11G023200	G7KC21_MEDTR	0	AT4G00430.1	0	plasma membrane intrinsic protein 1;4
Medtr_v1_030138	MEDTR5G013480_1	AT5G41150.1	0,00E+00	Glyma.01G204500	G7JYL5_MEDTR	2,00E-125	AT5G41150.2	1,00E-108	Restriction endonuclease, type II-like superfamily protein
Medtr_v1_050847	MEDTR8G075940_1	AT3G13080.1	5,00E-91	Glyma.08G101500	M7YF87_TRIUA	0	AT3G13080.1	0	multidrug resistance-associated protein 3

Table 29. Gene Ontology (GO) enrichment analysis of transcripts DEG with down-regulated during legume seeds maturation. Analysis was performed on the 223 genes using PlantGSEA, with annotations for Arabidopsis AGI number.

Gene Set Name (N°Genes)	GO_Description	Genes in Overlap	p-value	FDR
Response to stimulus(6222)	GO:0050896	103	1.83e-18	3.74e-15
Response to stress(4037)	GO:0006950	78	1.01e-18	2.76e-15
Response to chemical stimulus(3953)	GO:0042221	74	1.36e-16	1.39e-13
Response to abiotic stimulus(2615)	GO:0009628	56	4.41e-16	4.01e-13
Response to organic substance(2739)	GO:0010033	52	7.58e-12	3.45e-09
Response to biotic stimulus(1675)	GO:0009607	37	2.83e-11	9.64e-09
Cellular response to stimulus(2302)	GO:0051716	37	4.88e-06	6.31e-4
Defense response(1644)	GO:0006952	35	5.18e-10	1.46e-07
Multi-organism process(1804)	GO:0051704	35	1.96e-08	4.23e-06
Response to hormone stimulus(1364)	GO:0009725	34	9.39e-13	4.8e-10
Response to other organism(1411)	GO:0051707	32	2.84e-10	8.6e-08
Cellular response to chemical stimulus(1403)	GO:0070887	27	1.48e-06	2.34e-4
Response to temperature stimulus(955)	GO:0009266	26	2.45e-11	8.7e-09
Cellular nitrogen compound metabolic process(1369)	GO:0034641	26	3.47e-06	4.78e-4
Response to abscisic acid stimulus(618)	GO:0009737	23	9.45e-16	7.74e-13
Response to carbohydrate stimulus(811)	GO:0009743	22	1.24e-09	3.18e-07
Amine metabolic process(1030)	GO:0009308	21	7.95e-06	9.72e-4
Response to fungus(497)	GO:0009620	17	1.63e-10	5.14e-08
Response to chitin(421)	GO:0010200	16	2.11e-11	7.85e-09
Response to jasmonic acid stimulus(471)	GO:0009753	15	1.74e-08	4.07e-06
Hormone-mediated signaling pathway(589)	GO:0009755	15	3.5e-06	4.78e-4
Response to wounding(340)	GO:0009611	14	4.17e-11	1.37e-08
Hyperosmotic response(250)	GO:0006972	12	1.6e-11	6.23e-09
Response to heat(303)	GO:0009408	12	3.93e-09	9.75e-07
Response to ethylene stimulus(353)	GO:0009723	12	1.59e-07	3.1e-05
Defense response to fungus(342)	GO:0050832	11	1.82e-06	2.81e-4
Two-component signal transduction system (phosphorelay)(169)	GO:0000160	10	3.12e-12	1.5e-09
Abscisic acid mediated signaling pathway(252)	GO:0009738	10	1.05e-07	2.1e-05
Cellular response to abscisic acid stimulus(267)	GO:0071215	10	3.52e-07	6.54e-05
Intracellular signal transduction(222)	GO:0035556	9	3.69e-07	6.71e-05
Jasmonic acid biosynthetic process(135)	GO:0009695	7	1.91e-07	3.63e-05
Chlorophyll catabolic process(58)	GO:0015996	4	1.08e-05	1.25e-3

4.5.3 Characterization of genes that are differentially expressed in immature and mature legume seed during imbibition

The next step was to understand which mechanisms might be underlying the differences in seed quality. During imbibition, many genes are expressed that are not only related to seed(ling) performance, making it difficult to separate gene expression related to seed vigor from genes involved in other functions. A comparative analysis between the two legume species is expected to eliminate a large number of genes that are no specific. For this reason, changes in transcriptome of Medicago seeds were followed at immature (28DAP) seed dried and mature seed during different times of imbibition: 0h, 6h and 24h. The number of differential transcripts expressed, being up-regulated for expression level of $Log_2 > 1$; and being down-regulated for expression level of $Log_2 < -1$ are shown in the table 30.

During the 24h of imbibition, the total number of up- and down-DEG in mature seeds were 2.06-fold and 1.2 fold higher compared to immature seeds. The comparison between mature and immature at each time point of imbibition also indicated that a higher number of DEG in mature seeds compared to immature seeds. This difference was observed already in dry seeds and increased steadily during imbibition for up-regulated transcripts (Table 30). For the down regulated DEG, the number of transcripts decreased mostly between 6 and 24h. Like in soybean this suggests that the RNA metabolism during imbibition appears to be dependent on the maturity stage.

Table 30. Number of differential transcripts between immature dried (Imm d) and mature (Mat) seeds ranked according to the level of expression calculated on a Log_2 ratio. Time of imbibition: 0, 6 and 24 h in water at 20° C in the dark.

	Log_2(ratio)							
	> 1	> 2	> 3	> 4	< -1	< - 2	< - 3	< -4
Mat 0h/ Imm d 0h	640	72	7	1	1788	272	20	0
Mat 6h/ Imm d 6h	811	109	16	1	1502	271	25	6
Mat 24h/ Imm d 24h	1428	352	58	2	1375	116	6	0
Mat 24h/ Mat 0h	7038	3629	1473	351	5267	1710	440	59
Im d 24h/ Imm d 0h	4441	1161	366	90	4690	1307	318	59

To understand which mechanisms are associated with germination in both legumes, we built Venn diagrams using the genes that were differentially up-

regulated or down-regulated in soybean seed at18h of imbibition and in Medicago seed at 6h and 18h (Figure 23). From the 373 genes that had higher transcript levels in mature imbibed soybean seeds compared to immature seeds, 61 (16%) genes were also upregulated in mature compared to immature Medicago seeds (Figure 23 a). Transcripts of 35 genes were already differential after 6h of imbibition, out of which 14 were no longer differential at 24h of imbibition. For 26 genes, transcript levels became differential only after 24h of imbibition. Next, we examined the genes down-regulated during imbibition. From the 740 down-regulated genes between immature and mature imbibed soybean seeds, 129 (17%) genes were also downregulated in mature compared to immature Medicago seeds (Figure 23 b). Transcripts of 110 genes were already differential after 6h of imbibition, out of which 77 were no longer differential at 24h of imbibition. For 19 genes, transcript levels became differential only after 24h of imbibition. Since we are interested in genes that remain differential during imbibition between mature and immature legume seeds, further analyses were performed on the 47 genes with higher transcript levels in mature vs immature seeds, and the 52 genes with lower transcripts in mature compared to immature seeds. Their gene function is further analyzed in the following sections.

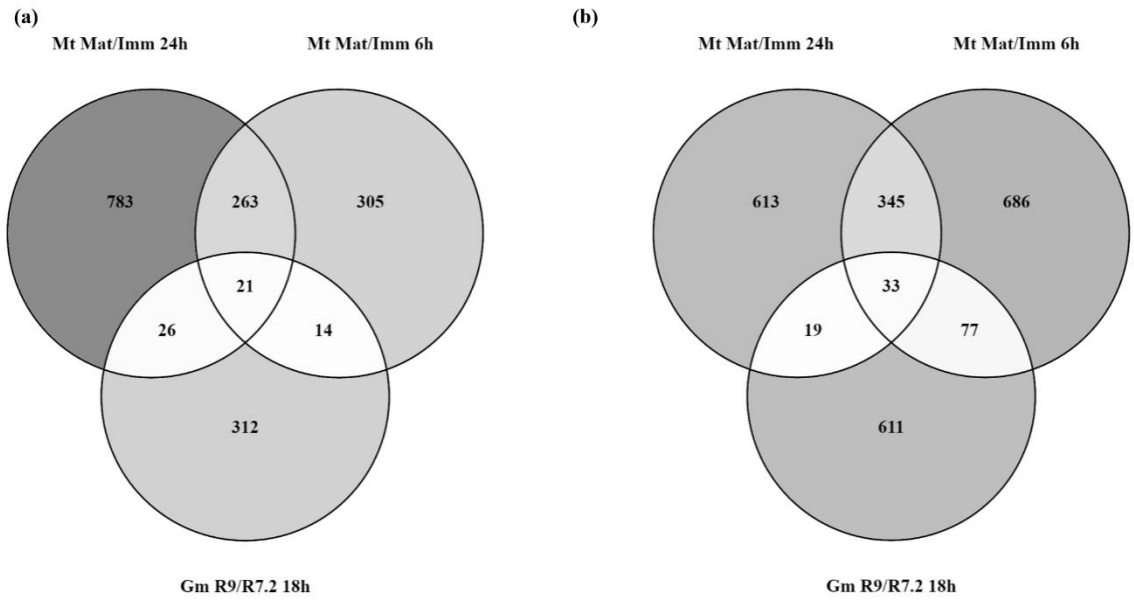

Figure 23. Venn diagrams of genes that are upregulated (A) or downregulated (B) in *Glycine max* and *Medicago truncatula* during imbibition. Gm R9/R7.2 18h: genes that are expressed significantly between immature (R7.2) and mature (R9) soybean seeds during

18h of imbibition. Mt Mat/Imm 6h or 24h: genes that are expressed significantly between immature (28DAP) and mature *M. truncatula* seeds during 6h and 24h of imbibition.

4.5.3.1 Analysis of genes with higher transcript levels in mature legume seeds during imbibition versus immature.

Considering our interest in genes that remain differential during imbibition between mature and immature legume seeds, we first analyzed in more details genes with higher transcript levels in both, soybean and Medicago shows 47 genes with higher transcript levels in mature vs immature seeds (Supplementary Table 5). For most of the genes, their expression level was not yet differential in dry seeds, but increased during imbibition. A part of the genes already showed a significant increase in transcript levels between mature and immature seeds at 6h, but for all genes, the differences were largest at 24h of imbibition. Owing that transcript levels are proportional to transcriptional activity, this suggests that these genes are more actively being transcribed in mature seeds compared to immature seeds during imbibition.

Several analyses were performed using PLantGSEA (YI et al., 2013). For the GO cellular component, 18 out of the 47 genes were assigned to the plasma membrane (enrichment with FDR of $1.35e^{-4}$) and 5 to the apoplast (FDR $1.93e^{-4}$). For the GO biological function (Table 31), gene involved in phototropism were highly overrepresented, as well as genes involved in seed coat development in relation to pigmentation and anthocyanin accumulation and genes involved in innate immune response and ion transport. A second analysis was performed using the curated gene sets. This revealed a signification overrepresentation of genes that were detected as down-regulated by light by in hypocotyl, cotyledon and roots of seedlings.

Table 31. Gene Ontology (GO) enrichment analysis of transcripts DEG with higher transcript levels in mature vs immature imbibed legume seeds. Analysis was performed on the 47 genes using PlantGSEA, with annotations for Arabidopsis AGI number.

Gene Set Name (N^0 Genes)	GO_Description	Genes in Overlap	*p*-value	FDR
Phototropism (13)	GO:0009638	2	3.57E-21	1.77E-17
Seed coat development (18)	GO:0010214	2	4.13E-16	3.42E-13
Pigmentation (113)	GO:0043473	3	9.92E-08	1.70E-05
Anthocyanin accumulation in tissues in response to uv light (113)	GO:0043481	3	9.92E-08	1.70E-05
Response to blue light (126)	GO:0009637	3	5.52E-07	8.85E-05
Defense response (1644)	GO:0006952	10	1.96E-05	2.46E-03
Tropism (166)	GO:0009606	3	2.11E-05	2.56E-03
Innate immune response (926)	GO:0045087	7	4.02E-05	4.65E-03
Ion transport (1016)	GO:0006811	7	1.38E-04	1.32E-02

4.5.3.2 Analysis of genes with lower transcript levels in mature legume seeds during imbibition versus immature.

There are 52 DEG transcripts down-regulated during legume seed imbibition (Supplementary Table 6). The detailed analysis of this list revealead four members of the “Cupin family protein”; two members“HSP20-like chaperones superfamily protein”; one “Raffinose synthase family protein” The Table 32 shows the main GO terms that support a role of RFO metabolism "Raffinose family oligosaccharide biosynthetic process, myo-inositol transport" and secondary metabolism “Phenylpropanoid metabolic process"; "Response to wounding", “flavonoid biosynthetic process”. Also GA metabolism appear to be significantly over-represented.

Table 32. Gene Ontology (GO) enrichment analysis of transcripts DEG with lower transcript levels in mature vs immature imbibed legume seeds. Analysis was performed on the 52 genes using PlantGSEA, with annotations for Arabidopsis AGI number.

Gene set name (n°.Genes)	Description	N°. Genes in Overlap	*p*-value	FDR
Regulation of secondary metabolic process(157)	GO:0043455	5	5.59e-15	1.68e-12
Flavonoid metabolic process(250)	GO:0009812	6	9.06e-14	2.14e-11
Phenylpropanoid metabolic process(400)	GO:0009698	7	1.17e-11	2.29e-09
Myo-inositol transport(2)	GO:0015798	1	1.7e-11	2.96e-09
Low affinity nitrate transport(2)	GO:0080055	1	1.7e-11	2.96e-09
Flavonoid biosynthetic process(225)	GO:0009813	5	1.85e-10	2.78e-08
Response to wounding(340)	GO:0009611	5	7.44e-07	7.24e-05
Rrna transport(5)	GO:0051029	1	2.52e-06	2.24e-4
Rrna export from nucleus(5)	GO:0006407	1	2.52e-06	2.24e-4
Small molecule biosynthetic process(1859)	GO:0044283	12	4.7e-06	3.89e-4
Dna-dependent transcription, elongation(133)	GO:0006354	3	4.95e-06	4.00e-04
Gibberellin biosynthetic process(54)	GO:0009686	2	1.03e-05	7.96e-4
Response to abiotic stimulus(2615)	GO:0009628	14	2.42e-05	1.7e-3
Response to abscisic acid stimulus(618)	GO:0009737	6	3.02e-05	2.08e-3
Polyamine metabolic process(60)	GO:0006595	2	3.19e-05	2.16e-3
Gibberellin metabolic process(61)	GO:0009685	2	3.78e-05	2.5e-3
Small molecule metabolic process(4053)	GO:0044281	18	5.17e-05	3.02e-3
Response to temperature stimulus(955)	GO:0009266	7	2.31e-4	0.0122
Generation of precursor metabolites and energy(730)	GO:0006091	6	2.41e-4	0.0124
Gibberellic acid mediated signaling pathway(81)	GO:0009740	2	4.69e-4	0.021
Gibberellin mediated signaling pathway(82)	GO:0010476	2	5.16e-4	0.0225
Cellular response to gibberellin stimulus(83)	GO:0071370	2	5.66e-4	0.0243
Raffinose family oligosaccharide biosynthetic process(10)	GO:0010325	1	6.44e-4	0.027
Glycoside biosynthetic process(215)	GO:0016138	3	7.37e-4	0.0305
Response to high light intensity(224)	GO:0009644	3	1.03e-3	0.041
Response to hormone stimulus(1364)	GO:0009725	8	1.05e-3	0.0415

4.5.4 Abscisic acid (ABA) and Gibberellin (GA) signalying and metabolism during legumes imbibition

It is known that ABA plays an important in the control of dormancy and stress response, particularly drought stress. Indition, ABA plays role during seed development, promoting protein synthesis; and inhibits germination in the mature seeds (reviewed in FINKELSTEIN et al., 2002). The hormonal balance between ABA and gibberellins (GAs) has been shown to act as an integrator of environmental cues to maintain dormancy or activate germination (FINKELSTEIN et al., 2008; NONOGAKI et al., 2010; WEITBRECHT et al., 2011; RAJJOU et al., 2012).

During imbibition legume seeds, we find genes related to signaling and metabolism of ABA and GA. Considering our interest in genes that remain differential during imbibition between mature and immature legume seeds, first moment we analyzed in detail the genes with higher transcript levels during imbibition of soybean seed mature and then comparing with higher levels during imbibition of soybean seed immature. However, we note that there were no differences between both stages during imbibition. The same was observed for the genes higher expressed (Table 33), as well as for the genes that were down express (Table 34).

Table 33. Positive expression of genes involved in signaling and matabolism of ABA and GA during imbibition of soybean seeds. Presents: Immature seeds (I), Mature seeds (M) or both stages (B).

Process involved	Gmax 2.0 ID	Tax Hit	*p*-value	Name (Soybase)	Arabidopsis	*p*-value	Presence
ABA signaling	Glyma.16G076600	Q0H212_PHAVU	0	Abscisic acid 8'-hydroxylase	AT4G19230	0	B
	Glyma.01G153300	B0L642_SOYBN	0	Abscisic acid 8'-hydroxylase	AT4G19230	0	B
Ga signaling	Glyma.04G150500	Q0HA68_MALDO	0	DELLA protein	AT1G14920	3,00E-158	M
	Glyma.11G216500	A5HVE5_PHAVU	0	DELLA protein	AT3G03450	0	B
	Glyma.06G312900	I6LZP2_SOYBN	0	GAMYB1 protein	AT3G11440	4,00E-102	B
	Glyma.04G125700	I6LZP2_SOYBN	0	GAMYB1 protein	AT5G06100.3	2,00E-102	B
	Glyma.13G187500	I6LZP2_SOYBN	0	GAMYB1 protein	AT5G06100.3	9,00E-120	B
	Glyma.15G225300	I6LZP2_SOYBN	0	GAMYB1 protein	AT5G06100.3	3,00E-118	B
	Glyma.09G157600	G7KNV2_MEDTR	0	Gibberellin receptor GID1	AT5G23530	5,00E-147	B
	Glyma.16G208100	G7KNV2_MEDTR	0	Gibberellin receptor GID1	AT5G23530	3,00E-144	M
GA metabolism	Glyma.02G257900	G7K254_MEDTR	1,00E-138	Gibberellin 2-beta-dioxygenase	AT3G11150	3,00E-97	I
	Glyma.09G149200	O04280_PHAVU	0	Gibberellin 20-oxidase	AT4G25420	0	B
	Glyma.05G062400	G7JHQ5_MEDTR	0	Gibberellin 2-beta-dioxygenase	AT4G16770	5,00E-129	B
	Glyma.13G361700	G3OX_PEA	0	Gibberellin 3-beta-dioxygenase 1	AT1G15550	1,00E-149	B
	Glyma.15G012100	G3OX_PEA	0	Gibberellin 3-beta-dioxygenase 1	AT1G15550	6,00E-150	B

Table 33. Continuation…

Process involved	Gmax 2.0 ID	Tax Hit	*p*-value	Name (Soybase)	Arabidopsis	*p*-value	Presence
ABA metabolism	Glyma.19G122800	Q2HUH2_MEDTR	3,00E-143	ABA-responsive element binding protein 3	AT3G56850	1,00E-102	B
	Glyma.03G003400	Q2HUH2_MEDTR	2,00E-141	ABA-responsive element binding protein 3	AT3G56850	2,00E-103	M
	Glyma.05G079800	Q2HUH2_MEDTR	2,00E-124	ABA-responsive element binding protein 3	AT3G56850	4,00E-90	M
	Glyma.19G067900	Q2HUH2_MEDTR	7,00E-127	ABA-responsive element binding protein 3	AT3G56850	8,00E-98	M
	Glyma.20G049200	G7K7P6_MEDTR	9,00E-113	ABA-responsive element binding protein 3	AT3G56850	1,00E-59	B
	Glyma.06G314400	D9ZIQ3_MALDO	7,00E-166	ABA-responsive element binding protein 3	AT3G56850	1,00E-91	M

Table 34. Negative expression of genes involved in signaling and matabolimo of ABA and GA during imbibition of soybean seeds. Presents: Immature seeds (I), Mature seeds (M) or both stages (B).

Process involved	Gmax 2.0 ID	Tax Hit	*p*-value	Name (Soybase)	Arabidopsis	*p*-value	Presence
ABA signaling	Glyma.14G162100	Q9ZPL9_LOTJA	7,00E-173	Protein phosphatase type 2C	AT2G29380.1	5,00E-84	B
	Glyma.19G069200	Q9ZPL9_LOTJA	4,00E-164	Protein phosphatase type 2C	AT1G07430.1	3,00E-85	I
	Glyma.01G225100	G7KAL2_MEDTR	9,00E-144	Protein phosphatase 2C	AT2G29380.1	7,00E-83	B
	Glyma.11G018000	G7KAL2_MEDTR	2,00E-148	Protein phosphatase 2C	AT2G29380.1	7,00E-81	B
	Glyma.09G282900	H2BIS8_CITSI	0	ABA 8'-hydroxylase	AT2G29090.2	0	B
Ga signaling	Glyma.10G190200	Q0HA68_MALDO	0	DELLA protein	AT1G14920.1	8,00E-151	B
	Glyma.20G200500	Q0HA68_MALDO	0	DELLA protein	AT1G14920.1	2,00E-137	B
GA metabolism	Glyma.04G041100	G7JMF0_MEDTR	5,00E-35	Gibberellin receptor GID1	AT1G47480.1	7,00E-21	B
	Glyma.07G211100	G7JMF0_MEDTR	4,00E-168	Gibberellin receptor GID1	AT1G47480.1	7,00E-104	M
	Glyma.02G084400	G8A030_MEDTR	2,00E-173	Gibberellin 20 oxidase 1-B	AT1G03400.1	5,00E-106	B
	Glyma.07G168500	G8A030_MEDTR	0	Gibberellin 20 oxidase 1-B	AT1G03400.1	2,00E-106	B
	Glyma.12G029800	G7K109_MEDTR	2,00E-120	Gibberellin 2-beta-dioxygenase	AT4G21200.1	3,00E-119	B
	Glyma.13G218200	G2OX_PHACN	0	Gibberellin 2-beta-dioxygenase	AT1G30040.1	2,00E-153	I
	Glyma.14G058600	G7K254_MEDTR	3,00E-137	Gibberellin 2-beta-dioxygenase	AT3G11150.1	4,00E-94	I
	Glyma.15G093900	G2OX_PHACN	0	Gibberellin 2-beta-dioxygenase	AT1G30040.1	1,00E-154	B
	Glyma.13G259400	Q4W8C3_PHAAN	0	Gibberellin 2-oxidase	AT1G78440.1	6,00E-130	B

Table 34. Continuation…

Process involved	Gmax 2.0 ID	Tax Hit	*p*-value	Name (Soybase)	Arabidopsis	*p*-value	Presence
ABA metabolism	Glyma.01G186200	I1LHE5_SOYBN	0	Zeaxanthin epoxidase, chloroplastic	AT5G67030.1	0	M
	Glyma.01G186200	I1LHE5_SOYBN	0	Zeaxanthin epoxidase, chloroplastic	AT5G67030.1	0	B
	Glyma.11G055700	I1LHE5_SOYBN	0	Zeaxanthin epoxidase, chloroplastic	AT5G67030.1	0	B
	Glyma.17G174500	I1MVX7_SOYBN	0	Zeaxanthin epoxidase, chloroplastic	AT5G67030.1	0	I
	Glyma.15G250100	NCED1_PHAVU	0	9-cis-epoxycarotenoid dioxygenase NCED1	AT3G14440.1	0	I

Reinforcing our interest in genes that remain differential during imbibition between mature and immature legume seeds, our final step was to analyze in detail the list of genes highly expressed of mature seeds in *Medicago truncatula* (#genes 7038) and immature seeds (# 4441) during imbibition. We noticed that for medicago seeds there is a difference to ABA/ GA metabolism according to maturation of the seeds (Table 35), i.e, in mature seed there are genes involved in ABA metabolism and signaling, such as CYP707A1, nine-cis-epoxycarotenoid dioxygenase 4, zeaxanthin epoxidase and ABA-responsive element binding protein 3

.

Table 35. Positive expression of genes involved in signaling and matabolism of ABA and GA during imbibition of Medicago truncatula seeds. Presents: Immature seeds (I) or Mature seeds (M).

Process involved	Probe_id	Seq_id Mt 4.0 v1	Arabidopsis	Name Arabdopsis gene	*p*-value	Presence
ABA signaling	Medtr_v1_030457	MEDTR5G017350_1	AT5G67030	ABA1,LOS6,NPQ2,ATABA1,ZEP,IBS3,ATZEP; zeaxanthin epoxidase (ZEP) (ABA1)	1,00E-130	M
	Medtr_v1_030456	MEDTR5G017330_1	AT5G67030	ABA1,LOS6,NPQ2,ATABA1,ZEP,IBS3,ATZEP; zeaxanthin epoxidase (ZEP) (ABA1)	1,00E-124	M
	Medtr_v1_030459	MEDTR5G017370_1	AT5G67030.2	ABA1,LOS6,NPQ2,ATABA1,ZEP,IBS3,ATZEP; zeaxanthin epoxidase (ZEP) (ABA1)	9,00E-15	M
Ga signaling	Medtr_v1_006731	MEDTR1G086550_1	AT1G02400	ATGA2OX4,ATGA2OX6,DTA1,GA2OX6; gibberellin 2-oxidase 6	1,00E-106	M
	Medtr_v1_013745	MEDTR2G102570_1	AT1G15550	GA4,ATGA3OX1,GA3OX1; gibberellin 3-oxidase 1	1,00E-119	M
	Medtr_v1_004175	MEDTR1G011580_1	AT1G15550	GA4,ATGA3OX1,GA3OX1; gibberellin 3-oxidase 1	2,00E-88	M
	Medtr_v1_009951	MEDTR2G019370_1	AT1G30040	ATGA2OX2,GA2OX2; gibberellin 2-oxidase	1,00E-123	M
	Medtr_v1_092485	IMGA|contig_247863_1.1	AT1G47990	ATGA2OX4,GA2OX4; gibberellin 2-oxidase 4	2,00E-59	M
	Medtr_v1_029438	MEDTR5G005570_1	AT4G21200	ATGA2OX8,GA2OX8; gibberellin 2-oxidase 8	1,00E-126	M
	Medtr_v1_089034	NA	AT4G25420	GA5,GA20OX1,AT2301,ATGA20OX1; 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily pro	1,00E-150	M
	Medtr_v1_026223	MEDTR4G096840_1	AT1G30040	gibberellin 2-oxidase	1,00E-113	I
	Medtr_v1_086002	IMGA|contig_71636_1.1	AT1G78440	Arabidopsis thaliana gibberellin 2-oxidase 1	5,00E-88	I
	Medtr_v1_010774	MEDTR2G034280_1	AT3G03450	RGA-like 2	4,00E-51	I
	Medtr_v1_010772	MEDTR2G034260_1	AT3G03450	RGA-like 2	1,00E-52	I
	Medtr_v1_088591	IMGA|contig_165928_1.1	AT5G51810	gibberellin 20 oxidase 2	1,00E-130	I
	Medtr_v1_083540	IMGA|contig_91823_1.1	AT1G78440	ATGA2OX1,GA2OX1; Arabidopsis thaliana gibberellin 2-oxidase 1	1,00E-109	I

Table 35. Continuation…

Process involved	Probe_id	Seq_id Mt 4.0 v1	Arabidopsis	Name Arabdopsis gene	*p*-value	Presence
GA metabolism	Medtr_v1_089118	IMGA|contig_52215_1.1	AT1G14920	GAI,RGA2; GRAS family transcription factor family protein	0	M
	Medtr_v1_050454	MEDTR8G069100_1	AT2G22475	GEM; GRAM domain family protein	2,00E-24	M
	Medtr_v1_019312	MEDTR3G096500_1	AT5G51810	GA20OX2,AT2353,ATGA20OX2; gibberellin 20 oxidase 2	1,00E-115	M
	Medtr_v1_088394	IMGA|contig_55897_1.1	AT1G14920	GRAS family transcription factor family protein	1,00E-133	I
ABA metabolism	Medtr_v1_031175	MEDTR5G025610_1	AT4G19230	CYP707A1; cytochrome P450	0	M
	Medtr_v1_050546	MEDTR8G072260_1	AT4G19230	CYP707A1; cytochrome P450	0	M
	Medtr_v1_031145	MEDTR5G025270_1	AT4G19170	nine-cis-epoxycarotenoid dioxygenase 4	0	I
	Medtr_v1_045419	MEDTR7G088090_1	AT3G56850	AREB3,DPBF3; ABA-responsive element binding protein 3	2,00E-81	M
	Medtr_v1_015247	MEDTR3G010660_1	AT3G56850	AREB3,DPBF3; ABA-responsive element binding protein 3	3,00E-71	M
	Medtr_v1_035019	MEDTR5G080680_1	AT1G07430	HAI2; highly ABA-induced PP2C gene 2	1,00E-118	M

5. DISCUSSION

5.1. CHARACTERIZATION OF ACQUISITION OF SEED QUALITY AND ABSCISIC ACID QUANTIFICATION DURING SOYBEAN SEED DEVELOPMENT.

The first work to characterize phenological scale of soybean plant development was proposed by Fehr and Caviness (1977). This phenological scale contains vegetative stages (V1, V2...Vn) and eight reproductive stages (R1, R2, R3…R8). Then an updated version was published by the same group of researchers (RITCHIE et al., 1982). In this new publication, the stage R7 was divided into R7.1 (50% yellowing of leaves and pods); R7.2 (between 50 and 75%) and R7.3 (above 75%). The R9 stage was proposed to describe the point of field maturation or harvest maturity, which happens when the seeds water content reaches approximately 12%. Since the scale proposed by Ritchie et al (1982) mostly consider changes in the plant, leaf and pod (Table 3) we further improved this scale by adding seed trait information like seed size and of the pod, changes in color of the cotyledons and embryonic axis, as well as changes of the hilum and pod color (Figure 6).

As expected, soybean seed development could be divided in several distinct phases ranging from tissue differentiation and growth to seed filling and de-greening followed by seed drying and harvest maturation. The seed filling started at stage R5.2. Fresh and dry weights increased steadily onwards until stage R7.2. Thereafter, dry weight was stable, indicating the end of the filling phase. De-greening starting in stage R7.1, the pod and the embryo turned yellow while everything else still green. In stage R7.2

de-greening increased and is mostly related to chlorophyll degradation, which is an important developmental event that is necessary to avoid the deleterious effects caused by the oxidative power of chlorophyll and its degradation products (TEIXEIRA et al., 2016). The direct consequence of chlorophyll retention in soybean seeds is the discarding of many lots, result of the decline of vigor and viability (ZORATTO et al., 2007; PÁDUA et al., 2007).

The chlorophyll content of mature soybean seeds is determined by the genotype, and therefore varies among the cultivars. This level, as the seed maturation stage, can be influenced by the drying conditions, and climate conditions affecting the normal ripening under field conditions (SINNECKER, 2002). In normal circumstances, during seed maturation, the chlorophyll is degraded. The retention can be caused by uneven maturation, due to insect attacks mainly bedbugs that cause retention of the stem green (CÂMARA and HEIFFIG, 2000). Other factors that predispose soybeans to the expression of green seed, according to França-Neto et al. (2005) are biotic and abiotic stresses, resulting in premature death of the plant or forced maturation. During the seed filling, period preceding the R7.2 stage, the soybean plant is sensitive to water stress. According to Farias et al. (2001) water availability, photoperiod and temperature are the factors that most affect the development and soybean productivity. Water stress during the seed filling could reduce soybean yield by accelerating leaf senescence, shortening the seed filling period and produce smaller seeds (BREVEDAN and EGLI, 2003).

During the development, germinability was acquired early, between R5.5 and R6. At stage R7.2, the seed water content decreased steadily until stage R9. Therefore, the maturation drying was considered to be from stage R7.2 to R9 since dry weight did not change significantly. The slow dehydration during the maturation process is necessary to produce viable soybean seeds, to result in chlorophyll degradation, and to allow genome changes to produce specific enzymes that act in germination process (ADAMS et al., 1982).

The abscisic acid was also quantified and we found a peak at stage R6, shortly before the induction of desiccation tolerance. However, whether this peak is necessary to induce desiccation tolerance remains to be investigated. Similarly, a decrease in the rate of dry weight accumulation was associated with a sharp decline in ABA concentration. Abscisic acid acts at different stages along the formation and maturation of the seed. In the early embryogenesis stage, the ABA is produced by the maternal tissue and

prevents vivipary (BEWLEY et al., 2013). During maturation, this hormone is produced by the embryonic tissue allowing protein and lipid synthesis (KERMODE 1990, GUTIERREZ et al 2007). Endogenous content in the seed undergoes considerable reduction after reserve accumulation and the beginning of the dehydration of the seed (BEWLEY and BLACK 1994). Thus, it is clear the importance and participation of ABA in seed formation. In addition to inhibiting germination, ABA also plays a role on desiccation tolerance and on the synthesis of late maturation proteins. According to Schussler et al. (1984), ABA may be involved in the stimulation of rapid unloading of sucrose into the testa of soybean and ABA in the cotyledons may enhance sucrose uptake by the cotyledons. Liu et al. (2010) reported that ABA concentration in seed was positively correlated with seed-filling rate.

Differently from previous reports, in which mostly plant characteristics where considered, we added seed physiological traits to further describe the soybean seed development. Interestingly, we found that the point where soybeans reach highest longevity occurred after the point of physiological maturity. We observed that soybean seeds need a step of further drying before harvest to achieve high longevity indexes. Thus, seed longevity is defined as the total time-span during which seeds remain viable, therefore, longevity is taken into account to define seed vigor. The time between the seed harvest and subsequent sowing may be a few days or extend over several months, according to cultivar, place of production, the prevailing environmental conditions and technology production. Therefore, it is extremely important for soybean producers have the assurance that the maximum longevity was acquired.

The physiological characterization shows that there is no phenotypic difference in germination, seedling emergence and seed vigor between stages R7.2 and R9, suggesting that physiological maturity in terms of seedling emergence was reached at R7.2 and not at stage R7, as previously proposed. Thus, the stage R7 is not sufficient to describe the acquisition of germination, desiccation tolerance and vigor. Our results are consistent with those proposed by Zanakis et al. (1994), confirmed that the late maturation and dehydration phase, after physiological maturity are important phases of soybean seed development when several quality components are acquired such as desiccation tolerance, vigor and longevity.

The identification of physiological maturity point during seed development remains controversial among researchers, by having two schools of thoughts.

The 'physiological maturity' has been defined by some as the point during development when seed dry weight has reached its maximum value (TEKRONY and EGLY, 1997) and which no longer occur significant increases in dry weight of the seeds (ELLIS and PIETA FILHO, 1992). Ellis and Pietá Filho (1992), reinforce that the term 'physiological maturity' must be used to refer the point of maximum physiological quality, while that for the maximum dry weight must be used as 'mass maturity'. Whereas that for some species, the highest physiological quality is achieved after maximum dry weight, i.e, is acquired progressively and later during development.

Seed longevity is defined as the ability to remain alive during storage under dry storage conditions. Thus, the last stages of seed maturation of soybean seeds are important to maximize seed quality. Between physiological maturity and harvesting maturity the longevity almost doubled, since P50 increased from 27 days at R 7.2 to 48 days at R9 (Figure 11). Therefore, our results confirm that seed quality characteristics are acquired sequentially after physiological maturity when the seeds lose water and are still in the mother plant (ELLIS et al. 1987; ELLIS et al. 1993; DEBEAUJON, 2000; PROBERT, 2007; CHATELAIN et al. 2012; VERDIER et al. 2013). The increase in physiological quality after physiological maturity is observed in beans, *Brassica napus*, arabidopsis, tomatoes, peppers, and melons (BEWLEY et al., 2013).

An important model in the studies of seed traits is the legume species *Medicago truncatula* (GALLARDO et al., 2003; CHANTELAIN et al., 2012; VERDIER et al., 2013). *M. truncatula* is closely related to economically relevant legumes like alfalfa (*Medicago sativa*), soybean (*Glycine max*), and pea (*Pisum sativum*) and can be used as a platform for physiological and molecular studies in legume species. Chatelain et al. (2012) observed that longevity (P50) in *M. truncatula* increased over 30-fold between 28DAP, corresponding to the end of seed filling and 44 DAP point of abscission, that corresponding to harvest maturity.

It was not found in the literature impact of genetic variation on the stability of the acquisition of physiological maturity. On the other hand, weather adverse conditions such as intense summer during periods of histodifferentiation and maturation of soybean, result in the forced maturation of seeds, resulting in low yields and occurrence of green seeds, which will give a marked reduction of their qualities and severe reduction in crop productivity (FRANÇA NETO et al., 2007).

High temperatures are considered the main responsible for the "forced" soybean maturation (FRANÇA NETO et al., 2005). The effects of environmental stresses, particularly those caused by temperature, are not yet fully known. Studies on the subject suggest that the temperature reaching values above 30^{o}C during the maturation of the seeds can cause severe damage to seed production and quality. In this condition the translocation of reserves occurs very rapid, preventing the complete chlorophyll degradation, causing the formation of green seeds of low quality and impaired performance (MARCOS FILHO, 2005). After physiological maturity, degenerative changes begin to occur in the seeds, so that the physiological quality can be maintained or may decrease, depending on environmental conditions in the period that preceding the harvest, in harvest process, drying, processing and conditions storage (DELOUCHE and BASKIN, 1973; MCDONALD, 1999). Therefore, it is very important that all stress-defense mechanisms are fully established and known.

5.2. TRANSCRIPTOME ANALYSIS DURING SOYBEAN SEED MATURATION

In soybean, transcriptome studies have generated a wealth of data describing seed development, mainly during embryogenesis and filling (HAJDUCH et al., 2005; HUDSON, 2010; JONES et al., 2010, LIBAULT, 2010; SEVERIN et al., 2010; ASAKURA et al., 2012; SHA et al., 2012, SHAMIMUZZAMAM and VODKIN, 2012; AGHAMIRZAIE et al., 2013).

However, transcriptome changes associated with the acquisition of seed physiological quality remains unknown. Furthermore, how maturation drying influences the transcriptome in preparation for the quiescent dry state is unknown. Using RNAseq, we compared two late maturation stages during which seed longevity doubled whereas no other physiological traits were acquired. This revealed three main processes associated with longevity: an increase in HSP transcripts and transcripts associated with light signaling and a decrease in transcripts associated with photosynthesis. Artificial drying at stage R7.2 was compared with natural maturation drying to check whether premature drying induced changes in the transcriptome.

In orthodox seeds, such as soybean, desiccation tolerance and maintenance of a quiescent state are associated with the presence of particular proteins

such as the LEA proteins, seed storage proteins and heat shock proteins (HSPs) (RAJJOU and DEBEAUJON, 2008). Our data shows, in association with the increase in longevity, an over-representation of terms associated with the synthesis of HSP (response to heat, protein-folding, response to ER stress). Indeed, 10 transcripts encoding small HSP were found in the top 30 highly expressed genes in dry seeds (Supplementary Table S1). This result is in agreement with previous transcriptome studies on soybean (JONES et al., 2010), *Medicago truncatula* (VERDIER et al., 2013) and *Arabidopsis thaliana* (KOTAK et al., 2007) in which late seed maturation was shown to be accompanied by an increase in HSP transcript levels. Using a gene co-expression network in developing Medicago seeds, Verdier et al (2013) found a module in which HSP were associated with longevity. It is likely that the proteins encoded by these genes are also synthesized because small HSP were found in the proteome in developing Arabidopsis seeds during late seed maturation (WEHMEYER and VIERLING, 2000). The role of HSP in protecting seeds against seed ageing was also demonstrated by Tejedor-Cano et al. (2010) who showed that a drastic reduction of small HSPs in transgenic tobacco seeds led to decreased stability during storage. Kaur et al (2015) revealed that small HSPs (OsHSP18.2 class II cytosolic HSP) have the ability to improve seed vigor and longevity by reducing deleterious ROS accumulation in Arapidopsis seeds. Interestingly Bettey et al (1988) reported that the amount of HSP correlated with seed germination performance in Brassica. Therefore, it would be interesting to evaluate whether the amount of HSP could be a marker for seed longevity in soybean.

Our transcriptome revealed also the presence of HSF, a family of transcription factors known to modulate the expression of HSP coding genes, such as two homologues of HSFA2 and one of HSFA6B. Seed-specific overexpression of the sunflower (*Helianthus annuus*) HSF *HaHSFA9* in tobacco (*Nicotiana tabacum*) led to increased stability against accelerated aging of seeds demonstrating the importance of such factors in seed longevity (PRIETO-DAPENA et al., 2006; PERSONAT et al., 2014). This transcription factor also interacts with the sunflower drought-responsive factor HaDREB2 (DEHYDRATION RESPONSIVE ELEMENT BINDING PROTEIN B2) in a seed-specific manner to enhance stability against accelerated aging (ALMOGUERA et al., 2009). Likewise in the legume *M. truncatula*, several members of the AP2/EREBP family were co-expressed with HSF in the longevity module (VERDIER et al. 2013). Our soybean transcriptome also showed several members of the AP/EREBP being differentially up and

down-regulated indicating that different members of this family might be required to regulate gene expression during soybean seed development. In Arabidopsis, ABI3 induces HSFA9 (KOTAK et al., 2007). Here, ABA was found to peak much mid-way through seed development shortly before the induction of desiccation tolerance. However, whether this peak is necessary to induce HSF remains to be investigated. More work is needed to understand the signaling pathways inducing HSF and longevity during maturation

Although the role of ABA in seed longevity is supported by numerous studies (OOMS et al. 1993; PARCY et al. 1994; MAO and SUN 2015) the implication of other hormones is less well documented. Auxin acts downstream of ABI3 to enhance seed longevity by alleviating HaIAA27-mediated HaHSFA9 repression in sunflower (CARRANCO et al. 2010). Longevity genes in *M. truncatula* were found to be enriched in binding sites for auxin-binding factors (RIGHETTI et al. 2015), implying that auxin also may have a role in seed longevity. Consistent with these observations, we found three members of the Auxin Responsive Factor family being up-regulated (homologues of Arabidopsis ARF19, ARF4 and ARF8) while none was found to be significantly down-regulated. Also, all members of AUX/IAA transcription factors were down-regulated or no significantly differentially expressed during the increase in seed longevity.

Besides the synthesis of HSP and the putative role of HSFs, our transcriptome raises the question whether red light influences seed longevity. We observed a high number of categories associated to “far-red light photoreceptor activity”, to “response to very low fluence”, to “red light stimulus”, to “photoreceptor activity” were up-regulated during the increase in longevity. A closer inspection of genes belonging to these categories revealed the presence of phytochrome A and PIF homologues (PIF7, Photoreceptor interacting factors). In our data, when mature and immature soybean seeds were imbibed in light and dark conditions, no difference in seed germination and hypocotyl elongation was found between mature and immature seeds (Figure 12) suggesting that regulatory factors controlling skotomorphogenesis and de-etiolation via the phytochromes are identical between stage R7.2 and R9. Flowering and maturity is highly controlled by major genes, in soybean two of them correspond two paralogous genes encoding a PhyA-like photoreceptor and strongly controls flowering time and time to physiological maturity according to the photoperiod (LIU et al., 2008). Therefore, our transcriptome could be showing a remnant of such behavior. Gene expression in developing soybean seeds is also controlled by the circadian clock with an organ-specific phasing, which could influence the

expression of light receptors (HUDSON et al., 2010). However, in our experiments, the seed material was always harvested rigorously at the same time of the day to avoid artifacts due to the operation of the circadian clock.

Developing seeds might respond to light cues such as photoperiod and light quality differently according to the stage of maturation. It is known that the light penetrating the fruit tissues in pea and soybean that reach the embryo is predominantly in the far red region (ALLEN et al., 2009). Interestingly, in lettuce seed longevity as assessed by accelerated aging was higher in seeds obtained from parent plants grown under a higher proportion of far-red light, indicating that seeds did respond to light and that red light was detrimental to longevity (CONTRERAS et al., 2009). Reports on Arabidopsis showed that transferring plants from different light environments during seed development strongly influenced seed germination (HAYES and KLEIN, 1974; He et al. 2014). A transfer before seed maturation had no significant effect, whereas a transfer during seed maturation induced a decrease in the germination in the dark, from 78 to 45 % when the transfer was performed when seeds started to dehydrate and from 78% to 0% when seed were dry. Further work is necessary to understand the role of light perception during maturation and a putative role of PhyA in the maturation of seeds.

Approximately 68% of the transcripts expressed decreased during the increase in seed longevity. The main categories that were over-represented are associated with photosynthesis (Table 11). This might be explained by the fact that chlorophyll degradation takes place in stage R7.2 and is fully degraded by stage R9 (Figure 6). In most species, chlorophyll degradation occurs at the final of seed maturation before the seed reaches the dry state (NAKAJIMA et al., 2012; TEIXEIRA et al., 2016). Chlorophyll retention in dry oily seeds appears to be detrimental to seed longevity. Seeds of the Arabidopsis *green-seeded* (*grs*) mutant contained twice more chlorophyll than wild-type seeds and exhibited reduced storability (CLERKX et al., 2003). Likewise, seeds of the *non-yellow coloring1 (nyc1)/nyc1-like (nol)* double mutant that are affected in the conversion of chlorophyll b to chlorophyll a, the first step of chlorophyll degradation, contained 10-fold more chlorophyll than wild-type and had a strongly reduced longevity (NAKAJIMA et al., 2012). However, the cause-effect relationship between longevity and chlorophyll remains elusive. It could be that the photosynthesis machinery needs to be degraded to avoid the potential damage imposed by the ROS produced by the photosynthetic electron transport chains during drying.

Although there are no differences in germination percentage between stage R7.2 and R9, it was possible to observe a variation of the transcripts according to drying process. Premature drying showed that an increase in transcripts negatively correlated with seed longevity, such as photosynthesis.

During the premature drying there is an increase in the level of transcripts associated with photosynthesis (Table 13) and all these transcripts are localized in the chloroplast. The chloroplasts are active metabolic centers that sustain life on earth by converting solar energy to carbohydrates through the process of photosynthesis and oxygen release. The chloroplast genome encodes many key proteins that are involved in photosynthesis and other metabolic processes, including biosynthesis of fatty acids, amino acids, pigments, and vitamins. The presence of these transcripts may be linked to the fact that drying of immature seeds leads to occurrence of green areas in the central regions of the cotyledons. Therefore, the permanency this small amount of chlorophyll may result in the high expression of these genes.

In general, loss of seed viability during drying may be accompanied by increased of lipid peroxidation and accumulation of free radicals. Several mechanisms related to the accumulation of free radicals indicate that the first production site is the chloroplast. In addition, an NAD(P)H dependent oxidase can be activated (VANDENABEELE et al., 2000) and lead to the production of superoxide. Antioxidants such as ascorbic acid and glutathione are essential for the protection of plants against oxidative stress (MITTLER, 2002). The cisteins, hydroquinones, mannitol, some carotenoires and b-carotene, are also important non-enzymatic antioxidants (SCANDALIOS, 1993). Here, during the premature drying there was increased level in transcripts encoding glutathione gamma-glutamylcysteinyltransferase-3 and glutatione peroxidase. These two enzymes play critical roles in protecting cells from oxidative damage and in maintaining redox homeostasis. Thus, this could be a symptom that premature drying imposes a stress to unripen seeds. Possibly premature drying increased the respiration rate of seeds and therefore was also observed increased in NADH dehydrogenase (ubiquinone) activity.

When we looked at the top down-regulated genes during premature drying, we noticed several transcripts involved in ethylene signaling. It is known the role of ethylene in ripening in the climatic fruits such as tomato (*Lycopersicon esculentum* Mill.), melon (*Cucumis melon* L.), and others crops (CHAVES and MELLO-FARIAS,

2006). It is still largely related to the vegetative growth, seed dormancy and germination (ABELES et al., 1992). However, the transcripts involved in signaling and rate of ethylene in soybean seeds in response to drying has not been well reported. Samarah et al. (2016) studied the ethylene evolution from soybean seed podded and depodded and relationship with tolerance to desiccation during seed maturation. They concluded that the release of ethylene decreased as tolerance to desiccation of seeds increased. Inhibition of the last step of the ethylene biosynthesis pathway resulting in a decrease in ethylene production may be required for seed desiccation as part of maturation phase (BOGATEK and GNIAZDOWSKA, 2012).

During the natural drying of soybean seeds we showed that increased transcripts associated with seed longevity, such as synthesis of protective HSP. The HSP are molecular proteins involved in response to stress. As mentioned before the function of the HSP has been related with the maintenance and repair of macromolecular structures during dehydration and rehydration, respectively (VERTUCCI and FARRANT, 1995). Although all organisms synthesize HSPs in response to heat balance synthesized proteins and the relative importance of individual families of HSPs in stress tolerance varies widely between organisms (QUEITSCH et al., 2000). Furthermore, natural drying induced a much greater loss of transcripts than artificial drying. We speculate that drying impacts of the processes leading to longevity by acting on the turnover of these transcripts. How this is achieved remains to be investigated. However, there is evidence showing that the rate of drying is important for seed longevity. For example, very recently, it was reported that in rice, seed longevity can increase 3-fold when immature seeds are dried fast using a warm air flow compared to regular drying (WHITEHOUSE et al. 2015).

5.3. TRANSCRIPTOME COMPARISONS OF IMMATURE SEEDS DRIED (STAGE R7.2d) AND MATURE SEEDS (STAGE R9) DURING IMBIBITION

5.3.1. Increased transcripts during imbibition

Germination is a complex process during which the seed must quickly recover from maturation drying, resume metabolism and complete essential cellular events to allow the embryo to emerge, and prepare for subsequent seedling growth (NONOGAKI et al., 2010). In soybean seeds, the molecular events leading to these

processes are far from being completely known. It has been shown that during late maturation, seeds are able to synthesize and store mRNAs. These stored mRNAs are thought to be used during seed imbibition to ensure germination (RAJJOU et al., 2004; GALLAND and RAIJOU, 2015). In our study, immature dried seeds and mature seeds did not differ in germination (Figure 9). Consistent with these observations, we found that the vast majority of transcripts in stages R7.2 and R9, representing 82% of the total DEG during imbibition, followed the same trend in both stages during imbibition, suggesting that the main factors regulating germination are probably already set in seeds at R7.2.

When imbibing seeds where compared to dried seeds and we observed an over-representation of biological processes associated with growth of different tissues (root tip, root hair, meristem, cell wall elongation; Table17). This probably reflects the activation of genes that are necessary for radicle growth during imbibition. Transcripts of these genes were very low in dried seeds regardless of the maturation stages and were not differentially expressed during drying. Therefore, it is difficult to reconcile these observations with the idea that mRNAs are stored in the dry state to prepare the seeds for germination and, because of that there is no need for transcriptional activity. In Arabidopsis, 500 μM α-amanitin did not inhibit the percentages of germination (RAJJOU et al., 2004). However, it decreased the speed of germination by 22% in wild-type seeds and about 350% in several testa mutants in Arabidopsis. Also, treated seeds originated radicles that were not vigorous compared to untreated seeds (RAJJOU et al., 2004). There is increasing evidence that the seed coat is essential for Arabidopsis germination by regulating ABA and GA signaling pathway in the embryo. Many reports showed that the endosperm layer is responsible for most, if not all, of the germination-repressive activity (BETHKE et al., 2007; LEE et al., 2012); KANG, 2015). Altogether, these data on Arabidopsis suggest that gene activation in the envelope is necessary for germination. In soybean, the testa is dead after maturation drying (RANATHUNGE et al., 2010). Therefore, regulation of seed germination must occur within the embryo. Because of the differences we observed between the transcript profile observed in imbibing seeds and the transcripts stored in dry seeds, we speculate that in contrast to Arabidopsis, gene activation is necessary for soybean seed germination. An experiment to check if α-amanitin or other transcription inhibitors have an effect on soybean germination would be of interest to clarify whether transcription is necessary for visible seed germination (radicle protrusion)

or if the fact that Arabidopsis seeds can germinate in the absence of transcription is just an exception.

Genes involved in lipid reserve mobilization appear to be activated early during imbibition in immature seeds (R7.2) and mature seeds (stage R9). Among the top differentially expressed transcripts during imbibition (Supplementary Table 4), we found transcripts encoding Phosphoenolpyruvate carboxykinase, two probes encoding an isocitrate lyase and several probes encoding lipases whose levels increased 50-100 fold in 18h imbibed seeds. In oily seeds, ATP-dependent phosphoenolpyruvate carboxykinase plays a fundamental role in the provision of sucrose during germination and early post-germinative seedling growth. This enzyme is important because it catalyzes the conversion of oxaloacetate to phosphoenolpyruvate in the gluconeogenic production of sugars from storage oil in germinating oilseeds (RYLOTT et al 2003).

In a more detailed analysis of GO enrichment (Table 17) five probes encoding five different ethylene receptors, including an Arabidopsis homologue of ESR1 were found to be up-regulated. Further inspection of the transcriptome revealed the many more transcripts involved in the ethylene-signaling pathway such as several homologues of ETHYLENE-RESPONSIVE TRANSCRIPTION FACTORS (including RAP2-3), ETHYLENE RESPONSIVE ELEMENT BINDING FACTOR 3 (EREBF3), EREBF1, ETHYLENE-OVERPRODUCTION PROTEIN 1, ETHYLENE OVERPRODUCER-LIKE PROTEIN (ETO1 LIKE-1), a negative regulator of ethylene synthesis in Arabidopsis. We also found that several probes of ACC SYNTHASE and ACC OXIDASE, two genes involved in the biosynthesis of ethylene were also strongly up-regulated, suggesting that ethylene was produced during germination in imbibing soybean seeds. Ethylene participates in the germination of seeds, for example, lettuce and chick-pea (MATILLA and MATILLA-VAZQUEZ, 2008). Since studies using inhibitors of ethylene biosynthesis or of ethylene action and analysis of mutant lines altered in genes involved in the ethylene signaling pathway (*etr1, ein2, ain1, etr1, and erf1*) demonstrate the involvement of ethylene in the regulation of germination and dormancy (reviewed in CORBINEAU et al. 2014). Ishibashi et al (2013) confirmed that soybean seed germination was regulated through ethylene production in response to reactive oxygen species (ROS). However, apart for several transcripts encoding a L-ascorbate oxidase-like protein, our transcriptome did not reveal that seeds were submitted to ROS. Our data suggest that ethylene plays an important role in soybean seed germination.

Ethylene is known to regulate germination and dormancy by controlling GA and ABA signaling pathway (reviewed in CORBINEAU et al. 2014). GA promotes germination whereas ABA inhibits it. Consistent with this, we found several transcripts associated with GA signaling and GA biosynthesis being upregulated and ABA signaling and synthesis being down-regulated. GA-related transcripts encoded several GA-regulated proteins of unknown function, GA-20 OXIDASE and GA-3 OXIDASE and the GA receptor GID1. Down-regulated ABA related transcripts were homologues of the signaling pathway (OST1, a member of ABA activated SNF1-related protein kinases (SnRK2), a PP2CA (PROTEIN PHOSPHATASE 2CA), ABI5, APF2, a bZIP factor interacting with ABI5, and KEG involved in degrading ABI5. Further work to determine the existence of a cross-talk between ethylene, GA and ABA in the regulation of soybean seed germination are of interest. We also found several Arabidopsis homologues of the ABA signaling pathway being up-regulated during imbibition, such as several receptors of ABA (including RCAR10, REGULATORY COMPONENTS OF ABA RECEPTOR 10, PYR1, PYR1-like), ABA signaling proteins such as ABI2 and ABI5-like and an enzyme involved in ABA synthesis (ABA4). Although the ABA signaling pathway bears the same characteristics between seeds and vegetative tissues, the components are different. Therefore, we speculate that during imbibition, soybean seeds replace its seed-specific ABA signaling pathway by a vegetative-specific signaling pathways. This is supported in our transcriptome by the upregulation of several CALCIUM-DEPENDENT PROTEIN KINASE (CPK2, CPK13, CDPK-RELATED KINASE 1).

Another feature of the transcriptome of germinating soybean seeds is the over-representation of processes related to organ formation, in particular in relation to light (red, far-red and blue), and meristem activity, such as flower morphogenesis, organ morphogenesis, cotyledon morphogenesis. Many of these transcripts were related to auxin signaling and auxin transport such as AUXIN RESPONSE FACTOR 5 (ARF5, Monopteros, involved in reprogramming cell identity during embryonic development, ARF8, auxin efflux carriers (PIN1 and PIN3). It also contained a homologue of HB-3, a WUS type homeodomain protein that is required for meristem growth and development and acts through positive regulation of WUS. This is intriguing, considering that embryogenesis has terminated before stage R5. Also, whole seeds were used for RNAseq, with the effect of strongly diluting transcripts that are specific to meristems. Therefore, these genes may play additional auxin-related roles as suggested by the Arabidopsis TAIR

annotation. Auxin by itself is not necessary for seed germination but may participate to the fine-tuning different aspects of cellular growth necessary for germination in relation with environmental cues such as light, gravitropism. For example, in Arabidopsis, AUXIN RESPONSE FACTOR10 (LIU et al., 2013) and ARF 13 regulate germination by controlling the expression of ABI3 during imbibition. Whether the identified genes in soybean play similar role remains to be determined.

Intriguingly, we found several transcription factors involved in similar process related to organ formation and meristem activity, which were significantly up-regulated in mature seeds but not in immature seeds during imbibition. This is the only profiles containing a significant enrichment of TF (Table 21). This includes a Class I KNOX box homeobox factor whose homolog in Arabidopsis is STM/BUM1. This TF is required for shoot apical meristem (SAM) formation during embryogenesis and for SAM function throughout the lifetime of the plant. We also noted the presence of GRAS-type TF, whose homologue is SGR7, (SHOOT GRAVITROPISM 7) and is essential for normal shoot gravitropism. There was also a transcript annotated as ABI3-like. In Arabidopsis, this AP2/B3-like transcriptional factor family is apparently not related to ABA signaling. Expressed in the lateral organ boundary region, it operates in a network of regulatory genes controlling leaf serration being all these processes regulated by auxin fluxes.

5.3.2. Decreased transcripts during imbibition

Transcript associated with phytate synthesis disappeared rapidly during imbibition. The phytic acid (myo-inositol-1-phosphate) is the main source of phosphorus stored in seeds during development (RABOY, 1987). In soybean seeds, phytic acid is deposited in protein bodies as a complex of chelated minerals and proteins known as phytin (PRATTLEY and STANLEY, 1982). Phytin represents approximately 1% of the total weight in dry seeds (HEGEMAN and GRABAU, 2001). Adequate levels of phosphorus are critical to the growth and development of all organisms for a range of functions such as macromolecular structure, energy generation, and metabolic regulation. While the demand for phosphorus increases dramatically during periods of rapid cell growth and division, such as seed germination (HEGEMAN and GRABAU, 2001), we found no evidence in our transcriptome of an activation of genes involved in phytin degradation. Probably, this occurs later after germination.

In the next step, we analyzed differential expression profiles during imbibition. In a more detailed inspection in profile L (Figure 15, Table 22) we noted that this profile contained mainly chloroplastic genes involved in transcription and elongation with the chloroplasts. These include ribosomal genes, genes encoding proteins functioning in the photosystem1 and electron transport chain. This shows that these transcripts were probably not degraded during maturation drying. Paradoxically, while chloroplasts will be resynthesized later after germination during the greening process, left-over transcripts associated with photosynthesis are first degraded even if they could be reused later. This is stressed by the up-regulation of transcripts encoding photosynthesis genes during imbibition of R9 seeds (such as PHOTOSYSTEM II SUBUNIT Q-2, PHOTOSYSTEM II REACTION CENTER PSB29 PROTEIN, PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 3 PHOTOSYSTEM I SUBUNIT PSAD). The same line of argument applies for transcripts encoding LEA and HSP proteins that are dramatically decreasing during imbibition, while it is known that osmotic stress reducing desiccation tolerance or priming induces the accumulation of these proteins (BOUDET et al., 2006, BUITINK et al., 2006).

5.3.3. Are repair processes associated with seed longevity?

The only phenotypic difference between harvested seeds at stage R7.2 and mature seeds, stage R9 is related to a reduced storability that was found in seeds harvest in stage R7.2. So, we hypothesized that a transcriptome approach comparing changes during early imbibition of two seeds lots differing in longevity could give access to repair mechanisms that would be synthesized during imbibition of seeds from stage R9. We previously mentioned that genes involved in meristem development and morphogenesis were highly expressed in imbibed seeds at stage R9 but not in immature seeds. For example, STM/BUM1 I KNOX KNOTTED1-like homeobox factor was strongly upregulated as confirmed by RT-qPCR. This gene is an important and conserved transcriptional mediator of meristematic potential (reviewed in SCOFIELD and MURRAY, 2006). It is essential for initial shoot meristem formation and maintenance of SAM organization. Since it is associated with the regulation of stem cell niches, we could speculate that inducing the expression of such regulator could be a preemptive measure to ensure proper organ formation during germination in case the meristem would be

damaged. Indeed, the expression of such genes could play an essential role in the regeneration of organs that could have suffered from deterioration during storage. Consistent with this, we also identified and confirmed by RT-QPCR the upregulation of the homologue of the AUXIN RESPONSE FACTOR 5, ARF5/MONOPTEROS. Its function, in conjunction with an auxin signal is to promote primordium founder cell fate to initiate flowers. Likewise, a transcript encoding a HD domain class transcription factor, which homologue in Arabidopsis play a similar role to KNOTTED-1 and is regulated by auxin was found. If we assume that these transcription factors are synthesized as a preemptive measure, then auxin should play a role in seed longevity. Evidence for this hypothesis comes from a recent work by Righetti et al (2015) in which they constructed a co-expression network of seed longevity. These authors reported that genes, which are present in the cross-species longevity module were highly enriched in binding sites for auxin response factors (ARFs) in their promoters. Instead of an expected 6% of genes, 60% of the *M. truncatula* genes in the longevity module contained the ARFAT cis-element (RIGHETTI et al 2015).

The same argument of a pre-emptive measure against potentially damaged tissues applied for the homologue of HSFA9 that was strongly upregulated in imbibing mature seeds (Figure 19). It could facilitate the production of chaperone to repair proteins that were damage during storage. Interestingly PMIT (PROTEIN-L-ISOASPARTATE METHYLTRANSFERASE 1), an ABA induced protein involved in repairing protein in aged seeds was down-regulated during imbibition

Our data from profile G also provides evidence of DNA repair that could be induced during imbibition. First, DNA2, a member of DNA replication helicase family protein (*Glyma.*06G310500), which Arabidopsis homologue was recently found to be associated with DNA repair and maintenance of the meristem. During DNA break repair, DNA2 mediates 5'-end resection of DNA by cleaving the 5'-single stranded DNA. The second probe is *Glyma.*18G194300, the arabidopsis homolog of a Werner syndrome-like exonuclease whose expression was significantly up-regulated in R9 during imbibition but not stage R7.2 (Figure 16). Only a limited amount of information on this type of exonuclease is available in plants. This enzyme is thought to channel DNA double stand-break repair processes into non-recombinogenic pathways, thus facilitating genomic stability. Consistent with these observations, our profile also contained a DEG encoding a MMS19 nucleotide excision repair protein-like protein (*Glyma.*13G354600).

5.4. COMPARATIVE ANALYSIS BETWEEN SOYBEAN AND *MEDICAGO TRUNCATULA* TO IDENTIFY GENES RELATED TO SEED VIGOR ACQUISITION DURING MATURATION

High seed quality is crucial for seedling establishment and yield. Seed quality characteristics such as germinability and longevity are acquired during seed maturation, in both soybean and Medicago. Our data on soybean revealed that complex transcriptomic features governing seed longevity and RNA metabolism during seed imbibition were dependent on the maturity stage. However, a cause effect relationship between transcriptome changes and expression of seed vigor in relation with the level of maturation was not secured. This was addressed here by a comparative analysis between soybean and *Medicago truncatula* at the same physiological stages

The first similarity that caught our attention was the presence of NFXL-1 that was up-regulated for both, soybean and medicago during maturation. Recently, Righetti et al. (2015) discovered the presence of NFXL-1 in an expression profile during seed maturation of both Medicago and Arabidopsis. This gene was part of the longevity module and a mutant affected in NFXL-1 displayed decreased longevity (RIGHETTI et al.,2015). Our analysis corroborate this observation and suggest that the function of this gene in regulating seed longevity is conserved between species. This gene, also present in humans encodes a NFX-1-type nuclear Zn-finger transcriptional repressor that binds a conserved cis-acting element, the X-box in promoters of class II MCH genes. Little is known regarding its function. In Arabidopsis, it promotes growth under salt and osmotic stress by activating the expression of ABA stress responsive genes such as LEA genes and H_2O_2 signaling (LISSO et al., 2006). Further work to understand the role of NFXL-1 in seed vigor would be of interest.

As mentioned above, during the late stage of maturation, seeds are able to synthesize mRNA in preparation for the dry state. These stored mRNAs are thought to be used during seed imbibition to ensure germination (RAIJOU et al., 2004; GALLAND and RAIJOU, 2015). To test whether this is also the case in soybean we compared the transcriptome of soybean imbibing seeds with results found in the analysis of medicago seeds. Analysis of transcripts corresponding to DEG up-regulated during legume seed imbibition revealed genes highly overrepresented that are involved in phototropism. Two genes were identified in the overrepresented group related to phototropism: the

phototropic-responsive NPH3 gene, which is a highly conserved gene between Arabidopsis (AT5G64330), *Medicago truncatula* (two genes: Medtr2g009340.1 and Medtr4g093850.5) and soybean (Glyma.07G180900), and a Phototropin 1/NPH1 gene (AT3G45780, Medtr4g061610, Glyma.12G074100), which is known to interact with the blue light photoreceptor NPH1. The NPH1/NPH3 signaling pathway stimulates the shoot ward auxin flux by modifying the subcellular targeting of PIN2 in the root apex transition zone, leading to blue-light-induced root phototropism (WAN et al., 2012). Considering that PHYA and PHYB were also revealed by our species comparison, it would be interesting to further investigate the link between light photoreception activity and seed vigor

Seed germination and innate immunity response both play a significant role in the plant life. Since it is necessary to provide defense mechanisms against various possible tensions that the seed/seedling can experience in the production field. Our data show an over-representation of GO terms involved in innate immune response in imbibing seeds. Seven genes that represent this GO were found. Among them it is worth mentioning the SnRK1 gene, which plays an important role in plant growth as a key regulator ofsugar signaling, stress, and seedling growth (XUE-FEI et al.; 2012).

Raffinose synthase Glyma.06G179200 was in the list of the 52 DEG transcripts that were down-regulated during legume seed imbibition, The content of RFO increased during soybean seed maturation and has been traditionally associated with desiccation tolerance (OBENDORF et al., 2009; VERDIER et al., 2013). RFOs are suggested to protect cellular integrity during desiccation by stabilizing membranes during dehydration. Whether RFO are an important source of energy during seed germination is unclear. The inhibition of the breakdown of RFOs caused significant delay in germination (BLÖCHL et al., 2007). However, according to DIERKING and BILYEU (2009) the reduced amount of raffinose and also stachyose, did not reduce or delay soybean seed germination.

It has been mentioned here the importance of the ABA, as a regulator of growth and development of plants, that participates in the induction and maintenance of seed dormancy (RAJJOU et al., 2012). ABA is required for the seed maturation program, acting via the master regulator TFs to control storage and late embryogenesis accumulating (LEA) protein deposition and the acquisition of desiccation tolerance (VERDIER et al., 2013). The decrease in ABA concentration after seed imbibition allows the progress of germination (NONOGAKI et al., 2010). Thus, during

germination, the seed must establish a specific catabolism, reducing their sensitivity and performing the biosynthesis of inhibitors to reduce ABA activity levels (RAJJOU et al., 2012; ARC et al., 2013).

The ABA catabolism may be performed by oxidation, or conjugation with glucose. The main catabolic route is through 8'-hydroxylation by cytochrome P450 monooxygenase subfamily of CYP707A (ARC et al., 2013). Some genes are directly related to the ABA synthesis, such as ABA 1, 9-cis-epoxycarotenoid dioxygenase (NCEDs) and ABA2/G1N1/SDR1 (BENTSINK et al., 2006; HOLDSWORTH et al.; 2008), which mutants exhibits low ABA concentrations during seed development, resulting in no dormancy seeds (HOLDSWORTH et al.; 2008).

The ABA accumulation is controlled by synthesis, metabolism and transportation. The NCEDs is an enzyme of the rate-limiting step in ABA biosynthesis. Another important enzyme for ABA metabolism is cytochrome P450 type enzyme CYP707A (reviewed in NAMBARA and MARION-POLL 2005). The metabolism is mainly controlled by the expression of the CYP707A gene in the stomata and vascular tissues of leaves under rehydrated conditions (OKAMOTO et al. 2009).

Considering the major roles of ABA and GA in seed dormancy and germination (BEWLEY and BLACK., 1994; NONOGAKI et al., 2007), in our transcriptome study during imbibition of *Medicago truncatula* mature seeds, were found genes related to ABA signaling and metabolism, such as zeaxanthin epoxidase (ZEP) (ABA1), CYP707A1, nine-cis-epoxycarotenoid dioxygenase and ABA-responsive element binding protein. In contrast, we found a number of genes involved in GA signaling, such as GA20OX2, GRAS family transcription factor family protein and GRAM domain family protein.

Seed germination is often associated with a sharp drop in the ratio of ABA and GA. The ABA inhibitory effect on germination can be reversed by GA_3, when used in concentrations that exceed their content, but from a critical level of ABA their action prevails inhibitory (CUNHA and CASALI, 1989).

6. CONCLUSIONS

- ✓ The capacity to germinate was progressively acquired during early seed filling, between stages R5.5 and stage R6. ABA levels increased from stage R5.1 until R6, corresponding to a maximum. Thereafter, it declined progressively;

- ✓ Desiccation tolerance was acquired at stage R7.2. All parameters used to assess seed vigor indicated that maturity was obtained at stage R7.2. Nevertheless, longevity was still not fully acquired at this stage as it nearly doubled between stage R7.2 and R9;

- ✓ The transcriptome analysis shows that there are differences at the molecular level between seeds of stage R7.2 (immature) and R9 (mature). Soybean mature seeds revealed a significant over-representation of genes related to response to heat, protein-folding, response to ER stress and genes related to light;

- ✓ Degradation and synthesis of transcripts during imbibition are similar between immature and mature seeds;

- ✓ Radicle growth competence through transcriptional regulation is activated early during seed imbibition;

- ✓ Transcript associated with ABA-induced stress response, Chl degradation, phytate synthesis and chaperone function disappear rapidly during imbibition;

- ✓ Chloroplast-encoded transcripts that should disappear during maturation are rapidly degraded during imbibition of immature seeds;

- ✓ Imbibition-induced transcriptome profile associated with longevity highlights a putative DNA repair;

- ✓ The comparison of the transcriptome profile between soybean and *Medicago truncatula* during maturation showed a significant over-representation of genes such as PHY A, PHYB, NFXL-1, ABI1-like 1, and SCARECROW-like 14

.

7. REFERENCES

ABELES, F. B.; MORGAN, P. W.; SALTVEIT JR, M. E. **Ethylene in Plant Biology**, 2. ed. New York: Academic Press, 1992.

ACKERSON, R. C. Abscisic acid and precocious germination in soybeans. **Journal Experimental Botany**, Lancaster, v. 35, p. 414-421, 1984.

ADAMS, C.; FJERSTAD, M. C.; RINNE, R. W. Characteristics of soybean seed maturation: necessity for slow dehydration. **Crop Science Society of America**, Madison, v. 23, n. 2, p. 265-267, 1982.

AGHAMIRZAIE, D. et al. Changes in RNA splicing in developing soybean (*Glycine max*) embryos. **Biology**, Switzerland, v. 2, p. 1311-1337, 2013.

ALDRICH, S. R. Maturity measurements in corn and an indication that the grain development continues after premature cutting. **Journal of the American Society of Agronomy**, Madison, v. 35, n. 8, p. 667-680, 1943.

ALLEN, D. K.; OHLROGGE, J. B.; SHACHAR-HILL, Y. The role of light in soybean seed filling metabolism. **The Plant Journal**, Chichester, v. 58, n. 2, p. 220-234, 2009.

ALMOGUERA, C. et al. The HaDREB2 transcription factor enhances basal thermotolerance and longevity of seeds through functional interaction with HaHSFA9. **BMC Plant Biology,** London, v. 9, n. 75, 2009.

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biology**, v. 11, R106, 2010.

ANDERSON, S. R. Development of pod and seed of birds foot trefoil as related to maturity and to seed yields. **Agronomy Journal**, v. 47, n. 10, p. 483-487, 1955.

ARC, E. et al. ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. **Frontiers in Plant Science.** v. 4, n. 63, 2013.

ASAKURA, T. et al. Global gene expression profiles in developing soybean seeds. **Plant Physiology Biochemistry**, v. 52, p.147-153, 2012.

BASBOUSS-SERHAL, I. et al. Germination potential of dormant and nondormant arabidopsis seeds is driven by distinct recruitment of messenger RNAs to polysomes. **Plant Physiology**, v. 163, n. 3, p. 1049-1065, 2015.

BAUD, S. et al. An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. **Plant Physiology and Biochemistry**. v. 40, n. 2, p.151-160, 2002.

BAUD, S. et al. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. **The Plant Journal**, v. 50, n. 5, p. 825-838, 2007.

BAZIN, J. et al. Role of relative humidity, temperature, and water status in dormancy alleviation of sunflower seeds during dry after-ripening. **Journal of Experimental Botany**, v. 62, n. 2, p. 627-640, 2011.

BELLALOUI, N. Soybean seed phenol, lignin, and isoflavones partitioning as affected by seed node position and genotype differences. **Food and Nutrition Sciences**, Stoneville, USA, n. 3, p. 447-454**,** 2012.

Bentsink, L. et al. Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. **Proceedings of the National Academy of Sciences of the United States of America**. v. 103, n. 45, p. 17042–17047, 2006.

BERGAMIN, M.; CANCIAN, M. A. E.; CASTRO, P. R. C. Ecofisiologia da soja. In: CASTRO, P. R. C.; KLUGE, R. A. (Org.). **Ecofisiologia de cultivos anuais**: trigo, milho, soja, arroz e mandioca. São Paulo: Nobel, p. 73-90. 1999.

BETHKE, P. C. et al. The arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. Plant Physiology. v. 143, n. 3, p. 1173-1188, 2007.

BETTEY, M.; FINCH-SAVAGE, W. E. Stress protein content of mature *Brassica* seeds and their germination performance**. Physiology & Biochemistry**, v. 8, n. 3, p. 347-355, 1998.

BEWLEY, J. D. et al. **Seeds: physiology of development, germination and dormancy.** 3rd ed. New York: Springer, 2013.

BEWLEY, J. D. Seed Germination and dormancy. **The Plant Cell**, v.9, p.1055-1066, 1997.

BEWLEY, J. D.; BLACK, M. **Physiology and biochemistry of seeds in relation to germination**. Development, germination and growth. v.1, Berlim: Springer Verlag, 1978, 306p.

BEWLEY, J. D.; BLACK, M. **Seeds**: physiology of development and germination. New York: Plenum Press, 1994. 445 p.

BLACKMAN, S.A. et al.. Maturation proteins associated with desiccation tolerance in soybean. **Plant Physiology**. v.96, p.868-874, 1991.

BLÖCHL, A.; PETERBAUER, T.; RICHTER, A. Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. **Journal of Plant Physiology**, v. 164, p. 1093-1096, 2007.

BOGATEK, R.; GNIAZDOWSKA, A. Ethylene in seed development, dormancy and germination. **Annual Plant Reviews**, v. 44, p. 189-218, 2012.

BOTHA, F. C.; POTGIETER, G. P.; BOTHA, A. M. Respiratory metabolism and gene expression during seed germination. **Journal of Plant Growth Regulation**, v. 11, n. 3, p. 211-224, 1992.

BOUDET, J. et al. Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. **Plant Physiology,** v. 140, n. 4, p. 1418-1436, 2006.

BOZATO, R; BOZATO A. L.V. **A soja no Brasil**: história e estatística. Londrina, EMBRAPA-CNPSo, 1987. 61p. (Documentos, 21).

BRASIL. Ministério da Agricultura e Reforma Agrária. **Regras para análise de sementes**. Brasília, DF: SNAD/DNDV/CLAV, 2009. 398 p.

BRAY, E.A. Molecular responses to water deficit. Plant Physiology. v.103, p. 1035-1040, 1993.

BREVEDAN, R. E.; EGLI, D.B. Short periods of water stress during seed filling, leaf senescence and yield of soybean. **Crop Science**. v. 43, p. 2083-2088, 2003.

BUCKERIDGE, M. S.; SANTOS, H. P.; TINÉ, M. A. S. Mobilisation of storage cell wall polysaccharides in seeds. **Plant Physiology and Biochemistry.** v. 38, p.141-156, 2000.

BUITINK, J. et al. Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. **The Plant Journal,** v. 47, n. 5, p. 735-750, 2006.

BUITINK, J.; LEPRINCE, O. Glass formation in plant anhydrobiotes: survival in the dry state. ***Cryobiology***, v. 48, n. 3, p. 215-228. 2004.

BUITINK, J.; LEPRINCE, O. Intracellular glasses and seed survival in the dry state. ***Comptes Rendus Biologies***, v. 331, n. 10, p. 788-795. 2008.

CÂMARA, G. M. S.; HEIFFIG, L. S. Fisiologia, ambiente e rendimento da cultura da soja. In: CÂMARA, G.M.S. **Soja**: tecnologia da produção II. Piracicaba: G.M.S. CÂMARA, 2000. p.81-120.

CAPELETI, I. et al. A new procedure for quantification of lignin in soybean (*Glycine max* (L.) Merrill) seed coat and their relationship with the resistance to mechanical damage, **Seed Science & Technoly**. v. 33, p. 511-515, 2005.

CARBONELL, S. A. M.; KRZYZANOWSKI, F. C. The pendulum test for screening soybean genotypes for seeds resistant to mechanical damage. **Seed Science and Technology**, v. 23, p. 331–339, 1995.

CARRANCO, R. et al. Repression by an auxin/indole acetic acid protein connects auxin signaling with heat shock factor-mediated seed longevity. **Proceedings of the National Academy of Sciences**, v. 10, n. 50, p. 21908-21913, 2010.

CARRERA, C. et al. Environmental variation and correlation of seed components in nontransgenic soybeans: protein, oil, unsaturated fatty acids, tocopherols, and isoflavones. **Crop Science Society of America**, v. 51, n. 2, p. 800-809, 2011.

CARVALHO, N. M.; NAKAGAWA, J. **Sementes**: ciência, tecnologia e produção. Jaboticabal: Funep, 2000. 588 p.

CENTRO DE PESQUISA METEOROLÓGICAS E CLIMÁTICAS APLICADAS A AGRICULTURA - CEPAGRI. **Clima dos municípios paulistas**: Botucatu. Disponível em: <http://www.cpa.unicamp.br/outras-informacoes/clima_muni_086.html.>.

CHACHALIS, D.; SMITH, M.L. Imbibition behavior of soybean (*Glycine max* (L.) Merril) accessions with different testa characteristics. **Seed Science andTechnology**, v. 28, n. 2, p. 321-331, 2000.

CHATELAIN, E. et al. Evidence for participation of the methionine sulfoxide reductase repair system in plant seed longevity. **PNAS,** v. 110, n. 9, p. 3633-3538, 2013.

CHATELAIN, E. et al. Temporal profiling of the heat stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. **Plant Cell and Environment**, v. 35, p. 1440-1455, 2012.

CHAVES, A. L. S.; MELLO-FARIAS, P. C. D. Ethylene and fruit ripening: from illumination gas to the control of gene expression, more than a century of discoveries. **Genetics and Molecular Biology**, v. 29, p. 508-515, 2006.

CHEN, F.; NONOGAKI, H.; BRADFORD, K, J. A gibberellin-regulated xyloglucan endotransyglycosylase gene is expressend in the endosperm cap during tomato seed germination. **Journal of Experimental Botany**, Oxford, v. 53, n. 367, p. 215-223, 2002.

CHEN, H. et al. Transcriptome-wide mapping of pea seed ageing reveals a pivotal role for genes related to oxidative stress and programmed cell death. **Plos One,** v. 8, n. 10, p. e78471, 2013.

CLERKX, E. J. M. et al. Characterization of *green seed*, an Enhancer of *abi3-1* in Arabidopsis That Affects Seed Longevity. Plant Physiology, v. 132, n. 2, p. 1077–1084, 2003.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileira**: grãos, quinto levantamento. Brasília, DF, 2016. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_02_04_11_21_34_boletim_graos_fevereiro_2016_ok.pdf>.

CONTRERAS, S. et. al. Red to far-red ratio during seed development affects lettuce seed germinability and longevity. **HortScience**, v. 44, n. 1, p. 130-134. 2009.

COOK, D. R. *Medicago truncatula*— a model in the making!: Commentary. **Current Opinion in Plant Biology**, v. 2, n. 4, p. 301-304, 1999.

CORBINEAU, F. et al. Ethylene, a key factor in the regulation of seed dormancy. **Front. Plant Science**, v.5.p.1-5, 2014.

COSTA, N. P. et al. Qualidade fisiológica, física e sanitária de sementes de soja produzidas no Brasil. **Revista Brasileira de Sementes**, v. 25, n. 1, p. 128-132, 2003.

COSTA, N. P. et al. Zoneamento ecológico do Estado do Paraná para a produção de sementes de cultivares precoces de soja. **Revista Brasileira de Sementes**, v. 16, n. 1, p. 12-19, 1994.

CROWE, A. J. et al. The seed-specific transactivator, ABI3, induces oleosin gene expression. **Plant Science**, v. 151, n. 2, p. 171-181, 2000.

CUNHA, R.; CASALI, V. W. D. Efeito de substâncias reguladoras de crescimento sobre a germinação de sementes de alface. **Revista Brasileira de Fisiologia Vegetal**, Londrina, v. 9, n. 2, p. 121-132, 1989.

DEBEAUJON, I.; LÉON-KLOOSTERZIEL, K. M.; KOORNNEEF, M. Influence of the Testa on Seed Dormancy, Germination, and Longevity in Arabidopsis. **Plant Physiology**, v. 122, p. 403-413, 2000.

DELLAGOSTIN, M. et al. Dissimilaridade genética em população segregante de soja com variabilidade para caracteres morfológicos de semente. Revista Brasileira de Sementes**, v. 33, n. 4, p. 689 - 698, 2011.**

DELOUCHE, J. C.; BASKIN, C. C. Accelerated aging techniques for predicting the relative storability of seed lots. **Seed Science Technology**. v.1, p. 427-452, 1973.

DIERKING, E. C.; BILYEU, K. D. Raffinose and stachyose metabolism are not required for efficient soybean seed germination. **Journal of Plant Physiology,** v. 166, n. 12, p. 1329-1335, 2009.

DU, Z. et al. AgriGO: a GO analysis toolkit for the agricultural community. **Nucleic Acids Research**, v. 38, W64-70, 2010.

ELLIS, R. H.; HONG, T. D.; JACKSON, M. T. Seed production environment, time of harvest, and the potential longevity of seeds of three cultivars of rice. **Annals of Botany**, v. 72, p.583-590, 1993.

ELLIS, R. H.; HONG, T. D.; ROBERTS, E. H. The development of desiccation tolerance and maximum seed quality during seed maturation of six grain legumes. **Annals of Botany**, v. 59, p. 23-29, 1987.

ELLIS, R. H.; PIETRA FILHO, C. The development of seed quality in spring and winter cultivars of barley and wheat. **Seed Science Research**, v. 2, n. 1, p. 9-15, 1992.

EMBRAPA, EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Centro Nacional de Pesquisa de Solos. **Sistema brasileiro de classificação de solos.** Brasília, DF, 2006.

EMBRAPA, EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Tecnologias de produção de soja:** região central do Brasil: 2011. Londrina, 2011. 255 p.

FARIAS, J.R.B. et al. Caracterização de risco de déficit hídrico nas regiões produtoras de soja no Brasil. **Revista Brasileira de Agrometeorologia**. Passo Fundo, v.9, n.3, p.415-421, 2001.

FARRANT, J. M. et al. An overview of mechanisms of desiccation tolerance in selected angiosperm resurrection plants. **Plant Stress**, v. 1, p. 72-84, 2007.

FEHR, W. R.; CAVINESS, C. E.; **Stages of soybean development**. Ames: Iowa State University of Science and Technology, 1977. 11 p.

FERREIRA, A. G.; BORGHETTI, F. **Germinação**: do básico ao aplicado. Porto Alegre: Artmed, 2004. 323 p.

FINCH-SAVAGE, W.E. et al. Seed dormancy release in Arabidopsis *Cvi* by dry after-ripening, low temperature, nitrate and light shows common quantitative patterns of gene expression directed by environmentally specific sensing. **Plant J**. v. 51, n. 1, p. 60–78, 2007.

FINKELSTEIN R. et al. Molecular aspects of seed dormancy. **Annual Review of Plant Biology**. n. 59, p. 387–415, 2008.

FINKELSTEIN, R. et al. Abscisic Acid Signaling in Seeds and Seedlings. **The Plant Cell**, v. 14, n. 1, S15-S45, Supplement 2002.

FRANÇA NETO, J. B. et al. **Tecnologia da produção de semente de soja de alta qualidade**. Londrina: Embrapa, 2007. 22 p. (Circular técnica, 40).

FRANÇA NETO, J. B. et al.**Tecnologia da produção de semente de soja de alta qualidade**. Embrapa- CNPSo, Londrina, 2007. 22 p. (Embrapa-CNPSo, circular técnica, 40).

FRANÇA NETO, J. B. Qualidade fisiológica da semente. In: FRANÇA NETO, J. B; HENNING, A. A. **Qualidade fisiológica e sanitária de semente de soja**. Londrina: EMBRAPA, 1984. p. 5-24. (Circular Técnica, 9).

FREY, A. et al. Maternal synthesis of abscisic acid controls seed development and yield in Nicotiana plumbaginifolia. **Planta**, n. 218, p. 958–964, 2004.

FRUGOLI, J.; HARRIS, J. *Medicago truncatula* on the Move! **The Plant Cell**, v. 13, n. 3, p. 458-463, 2001.

GALLAND, M. et al. Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. **Molecular & Cellular Proteomics**, v. 13, p. 252–268, 2014.

GALLAND, M.; RAJJOU, L. Regulation of mRNA translation controls seed germination and is critical for seedling vigor. **Frontiers in Plant Science**, v. 6, p. 1-3, 2015.

GALLARDO K. et al. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. **Plant Physiology,** v. 133, p. 664-682, 2003.

GALLARDO, K. et al. Proteomic analysis of Arabidopsis seed germination and priming. **Plant Physiology**, Rockville, v. 126, n. 2, p. 835-848, 2001.

GENTLEMAN, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, n. 10, R80, 2004.

GHASSEMI-GOLEZANI, K., LOTFI, R., NOROUZI, M. Seed quality of soybean cultivars affected by pod position and water stress at reproductive stages. **International Journal of Plant, Animal and Environmental Sciences**, v.3, p.119-125, 2012.

GHASSEMI-GOLEZANI, K.; MOUSABEYGI, T.; RAEY,Y.; AHARIZAD, S. Effects of water stress and pod position on the seed quality of chickpea (*Cicer arietinum* L.) cultivars. **Notulae Botanicar Horti Agrobotanici Cluj-Napoca**, v.38, n.1, 2010.

GOMES, M. S. O. et al. Effect of harvesting and drying conditions on chlorophyll levels of soybean (*Glycine max* L. Merr). **Journal of Agricultural and Food Chemistry**, v. 51, n. 6, p. 1634–9, 2003

GRELET, J. et al. Identification in Pea Seed Mitochondria of a Late-Embryogenesis Abundant Protein Able to Protect Enzymes from Drying. **Plant Biologists**, v. 137, p. 157-167, 2005.

GRIMES, H. D. et al. A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. **Plant Cell,** v. 4, p. 1561–1574, 1992.

GUTIERREZ, L. et al. Combined networks regulating seed maturation. **Trends in Plant Science**, v. 12, p. 294-300. 2007.

HAJDUCH, M. et al. A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. **Plant Physiology**, v. 137, n. 4, p. 1397-1419, 2005.

HAKE, S. et al. The role of KNOX genes in plant development. **Cell and Developmental Biology**, v. 20, p. 125- 51, 2004.

HAYES, R. G.; KLEIN**,** W**.** H. Spectral quality influence of light during development of *Arabidopsis thaliana* plants in regulating seed germination**. Plant &Cell Physiology,** v. 15, n. 4, p. 643-653**,** 1974.

HEGEMAN, C. E.; GRABAU, E. A. A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. **Plant Physiology**, v. 126, n. 4, p. 1598-1608, 2001.
HILHORST, H. W. M. A critical update on seed dormancy. I. Primary dormancy. **Seed Science Research**, Ardingly, Haywards Heath**,** v. 5, p. 61-73, 1995.

HOEKSLRA. E A. et al. Changes in soluble sugars in relation to desiccation tolerance in cauliflower seeds. **Seed Science Research**. v. 4, n. 2, p. 143-147, 1994.

HOEKSTRA, F. A. et al. Imbibitional leakage from anhydrobiotes revisited. **Plant, Cell and Environment**, v. 22, n. 9, p. 1121-1131, 1999.

HOEKSTRA, F.; GOLOVINA, E. A.; BUITINK, J. Mechanisms of plant desiccation tolerance. **Trends in Plant Science**, v. 6, n. 9, p. 431-438, 2001.

HOLDSWORTH, M.; BENTSINK, L.; SOPPE, W. Molecular networks regulating Arabidopsis seed maturation , after ripening , dormancy and germination. **New Phytologist**, v. 179, n. 1, p. 33– 54, 2008.

HOWELL, K. A. et al. Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. **Plant Physiology,** v. 149, n. 2, p. 961-980, 2009.

HSING, Y. C. et al. Premature drying and germination in wild soybean seeds. **Taiwania**, v. 40, p. 73-81, 1995.

HUBER, S. C. et al. Canopy position has a profound effect on soybean 1 seed composition. **PeerJ.** v.13, n.4: 32452, 2016.

HUDSON, K. A**.** The circadian clock-controlled transcriptome of developing soybean seeds. **Plant Genome**, v. 3, p. 3-13; 2010.

HUNDERTMARK, M. et al. The reduction of seed-specific dehydrins reduces seed longevity in *Arabidopsis thaliana*. **Seed Science Research**, v. 21, n. 3, p. 165-173, 2011.

HUNT, L.; HOLDSWORTH, M. J.; GRAY, J. E. Nicotinamidase activity is important for germination. **The Plant Journal**, v. 51, n. 3, p. 341-351, 2007.

ILLIPRONTI, R. A. et al. Time of pod set and seed position on the plant contribute to variation in quality of seeds within soybean seed lots. **Netherlands Journal of Agricultural Science**, v. 48, p. 165-180, 2000.

ISHIBASHI, Y. et al. Regulation of soybean seed germination through ethylene production in response to reactive oxygen species. **Annals Botany**, v. 111, n. 1, p. 95-102, 2013.

JIA, H.; SUZUKI, M.; McCARTY, D. R. Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. **Wiley Interdisciplinary Reviews Developmental Biology,** v. 3, n. 1, p. 135-145, 2014.

JONES S. I.; VODKIN, L. O; Using RNA-Seq to profile soybean seed development from fertilization to maturity. **PlosOne,** v. 8, n.3, p. e59270, 2013.

JONES, S. I.; GONZALEZ, D. O.; VODKIN, L. O. Flux of transcript patterns during soybean seed development. **BMC Genomics**, v. 11, p. 136, 2010.

KALEMBA, E. M.; JANOWIAK, F.; PUKACKA, S. Association of Protective Proteins with Dehydration and Desiccation of Orthodox and Recalcitrant Category Seeds of Three Acer Genus Species**. Journal of Plant Growth Regulation**, v. 31, p.351– 362, 2012.

KANE, M. V. et al. Early- maturing soybean cropping system: III. Protein and oil contents and oil composition. **Agronomy Journal**, v. 89, n. 3, p. 464-469, 1997.

KANG, J. et al. Abscisic acid transporters cooperate to control seed germination. **Nature Communications**, v. 6, p. 8113, 2015.

KARSSEN, C. D. et al. Induction of dormancy during seed development by endogenous abscisic acid: Studies of abscisic acid deficient genotypes of *Arabidopsis thaliana* (L.) Heynh. **Planta**, v. 157, p. 158-165, 1983.

KARSSEN, C. M. Hormonal regulation of seed development, dormancy, and germination studied by genetic control. In: KIGEL, J. D.; GALILI, G. (Eds) **Seed development and germination**. New York: Marcel Decker, p. 333-350, 1995.

KAUR, H. et al. Differentially expressed seed aging responsive heat shock protein OsHSP18.2 implicates in seed vigor, longevity and improves germination and seedling establishment under abiotic stress. **Frontiers Plant Science**, v. 6, p. 713, 2015.

KEIRSTEAD, C.H. **Marketing study of factors affecting the quantity and value of products obtained from soybeans**. Washington, DC: United States Departament of Agriculture, Production and Marketing Administration, Fats and Oils Branch, 1952.

KERMODE, A. R. Role of abscisic acid in seed dormancy. **Journal of Plant Growth Regulation**, Berlin, Germany, n. 24, p. 319–344, 2005.

KERMODE, A. R.; BEWLEY, J. D. The role of maturation drying in the transition from seed development to germination: acquisition of desiccation tolerance and germinability during development of *Ricinus communis* L. seeds. **Journal of Experimental Botany**, v. 36, p. 1906-1915, 1985.

KERMODE, A.R. Regulatory mechanisms involved in the transition from seed development to germination. **Critical Reviews in Plant Sciences. v.** 9, p. 155-195, 2990.

KERMODE, A.R.; FINCH-SAVAGE, W. E. Desiccation sensitivity in orthodox and recalcitrant seeds in relation to development. In: Black M, Pritchard HW, editors. **Desiccation and survival in plants: drying without dying**. Wallingford, Oxon: CABI Publishing; 2002. p. 149-184.

KHAN, M. M. et al. Free radical accumulation and lipid peroxidation in testas of rapidly aged soybean seeds: a light-promoted process. **Seed Science Research**, Wallingford, v. 6, n. 3, p. 101-107, 1996.

KOORNNEEF, M.; REULING, G.; KARSSEN, C.M. The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. **Physiol Plant**. v. 61, n.3, p.377-383, 1984.

KOTAK, S. et al. A Novel Transcriptional Cascade Regulating Expression of Heat Stress Proteins during Seed Development of *Arabidopsis*. **The Plant Cell**, v. 19, n. 1, p. 182-185, 2007.

KUCERA, B.; COHN, M. A.; LEUBNER-METZGER, G. Plant hormone interactions during seed dormancy release and germination. **Seed Science Research**, n. 15, 281–307, 2005.

LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, v. 10, n. 3, p.25, 2009.

LARRE, C. et al. Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. **Plant Physiology,** Rockville, v.140, p.1418-1436, 2006.

LE, B. H. et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. **Proceedings of the National Academy of Sciences**, v. 107, n. 18, p. 8063-8070, 2010.

LE, B. H. et al. Using genomics to study legume seed development. **Plant Physiology,** v. 144, n. 2, p. 562-574, 2007.

LEE, K. P. et al. A seed coat bedding assay shows that RGL2-dependent release of abscisic acid by the endosperm controls embryo growth in Arabidopsis dormant seeds. **Proceedings of the National Academy of Sciences**, v. 107, n. 44, p. 19108-19113, 2010.

LEPINIEC, L. et al. Genetics and biochemistry of seed flavonoids. **Annual Review of Plant Biology**, v. 57, p. 405-430, 2006.

LEPRINCE, O. et al. Respiratory pathways in germinating maize radicles correlated with desiccation tolerance and soluble sugars. **Physiologia Plantarum**. v. 85, n. 4, p.581-588, 1992.
LEPRINCE, O.; BUITINK, J. Desiccation tolerance: from genomics to the field. **Plant Science**, v. 179, n. 6, p. 554-564, 2010.

LEPRINCE, O.; HENDRY G.A.F.; McKERSIE, B.D. The mechanisms of desiccation tolerance in developing seeds.**Seed Science Research**, v.3, p.231-246, 1993.
LIBAULT, M. et al. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. **The Plant Journal**, v. 63, p. 86-99, 2010.

LINDQUIST, S.; CRAIG, E. The heat-shock proteins. **Annual Reviews Genet.** n. 22, p. 631-677, 1988.

LISSO, J.; ALTMANN, T.; MÜSSIG, C. The AtNFXL1 gene encodes a NF-X1 type zinc finger protein required for growth under salt stress. **FEBS Letters**, v. 580, n. 20, p. 4851-4856, 2006.

LIU, B. et al. Effects of enhanced UV-B radiation on seed growth characteristics and yield components in soybean. **Field Crops Research,** v. 154, p, 158-163, 2013.

LIU, B. et al. Endogenous hormones in seed, leaf, and pod wall and their relationship to seed filling in soybeans. **Crop &Pasture Science**, v. 61, p. 103–110, 2010.

LIU, B. et al. Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. **Genetics**, v. 180, p. 995-1007, 2008.

LIU, X. et al. Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. **Proceedings of the National Academy of Sciences**, v. 110, n. 38, p. 15485-15490, 2013.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2 -ΔΔCt Method. **Methods**, v. 25, p. 402-408, 2001.

LONG, S. R., DALE, R. M., SUSSEX, I. M. Maturation and germination of *Phaseolus vulgaris* embryonic axes in culture. **Planta**, v. 153, p. 405-415, 1981.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, p. 550, 2014.

MA, L. et al. Simultaneous determination of 15 plant growth regulators in bean sprout and tomato with liquid chromatography-triple quadrupole tandem mass spectrometry. **Food Analytical Methods**, v. 6, p. 941-951, 2013.

MACOVEI, A. et al. New insights on the barrel medic *MtOGG1* and *MtFPG* functions in relation to oxidative stress response in planta and during seed imbibition. **Plant Physiology and Biochemistry,** v. 49, n. 9, p. 1040–1050, 2011.

MAEO, K. et al. An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. **The Plant Journal**, v. 60, n. 3, p. 476-487, 2009.

MAO, Z.; SUN, W. Arabidopsis seed-specific vacuolar aquaporins are involved in maintaining seed longevity under the control of ABSCISIC ACID INSENSITIVE. **Journal of Experimental Botany**, v. 66, p. 4781-4794, 2015. doi:10.1093/jxb/erv244.

MARCOS FILHO, J. **Fisiologia de sementes de plantas cultivadas**. Piracicaba: FEALQ, 2005. 495 p.

MATILLA, A. J.; MATILLA-VAZQUEZ, M. A. *Involvement of ethylene in seed physiology*. **Plant Science**, v. 175**, p. 87-98, 2008.**

MAY, M. J. et al. Glutathione homeostasis in plants: implications for environmental sensing and plant development. **Journal Experimental Botany**, v. 49, n. 321, p. 649-667, 1998.

MCDONALD, M.B. 1999. Seed deterioration: physiology, repair and assessment. **Seed Science Technology**. v. 27, p. 177-237, 1999.
McWILLIAMS, D. A.; BERGLUND, D. R.; ENDRES, G. J. **Soybean Growth and Management**. North Dakota State University, 1999.

MEDIC, J.; ATKINSON, C.; HURBURGH JR., C., R. Current Knowledge in Soybean Composition. Journal of the American Oil Chemists Society, v. 91, n. 3, p. 363-384, 2014.

MERTZ, L.M.; HENNING, F.A.; CRUZ, H.L.; MENEGHELLO, G.E.; FERRARI, C.S.; ZIMMER, P.D. Diferenças estruturais entre tegumentos de sementes de soja com permeabilidade contrastante. **Revista Brasileira de Sementes**, v.31, n.1, p.23-29, 2009.

MISRA, S.; BEWLEY, J. D. Reprogramming of protein synthesis from a developmental to a germinative mode induced by desiccation of the axes of *Phaseolus vulgaris*. **Plant Physiology,** v. 78, p. 876-882, 1985.

MITTLERA, R. Oxidative stress, antioxidants and stress tolerance. **Trends in Plant Science,** Oxford, v. 7, n. 9, p. 405-410, 2002.

MORALES, M. A. P. et al. Transcriptome analyses and virus induced gene silencing identify genes in the *Rpp4*-mediated Asian soybean rust resistance pathway. **Functional Plant Biology**, v. 40, n.10, p. 1029-1-47, 2013.

MUDGE, J. et al. Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. **BMC Plant Biology**. v. 5, n.15, 2005.

NAGAO, R. T. et al. Analysis of multiple classes of soybean heat shock genes and proteins. p. 3–20. Physical stresses in plants: Genes and their products for tolerance. *In:* Biochemical and Cellular Mechanisms of Stress Tolerance in Plants: Proc. of the NATO Advanced Research Workshop, Maratea, Italy. 20–24 June 1994. Springer-Verlag, Berlin

NAKABAYASHI, K. et al. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. **The Plant Journal**, v. 41, p. 697-709, 2005.

NAKAJIMA, S. et al. Chlorophyll *b* Reductase Plays an Essential Role in Maturation and Storability of Arabidopsis Seeds. Plant Physiology, v. 160, n. 1, p. 261–273, 2012.

NAMBARA, E.; MARION-POLL, A. ABA action and interactions in seeds. **Trends in Plant Science**, v. 8, n. 5, p. 213-217, 2003.

NAMBARA, E.; MARION-POLL, A. Abscisic acid biosynthesis and catabolim. **Annual Review Plant Biologist,** v. 56, p. 165-185, 2005.

NONOGAKI, H.; BASSEL, G. W.; BEWLEY, J. D. Germination: still a mystery. **Plant Science**, v. 179, n. 6, p. 574-581, 2010.

OBENDORF, R. L. et al. Accumulation of soluble carbohydrates during seed development and maturation of low-raffinose, low-stachyose soybean. **Crop Science**, v. 49, p. 329-341, 2009.

OBENDORF, R. L. et al. Influence of seed maturation on germinability in soybean. **Crop Science**, v. 20, p. 483-486, 1980.

OBENDORF, R. L. et al. Soluble oligosaccharides and galactosyl cyclitols in maturing soybean seeds in planta and in vitro. **Crop Science**, v. 38, p. 78-84, 1998.

OGÉ, L. et al. Protein Repair L-Isoaspartyl Methyltransferase1 is involved in both seed longevity and germination vigor in Arabidopsis. **The Cell Plant**, v. 10, n. 11, p. 3022-3037, 2008.

OKAMOTO, M. et al. CYP707A1 and CYP707A2, which encode abscisic acid 8'-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. **Plant Physiology**, v. 141, n. 1, p. 97–107, 2006.

OOMS, J. et al. Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Abscisic Acid-Insensitive abi3 Mutants). **Plant Physiology**, v. 102, n. 4, p. 1185-1191, 1993.

ORNBOS Jr, D. L. Production environment and seed quality. p 119-152. In: Barsa, A.S., ed. **Seed Quality: Basic Mechanisms and Agricultural Implications**. Food Products Press, New York, NY, USA, 1995.

PÁDUA, G. P. et al. Incidence of green soybean seeds as a function of environmental stresses during seed maturation. **Revista Brasileira de Sementes**, v. 31, n. 3, p. 150–159, 2009.

PÁDUA, G. P. et al. Tolerance level of green seed in soybean seed lots after storage. **Revista Brasileira de Sementes**, v. 29, n. 3, p. 112–20, 2007.

PÁDUA, G. P. et al. Zoneamento agroclimático do estado de Minas Gerais para a produção de semente de soja de alta qualidade. **Journal of Seed Science**, Londrina, v. 36, n. 4, 2014.

PAMMENTER, N. W.; BERJAK, P. A review of recalcitrant seed physiology in relation to desiccation tolerance mechanisms. **Seed Science Research**, v. 9, n. 1, p.13-37, 1999.

PARCY, F. et al. The *ABSCISIC ACID-lNSENSITIVE3, FUSCA3,* and *LEAFY COTYLEDONf* Loci Act in Concert to Control Multiple Aspects of Arabidopsis Seed Development. **The Plant Cell**, v. 9, p. 1265-1277, 1997.

PELTIER, A.J.; HATFIELD, R.D.; GRAU, C.R. Soybean stem lignin concentration relates to resistance to *Sclerotinia Sclerotiorum*. **Plant Diseases**, v. 93, n. 2, p;149-154, 2009.

PERSONAT, J. M. et al. Co-overexpression of two Heat Shock Factors results in enhanced seed longevity and in synergistic effects on seedling tolerance to severe dehydration and oxidative stress. **BMC Plant Biology**, v. 14, n. 56, 2014.

PFAFFL, M. W.; HORGAN, G. W.; DEMPFLE L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. **Nucleic Acids Research**, v. 30, n. 9, 2002.

POPINIGS, F. **Fisiologia da semente.** 2. ed. Brasília: AGIPLAN, 1985. 289p.

POTTERS, G. et al. Ascorbate and glutathione: guardians of the cell cycle, partners in crime? **Plant Physiology and Biochemistry**, v. 40, n. 6-8, p. 537-548, 2002.

PRASAD, B.; PRASAD, R.; PRASAD, S. Seed quality variation in relation to pod and seed position on the mother plant of soybean [*Glycine max*. (L.) Merrill] seed Iot. **Journal of Crop and Weed**. v. 6, n. 2, p.1-5, 2010.

PRATTLEY, C. A.; STANLEY, D. W. Protein-phytate interactions in soybeans. I. Localization of phytate in protein bodies and globoids. **Journal of Food Biochemistry,** v. 6, n. 4, p. 243-253, 1982.

PRIETO-DAPENA, P. et al. Improved Resistance to Controlled Deterioration in Transgenic Seeds. **Plant Physiology,** v. 142, n. 3, p. 1102-1112, 2006.

PROBERT, R. et al. Seed quality for conservation is critically affected by pre-storage factors. **Australian Journal of Botany**, v. 55, p. 326-335, 2007.

QIU, L. J., CHANG, R. Z. The origin and history soybean. In: SINGH, G. (Ed.). **The soybean**: botany, production and uses. Ludhiana: CAB, 2010. p. 13-35.

QUEBEDEAUX, B.; SWEETSER, P. B.; ROWELL, J.C. Abscisic Acid Levels in Soybean Reproductive Structures during Development. **Plant Physiology**, v. 58, n. 3, p. 363-366, 1976.

QUEITSCH, C. et al. Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis.**Plant Cell**, v.12, n.4, p.479-492, 2000.

RABOY, V.; DICKINSON, D. B. The timing and rate of phytic acid accumulation in developing soybean seeds. **Plant. Physiology**, v. 85, p. 0841-844. 1987.

RAIJ, B. et al. Recomendações de adubação e calagem para o Estado de São Paulo**. Boletim Técnico do Instituto Agronômico de Campinas**, Campinas, n. 100, 1997.

RAJJOU, L. et al. Proteome-wide characterization of seed aging in Arabidopsis. A comparison between artificial and natural aging protocols. **Plant Physiology.** v. 148, p. 620-641, 2008.

RAJJOU, L. et al. Seed germination and vigor. **The Annual Review of Plant Biology**, v. 63, p. 507-533, 2012.

RAJJOU, L. et al. The effect ofα-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNA during germination. **Plant Physiology**, v. 134, p. 1518-1613, 2004.

RAJJOU, L.; DEBEAUJON, I. Seed longevity: survival and maintenance of high germination ability of dry seeds. ***Comptes Rendus Biologies***, v. 331, n. 10, p. 796–805, 2008.

RANATHUNGE, K. et al. Properties of the soybean seed coat cuticle change during development. **Planta**, v. 231, n. 5, p. 1171-1188, 2010.

REN, C.; BILYEN, K. D.; BEUSELINCK, P. R. Composition, vigor and proteome of mature soybean seeds developed under high temperature. **Crop Science**, v. 49, p. 1010-1022, 2009.

RIGHETTI, K. et al. Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. **The Plant Cell**, v. 27, p. 2692-2708, 2015.

RITCHIE, S. W.; HANWAY, J. J.; THOMPSON, H. E. How a soybean plant develops. Cidade: Iowa State University of Science and Technology, 1982. (Special Report, 53).

ROSENBERG, L. A.; RINNE, R. W. Moisture loss as a prerequisite for seedling growth in soybean seeds (*Glycine max* L. Merr.). **Journal of Experimental Botany**, v. 37, p. 1663-1674, 1986.

RUIJTER, J. M. et al. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. **Science Direct Methods**, v. 59, n. 1, p. 32-46, 2013.

RUSHTON, P. J. et al. WRKY transcription factors. **Trends in Plant Science**, v. 15, n. 5, p. 247-258, 2010.

RYLOTT, E.L.; GILDAY, A.D., GRAHAM, L.A. The gluconeogenic enzyme phospho*enol*pyruvate carboxykinase in arabidopsis is essential for seedling establishment. **Plant Physiology.** v. 131, n. 4, p.1834-1842, 2003.

SALDIVAR, X. et al. Changes in chemical composition during soybean seed development. **Food Chemistry**, v. 124, n. 4, p. 1369-1375, 2011.

SAMARAH, N. H. et al. Ethylene evolution from soybean seeds podded and depodded related to seed desiccation tolerance during maturation**. Seed Science and Technology.** v. 44, n. 1, p. 53-63, 2016.

SANDOVAL, J. A. et al. N-Acylphosphatidylethanolamine in dry and imbibing cotton seeds: amounts, molecular species, and enzymatic synthesis. **Plant Physiology**, v. 109, n. 1, p. 269-275, 1995.

SANHEWE, A. J.; ELLIS, R. H. Seed development and maturation in *Phaseolus vulgaris* II. Post-harvest longevity in air-dry storage. **Journal Experimental Botany**, v. 7, n. 7, p. 959-965, 1996.

SANO, N. et al. Staying alive: molecular aspects of seed longevity. **Plant and Cell Physiology**, p. 1-15, 2015.

SANTOS, E. L. et al. Qualidade fisiológica e composição química das sementes de soja com variação na cor do tegumento. **Revista Brasileira de Sementes**, v. 29, n. 1, p. 20-26, 2007.

SANTOS, T. L. Soja. In: CASTRO, P. R. C.; KLUGE, R. A.; SESTARI, I. **Manual de fisiologia vegetal: fisiologia dos cultivos.** Piracicaba: Editora Agronômica Ceres, p. 157-175, 2008.

SANTOS-MENDOZA, M. et al. LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. **FEBS Letters**, v. 579, n. 21, p. 4666-4670, 2005.

SCANDALIOS, J. G. Oxygen stress and superoxide dismutases. **Plant Physiology**, Bethesda, v. 101, n. 1, p. 7-12, 1993.

SCHMUTZ J. et al. Genome sequence of the palaeopolyploid soybean. **Nature.** v. 463, p. 178-183, 2010.

SCHUSSLER, J. R., BRENNER, M. L., BRUN, W. A. Abscisic acid and its relationship to seed filling in soybeans. **Plant Physiology**, v. 76, p. 301-306, 1984.

SCOFIELD, S.; MURRAY, J. A. KNOX gene function in plant stem cell niches. **Plant Molecular Biology**, v. 60, n. 6, p. 929-946, 2006.

SEDIYAMA, T.; PEREIRA, M. G.; SEDIYAMA, C. S.; GOMES, J. L. L. **Cultura da Soja, Parte I**. Viçosa: UFV, p.97, 1993.

SEVERIN, A. J. et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. **BioMed Central Plant Biology**, v. 10, n. 160, 2010.

SEVERIN, A. J. et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. **BioMed Central Plant Biology**, v. 10, n. 160, p. 1-16, 2010.

SHA, A. H. et al. Large-scale sequencing of normalized full-length cDNA library of soybean seed at different developmental stages and analysis of the gene expression profiles based on ESTs. **Molecular Biology Reports**, v. 39, n. 3, p. 2867-2874, 2012.

SHAMIMUZZAMAN, M.; VODKIN, L. Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. **BMC Genomics**, v. 13, p. 310, 2012.

SHARMA, S. et al. Positional effects on soybean seed composition during storage, **J Food Sci Technol,** v. 2, n. 50, p. 353-359, 2013.

SINNECKER, P. Degradação da clorofila durante a maturação e secagem de semente de soja. 2002. 103 p. **Tese** (Doutorado em Ciência dos Alimentos)-Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2002.

SOUZA, F. H. D.; MARCOS-FILHO, J. The seed coat as a modulator of seed-environment relationships in Fabaceae. **Revista Brasileira de Botânica,** v. 24, n.4. p. 365-375, 2001.

SU, T. et al. WRKY42 Modulates Phosphate Homeostasis through Regulating Phosphate Translocation and Acquisition in Arabidopsis. **Plant Physiology**, v. 167, n. 4, p. 1579-1591, 2015.

SUN, W. O.; LEOPOLD, A. D. Acquisition of desiccation tolerance in soybeans. **Physiologia Plantarium**, v. 87, p. 403-409, 1993.

TAIZ, L.; ZEIGER, E. **Fisiologia Vegetal**. 5. ed. São Paulo: Artmed, 2013. 918p.

TEIXEIRA, R. N. et al. Gene expression profiling of the green seed problem in Soybean. **BioMed Central Plant Biology**, v. 16, n. 37, p. 1-15, 2016.

TEJEDOR-CANO, J. et al. Loss of function of the HSFA9 seed longevity program. **Plant, Cell & Environment,** v. 33, n. 8, p. 1408-1417, 2010.

TEKRONY, D. M. et al. Effect of field weathering on the viability and vigour of soybean seed. **Agronomy Journal**, v. 72, p. 749-753, 1980.

TEKRONY, D. M. et al. Physiological maturity in soybean. **Agronomy Journal**, v. 71, p. 771-775, 1979.

TO, A. et al. A Network of Local and Redundant Gene Regulation Governs *Arabidopsis* Seed Maturation. **The Plant Cell**, v. 18, p. 1642-1651, 2006.

TSUCHIYA et al. The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis. **Plant Journal**, v. 37, n.1, p. 73-81, 2004.

TUNNACLIFFE, A.; WISE, M. The continuing conundrum of the LEA proteins..**The Science of Nature**. v. 94, n. 10, p. 791-812, 2007.

VANDENABEELE, J. D. S. et al. Dual action of the active oxigen species during plant stress responses. **Cellular and Molecular Life Science**, Basel, v. 57, n. 5, p. 779-795, 2000.

VERDIER, J. et al. A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. **Plant Physiology**, v. 163, p. 757-774, 2013.
VERTUCCI, C.W.; FARRANT, J. M. Acquisition and loss of desiccation tolerance. In: KIGEL, J.; GALILI, G. (Ed.). S**eed development and germination**. New York: Marcel Dekker, 1995. p.237-271.

VICENTE-CARBAJOSA, J. et al. Barley BLZ1: a bZIP transcriptional activator that interacts with endosperm-specific gene promoters. **The Plant Journal**, v. 13, n. 5, p. 629-640, 1998.

VICENTE-CARBAJOSA, J.; CARBONERO, P. Seed maturation: developing an intrusive phase to accomplish a quiescent state. **International Journal of Developmental Biology**, v. 49, n. 5-6, p. 645-651, 2005.
WALTERS, C. et al. Preservation of recalcitrant seeds. **Science**. v. 339, p. 915-916, 2013.

WAN, Y. et al. The signal transducer NPH3 integrates the phototropin1 photosensor with PIN2-based polar auxin transport in arabidopsis root phototropism. **The Plant Cell**. v. 24, p.551-565, 2012.

WANG, W.X.; VINOCUR, B.; ALTMAN,A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. **Planta**. v.218, n.1, p.1-14, 2003.

WATANABE, I.; NAGASAWA, T. Appearance and chemical composition of soybean seeds in germplasm collection of Japan: II. Correlation among protein, lipid and carbohydrate percentage. **Japanese Journal of Crop Science**, v. 59, p. 661–666, 1990.

WATERWORTH, W. M. et al. A plant DNA ligase is an important determinant of seed longevity. **The Plant Journal**, v. 63, n. 5, p. 848-860, 2010.

WEHMEYER, N.; VIERLING, E. The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. **Plant Physiology**, v. 1226, n. 1, p. 1099-1108, 2000.

WEITBRECHT, K.; KERSTIN MÜLLER, K.; LEUBNER-METZGER, G. First off the mark: early seed germination. **Journal of Experimental Botany**, v. 62, n. 10, p. 3289-3309, 2011.

WHITE, C.N.; RIVIN, C.J. Gibberellins and seed development in maize. II. Gibberellin synthesis inhibition enhances abscisic acid signaling in cultured embryos. **Plant Physiol.** v. 122, n. 4, p.1089-97, 2000.

WHITEHOUSE, K. J.; HAY, F. R; ELLIS, R. H. Increases in the longevity of desiccation-phase developing rice seeds: response to high-temperature drying depends on harvest moisture content. **Annals of Botany**, v. 117, n. 5, p. 1-13, 2016.

WILLARD, C. J. The time of harvesting soybeans for hay and seed. **Journal of the American Society of Agronomy**, v. 17, n. 2, p. 157-168, 1925.

WILSON, R. F. **Seed composition Soybeans**: Improvement, production and users. 3rd. ed. Madison, WI. : ASA, CSSA, SSSA, 2004.

WISE, M. J. LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. **BMC Bioinformatics**. v. 4, n. 1, p. 52-54, 2003.

XU, N.; BEWLEY, J. D. Sensitivity to abscisic acid and osmoticum changes during embryogenesis of alfafa (*Medicago Sativa*). **Journal of Expimental Botany**, v. 42, n. 6, p. 821-826, 1991.

XUE-FEI, D. et al. The SnRK protein kinase family and the function of SnRK1 protein kinase. **International Journal of Agriculture & Biology,** v. 14, n. 4, p. 575-579, 2012.

YI, X.; DU, Z.; SU, Z. PlantGSEA: a gene set enrichment analysis toolkit for plant community. **Nucleic Acids Research**. v. 41, W98–103. 2013.

ZANAKIS, G. N.; ELLIS, R. H.; SUMMERFIELD, R. J. Seed quality in relation to seed development and maturation in three genotypes of soybean (*Glycine max*). **Experimental Agriculture.** v. 30, p. 139-156, 1994.

ZENTELLA, R.; YAMAUCHI, D.; HO, T. H. D. Molecular dissection of the gibberellin/abscisic acid signaling pathways by transiently expressed RNA interference in Barley aleurone cells. **Plant Cell**. Rockville, v. 14, p. 2289-2301, 2002.

ZORATTO, M. F. et al. Presença de sementes esverdeadas em soja e seus efeitos sobre seu potencial fisiológico. **Revista Brasileira de Sementes,** v.29, n. 1, p. 11–7, 2009.

Supporting Information

Supplementary Table 1. List of the top 30 transcripts whose level increase the most during maturation.

***Gmax* 2.0 ID**	**Tax Hit**	***p*-value**	**Name**	**R90h/ R7.2d 0h**	**BH-value**
Glyma.06G142900	C7IVU8_9POAL	2,00E-25	Sorghum bicolor protein targeted either to mitochondria or chloroplast proteins T50848	4,11	1,06E-35
Glyma.17G224900	HSP21_SOYBN	3,00E-114	17.9 kDa class II heat shock protein	3,73	7,81E-15
Glyma.14G099900	HSP21_SOYBN	1,00E-93	17.9 kDa class II heat shock protein	3,69	3,70E-33
Glyma.17G077500	Q39889_SOYBN	0	Heat shock protein	3,51	1,01E-12
Glyma.14G222700	B9GQ52_POPTR	0	Prolyl oligopeptidase family protein	3,34	9,65E-10
Glyma.08G106700	G5ELZ2_TOBAC	0	FK506-binding protein	3,27	1,84E-15
Glyma.06G235100	-	-	-	3,25	4,59E-09
Glyma.12G200600	-	-	-	3,13	2,81E-16
Glyma.04G222500	C7IVU8_9POAL	2,00E-25	Sorghum bicolor protein targeted either to mitochondria or chloroplast proteins T50848	3,05	7,23E-21
Glyma.13G176200	HSP14_SOYBN	5,00E-106	17.5 kDa class I heat shock protein	2,90	7,22E-19
Glyma.09G040400	E9KT46_9FABA	7,00E-86	PR10.13.36 (Fragment)	2,88	6,30E-05
Glyma.18G061100	K7MQ84_SOYBN	0	*Asparagine synthetase*	2,82	4,74E-21
Glyma.07G267800	OPT3_ARATH	0	Oligopeptide transporter 3	2,79	6,70E-19
Glyma.06G134900	A9QVH3_9FABA	2,00E-125	Heat shock protein	2,79	2,32E-17
Glyma.12G200400	-	-	-	2,63	7,19E-10

Supplementary Table 1. Continuation …

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R90h/ R7.2d 0h	BH-value
Glyma.14G140400	A2TDC3_9ROSI	2,00E-104	Truncated hemoglobin	2,63	1,86E-07
Glyma.09G181200	G7KML0_MEDTR	0	Kinesin-like protein KIF3A	2,63	4,14E-12
Glyma.10G093100	E4MXB6_THEHA	2,00E-86	mRNA, clone: RTFL01-22-A04	2,63	7,59E-10
Glyma.07G080400	G7L1G0_MEDTR	0	Phospholipase D	2,56	1,90E-09
Glyma.18G259700	G7L1Y9_MEDTR	0	Cell division protease ftsH-like protein	2,50	1,88E-08
Glyma.12G073100	Q76LA6_SOYBN	2,00E-180	Cytosolic ascorbate peroxidase 2	2,46	2,39E-15
Glyma.09G283400	G7IBD4_MEDTR	0	Chaperone protein dnaJ	2,45	5,81E-10
Glyma.03G068100	D4N5G1_SOYBN	0	Rubisco activase	2,45	1,08E-07
Glyma.07G200200	HSP16_SOYBN	7,00E-110	18.5 kDa class I heat shock protein	2,44	2,81E-16
Glyma.06G114600	Q9XI02_ARATH	0	F8K7.18 protein	2,44	8,68E-06
Glyma.20G213900	HSP41_SOYBN	2,00E-136	22.0 kDa class IV heat shock protein	2,43	8,38E-16
Glyma.10G176800	B7FAD8_ORYSJ	2,00E-117	cDNA, clone: J100065N04, full insert sequence	2,41	3,21E-10
Glyma.03G189900	Q8H0I7_PETHY	0	Cytochrome P450	2,40	3,08E-09
Glyma.09G246300	G7KSH3_MEDTR	2,00E-134	Hydroxyproline-rich glycoprotein-like protein	2,37	1,19E-06
Glyma.16G178900	B9MX43_POPTR	8,00E-121	Glutamine cyclotransferase family protein	2,36	1,05E-07

Supplementary Table 2. .List of the top 30 transcripts whose level decrease the most during maturation.

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R9 0h/ R7.2d 0h	BH-value
Glyma.01G130200	I1J7M1_SOYBN	1,00E-76	Non-specific lipid-transfer protein	-3,76	6,16E-14
Glyma.02G061600	G7K9G6_MEDTR	3,00E-27	Protein disulfide-isomerase	-3,47	7,84E-11
Glyma.05G204800	P21_SOYBN	6,00E-128	Protein P21	-3,43	5,28E-20
Glyma.01G034700	TBL27_ARATH	6,00E-119	Protein ALTERED XYLOGLUCAN 4	-3,33	3,26E-07
Glyma.13G133000	G7L0M5_MEDTR	1,00E-49	Zinc finger protein	-3,32	5,53E-19
Glyma.10G033500	F4INW4_ARATH	0	Histone H1flk-like protein	-3,32	3,36E-07
Glyma.16G208900	Q43709_SOYBN	2,00E-64	Bowman-Birk proteinase isoinhibitor D-II	-3,30	3,58E-09
Glyma.18G127200	G7K9P5_MEDTR	0	Peptide transporter PTR1	-3,30	8,85E-25
Glyma.10G058200	I1L905_SOYBN	0	Phenylalanine ammonia-lyase	-3,30	6,77E-24
Glyma.19G052900	-	-	-	-3,21	2,45E-07
Glyma.06G160300	Q39894_SOYBN	1,00E-42	Seed-specific low molecular weight sulfur-rich protein	-3,19	8,39E-08
Glyma.08G250300	Q69F92_PHAVU	3,00E-51	B12D-like protein	-3,14	1,99E-07
Glyma.13G309900	A2Q4H0_MEDTR	1,00E-154	Mitogen-activated protein kinase kinase kinase A	-3,12	6,64E-08
Glyma.02G000500	K7K5Q4_SOYBN	2,00E-138	Auxin-induced protein	-3,11	7,94E-07
Glyma.05G236900	I1K5N0_SOYBN	0	Pectinesterase	-3,10	8,54E-06

Supplementary Table 2. Continuation…

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R9 0h/ R7.2d 0h	BH-value
Glyma.15G274900	G7JEB0_MEDTR	1,00E-27	Ycf68	-3,09	1,79E-19
Glyma.12G090600	S4Z2Z0_TRIMO	2,00E-26	Hypothetical chloroplast RF68	-3,08	1,43E-24
Glyma.U040600	Q8S8Z5_SOYBN	0	Syringolide-induced protein B13-1-1 (Fragment)	-3,04	4,87E-07
Glyma.11G114700	H6T057_9POAL	5,00E-67	Photosystem II CP43 chlorophyll apoprotein (Fragment)	-3,02	1,86E-13
Glyma.08G244500	G7L593_MEDTR	0	UDP-glucosyltransferase HRA25	-3,01	4,29E-11
Glyma.14G158500	-	-	-	-3,00	4,07E-28
Glyma.06G186200	Q9AUH7_PEA	2,00E-17	UVI1	-2,99	7,48E-12
Glyma.03G176300	T2DNZ6_PHAVU	4,00E-126	In2-1 protein	-2,98	1,15E-08
Glyma.18G228200	C1E094_MICSR	7,00E-60	Vacuolar iron family transporter	-2,97	1,75E-06
Glyma.20G054100	I1NE44_SOYBN	0	Lipoxygenase	-2,97	2,46E-06
Glyma.13G194400	G3FGW7_SOYBN	5,00E-97	Leginsulin 1	-2,94	7,90E-08
Glyma.14G132100	H8PI40_GLYSO	2,00E-76	Maturase K	-2,93	7,90E-08
Glyma.10G173200	I1LBY5_SOYBN	2,00E-73	Protein yippee-like	-2,93	3,40E-06
Glyma.05G057600	Q9SXE7_ARATH	1,00E-37	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	-2,90	5,57E-11
Glyma.08G241600	G7L453_MEDTR	0	Cyclic nucleotide-gated ion channel	-2,89	7,93E-08

Supplementary Table 3. List of the top 30 transcripts whose level increase the most during artificial drying.

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R7.2d/ R7.2f	BH-value
Glyma.05G236800	I1K5N1_SOYBN	0	Pectinesterase	5,40	9,5E-17
Glyma.10G206000	F4I9C1_ARATH	2,00E-07	Hydroxyproline-rich glycoprotein-like protein	5,21	3,6E-08
Glyma.11G068100	-	-	-	5,06	2,5E-20
Glyma.11G187000	PCS3_LOTJA	3,00E-120	Glutathione gamma-glutamylcysteinyltransferase 3	5,06	4,2E-09
Glyma.15G188400	PSBN_SOYBN	4,00E-22	Protein PsbN	5,02	1,4E-17
Glyma.08G229200	-	-	-	4,89	1,6E-17
Glyma.20G184500	S4WWG9_GLYSO	1,00E-139	Hydroxyproline-rich glycoprotein family protein (Fragment)	4,74	7,5E-09
Glyma.12G151700	ATL44_ARATH	4,00E-10	RING-H2 finger protein ATL44	4,73	1,2E-06
Glyma.05G236900	I1K5N0_SOYBN	0	Pectinesterase	4,69	6,0E-08
Glyma.09G171400	V9P892_9ROSI	3,00E-11	Photosystem II protein M	4,55	1,5E-09
Glyma.15G208300	PSBM_POPTR	2,00E-13	Photosystem II reaction center protein M	4,49	4,8E-14
Glyma.03G181000	-	-	-	4,45	3,8E-22
Glyma.07G079300	-	-	-	4,40	6,2E-42
Glyma.01G105100	-	-	-	4,35	7,4E-06
Glyma.17G245200	Q0PJI9_SOYBN	1,00E-79	MYB transcription factor MYB124	4,34	6,1E-07

Supplementary Table 3. Continuation…

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R7.2d/ R7.2f	BH-value
Glyma.09G171300	C1K530_9ASTR	8,00E-08	Cytochrome b6-f complex subunit 8	4,29	4,7E-13
Glyma.16G038100	-	-	-	4,24	8,3E-11
Glyma.13G313100	Q2HSI2_MEDTR	0	Protein kinase	4,09	1,0E-04
Glyma.06G186200	Q9AUH7_PEA	2,00E-17	UVI1	4,08	1,1E-19
Glyma.18G132800	R9ZV93_GLYTO	7,00E-33	Protein Ycf2	4,07	2,5E-05
Glyma.07G143800	V9Z7G6_9ROSI	6,00E-57	Photosystem I P700 apoprotein A1 (Fragment)	4,07	2,9E-14
Glyma.11G114700	H6T057_9POAL	5,00E-67	Photosystem II CP43 chlorophyll apoprotein (Fragment)	4,03	4,2E-22
Glyma.04G095000	A0AS46_MEDVI	9,00E-96	Photosystem II D2 protein (Fragment)	4,01	8,5E-22
Glyma.05G074900	Q66VR0_9ERIC	5,00E-29	PsbC (Fragment)	3,95	6,4E-15
Glyma.03G176300	T2DNZ6_PHAVU	4,00E-126	In2-1 protein	3,95	8,8E-13
Glyma.09G185400	-	-	-	3,94	2,6E-08
Glyma.06G224500	K4HGK5_CAMSI	2,00E-54	Photosystem I P700 apoprotein A1 (Fragment)	3,92	3,3E-13
Glyma.13G068600	NU3C_SOYBN	2,00E-75	NAD(P)H-quinone oxidoreductase subunit 3, chloroplastic	3,87	8,1E-19
Glyma.U013900	-	-	-	3,87	9,8E-05
Glyma.06G171100	-	-	-	3,86	3,2E-18

Supplementary Table 4. List of the top 30 transcripts whose level increase the most during imbibition for both stages (profile A).

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R918h/ R90h	BH-value	R7.2d18h/ R7.2d0h	BH-value
Glyma.01G019200	Q94LX7_FLAPR	0	Phosphoenolpyruvate carboxykinase	8,12	1,95E-49	6,01	2,63E-37
Glyma.18G210900	C6SWY4_SOYBN	9,00E-79	Non-specific lipid-transfer protein	8,08	3,53E-44	7,75	9,18E-31
Glyma.14G191700	Q0PJL5_SOYBN	0	MYB transcription factor MYB57	8,03	4,86E-31	8,22	7,20E-23
Glyma.02G028100	B6CAM2_SOYBN	0	Matrix metalloproteinase	7,85	2,82E-32	9,18	2,54E-117
Glyma.02G028000	B6CAM2_SOYBN	0	Matrix metalloproteinase	7,81	5,97E-18	6,45	1,30E-15
Glyma.17G019300	I1MRA7_SOYBN	0	Amine oxidase	7,79	5,39E-40	8,71	2,94E-23
Glyma.17G063900	G7JVH0_MEDTR	0	Mitochondrial chaperone BCS1	7,63	8,18E-17	6,95	1,01E-18
Glyma.06G193800	M1GM36_GOSHI	2,00E-49	GA-stimulated transcript-like protein 2	7,58	2,83E-27	8,04	1,39E-21
Glyma.20G248100	G7IB14_MEDTR	2,00E-84	Transcription factor bHLH128	7,49	2,34E-66	8,39	1,23E-40
Glyma.14G141000	B2ZPK6_SOYBN	0	BURP domain-containing protein	7,47	3,72E-21	6,51	3,53E-15
Glyma.17G012300	G7JJ92_MEDTR	0	L-ascorbate oxidase-like protein	7,46	6,06E-16	7,39	2,56E-13
Glyma.07G096000	Q5N7Q6_ORYSJ	9,00E-24	Receptor-like protein kinase-like	7,45	7,14E-16	6,41	6,51E-15
Glyma.02G036500	E2IXG6_MEDTR	0	Leucine-rich repeat receptor-like kinase	7,38	2,66E-30	7,16	2,27E-22
Glyma.19G219500	G7KTQ1_MEDTR	0	Boron transporter	7,35	1,15E-19	5,94	3,72E-10
Glyma.03G202400	D7L4L9_ARALL	0	Protein binding protein	7,34	5,89E-30	7,13	3,89E-32

Supplementary Table 4. Continuation…

Gmax 2.0 ID	Tax Hit	*p*-value	Name	R918h/ R90h	BH-value	R7.2d18h/ R7.2d0h	BH-value
Glyma.17G015400	G7JI12_MEDTR	0	Cytochrome P450	7,17	3,32E-53	8,22	2,37E-32
Glyma.13G112200	U5G769_POPTR	3,00E-80	Senescence-associated family protein	7,16	4,50E-68	7,26	6,60E-44
Glyma.02G099500	G7JY94_MEDTR	0	AP2 domain-containing transcription factor	7,12	4,28E-14	3,68	8,40E-11
Glyma.13G347700	B3TDK7_SOYBN	0	Lipoxygenase	7,07	4,55E-43	6,50	7,42E-18
Glyma.10G172200	B2LUN7_SOYBN	0	Granule bound starch synthase Ia	7,06	2,64E-62	6,69	5,15E-86
Glyma.13G069900	G7KIU9_MEDTR	9,00E-31	Gibberellin regulated protein	7,05	2,72E-18	5,84	1,57E-14
Glyma.02G224900	Q0PJL5_SOYBN	0	MYB transcription factor MYB57	6,86	6,56E-34	7,85	3,86E-25
Glyma.04G023900	T2DNJ3_PHAVU	0	Tubulin beta-1 chain-like protein	6,85	4,51E-17	6,15	3,19E-13
Glyma.14G036900	G7KCU0_MEDTR	2,00E-113	Early nodulin-like protein	6,85	2,54E-21	6,39	1,36E-21
Glyma.15G142400	F8SMB8_SOYBN	0	Acidic glucanase	6,83	5,72E-17	5,41	1,96E-13
Glyma.13G248000	D9ZIW8_MALDO	3,00E-33	COL domain class transcription factor	6,82	8,55E-25	8,32	2,91E-18
Glyma.10G047100	G7L0N3_MEDTR	5,00E-171	Protein ALUMINUM SENSITIVE	6,79	1,80E-12	4,48	1,66E-12
Glyma.06G027700	T2DN49_PHAVU	0	Cysteine proteinase RD19a-like protein	6,79	1,47E-36	6,06	6,26E-26
Glyma.13G304400	T2DP41_PHAVU	0	Xyloglucan endotransglucosylase/ hydrolase protein 9-like protein	6,76	8,89E-25	8,50	2,38E-19
Glyma.20G033900	Q8L5G4_CICAR	3,00E-168	Alpha-expansin 4	6,75	4,23E-35	5,17	1,44E-21

Supplementary Table 5. Genes up-regulated in mature legume seeds during imbibition versus immature.

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*p*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_080740	Medtr3g024510.1	AT1G29670.1	1,00E-150	Glyma.13G223100	G7J1P3_MEDTR	1,00E-134	AT1G29670.1	GDSL-like Lipase/Acylhydrolase superfamily protein	2,00E-103
Medtr_v1_072741	Medtr4g011630.1	AT1G15520.1	0,00E+00	Glyma.13G361900	G7IMF4_MEDTR	0,00E+00	AT1G15520.1	pleiotropic drug resistance 12	0,00E+00
Medtr_v1_072741	Medtr4g011630.1	AT1G15520.1	0,00E+00	Glyma.15G011900	G7IMF2_MEDTR	0,00E+00	AT1G15520.1	pleiotropic drug resistance 12	0,00E+00
Medtr_v1_072741	Medtr4g011630.1	AT1G15520.1	0,00E+00	Glyma.15G012000	G7IMF2_MEDTR	0,00E+00	AT1G15520.1	pleiotropic drug resistance 12	0,00E+00
Medtr_v1_087003	Medtr0565s0010.1	AT1G29670.1	1,00E-155	Glyma.13G223100	G7J1P3_MEDTR	1,00E-134	AT1G29670.1	GDSL-like Lipase/Acylhydrolase superfamily protein	2,00E-103
Medtr_v1_087027	Medtr4g062510.1	AT5G06900.1	1,00E-114	Glyma.19G146800	C93A1_SOYBN	0,00E+00	AT5G06900.1	cytochrome P450, family 93, subfamily D, polypeptide 1	0,00E+00
Medtr_v1_010809	Medtr2g034720.1	AT5G49360.1	0,00E+00	Glyma.15G143700	G7IMV1_MEDTR	0,00E+00	AT5G49360.1	beta-xylosidase 1	0,00E+00
Medtr_v1_092529	Medtr4g062500.1	AT5G06900.1	1,00E-119	Glyma.03G143700	C93A1_SOYBN	0,00E+00	AT5G06900.1	cytochrome P450, family 93, subfamily D, polypeptide 1	0,00E+00
Medtr_v1_016378	Medtr3g034290.1	AT3G63470.1	0,00E+00	Glyma.09G226700	G7IZJ6_MEDTR	0,00E+00	AT3G63470.1	serine carboxypeptidase-like 40	0,00E+00
Medtr_v1_029654	Medtr5g008010.1	AT4G33270.1	0,00E+00	Glyma.08G228100	U5FR10_POPTR	0,00E+00	AT4G33270.1	Transducin family protein / WD-40 repeat family protein	0,00E+00
Medtr_v1_076047	#N/D	AT2G27740.1	5,00E-54	Glyma.12G086500	-	-	AT2G27740.1	Family of unknown function (DUF662)	4,00E-77
Medtr_v1_034144	Medtr5g069310.1	AT3G47340.1	1,00E-52	Glyma.11G171400	I1LLM2_SOYBN	0,00E+00	AT3G47340.1	glutamine-dependent asparagine synthase 1	0,00E+00
Medtr_v1_028182	Medtr4g133750.1	AT3G26300.1	1,00E-111	Glyma.01G179400	G7KAT1_MEDTR	0,00E+00	AT3G26300.1	cytochrome P450, family 71, subfamily B, polypeptide 34	2,00E-133
Medtr_v1_074514	Medtr3g073730.1	AT5G10180.1	0,00E+00	Glyma.14G169300	G7KA22_MEDTR	0,00E+00	AT5G10180.1	slufate transporter 2;1	0,00E+00
Medtr_v1_048319	Medtr8g006270.1	AT4G01070.1	1,00E-109	Glyma.05G181100	I2BH27_LINUS	2,00E-172	AT4G01070.1	UDP-Glycosyltransferase superfamily protein	1,00E-133
Medtr_v1_034662	Medtr5g076140.1	AT5G37010.1	3,00E-61	Glyma.18G048000	D0ABF7_9ORYZ	5,00E-61	AT5G37010.1	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: N-terminal protein myristoylation; EXPR	2,00E-69
Medtr_v1_018211	Medtr3g080990.1	AT4G10850.1	2,00E-77	Glyma.13G002700	M7Z837_TRIUA	3,00E-96	AT4G10850.1	Nodulin MtN3 family protein	1,00E-93
Medtr_v1_078735	Medtr6g073013.1	AT1G14440.2	4,00E-34	Glyma.17G002200	B0LK13_SARHE	3,00E-95	AT1G14440.2	homeobox protein 31	9,00E-72
Medtr_v1_048729	Medtr8g018130.1	AT5G17680.1	1,00E-131	Glyma.08G190600	G7JUR1_MEDTR	0,00E+00	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative	3,00E-103
Medtr_v1_001536	Medtr3g102810.1	AT5G67360.1	5,00E-30	Glyma.06G045100	D3YBD3_TRIRP	0,00E+00	AT5G67360.1	Subtilase family protein	0,00E+00
Medtr_v1_032102	Medtr5g037080.1	AT1G70000.2	5,00E-62	Glyma.02G026300	Q0PJI4_SOYBN	0,00E+00	AT1G70000.2	myb-like transcription factor family protein	4,00E-77
Medtr_v1_092069	Medtr6g033675.1	AT3G16520.3	1,00E-111	Glyma.16G175400	C6ZJB3_PUEML	0,00E+00	AT3G16520.3	UDP-glucosyl transferase 88A1	2,00E-124

Supplementary Table 5. Continuation…

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*p*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_069967	Medtr4g005360.1	AT3G20270.3	3,00E-73	Glyma.08G224700	B6UI71_MAIZE	2,00E-80	AT3G20270.3	lipid-binding serum glycoprotein family protein	5,00E-77
Medtr_v1_031602	Medtr5g030950.1	AT4G37930.1	0,00E+00	Glyma.09G202000	I1L4U2_SOYBN	0,00E+00	AT4G37930.1	serine transhydroxymethyltransferase 1	0,00E+00
Medtr_v1_045883	Medtr7g093830.1	AT5G42500.1	1,00E-42	Glyma.03G147700	G7KWC3_MEDTR	2,00E-110	AT5G42500.1	Disease resistance-responsive (dirigent-like protein) family protein	3,00E-52
Medtr_v1_048315	Medtr8g006220.1	AT4G01070.1	1,00E-111	Glyma.05G181100	I2BH27_LINUS	2,00E-172	AT4G01070.1	UDP-Glycosyltransferase superfamily protein	1,00E-133
Medtr_v1_027024	Medtr6g465230.1	AT5G22990.1	2,00E-17	Glyma.10G295200	G7JE49_MEDTR	4,00E-19	AT5G22990.1	C2H2-like zinc finger protein	5,00E-22
Medtr_v1_051967	Medtr8g098360.1	AT5G12250.1	0,00E+00	Glyma.05G207500	G7LHC6_MEDTR	0,00E+00	AT5G12250.1	beta-6 tubulin	0,00E+00
Medtr_v1_051967	Medtr8g098360.1	AT5G12250.1	0,00E+00	Glyma.08G014200	TBB1_SOYBN	0,00E+00	AT5G12250.1	beta-6 tubulin	0,00E+00
Medtr_v1_076300	#N/D	AT3G16520.3	1,00E-106	Glyma.16G175400	C6ZJB3_PUEML	0,00E+00	AT3G16520.3	UDP-glucosyl transferase 88A1	2,00E-124
Medtr_v1_035118	Medtr5g081860.1	AT2G38090.1	1,00E-69	Glyma.03G078000	Q0PJE0_SOYBN	1,00E-160	AT2G38090.1	Duplicated homeodomain-like superfamily protein	3,00E-120
Medtr_v1_072909	Medtr6g033740.1	AT4G28000.1	0,00E+00	Glyma.09G127000	U5FLT2_POPTR	0,00E+00	AT4G28000.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0,00E+00
Medtr_v1_099143	Medtr5g017210.1	AT3G50330.1	4,00E-42	Glyma.06G040800	Q00M67_SOYBN	0,00E+00	AT3G50330.1	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	5,00E-45
Medtr_v1_025533	Medtr4g082820.1	AT4G01070.1	3,00E-50	Glyma.05G181100	I2BH27_LINUS	2,00E-172	AT4G01070.1	UDP-Glycosyltransferase superfamily protein	1,00E-133
Medtr_v1_084463	Medtr0021s0360.1	AT1G68710.1	0,00E+00	Glyma.01G092900	ALA10_ARATH	0,00E+00	AT1G68710.1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein	0,00E+00
Medtr_v1_067075	#N/D	AT5G33320.1	1,00E-135	Glyma.17G088500	P93390_TOBAC	0,00E+00	AT5G33320.1	Glucose-6-phosphate/phosphate translocator-related	1,00E-171
Medtr_v1_027067	Medtr4g115360.2	AT2G10940.2	2,00E-43	Glyma.13G160400	B6TLL4_MAIZE	4,00E-43	AT2G10940.2	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	9,00E-50
Medtr_v1_025449	Medtr4g081230.1	AT5G17680.1	1,00E-126	Glyma.08G190600	G7JUR1_MEDTR	0,00E+00	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative	3,00E-103
Medtr_v1_084574	Medtr3g465430.1	AT3G51770.1	0,00E+00	Glyma.18G059700	F4J4I9_ARATH	0,00E+00	AT3G51770.2	tetratricopeptide repeat (TPR)-containing protein	0,00E+00
Medtr_v1_070121	Medtr7g102070.1	AT2G36910.1	0,00E+00	Glyma.19G184300	U5GAD2_POPTR	0,00E+00	AT2G36910.1	ATP binding cassette subfamily B1	0,00E+00
Medtr_v1_019895	Medtr3g105610.1	AT2G29120.1	0,00E+00	Glyma.13G233300	G7IK74_MEDTR	0,00E+00	AT2G29120.1	glutamate receptor 2.7	8,00E-127
Medtr_v1_027512	Medtr4g123880.1	AT1G56720.3	0,00E+00	Glyma.03G228800	B9N417_POPTR	0,00E+00	AT1G56720.3	Protein kinase superfamily protein	0,00E+00
Medtr_v1_052567	Medtr8g107510.1	AT1G70300.1	0,00E+00	Glyma.01G031800	I1J5A7_SOYBN	0,00E+00	AT1G70300.1	K+ uptake permease 6	0,00E+00

Supplementary Table 5. Continuation…

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*p*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_087263	Medtr0451s0010.1	AT3G26300.1	1,00E-110	Glyma.01G179400	G7KAT1_MEDTR	0,00E+00	AT3G26300.1	cytochrome P450, family 71, subfamily B, polypeptide 34	2,00E-133
Medtr_v1_036075	Medtr5g092990.1	AT5G17680.1	2,00E-33	Glyma.08G190600	G7JUR1_MEDTR	0,00E+00	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative	3,00E-103
Medtr_v1_010071	Medtr2g020710.1	AT5G18840.1	0,00E+00	Glyma.15G099500	G7IFL9_MEDTR	0,00E+00	AT5G18840.1	Major facilitator superfamily protein	0,00E+00
Medtr_v1_015300	Medtr3g011390.1	AT3G07040.1	1,00E-127	Glyma.18G093300	G7J1I4_MEDTR	0,00E+00	AT3G07040.1	NB-ARC domain-containing disease resistance protein	4,00E-100
Medtr_v1_077258	Medtr4g056520.1	AT2G27920.1	1,00E-167	Glyma.11G191200	SCP51_ARATH	0,00E+00	AT2G27920.1	serine carboxypeptidase-like 51	0,00E+00
Medtr_v1_074584	Medtr3g464580.1	AT3G47340.1	0,00E+00	Glyma.11G171400	I1LLM2_SOYBN	0,00E+00	AT3G47340.1	glutamine-dependent asparagine synthase 1	0,00E+00
Medtr_v1_024375	Medtr4g061610.1	AT3G45780.2	1,30E-01	Glyma.12G074100	Q5DW44_PHAVU	0,00E+00	AT3G45780.2	phototropin 1	0,00E+00
Medtr_v1_020096	Medtr3g107980.1	AT5G25510.1	0,00E+00	Glyma.04G071700	G7JBQ2_MEDTR	0,00E+00	AT5G25510.1	Protein phosphatase 2A regulatory B subunit family protein	0,00E+00
Medtr_v1_027626	Medtr4g125180.1	AT2G05920.1	0,00E+00	Glyma.09G070000	G7IGE1_MEDTR	0,00E+00	AT2G05920.1	Subtilase family protein	0,00E+00
Medtr_v1_077859	Medtr4g126270.1	AT5G56890.1	0,00E+00	Glyma.13G104300	G7III2_MEDTR	0,00E+00	AT5G56890.1	Protein kinase superfamily protein	0,00E+00
Medtr_v1_040475	Medtr6g086115.1	AT3G24480.1	1,00E-159	Glyma.05G232000	LRX4_ARATH	0,00E+00	AT3G24480.1	Leucine-rich repeat (LRR) family protein	0,00E+00
Medtr_v1_040475	Medtr6g086115.1	AT3G24480.1	1,00E-159	Glyma.08G039400	LRX4_ARATH	0,00E+00	AT3G24480.1	Leucine-rich repeat (LRR) family protein	0,00E+00
Medtr_v1_025546	Medtr4g082950.1	AT3G29320.1	3,00E-49	Glyma.20G026700	K7N0Z8_SOYBN	0,00E+00	AT3G29320.1	Glycosyl transferase, family 35	0,00E+00
Medtr_v1_046321	Medtr7g104050.1	AT3G52910.1	4,00E-71	Glyma.19G192700	G7L5Z0_MEDTR	3,00E-157	AT3G52910.1	growth-regulating factor 4	9,00E-89
Medtr_v1_026135	Medtr4g093850.5	AT5G64330.1	0,00E+00	Glyma.07G180900	G7JV64_MEDTR	0,00E+00	AT5G64330.1	Phototropic-responsive NPH3 family protein	0,00E+00
Medtr_v1_085553	#N/D	AT4G13870.1	2,00E-15	Glyma.18G194300	E5GBP0_CUCME	2,00E-57	AT4G13870.2	Werner syndrome-like exonuclease	2,00E-23
Medtr_v1_048739	Medtr8g018260.1	AT5G17680.1	1,00E-144	Glyma.08G190600	G7JUR1_MEDTR	0,00E+00	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative	3,00E-103
Medtr_v1_048740	Medtr8g018270.1	AT5G17680.1	1,00E-124	Glyma.08G190600	G7JUR1_MEDTR	0,00E+00	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative	3,00E-103
Medtr_v1_009287	Medtr2g009340.1	AT5G64330.1	0,00E+00	Glyma.07G180900	G7JV64_MEDTR	0,00E+00	AT5G64330.1	Phototropic-responsive NPH3 family protein	0,00E+00
Medtr_v1_033153	Medtr5g055310.1	AT1G25390.1	1,00E-121	Glyma.10G271400	G7IC81_MEDTR	0,00E+00	AT1G25390.1	Protein kinase superfamily protein	2,00E-156
Medtr_v1_049375	Medtr4g043630.2	AT5G17680.1	0,00E+00	Glyma.08G190600	G7JUR1_MEDTR	0,00E+00	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative	3,00E-103

Supplementary Table 6. Genes down-regulated in mature legume seeds during imbibition versus immature.

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*e*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_031642	Medtr5g031430.1	AT4G39330.1	1,00E-117	Glyma.05G187700	F8SMC1_SOYBN	0,00E+00	AT4G39330.1	cinnamyl alcohol dehydrogenase 9	0,00E+00
Medtr_v1_086340	#N/D	AT4G25640.2	6,00E-11	Glyma.10G232200	I1LDN0_SOYBN	0,00E+00	AT4G25640.1	detoxifying efflux carrier 35	0,00E+00
Medtr_v1_101150	Medtr7g101425.1	AT2G37040.1	7,00E-77	Glyma.10G058200	I1L905_SOYBN	0,00E+00	AT2G37040.1	PHE ammonia lyase 1	0,00E+00
Medtr_v1_046955	Medtr7g114990.1	AT1G60710.1	1,00E-113	Glyma.03G248700	AKR1_SOYBN	3,00E-141	AT1G60710.1	NAD(P)-linked oxidoreductase superfamily protein	2,00E-94
Medtr_v1_042249	Medtr7g016820.1	AT5G13930.1	0	Glyma.01G228700	CHS7_SOYBN	0,00E+00	AT5G13930.1	Chalcone and stilbene synthase family protein	0,00E+00
Medtr_v1_001801	Medtr1g040875.1	AT3G23000.1	1,00E-136	Glyma.10G001700	C6F1T7_SOYBN	0,00E+00	AT3G23000.1	CBL-interacting protein kinase 7	3,00E-167
Medtr_v1_050434	Medtr8g068870.1	AT5G54160.1	6,00E-98	Glyma.04G227700	F8SMC0_SOYBN	0,00E+00	AT5G54160.1	O-methyltransferase 1	0,00E+00
Medtr_v1_086822	Medtr7g096990.1	AT5G44120.3	8,00E-23	Glyma.19G164800	Q9FEC5_SOYBN	0,00E+00	AT5G44120.3	RmlC-like cupins superfamily protein	3,00E-50
Medtr_v1_068697	Medtr8g096910.1	AT4G11650.1	7,00E-49	Glyma.05G204800	P21_SOYBN	6,00E-128	AT4G11650.1	osmotin 34	3,00E-111
Medtr_v1_044768	Medtr7g079820.1	AT3G22640.1	6,00E-68	Glyma.10G028300	SBP_SOYBN	0,00E+00	AT3G22640.1	cupin family protein	6,00E-48
Medtr_v1_087438	Medtr1g103400.1	AT3G22640.1	1,00E-61	Glyma.20G148300	Q94LX2_SOYBN	0,00E+00	AT3G22640.1	cupin family protein	2,00E-65
Medtr_v1_078904	Medtr4g025670.1	AT3G60730.1	1,00E-139	Glyma.03G028900	I1JKR6_SOYBN	0,00E+00	AT3G60730.1	Plant invertase/pectin methylesterase inhibitor superfamily	0,00E+00
Medtr_v1_013494	Medtr2g099570.1	AT1G55020.1	0	Glyma.13G347600	B3TDK4_SOYBN	0,00E+00	AT1G55020.1	lipoxygenase 1	0,00E+00
Medtr_v1_044759	Medtr7g079730.1	AT3G22640.1	3,00E-64	Glyma.10G246500	Q948Y0_SOYBN	0,00E+00	AT3G22640.1	cupin family protein	6,00E-62
Medtr_v1_033736	Medtr5g064060.1	AT1G07400.1	5,00E-57	Glyma.12G058200	G8DZS0_9ROSI	1,00E-39	AT1G07400.1	HSP20-like chaperones superfamily protein	5,00E-26
Medtr_v1_006089	Medtr1g072600.1	AT5G44120.3	5,00E-47	Glyma.10G037100	Q9SB11_SOYBN	0,00E+00	AT5G44120.3	RmlC-like cupins superfamily protein	7,00E-54

Supplementary Table 6. Continuation…

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*e*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_005824	Medtr1g061760.1	AT5G03795.1	1,00E-167	Glyma.03G189100	T2DLF4_PHAVU	0,00E+00	AT5G03795.1	Exostosin family protein	0,00E+00
Medtr_v1_032184	Medtr5g038060.1	AT1G69870.1	0	Glyma.18G127200	G7K9P5_MEDTR	0,00E+00	AT1G69870.1	nitrate transporter 1.7	0,00E+00
Medtr_v1_046844	Medtr7g113790.1	AT1G03220.1	9,00E-80	Glyma.03G239700	7SB1_SOYBN	0,00E+00	AT1G03220.1	Eukaryotic aspartyl protease family protein	4,00E-84
Medtr_v1_035086	Medtr5g081530.1	AT1G53540.1	1,00E-60	Glyma.02G076600	Q9XET1_SOYBN	7,00E-108	AT1G53540.1	HSP20-like chaperones superfamily protein	2,00E-47
Medtr_v1_074228	Medtr4g056590.1	AT2G27940.1	3,00E-23	Glyma.11G191800	ATL57_ARATH	2,00E-18	AT2G27940.1	RING/U-box superfamily protein	3,00E-21
Medtr_v1_044473	Medtr7g076320.1	AT2G01770.1	6,00E-18	Glyma.18G228200	C1E094_MICSR	7,00E-60	AT2G01770.1	vacuolar iron transporter 1	6,00E-20
Medtr_v1_019599	Medtr3g100650.1	AT4G24660.1	1,00E-55	Glyma.09G170500	Q5IR71_SOYBN	3,00E-96	AT4G24660.1	homeobox protein 22	2,00E-41
Medtr_v1_029725	Medtr5g008850.1	AT4G10020.1	1,00E-140	Glyma.11G015100	G7K984_MEDTR	0,00E+00	AT4G10020.1	hydroxysteroid dehydrogenase 5	1,00E-177
Medtr_v1_095133	#N/D	AT2G20770.1	0,37	Glyma.12G111000	G7KIT8_MEDTR	0,00E+00	AT2G20770.1	GCR2-like 2	0,00E+00
Medtr_v1_039918	Medtr2g008120.1	AT1G43710.1	1,00E-122	Glyma.15G050600	G7IL92_MEDTR	0,00E+00	AT1G43710.1	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein	0,00E+00
Medtr_v1_000534	Medtr8g016520.1	ATCG00170.1	1,00E-124	Glyma.02G221400	RPOC2_SOYBN	3,00E-84	ATCG00170.1	DNA-directed RNA polymerase family protein	1,00E-73
Medtr_v1_023449	Medtr4g051110.1	ATCG00160.1	1,00E-110	Glyma.03G031100	RR2_SOYBN	1,00E-83	ATCG00160.1	ribosomal protein S2	3,00E-70
Medtr_v1_017981	Medtr3g077280.1	AT5G40390.1	0	Glyma.06G179200	G7J4U7_MEDTR	0,00E+00	AT5G40390.1	Raffinose synthase family protein	0,00E+00
Medtr_v1_004965	Medtr1g025490.1	AT1G75810.1	1,00E-24	Glyma.17G258800	Q9ZQY8_ARATH	1,00E-15	AT1G75810.1	unknown protein	1,00E-34
Medtr_v1_086153	#N/D	AT1G61070.1	0,000001	Glyma.13G278000	Q2KM12_9FABA	1,00E-42	AT1G61070.1	low-molecular-weight cysteine-rich 66	2,00E-08
Medtr_v1_012500	Medtr2g081590.1	AT5G25610.1	1,00E-24	Glyma.12G217400	O24482_SOYBN	0,00E+00	AT5G25610.1	BURP domain-containing protein	9,00E-32
Medtr_v1_027908	Medtr4g130540.1	AT1G16030.1	0	Glyma.17G072400	HSP70_SOYBN	0,00E+00	AT1G16030.1	heat shock protein 70B	0,00E+00

Supplementary Table 6. Continuation…

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*e*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_085808	#N/D	AT5G03850.1	2,00E-16	Glyma.03G205800	G7KLD3_MEDTR	6,00E-37	AT5G03850.1	Nucleic acid-binding, OB-fold-like protein	9,00E-34
Medtr_v1_066901	Medtr2g090605.1	AT1G03790.1	2,00E-71	Glyma.12G116900	U5FMC6_POPTR	2,00E-100	AT1G03790.1	Zinc finger C-x8-C-x5-C-x3-H type family protein	2,00E-73
Medtr_v1_044763	Medtr7g079770.1	AT3G22640.1	2,00E-68	Glyma.20G148400	Q94LX2_SOYBN	0,00E+00	AT3G22640.1	cupin family protein	2,00E-65
Medtr_v1_030731	Medtr5g020630.1	AT5G48330.1	1,00E-143	Glyma.01G173400	G7KEX6_MEDTR	0,00E+00	AT5G48330.1	Regulator of chromosome condensation (RCC1) family protein	1,00E-163
Medtr_v1_035904	Medtr5g090990.1	AT1G06570.1	1,00E-175	Glyma.02G284600	A5Z1N7_SOYBN	0,00E+00	AT1G06570.2	phytoene desaturation 1	4,00E-159
Medtr_v1_051080	Medtr8g079680.1	AT5G66860.1	1,00E-84	Glyma.05G081300	G7LGD1_MEDTR	1,00E-137	AT5G66860.1	Ribosomal protein L25/Gln-tRNA synthetase, anti-codon-binding domain	9,00E-106
Medtr_v1_034784	Medtr5g077580.1	AT1G30220.1	0	Glyma.09G087400	G7IGM5_MEDTR	0,00E+00	AT1G30220.1	inositol transporter 2	0,00E+00
Medtr_v1_027404	Medtr4g121940.1	AT3G16990.1	2,00E-79	Glyma.17G032700	PM36_SOYBN	6,00E-169	AT3G16990.1	Haem oxygenase-like, multi-helical	2,00E-98
Medtr_v1_031594	Medtr5g030870.1	AT2G32210.1	0,0001	Glyma.08G171400	-	-	AT2G32210.1	unknown protein	2,00E-15
Medtr_v1_007348	Medtr1g098250.1	AT4G09830.1	3,00E-52	Glyma.08G144000	K0A0U9_ELAGV	2,00E-79	AT4G09830.1	Uncharacterised conserved protein UCP009193	2,00E-57
Medtr_v1_031272	Medtr5g026780.1	AT1G31830.1	0	Glyma.20G106200	D7KGN1_ARALL	0,00E+00	AT1G31830.2	Amino acid permease family protein	0,00E+00
Medtr_v1_026593	Medtr4g104210.1	AT3G27160.1	6,00E-21	Glyma.05G044800	G7JMN6_MEDTR	7,00E-34	AT3G27160.1	Ribosomal protein S21 family protein	2,00E-35
Medtr_v1_068329	Medtr4g037465.1	AT3G06240.1	1,00E-20	Glyma.17G017600	D3YBB5_TRIRP	1,00E-95	AT3G06240.1	F-box family protein	8,00E-23
Medtr_v1_013048	Medtr2g089580.1	ATCG00020.1	1,00E-56	Glyma.11G162200	Q85V85_9POAL	7,00E-133	ATCG00020.1	photosystem II reaction center protein A	7,00E-135
Medtr_v1_044022	Medtr7g070020.1	AT4G32520.2	5,00E-27	Glyma.13G222300	C6ZJY9_SOYBN	0,00E+00	AT4G32520.2	serine hydroxymethyltransferase 3	0,00E+00
Medtr_v1_074217	Medtr8g033390.1	AT2G45630.2	2,00E-90	Glyma.07G082000	E6NU30_JATCU	1,00E-127	AT2G45630.2	D-isomer specific 2-hydroxyacid dehydrogenase family protein	2,00E-115
Medtr_v1_009590	Medtr2g013460.1	AT4G14440.1	2,00E-79	Glyma.10G007900	G7IPS1_MEDTR	6,00E-93	AT4G14440.1	3-hydroxyacyl-CoA dehydratase 1	2,00E-90

Supplementary Table 6. Continuation…

Probe_id	Seq_id Mt4.0v1	Arabidopsis	*e*-value	id *Gmax*	Tax Hit	*p*-value	Arabidopsis	Name Arabdopsis gene	*p*-value
Medtr_v1_068217	#N/D	AT1G71190.1	9,00E-79	Glyma.08G131300	K7V5D8_MAIZE	7,00E-104	AT1G71190.1	senescence associated gene 18	2,00E-111
Medtr_v1_089600	Medtr3g113270.1	AT3G27230.1	1,00E-100	Glyma.04G025700	D7MJ22_ARALL	0,00E+00	AT3G27230.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	0,00E+00
Medtr_v1_010187	Medtr2g025050.1	AT4G04020.1	1,00E-103	Glyma.09G008100	Q2HTH4_MEDTR	7,00E-149	AT4G04020.1	fibrillin	6,00E-131
Medtr_v1_101070	#N/D	AT1G55850.1	9,00E-15	Glyma.08G330600	G7IV14_MEDTR	0,00E+00	AT1G55850.1	cellulose synthase like E1	0,00E+00
Medtr_v1_050686	Medtr8g074040.1	AT2G17650.1	0	Glyma.02G042200	G7L9H4_MEDTR	0,00E+00	AT2G17650.1	AMP-dependent synthetase and ligase family protein	0,00E+00
Medtr_v1_077620	#N/D	AT3G20390.1	5,00E-64	Glyma.08G316200	B4UWE2_ARAHY	5,00E-104	AT3G20390.1	endoribonuclease L-PSP family protein	4,00E-79