Deciphering the impacts of heat stress on the acquisition of seed quality traits in Medicago truncatula

Zhijuan Chen

DOCTORAT ECOLOGIE BRETAGNE GEOSCIENCES LOIRE AGRONOMIE ALIMENTATION

l'institut Agro agriculture • alimentation • environnement

AGRO CAMPUS OUEST

THESE DE DOCTORAT DE

L'INSTITUT NATIONAL D'ENSEIGNEMENT SUPERIEUR POUR L'AGRICULTURE, L'ALIMENTATION ET L'ENVIRONNEMENT
ECOLE INTERNE AGROCAMPUS OUEST

ECOLE DOCTORALE N° 600
Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation
Spécialité : Biochimie, biologie moléculaire et cellulaire

Par
Zhijuan CHEN

Deciphering the impacts of heat stress on the acquisition of seed quality traits in *Medicago truncatula*

Thèse présentée et soutenue à Angers, le 11 Mai 2021
Unité de recherche : Institut de Recherche en Horticulture et Semences (IRHS)
Thèse N° : C155 2021-13

Rapporteurs avant soutenance :

Petr SMYKAL	Associate Professor, Department of Botany, Faculty of Science, Palacky University
Vanessa VERNOUD	Chargée de recherche, INRAE Centre de Dijon, UMR Agroécologie

Composition du Jury :

Président :	Emmanuel GEOFFRIAU	Professeur, Institut Agro - Agrocampus Ouest, UMR IRHS
Examinateurs :	Véronique GRUBER	Professeure, Université de Paris, UMR IPS2
	Bertrand DUBREUCQ	Directeur de recherche, INRAE Centre Ile de France, UMR IJPB
Dir. de thèse :	Olivier LEPRINCE	Professeur, Institut Agro - Agrocampus Ouest, UMR IRHS
Co-encadrant :	Jerome VERDIER	Chargé de recherche, INRAE UMR IRHS

Thesis Submitted to

AGROCAMPUS OUEST

Ecole Doctorale: ECOLOGIE, GEOSCIENCE, AGRONOMIE, ALIMENTATION (EGAAL)

Unité de recherche: UMR 1345 Institut de Recherche en Horticulture et Semences (IRHS)

By

Zhijuan CHEN

DECIPHERING THE IMPACTS OF HEAT STRESS ON THE ACQUISITION OF SEED QUALITY TRAITS IN *MEDICAGO TRUNCATULA*

Thesis Defense in Angers (France) on May 11, 2021

Petr SMYKAL	Associate Professor, Department of Botany, Faculty of Science, Palacky University	Rapporteur
Vanessa VERNOUD	Chargée de recherche, INRAE Centre de Dijon, UMR Agroécologie	Rapportrice
Emmanuel GEOFFRIAU	Professeur, Institut Agro - Agrocampus Ouest, UMR IRHS	Président
Véronique GRUBER	Professeure, Université de Paris, UMR IPS2	Examinatrice
Bertrand DUBREUCQ	Directeur de recherche, INRAE Centre Ile de France, UMR IJPB	Examinateur
Olivier LEPRINCE	Professeur, Institut Agro - Agrocampus Ouest, UMR IRHS	Dir. de thèse
Jerome VERDIER	Chargé de recherche, INRAE Centre Pays de la Loire, UMR IRHS	Co-encadrant

ACKNOWLEDGEMENT

Three years PhD study is finishing, but the scene of just coming to the lab is still fresh in my mind. The PhD study made me learn a lot in a new place and a new research field. Thanks to Agrocampus Ouest and the Research Institute of Horticulture and Seeds for giving me the opportunity to study in France and the Chinese Scholarship Council for providing me financial support.

I would like to express my deepest thanks to Professor Oliver Leprince and Dr. Jerome Verdier for the guidance, inspiration and support during my thesis study. It was an invaluable experience to work together with them. I was deeply impressed by Olivier's rigorous academic attitude. His expert advices on seed physiology helped me solve the difficulties that I met in the project. I am extremely grateful to Jerome for his help over the past few years, since I did the master rotation in his lab in the Shanghai Center for Plant Stress Biology (PSC). With his elaborate guidance and patience, I learned a lot and was able to complete my study and this thesis. Sincerest thanks to Dr. Julia Buitink for all of her advice and care. Every time we talked, I felt calm and full of energy. Their innovative scientific mind and diligent work style inspired me, which is also the direction of my future study and work.

I would like to express my heartfelt thanks to my thesis committee members, Richard Thompson, John Harada and Etienne Bucher, for their helpful ideas and suggestions to help me move forward.

I am sincerely grateful to the jury members of my thesis, Dr. Petr Smykal and Dr. Vanessa Vernoud for accepting to being the rapporteurs of my thesis and Prof. Emmanuel Geoffriau, Prof. Veronique Gruber and Dr. Bertrand Dubreucq for being the examiners.

Thank you very much, all the members in SEED lab, who gave me a lot of help and suggestions on the lab work in the process of my doctoral study. I would like to thank Benoit for his help in the germination experiments and dissecting seed tissues, by the way improving my French. Thank Joseph for taking care the plants, especially for large population of *Medicago* HapMap. Thanks to David Lalanne for introducing the laboratory organization and teaching me some molecular experiments. Thanks to David Windels for helping me with the ChIP experiment. Thanks to Jaiana for proofreading and suggestions for this thesis. Thank Martine and Thu for giving me help when I needed them. I am lucky to carry out my PhD study in a well-organized

and friendly group. I also would like to express thanks to Daniel Beucher for managing the growth chamber. Thanks also to Remi Gardet and Daniel Sochard from ImHorPhen team for help in greenhouse. And thank all IRHS members for providing me kindly technical support.

I also thank Elise Bizouerne, Elise Rethore, Jean Baptiste, Marthe and Martin in our doctoral students' office. It is a good memory to stay with you during my doctoral study. Thank you for all of your help in my daily life and wish all of you have a bright future.

Special thanks to my family and boyfriend for their understanding and support, they gave me courage and strength to continue to move forward! Thank my dear friends in France and in China for being with me when I was happy or depressed. This period of study and life in Angers is happy and impressive I wish all you a better life and work in the future!

ABSTRACT

Keywords: Seed maturation, Heat stress, Seed quality, Transcriptomics, Epigenetics, GWAS

Legume seeds represent a crucial source of human food and animal feed to ensure global food security. However, global warming and climate change endanger agricultural productions by impacting seed yield and seed quality. In *Medicago truncatula*, a model plant for legumes, the timing of seed maturation is greatly affected by adverse environmental growing conditions, leading to alterations of final seed traits. In this study, we intend to explore and describe seed molecular processes underlying the alteration of seed traits in response to heat stress to identify candidate genes regulating stress response and phenotypic plasticity of seed quality traits.

Two complementary approaches were implemented to identify candidate genes: (i) deciphering the (epi)genetic regulations occurring in developing seed tissues in response to heat stress in the reference genotype A17, and (ii) exploiting the natural diversity of the *Medicago truncatula* HapMap population using genome-wide association study approaches. As a result, we observed an intense molecular reprogramming of seed maturation processes via the alteration of the main regulators of seed development. A set of candidate genes potentially regulating heat stress response and plasticity in seeds was identified. Among them, two genes were selected for functional characterizations and based on preliminary results, *MtHAP3* showed a potential role in the reprogramming of early seed maturation determining seed size and seed longevity in response to heat stress; and *MtMIEL1* acted as a regulator of seed germination plasticity in response to heat stress.

French summary: Impacts du stress thermique sur l'acquisition des caractères de qualité des semences chez *Medicago truncatula*

Résumé: Les graines de légumineuses sont une source de nourriture importante pour garantir la sécurité alimentaire mondiale. Cependant, le réchauffement climatique met en danger les productions agricoles en affectant le rendement et la qualité des semences. Chez *M. truncatula*, une plante modèle des légumineuses, la maturation des graines est fortement affectée par les conditions environnementales, conduisant à l'altération de la qualité des graines. Dans cette étude, nous proposons d'explorer les processus moléculaires des semences aboutissant à la modification de la qualité des graines suite au stress chaleur pour identifier des gènes candidats régulant la réponse au stress et la plasticité phénotypique des caractères de qualité de semences.

Deux approches ont été développées pour identifier les gènes candidats: (i) décrire les régulations (épi)génétiques intervenant lors du développement des tissus des graines en réponse au stress thermique dans le génotype de référence *M. truncatula* A17, et (ii) exploiter la diversité naturelle présente dans la population HapMap de *M. truncatula* en utilisant des approches d'association pangénomique. Un ensemble de gènes candidats potentiellement régulant la réponse au stress thermique et la plasticité des graines a été identifié. Parmi eux, deux gènes ont été sélectionnés pour être caractérisés fonctionnellement et sur la base de résultats préliminaires, *MtHAP3* pourrait avoir un rôle dans la reprogrammation de la maturation des graines en réponse au stress thermique ; et *MtMIEL1* pourrait agir en régulateur de la plasticité de germination des graines en réponse à un stress thermique.

Mots clefs: Maturation des graines, Stress thermique, Qualité des semences, Transcriptomique, Epigénétique, GWAS

I. Introduction

Les graines de légumineuses et de céréales ont une importance agronomique forte car elles présentent des teneurs élevées en protéines et en glucides, ainsi qu'en fibres et en minéraux (Deshpande, 1992; McKevith, 2004). Ces deux grandes familles de plantes, les Fabacées et les graminées, sont à l'origine de la plupart des aliments destinés aux humains et aux animaux, estimant que les graines fournissent directement 70% des calories humaines (Bewley and Black, 1994). En tant que légumineuse modèle, *Medicago truncatula* a un petit génome (environ 375 Mb) diploïde (2n = 16), et un cycle de vie court, présentant un délai de 6-7 mois de graine à graine. *Medicago truncatula* est une plante d'origine méditerranéenne et une plante modèle des légumineuses depuis 1990, date à laquelle elle a été

utilisée pour la première fois pour étudier la symbiose rhizobium-légumineuse (Barker et al., 1990). Cette espèce est une plante modèle précieuse et utile dans le domaine de la recherche sur les légumineuses avec de nombreuses études génétiques et génomiques publiées dans tous les domaines de la biologie végétale (pour revue Kang et al., 2016). Le génome de *M. truncatula* a été séquencé pour la première fois en 2011 (Young et al., 2011) et est toujours en cours de développement avec une cinquième version (Pecrix et al., 2018). Des populations mutantes de *M. truncatula* ont été développées par différentes approches et servent a explorer les domaines de la recherche biologique, notamment la biologie des semences, la fixation symbiotique de l'azote et les réponses au stress abiotique (Kang et al., 2016).

Les graines de *Medicago*, comme les autres graines, se composent de trois tissus différents (c'est-à-dire l'embryon, l'albumen et le tegument), qui jouent des rôles différents pendant le développement des graines. L'embryon est le zygote diploïde issu de la pollinisation contenant un allèle mâle et un allèle femelle. L'embryon stockera les molécules de stockage comme dans la plupart des espèces de dicotylédones puis se dessèchera finalement pour atteindre moins de 10% de teneur en eau par rapport au poids sec de la graine afin d'être conservée pendant une période de temps prolongée. L'embryon occupe la majeure partie de l'espace d'une graine mature et deviendra la future plante. L'albumen est un tissu triploïde car il provient de la double fécondation d'un gamete male avec les deux noyaux polaires de la grande cellule centrale (Bewley et al., 2013). L'albumen servira de tissu de stockage transitoire et d'intégrateur central du développement des graines. Au cours du développement de la graine de *Medicago*, l'albumen est absorbé par l'embryon. Dans les graines matures de *Medicago*, l'albumen residuel est limité à quelques couches cellulaires entourant l'embryon et la plupart des cellules sont mortes (pour revue Verdier et al., 2019). Concernant l'enveloppe de la graine, les teguments, c'est un tissu diploïde d'origine maternelle qui protège l'embryon et l'albumen et qui importera des nutriments à l'embryon. Pour produire des graines saines, la croissance et le développement des trois tissus de la graine sont fortement coordonnés (pour revue Bewley et al., 2013).

Le développement des semences est un processus de développement complexe qui commence par la fertilisation. Chez *M. truncatula*, le développement des graines peut être divisé en trois phases: (i) l'embryogenèse, (ii) le remplissage des graines et (iii) la maturation tardive. Au cours du développement de la graine, le poids sec de la graine augmente progressivement et s'accompagne d'une perte de teneur en eau. L'embryogenèse est définie par le développement de l'embryon à partir de la double fécondation (c'est-à-dire 0 jour après la pollinisation) jusqu'à 12 jours après la pollinisation (JAP). Le développement embryonnaire subit des divisions cellulaires, et chaque stade de développement sera caractérisé par son aspect morphologique: stades globulaire, cœur, torpille, cotylédon courbé et enfin le stade de la graine sèche ou mature (Bewley and Black, 1994). Le remplissage des graines et la maturation tardive se chevauchent partiellement. Le remplissage des graines concerne principalement l'allongement des cellules embryonnaires et l'accumulation de molécules de stockage induisant

l'augmentation du poids des graines. Lors de la phase de remplissage de la graine, la graine accumule des molécules de stockage accompagnées d'une diminution drastique de la teneur en eau (pour revue Verdier et al., 2019). La maturation tardive est une période identifiée de 24 JAP à la maturité des graines, au cours de laquelle les graines acquièrent les caractères de qualité germinative, tels que la la vigueur des graines et la longévité (Verdier et al., 2013). La période de maturation tardive des graines varie selon les espèces. Par exemple, la maturation tardive chez *Arabidopsis* est parmi l'une des plus courtes avec un chevauchement entre les phases de remplissage et de maturation tardive des graines, ce qui explique que chez cette espèce, on distingue généralement deux phases: l'embryogenese et la maturation, tandis que *M. truncatula* présente une période relativement longue de la phase de maturation tardive par rapport au temps total de développement de la graine (pour revue Leprince et al., 2017).

Dans un environnement en constant changement, les plantes doivent réagir et s'adapter à différentes conditions environnementales. L'augmentation des températures moyennes renforce la nécessité de comprendre comment les plantes réagissent au stress thermique pour améliorer la tolérance à la chaleur des cultures. Au cours des dernières années, de nombreuses études ont rapporté comment les plantes ressentent, réagissent et s'adaptent au stress thermique (revue de J. Liu et al., 2015; Ohama et al., 2017). La voie de régulation transcriptionnelle de la réponse au stress thermique a été principalement étudiée dans les tissus végétatifs tels que les racines et les feuilles. Jusqu'à présent, très peu d'études se sont concentrées sur les impacts des stress pendant le développement des semences et leurs impacts sur les qualités des semences. Dans une étude précédente, il a été démontré que le froid, la chaleur et le stress osmotique ont des effets significatifs sur le développement des graines (Righetti et al., 2015). Dans ces différentes conditions environnementales, la durée du stade d'embryogenèse était relativement stable, alors que la durée du stade de maturation était sévèrement affectée. Les graines qui se sont développées dans des conditions de stress thermique ont été les plus touchées avec une réduction importante de la durée de la phase de maturation, condensée de 30 jours dans des conditions témoins à environ 10 jours pendant un stress thermique (Righetti et al., 2015). Grâce à la caractérisation des processus physiologiques, il a été observé que l'acquisition de la tolérance à la dessiccation était un processus robuste, sans impact majeur des différentes conditions environnementales. À l'inverse, l'acquisition de la longévité etait fortement affectée par différentes conditions environnementales (Righetti et al., 2015). De nombreuses études sur différentes espèces ont montré des résultats similaires. Chez le riz, une température élevée raccourcissait la période de reproduction (He et al., 2014), la longévité potentielle des graines produites dans des conditions plus chaudes était inférieure à celle des environnements plus froids (Ellis et al., 1993) et des plantes exposées à des conditions transitoires modérées. Un stress thermique (35°C) au début du développement des graines a produit des graines matures mais avec un taux de germination different que les conditions témoins en raison de l'altération de la biosynthèse de l'ABA et du GA (Begcy et al., 2018). Dans le soja, le stress thermique pendant le développement des graines a entraîné une germination plus faible des graines, une diminution du poids

des graines et plusieurs changements dans le contenu métabolique (Chebrolu et al., 2016). Une étude sur des plantes alpines de lit de neige a montré que les graines de plantes exposées à des températures chaudes peuvent vivre plus longtemps que les graines produites à l'état naturel (Bernareggi et al., 2015). Toutes ces études démontrent que le stress thermique pendant le développement des graines affecte la durée de la maturation des graines et a un impact sur des caractéristiques importantes des graines telles que la longévité et la vigueur des graines. Cependant, la caractérisation des mécanismes moléculaires impliqués dans la réponse au stress thermique des graines est encore mal comprise, en particulier pour les espèces de légumineuses.

Au cours des dernières décennies, la régulation épigénétique est de plus en plus etudiee dans différentes espèces liées à divers processus biologiques. La méthylation de l'ADN et la modification des histones sont les principales modifications épigénétiques pour réguler l'expression des gènes. Chez les plantes, la régulation épigénétique se produit non seulement dans diverses phases de développement des plantes, mais également pendant les réponses aux stress biotiques et abiotiques (Chinnusamy and Zhu, 2009 ; Ahmad et al., 2010). Au cours de ma thèse, nous avons rassemble dans une revue les connaissances sur les mécanismes épigénétiques impliqués dans les processus de développement et le stress (a)biotique chez les légumineuses (Windels et al., 2020). Lorsque les plantes modifient leur programme de développement dans les phases dites de transition, telles que la floraison au développement des graines ou le développement des graines à la germination, les protéines du groupe polycomb (PcG) jouent un rôle crucial dans le contrôle de ces processus (pour revue Yang et al., 2017; Yan et al., 2020). Les protéines PcG forment généralement différents complexes proteiques qui sont organisés en complexe répressif polycomb 1 (PRC1) et complexe répressif polycomb 2 (PRC2) en fonction de leur action pour condenser des régions de chromatine spécifiques via différentes modifications d'histones. Les complexes PRC1 condensent la chromatine par dépôt d'ubiquitine sur la lysine 119 de l'histone 2A (H2AK119Ub). En revanche, les complexes PRC2 sont capables de condenser la chromatine en ciblant la lysine 27 sur l'histone 3 par triméthylation (H3K27me3). Cette triméthylation de H3 en lysine 27 est une forte marque répressive, largement distribuée dans le génome (c'est-à-dire environ 20% des gènes d'*Arabidopsis* ont été trouvés liés à la marque H3K27me3), qui se dépose dynamiquement et est éliminée au cours du développement de la plante, conduisant à la condensation de la chromatine (Zheng and Chen, 2011). Les complexes PRC2 régulent l'initiation de la phase de développement des graines et sont nécessaires dans la transition entre l'embryogenèse et la maturation des graines. En effet, il a été montré qu'en l'absence de fertilisation, FIS-PRC2 est impliqué dans la répression du développement de l'albumen, tandis que VRN-PRC2 et EMF-PRC2 dans la répression du développement du tégument de la graine (Lau et al., 2012; Figueiredo and Köhler, 2018). De plus, des mutations dans le complexe FIS-PRC2 ont altéré le développement des trois tissus de la graine, conduisant à l'avortement de la graine (Robert et al., 2018; Robert, 2019).

Au cours de la maturation des graines, une augmentation de la condensation de la chromatine a été observée, agissant potentiellement via ABI3 (Van Zanten et al., 2011). La taille des noyaux et l'état de la chromatine sont récupérés lors de la germination des graines. La tolérance à la dessiccation, la longévité et les qualités germinatives sont acquises au cours de la maturation des graines et marque répressive H3K27me3 est responsable du maintien de la compaction de la chromatine et de la répression de la transcription des gènes impliqués dans le développement des graines, dans la présente étude, nous proposons d'explorer si la condensation de la chromatine est impliquée dans l'acquisition de la longévité et des qualités germinatives. En outre, cette étude examinera également si les marques répressives H3K27me3 sont affectées par les conditions de stress thermique pendant le développement de la graine, et si ces changements sont associés à la réponse physiologique de la graine au stress thermique.

Comme développé dans les sections précédentes, dans un contexte d'urgence mondial, le stress thermique a un impact dramatique sur le développement des plantes, y compris l'acquisition de caractéristiques de qualité des semences, ce qui représente une menace importante pour la sécurité alimentaire. Les réponses moléculaires à la chaleur ont été étudiées de manière intensive dans les parties végétatives des plantes, mais on en sait peu sur les processus moléculaires se produisant pendant le développement des graines.

L'objectif de ce projet de doctorat est d'explorer et de décrire les processus moléculaires des semences affectés lors d'un stress thermique afin d'identifier des gènes candidats susceptibles d'alerter ou de réguler la réponse au stress et d'expliquer la plasticité phénotypique des caractères de qualité des semences suite à un stress thermique.

Pour atteindre cet objectif, ce projet de thèse a été initialement organisé en trois parties (WP).

WP1 a l'intention de démêler les mécanismes moléculaires se produisant dans les tissus des graines sous-jacents aux changements physiologiques tels que le poids, la longévité et la vigueur des graines lorsque les graines ont subi un stress thermique au cours de leur développement. Dans cette partie, les changements physiologiques, transcriptomiques et épigénétiques (c.-à-d. Méthylation de l'ADN et H3K27me3) en réponse au stress thermique dans les semences seront explorés en utilisant le génotype de référence A17 de *Medicago truncatula* pour (i) identifier les gènes candidats affectant le développement des semences pendant le stress thermique, puis (ii) découvrir le niveau de (épi) régulation (épi)génétique de ces gènes candidats.

En parallèle, WP2 propose d'utiliser la diversité génétique naturelle présente dans la collection *Medicago truncatula* HapMap pour identifier les loci/gènes potentiellement impliqués dans la plasticité

des traits de semences en réponse au stress thermique en utilisant une approche d'étude d'association à l'échelle du génome (GWAS).

Enfin, WP3 correspond à un ensemble intégratif de WP1 et WP2 pour sélectionner une courte liste de gènes candidats pertinents, qui pourraient être impliqués dans la plasticité des traits de semences en réponse au stress thermique et agissant à différents niveaux de régulation (épi)génétique. Cette partie comprend également une caractérisation fonctionnelle préliminaire de gènes candidats à l'aide des populations mutantes d'insertion *Medicago Tnt1* et *Arabidopsis T-DNA* afin de sélectionner les gènes les plus pertinents pour les études futures.

En raison de laboutissement partiel du WP3 en raison de la pandémie Covid-19 et des différents problemes qui ont ralenti notre progression, nous avons décidé de reformater ce manuscrit de thèse en deux parties: un chapitre 2 correspondant aux analyses à l'échelle du génome de la dynamique du transcriptome, du méthylome et de la chromatine de tissus isoles de la graine dans des conditions de stress thermique et un chapitre 3 correspondant à l'étude de la régulation des caractères de la graine dans des conditions optimales et de stress thermique à l'aide des accessions naturelles de *Medicago truncatula* et des études d'association à l'échelle du génome. L'intégration des deux WP n'a pas été effectuée ; cependant, quelques analyses fonctionnelles préliminaires de gènes candidats identifiés dans les chapitres respectifs sont présentées à la fin des deux chapitres.

II. Analyses à l'échelle du génome de la dynamique du transcriptome, du méthylome et de la chromatine de tissus de semences isolés dans des conditions de stress thermique

Dans ce chapitre, nous avons voulu décrire les réponses physiologiques et moléculaires au stress thermique pendant le stade de reproduction des plantes, encore inexploré, correspondant au développement/maturation des graines afin d'identifier les impacts du stress sur l'acquisition des caractères de qualité des graines. Cette étude exploratoire utilisera des analyses à l'échelle du génome pour définir la dynamique du transcriptome, du méthylome de l'ADN et de la chromatine à partir de tissus isolés de graines en réponse à des conditions de stress thermique à quatre stades clés de la maturation des graines en utilisant le génotype de référence A17 de *Medicago truncatula*. Cette analyse vise à identifier les processus moléculaires pertinents et les gènes candidats régulant les traits physiologiques des semences à partir de différents tissus de semences et à différents niveaux de régulation (épi)génétique.

Après la récolte de graines matures développées dans des conditions contrôle (20°C) et de stress thermique (26°C), les graines ont été immédiatement séchées à 44% d'humidité relative (HR) (en utilisant une solution saturée de K_2CO_3 à 20°C) pendant trois jours. La dormance physiologique a été libérée en utilisant un traitement post-récolte par stockage de graines matures pendant 6 mois à température ambiante, puis les graines ont été ensuite stockées à 4°C dans l'obscurité pour des analyses ultérieures. La première caractérisation morphologique à consister à mesurer le poids / la taille des graines. En effet, nous avons observé une différence visible de poids / taille des graines entre les graines matures produites à 20°C et 26°C. Les plantes cultivées dans des conditions de stress thermique ont produit des graines plus petites que les graines de la condition témoin dans le génotype A17 de référence de *Medicago truncatula*. La longévité et la germination des graines sont des caractéristiques importantes de la qualité des graines. Les deux sont acquis pendant le développement de la graine, plus précisément pendant la phase de maturation chez *M. truncatula*. Tout d'abord, afin de vérifier si le stress thermique influence la longévité des graines, nous avons utilisé les conditions de vieillissement artificiel de 75% d'humidité relative à 35°C pour accélérer le vieillissement des graines. Nous avons observé que les graines produites à 26°C présentaient une plus grande capacité de longévité et étaient capables d'être viables plus longtemps dans des conditions de vieillissement artificiel par rapport aux graines produites à 20°C. Concernant la germination des graines dans des conditions de germination à basse température (10°C), il n'y avait pas de différence concernant le taux de germination final, qui atteignait 100% dans les deux lots de semences. Cependant, nous avons observé une diminution de la vitesse de germination des graines soumises à un stress thermique par rapport aux graines témoins.

Démêler les mécanismes moléculaires qui sous-tendent les changements physiologiques tels que le poids, la longévité et la vigueur des graines lorsque les graines ont subi un stress thermique au cours de leur développement chez *M. truncatula* A17. Nous avons combiné l'acquisition de ces processus physiologiques majeurs avec des données moléculaires obtenues à partir des données d'expression des stades de développement des graines produites dans des conditions de stress thermique et disponibles dans Righetti et al. (2015) en générant une matrice de corrélation pour identifier les stades correspondants en fonction des changements transcriptomiques globaux. En croisant ces informations, nous avons décidé de sélectionner quatre stades, nommées S1, S2, S3 et S4, correspondant à i) le début de la maturation des graines (S1: 17 JAP à partir de graines à 20°C et 14 JAP à partir de graines à 26°C), ii) après l'acquisition de la tolérance à la dessiccation et lors du remplissage des semences (S2: 26 JAP à partir de semences à 20°C et 17 JAP à partir de semences à 26°C), iii) début de l'acquisition de la longévité et à la fin du remplissage des semences (S3: 36 JAP à partir de graines à 20°C et 22 JAP à partir de graines à 26°C) et vi) à la maturité des graines (c.-à-d. graines sèches) (S4: 44 JAP à partir de graines à 20°C et 28 JAP à partir de graines à 26°C). Pour révéler les mécanismes moléculaires sous-jacents de ces quatre stades affectés par le stress thermique, nous avons d'abord analysé les changements

du transcriptome des trois tissus isolés de graines à ces quatre stades de développement au cours de la maturation des graines dans des conditions de contrôle et de stress thermique.

Par des analyses de transcriptome utilisant la comparaison des cinétiques de developpement pour l'embryon et l'albumen et des comparaisons « deux a deux » pour le tegument de la graine entre les conditions de contrôle et de stress dans les tissus de graines isolés, 12766, 11430 et 5761 gènes ont été identifiés comme différentiellement exprimés (DEG, False Discovery Rate (FDR) <5%) dans l'embryon, l'albumen et l'enveloppe de la graine respectivement et un ensemble commun de 1274 DEG dans les trois tissus de la graine ont été mis en évidence. Pour mieux comprendre le role de ces gènes nous avons regardé les classes fonctionnelles correspondant aux gènes différentiellement exprimés grace a des analyses de sur-représentation sur ces listes de DEG. L'une des classes fonctionnelles les plus surreprésentées était 'stress.abiotique.Heat' dans les trois tissus de la graine, qui a validé nos conditions de croissance et confirmé que l'intensité de stress thermique que nous avons appliquée pendant le développement des graines était suffisante pour induire une réponse au stress. Les classes fonctionnelles liées à 'DNA synthesis.Chromatin structure', 'Photosynthesis.Light reaction' et 'development.storage protein' se sont également révélés enrichis en DEG communs dans les trois tissus de graines. Outre l'intérêt pour les DEG communs lors de stress thermique dans tous les tissus de la graine, nous avons cherché à déterminer si et comment chaque tissu de graine réagirait au stress thermique. Pour cela, nous avons également effectué des analyses de sur-représentation sur les DEG spécifiques aux tissus, dont les expressions étaient statistiquement régulées à la hausse ou à la baisse dans l'un des trois tissus de la graine.

Étant donné que le stress thermique influence la duree du développement des graines et les caractéristiques de maturation des graines chez *Medicago truncatula*, nous avons examiné si les gènes régulateurs clés du développement des graines, *LEC1, ABI3, FUS3 et LEC2* (c'est-à-dire les gènes *LAFL*), étaient affectés par le stress thermique pendant le développement des graines. En utilisant l'annotation du génome de *Medicago* (version 5) et l'outil de recherche d'alignement local de base (BLAST), nous avons identifié des orthologues putatifs de ces gènes et extrait leurs profils d'expression au cours du développement des graines de *M. truncatula*. En outre, nous avons également examiné les profils d'expression de deux facteurs transcriptionnels important, *MtABI4* (MtrunA17_Chr5g0437371) et *MtABI5* (MtrunA17_Chr7g0266211), qui jouent un rôle important dans le développement des semences et sont impliqués dans la voie de signalisation ABA (Finkelstein et al., 1998 ; Finkelstein and Lynch, 2000). *ABI4* est un régulateur positif de la dormance primaire des graines chez *Arabidopsis* (Shu et al., 2013). *ABI5* régule non seulement la dormance et la germination des graines, mais joue également un rôle dans la longévité des graines chez *Medicago truncatula* (Zinsmeister et al., 2016). Les résultats ont montré que les niveaux de transcription de ces régulateurs clés du développement des graines sont affectés par le stress thermique et qu'en raison de leur rôle pléiotrope dans les mécanismes de maturation des graines, ils pourraient représenter de bons gènes candidats pour expliquer les phénotypes de graines

soumis à un stress thermique. Cela suggère également que le stress thermique induit une reprogrammation transcriptionnelle intense du développement/maturation des graines en modifiant les expressions géniques des gènes régulateurs centraux. Fait intéressant, lors de la recherche de l'orthologue du gène *LEC1*, nous avons identifié un gène très étroitement apparenté qui correspond à un gène *HAP3/NF-YB6* annoté comme *MtLEC1-LIKE* (*MtLIL*, MtrunA17_Chr4g0076381). Chez *Arabidopsis*, *L1L* est capable de compléter la mutation *lec1* et est nécessaire pour le développement de l'embryon et l'initiation du programme de maturation chez *Arabidopsis* (Kwong et al., 2003). Chez *M. truncatula*, ce gène *HAP3/MtL1L* était fortement exprimé en S1 dans l'embryon et en S1-S2 dans l'albumen, plus tardivement que *MtLEC1*. C'était egalement l'un des gènes les plus statistiquement differentiellement exprimés après un stress thermique, ce qui nous a amenés à considérer ce gène comme un gène candidat pour une validation fonctionnelle ultérieure.

Afin de se concentrer sur l'effet du stress thermique dans les mécanismes moléculaires de l'embryon pour identifier les gènes candidats qui expliquent le changement des caractéristiques physiologiques, nous avons effectué une analyse d'inference du réseau de corrélation des gènes (WGCNA, Langfelder and Horvath, 2008) en intégrant des données transcriptomiques et physiologiques pour mettre en évidence des gènes candidats potentiellement regulant ces caracteres physiologiques. Nous avons combiné les transcriptomes normalisés montrant des expressions différentielles dans l'embryon et les traits physiologiques liés à la cinétique du poids des graines et à l'acquisition de la longévité des graines obtenues aux mêmes stades que ceux utilisées dans l'analyse du transcriptome dans les deux conditions à 20°C et 26°C. Deux modules présentant des corrélations très significatives avec nos traits de semence observés ont été identifiés: un module, appelé coral3, montrant une corrélation de 0,84 (p-value de $3,10^{-7}$) avec l'acquisition de la longévité (P50) et dans une moindre mesure un module, appelé palevioletred3, montrant une corrélation de 0,67 (valeur p-value de $3,10^{-4}$) avec l'augmentation du poids des graines (DW). Les gènes ayant une appartenance élevée au module (ou une connectivité basée sur l'eigengène, MM>0,9) et la signification du gène (GS>0,9) dans ces deux modules ont été sélectionnés pour des gènes candidats potentiellement importants pour réguler la longévité des graines et le poids sec dans des conditions de stress thermique, ce qui contiennent respectivement 154 et 15 gènes.

La dynamique de la méthylation de l'ADN pendant le développement de l'embryon dans des conditions de culture contrôle et de stress thermique a egalement été étudiée. Le pourcentage global de méthylcytosines identifiées dans chaque contexte entre les conditions temoins et stressés n'ont pas radicalement changé entre les stades de développement S2, S3 et S4, mais nous avons observé une augmentation lors de S1, comme observé chez *Arabidopsis* (Bouyer et al., 2017). Les taux globaux moyens de méthylation étaient également similaires à ceux observés dans les graines d'*Arabidopsis*. Les pourcentages de méthylation dans les contextes CHG et CHH à S1 ont montré un niveau plus élevé dans les embryons produits dans des conditions de stress thermique. Pour avoir un aperçu des régions

differentielllement méthylées après un stress thermique, nous avons identifié de 1 à 2175 régions méthylées différentiellement (DMR) en fonction des stades de développement de l'embryon et des contextes de méthylation comprenant des régions hypo-méthylées (c'est-à-dire une diminution de la méthylation de l'ADN due au stress thermique) ou régions hyper-méthylées (c'est-à-dire une augmentation de la méthylation de l'ADN due au stress thermique). En cartographiant les DMR dans les régions génomiques pour distinguer les DMR situés dans les promoteurs de 1 kb, les séquences codant pour les gènes et les éléments transposables (TE), nous avons observé une transition des distributions de DMR dans les régions génomiques entre S1 et les stades ultérieurs. L'augmentation de DMR localisées dans les TE (60%) et dans les promoteurs de 1 kb (23%) à S1 a d'abord suggéré que l'augmentation globale de methylcytosine survenant pendant le stress thermique à S1 pourrait être due à une augmentation des régions méthylées dans les TE et les promoteurs de 1 kb. Pour mieux comprendre le rôle de la méthylation différentielle de l'ADN en S1 et son rôle potentiel sur la régulation transcriptionnelle, nous avons combiné les données de nos méthylomes et transcriptomes en sélectionnant les régions hypo-méthylées situées dans des promoteurs de 1 kb avec les gènes régulés à la hausse correspondants et inversement les régions hyper-méthylées situées dans des promoteurs de 1 kb avec des gènes régulés à la baisse correspondants dans des conditions de stress thermique. En conséquence, nous avons identifié 97 gènes qui pourraient potentiellement être transcriptionellement régulés via la méthylation de l'ADN dans l'embryon au stade S1 après un stress thermique.

En ce qui concerne la dynamique H3K27me3 se produisant pendant le développement de l'embryon à partir de graines produites pendant les conditions temoin et de stress thermique, nous avons observé une diminution constante des régions genomiques de fixation avec H3K27me3 en condition de contrôle pendant le développement des graines. Fait intéressant, dans les graines soumises à un stress thermique, le nombre de sites génomiques fixés à H3K27me3 était encore très élevé au cours du développement des graines, ce qui suggère qu'à ces stades les marques H3K27me3 étaient dépendantes du stress et qu'elles pourraient jouer un rôle important dans la régulation du développement des graines soumises à un stress thermique. L'analyse de la dynamique de la marque H3K27me3 au cours du développement de l'embryon sous contrôle en conditions de stress thermique doit être analysée plus en détail.

Ce chapitre se termine par l'analyse fonctionnelle préliminaire des mutants *l1l* chez *Medicago*, nous n'avons obtenu que des résultats prometteurs mais préliminaires concernant le potentiel et le rôle spécifique de *MtL1L* dans la régulation du développement des graines sous stress thermique. *MtL1L* pourrait être un gène candidat prometteur avec un rôle spécifique dans la régulation du développement des graines dans des conditions de stress thermique chez *Medicago truncatula*, ce qui pourrait expliquer sa forte régulation à la baisse (c'est-à-dire environ 10 fois) chez l'embryon à S1 sous stress thermique. Suite à cette expérience, nous avons re-criblé la population mutante *Medicago Tnt1* et identifié deux

nouvelles lignées mutantes d'insertion. Les quatre lignées mutantes sont actuellement en croissance et seront utilisées pour confirmer les résultats préliminaires.

III. Étude de la régulation des caractères des semences dans des conditions optimales et de stress thermique à l'aide des accessions naturelles de *Medicago truncatula* et des études d'association à l'échelle du génome

Des accessions naturelles de *Medicago truncatula* provenant de différentes zones géographiques principalement distribuées dans le bassin méditerranéen et de divers environnements climatiques ont été collectées pour constituer une collection d'accessions de *Medicago truncatula* (Ronfort et al., 2006). Récemment, avec le développement des technologies de séquençage haut-debit et des plates-formes de phénotypage, l'étude d'association à l'échelle du génome (GWAS) est devenue une méthode populaire pour étudier le role des polymorphismes de sequence en lien avec divers traits. À cette fin, un sous-ensemble pertinent de ces accessions a été, par la suite, sélectionné en fonction de la structuration de la population, et 288 accessions de *Medicago* ont été séquencées à l'aide de technologies de séquençage de nouvelle génération afin d'identifier les polymorphismes de nucléotides uniques (SNP) existant au sein de cette population (http://www.medicagohapmap.org/home/view). Cette ressource d'accessions aux polymorphismes connus, appelée population de carte haplotypique (HapMap) de *Medicago truncatula*, représente une ressource génétique précieuse en biologie des légumineuses pour identifier les gènes candidats associés à l'acquisition de caractères agronomiques (pour revue Mammadov et al., 2012 ; Govindaraj et al., 2015).

Dans ce chapitre, nous avons profité de la population de *M. truncatula* HapMap pour réaliser des études d'association à l'échelle du génome. Une première approche nous a permis de comparer des modèles linéaires mixtes classiques à locus unique tels que EMMA (Hyun et al., 2008) avec des modèles multi-locus plus récents tels que le modèle fixe et aléatoire d'unification des probabilités circulantes (FarmCPU) (Liu et al., 2016) en utilisant 162 accessions *Medicago* HapMap pour identifier les locus candidats régulant la taille et la composition des semences produites dans des conditions optimales. En parallèle, une deuxième approche consistait à cultiver 200 accessions HapMap et à produire des graines matures dans des conditions optimales et de stress thermique pour obtenir des loci/gènes candidats putatifs impliqués dans le contrôle de la taille des graines et des caractères de vigueur des graines dans des conditions optimales et de stress, ainsi que des loci/gènes contrôlant la plasticité phénotypique de ces caractéristiques de performance des graines en réponse à un stress thermique.

Dans la première section, deux modèles différents pour les prédictions d'association GWAS à l'échelle du génome ont été appliqués pour les données phénotypiques normalisées: un modèle linéaire

mixte classique à locus unique (EMMA) avec prise en compte de la parenté et de la structure de population, et un modèle multi-locus (FarmCPU) avec correction de la structure de la population. Lors de l'exécution du modèle FarmCPU multi-locus, nous avons observé des graphiques quantile-quantile (QQ) avec un meilleur ajustement entre les résultats attendus et observés suivant la distribution d'hypothèse nulle attendue des p valeurs. Ces graphiques QQ reflètent que la plupart des SNP testés n'ont pas de p valeurs significatives, à l'exception de quelques SNP qui ont un effet fort et significatif. De plus, les graphiques QQ obtenus après avoir exécuté l'algorithme EMMA ont généralement montré une courbe correspondant aux résultats observés en dessous de la courbe théorique (c.-à-d. Courbe dégonflée), ce qui suggére que ce modèle n'était pas le plus approprié pour cette étude d'association. Concernant les graphiques Manhattan obtenus à partir de différents modèles, nous avons également observé des différences entre EMMA et FarmCPU. En général, nous avons obtenu moins de bruit de fond avec FarmCPU, avec une localisation plus précise et des valeurs p-valeurs plus faibles des SNP que celles obtenues à partir du modèle linéaire mixte (MLM), en particulier lorsque l'analyse statistique a montré des SNP très significatifs. Les graphiques Manhattan obtenues à partir du MLM présentaient des « pics » plus larges constitués de plusieurs SNP significatifs (c'est-à-dire des clusters de SNP). Dans l'ensemble, nous avons noté que la plupart des SNP les plus significatifs ont été identifiés dans les deux méthodes, mais FarmCPU a fourni plus de sensibilite de détection, de puissance et de précision pour identifier les nucléotides à caractères quantitatifs (QTN). Par conséquent, nous avons décidé de nous concentrer sur le modèle mixte multi-locus avec FarmCPU dans les analyses ultérieures. Dans cette étude, nous avons tout de meme identifié des gènes/loci potentiellement impliqués dans à la fois le contrôle de la taille et de la teneur en protéines des graines, ce qui pourrait potentiellement permettre l’amelioration simultanément des valeurs nutritionnelles et des performances agronomiques des graines.

Dans la deuxième section, 200 accessions naturelles sélectionnées de la collection *Medicago truncatula* HapMap ont été cultivées en serre dans des conditions optimales et de stress thermique, mais en utilisant la même intensité lumineuse, photopériode. Des triplicats des 200 accessions ont été cultivés dans des conditions témoins à une température minimale de 20°C jour / 18°C nuit dans le cadre du projet ANR REGULEG (n°ANR-15-CE20-0001). Dans le cadre de mon projet de doctorat, trois replicats des 200 accessions ont été cultivés dans des conditions de stress thermique. Pour produire des graines matures soumises à un stress thermique, nous avons d'abord suivi la même procédure que les plantes témoins. Puis, au stade de la floraison, les plantes ont été déplacées vers une serre voisine avec la même photopériode et les mêmes conditions d'éclairage mais avec des températures minimales de 26°C jour et 24°C nuit jusqu'à atteindre la maturité des graines (c.a.d. conditions de stress thermique pour *Medicago*). Les graines produites à la fois dans des conditions optimales et de stress thermique ont été récoltées pour des analyses de caractères de qualité des graines.

Nous avons d'abord observé une diminution significative du rendement en graines (c.-à-d. nombre de gousses et de graines) des plantes cultivées dans des conditions de stress thermique. Le

nombre de graines limitées produites dans des conditions de stress de chaleur pour certaines accessions a impacté le nombre d'accessions disponibles pour l'analyse GWAS des caractères de qualité des semences. Par exemple, l'accession HM059 n'a pas produit de semences et n'a donc pu etre utilisé. En ce qui concerne les performances de germination et de longévité des graines, nous avons caractérisé phénotypiquement la longévité des graines et la germination des graines matures produites à partir de 151 et 112 accessions, respectivement. Les résultats ont montré que ce stress modéré appliqué à la floraison a impacté le poids final des graines, mais egalement la germination et la longevité. Les analyses d'association pangénomiques (GWAS) ont été réalisées pour identifier les loci putatifs ou les gènes impliqués dans la régulation des traits de semences et de leur plasticité en réponse au stress thermique. Nous avons identifié de nombreux QTNs forts et/ou genes candidats potentiels impliqués dans la régulation de ces traits sous stress thermique en utilisant des analyses combinées avec nos données de transcriptomiques. En utilisant des analyses GWAS et postGWAS en combinaison avec des données adéquates transcriptomiques nous avons pu identifier des gènes candidats solides régulant potentiellement les différents traits de semences. Parmi eux, *MtMIEL1*, un gène de type zinc-finger de la famille des domaines RING, a été validé comme reguler transcritionnellement et associée à la vitesse de germination dans des graines stressées par le chaud. Dans *Medicago*, nous avons montré que *MtMIEL1* etait transcriptionnellement regulée lors de la production de semences soud l'effet de la chaleur, et que son profil d'expression dans quatre différentes accessions de *Medicago* était positivement associé à leur vitesse de germination. Enfin, une analyse de perte de fonction de l'orthologue de *MIEL1* chez *Arabidopsis* a révélé son rôle de régulateur de la plasticité de la germination des graines en réponse au stress thermique.

IV. Conclusions et perspectives

Ce projet de thèse de trois ans visait à explorer les processus moléculaires des semences affectés par le stress thermique dans des tissus isolés de graines afin d'identifier les gènes candidats régulant potentiellement la réponse au stress chez *Medicago truncatula*. L'objectif principal a été atteint avec une caractérisation exhaustive des changements du transcriptome dans ces trois tissus de la graine tout au long du développement, ainsi qu'une caractérisation de la dynamique du méthylome et de la chromatine chez l'embryon après un stress thermique. De plus, plusieurs gènes candidats ont été identifiés et des analyses préliminaires de mutants ont été initiées pour valider nos choix.

L'une des principales réalisations de ce travail de doctorat a été de développer cette carte de la régulation (épi)génétique due à la réponse au stress thermique. Même si elles ne sont pas entièrement exploitées au terme de ces trois années, ces connaissances représentent un outil précieux pour les projets actuels/futurs de l'équipe SEED. Par exemple, ces connaissances ont déjà été utilisées dans d'autres projets tels que les projets ANR DESWITCH (2020-2023) ou RFI epiDT (2020-2021). Elles

deviendront encore plus précieuses dans les années à venir car ce projet de thèse a ete le projet fondateur de notre thematique sur le «stress des semences», mais maintenant plusieurs autres projets sont en cours liés au stress thermique chez différentes espèces mais aussi à différents stress (a)biotiques chez *Medicago truncatula*. Ensemble, ces projets devraient fournir, à moyen terme, un aperçu précis de la façon dont les semences perçoivent, répondent et s'adaptent aux stress au cours de leur développement. L'un des objectifs de ce travail de thèse est de valoriser ces données et de les rendre accessibles à la communauté scientifique. Dans cette perspective, nous collaborons actuellement avec le LIPM (Toulouse) pour développer un serveur web Medicago Gene Expression Atlas basé sur des données RNA-seq, qui contient déjà toutes les données transcriptomiques générées dans cette thèse et offre de nombreux outils de visualisation et d'analyse à partager pour pleinement exploiter ces données. En ce qui concerne les résultats des approches GWAS et la dynamique de l'épigénome (c'est-à-dire le méthylome et le H3K27me3), nous avons déjà mis en ligne toutes les données dans notre Jbrowse interne (c'est-à-dire le navigateur du génome), qui sera accessible au public peu après la publication de ces résultats.

La deuxième réalisation principale de ce projet a été d'identifier les gènes candidats pertinents pour mieux comprendre comment les graines détectent, réagissent et s'adaptent au stress thermique. Même si l'intégration des données des différentes parties de ce travail de thèse n'a pas été achevée en temps opportun, nous avons identifié des listes de gènes candidats dans chacune des sections et ces choix semblaient pertinents sur la base des résultats préliminaires des analyses de mutants (par exemple, pour *MtMIEL1* et *HAP3/MtL1L*). Bien entendu, le plan initial d'intégration de toutes les données OMICS afin d'affiner la liste des gènes candidats est toujours en cours avec certaines solutions déjà testées mais non discutées dans ce rapport comme CAMOCO qui nous a permis d'intégrer la transcriptomique et les données GWAS (Schaefer et al., 2018). À ce jour, deux gènes candidats solides sont à l'étude, *MtMIEL1* (MtrunA17_Chr2g0286331) et *HAP3/MtL1L* (MtrunA17_Chr4g0076381), et la caractérisation fonctionnelle de ces deux gènes se poursuivra. De plus, nous avons maintenant des lignées mutantes homozygotes pour plus de candidats tels que *MtABI5* (MtrunA17_Chr7g0266211), *MtDASH* (MtrunA17_Chr2g0282441) et *MtHAP2* (MtrunA17_Chr2g0300261). Même si *MtDASH* et *MtABI5* sont déjà des gènes publiés connus pour être impliqués dans les processus de développement des semences, leur régulation par les stress (thermiques) est intéressante à élucider, contrairement à *MtHAP2*, aucun rapport n'a encore été fait sur ce gene, donc cela pourrait représenter un candidat intéressant dans la réponse au stress thermique mais aussi dans le processus de maturation des graines.

TABLE OF CONTENTS

CHAPTER 1: LITERATURE REVIEW

1. *Medicago truncatula* as a model plant to study seed biology in legumes

Legume and cereal seeds have significant agronomic importance as they present high protein and carbohydrate contents, as well as fibers and minerals (Deshpande, 1992; McKevith, 2004). These two major plant families, the Fabaceae and Poaceae, are the source of most of human food and animal feed, estimating that seeds directly provide 70% of human calories (Bewley and Black, 1994). As a model legume plant, *Medicago truncatula* has a diploid (2n=16) and small genome (about 375 Mb), and short life cycle, presenting a 6 months' time from seed to seed. *Medicago truncatula* is a Mediterranean originated plant and has been a model plant of legumes since 1990, when it was first used to study rhizobia-legume symbiosis (Barker *et al.*, 1990). This species is a valuable and useful model plant in the field of legume research with many genetic and genomic studies published in all area of plant biology (for review Kang *et al.*, 2016). The genome of *M. truncatula* was first sequenced in 2011 (Young *et al.*, 2011) and has still been under development with a recent fifth release (Pecrix *et al.*, 2018). Mutant populations of *M. truncatula* have been developed by different approaches to be applied in biological research fields, including seed biology, symbiotic nitrogen fixation and abiotic stress responses (Kang *et al.*, 2016).

Medicago seeds consist of three different tissues (*i.e.* embryo, endosperm and seed coat), which play different roles during seed development. The embryo is the diploid zygote resulting from the pollination containing one male allele and one female allele. The embryo will store storage molecules like in most dicot species then finally desiccate to reach less than 10% of water content with respect to seed dry weight in order to be stored for extended period of time (Verdier *et al.* 2013). The embryo occupies most of the space in a mature seed and will become the future plant. The endosperm is a triploid tissue since it originates from the double fertilization of one sperm cell fuses with the two polar nuclei of the large central cell (Bewley *et al.*, 2013). The endosperm will serve as a transient storage tissue and central integrator of seed development. During *Medicago* seed development, the endosperm is absorbed by the embryo. In *Medicago* mature seeds, the remaining endosperm is limited to few cell layers surrounding embryo and most of the cells are dead (for review Verdier *et al.*, 2019). Concerning seed coat, it is a diploid tissue from maternal origin that protects embryo and endosperm and

that will import nutrients to the embryo. To produce healthy seeds, the growth and development of all three seed tissues are coordinated (for review Bewley *et al.* 2013).

2. Seed development and acquisition of seed qualities during maturation stage

Seed development is a complex developmental process that starts with fertilization. In *M. truncatula,* seed development could be divided in three phases: (i) embryogenesis, (ii) seed filling and (iii) late maturation. During seed development, seed dry weight is increasing gradually and is accompanied by the loss of water content (Figure 1.1). Embryogenesis is defined by embryo development from double fertilization (*i.e.* 0 day after pollination) until 12 days after pollination (DAP). Embryo development undergoes cell divisions, and each developmental stage will be characterized by its morphological aspect: globular, heart, torpedo, bent cotyledon stages and finally reaching the dry or mature seed stage (Bewley and Black, 1994). Seed filling and late maturation are partially overlapping. Seed filling mainly concerns embryo cell elongation and accumulation of storage molecules inducing the increase of seed weight. During the seed filling stage, seed accumulates storage molecules accompanying by a drastic decrease of water content (for review Verdier *et al.*, 2019). Late maturation is a period identified from 24 DAP to seed maturity, during which seeds acquire the germinative quality traits (Figure 1.1), such as the capacity to germinate, seed vigor, desiccation tolerance and longevity (Verdier *et al.*, 2013). The time frame of late seed maturation varies in different species. For example, the late maturation in *Arabidopsis* is among one of the shortest and to a large content overlap with seed filling phase, which explains that in this species we usually distinguish two phases: embryogeneisis and maturation, while *M. truncatula* exhibits a relatively long period of late maturation phase compared to the total seed developmental time (reviewed in Leprince *et al.* 2017).

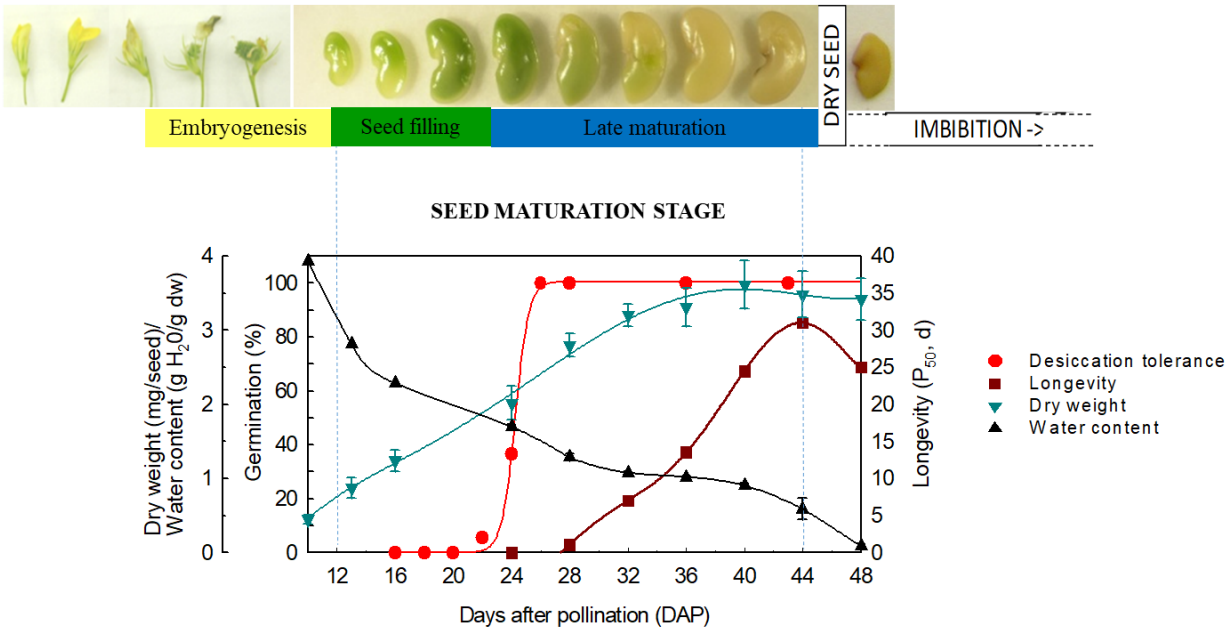


Figure 1.1 Physiological processes during seed maturation stage in *M. truncatula* (Adapted from Verdier *et al.*, 2013). Changes in seed dry weight, water content, acquisition of desiccation tolerance and acquisition of longevity are indicated in the curves following the days after pollination (DAP).

2.1 *LAFL* genes: keystone regulators of seed development

The coordinated seed development is controlled at different levels by metabolomic, hormonal and genetic factors such as several key transcription factors (TFs). In *Arabidopsis*, genetic analyses and molecular studies demonstrated that the LAFL genes, that are *LEAFY COTYLEDON1* (*LEC1*), *ABSCISIC ACID INSENSITIVE3* (*ABI3*), *FUSCA3* (*FUS3*) and *LEAFY COTYLEDON2* (*LEC2*), are master transcriptional regulators controlling many seed developmental processes, including embryogenesis and seed maturation (reviewed in Santos-Mendoza *et al.*, 2008; Verdier and Thompson, 2008; Braybrook and Harada, 2008; Fatihi *et al.*, 2016; Boulard *et al.*, 2017). *LEC1* is a CCAAT-box binding factors (CBF, also called Heme Activator Protein, HAP) gene family, which is required for both embryogenesis and maturation phase (West *et al.*, 1994; Lotan *et al.*, 1998; Kwong *et al.*, 2003; Jo *et al.*, 2019). This gene family encodes histone-like proteins, which function in a heterotrimer complex, with a NF-YA (CBF-B/HAP2) and a NF-YC (CBF-C/HAP5) subunits to be functional and bind the CCAAT box in the promoter regions of target genes. Recently, *LEC1* was identified as a pioneer transcription factor in resetting chromatin state from repressed state (marked by H3K27me3) to active state (H3K36me3) to activate the expression of *FLOWERING LOCUS C* (*FLC*), a

repressor of floral transition (Tao *et al.*, 2017). *LEC2*, *ABI3* and *FUS3* encode another class of transcription factors related to plant-specific B3 domain TFs (Giraudat *et al.*, 1992; Luerßen *et al.*, 1998; Stone *et al.*, 2001) and also named as "AFL" genes (Roscoe *et al.*, 2015). The expression of *LAFL* genes has a specific and highly regulated temporal and spatial pattern during seed development (Bewley *et al.*, 2013; Fatihi *et al.*, 2016) (Figure 1.2A). *LEC1* and *LEC2* are transiently expressed in early developing seed, with *LEC2* expressed later than *LEC1*, and their transcripts cannot be detected in mature seed. *FUS3* and *ABI3* are expressed later during seed development and transcripts of both genes are present in mature seed (Figure 1.2A).

Mutations of these master regulators results in dramatic phenotypes in seeds, including precocious germination, desiccation sensitivity and reduced storage proteins (Keith *et al.*, 1994; Meinke *et al.*, 1994; West *et al.*, 1994). Single and multiple mutant analyses demonstrated that LAFL genes function redundantly to control seed maturation, with *LEC1* acting upstream to regulate *LEC2*, *FUS3* and *ABI3* (Meinke *et al.*, 1994; Kagaya *et al.*, 2005; Roscoe *et al.*, 2015). Different studies showed that ectopic expressions of *LEC1* or *LEC2* could lead to somatic embryogenesis (Lotan *et al.*, 1998; Stone *et al.*, 2001) and induce the expression of *FUS3* and *ABI3* (Kagaya *et al.*, 2005; Santos Mendoza *et al.*, 2005). The regulatory network between the LAFL genes and their target genes is summarized in the simplified Figure 1.2B (for reviews Fatihi *et al.*, 2016; Lepiniec *et al.*, 2018). *LEC1* and *LEC2* positively regulate *FUS3* and *ABI3* and also active each other, in addition, *FUS3* and *ABI3* also regulate themselves (To *et al.*, 2006; Stone *et al.*, 2008). *WRINKLED1* (*WRI1*), which encodes an AP2/EREB domain protein regulating fatty acid synthesis, is a direct target of *LEC2* (Cernac and Benning, 2004; Baud *et al.*, 2007). *MYB118* is an endosperm-specific gene, which is activated by LEC2, but subsequently act as repressor of *LEC2* expression and downstream seed maturation genes (Barthole *et al.*, 2014).

The orthologs of these *LAFL* genes have been identified in most angiosperms, including dicots and monocots, particularly in the crop species and their functions seemed to be conserved (Verdier and Thompson, 2008; Peng and Weselake, 2013; Zhiguo *et al.*, 2018).

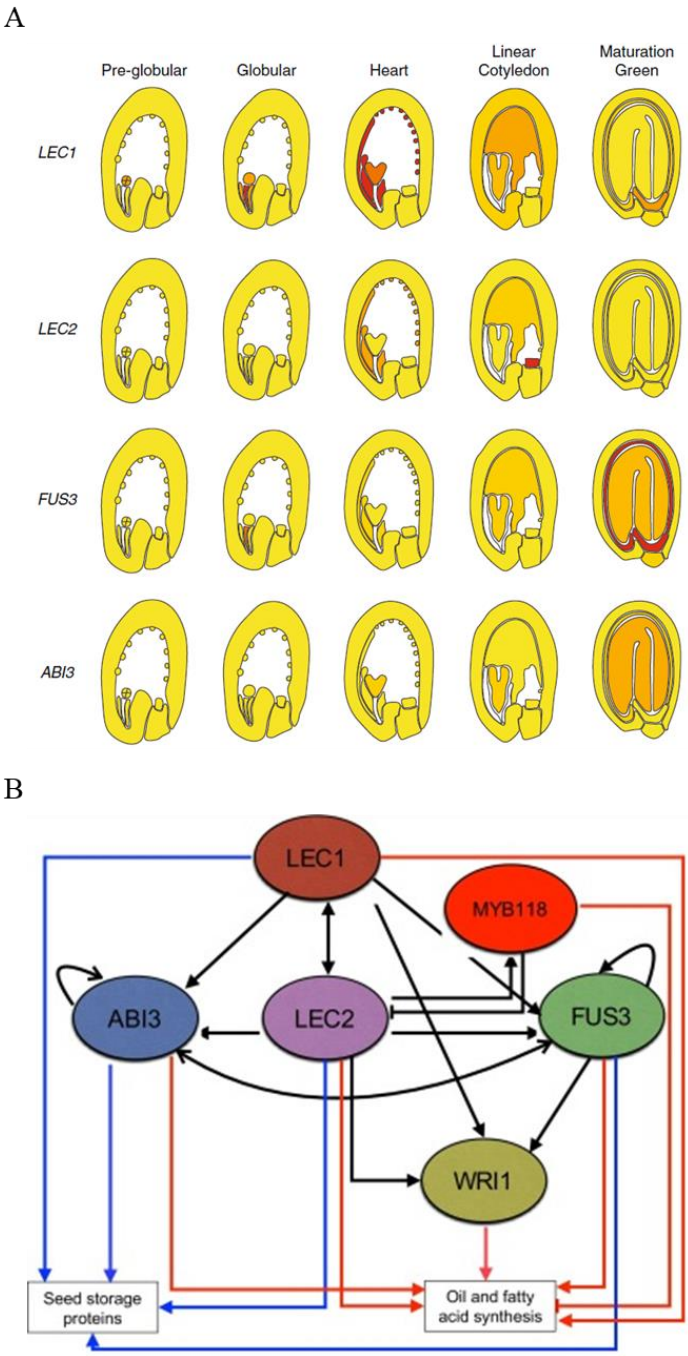


Figure 1.2 Expression patterns (A) and regulatory network (B) of *LAFL* genes (*LEC1*, *ABI3*, *FUS3* and *LEC2*) involved in seed development (Fatihi *et al.*, 2016). Expression levels of LAFL genes in different tissues at different stages are indicated with the color intensity from yellow (absent or low) to red (high). Expression patterns are from Arabidopsis eFP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi?dataSource=Seed) based on the transcriptomic data from Belmonte *et al.* (2013).

2.2 Seed yield

Plants produce various seed yield depending of the environmental conditions encountered. Seed number and seed size are the two components that determine the seed yield. The final seed size is determined during seed development as a result of the coordinated development and growth of three seed tissues (Chaudhury *et al.*, 2001; Li *et al.*, 2019). Indeed, the embryo and endosperm develop within the integuments, and all three tissues will affect the final seed size.

First, in the embryo of dicotyledonous plants such as *Arabidopsis thaliana* and *Medicago truncatula*, endosperm reserves are absorbed during the maturation stage by the embryo. Several studies showed that the mature seed size is determined by embryo size in the dicots (Sundaresan, 2005) and that embryo cell division during late embryogenesis is critical for the final embryo size (Lemontey *et al.*, 2000). This role of cell division during embryogenesis has also been demonstrated in two ecotypes of *M. truncatula* (W6-6016 and W6-6018), which exhibited larger seeds than the reference accession (A17), due to a longer cell division period during embryogenesis, potentially explained by a difference in hormonal balance in these ecotype seeds (Bandyopadhyay *et al.*, 2016).

Second, the development and growth of endosperm have also been showed to have a role in regulating seed size. The growth of endosperm is influenced by "parent-of-origin" effects (Scott *et al.*, 1998) and parent ploidy. Indeed, crosses between diploid and tetraploid plants produced plants with seeds that developed abnormally with change in size. The imbalance of the parental genome in seed results either in an enlarged endosperm or a reduced endosperm. With paternal genome in excess, growth of the endosperm is promoted, therefore it tends to produce larger embryo with increased final seed size. On the opposite, an excess of the maternal genome results in inhibition of endosperm growth, which leads to reduction of final seed size (Bradford and Nonogaki, 2007). Moreover, the timing of endosperm cellularization phase affects seed development and final seed size. Early endosperm cellularization results in small seeds, while delayed endosperm cellularization leads to bigger seeds (Garcia *et al.* 2003; Berger *et al.* 2006). *DNA METHYLTRANSFERASE 1* (*MET1*) is a DNA methyltransferase that maintains DNA methylation at CG content in *A. thaliana* (Jean finnegan and Dennis, 1993). Reciprocal crosses between the antisense *MET1* transgenic and wild-type lines showed that DNA hypomethylation has a parent-of-origin effect on seed size (Adams *et al.*, 2000). Reciprocal crosses between *met1-6* mutant and wild type demonstrated

that the hypomethylation in maternal and paternal genomes greatly affected the F1 seed size. F1 generation from homozygous *met1-6* mutation on the maternal genome produced larger seeds, while F1 generation from homozygous *met1-6* mutation on the paternal genome of homozygous *met1-6* mutation produced smaller seeds (Xiao *et al.*, 2006). In the past decades, several genes involved in different signaling pathways have also been identified as controlling seed size by regulating endosperm development. The IKU pathway was reported in several studies and includes *HAIKU1* (*IKU1*), *IKU2*, and *MINISEED3* (*MINI3*) genes (Garcia *et al.*, 2003; Luo *et al.*, 2005; Wang *et al.*, 2010). Mutants of these genes produced reduced seed size, due to precocious cellularization during endosperm development. The *SHORT HYPOCOTYL UNDER BLUE 1* (*SHB1*) was showed to act upstream to *IKU2* and *MINI3* by direct binding to their promoters to regulate endosperm growth (Zhou *et al.*, 2009; Kang *et al.*, 2013). The role of endosperm growth in regulating seed size pathway is also influenced by phytohormones. *Cytokinin oxidase 2* (*CKX2*) in cytokinin signaling pathway is the target of *MINI3* which can bind to *CKX2* promoter and participate to seed size regulation (Li *et al.*, 2013). In *Medicago*, *DASH (DOF Acting in Seed embryogenesis and Hormone accumulation)*, endosperm specific, gene involved in auxin homeostasis was shown to be involved in final seed size detmernination (Noguero *et al.* 2015)

Finally, during seed development, it has been showed that the integuments that constitute the seed coat are responsible for establishing the volume of seeds. Seed coat is exclusively a maternal tissue and the cell division and elongation rates controlled by maternal factors have been shown to control final seed size (reviewed in Li and Li 2015; Li, Xu, and Li 2019). In *A. thaliana*, *Transparent Testa Glabra 2* (*TTG2*) gene encodes a WRKY transcription factor that controls seed coat pigmentation. The *ttg2* mutants also caused reduced elongation of integument cells, which is coordinated with reduction of endosperm growth, leading to smaller seeds (Johnson *et al.*, 2002). Reciprocally, we observed that in *iku2* mutants, that displayed smaller seed phenotype by controlling endosperm growth, a modulation of the integument cell elongation (Garcia *et al.*, 2003). Moreover, to highlight this interplay between seed coat and endosperm, the *ttg2 iku2* double mutant produced smaller seeds compared to the single mutants seeds of *ttg2* and *iku2* (Garcia *et al.*, 2005). The role of maternal tissues to control seed size was also shown to be associated with different pathways such as ubiquitin pathway, transcriptional regulatory factors, G-protein signaling, phytohormones and other plant growth substances, as summarized in Figure 1.3 (reviewed in Li and Li, 2015; Li *et al.*, 2019).

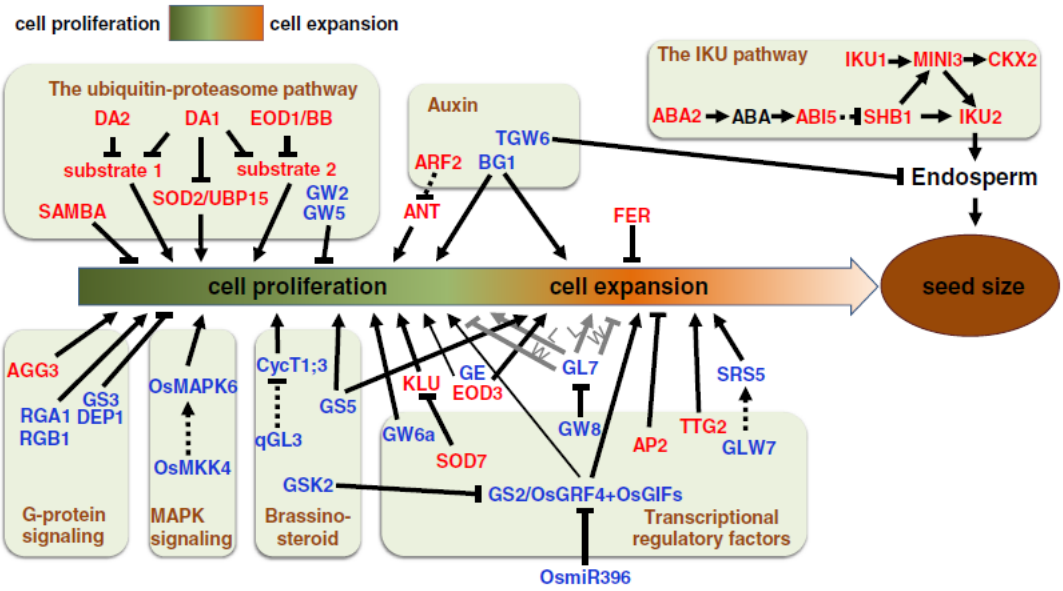


Figure 1.3 The major signaling pathways involved in regulation of seed size in *Arabidopsis* and rice (Li and Li, 2016). Thin lines which are compared with thick lines represent weak effects. Dashed lines indicate that genetic relationships are not clear. The regulators to controlling seed size in *Arabidopsis* and rice are shown in red and blue, respectively.

2.3 Seed desiccation tolerance

Desiccation tolerance (DT) is the capacity of an organism or tissue to endure extreme desiccation and yet survive and thrive. Desiccation tolerant tissues can undergo a dehydration corresponding to less than 0.1 gram of H_2O per gram of dry weight, or less than 10% of their dry weight, without lethal damages (Ooms *et al.*, 1993; Alpert, 2005). DT is an ancient trait that spread widely in ferns, mosses, and seeds of higher plants (review in Oliver *et al.* 2020). During seed development, the water content of seed is progressively declining along with the deposition of storage reserves. At the maturity stage of *Medicago* seeds, water content reaches 10% of seed dry weight and seed enters into a glassy state at maturity to sustain this extreme dehydration (Verdier *et al.*, 2013). Orthodox seeds (*i.e.* desiccation-tolerant seeds) can survive and be stored in this extreme condition due to activation of specific protective processes occurring during seed development. In *Medicago*, acquisition of desiccation tolerance in seed was determined during development by collecting immature seeds at different developmental stages from 16 to 48 DAP to check their percentages of viability (*i.e.* germination) after a drying period. We observed that 20 DAP immature seeds were not able to germinate after drying. In contrast, 24 DAP immature seeds germinated around 40% after drying, and from 28 DAP seeds germinated at 100% (Verdier *et al.* 2013). This result showed that acquisition of desiccation

tolerance of *Medicago* seed was acquired rapidly within 5-10 days, early during seed maturation between 20 and 28 DAP (Figure 1.1).

Survival in dry state implies a series of cellular and molecular processes to prevent seed tissue damage (Buitink and Leprince, 2004). During their development, seeds accumulate non-reducing sugars and oligosaccharides and Late Embryogenesis Abundant (LEA) proteins that contribute to protect cellular structures and compounds against desiccation. With the loss of water, the sugars and oligosaccharides accumulate and replace water in the cytoplasm resulting in viscosity increase (Leprince *et al.*, 2017). Moreover, LEA proteins accumulate to high content during seed maturation (Cuming, 1999). LEA proteins were first identified in cotton embryos (Galau *et al.*, 1987). They are hydrophilic, unstructured and heat-stable proteins (Bies-Ethève *et al.*, 2008; Hundertmark and Hincha, 2008). In *Medicago truncatula*, some candidate LEA genes including *MtEm6* and *MtPM25* were associated with acquisition of desiccation tolerance in comparative proteomic analysis (Boudet *et al.*, 2006). LEA proteins are generally related to desiccation tolerance, but they form a large family of multidomain and multifunctional proteins and for many of them their protective roles remain to be elucidated (Manfre *et al.*, 2006; Olvera-Carrillo *et al.*, 2010; Banerjee and Roychoudhury, 2016). Furthermore, the phytohormone abscisic acid (ABA), shown to be a key regulator of the global seed development and maturation, plays a critical role in the acquisition of desiccation tolerance via, for instance, a central regulator gene *ABI3*. In *Medicago*, mature seeds of loss-of-function *abi3* mutants showed intolerance to desiccation and down-regulation/decreased of RFO and LEA transcripts/proteins (Delahaie *et al.*, 2013).

2.4 Seed longevity

Desiccation tolerance enables seed to survive at dry state, but seed longevity or storability is another seed quality trait defined by the capacity of a seed lot to survive in the dry state during an extended storage period, which is highly dependent of seed storage conditions. Indeed, humidity and temperature are crucial factors to influence longevity (*i.e.* high temperature and relative humidity accelerate seed cellular mobility and decrease longevity). To evaluate seed longevity, an artificial treatment is used to accelerate seed ageing in controlled temperature and humidity conditions. This treatment consists to store seeds under moderate humidity and temperature conditions, without causing seed deterioration. Accelerated ageing treatments may depend on the seeds and the plant species but regarding *Arabidopsis* and

Medicago seeds, the most common storage conditions are 35°C with 75% relative humidity as standard treatment, and 35°C under 60% relative humidity as mild ageing treatment (Zinsmeister *et al.*, 2020). Seed lot is stored in artificial ageing conditions during different storage time intervals and seed viability of aged seeds is determined using percentage of germination. When viability of aged seeds was evaluated at sufficient storage time intervals, we assessed the seed longevity curve, which is characterized by the loss of seed viability along different ageing time (Figure 1.4). This longevity curve is then used to characterize a representative value of longevity for a seed lot, called P50, which corresponds to the storage (artificial ageing) time (in days) to lose 50% germination. The P50 is used as a reference value to compare the longevity performances between different seed lots. Seed longevity is also a seed trait that is acquired during seed late maturation in *Medicago*, and is greatly affected by environmental conditions in which seeds were produced (Righetti *et al.*, 2015). By using longevity curves determining P50 values from immature seeds collected from 24 to 48 DAP, it was determined that longevity is acquired progressively from 28 DAP and reaches a maximum around 44 days after pollination in *Medicago* seeds (Verdier *et al.*, 2013).

As discussed earlier, desiccation tolerance is a prerequisite for seed longevity but seeds acquire germination capacity and desiccation tolerance before longevity, which implies that these two processed are controlled by distinct mechanisms. In legumes, we observed accumulation of soluble non-reducing sugars such raffinose family oligosaccharides (RFO) concomitant with the acquisition of seed longevity. The RFO helps to form the glassy state in dry seed but their precise role in desiccation tolerance and longevity are still unclear (for review Leprince *et al.* 2017). Indeed, seeds without accumulation of oligosaccharides can still acquire desiccation tolerance (Black *et al.*, 1999) and, to date, there is no direct evidence showing the relationship between RFO and longevity. Another example of common points between DT and longevity are the LEA proteins. In *M. truncatula*, 35 polypeptides were identified in seeds, which were encoded by 16 *LEA* genes. The expression profiles of LEA polypeptides during seed maturation phase showed two different patterns, which suggested distinct regulatory mechanisms (Chatelain *et al.*, 2012). Another study showed accumulation of some of LEA proteins together with acquisition of desiccation tolerance, while most LEA proteins accumulate with the acquisition of longevity (for review Leprince *et al.*, 2017). Moreover, many LEA proteins are regulated by ABA and specific regulators of seed maturation. Indeed, *Arabidopsis* mutants such as *abi3 (ABSCISIC ACID INSENSITIVE 3), abi5 (ABSCISIC ACID INSENSITIVE 5), fus3 (FUSCA3)* and *lec2* (*LEAFY COTYLEDON 2*) displayed

downregulation of different sets of *LEA* genes specific of mutant lines (Bies-Ethève *et al.*, 2008). As expected, the interplay of ABA signaling pathway, via the *ABI3* pathway, was also showed in many studies as intermediate in the regulation of seed longevity, such via the *HEAT SHOCK TRANSCRIPTION FACTOR 9* (*HSFA9*), which is regulated upstream by *ABI3* in *Arabidopsis* and downstream able to control the expression of heat shock proteins genes (*HSPs)* (Kotak *et al.*, 2007). *HSFA9* is specifically expressed in seeds and involved in a genetic program to control seed longevity in tabacco (Prieto-Dapena *et al.*, 2006; Tejedor-Cano *et al.*, 2010). However, in *Medicago*, *MtHSFA9* did not appear to be implicated in *Medicago* seed longevity but rather in seed dormancy (Zinsmeister *et al.*, 2020). Finally, chlorophyll and degradation of chlorophyll via the *ABI3* pathway was showed to play a role in modulating seed longevity in *Medicago* (Zinsmeister *et al.*, 2016). To illustrate this relationship between acquisition of longevity and chlorophyll retention, many mutants of chlorophyll degradation enzymes, activated by ABI3, such as *NON-YELLOW COLORING1* (*NYC1*), *NON-YELLOW COLORING1-like* (*NOL*) and *STAY-GREEN2* (*SGR2*), showed reduced longevity (Nakajima *et al.*, 2012; Dekkers *et al.*, 2016; Zinsmeister *et al.*, 2016). Recently, other hormonal pathways with interplay of ABA such as auxin showed a relation to seed longevity in *Arabidopsis* but need to be elucidated (Pellizzaro *et al.*, 2020).

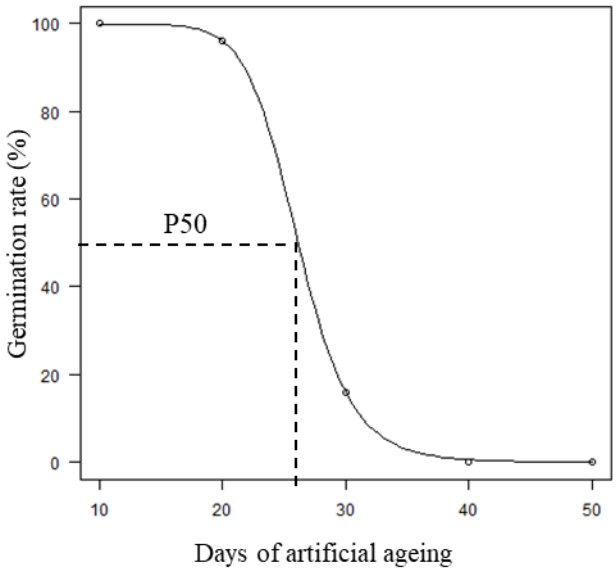


Figure 1.4 Measurement for seed longevity. Survival curve based on the germination rate of seeds during storage and measurement of P50 (*i.e.* the storage time (in days) to loss 50% germination) indicated by the dashed lines.

2.5 Seed germination and dormancy

Germination is an essential ability of seeds to perpetuate the future of species. The germination process starts at seed imbibition with the absorption of water until seed coat rupture and the emergence of radicle (Bewley *et al.*, 2013).

2.5.1 Seed germination assessment

The above description refers to whether a single seed germinated or not, thus for assessing the germination capacity of a population of seeds, the percentage of germinated seeds is measured (*i.e.* protruding radicles > 1mm) within an indicated time period, which corresponds to final germination percentage. Plotting the germination percentage of a seed lot at regular intervals allows to determine the kinetic of seed germination, which is crucial to calculate representative values of germination capacity of a seed lot such as the germination speed (T50, *i.e.* time for the seed lot to reach 50% germination) and germination homogeneity (T80-T20, *i.e.* time difference between 80% and 20% germination) (Figure 1.5).

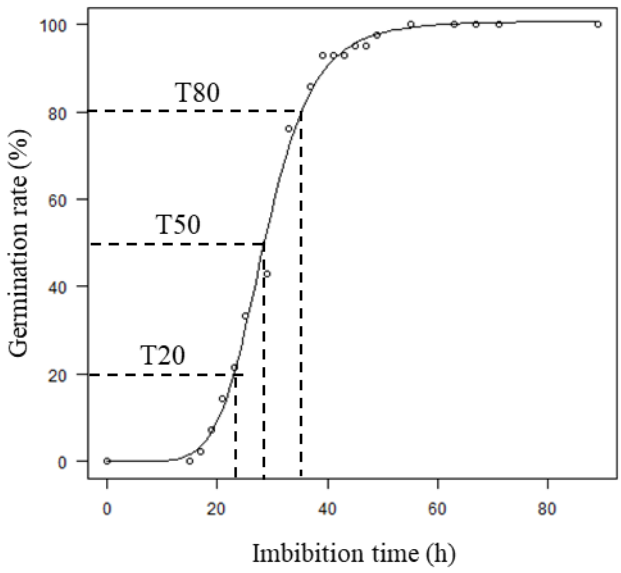


Figure 1.5 Measurement for seed germination. Germination curve and measurement of germination speed (T50) and homogeneity (T80-T20). T50, time to 50% germination; T20, time to 20% germination; T80, time to 80% germination.

These germination characteristics are important to define the seed vigor, which includes homogeneous and rapid germination to allow seedling establishment under a range of contrasted environmental (*i.e.* stress) conditions (Finch-Savage and Bassel, 2016). As shown

in Figure 1.6, *Medicago* seeds acquire the ability to germinate in optimal conditions early during seed development, before the acquisition of DT, at 16 DAP seeds are able to germinate. However, in *M. truncatula* (Verdier *et al.*, 2013; Righetti *et al.*, 2015), like other legumes (Pereira Lima *et al.*, 2017), the different vigor traits (*i.e.* ability to germinate under adverse conditions) are acquired sequentially, from seed filling until the late phase of seed maturation (reviewed in Leprince *et al.*, 2017). As for other seed performance traits, in *Medicago* (Righetti *et al.*, 2015), like many other species (Finch-Savage and Bassel, 2016; Penfield and MacGregor, 2017), seed vigor is also drastically affected by environmental conditions occurring during seed development. This highly plastic seed vigor response to environment is considered as a bet-hedging strategy to ensure dissemination of the species. In this respect, one of the most studied germination vigor traits is dormancy (for review Penfield and MacGregor, 2017).

2.5.2 Seed dormancy

Seed dormancy is defined as the temporary failure of viable seeds to complete germination under favorable conditions (Bewley *et al.*, 2013). Dormancy is a widespread characteristic of seeds, existing in many species to ensure survival in unfavorable conditions and guarantee offspring dissemination. Different types of dormancy were described in different species and Baskin and Baskin (2004) categorized dormancy into five classes, including physiological dormancy, morphological dormancy, physical dormancy and combinational dormancies such morphophysiological and physical-physiological dormancies. *M. truncatula* seed is a typical example of combinational dormancy, displaying both physical and physiological dormancies, as well as many legume seeds.

Physical dormancy is mainly controlled by the seed coat permeability (*i.e.* seed coat composition), which prevents seed imbibition. It can be released by natural seed (coat) scarification due to mechanical friction or acidic environments such animal digestive system. Physical dormancy is dependent of environmental conditions occurring during seed development. In *M. truncatula*, a slight change in the seed coat properties regulating seed imbibition and physical dormancy was observed when plants were grown in 35°C/15°C compared to 25°C/15°C conditions (Renzi *et al.*, 2020).

In contrast to physical dormancy, physiological dormancy is regulated by the embryo and endosperm molecular signals, via the ratio of abscisic acid (ABA) and gibberellic acid (GA) contents. ABA has a repressive effect on seed germination (Finkelstein *et al.*, 2002), but GA

stimulates germination (Olszewski *et al.*, 2002). Seeds acquire physiological dormancy during seed development by modulating the ABA/GA ratio and freshly harvested seeds display a dormant state that is called primary dormancy (Bewley and Black, 1994). During *Medicago* seed development, seeds start to acquire physiological dormancy around 22 DAP and reach the deepest dormancy level around 32 DAP, followed by a slight decrease in dormancy level until seed maturity (Figure 1.6) (Verdier *et al.*, 2013). Seed primary dormancy can be released by after-ripening treatment, in which seeds are stored in dry conditions at room temperature for an extended period of time, which differs between species but stands around six months for *Arabidopsis* or *Medicago* or by stratification treatment such as exposition to cold temperature for few days. As mentioned earlier, acquisition and depth of physiological dormancy are strongly influenced by environmental conditions during seed development, which modulates the ABA/GA ratio and the accumulation of ABA in mature seeds determing the depth of dormancy (Finch-Savage and Leubner-Metzger, 2006).

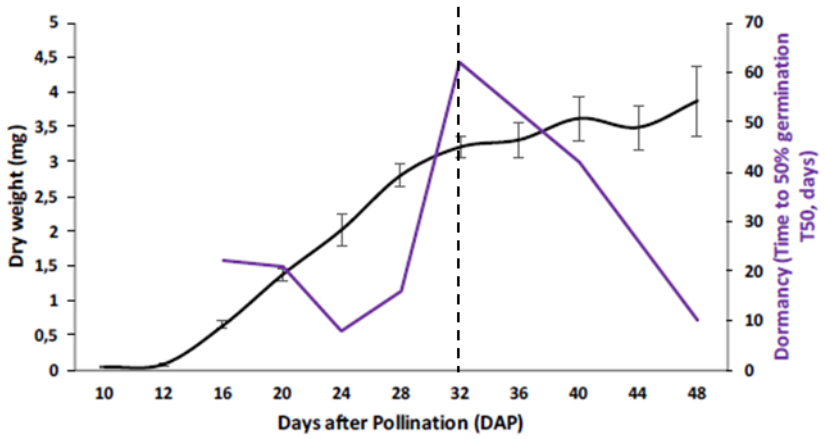


Figure 1.6 Acquisition of germination capacity and physiological dormancy during seed maturation in *Medicago truncatula* (Figure from Verdier, Leprince, and Buitink 2019). Acquisition of physiological dormancy is marked by the increase of T50 (*i.e.* time of the immature seed lot to reach 50% germination) of seed collected at different stages during seed maturation. The dashed line indicates the highest level of dormancy at 32 DAP.

2.5.3 Regulation of seed physiological dormancy and germination

The balance of ABA and GA plays important roles in regulation of seed dormancy and germination. Induction of primary dormancy is associated with the increase of ABA content during seed development that prevents precocious germination (Bradford and Nonogaki, 2007). In most cases, ABA accumulates during early seed development due to ABA synthesis in both

embryo and endosperm. Then, ABA content decreases during late maturation stage but maintaining a certain level in mature seeds to ensure dormancy. ABA biosynthesis is controlled by *zeaxanthin epoxidase* (*ZEP*) and *9-cis-epoxycarotenoid dioxygenase* (*NCED*) family genes. ABA1/ZEP is the first enzyme identified in ABA biosynthesis pathway, which is expressed ubiquitously in *Arabidopsis* seeds (Marin *et al.*, 1996; Audran *et al.*, 2001). Overexpression of *ZEP* gene induces higher ABA content in mature seeds and delays seed germination (Frey *et al.*, 1999). *NCED* family is composed of several related genes in different plant species that are responsible for ABA biosynthesis (Tan *et al.*, 2003). In *Arabidopsis*, nine *NCED* genes show tissue-specific expression pattern. *AtNCED6* and *AtNCED9,* which express specifically in seed, are necessary for ABA synthesis in dormancy induction. The double mutant, *Atnced6/nced9*, exhibited reduced ABA level and seed dormancy (Lefebvre *et al.*, 2006). Additionally, the short-chain dehydrogenase/reductase (ABA2/SDR) and abscisic-aldehyde oxidase (AAO) enzymes act downstream of the ABA biosynthesis pathway (Seo and Koshiba, 2002). The cytochromes P450 *CYP707* genes were identified as key regulators encoding ABA 8'-hydroxylases in ABA catabolism. Four of *CYP707* genes have been showed to play distinct roles in controlling ABA level in *Arabidopsis* seeds (Okamoto *et al.*, 2006). The dormancy depth caused by residual ABA content in mature seeds is controlled by the interplay of enzymes involved in both ABA biosynthesis and catabolism (Finkelstein *et al.*, 2008).

GA is a plant hormone involved in various processes and, in seeds, it behaves antagonistically to ABA to release dormancy and promote germination (Olszewski *et al.*, 2002). GA promotes seed germination by inducing hydrolytic enzymes to weaken embryonic surrounding tissues and stimulate the mobilization of reserves (Bewley and Black, 1994). The application of paclobutrazol in seeds, an inhibitor of GA biosynthesis, resulted in reduced seed germination (Norman *et al.*, 1986). GA accumulation during embryo development could result in precocious germination in absence of normal accumulation of ABA causing viviparous phenotype. Major enzymes controlling GA content are biosynthesis enzymes such as GA3ox (GIBBERELLIN 3 OXIDASE) and GA20ox (GIBBERELLIN 20 OXIDASE), in contrast to catabolic enzyme such as GA2ox (GIBBERELLIN 2 OXIDASE) (Seo *et al.*, 2006).

During stratification and after-ripening release of dormancy, ABA level declines while GA level increases (Ali-Rachedi *et al.*, 2004; Yamauchi *et al.*, 2004). Nevertheless, seed dormancy is not only affected by the environmental conditions occuring during seed development but also influenced by storage conditions and imbibition environment via regulation of ABA and GA content in response to environmental conditions. The balance of

ABA/GA ratio during imbibition is controlled by the dynamic biosynthesis and catabolism of ABA and GA (Figure 1.7) (Finch-Savage and Leubner-Metzger, 2006).

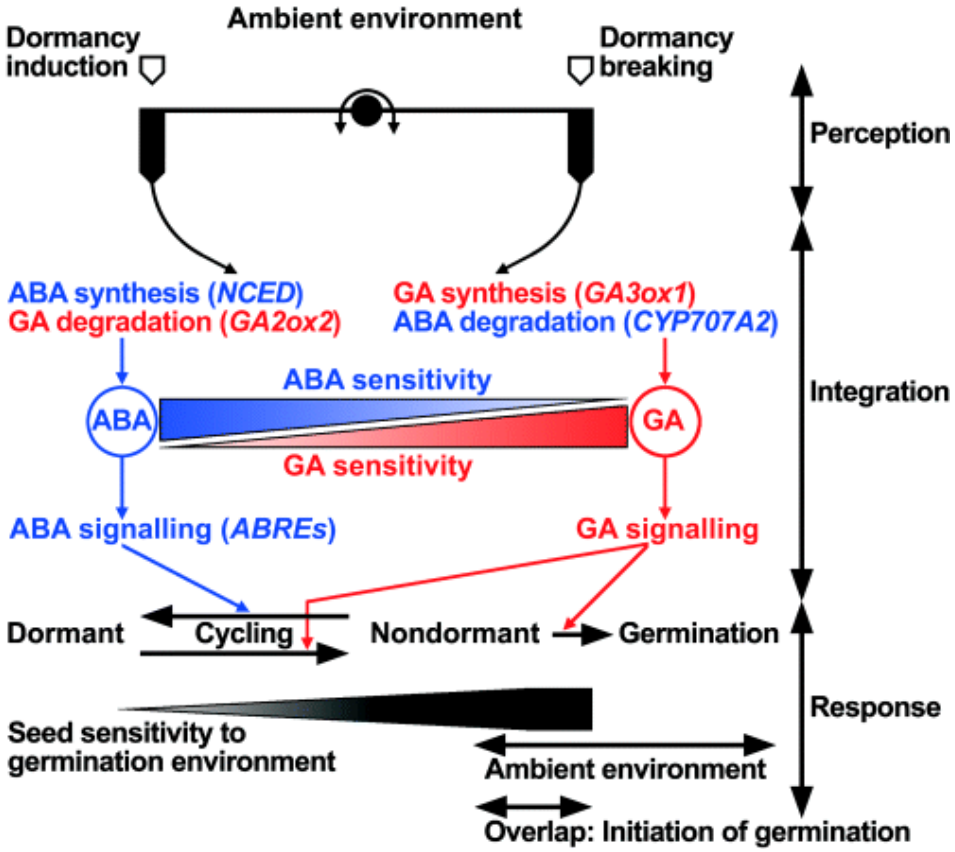


Figure 1.7 Model related to regulation of seed dormancy and germination at imbibition stage by moderating synthesis and catabolism of abscisic acid (ABA) and gibberellins (GA) combining with the environmental conditions (Finch-Savage and Leubner-Metzger, 2006). Ambient environmental factors such as temperature and light influence the balance the ABA/GA. In response to the environment, changes in hormone synthesis and degradation cause dormancy cycling or transition to germination.

In addition, many other hormones interplayed with ABA and GA to control seed dormancy and germination such as auxin, ethylene, brassinosteroid (BR) and cytokinin. The BR biosynthetic and responsive mutants such as *det2* and *bri1* were more sensitive to ABA compared to wild type *Arabidopsis*, which suggested that BR signaling is required for overcoming the ABA inhibition of seed germination (Steber and McCourt, 2001). Ethylene also plays a positive role in seed germination. In *Arabidopsis*, *ethylene resistant 1* (*etr1*) mutant and *ethylene insensitive 2* (*ein2*) mutant exhibited increased dormancy and more sensitivity to ABA during seed germination (Beaudoin *et al.*, 2000). Another example is the role of *miR160*,

which inhibits auxin-related gene expression during germination resulting in modulating ABA sensitivity during germination (Liu *et al.*, 2007). Finally, cytokinin was demonstrated to play a role in the transition between dry seed and seedling in concert with ABA via *ABI5* gene regulation (Wang *et al.*, 2011).

Some signaling molecules such as reactive oxygen species (ROS) and nitrogen-containing compounds also participate in the regulation of seed dormancy and germination, via the modulation of the ABA/GA ratio. Nitrogen-containing compounds including nitrite (NO_2^-), nitrate (NO_3^-), and nitric oxide (NO) can break seed dormancy in *Arabidopsis* (Bethke *et al.*, 2006), barley (Bethke *et al.*, 2004) and many other plant species (for review Bradford and Nonogaki 2007). Indeed, it has been shown that the enhancement of ABA catabolism requires the presence of nitric oxide (NO) to release dormancy (Liu *et al.*, 2010). Moreover, ROS molecules such as H_2O_2 enhance ABA catabolism and GA synthesis during seed imbibition by promoting the expression of *CYP707A* genes and GA biosynthesis genes, respectively. Finally, some key genes involved in the ABA or GA pathway play important role in dormancy and germination such as *DELAY OF GERMINATION 1* (*DOG1*), a seed-specific gene that controls seed dormancy (Bentsink *et al.*, 2006). In *dog1* mutant ABA content is reduced while GA is increased. The abundance of DOG1 in freshly harvested seeds is correlated with depth of dormancy (Nakabayashi *et al.*, 2012). *DOG1* not only acts in seed dormancy but also in seed maturation in concert with ABA signaling components *ABI3* and *ABI5* (Dekkers *et al.*, 2016).

After dormancy release, upon the imbibition of dry and viable seeds, a series of physiological events occur to complete germination including DNA repair, translation and degradation of stored mRNA, reserve mobilization and cell division (for review Bradford and Nonogaki 2007). There is an abundance of mRNAs that were stored in dry seeds during late seed development that now serves during early seed germination. The radicle appearance through endosperm and seed coat is considered as the completion of germination, and for definition, seed germination phase does not include the phase of seedling growth after the emergence of radicle. During this phase seeds continue to absorb water as well as molecular process occur for seedling establishment (for review Bewley *et al.* 2013).

3. Heat stress response in plant vegetative tissues and seed

Within a changing environment, plants need to response and adapt to different environmental conditions. The increasing average temperatures reinforce the need for understanding how plants response to heat stress to improve heat tolerance in crops. In the past years, many studies reported how plants sense, response and adapt to heat stress (reviewd in J. Liu *et al.* 2015; Ohama *et al.* 2017). Generally, under heat stress, plants have defect in photosynthetic activity, and the reduced water content caused by heat negatively impacts cell division and growth (Hasanuzzaman *et al.*, 2013). The most noticeable phenotypes following a reasonable heat stress intensity are a loss of vegetative biomass with bud abortion and a decrease of seed yield due to pollination problem (*i.e.* often linked to defect in germination of pollen grain) (Guilioni *et al.*, 1997; Qu *et al.*, 2013). During the period of stress, plants initiate heat stress molecular mechanisms to enhance plant thermotolerance, which aims to prevent water loss, protect cellular proteins and maintain cellular structure to allow plant survival. The two main protective mechanisms enhanced following a heat stress are the increase of heat shock proteins (HSPs) and reactive oxygen species (ROS)-scavenging enzymes. HSPs, including small HSPs, are present during normal growth and development in plants (Vierling, 1991; Waters *et al.*, 1996). However, they are over-produced during heat stress due to their role as molecular chaperones to protect or repair proteins damaged by heat stress. Even if these HSPs are essential in thermotolerance mechanisms, such as sHSP21 required for chloroplast development (Zhong *et al.*, 2013), their specific functions and targets are still largely unknown. For instance, some functional validation studies showed that overexpression of *AtHSP17.6A* can promote salt and drought tolerance of *Arabidopsis* plants (Sun *et al.*, 2001), while overexpression of *sHSP17.7* led to rice plants more tolerant to heat treatment and UV-B (Murakami *et al.*, 2004). On the other hand, ROS-scavenging enzymes, such as ascorbate peroxidase (APX) and catalase (CAT), also play a crucial role in cellular detoxification of ROS. Indeed, ROS molecules such as H_2O_2, O_2^-, and $^1O_2$ are over-produced during (heat) stress and serve as signal molecules to enhance heat stress response but need to be recycled to prevent cell death (Suzuki and Mittler, 2006; Baxter *et al.*, 2014).

3.1 Plant heat stress response network

Recent studies in *Arabidopsis thaliana* unravelled some parts of the transcriptional networks induced by heat stress (Figure 1.8) (for review Ohama *et al.* 2017). In this complex

network, many HEAT SHOCK TRANSCRIPTION FACTORS (HSFs) act as the core regulators in heat stress response to activate downstream genes. In *Arabidopsis*, 21 HSFs were identified and classified into three classes (A, B and C) and 14 groups (Nover *et al.*, 2001).

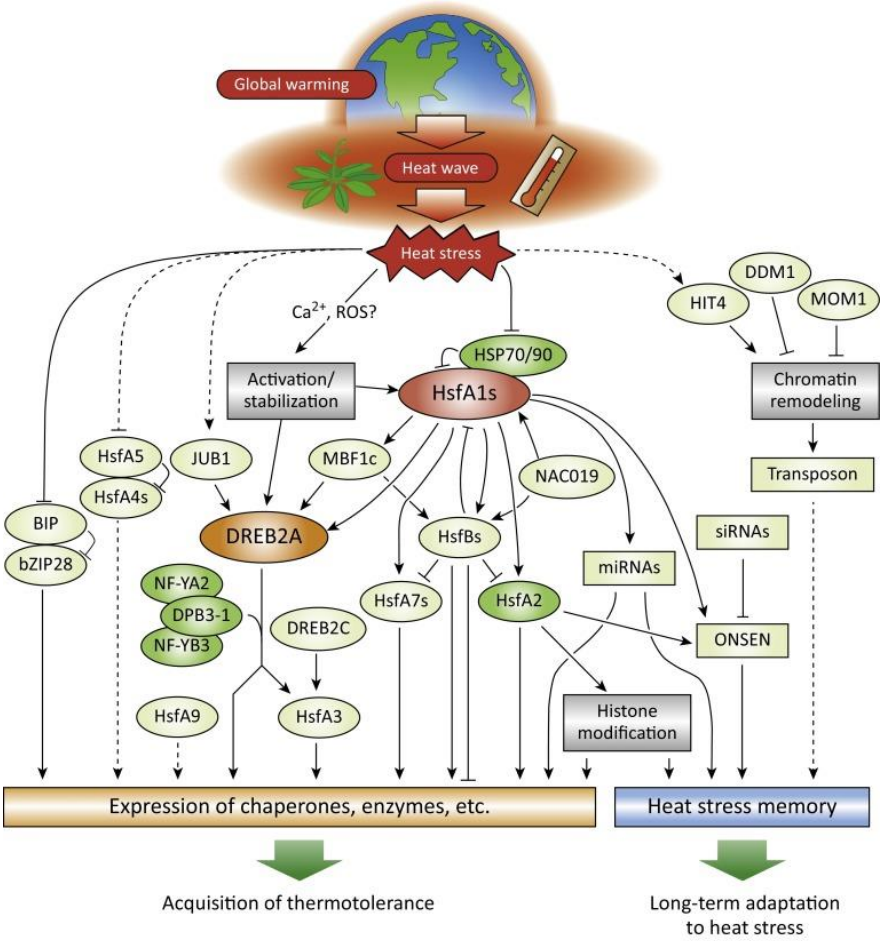


Figure 1.8 The regulatory network during heat stress response (Ohama *et al.*, 2017). The dashed lines indicate to be further verified.

HSFA1 gene group has been shown as the master regulators in heat stress response. Indeed, *hsfa1* mutants displayed high sensitivity to heat stress (Mishra *et al.*, 2002) and using *hsfa1a* and *hsfa1b* mutants in *Arabidopsis*, they showed that these two genes modulate the timing of heat stress response by controlling the expression of relevant genes (Lohmann *et al.*, 2004; Busch *et al.*, 2005). Finally, the quadruple mutant *hsfA1a/b/d/e* showed reduced thermotolerance of seedlings compared to triple mutants *hsfA1a/b/d as weel as* reduced seed germination rate. These results suggested that *HSFA1* genes work partially redundantly as thermotolerance master regulators (Liu *et al.*, 2011; Yoshida *et al.*, 2011). *HSFA2* is a heat-induced transcription factor, a direct target of *HSFA1* and functions as a key regulator to

acquire/extend thermotolerance (Charng *et al.*, 2007), but also involved in the regulation of different environmental stress responses (Nishizawa *et al.*, 2006). The heat-induced expression of *HSFA3* is directly regulated by *DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2A* (*DREB2A*), which directly binds to the *HSFA3* promoter (Schramm *et al.*, 2008), and is involved in the response to drought, salt and heat stresses (Rizhsky *et al.*, 2004; Sakuma *et al.*, 2006). Indeed, DREB2A, an APETALA2/ethylene-responsive element-binding factor-type (AP2/ERF-type) transcription factor, is a well-documented protein able to bind to the drought-responsive elements (DRE) to regulate the expression of drought related genes but also heat stress-responsive genes (Liu *et al.*, 1998). *HSFA4*, which acts antagonistically to *HSFA5* (Baniwal *et al.*, 2007), has been suggested to play a role in the ROS signaling pathway to prevent plant cell damage from various stress (Yamanouchi *et al.* 2002; Davletova *et al.* 2005).

Compared to HSF class A, HSFBs and HSFCs are relatively less understood. In tomato *HSFB1* works together with *HSFA1a* and *HSFA2* to active the expression of heat responsive genes (Bharti *et al.*, 2004). However, in *Arabidopsis* HSFBs such as *HSFB1* and *HSFB2b* were showed to inhibit expression of heat responsive genes (Kumar *et al.*, 2009). Future studies on these HSF classes are needed to shed light on their importance in the heat stress response and thermotolerance.

3.2 Seed-specific heat stress transcription factor

HSFA9 is a specific transcription factor among HSF family members that is only expressed during seed development (Kotak *et al.*, 2007). In sunflower, *HaHSFA9* is embryo-specific and acts as a transcriptional activator of *Hsp17.7CI* and *DREB2* (Concepción Almoguera *et al.* 2002; Díaz-Martín *et al.* 2005). Overexpression of *HaHSFA9* in tobacco seeds induced abundant accumulation of sHSP and HSP101, which increased seed longevity and thermotolerance of seedlings (Prieto-Dapena *et al.*, 2006, 2008). Co-overexpression of *HaDREB2* and *HaHSFA9* in tobacco seeds produced plants with even higher accumulation of sHSP and increased thermotolerance and longevity compared to single overexpression of *HSFA9* (Almoguera *et al.*, 2009). In addition, loss of function of *HSFA9* resulted in reduction of HSP accumulation and reduction of seed longevity (Tejedor-Cano *et al.*, 2010). Moreover, *ABI3* was showed to control the expression of *HSFA9,* which will consequently induce the expression of *Hsp17.4-CI* and *Hsp17.7-CII* during the seed maturation in *Arabidopsis* (Kotak *et al.*, 2007). In *M. truncatula,* a regulatory network analysis revealed a link between *MtHSFA9*

and *MtABI3* with seed desiccation tolerance and longevity (Verdier *et al.*, 2013). However, a recent study of *HSFA9* in *M. truncatula* did not observed any phenotype of *hsfa9* mutants in seed longevity, but instead a role of *MtHSFA9* in the negative regulation of seed dormancy via change in ABA homeostasis (Zinsmeister *et al.*, 2020a).

3.3 Impacts of heat stress on seed development and quality

The above transcriptional regulatory pathway of heat stress response was mainly studied in vegetative tissues such as roots and leaves. So far, very few studies focused on the impacts of stresses during the seed development and their impacts on seed qualities. In a previous study, it was shown that cold, heat and osmotic stresses had significant effects on seed developmental timing (Righetti *et al.*, 2015). Under these different environmental conditions, the duration of embryogenesis stage was relatively stable, whereas the duration of maturation stage was severely affected. Seeds that developed under heat stress condition were the most impacted with a severe reduction of the maturation phase duration, condensed from 30 days in control conditions to around 10 days during heat stress (Righetti *et al.*, 2015). Through the characterization of physiological processes, we observed that acquisition of desiccation tolerance was a robust process, with no major impact from different environmental conditions. On the opposite, acquisition of longevity was highly affected by different environmental conditions (Righetti *et al.*, 2015). Many studies in different species showed similar results. In rice, high temperature shortened the reproductive period (He *et al.*, 2014), the potential longevity of seeds produced in the warmer condition was lower than that of the cooler environments (Ellis *et al.*, 1993) and plants exposed to transient moderate heat stress (35°C) during early seed development displayed mature seeds with higher germination rate than control conditions due to the alteration of ABA and GA biosynthesis (Begcy *et al.*, 2018). In soybean, heat stress during seed development resulted in lower seed germination, decreased seed weight and several changes in metabolites (Chebrolu *et al.*, 2016). A study in alpine snowbed plants showed that seeds from plants exposed to warm temperature can live longer than seeds produced in natural condition (Bernareggi *et al.*, 2015). All these studies demonstrate that heat stress during seed development affects the duration of seed maturation and impact important seed traits such as seed longevity and vigor. However, the characterization of the molecular mechanisms involved in heat stress response in seeds is still poorly understood, especially for legumes species.

4. Epigenetic regulation involved in legumes and general seeds

In the past decades, epigenetic regulation is being well studied in different speciesrelated to varous biological processes. DNA methylation and histone modification are the major epigenetic modifications to regulate gene expression. In plants, epigenetic regulation not only occurs in various plant development phases but also during the biotic and abiotic stress responses (Chinnusamy and Zhu, 2009; Ahmad *et al.*, 2010). During my PhD, we have been invited to review distinct epigenetic mechanisms involved in developmental processes and (a)biotic stress in legumes (Windels *et al.*, 2020), my contribution to this review (presented below) was in the writing of the introduction part.

4.1 Epigenetic regulation in legumes (published review)

(This section was published in journal of Legume Science.)

Received: 1 March 2020 | Revised: 2 July 2020 | Accepted: 3 August 2020

DOI: 10.1002/leg3.60

REVIEW

Legume Science Open Access WILEY

Snapshot of epigenetic regulation in legumes

David Windels | Thu Thi Dang | Zhijuan Chen | Jerome Verdier

IRHS (Institut de Recherche en Horticulture et Semences), UMR 1345, INRAE, Agrocampus-Ouest, Université d'Angers, Angers, France

Correspondence
Jerome Verdier, IRHS (Institut de Rechercheen Horticulture et Semences), UMR 1345, INRAE, Agrocampus-Ouest, Université d'Angers, SFR 4207 QuaSaV, Angers 49071,Beaucouzé, France.
Email: jerome.verdier@inrae.fr

Abstract

In the current context of food security and increase in plant protein demand, legumes have an important role in facing challenges of climate change. With the recent improvement of sequencing technologies and the emergence of new knowledge related to plant epigenetic regulation in response to developmental and environmental changes, legume epigenetics is an emerging field with high potential for improving legume crop productivity and adaptability. The objective of this review is to providea snapshot of epigenetic studies in different legume species. We have summarized the state-of-the-art regarding legume epigenetic regulation controlling or participating in developmental aspects such as nodule, flower, and seed development and related to biotic and abiotic stresses. This extensive view of the different studies onlegume epigenetics provides a baseline for identifying common and distinct mechanisms, and key players in epigenetic regulation from those of model species, such Arabidopsis, and highlights the impact that a better understanding of these mechanisms in legumes could have in order to improve plant productivity and adaptability.

KEY WORDS

DNA methylation, epigenetics, histone marks, legumes

1 | INTRODUCTION

Knowledge in the field of epigenetics has rapidly increased over the past decade. Many epigenetic mechanisms have been identified, although from a very limited number of model species in mammals, insects, or plants such as Arabidopsis. Recently, the rapid increase of genomic technologies has allowed the decryption of many genomes, rendering possible the transfer of this knowledge to non-model organisms.

Since the introduction of the term epigenetic by Conrad Waddington in the 1940s, epigenetic concepts have radically changed, and nowadays, its exact definition is still being debated within the scientific community (Deans & Maggert, 2015). A general and commonly accepted definition refers as epigenetics (that literally means "abovegenetics"), heritable changes that do not alter the genetic code but could lead to modification of gene expression and phenotypic changes. Indeed, epigenetic changes do not change the DNA sequence, but by modifying the chromatin structure, they will affecthow cells transcribe their genes. The two main epigenetic mechanisms are DNA methylation and post-translational histone modifications (PHM). These two mechanisms have been intensively studied in the past years, mainly in Arabidopsis, uncovering some aspects of their function and regulation as well as their influence on each other.

DNA methylation is a conserved epigenetic modification in plantand mammals. It directly impacts DNA by adding a methyl group from S-adenosyl-L-methionine to the fifth carbon position of a cytosine ring to generate a 5-methylcytosine (5mC). While restricted to ^{m}CG sequences in mammals, plant DNA methylation is found in three contexts: CG, CHH, and CHG, in which H can be any base except for a guanine.

In plants, these methylation marks are regulated by DNA methyltransferases such as METHYLTRANSFERASE1 (MET1) for the CG

context, CHROMOMETHYLASE 2, and 3 (CMT2 and CMT3) for the CHG context, DOMAIN REARRANGED METHYLASE2 (DRM2) and CMT2 for the CHH context. These enzymes have a role in DNA meth ylation maintenance, but some of them have specific role in de novo methylation establishment through the DRM2- and CMT2-dependent pathways (for review Law & Jacobsen, 2010). Indeed, the DRM2-dependent pathway or de novo RNA-directed DNA methylation (RdDM) pathway relies on small RNAs, especially small interfering RNAs (siRNAs) to serve as guides to methylate specific DNAsequences. The RdDM pathway involves two RNA polymerases specific to plants: POL IV, necessary for siRNA production (Herr, 2005;Onodera et al., 2005; Pontier et al., 2005) and POL V needed to guide ARGONAUTE 4 (AGO4) to the chromatin (Wierzbicki, Ream, Haag, &Pikaard, 2009) through a well-established process. These 24-nt siRNAs are produced through POL IV, and its associated transcrip- tional complex including the RNA-DEPENDENT RNA POLYMERASE 2 (RDRP2 or RDR2), which generates double stranded RNA (dsRNA), ultimately cleaved by DICER-LIKE PROTEIN 3 (DCL3) into siRNAs. These siRNAs are, then, loaded onto ARGONAUTE proteins (mainly AGO4 et AGO6) and paired with scaffold RNA produced by POL Vwith the help of the DDR protein complex. The DDR complex is composed of DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), RNA-DIRECTED DNA METHYLATION 1 (RDM1), and DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) proteins, which stabilize the POL V chromatin interaction with the help of the MORC protein complex. AGO4, with associated RNAs, interacts with the DNA methyltransferase DRM2 with possible assistance of RNA-DIRECTED DNA METHYLATION 3 (RDM3) to methylate targeted regions (for review Matzke & Mosher, 2014; Zhang & Zhu, 2011). Finally, after DNA methylation, another protein complex belonging to the INVOLVED IN DE NOVO2-IDN2 PARALOGUE (IDN-IDP) complex may stabilize the siRNA/scaffold RNA to interact with the SWI/SNF chromatin remodeling complex that will change nucleosome positioning to silence transcription (Finke, Kuhlmann, & Mette, 2012; Zhu, Rowley, Böhmdorfer, & Wierzbicki, 2013). It is still unclear to which extent the RdDM pathway is involved in direct gene methyla- tion and regulation, but it is crucial for targeting specific repetitive sequences and transposable elements (TEs), which may indirectly control nearby gene activation or repression (for review Sigman & Slotkin, 2016). Partially redundant with the RdDM pathway, the CMT2-dependent pathway is involved in de novo CHH methylation of heterochromatin regions and more specifically TEs (Zemachet al., 2013). This pathway, less described than the RdDm pathway, occurs through a siRNA independent manner and relies on DECREASED IN DNA METHYLATION 1 (DDM1) chromatin remodeler.

To counterbalance methylation activity, DNA demethylation occurs either passively due to failure in DNA methylation mainte- nance following replication or actively by regulation of methylation levels by DNA glycosylases. In Arabidopsis, four 5mC DNA glycosylases are able to excise methyl groups from all three DNA methylation contexts (i.e., DEMETER [DME], DEMETER-LIKE PROTEIN2 [DML2], DEMETER-LIKE PROTEIN3 [DML3], and REPRESSOR OF SILENCING1 [ROS1]; see review Zhang, Lang, & Zhu, 2018). Several studies showed a coordination between DNA methylation and active demethylation by an antagonistic effect of RdDM and ROS1 activity to prevent hypermethylation at specific loci (Tang, Lang, Zhang, & Zhu, 2016). A 39-nt specific regulatory element in the *ROS1* promoter, called a DNA monitoring methylation sequence (MEMS), has been identified to serve as a putative sensor of MET1 and RdDMpathway activities. Indeed, high MET1 and RdDM activities lead to hypermethylation of this sensor, which activates ROS1 demethylaseexpression to regulate the genome-wide DNA methylation (Leiet al., 2015).

DNA methylation is usually described as a repressive modification of heterochromatin and pericentromeric regions, associated with gene and transposon silencing. However, it seems to play different roles depending on methylation locations. Methylation of promoter (or intergenic) regions has been proposed to regulate gene expression by inhibiting the binding of transcriptional activators/repressors, therefore activating or repressing transcription. It was shown to completely repress gene expressions, for tissue-specific genes (Johnson et al., 2007; Zhang et al., 2006). Methylation in promoter regions has been shown to be involved in gene regulation of specific processes such as imprinting during seed development and regulationof some immune-responsive genes (for review Matzke &Mosher, 2014; Zhang, Su, Hu, & Li, 2018). Methylation promoters appeared to be the consequence of the spreading of methylation from closely located TEs. In Arabidopsis, only 5% of promoter regions are methylated, but it does not reflect the situation of plant species with larger genome, such as legume crops, which contain many transpo- sons and repeat elements with possible impacts on nearby gene expression through promoter methylation. Role of DNA methylation within gene bodies is still unclear. In contrast to methylation of pro- moters, gene body methylation occurs in 30% of Arabidopsis genesbut with relatively low methylation levels (Zhang et al., 2006). Some correlations revealed that body-methylated genes were enriched in GC context and that these genes were often associated with high and/or constitutive expression of such housekeeping genes (Zhang et al., 2006). A recent study revealed that gene body methylation levels were not associated with highly expressed genes but rather with long and slowly evolving genes (Kawakatsu et al., 2016). To date, two hypotheses regarding the role of methylation in gene bodies have been proposed: (a) it could mask cryptic transcription sites, and it could help splicing of isoforms (Neri et al., 2017), (b) it could reduce variation of gene expression by excluding H2A.Z from the nucleosome, whose binding to gene bodies is anticorrelated to methylation but correlates to gene responsiveness to the environment (Zilberman, Coleman-Derr, Ballinger, & Henikoff, 2008). The role of methylation in TEs is much clearer; it acts as a repressor of the transposition activity inducing TE silencing. TEs are heavily methylated in all contextsand methylation maintenance involves mainly MET1, CMT3, DRM2 and relies on the RdDM pathway (Zhang et al., 2006). TEs and repetitive elements represent a large proportion of most plant genomes, active TEs could insert within or around protein sequences disrupting normal genome function and threatening genome stability. To prevent

this phenomenon, hypermethylation of TEs will silence and immobilize transposons in order to prevent disruption of normal gene functions and enhance genome stability (Mlura et al., 2001; Suzuki & Bird, 2008; Sekhon & Chopra, 2009; for review Sigman & Slotkin, 2016). PHM are a conserved epigenetic mechanism controlling recruitment of chromatin remodeling proteins via modification of the nucleosome structure. Indeed, the nucleosome is an important structure controlling access and binding of regulatory factors (Berger, 2007). Eight histone proteins form the nucleosome with two copies of eachof H2A, H2B, H3, and H4 proteins, around which is wrapped 147BP of DNA (Peterson & Laniel, 2004). Amino acids of the N-terminal tails of histones H3 and H4 are easily modified by methylation, acetylation, phosphorylation, ubiquitination, ribosylation, or biotinylation. These modifications will affect inter-nucleosomal interactions and permit recruitment of chromatin remodeling enzymes, leading to chromatin structure change. Histone modifications can activate genes through acetylation, phosphorylation, and ubiquitination and mostly repressgenes through methylation, with some exceptions (Table 1). Repressive marks such as H3K27me3, H3K9me3, H4K20me have also been associated with heterochromatin-associated histone modification (Zhao, Zhan, & Jiang, 2019). Acetylation of lysines on H3 and H4 his- tones is controlled by multiple histone acetyltransferases (HATs) and histone deacetylases (HDACs). Methylation of lysines on H3 and H4 histones is controlled by histone methyltransferases (HMTs) and histone demethylases (HDMs). Regarding methylation, lysine residues can be mono-, di-, or tri-methylated, which confer different transcriptional roles with marks such as H3K4me2 and H3K36me3 acting in gene activation, whereas others such as H3K27me3 and H3K9me2 are repressive (see Table 1).

Although histones are highly conserved proteins, plants have developed structurally and functionally distinct classes of Histone 2A (i.e., H2A.X, H2A.Z) and H3 (i.e., H3.3) variants, which play important roles in the dynamics of association with DNA (see review Deal & Henikoff, 2011). H2A.Z, for instance, is a variant mainly found in gene bodies and around transcriptional start site of genes, acting with the SWR1 remodeling complex, and highly responsive to heat stress, which induces nucleosome dissociation from DNA, activating gene expression (Kumar & Wigge, 2010; Sura et al., 2017).

Finally, several studies have provided evidence of the interplay between DNA methylation and modification of histone marks to modify chromatin structure. As an example, it was recently shown how DNA demethylase ROS1 is recruited to target specific loci via two bromodomain-containing proteins, essential for recruiting the SWR1 remodeling complex through recognition of histone acetylation, which enhances active demethylation and deposition of H2A.Z histone variants (Nie et al., 2019).

Starting from the state of the art, mainly obtained in Arabidopsis,several recent articles have deciphered and compared epigenetic mechanisms in other plant species. In this review, we intend to provide a snapshot of epigenetic studies in legumes with a specific focuson epigenetic roles in developmental processes such as nodule, flower, pod and seed development, and responses to biotic and abiotic stresses.

TABLE 1 Summary of some major histone modifications with their preferential binding locations and their transcriptional roles

Histone marks	Transcriptional role	Position
H3K4me	Activation/repression	Entire transcribed regions
H3K4me2	Activation/repression	5′ end of gene body
H3K4me3	Activation	5′end of gene body
H3K9me2	Repression	entire transcribed regions
H3K9me3	Repression	5′ end of gene body
H3K27me3	Repression	5′ end of gene body
H3K36me3	Activation	5′ end of gene body
H4K20me	Repression	-
H3K9ac	Activation	5′ end of gene body
H3K27ac	Activation	5′ end of gene body
H3K36ac	Activation	5′ end of gene body
H4K5ac	Activation	-
H4K8ac	Activation	-
H4K12ac	Activation	-
H4K16ac	Activation	-

2 | DEVELOPMENTAL PROCESSES

Nodule development is mainly controlled by nodule-specific genes including cysteine-rich genes (NCRs), which are specific to legumes producing indeterminate nodules, such as Medicago. In this species, nodule zones represent the temporal developmental stages and are composed of the meristem (or apical meristem, ZI), the invasion zone (ZII), and the nitrogen-fixing zone (ZIII), which display specific ploidy levels ranging from 2C/4C (ZI), 4C/8C (ZII), and up to 32C/64C (ZIII; Vinardell et al., 2003). Moreover, these ploidy levels are correlated with expression of NCR genes in different nodule zones (Nagymihályet al., 2017). The proportion of these zones will define nodule maturity from immature nodule, when ZI and ZII are predominant, to mature nodule, when ZIII is well expanded. The first correlative evidence of the importance of DNA methylation for nodule development was a differential expression of methylases and demethylases between nodule zones, with higher expression of DNA methylase genes such *MET1*, *CMT2*, and *CMT3* in the nodule apex (ZI) and in contrast, demethylation genes, such as *DME*, which was more expressed in proximal part of invasion zone (ZIIp; Satgé et al., 2016). To validate the role of DNA methylation in nodule development, DME was silenced by RNA interference. This led to abnormal developmentof the nitrogen-fixing zone, which was unable to fix nitrogen, indicating that DME control of demethylation is required for forming a functional nodule. DNA capture was performed to detect regions with high gene expression in immature and mature nodules. Four hundred seventy-four of highly expressed regions showed a correlated variation of methylation levels in CG and CHG contexts, whereas the level of methylation in CHH context was stable along nodule development

(Satgé et al., 2016). This result was confirmed by methylation level analyses between 4C and 32C cells, where 79% of DMR and 74% of DMR-associated to genes were found in CG context (Nagymihályet al., 2017). Interestingly, both studies showed that CG-DMR-associated to genes were located in NCR genes, which were more expressed in mature nodule. In parallel, Nagymihály et al. (2017) showed that 11% of coding genes were differentially methylated between 32C and 4C (6% hypermethylated and 5% hypomethylated). These hypomethylated genes were overrepresented by NCR genes and nodule-specific genes. Methylation analyses of 375 NCR genes showed that 44% were hypomethylated and 4% hypermethylated, but unfortunately, no correlation between the level of methylation and the expression of NCR genes was observed in this study. Indeed, methylated NCR genes were expressed at the same level as hypo-methylated NCR genes, indicating that methylation might be involvedin activation but another mechanism such as histone modifications and/or chromatin accessibility could be implicated in gene expression. To investigate this hypothesis, DNA accessibility during nodule development was performed by ATAC-seq analyses. It revealed that high DNA accessibility was correlated to high expression of NCR genes butonly for the late stages of development and independently of methylation state, indicating that chromatin opening can occur without genedemethylation (Nagymihály et al., 2017). The same study analyzed thepresence of the repressive histone mark H3K27me3 and the active mark H3K9ac on NCR genes. They showed that H3K27me3 was massively present on the promoter and gene body of NCR genes, whichwere not expressed and correlated with high level of chromatin compaction. Inversely, highly expressed NCR genes associated withH3K9ac on gene body and coincided with opened chromatin.

The initiation of flower and pod development is highly regulatedand depends on environmental and developmental cues. In Arabidopsis, a strong epigenetic control has been highlighted for flowerdevelopment (Groszmann et al., 2011; Liu & Wendel, 2003; Saze, Scheid, & Paszkowski, 2003; Simpson, 2004; for review Whittaker &Dean, 2017). Several studies also confirmed the role of epigenetics in legumeflower development, highlighting some similarities and differences with Arabidopsis, despite the fact that in soybean, flower initiation isknown to be induced by short day conditions and does not need vernalization, unlike Arabidopsis. Liew, Singh, and Bhalla (2013) identified 124 histone modifiers and analyzed their expression profiles during soybean flowering initiation. Interestingly, the majority of these geneswere found to be highly expressed in the shoot apical meristem (SAM), suggesting an active role of histone modifications in regulatinggene expression in this tissue. Fourteen histone acetylases (HAT) fromthree families were identified with three from the CBP family, nine from the GNAT/MYST family and two from the $TAF_{II}250$ family. Most of them displayed higher expression in SAM with a peak of expression at 1 day after short days, which is coherent with induction of light-gene expression by two $TAF_{II}250$ HATs as shown in Arabidopsis (Benhamed, Bertrand, Servet, & Zhou, 2006). Twenty-four histone deacetylases (HDACs) were identified, distributed in three classes (i.e., *HD2*, *SIR2*, and *HDA* with, respectively, 6, 4, and 14 genes).

Among the four SIR2 members, two showed homologies with AtSRT2 and two others with AtSRT1. Interestingly, *SRT2* genes in soybean were more expressed in leaf whereas *SRT1* genes were more expressed in SAM, suggesting different and specific regulations. Regarding the HD2 class of histone deacetylases, in Arabidopsis, high levels of ABA repress *HD2* expression (Luo et al., 2012). In soybean, the two HD2 orthologs of AtHD2 showed the same behavior as in Arabidopsis being highly expressed in the SAM before short-day condition, followed by decreased expression at the onset of short-day treatment, which induces ABA production (Wong, Singh, & Bhalla, 2013). Finally, among the HDA class of histone deacetylases, 14 genes were identified in soybean. The most represented, with four members, was the HDA6 family, which is known to deacetylate at various lysine residues (Chen & Wu, 2010; Krogan, Hogan, & Long, 2012; Zhou, Zhang, Duan, Miki, & Wu, 2005). Interestingly, the functional analysis of the coding sequence of the HDAC family in soybean showed that the histone deacetylase domain is highly conserved. Surprisingly, one soybean HDA member, putative ortholog of AtHDA19, contained a zf-RVT domain in the HDAC domain, which is typically present in reverse transcriptase. The combination of HDAC and zf-RVT domains had not been described in other species and could indicate a specific function which remains to be discovered. Several histone (de)methylases were identified, among them 47 SET proteins (SDG for set domain group, histone methyltransferases), 15 protein arginine methyltransferases (PRMTs) and 24 JmjC demethylases insoybean. These genes were shown to be more expressed in SAM during flowering initiation, suggesting their potential role during this transition. Five classes of SDG genes are defined in Arabidopsis. Two members of class I, SWINGER (SWN) and CURLY LEAF (CLF), are two histone methyltransferases implicated in methylation of H3K27 and involved in the regulation of *FT* and *FLC* genes, therefore crucial in controlling flowering time in Arabidopsis (Jiang, Wang, Wang, & He, 2008). In soybean, two CLF and two SWN orthologs were identified but they did not display the same expression patterns. *CLFs* were more expressed in SAM and SWNs in leaf, suggesting a specific role of CLF in flowering. Interestingly, two paralogues of AtREF6, a H3k27me3 demethylase, were identified in soybean and displayed opposite expression profiles of CLF, indicating that GmREF6 couldplay an antagonistic role to CLF to control FT expression in SAM. The SDG class III genes composed of *ATX1*, *ATX2*, and *ATXR7 (*known to methylate H3K4 and H3K36 residues) were shown to prevent flowering before vernalization by activation of *FLC* expression in Arabidopsis (Pien et al., 2008; Tamada, Yun, Woo, & Amasino, 2009). In soybean, the orthologous genes were upregulated in the SAM, even after vernalization, which suggests a different role for these genes inthe SAM regarding flowering initiation. Interestingly, SDG class V genes, such as *AtSUVR* and *AtSUVH*, constitute the largest group of histone methyltransferases in Arabidopsis, and the SUVH subgroupwas also highly represented in soybean with 15 genes (compared to nine in Arabidopsis). However, no ortholog of the SUVR subgroupwas identified in soybean. Regarding PRMTs, 15 were identified in soybean. Three of these PRMTs, including PRMT10, were shown to regulate FLC in response to vernalization in Arabidopsis. In soybean,

an ortholog of PRMT10 was identified which displayed higher expression in the SAM after 2 days in short-day condition. Because soybean does not need vernalization, GmPRMT10 is probably involved in flowering initiation but could be activated through a different pathway. Regarding JmjC demethylases, 16 (out of a total of 24 in soybean) displayed a peak of expression 1 day after short-day treatment, including two orthologs of *AtEF6* (*EARLY FLOWERING 6*), which is implicated in repressing *FT* through demethylation of H3k4me3 in Arabidopsis (Jeong et al., 2009; Lu, Cui, Zhang, Liu, & Cao, 2010).

Regarding late flower development and pod initiation, Wang et al. (2018) showed differential expression of genes involved in DNA methylation during the three stages of flower development: S1 (ground green gynophores), S2 (white gynophores, 3 days), and S3 (gynophores enlarged, 9 days) in peanuts. Orthologous methylation genes such as DRM2 and MET1 DNA methylases and *DMS3* showed higher expression in S2 compared to S1, whereas *DRD1*, *MORC1*, and *IDN2* were more expressed in S3. Almost all genes implicated in the RdDM pathway were transcriptionally stable during all stages of pod devel- opment except for *NRPE5*, *HDA6*, *LDL1*, and *DMS3*. Then, authors analyzed the correlation between DNA methylation, expressionof transcripts and 24-nucleotide siRNAs or miRNAs. First, only half of transcript expressions was negatively correlated with the level of methylation. Second, a positive correlation was found between abun- dance of siRNAs and miRNAs and methylation level. These observations suggest a potential role of the RdDM pathway in DNA methylation regulation and regulation of peanut flower/pod development. It is to note that the role of the RdDM pathway in flower development is nonessential as many RdDM mutants displayed proper flower development in other species but could act in flowering time regulation via methylation of the *FWA* promoter region (Chan, Zhang, Bernatavichute, & Jacobsen, 2006).

Regarding seed development, An et al. (2017) analyzed the DNA methylation pattern in cotyledon during three stages of seed maturation in soybean, from early (S2) to middle (S6) and late seed maturation (S8). Global DNA methylation was mainly identified in CG (66%), then CHG (45%) and CHH (9%) contexts. However, the global CG andCHG levels were unchanged during seed maturation, whereas CHH level increased during seed development from 6% in S2 to 9% in S8. Lin et al. (2017) confirmed these previous observations by a more comprehensive analysis of DNA methylation during soybean seed development and within dissected seed tissues from post fertilization to germination. Indeed, global methylation levels in CG and CHG contexts were slightly decreased but changed little during the studied stages, whereas CHH methylation level greatly increased during seed maturation (between early-maturation stage and late-maturation), then dropped drastically at germination. Although the average global CHH methylation level across all samples was globally low (2%) compared to CG (57%) and CHG (37%), it increased more than three-fold during seed maturation then dropped by almost two-fold during germination. The mechanism involved in the variation of CHH methylation during seed development is still unclear because the authors did not observe any changes in *MET1*, *CMTs*, or *DRMs* expression. To havea better understanding of the role of CHH methylation during seed development, they analyzed the Arabidopsis *ddcc* quadruple mutant (i.e., *drm1drm2cmt2cmt3*), which is deficient in all methyltransferases involved in CHH and CHG methylation. Interestingly, the *ddcc* mutantdid not show any major seed developmental defect or major changes in gene expression, suggesting that CHH does not play a fundamental role in proper embryo or seed development. This hypothesis was confirmed by the analysis of methylation levels within or closely relatedto genes essential in seed development and seed germination such as storage proteins, oil biosynthesis, master regulators of seed development, and germination-enhanced proteins. The authors revealed that almost 50% of these genes were localized in regions poor in methylation, called demethylated valleys (DMVs). The seed DMVs represented 21% of the genome and appeared to be consistent in all tested seed tissues. These DMVs also appeared to be enriched in transcription factors (TFs, with 46% of them), and in genes involved in embryo formation and seed development (e.g., *WOX*, *CUC*, *CLAVATA*, *PIN1* genes). These hypomethylated regions did not show any variation in methylated state during seed development, whereas the genes contained in these regions were highly transcriptionally active and tightly regulated. To explain this gene regulation, they observed that the repressive histone mark H3K27m3 and the bivalent marks H3K27me3/H3K4me3 showed some modulation in these regions during seed development that appeared to be correlated with TF gene expression, suggesting a regulation of these TF expressions via histone mark modifications rather than DNA methylation. Finally, in the same study, they revealed that CHH methylation and also the CHG methylation were concentrated upstream and downstream of the coding sequence and within TEs, but very low in gene bodies; in contrast to CG methylation, which was mainly located in gene bodies. Changes in differentially methylated regions between the three developmental stages appeared to be enriched in CHH methylation sites. Indeed,97% of DMRs were linked to CHH context, and 65% of these CHH-DMRs were found to be differentially methylated between the three stages and located close to transcribed genes. In the *ddcc* mutant, transposases of 106 TEs were upregulated and showed a high densityof CHH methylation sites, suggesting that CHH methylation couldplay a role in repression of TEs during seed development. Although most of these results were obtained in Arabidopsis, the authors mentioned that the overall regulation of methylation during seed development seems highly similar in Arabidopsis and soybean. Therefore,results obtained from the Arabidopsis *ddcc* mutant could be extrapolated to soybean seed development.

3 | STRESS AND ADAPTABILITY

3.1 | Biotic stress

DNA (de)methylation was found to play critical roles in defense responses against a wide variety of pathogens (for review Deleris, Halter, & Navarro, 2016). In plant defense, resistance (R) genes encode Nucleotide binding Leucine rich Repeat (NLR) proteins, which play critical roles in effector perception for triggering immunity. Under

normal growth conditions, R proteins are maintained at low steady state levels and require a high degree of control to prevent fitness costs (Shirasu, 2009). In common bean, out of 197 CG-methylatedNLR genes, 172 (87.3%) showed low to undetectable expression levels, suggesting that DNA methylation could be an alternative way of transcriptionally silencing R proteins under normal growth conditions to avoid fitness cost due to their unnecessary accumulation. NLR proteins are organized in clusters that are often located close to the terminal knobs containing the satellite DNA *khipu.* Following thisobservation, Richard et al. (2018) suggested that methylation of NLR genes could result from the spreading of DNA methylation from *khipu* in common bean. In addition, it was shown that 24 nt siRNAs targeted 24% of NLR genes, which were identified as methylated, validating a potential regulation of NLR expression through RNA-directed DNA methylation (RdDm) pathway (Richard et al., 2018).

RNA silencing mediated by siRNA is employed in plant defense against a variety of pathogens from bacteria, to fungi and viruses. To better understand the importance of this mechanism in plant defense, Garg et al. (2017) analyzed correlations between transcript levels of DICER-LIKE (*DCL*), ARGONAUTE (*AGO*), and RNA-dependent RNA polymerase (*RDR*) gene families in Chickpea infected by *Ascochyta*Blight (AB, *Ascochyta rabiei*), pigeon pea infected by Sterility Mosaic Disease (SMD) and groundnut subjected to rust (*Puccinia arachidis)* and late leaf spot fungus (*Phaeoisariopsis personata).* A general trend of upregulation of siRNA biogenesis genes in resistant genotypes and downregulation of genes was observed in susceptible genotypes, including *RDR2*, *AGO7* in chickpea, *DCL2*, *DCL4*, *RDRs* genes in pigeonpea, and *DCL2* in groundnut, (Garg et al., 2017). A specific focus hasbeen done on these genes as Arabidopsis *ago7* and *rdr* and Tomato *dclb2* mutants were found to be more susceptible to fungal and viral pathogens (Ellendorff, Fradin, de Jonge, & Thomma, 2009; Wanget al., 2018).

Several studies also demonstrated that histone methylation and histone acetylation play a role in plant immunity. ChIP-seq experiments of the repressive mark H3K9me2 and active mark H4K12Ac combined with RNA sequencing in common bean at different stages of *Uromyces appendiculatus* infection revealed key genes related to the bean-rust interaction. Expression profiles of genes such as defense response genes (e.g., low molecular weight cysteine 68, GIGANTEA protein and DnaJ-domain chaperone superfamily), R proteins (e.g., Pleiotropic drug resistance protein 12, MATE efflux family and NB-ARC domain-containing) were correlated with changes of histone methylation and histone acetylation modification (Ayyappan et al., 2015).

Defense priming is an intrinsic protective mechanism, in which plants prime their defense mechanisms after a first attack/infection in order to defend themselves more rapidly in subsequent interactions with pathogens (Mauch-Mani, Baccelli, Luna, & Flors, 2017). It hasbeen shown that this phenomenon is related to the dynamics of chromatin structure. In Common bean, (pre)treatment with salicylic acid analogs such as BABA or INA enhanced resistance against *P. syringae pv. Phaseolicola,* with a protective effect transmitted to the next generation. It has been shown that this effect was due to the induction of pathogen-associated genes such as *PR1*, *PR4*, *NPR1*, and *WRKY29*, *WRKY53*, *WRKY6*, correlated with enrichment of the active histone mark H3K4me3 at the junction between promoter and coding regions in these genes (Martínez-Aguilar, Ramírez-Carrasco, Hernández-Chávez, Barraza, & Alvarez-Venegas, 2016).

Effectors secreted by pathogens are also known to target the components of HAT or HDAC complexes, thereby manipulating plant immunity. The ADH2 and GCN5 subunits of the SAGA complex (i.e., multi-protein chromatin modifying complex) are essential for HAT activity, which activates gene expression via acetylation of H3K9. Two robust studies in soybean highlighted the action of pathogen effectors in modulating plant immunity. The *Phytophthora sojae* effector PsAvh23 has been shown to bind to GmGNC5, which disrupts its assembly with ADH2, thereby decreasing H3K9 acetylation and resulting in transcriptional repression of soybean defense genes (Kong et al., 2017). Similarly, PsAvh52, an effector at the early stage of infection, interacts with GmTAP1, an acetyltransferase, regulating histone acetylation and promoting expression of susceptibility genes (Li et al., 2018).

Finally, WRKY TFs are well-documented players in plant defense, regulating transcript levels of many defense-related genes (Birkenbihl, Liu, & Somssich, 2017; Pandey & Somssich, 2009). In chickpea, following *Fusarium oxysporum f. sp. Cicero Race 1* infection, high expression of *WRKY40* in resistant plants was associated with high enrichment of the active mark H3K9Ac in its promoter region, which could suggest a role of histone activation marks in increasing resistance to this pathogen (Chakraborty, Ghosh, Sen, & Das, 2018).

3.2 | Abiotic stress

Several studies have shown correlations between changes in methylation levels and environmental stresses, suggesting potential involvement of epigenetic mechanisms in plant adaptability. For instance, drought stress in faba bean and water deficit in pea were associated with an overall increase of DNA methylation in both tolerant and sensitive genotypes (Abid et al., 2017; Labra et al., 2002). In contrast, salt stress in pigeon pea induced a global decrease of DNA methylation in shoot (Awana et al., 2019). On a longer term, continuous stress increased global DNA demethylation mainly in a tolerant soybean genotype. This increased demethylation was consistent with increased expression of DNA demethylases such as DML and ROS1. The demethylation analysis revealed that CG and CHG contexts within gene regulatory regions were more critical than CHH in soybean adaptation to stress (Liang et al., 2019). In contrast, salinity stress in *Medicago truncatula*, induced up to 77% of changes in CHH context, with only 9.1% and 13.9% in CHG and CG, respectively. However, no correlation between transcript level and DNA methylation pattern of some key genes known to be involved in salinity stress was reported, implying that these genes might be regulated by other epigenetic mechanisms (Yaish, Al-Lawati, Al-Harrasi, & Patankar, 2018). In contrast, Song et al. (2012) showed that, in soybean, among four TFs induced under salt stress, three were

demethylated in CG and non-CG contexts, preceding enrichments of active histone marks (H3K4me3 and H3K9ac) and decrease of the repressive mark H3K9me2, leading to gene upregulation, suggesting the possible interplay between DNA methylation and histone modification in stress response.

A growing body of evidence assigns crucial roles of histone acetylation and histone methylation in plant responses to external stress. Changes in DNA methylation, histone methylation, and histone acetylation were observed in soybean root meristems growing at different temperatures. Immunostaining patterns indicated that 5-methylcytidine (i.e., a marker of methylated DNA) and H3K9me2, mainly located in the heterochromatin, were more abundant in soybean during chilling stress than during recovery. In contrast, H3K9ac, H4K12ac, and H3K4me, indicators of permissive chromatin, were weakly labeled in the euchromatin of stressed plants, but stronger during the recovery process (Stępiński, 2012). Interestingly, crosstalk between histone methylation and histone acetylation was also reported in soybean subjected to salinity stress. Wu et al. (2011)

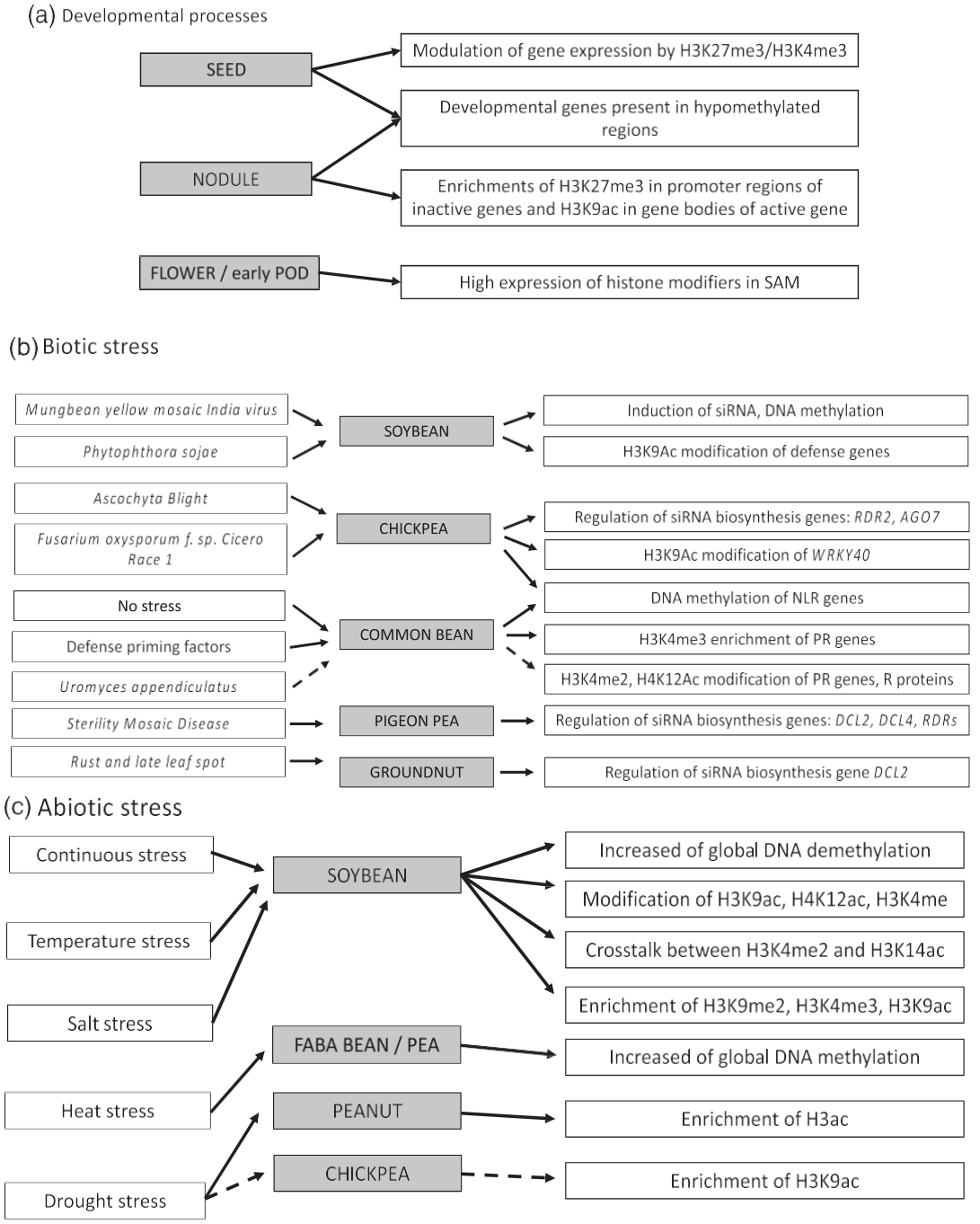


FIGURE 1 Summary of (a) developmental processes, (b) biotic, and (c) abiotic stress responses to different legumes species and their corresponding epigenetic mechanisms

proposed that the salinity stress-inducible plant homeodomain TF, GmPHD5, could bind salt-induced H3K4me2 marks. This binding allowed recruitment of a complex involved in gene activation with non-histone proteins such as GmISWI, a chromatin remodeling factor and GmGNAT1, an acetyl transferase, which can preferentially acetylate H3K14 to further activate expression of salinity-induced genes.

In peanut, another study showed a regulation of gene expression of the *AhDREB1* gene. The regulation, by acetylation of H3 enabled this member of the AP2/ERF TF family to positively regulate drought stress related genes, under PEG osmotic stress. Indeed, expression of *AhDREB1* was showed to be higher using trichostatin (TSA), an inhibitor of HDAC (histone deacetylase), eventually inducing drought resistance (Zhang, Su, et al., 2018). Salt and drought stresses have also been showed to induce the activation of *CaHDZ12*, a HD-Zip TF, in chickpea, whose expression was correlated with acetylation of H3K9ac in the promotor region (Sen, Chakraborty, Ghosh, Basu, & Das, 2017).

4 | PERSPECTIVES

From this extensive review of legume epigenetic studies, we can clearly appreciate the importance of epigenetic regulations in devel-opmental and stress-related processes (summarized in Figure 1). Con- sidering legume crops not only with their large genomes containing many TEs and repeat regions but also their genes with high copy num- ber (e.g., as described in this review with histone modifiers), their numerous small RNAs, and their specific legume processes (e.g., nodulation), there is no doubt that we will observed a growing interest in legume epigenetic studies in order to understand specific developmental processes and adaptative responses to environmental constraints in legumes. To conclude, epigenetic studies in legumes are still at an early developing stage and have been predominantly focus- ing on identification of key epigenetic players in different plant devel-opmental or stress-related processes. This initial descriptive step is essential as most legume genomes are still poorly annotated and con-tains many genes with high copy number that could have overlapping or distinct functions. Perspectives will be an increase of functional studies of these key epigenetic players that could be enhanced by the rapid development of CRISPR technologies to generate collection of epigenetic mutants in major legume crops. From this perspective, a better understanding of epigenetic mechanisms and the identification of epialleles in legumes will potentially boost plant crop improvement and stress adaptation.

ACKNOWLEDGMENTS

This work was conducted in the framework of the regional programme "Objectif Végétal, Research, Education and Innovation in Pays de la Loire," supported by the French Region Pays de la Loire, Angers LoireMetropole, and the European Regional Development Fund.

CONFLICT OF INTEREST

The authors have no conflict of interest.

AUTHOR CONTRIBUTION

TTD, ZC, DW, and JV wrote, reviewed and approved the finalmanuscript.

ETHICS STATEMENT

This manuscript does not contain any studies with human or animal subjects.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Jerome Verdier https://orcid.org/0000-0003-3039-2159

REFERENCES

Abid, G., Mingeot, D., Muhovski, Y., Mergeai, G., Aouida, M., Abdelkarim, S., … Jebara, M. (2017). Analysis of DNA methylation pat- terns associated with drought stress response in faba bean (Vicia faba L.) using methylation-sensitive amplification polymorphism (MSAP). Environmental and Experimental Botany., 142, 34–44. https://doi.org/ 10.1016/j.envexpbot.2017.08.004

An, Y. Q. C., Goettel, W., Han, Q., Bartels, A., Liu, Z., & Xiao, W. (2017). Dynamic changes of genome-wide DNA methylation during soybean seed development. Scientific Reports., 7(1), 12263–12263. https://doi. org/10.1038/s41598-017-12510-4

Awana, M., Yadav, K., Rani, K., Gaikwad, K., Praveen, S., Kumar, S., & Singh, A. (2019). Insights into salt stress-induced biochemical, molecu- lar and epigenetic regulation of spatial responses in pigeonpea (Cajanus cajan L.). Journal of Plant Growth Regulation., 38, 1545–1561. https://doi.org/10.1007/s00344-019-09955-4

Ayyappan, V., Kalavacharla, V., Thimmapuram, J., Bhide, K. P., Sripathi, V. R., Smolinski, T. G., & Kingham, B. (2015). Genome-wide profiling of histone modifications (H3K9me2 and H4K12ac) and gene expression in rust (Uromyces appendiculatus) inoculated common bean (Phaseolus vulgaris L.). PLOS ONE, 10(7), e0132176. https://doi.org/10. 1371/journal.pone.0132176

Benhamed, M., Bertrand, C., Servet, C., & Zhou, D. X. (2006). Arabidopsis GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for liqht-responsive gene expression. Plant Cell., 18, 2893–2903. https://doi.org/10.1105/tpc.106.043489

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. Nature., 447, 407–412. https://doi.org/10.1038/ nature05915

Birkenbihl, R. P., Liu, S., & Somssich, I. E. (2017). Transcriptional events defining plant immune responses. Current Opinion in Plant Biology, 38, 1–9. https://doi.org/10.1016/j.pbi.2017.04.004

Chakraborty, J., Ghosh, P., Sen, S., & Das, S. (2018). Epigenetic and tran-scriptional control of chickpea WRKY40 promoter activity under Fusarium stress and its heterologous expression in Arabidopsis leads to enhanced resistance against bacterial pathogen. Plant Science, 276, 250–267. https://doi.org/10.1016/j.plantsci.2018.07.014

Chan, S. W. L., Zhang, X., Bernatavichute, Y. V., & Jacobsen, S. E. (2006). Two-step recruitment of RNA-directed DNA methylation to tandem repeats. PLoS Biology, 4(11), e363. https://doi.org/10.1371/journal. pbio.0040363

Chen, L.-T., & Wu, K. (2010). Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. Plant Signal Behav, 5, 1318–1,320. https://doi.org/10.4161/psb.5.10.13168 PubMed: 20930557

Deal, R. B., & Henikoff, S. (2011). Histone variants and modifications in plant gene regulation. Current Opinion in Plant Biology., 14, 116–122. https://doi.org/10.1016/j.pbi.2010.11.005

Deans, C., & Maggert, K. A. (2015). What do you mean, "Epigenetic"? Genetics., 199, 887–896. https://doi.org/10.1534/genetics.114.173492

Deleris, A., Halter, T., & Navarro, L. (2016). DNA methylation and demethylation in plant immunity. Annual Review of Phytopathology, 54(1), 579–603. https://doi.org/10.1146/annurev-phyto-080615-100308

Ellendorff, U., Fradin, E. F., de Jonge, R., & Thomma, B. P. H. J. (2009). RNA silencing is required for Arabidopsis defence against Verticillium wilt disease. Journal of Experimental Botany, 60(2), 591–602. https://doi.org/10.1093/jxb/ern306

Finke, A., Kuhlmann, M., & Mette, M. F. (2012). IDN2 has a role down- stream of siRNA formation in RNA-directed DNA methylation. Epige- netics., 7, 950–960. https://doi.org/10.4161/epi.21237

Garg, V., Agarwal, G., Pazhamala, L. T., Nayak, S. N., Kudapa, H., Khan, A. W., … Varshney, R. K. (2017). Genome-Wide Identification, Characterization, and Expression Analysis of Small RNA Biogenesis Purveyors Reveal Their Role in Regulation of Biotic Stress Responses in Three Legume Crops. Frontiers in Plant Science, 8. https://doi.org/ 10.3389/fpls.2017.00488

Groszmann, M., Greaves, I. K., Albertyn, Z. I., Scofield, G. N., Peacock, W. J., & Dennis, E. S. (2011). Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. Proc Natl Acad Sci USA, 108, 2617–2,622. https://doi.org/10. 1073/pnas.1019217108 PubMed: 21266545

Herr, A. J. (2005). Pathways through the small RNA world of plants. FEBS Letters., 579, 5879–5888. https://doi.org/10.1016/j.febslet.2005.08.040

Jeong, J.-H., Song, H.-R., Ko, J.-H., Jeong, Y.-M., Kwon, Y. E., Seol, J. H., Amasino, … Noh, Y.-S. (2009). Repression of FLOWERING LOCUS T Chromatin by Functionally Redundant Histone H3 Lysine 4 Demethylases in Arabidopsis. PLoS ONE, 4(11), e8033. https://doi. org/10.1371/journal.pone.0008033

Jiang, D., Wang, Y., Wang, Y., & He, Y. (2008). Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb Repressive Complex 2 Components. PLoS ONE, 3(10), e3404. https://doi.org/10.1371/journal.pone.0003404

Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., & Jacobsen, S. E. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current Biology., 17, 379–384. https://doi.org/10.1016/j.cub.2007.01.009

Kawakatsu, T., Huang, S. s. C., Jupe, F., Sasaki, E., Schmitz, R. J. J., Urich, M. A. A., … Ecker, J. R. (2016). Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. Cell., 166, 492–505. https://doi.org/10.1016/j.cell.2016.06.044

Kong, L., Qiu, X., Kang, J., Wang, Y., Chen, H., Huang, J., … Wang, Y. (2017). A phytophthora effector manipulates host histone acetylation and reprograms defense gene expression to promote infection. Current Biology, 27(7), 981–991. https://doi.org/10.1016/j.cub.2017.02.044

Krogan, N. T., Hogan, K., & Long, J. A. (2012). APETALA2 negatively regu- lates multiple floral organ identity genes in Arabidopsis by recruiting the corepressor TOPLESS and the histone deacetylase HDA19. Devel- opment, 139, 4180–4,190. https://doi.org/10.1242/dev.085407 PubMed:23034631

Kumar, S. V., & Wigge, P. A. (2010). H2A.Z-containing nucleosomes medi- ate the thermosensory response in arabidopsis. Cell., 140, 136–147. https://doi.org/10.1016/j.cell.2009.11.006

Labra, M., Ghiani, A., Citterio, S., Sgorbati, S., Sala, F., Vannini, C., … Bracale, M. (2002). Analysis of cytosine methylation pattern in response to water deficit in pea root tips. Plant Biology, 4(6), 694–699. https://doi.org/10.1055/s-2002-37,398

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modify- ing DNA methylation patterns in plants and animals. Nature Reviews Genetics., 11, 204–220. https://doi.org/10.1038/nrg2719

Lei, M., Zhang, H., Julian, R., Tang, K., Xie, S., & Zhu, J. K. (2015). Regula- tory link between DNA methylation and active demethylation in Ara- bidopsis. Proceedings of the National Academy of Sciences of the United States of America., 112, 3553–3557. https://doi.org/10.1073/pnas.1502279112

Li, H., Wang, H., Jing, M., Zhu, J., Guo, B., Wang, Y., … Wang, Y. (2018). A Phytophthora effector recruits a host cytoplasmic transacetylase into nuclear speckles to enhance plant susceptibility. eLife, 7. https://doi.org/10.7554/elife.40039

Liang, X., Hou, X., Li, J., Han, Y., Zhang, Y., Feng, N., … Fang, S. (2019). High-resolution DNA methylome reveals that demethylation enhances adaptability to continuous cropping comprehensive stress in soybean. BMC Plant Biology, 19 (1). https://doi.org/10.1186/s12870-019-1670-9

Liew, L. C., Singh, M. B., & Bhalla, P. L. (2013). An RNA-Seq Transcriptome Analysis of Histone Modifiers and RNA Silencing Genes in Soybean during Floral Initiation Process. PLoS ONE, 8(10), e77502. https://doi.org/10.1371/journal.pone.0077502

Lin, J. Y., Le, B. H., Chen, M., Henry, K. F., Hur, J., Hsieh, T. F., … Goldberg, R. B. (2017). Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proceedings of the National Academy of Sciences of the United States of America., 114, E9730–E9739. https://doi.org/10. 1073/pnas.1716758114

Liu, B., & Wendel, J. F. (2003). Epigenetic phenomena and the evolution of plant allopolyploids. Mol Phylogenet Evol, 29, 365–379. https://doi.org/10.1016/S1055-7903(03)00213-6 PubMed: 14615180

Lu, F., Cui, X., Zhang, S., Liu, C., & Cao, X. (2010). JMJ14 is an H3K4 demethylase regulating flowering time in Arabidopsis. Cell Res, 20, 387–390. https://doi.org/10.1038/cr.2010.27 PubMed: 20177424

Luo, M., Wang, Y.-Y., Liu, X., Yang, S., Lu, Q., Yuhai, C., Keqiang, W. (2012). HD2C interacts with HDA6 and is involved in ABA and salt stress response in Arabidopsis. J Exp Bot, 63, 3297–3,306. https://doi.org/10. 1093/jxb/ers059 PubMed: 22368268

Martínez-Aguilar, K., Ramírez-Carrasco, G., Hernández-Chávez, J. L., Barraza, A., & Alvarez-Venegas, R. (2016). Use of BABA and INA As Activators of a Primed State in the Common Bean (Phaseolus vulgaris L.). Frontiers in Plant Science, 7. https://doi.org/10.3389/fpls.2016. 00653

Matzke, M. A., & Mosher, R. A. (2014). RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. Nature Reviews Genetics., 15, 394–408. https://doi.org/10.1038/nrg3683

Mauch-Mani, B., Baccelli, I., Luna, E., & Flors, V. (2017). Defense priming: An adaptive part of induced resistance. Annual Review of Plant Biology, 68(1), 485–512. https://doi.org/10.1146/annurev-arplant-042916-041132

Mlura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., & Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature., 411, 212–214. https://doi.org/10.1038/35075612

Nagymihály, M., Veluchamy, A., Györgypál, Z., Ariel, F., Jégu, T., Benhamed, M., … Kondorosi, É. (2017). Ploidy-dependent changes in the epigenome of symbiotic cells correlate with specific patterns of gene expression. Proceedings of the National Academy of Sciences of the United States of America., 114, 4543–4548. https://doi.org/10.1073/ pnas.1704211114

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., … Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature., 543, 72–77. https://doi.org/10.1038/ nature21373

Nie, W. F., Lei, M., Zhang, M., Tang, K., Huang, H., Zhang, C., … Zhu, J. K. (2019). Histone acetylation recruits the SWR1 complex to regulate active DNA demethylation in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America., 116, 16641–16650. https://doi.org/10.1073/pnas.1906023116

Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., & Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell., 120, 613–622. https://doi.org/10.1016/j.cell.2005.02.007

Pandey, S. P., & Somssich, I. E. (2009). The role of WRKY transcription fac- tors in plant immunity: Figure 1. Plant Physiology, 150(4), 1,648–1,655. https://doi.org/10.1104/pp.109.138990

Peterson, C. L., & Laniel, M. A. (2004). Histones and histone modifications. Current Biology: CB., 14, R546–R551. https://doi.org/10.1016/j.cub. 2004.07.007

Pien, S., Fleury, D., Mylne, J. S., Crevillen, P., Inzé, D., Avramova, Z., … Grossniklaus, U. (2008). ARABIDOPSIS TRITHORAX1 Dynamically Regulates FLOWERING LOCUS C Activation via Histone 3 Lysine 4 Trimethylation. The Plant Cell, 20(3), 580–588. https://doi.org/10. 1105/tpc.108.058172

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M. A., … Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes and Development., 19, 2030–2040. https://doi.org/10.1101/gad.348405

Richard, M. M. S., Gratias, A., Thareau, V., Do Kim, K., Balzergue, S., Joets, J., … Geffroy, V. (2018). Genomic and epigenomic immunity in common bean: The unusual features of NB-LRR gene family. DNA Research., 25, 161–172. https://doi.org/10.1093/dnares/ dsx046

Satgé, C., Moreau, S., Sallet, E., Lefort, G., Auriac, M.-C., Remblière, C., … Gamas, P. (2016). Reprogramming of DNA methylation is critical for nodule development in Medicago truncatula. Nature Plants, 2(11). https://doi.org/10.1038/nplants.2016.166

Saze, H., Scheid, O. M., & Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant game- togenesis. Nat Genet, 34, 65–69. https://doi.org/10.1038/ng1138 PubMed:12669067

Sekhon, R. S., & Chopra, S. (2009). Progressive loss of DNA methylation releases epigenetic gene silencing from a tandemly repeated maize Myb gene. Genetics., 181, 81–91. https://doi.org/10.1534/genetics.108.097170

Sen, S., Chakraborty, J., Ghosh, P., Basu, D., & Das, S. (2017). Chickpea WRKY70 regulates the expression of a homeodomain-leucine zipper (HD-Zip) I transcription factor CaHDZ12, which confers abiotic stress tolerance in transgenic tobacco and chickpea. Plant and Cell Physiology, 58(11), 1934–1952. https://doi.org/10.1093/pcp/pcx126

Shirasu, K. (2009). The HSP90-SGT1 chaperone complex for NLR immune sensors. Annual Review of Plant Biology, 60(1), 139–164. https://doi. org/10.1146/annurev.arplant.59.032607.092906

Sigman, M. J., & Slotkin, R. K. (2016). The first rule of plant transposable element silencing: Location, location, location. The Plant Cell, 28(2), 304–313. https://doi.org/10.1105/tpc.15.00869

Simpson, G. G. (2004). The autonomous pathway: epigenetic and posttran- scriptional gene regulation in the control of Arabidopsis flowering time. Curr Opin Plant Biol, 7, 570–574. https://doi.org/10.1016/j.pbi.2004.07.002 PubMed: 15337100

Song, Y., Ji, D., Li, S., Wang, P., Li, Q., & Xiang, F. (2012). The Dynamic Changes of DNA Methylation and Histone Modifications of Salt Responsive Transcription Factor Genes in Soybean. PLoS ONE, 7(7), e41274. https://doi.org/10.1371/journal.pone.0041274

Stępin´ski, D. (2012). Levels of DNA methylation and histone methylation and acetylation change in root tip cells of soybean seedlings grown at different temperatures. Plant Physiology and Biochemistry, 61, 9–17. https://doi.org/10.1016/j.plaphy.2012.09.001

Sura, W., Kabza, M., Karlowski, W. M., Bieluszewski, T., Kus- Slowinska, M., Pawełoszek, Ł., … Ziolkowski, P. A. (2017). Dual role of the histone variant H2A.Z in transcriptional regulation of stress- response genes. Plant Cell., 29, 791–807. https://doi.org/10.1105/ tpc.16.00573

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: Provocative insights from epigenomics. Nature Reviews Genetics., 9, 465–476. https://doi.org/10.1038/nrg2341

Tamada, Y., Yun, J.-Y., Woo, S. C., & Amasino, R. M. (2009). ARABIDOPSIS TRITHORAX-RELATED7 is required for methylation of lysine 4 of his- tone H3 and for transcriptional activation of FLOWERING LOCUS C. The Plant Cell, 21(10), 3,257–3,269. https://doi.org/10.1105/tpc.109.070060

Tang, K., Lang, Z., Zhang, H., & Zhu, J. K. (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. Nature Plants, 2 16169https://doi.org/10.1038/nplants.2016.169

Vinardell, J. M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., … Kondorosi, E. (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in Medicago truncatula nodules. Plant Cell., 15, 2093–2105. https://doi.org/10.1105/tpc.014373

Wang, P., Shi, S., Ma, J., Song, H., Zhang, Y., Gao, C., … Wang, X. (2018). Global Methylome and gene expression analysis during early peanut pod development. BMC Plant Biology., 18. https://doi.org/10.1186/ s12870-018-1546-4

Wang, T., Deng, Z., Zhang, X., Wang, H., Wang, Y., Liu, X., & Zhu, H. (2018). Tomato DCL2b is required for the biosynthesis of 22-nt small RNAs, the resulting secondary siRNAs, and the host defense against ToMV. Horticulture Research, 5(1). https://doi.org/10.1038/s41438- 018-0073-7

Whittaker, C., & Dean, C. (2017). The FLC locus: A platform for discoveries in epigenetics and adaptation. Annual Review of Cell and Developmental Biology, 33(1), 555–575. https://doi.org/10.1146/annurev-cellbio-100,616-060546

Wierzbicki, A. T., Ream, T. S., Haag, J. R., & Pikaard, C. S. (2009). RNA poly- merase v transcription guides ARGONAUTE4 to chromatin. Nature Genetics., 41, 630–634. https://doi.org/10.1038/ng.365

Wong, C. E., Singh, M. B., Bhalla, P. L. (2013). The Dynamics of Soybean Leaf and Shoot Apical Meristem Transcriptome Undergoing Floral Initi- ation Process. PLoS ONE, 8(6), e65319. https://doi.org/10.1371/ journal.pone.0065319

Wu, T., Pi, E.-X., Tsai, S.-N., Lam, H.-M., Sun, S.-M., Kwan, Y., & Ngai, S.-M. (2011). GmPHD5 acts as an important regulator for crosstalk between histone H3K4 di- methylation and H3K14 acetylation in response to salinity stress in soybean. BMC Plant Biology, 11(1), 178. https://doi. org/10.1186/1471-2229-11-178

Yaish, Mahmoud W., Al-Lawati, Abbas, Al-Harrasi, Ibtisam, Patankar, Himanshu Vishwas (2018) Genome-wide DNA Methylation analysis in response to salinity in the model plant caliph medic (Medicago truncatula). BMC Genomics, 19 (1), https://doi.org/10. 1186/s12864-018-4484-5

Zemach, A., Kim, M. Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., … Zilberman, D. (2013). The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell, 153(1), 193–205. https://doi. org/10.1016/j.cell.2013.02.033

Zhang, B., Su, L., Hu, B., & Li, L. (2018). Expression of AhDREB1, an AP2/ERF Transcription Factor Gene from Peanut, Is Affected by Histone Acetylation and Increases Abscisic Acid Sensitivity and Tol- erance to Osmotic Stress in Arabidopsis. International Journal of Molecular Sciences, 19(5), 1441. https://doi.org/10.3390/ ijms19051441

Zhang, H., Lang, Z., & Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. Nature Reviews Molecular Cell Biology., 19, 489–506. https://doi.org/10.1038/s41580-018-0016-z

Zhang, H., & Zhu, J. K. (2011). RNA-directed DNA methylation. Current Opinion in Plant Biology., 14, 142–147. https://doi.org/10.1016/j.pbi. 2011.02.003

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., …Ecker, J. R. R. (2006). Genome-wide high-resolution mapping and functional analysis of dna methylation in Arabidopsis. Cell., 126, 1189–1201. https://doi.org/10.1016/j.cell.2006.08.003

Zhao, T., Zhan, Z., & Jiang, D. (2019). Histone modifications and their regu- latory roles in plant development and environmental memory. Journal of Genetics and Genomics, 46(10), 467–476. https://doi.org/10.1016/j. jgg.2019.09.005

Zhou, C., Zhang, L., Duan, J., Miki, B., & Wu, K. (2005). HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in Arabidopsis. Plant Cell Online, 17, 1196–1204. PubMed: 15749761

Zhu, Y., Rowley, M. J., Böhmdorfer, G., & Wierzbicki, A. T. (2013). A SWI/- SNF chromatin-resmodeling complex acts in noncoding RNA-mediated transcriptional silencing. Molecular Cell., 49, 298–309. https://doi.org/ 10.1016/j.molcel.2012.11.011

Zilberman, D., Coleman-Derr, D., Ballinger, T., & Henikoff, S. (2008). His- tone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature., 456, 125–129. https://doi.org/10.1038/nature07324

How to cite this article: Windels D, Dang TT, Chen Z, Verdier J. Snapshot of epigenetic regulation in legumes. *Legume Science*. 2020;1-11. https://doi.org/10.1002/leg3.60

4.2 Polycomb Repressive Complexes and the role of histone H3 Lysine 27 trimethylation (H3K27me3) in seeds

As described in the previous section, gene expression is controlled by genetic factors and epigenetic modifications. When plants switch their developmental program in the so-called transition phases, such as flowering to seed development or seed development to germination, polycomb group (PcG) proteins play a crucial role in controlling these processes (for review Yang *et al.* 2017; Yan *et al.* 2020). PcG proteins usually form different complexes that are organized into polycomb repressive complex 1 (PRC1) and polycomb repressive complex 2 (PRC2) depending on their action to silence specific chromatin regions via different histone modifications. PRC1 complexes silence chromatin by deposition of ubiquitin of lysine 119 on histone 2A (H2AK119Ub). In contrast, PRC2 complexes are able to silence chromatin by targeting the lysine 27 on histone 3 by trimethylation (H3K27me3). This trimethylation of H3 at lysine 27 is a strong repressive mark, widely distributed in the genome (*i.e.* about 20% of *Arabidopsis* genes have been found linked to H3K27me3 mark), which is dynamically deposited and removed during plant development, leading to chromatin condensation (Zheng and Chen, 2011). Interestingly, PRC2 (as well as PRC1) complexes are highly conserved in plants, even more broadly from Drosophila to mammals (a comparison of PcG proteins in different organisms is available in Yan *et al.* 2020). In *Arabidopsis*, the four core components of PRC2 are (i) *CURLY LEAF* (*CLF*), *SWINGER* (*SWN*) and *MEDEA*, which have the histone methyltransferase (HMT) activity to trimethylate H3K27, (ii) *VERNALIZATION2* (*VRN2*), *FERTILIZATION INDEPENDENT SEEDS2* (*FIS2*), and *EMBRYONIC FLOWER2* (*EMF2*), which confer target specificity of the PRC2 complexes, (iii) *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*), which stabilizes and enhances the HMT activity and (iv) five *MULTICOPY SUPPRESSOR OF IRA* (*MSI1* to *MSI5*), which have a binding activity to histones and HMT. As three proteins exist in *Arabidopsis* to confer target specificity, it has been suggested that at least three PRC2 complexes (*i.e.* EMF-PRC2, VRN-PRC2, and FIS-PRC2) act in specific development processes to regulate different gene sets in *Arabidopsis* (Schuettengruber *et al.*, 2007; Zheng and Chen, 2011).

PRC2 complexes regulate the initiation of seed development phase and are required in the transition between embryogenesis and seed maturation. Indeed, it has been shown that, in absence of fertilization, FIS-PRC2 is involved in repressing endosperm development, whereas

VRN-PRC2 and EMF-PRC2 in repression of seed coat development (Lau *et al.*, 2012; Figueiredo and Köhler, 2018). Moreover, mutations in FIS-PRC2 complex impaired development of the three seed tissues, leading to seed abortion (Robert *et al.*, 2018; Robert, 2019). Another example, in rice, the loss-of-function mutants of *fie1* displayed delayed embryo development and decreased seed size. Interestingly, *OsFIE1* regulated target genes related to heat stress by altering their H3K27me3 levels (Huang *et al.*, 2016). As expected, PRC2 complexes also interplay with the *LAFL* genes. As major regulators of seed development, *LEC1* and the three B3 domains have to be tightly regulated regarding their temporal and spatial expressions by chromatin dynamics, even if the precise role of PRC2 complexes on LAFL regulation is still unclear. During the embryogenesis-maturation transition, it has been reported that *FUS3* is repressed by PcG during early embryogenesis (Makarevich *et al.*, 2006) and that, in *fie* mutants, *ABI3*, *FUS3*, and *LEC2* genes were upregulated (Bouyer *et al.*, 2011; Kim *et al.*, 2012). However, when and how the seed development genes switch from the repressive state to active state during seed development is still not completely understood. Regarding the second transition phase in seeds between maturation to germination, several reports showed the involvement of the PRC2 complexes in the repression of the *LAFL* genes and regulation of germination. For instance, *ABI3* and *LEC2* gene sequences are associated with the active histone marks (H3K4me3) before germination, which turned to be replaced by repressive H3K27me3 marks upon germination (Bouyer *et al.*, 2011; Müller *et al.*, 2012; Molitor *et al.*, 2014). Another example is the implication of *FIE* as binding factor of genes encoding positive regulators of ABA signaling pathway, such *ABI3* and *ABI4*, but also negative regulators of GA signaling pathway, such *RGA-Like 3* (*RGL3*) and finally, positive regulators of maturation such as *LEC2* which promote embryo development and maturation to repress their expressions (Bouyer *et al.*, 2011; Kim *et al.*, 2012; Deng *et al.*, 2013). Finally, following the action of PRC2 complexes in repressing seed maturation genes using H3K27me3, the role of the PRC1 complex is to recognize and bind to H3K27me3 to maintain the condensed chromatin state (for review Zheng and Chen, 2011). In *Arabidopsis*, the VAL (VP1/ABI3-LIKE) 1/2/3 proteins have been identified as partners of the PRC1 complex to silence *ABI3*, *FUS3*, and *LEC2* expressions to initiate germination and vegetative development (Yang *et al.*, 2013).

During seed maturation, an increase of chromatin condensation was observed, potentially acting through ABI3 (Van Zanten *et al.*, 2011). The nuclei size and chromatin state are recovered during seed germination. Desiccation tolerance, longevity and germinative qualities are acquired during seed maturation and as the repressive histone marks H3K27me3

are responsible for maintenance of chromatin compaction and repression of transcription of genes involved in seed development, in the present study, we propose to explore whether chromatin condensation is involved in acquisition of longevity and germinative qualities. Furthermore, this study will also investigate whether H3K27me3 repressive marks are impacted by heat stress conditions during seed development, and if these changes are associated with the seed physiological response to heat stress.

5. Aim of the thesis project

As developed in the previous sections, in a global warning context, heat stress has dramatic impact on plant development including acquisition of seed quality traits, which represents an important threat to food security. Molecular responses to heat have been intensively studied in plant vegetative parts but little is known about molecular processes occurring during seed development.

The aim of this PhD project is to explore and describe seed molecular processes affected during heat stress in order to identify candidate genes potentially sensing or regulating stress response and explaining the phenotypic plasticity of seed quality traits following heat stress.

To achieve this goal, this thesis project was initially organized in three work packages (WP).

WP1 intends to unravel the molecular mechanisms occurring in seed tissues underlying the physiological changes such as seed weight, longevity and vigour when seeds suffered from heat stress during their development. In this part, the physiological, transcriptomic and epigenetic (*i.e.* DNA methylation and H3K27me3) changes in response to heat stress in seed will be explored using the *Medicago truncatula* reference genotype A17 to (i) first identify candidate genes affecting proper seed development during heat stress, then (ii) uncover the level of (epi)genetic regulation of these candidate genes.

In parallel, WP2 proposes to use the natural genetic diversity present in the *Medicago truncatula* HapMap collection to identify loci/genes potentially involved in seed trait plasticity in response to heat stress using a genome-wide association study (GWAS) approach.

Finally, WP3 corresponds to an integrative package of WP1 and WP2 to select a short list of relevant candidate genes, which could be involved in seed trait plasticity in response to heat stress and acting at different levels of (epi)genetic regulation. This work package also includes a preliminary functional characterization of candidate genes using *Medicago Tnt1* and *Arabidopsis* T-DNA insertional mutant populations to select the most relevant genes for future studies.

Due to the truncation of WP3 because of the Covid-19 pandemic and different lockdowns that slowed down our progress, we decided to reformat this thesis manuscript into two parts: Chapter 2 corresponding to **Genome-wide analyses of transcriptome, methylome and chromatin dynamics of isolated seed tissues in heat stress conditions** and Chapter 3 corresponding to **Study of the regulation of seed traits in optimal and heat stress conditions using natural *Medicago truncatula* accessions and genome-wide association studies**. The integration of both WPs was not performed; however, some preliminary functional analyses of candidate genes are introduced at the end of both chapters.

CHAPTER 2: GENOME-WIDE ANALYSES OF TRANSCRIPTOME, METHYLOME AND CHROMATIN DYNAMICS OF ISOLATED SEED TISSUES IN HEAT STRESS CONDITIONS

AIM: the aim of this first section is to unravel the molecular mechanisms occurring in seed tissues underlying the physiological changes such as seed weight, longevity and vigour when seeds suffered from heat stress during their development using the *Medicago truncatula* reference genotype A17. In this part, the physiological, transcriptomic and epigenetic (*i.e.* DNA methylation and H3K27me3) changes in response to heat stress in seed will be explored to (i) first identify candidate genes affecting proper seed development during heat stress, then (ii) uncover the level of (epi)genetic regulation of these candidate genes.

MAIN RESULTS:

- High transcriptome dynamics in isolated seed tissues during seed development under control and heat stress conditions and identification of molecular heat stress responses.
- Impacts of heat stress on the acquisition of seed quality traits, leading to impairments in seed weight, seed longevity and seed vigour traits such as germination speed.
- Complete transcriptional reprogramming of seed maturation during heat stress with changes in expressions of *LAFL* genes and other major regulators of seed development.
- Descriptive map of (epi)genetic (transcript, DNA methylation and H3K27me3 dynamics) regulations during embryo development under control and heat stress conditions
- Identification of a possible role of DNA methylation in heat stress response at early seed maturation with a massive increase in differentially methylated regions at S1
- Identification of a possible role of H3K27me3 dynamic in heat stress response during mid and late seed maturation with an important increase of H3K27me3 genomic binding sites at S2, S3 and S4.
- Identification of candidate genes potentially involved in phenotypic changes following heat stress (*i.e.* mainly seed longevity and seed weight) with some already known functions in seed development but others with unknown functions.

- Preliminary results from functional analysis showing a seed-specific *HAP3* transcription factor with a potential role in regulating seed maturation following heat stress.

1. Introduction

The effect of global warming has tremendous impacts on plant growth such as earlier flowering time, modification of plant architecture, decrease of seed yield and alteration of seed properties (Wahid *et al.*, 2007; Long and Ort, 2010; Lobell *et al.*, 2011; Li *et al.*, 2014), which represents a threat to food security and agricultural production. More specifically on seed qualities/performances, studies showed the impact of climate change and heat stress on seed development affecting the seed qualities (Righetti *et al.*, 2015). Along with physiological changes, molecular responses to heat stress have been extensively studied in the past decades in plant vegetative organs. Heat stress factors (HSF), heat shock proteins (HSP), calcium signaling, phytohormones, chaperone proteins and secondary metabolites play an important role in heat stress responses (Bokszczanin and Fragkostefanakis, 2013; Ohama *et al.*, 2017). Indeed, an important transcriptional reprogramming allows plants to cope with heat stress by the induction of thermotolerance induced genes and the repression of growth-related genes. Moreover, gene regulation is strongly related to epigenetic regulation based on chromatin modifications (*e.g.* histone acetylation, methylation and ubiquitylation) and DNA modifications (*e.g.* cytosine methylation). These modifications can be triggered by developmental or environmental factors and by consequence, they modulate chromatin architecture without changing the genomic sequences, but only the accessibility of transcriptional machinery to specific genome regions (Lukens and Zhan, 2007; Chinnusamy and Zhu, 2009).

In this chapter, we intend to describe physiological and molecular responses to heat stress during the, yet unexplored, plant reproductive stage corresponding to seed development/maturation to decipher the impacts of stress on the acquisition of seed quality traits. This exploratory study will use genome-wide analyses to define the transcriptome, DNA methylome and chromatin dynamics from isolated seed tissues in response to heat stress conditions at four key stages during seed maturation using the *Medicago truncatula* reference genotype A17. This analysis intends to identify relevant molecular processes and candidate genes regulating seed physiological traits from different seed tissues and at different level of (epi)genetic regulation.

2. Results and discussion

2.1 Physiological and morphological characterizations of mature seeds developed under heat stress conditions

After harvest of mature seeds developed under control (20°C) and heat stress (26°C) conditions, seeds were immediately dried at 44% relative humidity (RH) (using a saturated solution of K_2CO_3 at 20°C) for three days. Physiological dormancy was released using post-harvest treatment by storage of mature seeds for 6 months at room temperature, then seeds were subsequently stored at 4°C in the dark for subsequent analyses.

The first morphological characterization to measure was the seed weight/size. Indeed, we observed a visible difference in seed weight/size between mature seeds produced at 20°C and 26°C. To evaluate this difference, we measured the dry seed weights of ten seed lots containing 30 seeds from each seed batch produced in both conditions. Results showed that the individual dry weight of mature seeds produced at 26°C was significantly lower than mature seeds produced at 20°C (*i.e.* decrease of about 25% of seed weight from 26°C, Figure 2.1A). To consolidate this observation, we characterized the morphological features by image analysis using ImageJ software. The representative parameter such as seed area was automatically measured for each seed. The average areas confirmed a significant decrease in size of seeds produced under heat stress conditions (Figure 2.1B). All together, these results suggested that plants grown in heat stress conditions produced smaller seeds than the control condition seeds in the *Medicago truncatula* reference genotype A17.

Seed longevity and germination are important traits of seed quality. Both are acquired during seed development, more precisely during the maturation phase in *M. truncatula*. Firstly, in order to verify whether heat stress influences seed longevity, we used the artificial ageing conditions of 75% relative humidity at 35°C to accelerate seed ageing in order to observe any change in seeds produced in both conditions. We artificially aged mature seeds during 10, 20, 30, 40 and 50 days (called days of artificial ageing, DAA), and for each time point we evaluated the germination percentage of seeds differentially aged. From this experiment, we observed a decrease of germination capacity of seeds artificially aged for long periods. Interestingly, from the survival curves, we observed that seeds produced at 26°C showed more longevity capacity and were capable to be viable for longer time in artificial ageing conditions compared to 20°C seeds (Figure 2.1C). The P50, a representative value for longevity which is the storage time to

lose 50% of seed viability during the controlled ageing condition, was calculated for both seed lots (produced at 20°C and 26°C). The P50 of 26°C seeds was significantly higher than control seeds by almost 10 days, suggesting that seeds produced in heat stress condition had higher longevity (Figure 2.1D).

Seed germination plays a vital role in the transition from seed to seedling establishment. In field conditions, seed germination happens at low temperature during winter or spring. Thus, to investigate if seed germination was affected by the heat stress during seed development/maturation, we evaluated germination performances at both low (10°C) and optimal (20°C) conditions using mature seeds produced at 20°C and 26°C. In optimal germination conditions, we did not observe any significant difference between seeds produced at 20°C and 26°C. In low-temperature germination conditions (10°C), there was no difference regarding the final germination rate, which reached 100% in both seed lots. However, we observed a decrease in germination speed of heat stressed seeds when compared to control seeds from the germination kinetics (Figure 2.1E). The time to reach 50 % of total germination rate, which is called T50, was longer from stressed seeds compared to control (around 3 hours slower, Figure 2.1F), which suggested that heat stress during seed production ultimately decreased seed germination speed at low temperature and impaired seed vigor.

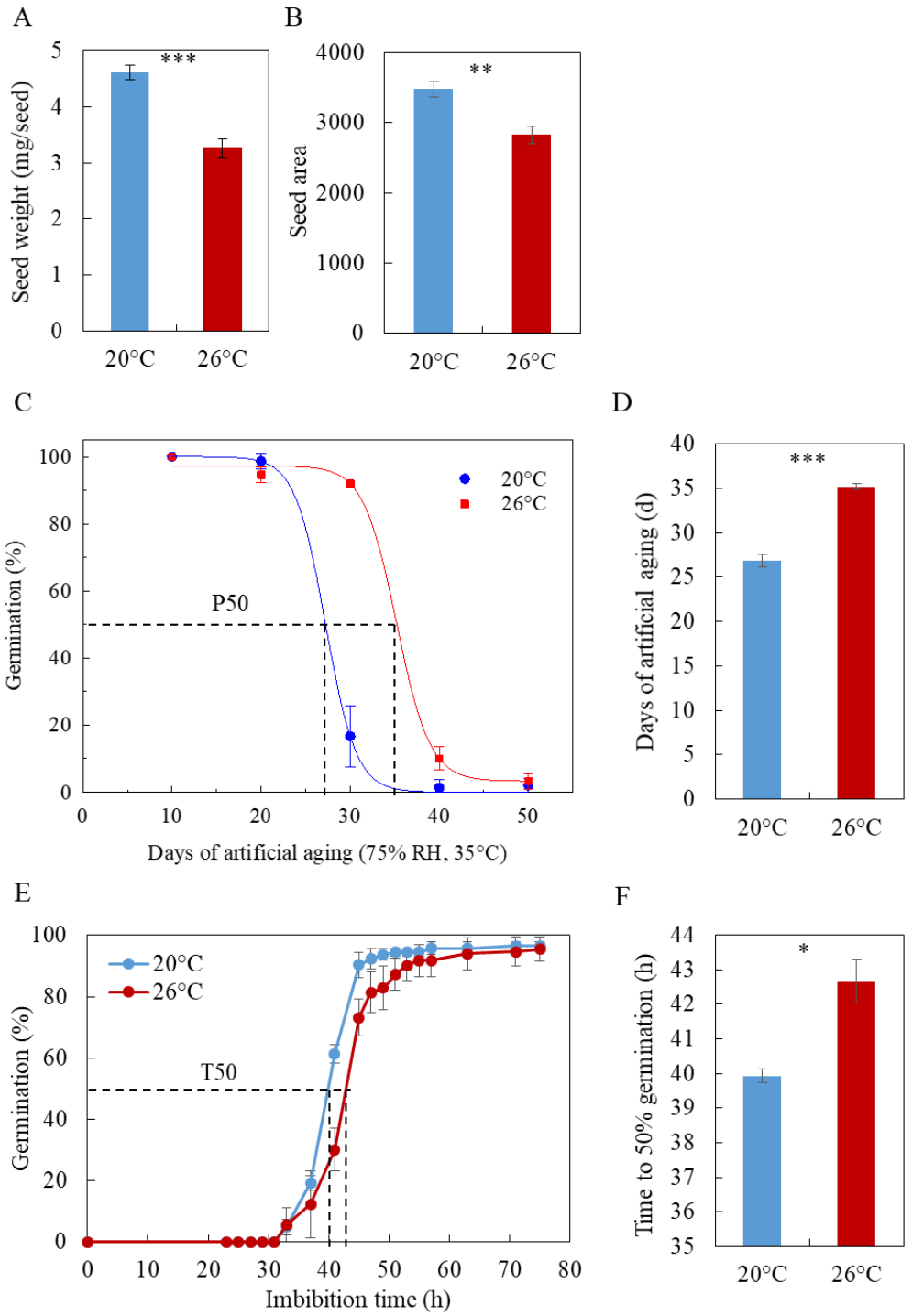
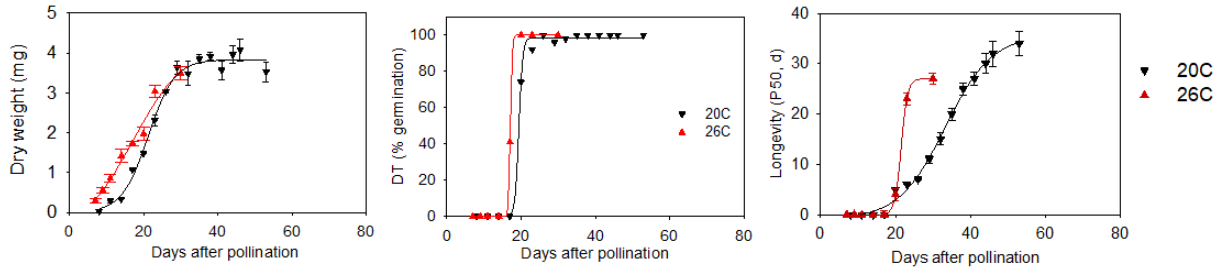


Figure 2.1. Physiological and morphological characterizations of seeds produced in control and heat stress conditions. (A-B) Dry weight and area of 20°C and 26°C mature seeds. (C) The survival curves of 20°C and 26°C mature seeds during artificial ageing. (D) P50 of 20°C and 26°C mature seeds. (E) Germination curves of 20°C and 26°C mature seeds at 10°C. (F) Time to 50% germination (T50) of 20°C and 26°C mature seeds. 20°C seed (control) is indicated in bleu, 26°C seed (heat stress) is indicated in red. *, 0.01<p-value<0.05; **, 0.001<p-value<0.01; ***, p-value<0.001.

2.2 Global transcriptome changes in response to heat stress in developing *M. truncatula* seeds

As shown in Righetti *et al.* (2015), heat stress during seed development has a dramatic influence on seed developmental timing, ranging from 28 to 48 days of seed development duration to reach complete maturity. In order to select appropriate and corresponding seed developmental stages between seeds produced at 20°C and 26°C, we used two different criteria. First, a physiological criterion by analysing the acquisition of two major processes during seed development: desiccation tolerance and longevity. For that purpose, we selected stages before and after the acquisition of these processes (Figure 2.2A). Second, we also combined these physiological characterizations with molecular data obtained from the microarray data of developmental stages of seeds produced in heat stress conditions and available in Righetti *et al.* (2015) by generating a correlation matrix to pinpoint corresponding stages based on global transcriptomic changes (Figure 2.2B). By crossing these information, we decided to select four stages, named S1, S2, S3 and S4, corresponding to i) the beginning of seed maturation (S1: 17DAP from 20°C seeds and 14DAP from 26°C seeds), ii) after acquisition of desiccation tolerance and during seed filling (S2: 26DAP from 20°C seeds and 17DAP from 26°C seeds), iii) beginning of the acquisition of longevity and at the end of seed filling (S3: 36DAP from 20°C seeds and 22DAP from 26°C seeds) and vi) at seed maturity (*i.e.* dry seed) (S4: 44DAP from 20°C seeds and 28DAP from 26°C seeds) (Figure 2.2C). To reveal the underlying molecular mechanisms of these four stages impacted by heat stress, we first analysed the transcriptome changes of the three isolated seed tissues at these four developmental stages during seed maturation under control and heat stress conditions.

A



B

CORRELATION MATRIX		20°C														
		8 DAP	*11 DAP*	*14 DAP*	*17 DAP*	*20 DAP*	*23 DAP*	*26 DAP*	*29 DAP*	*32 DAP*	*35 DAP*	*38 DAP*	*41 DAP*	*44 DAP*	*ABS*	*MAT*
26 °C	7 DAP	0.940	0.956	0.957	0.897	0.890	0.867	0.848	0.811	0.837	0.818	0.798	0.791	0.789	0.804	0.778
	9 DAP	0.924	0.950	0.957	0.945	0.934	0.912	0.888	0.851	0.876	0.850	0.830	0.815	0.806	0.821	0.794
	11 DAP	0.917	0.937	0.947	0.953	0.944	0.928	0.908	0.876	0.888	0.867	0.853	0.831	0.826	0.838	0.811
	14 DAP	0.889	0.890	0.895	0.927	0.937	0.944	0.937	0.926	0.925	0.910	0.908	0.869	0.857	0.870	0.847
	17 DAP	0.886	0.890	0.885	0.925	0.941	0.950	0.950	0.940	0.957	0.941	0.934	0.900	0.873	0.889	0.870
	20 DAP	0.862	0.858	0.851	0.884	0.901	0.915	0.928	0.929	0.948	0.955	0.953	0.929	0.893	0.910	0.899
	ABS	0.779	0.771	0.759	0.780	0.775	0.814	0.829	0.861	0.852	0.887	0.903	0.920	0.920	0.913	0.941
	MAT	0.780	0.773	0.761	0.781	0.778	0.816	0.832	0.861	0.854	0.892	0.902	0.922	0.921	0.913	0.944

C

	S1: beginning of seed maturation	S2: after acquisition of desiccation tolerance and during seed filling	S3: beginning the acquisition of longevity and at the end of seed filling	S4: mature seed
	STAGE1	**STAGE2**	**STAGE3**	**STAGE4**
20°C	**17DAP**	**26DAP**	**36DAP**	**44DAP**
26°C	**14DAP**	**17DAP**	**22DAP**	**28DAP**

Figure 2.2. Characterizations of physiological processes and developmental stages of *Medicago* seed produced in control and heat stress conditions. (**A**) Physiological characterizations of seeds produced in control and heat stress conditions regarding acquisitions of dry weight, desiccation tolerance (DT) and longevity (P50). (**B**) Correlation matrix between transcriptome changes obtained from developmental stages of seeds produced at 20°C and 26°C based on microarray data generated in Righetti *et al.* (2015). (**C**) Summary of selected stages and corresponding seed physiological processes.

2.2.1 RNA sequencing data for heat stress response in isolated *Medicago truncatula* seed tissues

(This section was published in journal of Data in Brief.)

Data in Brief 35 (2021) 106726

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

RNA sequencing data for heat stress response in isolated *medicago truncatula* seed tissues

Zhijuan Chen, Benoit Ly Vu, Olivier Leprince, Jerome Verdier*

Institut de Recherche en Horticulture et Semences-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071, Beaucouzé, France

A R T I C L E I N F O

Article history:
Received 4 November 2020
Revised 8 December 2020
Accepted 6 January 2021
Available online 21 January 2021

Keywords:
RNA sequencing
Seed maturation
Heat stress
Seed quality
Embryo
Endosperm
Seed coat

A B S T R A C T

Legumes are important crop species as they produce highly nutritious seeds for human food and animal feed. In grain legumes, sub-optimal conditions affect seed developmental timing leading to impairment of seed quality traits acquired during seed maturation. To understand the molecular mechanisms of heat stress response in legume seeds, we analysed transcriptome changes of three seed tissues (i.e. em bryo, endosperm and seed coat) at four developmental stages, during seed maturation, from seed filling to mature dry seeds, collected under optimal and heat stress conditions in the model legume, *Medicago truncatula* (reference genotype A17). The total RNA sequencing generated a dataset of 48 samples, representing more than 57 Gb fastq raw data. Mapping, quantification and annotation of the data were based on fifth release of *Medicago truncatula* genome and provided expression profiles of 44,473 transcripts in seed tissues at different developmental stages and under optimal and stress conditions. Time-course and pairwise comparisons between optimal and stress conditions showed that 9182, 8315 and 3481 genes were differentially expressed due to heat stress in embryo, endosperm and seed coat respectively. Moreover,

* Corresponding author.
E-mail address: jerome.verdier@inrae.fr (J. Verdier).
Social media: (J. Verdier)

https://doi.org/10.1016/j.dib.2021.106726

it highlighted a common set of 975 genes that were differentially expressed in all the seed tissues.

Specifications Table

Subject	Agricultural and Biological Sciences
Specific subject area	Omics: Transcriptomics Plant Science: Plant Physiology
Type of data	Tables Figures
How data were acquired	Total RNA samples were sent to BGI, Hong Kong, for library preparation. Libraries were constructed following a custom protocol from samples that passed quality controls (mass > 2 µg, concentration>80 ng/microl, OD260/280 ≅ 2.00, OD260/230 ≅ 2.20, RIN>6.5, 28S/18S<1.0, baseline smooth). After mRNA enrichment by rRNA depletion and oligo dT selection, RNA was fragmented and reverse transcribed to double-strand cDNA (dscDNA) by N6 random primer. The synthesized cDNA was subjected to end-repair and then was 3′ adenylated. Adaptors were ligated to the ends of these 3′ adenylated cDNA fragments. The ligation products were purified and many rounds of PCR amplification were performed to enrich the purified cDNA template using PCR primer, splint oligo and DNA ligase, followed by sequencing on BGISEQ-500 platform, generating an average 24 M reads of 50 bp per sample.
Data format	Filtered raw reads (FASTQ) Analysed RNA-seq data files (counts and DEG lists)
Parameters for data collection	Total RNAs were extracted from isolated embryo, endosperm and seed coat of four developmental stages of *Medicago truncatula* (A17) seeds that were harvested during maturation phase before and after acquisition of desiccation tolerance (respectively S1 and S2) and at the onset and after longevity acquisition (respectively S3 and S4) under standard temperature (20 °C day/ 18 °C night) and under high temperature (26 °C/24 °C).
Description of data collection	RNA-seq dataset was collected from single-end sequencing of cDNA libraries using BGISeq500 platform with 50 bp reads. Raw reads were filtered to remove adapters and low-quality reads, then mapped to *Medicago truncatula* reference transcriptome (version 5). Total mapped reads and number of transcripts (counts and TPM) were estimated using Salmon algorithm. Differential expressions of genes between standard and heat stress conditions were calculated using ImpulseDE2 and DEseq2 algorithms.
Data source location	Institution: Growth chambers from the Institut de Recherche en Horticulture et Semences, INRAE City: Beaucouzé Country: France
Data accessibility	Public Repository: Repository name: NCBI GEO Data identification number: GSE160725 Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE160725

Value of the Data

- These data represent valuable seed transcriptome dataset of heat stress in the model legume *Medicago truncatula* because it has been generated from isolated seed tissues (embryo, endosperm and seed coat) along the whole seed maturation.
- These data are useful resources for scientific communities working on seed and legume seed quality but also on plant stress biology to understand specific and common stress response pathways.

- These data provide new insights about molecular processes affected during heat stress in different seed tissues and candidate genes and molecular markers to predict seed quality in sub-optimal conditions.

1. Data Description

This manuscript presents a transcriptomic dataset obtained from dissected seed tissues (embryo, endosperm and seed coat) at four developmental stages of *Medicago truncatula* (A17) seeds during maturation phase [i.e. before and after desiccation tolerance (respectively S1 and S2) and at the onset and after longevity acquisition (respectively S3 and S4)] produced under standard temperature (20 °C day/ 18 °C night) and heat stress (26 °C/24 °C). Table 1 displays all sample names with information regarding seed tissues, treatments, developmental stages and numbers of 50 bp reads sequenced per sample. Figure S1 provides sequence quality histograms obtained from FastQC [1] and merged into a single graphic using MultiQC [2]. All 48 samples displayed Phred quality scores about 35, which corresponds to a base calling accuracy of about 99.95%. Salmon algorithm [3] was used to map and quantify raw reads to the reference *Medicago truncatula* transcriptome version 5 and corresponding count table is provided as Table S1. After mapping and quantification, these 48 samples were, then, hierarchically clustered based on their dissimilarity scores to validate reproducibility of replicates (Fig. 1). From this count table, differentially expressed genes were identified using time-course comparisons of the four developmental stages for embryo (Table S2) and endosperm (Table S3); and using pairwise comparisons of the two developmental stages for seed coat (Table S4). Fig. 2 summarizes the differentially expressed genes (DEGs) in the three seed tissues between optimal and heat stress conditions (False Discovery Rate, FDR<1%) and highlights a common set of 975 DEGs within the three seed tissues.

Table 1
Summary of sample files with corresponding information related to seed tissues, seed developmental stages, growth conditions and numbers of cleaned 50 bp reads used for RNA-seq mapping.

Tissues	Sample	Numbers of 50 bp reads	Sample	Numbers of 50 bp reads	Stages
Embryo	R20E-17–1	20,452,270	R26E-14–1	20,822,668	S1
Embryo	R20E-17–2	20,465,064	R26E-14–2	20,786,562	S1
Embryo	R20E-17–3	20,364,545	R26E-14–3	20,718,145	S1
Embryo	R20E-26–1	20,452,569	R26E-17–1	20,666,800	S2
Embryo	R20E-26–2	20,137,018	R26E-17–2	20,704,789	S2
Embryo	R20E-26–3	21,967,973	R26E-17–3	20,753,459	S2
Embryo	R20E-36–1	22,003,911	R26E-22–1	20,717,753	S3
Embryo	R20E-36–2	22,156,003	R26E-22–2	20,786,906	S3
Embryo	R20E-36–3	20,183,128	R26E-22–3	20,644,844	S3
Embryo	R20E-44–1	20,760,225	R26E-28–1	20,649,142	S4
Embryo	R20E-44–2	20,897,210	R26E-28–2	20,749,797	S4
Embryo	R20E-44–3	20,744,207	R26E-28–3	20,778,364	S4
Embryo	R20Eo-17–1	20,677,185	R26Eo-14–1	20,303,572	S1
Endosperm	R20Eo-17–2	20,659,667	R26Eo-14–2	20,257,351	S1
Endosperm	R20Eo-26–1	20,815,605	R26Eo-17–1	20,342,065	S2
Endosperm	R20Eo-26–2	20,174,987	R26Eo-17–2	20,221,552	S2
Endosperm	R20Eo-36–1	20,332,236	R26Eo-22–1	20,288,004	S3
Endosperm	R20Eo-36–2	20,129,499	R26Eo-22–2	20,463,956	S3
Endosperm	R20Eo-44–1	20,829,827	R26Eo-28–1	20,371,220	S4
Endosperm	R20Eo-44–2	20,810,635	R26Eo-28–2	20,405,449	S4
Seed Coat	R20SC-17–1	40,072,723	R26SC-14–1	41,944,942	S1
Seed Coat	R20SC-17–2	56,846,997	R26SC-14–2	42,596,392	S1
Seed Coat	R20SC-26–1	39,483,027	R26SC-17–1	22,920,874	S2
Seed Coat	R20SC-26–2	48,482,527	R26SC-17–2	34,721,666	S2
	CONTROL CONDITION		STRESS CONDITION		

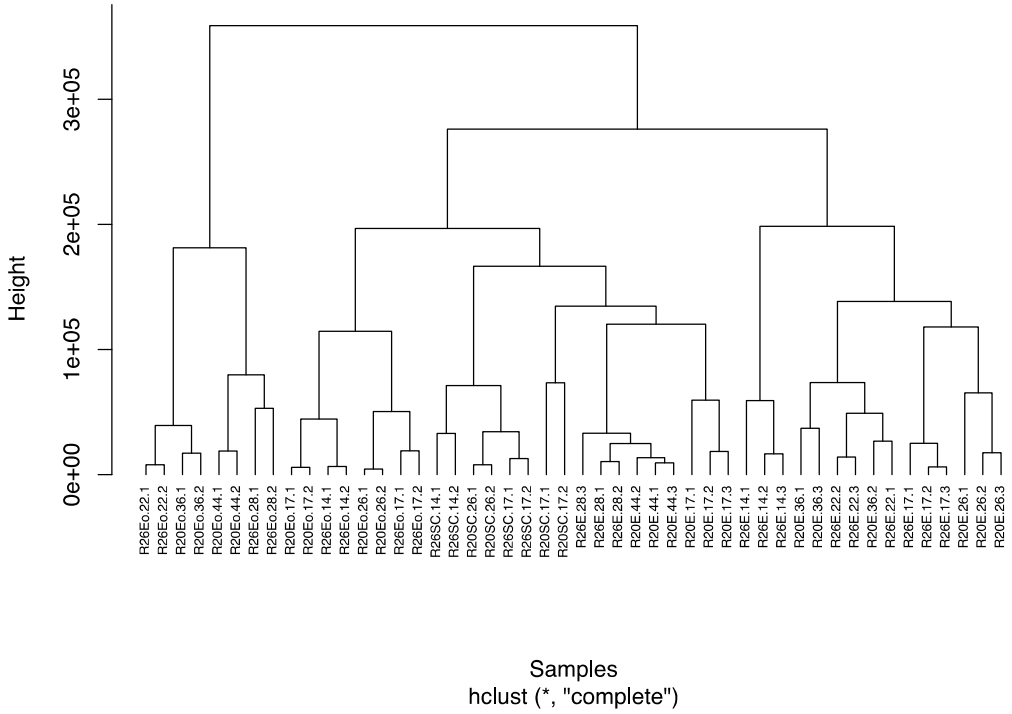


Fig. 1. Hierarchical cluster analysis based on dissimilarity scores obtained from the TPM values of the 48 samples to validate reproducibility of replicates.

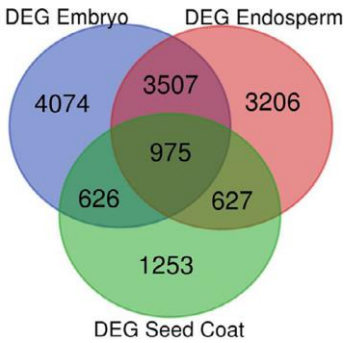


Fig. 2. Venn diagram with differentially expressed genes (DEGs, adjusted p-values below 1%) in the three seed tissues between optimal and heat stress conditions.

2. Experimental Design, Materials and Methods

2.1. Plant growth conditions and seed tissue sampling

Medicago plants were grown under standard conditions (20 °C/18 °C, 16 h photoperiod) in growth chamber. At flowering time, half of plants were kept at same optimal conditions (20 °C/18 °C, 16 h photoperiod) and half were grown under heat stress condition (26 °C/24 °C, 16 h photoperiod). Developing and mature seeds were collected from standard and stress conditions, seed tissues were quickly dissected then immediately frozen in liquid nitrogen before RNA extraction. According to our previous study [4], four seed developmental stages were collected under standard conditions at 17 days after pollination (DAP), 26 DAP, 36DAP and 44 DAP and

under heat stress conditions at corresponding developmental stages at 14 DAP, 17 DAP, 22 DAP and 28 DAP. These four developmental stages were shown to be the beginning of seed maturation and before acquisition of desiccation tolerance (S1), after the acquisition of desiccation tolerance and during seed filling (S2), at the onset of the acquisition of longevity and at the end of seed filling (S3) and mature seed (S4) according to our previous study [4].

2.2. RNA isolation and sequencing

Total RNAs were extracted from embryo and endosperm that were collected at four stages during seed maturation under standard and heat stress conditions. RNAs were only extracted from S1 and S2 for seed coat, as during late maturation stage seed coat is dying and does not provide good quality RNA. All RNA extractions were performed using MACHEREY-NAGEL NucleoSpin® RNA Plus kit. RNA quality was checked using a nanodrop spectrophotometer ND-1000 (NanoDrop Technologies) and a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All samples with good qualities (260/280 and 260/230 absorbance ratio >2; RNA Integrity Number, RIN>8; 28S/18S>1.7) were sent to Beijing Genomics Institute (https://www.bgi.com) (Hong Kong) for library preparation and sequencing on BGISEQ-500 platform, generating an average of 20 M reads of 50 bp per sample.

2.3. RNA-seq data analyses

After quality control of fastq files using FastQC [1], high-quality reads were mapped onto the reference Medicago A17 transcriptome version 5 (https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/downloads/1.7/MtrunA17r5.0-ANR-EGN-r1.7.fastaFiles.zip) (Table S5) [5] and transcript abundances were quantified with Salmon algorithm (version 0.14.1) [3] using the quasi-mapping mode and the '–validateMappings', '–useVBOpt' and '–seqBias' options. Reproducibility of replicates was validated by a hierarchical cluster analysis based on dissimilarity scores obtained from the normalized raw count values of the 48 samples using the 'cpm', 'dis' and 'hclust' functions in R. Differentially expressed genes (DEGs) were identified by time-course comparisons of the four developmental stages of embryo and endosperm using ImpulseDE2 [6] and by pair-wise comparisons of the two developmental stages of seed coat using DESeq2 [7]. All transcripts with change in expression showing adjusted p-values below 1% (no fold change cutoff) were considered as differentially expressed between standard and heat stress conditions and displayed on a Venn Diagram performed from the website: http://bioinformatics.psb.ugent.be/webtools/Venn/.

Ethics Statement

This work does not contain any studies with human or animal subjects.

CRediT Author Statement

Zhijuan Chen: Investigation, Methodology, Visualization, Writing - Original draft preparation; **Benoit Ly Vu:** Methodology; **Olivier Leprince:** Conceptualization, Reviewing and Editing; **Jerome Verdier:** Supervision, Conceptualization, Data curation, Writing - Reviewing and Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

This research was conducted in the framework of the regional program “Objectif Végétal, Research, Education and innovation in Pays de la Loire”, supported by the French Region Pays de la Loire, Angers Loire Métropole.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.106726.

References

[1] S. Andrews, FastQC - A quality control tool for high throughput sequence data, Babraham. Bioinforma (2010) http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[2] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics (2016), doi:10.1093/bioinformatics/btw354.
[3] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, Nat. Methods. (2017), doi:10.1038/nmeth.4197.
[4] J. Verdier, D. Lalanne, S. Pelletier, I. Torres-Jerez, K. Righetti, K. Bandyopadhyay, O. Leprince, E. Chatelain, B.L. Vu, J. Gouzy, P. Gamas, M.K. Udvardi, J. Buitink, A Regulatory Network-Based Approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of medicago truncatula Seeds, PLANT Physiol 163 (2013) 757–774, doi:10.1104/pp.113.222380.
[5] Y. Pecrix, S.E. Staton, E. Sallet, C. Lelandais-Brière, S. Moreau, S. Carrère, T. Blein, M.-F. Jardinaud, D. Latrasse, M. Zouine, M. Zahm, J. Kreplak, B. Mayjonade, C. Satgé, M. Perez, S. Cauet, W. Marande, C. Chantry-Darmon, C. Lopez-Roques, O. Bouchez, A. Bérard, F. Debellé, S. Muños, A. Bendahmane, H. Bergès, A. Niebel, J. Buitink, F. Frugier, M. Benhamed, M. Crespi, J. Gouzy, P. Gamas, Whole-genome landscape of Medicago truncatula symbiotic genes, Nat. Plants. 4 (2018) 1017–1025, doi:10.1038/s41477-018-0286-7.
[6] D.S. Fischer, F.J. Theis, N. Yosef, Impulse model-based differential expression analysis of time course sequencing data, Nucleic Acids Res (2018), doi:10.1093/nar/gky675.
[7] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol 15 (2014) 550, doi:10.1186/s13059-014-0550-8.

2.2.2 Over-representation analyses of differentially expressed genes in seeds

To gain insight into the functional classes corresponding to differentially expressed genes between control and heat stress conditions in different seed tissues, we performed over-representation analyses on the Differentially Expressed Genes (DEGs) that are common or specific to the three seed tissues (*i.e.* embryo, endosperm and seed coat). DEGs (*i.e.* both up- and down-regulated genes) displaying adjusted p-values below 5% according to ImpulseDE2/DEseq2 algorithms, which resulted in 13,811 genes (Supplementary Table S1), were selected to perform gene set enrichment analyses. Mapman functional annotations (Schwacke *et al.*, 2019) of the fifth release of *Medicago truncatula* genome (Pecrix *et al.*, 2018) were used in this analysis. The major significantly over-represented functional classes of different sets of DEGs were shown in Figure 2.3.

2.2.2.1 Enrichment of functional classes common to seed tissues following heat stress

When observing genes that are differentially expressed in all seed tissues due to heat stress, we first selected common DEG in all three seed tissues as general seed response to this abiotic stress. Following heat stress, it was not surprising to observe that one of the most over-represented functional classes was ‘stress.abiotic.Heat’ with 231 annotated DEGs in the three seed tissues (Figure 2.3A). This result validated our growing conditions and confirmed that the heat stress intensity we applied during seed development was sufficient to induce a stress response. Out of these heat stress response genes, we found 14 HSFs mainly from groups A, B and C. However, no ortholog from the *HSF1A* gene family showed differential expression in seeds, which differs from vegetative tissues analysis, where *HSF1A* genes play roles as master regulators of heat stress response (Mishra *et al.*, 2002). Except for *HSFA3 and HSFA5*, which displayed strong differential expression in both seeds and vegetative tissues, other HSFs identified in seeds were not the most affected by heat in vegetative tissues, such as *HSFA8, HSFB2*. This obervertion could suggest a partially specific response to heat stress in seeds compared to vegetative tissues, which needs to be analysed and explored deeply. Out of these stress related genes, we also identified many HSP and sHSP, such as otholog to *HSP17.7,* which enhances heat tolerance in carrot (Malik *et al.*, 1999) and rice (Murakami *et al.*, 2004),

but also orthologs of *HSP70* and *HSP90,* which act as chaperones during heat stress response (Pratt and Toft, 2003).

The second functional class that was enriched from the DEGs common to seed tissues was the 'DNA synthesis.Chromatin structure' with 194 annotated DEGs, which could reflect an intense chromatin dynamic following heat stress (Figure 2.3A). 'Photosynthesis.Light reaction' functional class with many DEGs related to photosystem I & II was also found to be part of the general seed stress response (Figure 2.3A). Indeed, photosynthesis is heat-sensitive and photosystem I & II activities are reduced under heat stress conditions (reviewed in Hasanuzzaman *et al.* 2013). Finally, the functional class related to 'development.storage protein' was also found to be enriched with 30 DEGs following heat stress, which encode for all storage protein families (*i.e.* albumin, vicilin or legumins), suggesting an alteration of nutritional composition of seeds subjected to heat stress (as described in Sehgal *et al.* 2018).

2.2.2.2 Enrichment of functional classes specific to seed tissues

Besides the interest in heat stress DEGs that were commonly expressed differentially in all seed tissues, we aimed to investigate if and how each different seed structure would response to heat stress. For that we gathered tissue-specific DEGs, whose expressions were exclusively up or down-regulated in one of the three seed tissues, *e.g.* seed coat or endosperm or embryo.

In the seed coat we observed specific DEGs representing enrichment of 'RNA regulation. DOF zinc finger' and 'signalling. Calcium' functional classes (Figure 2.3A). The DOF gene family is involved in diverse essential processes in plants, such as seed maturation, seed germination and plant development (Noguero *et al.*, 2013). We identified around 12 DOF transcription factors differentially expressed specifically in seed coat following heat stress. One of them (MtrunA17_Chr8g0346551) is closely related to *AtDAG1* (*DOF AFFECTING GERMINATION 1*), which in *Arabidopsis*, regulates the transition seed to seedling (Papi *et al.*, 2000). Indeed, *dag1* mutant seeds displayed a faster germination due to reduced dormancy (Papi *et al.*, 2000). The action of *DAG1* on seed dormancy and germination was shown to modulate the ABA/GA ratio by regulating the expression of ABA catabolic gene *CYP707A2* and GA biosynthetic gene *GA3ox1* (Boccaccini *et al.*, 2016). We observed that the transcript level of *DAG1* ortholog in *Medicago* was up-regulated in seed coat during early seed

development under heat stress conditions, which could explain the slow germination phenotype of stressed seeds. Concerning the enriched calcium signaling class, calcium is very well-known signal molecules for many biological processes and has been showed to be involved in heat stress signaling (Gong *et al.,* 1998).

In the endosperm, specific DEGs showed enrichments of the 'Polycomb group (PcG)' and 'protein degradation.ubiquitin.E3 RING' functional classes (Figure 2.3A). As explained in the introduction, the PcG proteins are implicated in genomic imprinting in endosperm, which control endosperm and seed development (Ingouff *et al.*, 2005; Huh *et al.*, 2007). In our dataset, four out of the ten genes annotated as PcG proteins showed differential expression in endosperm from seeds produced between control and heat stress conditions, and encode orthologs of *SWINGER* (*SWN*, MtrunA17_Chr1g0194631), *EMBRYONIC FLOWER 2* (*EMF2*, MtrunA17_Chr1g0196821), and two other potential histone methyltransferases (MtrunA17_Chr7g0235081, MtrunA17_Chr7g0235091). Regarding the E3 ubiquitin protein ligase RING, they are a large gene family that can transfer ubiquitin to (i) silence chromatin via the mono-ubiquitination of lysine 19 of histone H2A by the PRC1 complex or (ii) degrade proteins as part of the ubiquitin proteasome system. Furthermore, in the DEG list specific to endosperm , we noted the presence of a DOF transcriptional factor, already identified in *Medicago* and called *DASH* (*DOF Acting in Seed embryogenesis and Hormone accumulation*, MtrunA17_Chr2g0282441) (Noguero *et al.*, 2015). In *dash* mutants, embryo development was shown to be impaired resulting in the reduction of seed size. In our dataset, we observed a down-regulation of *MtDASH* transcript level following a heat stress, which could potentially explain the decrease of mature seed size of stressed seeds.

Finally, embryo-specific DEGs were enriched in genes belonging to many functional classes such as 'cell division', 'G-proteins', 'RNA binding', 'RNA helicase', 'RNA processing', 'RNA regulation of transcription', 'RNA processing splicing' (Figure 2.3A), which suggested that RNA processing could play an important role in heat stress response in embryo.

To further investigate the heat stress response from each seed tissue, we evaluated the functional classes of genes commonly expressed in two tissues, *e.g.* embryo and endosperm (E-Eo), embryo and seed coat (E-SC), endosperm and seed coat (Eo-SC) (Figure 2.3B). Interestingly, common DEGs from embryo and endosperm present some of the same enriched classes as the specific embryo DEGs. These classes were all related to RNA processing which

showed that the embryo RNA processing response against heat stress is also found on the endosperm.

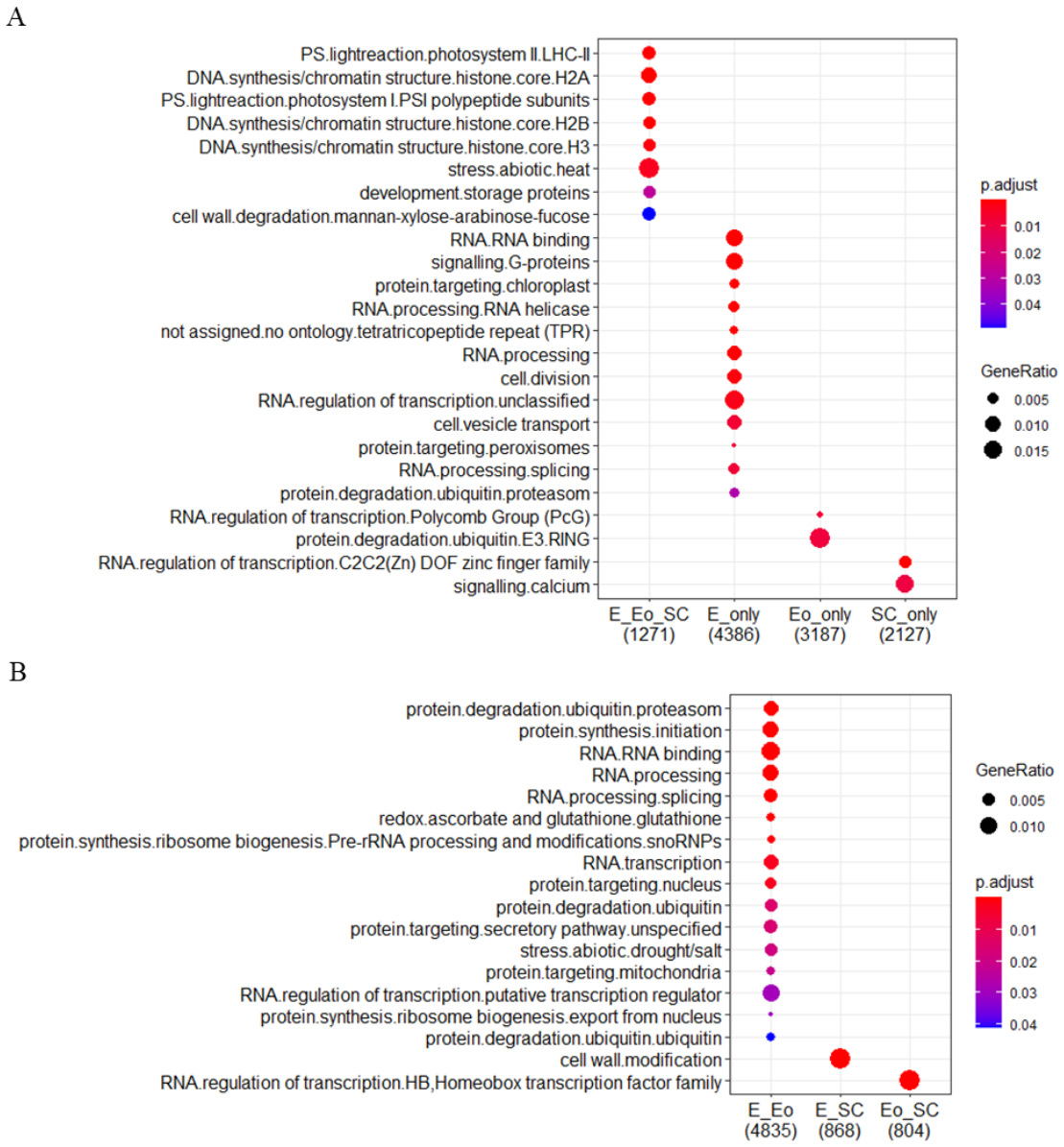


Figure 2.3. Over-representation of functional classes of DEGs identified in the embryo (E), endosperm (Eo) and seed coat (SC). Enriched terms were identified using ClusterProfiler with enricher function and a Bonferroni adjusted p-value threshold of 0.05. Numbers that are indicated at the bottom are the gene numbers used to perform enrichment analyses.

2.2.3 Transcript level of key regulatory genes during seed development under optimal and heat stress conditions

Several central regulatory genes have been identified as major regulators of seed developmental processes, including *LEC1*, *ABI3*, *FUS3* and *LEC2* (*i.e. LAFL* genes). Since heat stress influences seed developmental timing and seed maturation characteristics in *Medicago truncatula*, we investigated whether the key regulatory genes of seed development were affected by heat stress during seed development. Using *Medicago* genome annotation (version 5) and basic local alignment search tool (BLAST), we identified putative orthologs of these genes and extracted their expression profiles during *M. truncatula* seed development (Figure 2.4). *LEC1* is known to be expressed in early seed developmental stage during embryogenesis, thus, we only detected relatively low transcript levels of *MtLEC1* (MtrunA17_Chr1g0165041) at stage 1 during seed development in *Medicago truncatula* (Figure 2.4). In our transcriptomic data, *MtLEC2* (MtrunA17_Chr4g0047031) transcripts were only detected at low levels in stage 1 in endosperm under optimal conditions but no transcript was detected in seed tissues from heat stress conditions. Due to the early expression of these genes during seed development, we did not observe statistical differences in their transcript levels between seeds produced in control and heat stress conditions (Table 1). In contrast, we observed differential expression levels for the other key regulatory genes. Indeed, *MtFUS3* (MtrunA17_Chr7g0251881) showed differential expression in embryo and endosperm due to heat stress and was expressed in embryo and similarly *MtABI3* (MtrunA17_Chr7g0237841) transcript levels were differentially expressed in embryo, endosperm and seed coat between both conditions. In addition, we also looked at the expression profiles of two seed-preferentially transcriptional factors, *MtABI4* (MtrunA17_Chr5g0437371) and *MtABI5* (MtrunA17_Chr7g0266211), which play important roles in seed development. *ABI4*, containing the APETALA2-like domain, and *ABI5*, containing a basic leucine zipper (bZIP) domain, respectively, are implicated in ABA signalling pathway (Finkelstein *et al.*, 1998; Finkelstein and Lynch, 2000). *ABI4* is a positive regulator for primary seed dormancy in *Arabidopsis* (Shu *et al.*, 2013). *ABI5* not only regulates seed dormancy and germination but also plays a role in seed longevity in *Medicago truncatula* (Zinsmeister *et al.*, 2016). In our dataset, we observed that both *MtABI4* and *MtABI5* transcripts were differentially in embryo and endosperm between seeds produced in control and heat stress conditions (Table 1). Taken

together, these results showed that the transcript levels of the key regulators of seed development are impacted by heat stress, and due to their pleiotropic role in seed maturation mechanisms they could represent good candidate genes to explain heat-stressed seed phenotypes. It also suggested that heat stress induced an intense transcriptional reprogramming of seed development/maturation by altering gene expressions of the central LAFL regulatory genes.

Interestingly, when searching for the ortholog of *LEC1* gene, we identified a very closely related gene that corresponds to a *HAP3/NF-YB6* gene annotated as *MtLEC1-LIKE* (*MtLIL*, MtrunA17_Chr4g0076381). In *Arabidopsis*, *LIL* is able to complement *lec1* mutation and is necessary for embryo development and initiation of maturation program in *Arabidopsis* (Kwong *et al.*, 2003). In *M. truncatula*, this *HAP3/MtL1L* gene was highly expressed at S1 in embryo and S1-S2 in endosperm, later than *MtLEC1*. It is one of the most statistically differentially expressed genes following heat stress, which made us consider this gene as candidate gene for further functional validation.

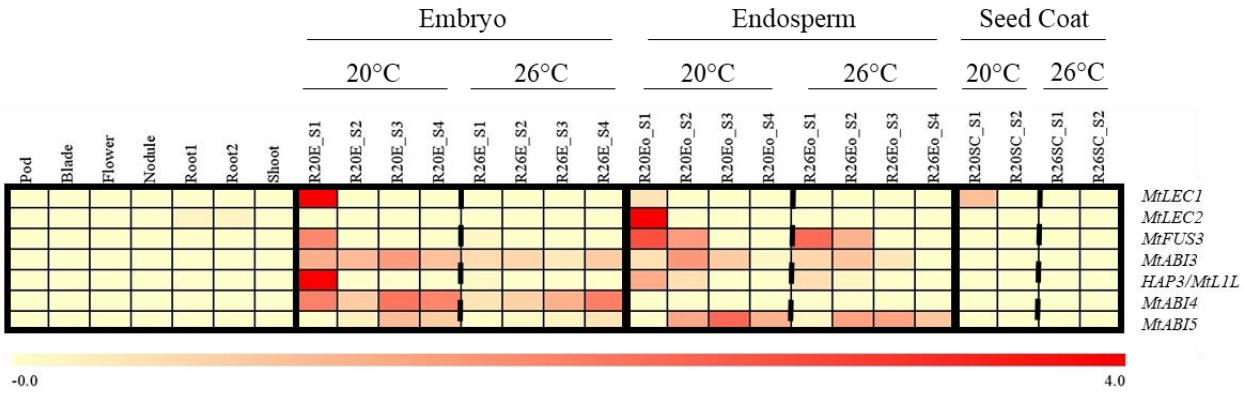


Figure 2.4. Expression profiles of major known regulatory genes in different *Medicago truncatula* organs and seed tissues during seed development. Seeds tissues (E: Embryo; Eo: Endosperm; SC: Seed Coat) were harvested during seed development (S1-4: stage 1 to 4) under control (20°C) and heat stress (26°C) conditions. Expression values are expressed in TPM (transcript per million) values after a z-score normalization.

Table 1. Summary of time-course (ImpulseDE2) and pairwise (DEseq2) comparisons between control and stress conditions of major known regulatory genes (with adjusted p-values below 5%) in isolated *Medicago truncatula* seed tissues during seed development.

Gene annotation				ImpulseDE2		DEseq2									
Name	Gene ID V5	Gene ID V4	SS	DEG_E	DEG_Eo	E_S1	E_S2	E_S3	E_S4	Eo_S1	Eo_S2	Eo_S3	Eo_S4	SC_S1	SC_S2
MtLEC1	MtrunA17_Chr1g0165041	Medtr1g039040	Yes	No	No	-4.74				-5.49				-3.98	2.28
MtLEC2	MtrunA17_Chr4g0047031	Medtr4g088605	Yes	No	No					-7.84					3.33
MtFUS3	MtrunA17_Chr7g0251881	NA	Yes	Yes	Yes	-0.88	-0.64	-0.47	2.32	-0.08	-0.45	-0.02	1.61	-1.77	-0.37
MtABI3	MtrunA17_Chr7g0237841	Medtr7g059330	Yes	Yes	Yes	0.30	-0.32	-0.12	-0.31	0.38	-0.28	-0.13	-0.23	-0.73	-1.02
HAP3/MtL1L	MtrunA17_Chr4g0076381	Medtr4g133952	Yes	Yes	Yes	-2.51	-0.35	-0.48	0.91	-0.69	-0.46	-0.70	1.89	-3.38	-1.67
MtABI4	MtrunA17_Chr5g0437371	Medtr5g082950	Yes	Yes	Yes	-0.47	-0.09	0.13	-0.13	1.33	3.16	3.27	-0.33	-1.92	-0.57
MtABI5	MtrunA17_Chr7g0266211	Medtr7g104480	No	Yes	Yes	1.73	-0.12	-0.08	-0.52	2.44	0.21	-0.23	-0.26	1.14	-0.88

Note: SS, Seed Specificity; E, Embryo; Eo, Endosperm; SC, Seed Coat. The genes that are down-regulated or up-regulated in heat stress condition at each stage are indicated by green or red. Values indicate Log2 (fold change).

2.3 Focus on embryo: Global transcriptome and epigenome changes in response to heat stress in *M. truncatula* embryo

To characterize the molecular impacts of heat stress during seed development and identify key regulator genes leading to phenotypic changes, we decided in the subsequent analyses to focus on the embryo.

2.3.1 Description of molecular processes impacted by heat stress in the embryo

In order to focus on the effect of heat stress in the embryo molecular mechanisms, we re-analysed our transcriptomic data with DESeq2 to perform pair-wise comparisons at the four embryo stages aiming to identify DEG between seeds developed under control and heat stress. Using an adjusted p-value threshold of 5%, we identified and annotated 11,207 DEGs due to heat stress with 5907 at S1, 968 at S2, 1717 at S3 and 2615 at S4. As described earlier, we performed gene set enrichment analyses to visualize the major molecular processes affected during the stress at different developmental stages in embryo (Figure 2.5). As observed, for the whole seed, stage S1 was the most impacted stage with 2,637 up-regulated and 3270 down-regulated genes. Down-regulated genes showed enrichment in functional classes related to 'DNA synthesis.chromatin structure', suggesting an important chromatin dynamics at this transition stage between the end of embryogenesis and beginning of maturation program. This transition stage is also characterized by enrichments of 'cell division' and 'cell cycle' classes, as well as many classes related to protein targeting/synthesis/degradation, suggesting a decrease of cell division and a cellular reprogramming of the molecular processes illustrated by these essential cellular activities. On the other side, at S1, upregulated genes mainly belong to metabolomic changes with enrichment of classes related to secondary metabolisms, storage proteins, starch and lipids. At S2, we did not observe many DEG due to heat stress, and, therefore, few functional classes were enriched except 'PS.photosystem II' in up-regulated genes and 'storage.proteins', 'abiotic.stress' classes. At S3, we still observed enrichment of classes related to photosynthesis and photosystems I&II, as well as cytokinin metabolism from up-regulated genes. Down-regulated genes were mainly associated with classes related to regulation of transcription, with many transcriptional regulators and more specifically auxin response factors in the list of DEGs. Finally, at seed maturity (S4), which corresponds to the

final stage of seed development, we identified 'DNA synthesis.chromatin structure' class and many RNA related classes such as 'RNA.Chromatin remodeling factors', or 'RNA.helicase', 'RNA transcription' and 'RNA binding', which suggested that at the final stage of seed development, many important regulatory processes are occurring regarding chromatin and transcriptional dynamics.

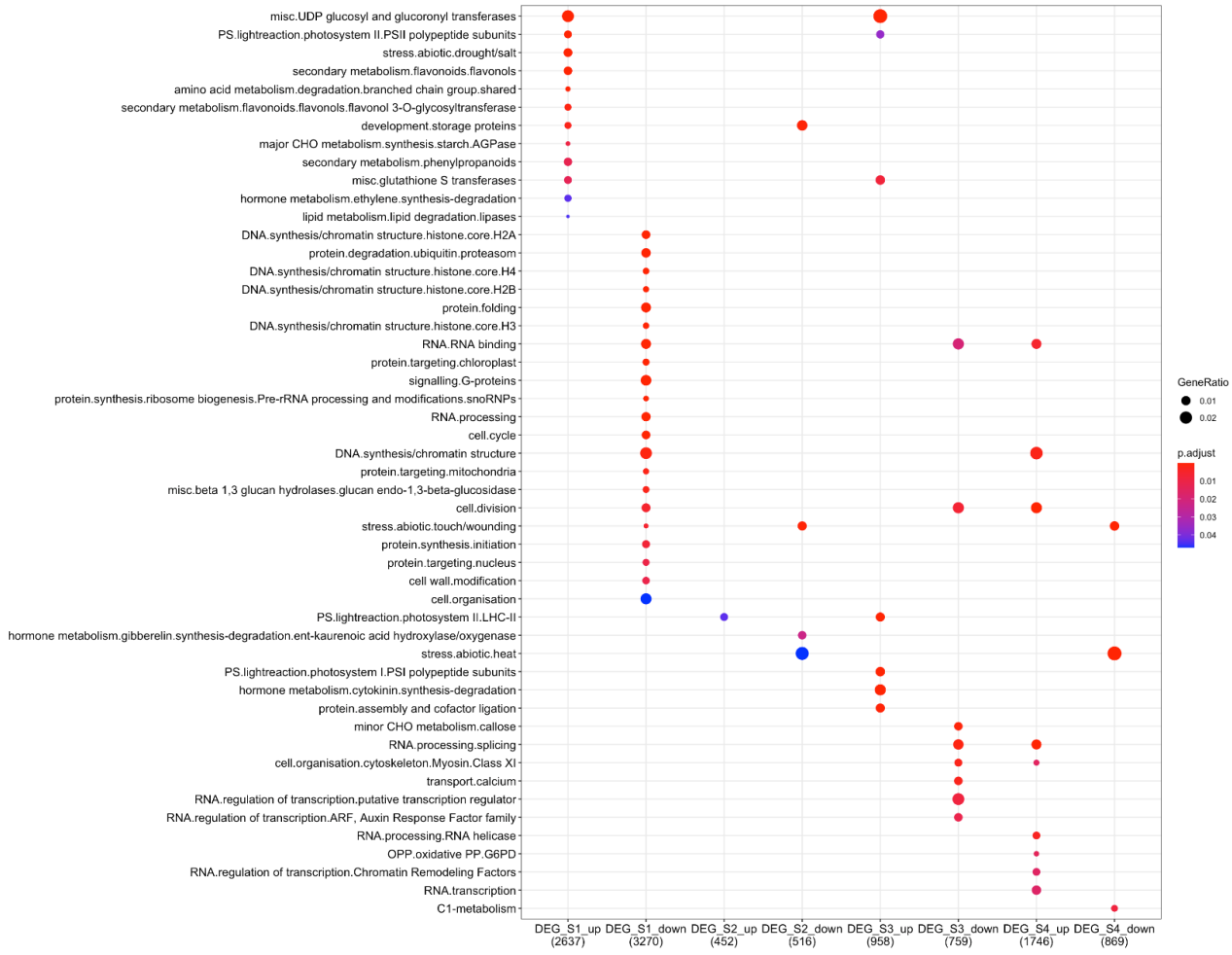


Figure 2.5. Over-representation of functional classes of Up- and Down-regulated genes in the embryo (E), due to heat stress at the four seed developmental stages (S1, S2, S3 and S4). Enriched terms were identified using ClusterProfiler with enricher function and a Bonferroni adjusted p-value threshold of 0.05. Numbers that are indicated at the bottom are the gene numbers used to perform enrichment analyses.

2.3.2 Identification of candidate genes associated with acquisitions of seed weight and longevity during heat stress in embryo

2.3.2.1. Weighted Gene Correlation Network analysis and gene module identification

To identify candidate genes that explain the change of physiological characteristics of seeds grown in heat stress conditions, we decided to perform a weighted gene correlation network analysis (WGCNA, Langfelder and Horvath, 2008). WGCNA is a data exploratory method, which integrates transcriptomic and physiological data to highlight candidate genes correlated between these two datasets. In order to relate transcriptomic and physiological data, we analyzed the acquisition of seed weight and P50 (longevity) (Figure 2.2A) at the same four developmental stages (S1 to S4) as the ones used for transcriptomic analyses. Regarding the transcriptomic data, we selected the 9,182 genes showing significant statistical changes of expression between embryo developed at 20°C and 26°C, identified from time-course comparison using ImpulseDE2 algorithm (Chen *et al.*, 2021b). By running WGCNA algorithm, we plan to assess pair-wise correlations between gene expression profiles in order to take into account their relationship, then determine relevant gene modules associated to the studied traits and finally calculate intermodular-connectivity and gene significance within the relevant module(s) to provide lists of candidate genes potentially regulating these physiological characteristics.

Before the construction of the co-expression network, we normalized the 9,182 transcripts showing differential expressions in embryo using Z-score and we combined them with the physiological traits related to the seed weight kinetic and the acquisition of seed longevity obtained at the same stages that were used in the transcriptome analysis in both conditions at 20°C and 26°C (Figure 2.6A). In this analysis, only acquisition of physiological traits characterized at the same stages that transcriptomic data could be used so we decided to focus on seed weight and seed longevity that were analyzed during seed. Due to the absence of transcriptomic data during seed germination, we did not include this physiological trait in this specific analysis. Then, to construct the weighted gene co-expression network, we determined the soft-threshold power (β) at 34 to satisfy a scale-free topology of the network with an R-squared of 0.9 (Figure 2.6B). This soft threshold power was used to calculate the adjacency

(*i.e.* defined as $ADJ_{ij} = abs(cor(x_i,x_j))^{\beta}$) of the unsigned network and identify relevant gene modules (*i.e.* clusters of highly interconnected genes). Then, using topological overlap matrix (TOM), we clustered genes into modules, which were eventually merged when highly correlated (*i.e.* merged if above 80% correlations, Figure 2.6C-D). Finally, we obtained 11 gene modules, which were determined by their module eigengenes (*i.e.* representative of the expression profiles of all genes for each module). We, then, calculated the eigengene significances (*i.e.* the correlation between physiological traits and module eigengenes) and their corresponding p-values for each module to obtain the module-trait relationship (Figure 2.6E). From this figure, we observed that two modules displayed highly significant correlations with our observed seed traits: a module, called coral3, showing a correlation of 0.84 (p-value of 3.10^{-7}) with the acquisition of longevity (P50) and to less extent a module, called palevioletred3, showing a correlation of 0.67 (p-value of 3.10^{-4}) with the increase of seed weight (DW) (Figure 2.6E).

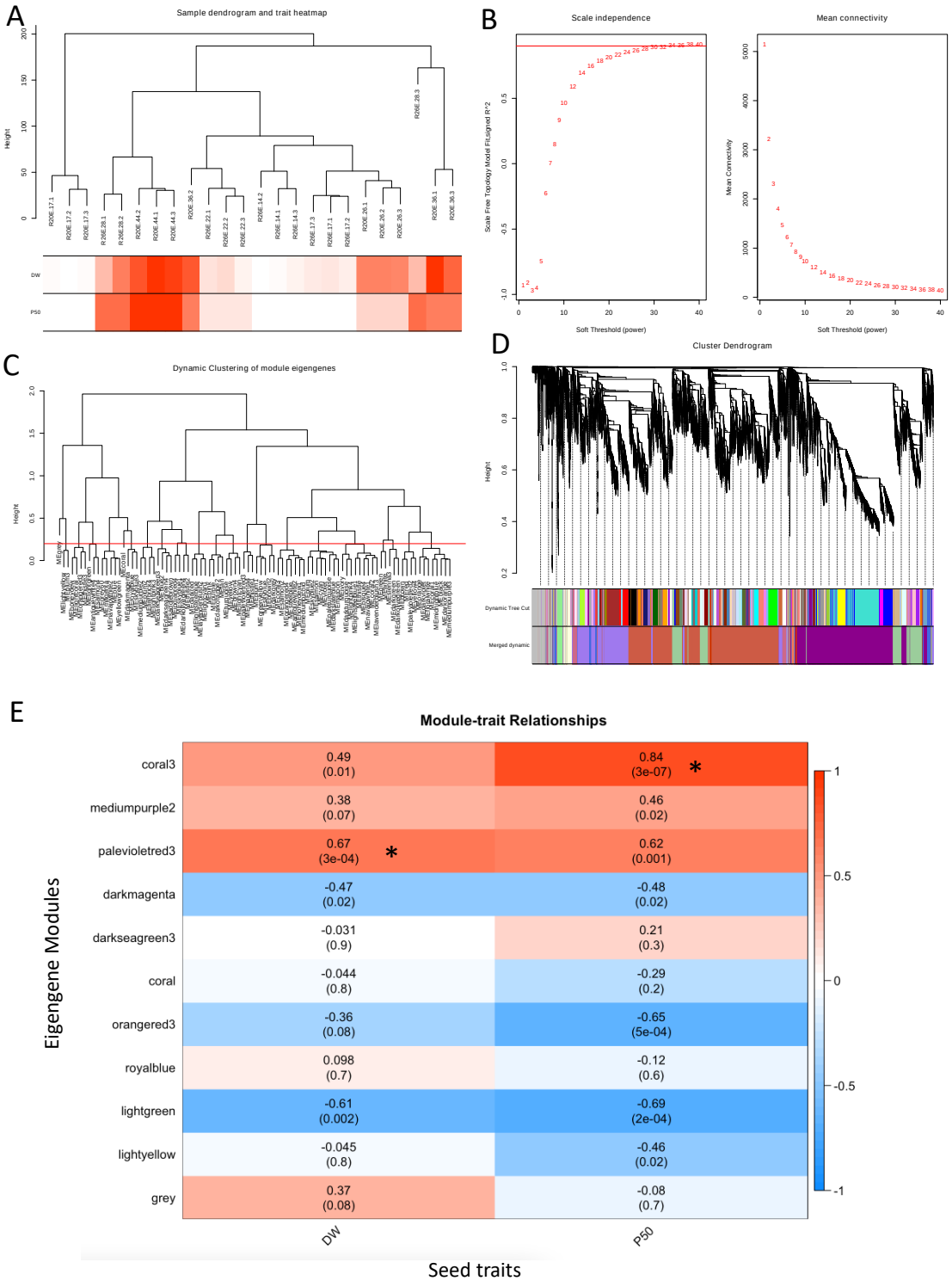


Figure 2.6. Weighted Gene Correlation Network Analysis (WGCNA). Different steps for WGCNA module identification. (**A)** Sample dendrogram and trait heatmap. (**B)** Scale dependence and mean connectivity to determine the soft threshold. (**C)** Clustering of different module eigengenes with the cutHeight threshold of 0.2. (**D)** Cluster dendrogram with modules merged when correlation above 80%. (**E)** Correlation of the identified modules with the two seed traits, Longevity (P50) and seed weight (DW). Modules significantly associated with the traits are indicated by asterisks (*) and correlation coefficients and p-values are indicated. Red and blue color

represents positive and negative correlations between traits and representative eigengene expression for each module.

2.3.2.2. Candidate genes involved in change in the acquisition of seed longevity during heat stress

By dissecting the 'Coral3' module, which showed a strong correlation with the acquisition of seed longevity trait, we identified a list of 3,675 genes, making this module one of the largest. To reduce and refine the final list of candidate genes within this module, we calculated two parameters for each gene: the module membership (or eigengene-based connectivity, MM) and the gene significance (GS). Module membership is defined by how correlated each gene is to its corresponding module eigengenes. In other words, if MM is close to 0, the gene is not part of the module and if MM is close to 1, it is highly connected to the module genes. MM reveals the intramodular connectivity of genes and genes having high module membership represent intramodular hub genes (*i.e.* highly connected) in the respective module. The Gene significance is a gene parameter related to the incorporation of physiological observations and reflects the biological significance of genes (*i.e.* genes with GS equal to zero are not relevant to the physiological traits, at the opposite high GS tend to highlight genes with potential biological relevance to the traits).

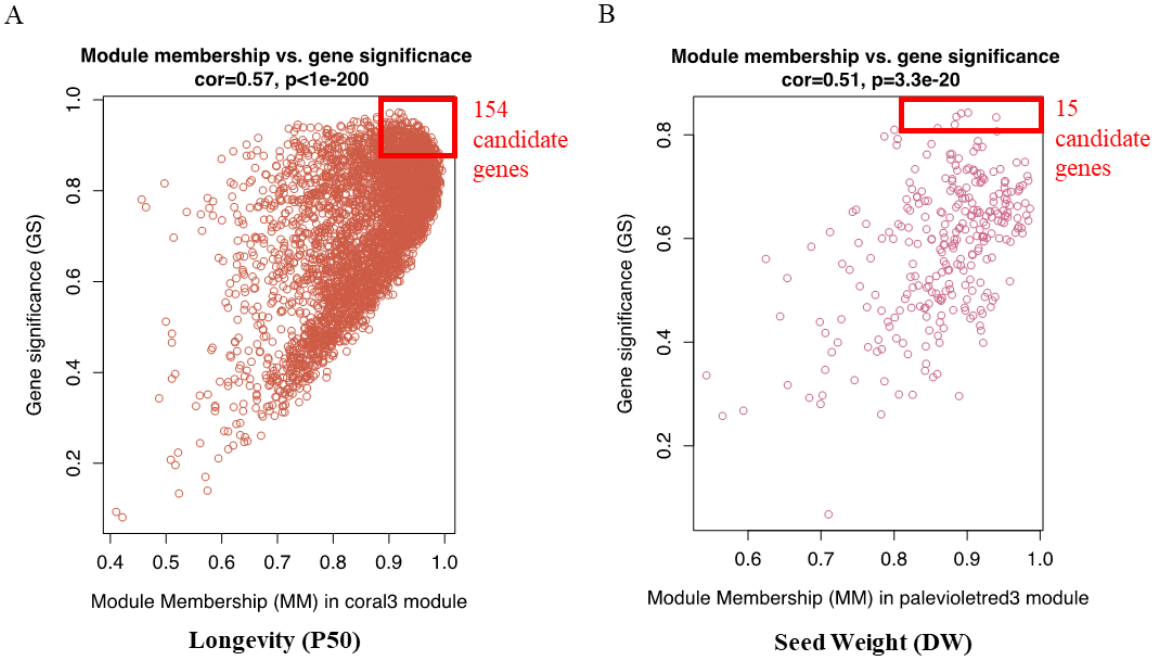


Figure 2.7. Analysis of modules correlated to longevity (coral3 module) and seed weight (palevioletred3 module). (**A)** Relationship between module membership of 'coral3' and gene significance for longevity (P50). (**B)** Relationship between module membership of 'palevioletred3' and gene significance for seed dry weight (DW). Red frames correspond to interesting candidate genes showing high degree of connection (MM) and high correlation with physiological traits (GS).

Figure 2.7A represents the 'Gene Significance versus the module membership' of genes belonging to the coral3 module correlated to the acquisition of seed longevity. To identify key regulators of seed longevity traits during heat stress conditions, we extracted genes showing high MM (>0.9) and high GS (>0.9) from this module, which refined the gene list to 154 candidate genes potentially important to regulate the longevity during heat stress conditions (Supplementary Table S2). Out of this list and based on gene annotations, we could clearly identify interesting candidate genes potentially modulating seed longevity during heat stress such as many annotated genes involved in cell redox regulation encoding thioredoxins (MtrunA17_Chr5g0403061), glutathione-S-transferases (MtrunA17_Chr3g0142901 and MtrunA17_Chr1g0181861) and glutathione peroxidase (MtrunA17_Chr1g0150061) (Bailly *et al.*, 1996). Interestingly, we also have two candidate genes, which have been demonstrated to have a role in modulating seed longevity: a Galactinol synthase (MtrunA17_Chr7g0271581) and an L-isoaspartyl methyltransferase2 (*PIMT2*, MtrunA17_Chr3g0116621) genes. The galactinol has been showed to be a biomarker of seed longevity in tomato, Brassicaceae and chickpea, and the galactinol synthase is a key enzyme of the Raffinose Family Oligosaccharide (RFO) pathway, which control accumulation of galactinol and the subsequent synthesis of RFOs (de Souza Vidigal *et al.*, 2016; Salvi *et al.*, 2016). This enzyme was already co-localized with a QTL of seed longevity in tomato and could play a role in modulating longevity in *Medicago* as well. Similar situation regarding PIMT2 protein that catalyzes the conversion of abnormal to normal L-isoaspartyl residues from proteins. In this way, the PIMT repair system serves as a marker for seed ageing with a clearly documented role in longevity in *Arabidopsis* and Chickpea by repairing damaged proteins (Ogé *et al.*, 2008; Verma *et al.*, 2013). Many other genes are present in this list correlated to the acquisition of seed longevity and could represent good candidates to understand the change in seed longevity following heat stress.

2.3.2.3. Candidate genes involved in seed weight change during heat stress

A similar approach than described previously for longevity was performed to refine the list of candidate genes from the palevioletred3 module, which was the most correlated to seed weight. This module, initially, contained 284 genes but by extracting genes showing high MM (>0.8) and high GS (>0.8) (Figure 2.7B), we refined this list to 15 candidate genes regulating the change in seed weight during heat stress conditions (Supplementary Table S2). Out of these 15 candidate genes, we could identify two potential good candidates: *MtABI5* and a *bHLH18*. These two genes were identified as the most statistically differentially expressed genes following heat stress (*i.e.* in the top100 of lowest adjusted p-values from ImpulseDE2). The bHLH gene has not been documented but appears to be seed-specific according to our *Medicago* plant organ transcriptomic data (Chen *et al.*, 2021a) and *MtABI5* has already been documented to its role in seed maturation, dormancy and longevity in *Medicago* (Finkelstein and Lynch, 2000; Zinsmeister *et al.*, 2016). *ABI5* is seen as an integrator of ABA and other phytohormone signaling (*i.e.* IAA, CK, JA, BR) during abiotic stress (Skubacz *et al.*, 2016). Its role during seed development is mainly related to seed vigor (*i.e.* dormancy and longevity) but a recent study identified a relationship between *ABI5* and *TERMINAL FLOWER1* (*TFL1*), which acts as a regulator of endosperm cellularization determining final seed size (Zhang *et al.*, 2020). These two previous genes could represent good candidates for functional validation regarding their role in modulating seed size in heat-stressed seeds.

2.3.3 Methylation dynamics during embryo development under heat stress conditions

Our transcriptomic data revealed a strong impact of heat stress on twelve genes annotated as DNA (cytosine-5)-methyltransferases and DNA demethylases showing differential expressions (Supplementary Table S1). Therefore, we analyzed the DNA methylome dynamics following heat stress at the same S1, S2, S3 and S4 developmental stages in the embryo tissues using whole-genome bisulfite sequencing (WGBS). To characterize the genome-wide DNA methylation state and understand the precise role of DNA methylation on seed response to heat stress, we isolated DNA from the four embryo developmental stages produced in control and heat stress conditions. Treatment of DNA with bisulfite converted non-methylated cytosine residues to uracil, but conserved 5-methylcytosine residues unchanged. Bisulfite treated DNA was then sent for sequencing allowing us to identify methylation levels of each cytosine present in the genome. About 60 million high-quality paired-end reads were generated for each sample and mapped uniquely to the *Medicago* genome version 5. Mapping,

deduplication and methylation calling steps were performed using Bismark software (Krueger and Andrews, 2011). First, we observed that the percentage of methylcytosines identified in each context in control conditions did not drastically change between developmental stages S2, S3 and S4, except an increase between S1 and latter stages, as observed in *Arabidopsis* (Bouyer *et al.*, 2017). Global methylation ratios were also similar, even slightly lower, to those observed in *Arabidopsis* embryo with about 78% of methylated CG sites, 50% of methylated CHG sites and 15% of methylated CHH sites, with respect to about 85%, 55% and 25% respectively in *Arabidopsis* embryo. All these common observations validated our analysis pipeline. Then, regarding the impact of heat stress on the global methylation dynamics, we observed similar percentages of global methylation in different contexts in both control and heat stress conditions, with the exception of S1, which displayed a higher content of methylated CHG and CHH sites in embryos produced in heat stress conditions (Figure 2.8).

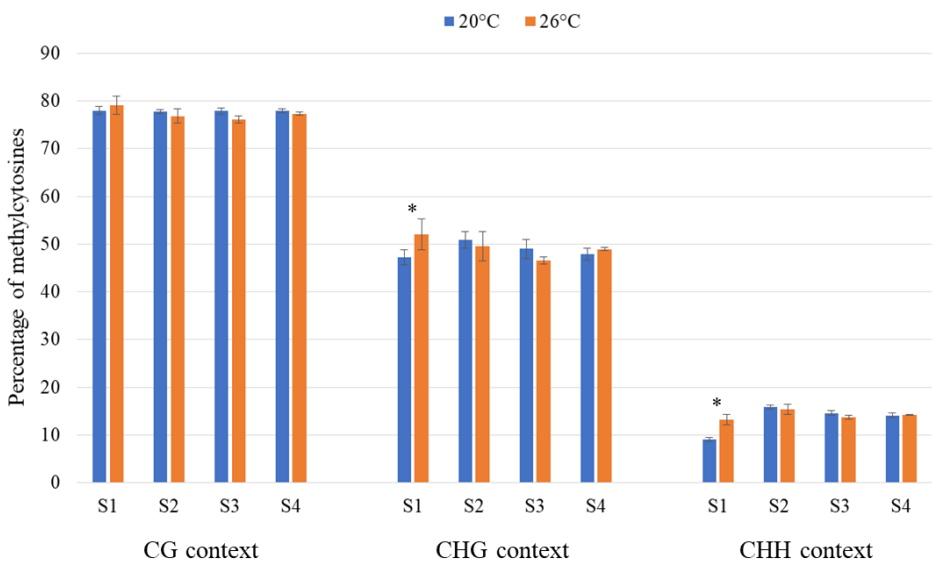


Figure 2.8. Global DNA methylation dynamics during seed development from bisulfite sequencing experiments. Percentages of cytosines that were identified as methylated in different sequence contexts (CG, CHG, and CHH), at the four different stages of seed development and under control and heat stress conditions. *, 0.01<p-value<0.05.

To get insight into the specific methylated regions following heat stress, we used DMRCaller (Catoni *et al.*, 2018) and an adjusted p-value threshold of 5% to identify specific differentially methylated regions (DMR) due to heat stress in embryos. As a result, we identified from 1 to 2175 DMRs depending on embryo developmental stages and methylation

contexts with either hypo-methylated regions (*i.e.* decrease of DNA methylation due to heat stress) or hyper-methylated regions (*i.e.* an increase of DNA methylation due to heat stress) (Figure 2.9) (Supplementary Table S3).

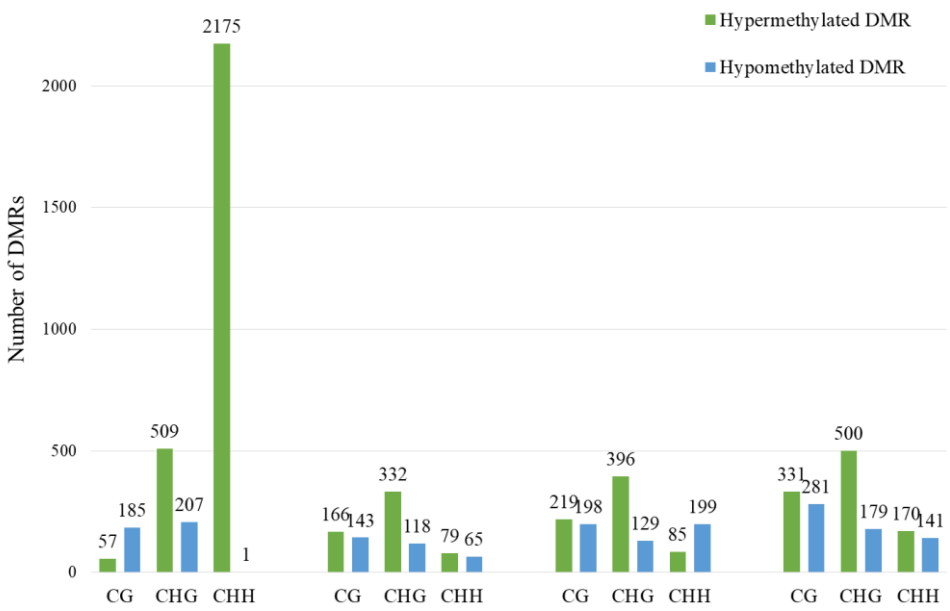


Figure 2.9. Identification of Differentially Methylated Regions (DMR) between heat stress (26°C) versus control (20°C) conditions in different sequence contexts and during seed development using DMRCaller. Numbers of hypo- or hyper-methylated DMR are indicated.

Interestingly, we observed an usual proportion of hypermethylated regions in the CHH context at S1 with 2175 DMRs, with respect to other stages and contexts, suggesting an intense DNA methylation dynamic at S1 following heat stress. To understand the role of DNA methylation changes during heat stress, we mapped the DMRs into genomic regions to distinguish DMR located in 1kb promoters, gene coding sequences and transposable elements (TE) (Figure 2.10). We observed similar DMR distributions in genomic regions at S2, S3 and S4 with 40-45% of DMR located in TEs, 17-19% in 1kb promoters, 18-21% in coding sequences and 18-24% in other genomic regions. In contrast at S1, we observed an increase of DMR located in TEs (60%) and in 1kb promoters (23%), suggesting first that the increase of DMR occurring during heat stress at S1 could be due to an increase of methylated regions in TEs and 1kb promoters, and second that this DNA methylation could play an important role in heat stress response at S1.

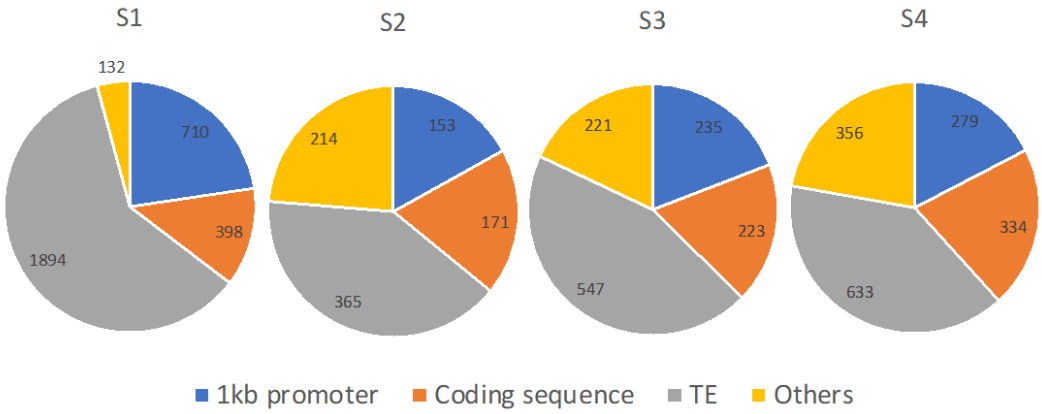


Figure 2.10. Distribution of DMRs over genomic regions. Numbers of DMRs located in 1kb promoters, coding regions, transposable elements (TE) and others are indicated at different developmental stages in embryo.

To get a better understanding of the role of differential DNA methylation at S1 and its potential role on transcriptional regulation, we combined data from our methylomes and transcriptomes by selecting hypo-methylated regions located in 1kb promoters with corresponding up-regulated genes and hyper-methylated regions located in 1kb promoters with corresponding down-regulated genes in heat stress conditions. As a result, we identified 97 genes that could potentially be transcriptionally regulated via DNA methylation in embryo at stage S1 following heat stress (Supplementary Table S3). Out of these 97 genes, and based on their tentative annotations, we could distinguish four main molecular functions potentially regulated by DNA methylation with 13 genes annotated as stress response genes such as *MtLEA-D34* (MtrunA17_Chr3g0090511), *MtHSP70* (MtrunA17_Chr5g0394391) and *MtERD4* (MtrunA17_Chr3g0084131), which displayed a seed-specific expressions; 8 genes annotated as signaling genes including four G-proteins involved in transducing signals from cell receptors to intracellular effectors; 15 genes annotated as involved in protein degradation/synthesis/modification including four E3 ubiquitin protein ligase RING types, already described earlier as involved in transferring ubiquitin either to histone H2A (via the PRC1 complex) or directly to protein to direct to the ubiquitin proteasome system; and finally 12 genes annotated as regulation of transcription including four zinc finger transcription factors and one GRAS TF (MtrunA17_Chr4g0057291), ortholog of Arabidopsis *SCARECROW-LIKE 28*, which is expressed in embryo and endosperm and recently characterize as a regulator of the G2/M phase of the mitotic cell cycle (Goldy *et al.*, 2021). Interestingly, it has to be noted that this list of 97 genes contains a DNA (cytosine-5)-methyltransferase (MtrunA17_Chr4g0007641), ortholog to *CMT3* in *Arabidopsis*, which displayed a down-

regulation of its expression and a hypermethylation of its promoter sequence at S1 during heat stress. In *Arabidopsis*, *CMT3* has been shown to be involved in DNA methylation of non-CG sites and preferentially in transposon related sequences (Bartee *et al.*, 2001; Lindroth *et al.*, 2001), which suggested that this DNA methyltransferase that could play a role in the decrease of the CHH-methylated regions mainly occurring in the TEs between Stage S1 and S2 in *Medicago*, could also be transcriptionally regulated by methylation of its promoter during heat stress (Figure 2.9 and 2.10).

Regarding DNA methylation dynamics, we also focused our analysis on stage S4, corresponding to mature seeds, which define the final developmental stage preceding the germination. In order to determine if DNA methylation dynamic occurring in the 1kb promoters in mature seeds (stage 4) could impact the expression of genes involved in germination of seeds produced under heat stress, we also combined transcriptome and methylome data. We, first, identified DEGs during the *Medicago* germination process by analyzing transcriptomes of three germination timepoints: mature seeds, 2mm radicle and 4 mm radicle to reveal that 20,272 genes showed differential expression during this germination time course (out of a total of 50,165 in *Medicago*) (unpublished data). Then, we combined the expression of these DEG during germination with the 279 genes displaying DMR in their 1kb promoters following heat stress and identified 99 genes in common. In conclusion, these 99 genes are differentially expressed during germination and displayed a DMR in their promoter regions following heat stress, which represents an interesting candidate gene list of genes potentially necessary for proper germination that could be transcriptionally regulated by DNA methylation, and thus involved in the delay of germination of seeds produced under heat stress.

Finally, the *LAFL* genes, key regulators of seed development, are not located in densely methylated regions, in consequence, we did not observe any major change in the methylation levels surrounding their genomic locations in control and stress conditions, which indicates that they may not be developmentally regulated by DNA methylation, such as suggested in *Arabidopsis* (Lin *et al.*, 2017), but either stress-regulated by DNA methylation.

2.3.4 Dynamics of the H3K27me3 marks during embryo development under heat stress conditions

Due to the importance of the PRC2 complexes during seed development and the impact of *HSFA2* on the regulation of H3K27me3 marks (Ohama *et al.*, 2017; Liu *et al.*, 2019a), we analyzed the H3K27me3 dynamics following heat stress at S1, S2, S3 and S4 developmental stages in the embryo tissues using Chromatin Immunoprecipitation Sequencing (ChIP-seq). DNA was cross-linked to proteins then nuclei were isolated from the embryos at different developmental stages o from both control and heat stress conditions. Histones H3 containing a trimethylation at lysine 27 (*i.e.* H3K27me3 marks) were immunoprecipitated using a specific antibody. During this immunoprecipitation, we also obtained enrichment in genomic sequences attached to H3K27me3, which are subsequently sequenced. A minimum of 80 million high-quality read pairs were generated for each IP sample (*i.e.* immunoprecipitated sample) and INPUT sample (*i.e.* control sample without the immunoprecipitation), then mapped to Medicago genome version 5 using the STAR mapper (Dobin *et al.*, 2013). Deduplication was done using Picard tool and peak calling and differential binding analysis from INPUT and IP samples were performed using MACS2 (Zhang *et al.*, 2008). It has to be noted that as a first approach and due to the low replicate number, we performed a qualitative analysis of H3K27me3 marks showing the presence/absence of these marks, but does not reflect their relative abundance.

Regarding the H3K27me3 dynamics occurring during embryo development from seeds produced during control and heat stress conditions (Figure 2.11A), we identified many genomic binding regions at S1 in both control and heat stressed seeds with more than 4,700 and 5,600 binding sites respectively. This result suggested that at stage 1, H3K27me3 repressive marks are highly present in the genome but may not be strongly regulated by heat stress. On the opposite, at stages S2, S3 and S4, we observed an important decrease of H3K27me3 binding regions in control seeds, which was constant during seed development with around 1,800 binding genomic sites at S2 then 1,500 at S3 and 900 at S4. Interestingly, we observed that in heat-stressed seeds, the number of genomic binding sites were still very high (*i.e.* more than 4,500 binding sites), suggesting that at these stages H3K27me3 marks were stress dependent and that they could play an important role in developmental regulation of heat-stressed seeds.

To get an overview of the role of these repressive marks during embryo development, we selected H3K27me3 binding sites that were located around coding sequences (*i.e.* located in 1kb promoter regions and gene sequences). Then, we differentially analyzed these gene binding sites in three categories: gene binding sites common to control and heat-stressed seeds (called common) and gene binding sites specific to control seeds (called spec20) and specific to heat stressed seeds (called spec26). Finally, we performed a gene set enrichment analysis of these gene binding sites regarding their potential functions (Figure 2.11B). We did not observed enrichment of many functional classes but at S1, S3 and S4, the common gene binding sites repressed by the H3K27me3 marks were mainly related to two classes: photosynthesis and mitochondrial electron transport/ATP synthesis. Regarding gene binding sites specific to heat stressed seeds, we observed enrichments of AP2/EREBP at S1 and NIN-like bZIP related transcription factors at S2, that could have pleiotropic effect on embryo and seed development during heat stress. This analysis did not clearly answer the question about the role of H3K27me3 marks during embryo development under heat stress and need to be analyzed in more details.

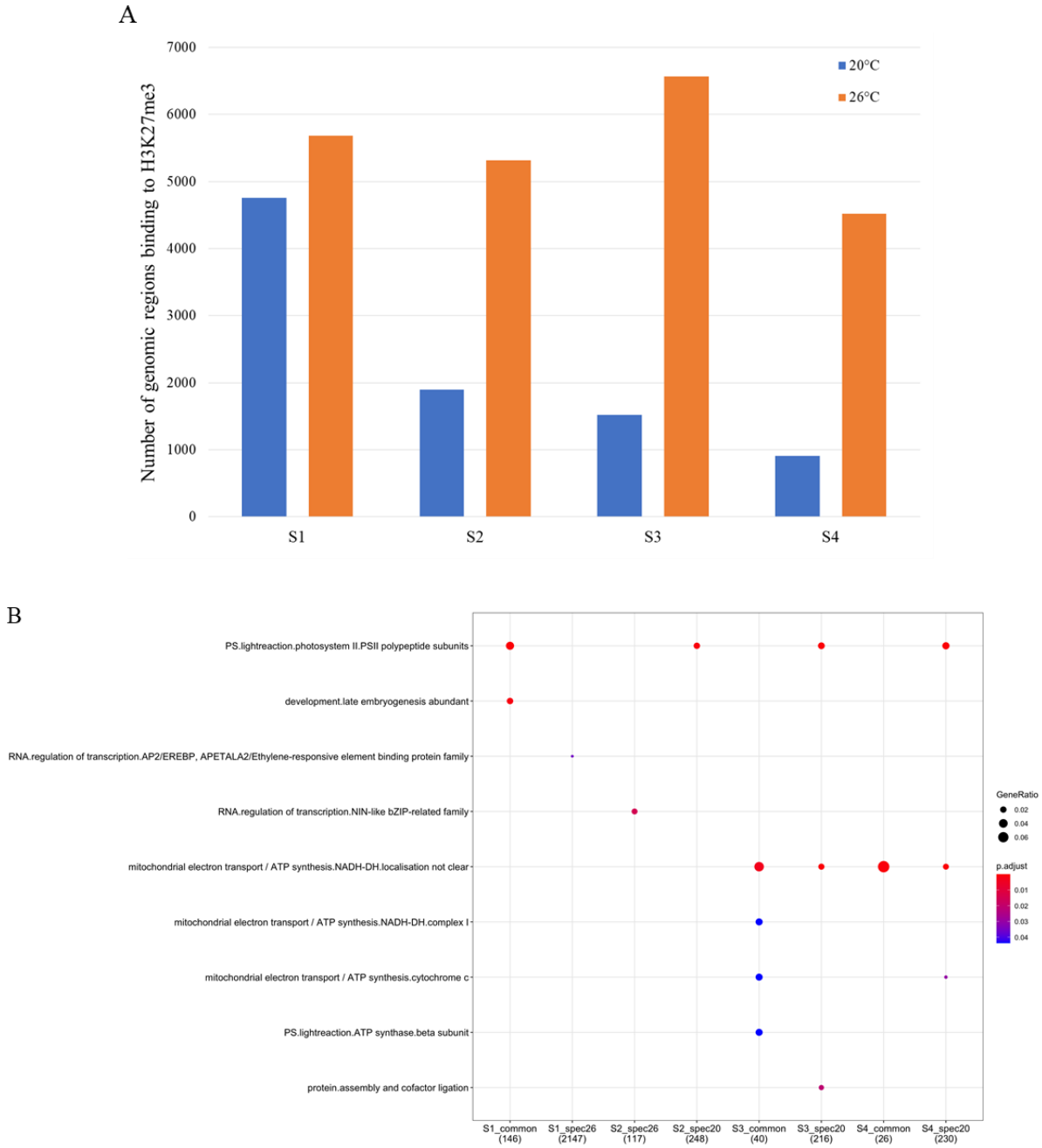


Figure 2.11. **(A)** Global H3K27me3 dynamics with numbers of total genomic binding regions to H3K27me3 at the four seed developmental stages and occurring under control and heat stress conditions. **(B)** Over-representation of functional classes of H3K27me3 genomic binding sites located in 1kb-promoter and coding sequence regions, which are common in control and stressed seeds (_common), specific to control (_spec20) and specific to heat stress (_spec26) conditions at the four seed developmental stages. Enriched terms were identified using ClusterProfiler with enricher function and a Bonferroni adjusted p-value threshold of 0.05. Numbers that are indicated at the bottom are the gene numbers used to perform enrichment analyses.

2.4 Preliminary results obtained from the functional validation of the candidate *HAP3 (MtL1L)* gene

As previously described above, the analyses of transcriptome and epigenome changes of embryo under heat stress allowed us to identify many candidate genes that could be involved in regulating seed development and explain the phenotypic impacts of heat stress. A short list of candidate genes has been selected to elucidate their specific role in relation to heat stress, including some already documented genes important for seed development such as *MtDASH* (MtrunA17_Chr2g0282441, Noguero *et al.*, 2015) and *MtABI5* (MtrunA17_Chr7g0266211, Zinsmeister *et al.*, 2016). Moreover, some candidate genes with unknown functions were also selected such as a seed-specific *bHLH* gene (MtrunA17_Chr2g0333541) and a *HAP3* gene (MtrunA17_Chr4g0076381) closely related to *AtLIL*. Due to the time-consuming process of growing *Medicago* plants to obtain homozygote mutant plants, we only obtained preliminary results regarding functional characterization of the *MtHAP3 (MtL1L)* candidate gene.

As described previously, we identified a candidate gene, encoding a HAP3 protein (MtrunA17_Chr4g0076381*)*, which displayed a high differential expression in embryo, endosperm and seed coat in seeds under heat stress (Figure 2.12A). This gene was also highlighted to have a seed specific expression in *Medicago* with a transiently expression between embryogenesis and maturation (at S1). This gene annotated as putative transcription factor HAP3/NF-YB family, and showed closely related sequence to Arabidopsis *NF-YB6* (AT5G47670) encoding LEC1-LIKE (L1L), which was shown to function as a regulator of embryo development and initiation of seed maturation (Kwong *et al.*, 2003). This gene family encodes interesting proteins that have a very conserved histone fold domain, which functions in protein-DNA and protein-protein interactions (Gnesutta *et al.*, 2017), and are also called CCAAT Binding Factor (CBF) or Heme Activator Protein (HAP). To be functional, NF-YB (CBF-A/HAP3) needs to be associated in a heterotrimer complex, with a NF-YA (CBF-B/HAP2) and a NF-YC (CBF-C/HAP5) subunits to bind the CCAAT box in the promoter regions of target genes. Plants have large NF-Y complexity with many genes encoding each subunit, which provides many combinations of the NF-Y complexes, allowing specific binding specificity in different developmental stages or under certain conditions. In *Medicago*, *MtL1L* was the only NF-YB gene differentially expressed in heat stress conditions and showing a seed specific expression. Moreover, due to its histone-like protein structure, and as demonstrated

for *AtLEC1* (Tao *et al.*, 2017), it could act as pioneer gene to differentially regulate seed development upon heat stress.

From the database of the Medicago *Tnt1*-insertional mutant population (D'Erfurth *et al.*, 2003; Tadege *et al.*, 2008) we identified two independent mutant lines of *MtL1L*: NF7602 with the *Tnt1* insertion within the intron and NF4536 within the second exon sequence (Figure 2.12B). From the heterozygous lines obtained from the Noble research Institute, we obtained homozygous lines at their second generations with their corresponding sibling wild-types (*i.e.* lines containing similar insertional background but without *Tnt1* insertion within the candidate gene sequence). Mutant and the sibling wild-type (WT) plants were grown in control conditions, then after apparition of five flowers, half of them were grown under heat stress conditions to analyze the impact of *l1l* mutation on heat stressed seeds. Mature seeds were harvested from control and heat stress conditions at the pod abscission stage and then dried at 44% relative humidity and 20°C. Seeds were stored at room temperature for one month before using for analyses.

Phenotyping of mutant seeds grown in control conditions did not display any statistically difference regarding to seed size compared to the sibling WT seeds (Figure 2.12C-D). This result is important for the subsequent analysis because even if MtL1L shared high sequence homology with AtL1L, their roles appear to be different. Indeed, in *Arabidopsis AtL1L* is crucial for embryo development and seed maturation, which was highlighted by the drastic phenotypes of *l1l* mutants displaying defective embryos (Kwong *et al.*, 2003). In *Medicago*, *MtL1L* does not appear to be necessary for proper embryo development when seeds are produced in normal conditions.

To continue with the functional characterization of *l1l* mutant in *Medicago*, we characterized the phenotype of mutant and WT seeds produced under heat stress. The seed weight of one of the *l1l* mutant lines (NF4536) displayed a significant reduction compared to the seeds of the sibling WT (Figure 2.12E). Regarding the second *l1l* mutant line (NF7602) with the insertion in the intronic sequence, we did not observe a significant difference in seed weight compared to the sibling WT (Figure 2.12F). This preliminary result based on limited number of seeds needs to be confirmed using larger seed lots, which are currently under production. Regarding seed longevity, we preliminary characterized NF7602 mutant line and corresponding WT using an artificial ageing experiment. Triplicates of 30 mature seeds produced under heat stress conditions were stored at 75% relative humidity and 35°C for 34 days. After this period of

accelerated ageing, we observed that the germination rate of *l1l* mutant (NF7602) seeds was higher than the one from wild-type seeds, with 40% germination of *l1l* mutants compared to less than 10% in the sibling WT seeds, indicating a higher longevity of the *l1l* mutants (Figure 2.12G). Unfortunately, the NF4536 mutant line didn't produce enough seeds from heat stress conditions to perform the longevity experiment.

In conclusion to this preliminary functional analysis of *l1l* mutants, we only obtained promising but preliminary results regarding the potential and specific role of *L1L* in regulating seed development under heat stress. *MtL1L* could be a promising candidate gene with a specific role in regulation of seed development under heat stress conditions in *Medicago truncatula*, which could explain its high downregulation (*i.e.* around 10 fold) in embryo at S1 under heat stress (Figure 2.12A). Following this experiment, we re-screened the *Medicago Tnt1* mutant population and identified two novel insertional mutant lines. The four mutant lines are currently growing and will be used to confirm preliminary results.

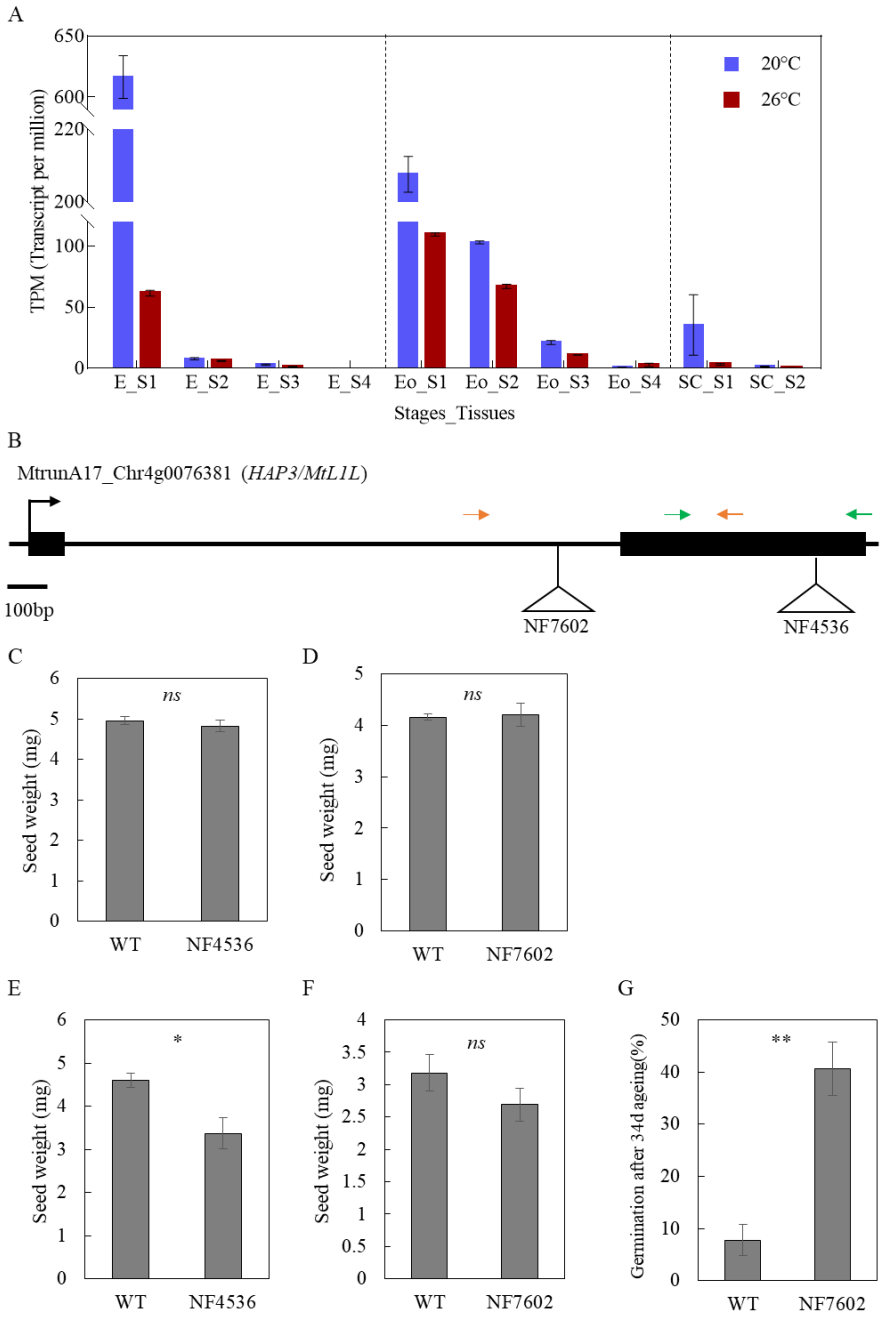


Figure 2.12. Characterization of the *l1l* mutants. (A) Transcript level of *HAP3*/*MtL1L* in three seed tissues in *Medicago truncatula* A17 at different stages (S1-4: stages 1 to 4) during seed development in control (20°C) and heat stress (26°C) conditions. E: embryo; Eo: endosperm; SC: seed coat. (B) Schematic gene structure of *HAP3*/*MtL1L* and *Tnt1* insertion sites of NF4536 and NF7602. Green arrows indicate the locations of primers used for NF4536 genotyping. Orange arrows indicate the locations of primers used for NF7602 genotyping. (C-D) Dry weight of NF4536 and NF7602 homozygous mutant and the corresponding sibling WT mature seeds produced under control (20°C) condition. (E-F) Dry weight of NF4536 and NF7602 homozygous mutant and the corresponding sibling WT mature seeds produced under heat stress (26°C) condition. (G) Germination rates of NF7602 aged seeds after 34 days artificial ageing (75% relative humidity and 35°C). *, 0.01<p-value<0.05; **, 0.001<p-value<0.01; *ns*, not significant.

3. Materials and methods

3.1 Plant materials and growth conditions

Medicago plants from the reference A17 genetic background were grown in controlled growth chambers following optimal conditions corresponding to 20°C/18°C (day/night) temperature and 16-h photoperiod with a light intensity of 150 $mmol.m^{-2}.s^{-1}$. For heat stress treatment during seed development, from the apparition of five flowers, plants were transferred to growth chambers with the same light, photoperiod, humidity and watering conditions, but with 26°C/24°C (day/night) temperature range.

Two independent *Tnt1* insertional *Mtl1l* mutants (NF4536 and NF7602) in *Medicago* R108 background were obtained from the Noble Research Institute (Ardmore, OK, USA) using the dedicated website (http://bioinfo.noble.org/mt_insertion/). Plants of homozygote mutants and sibling wild-types were first grown under optimal conditions (20°C/18°C, 16-h photoperiod) in growth room. At flowering time, half of the plants were kept at the same optimal conditions and half were moved to heat stress conditions (26°C/24°C, 16-h photoperiod) until seed maturity. Mature seeds were collected from 20°C and 26°C at the pod abscission stage and the pods were dried at 44% relative humidity and room temperature.

3.2 Seed weight and size measurement

Seed weight was measured from ten seed lots of 30 seeds using a precision balance, then the average individual seed weight was calculated by dividing the total seed weight by the number of seeds. To measure the seed size by image analysis, pictures were taken for triplicate of 50 seeds with a black background to improve the contrast. The image analysis was performed using ImageJ software (Schneider *et al.*, 2012) to automatically measure the individual seed area.

3.3 Seed longevity and germination assays

Artificial ageing experiments were performed using scarified *Medicago* mature seeds stored at 35°C and 75% relative humidity during different durations of 10, 20, 30, 40 and 50 days.

Because of the limited seed number obtained from *Medicago l1l* mutants, mature seeds were artificially aged in the same conditions for 34 days. Triplicates of 50 seeds were retrieved from different intervals of storage and imbibed in 5ml water in 5cm petri dishes containing a Whatman filter paper at 20°C in the dark. The percentage of viable seeds was calculated as the seed germination percentage after nine days of imbibition. The survival curve was made based on the germination percentage of viable seeds at each time point. From the survival curve, we determine the P50 value, which corresponds to the storage time to lose 50% viability of the seed lot.

To perform germination assays, triplicates of 50 seeds were first scarified and imbibed in 5ml water in 5cm petri dishes with a Whatman paper at 15°C in the darkness. Germinated seeds (*i.e.* protruding radicles > 1 mm) were counted every four hours. Germination speed representative value, T50, was calculated from the sigmoidal regression of each replicate as the time to reach 50% germination.

3.4 Gene set enrichment analysis (GSEA)

The enrichment analyses of gene lists were performed using the Clusterprofiler package (Yu *et al.*, 2012) in R using hypergeometric test followed by Benjamini-Hochberg false discovery rate correction, providing a q-values or adjusted p-values. Mapman functional classes were obtained by annotation of *Medicago* proteins from the *Medicago* genome version 5 using Mercator v.4 (Schwacke *et al.*, 2019).

3.5 Gene co-expression network analyses

Differentially expressed genes in embryo between control and heat stress conditions (about 9k genes) identified from ImpulseDE2 analysis were transformed using z-score normalization and used to construct the co-expression network. Co-expression network modules and candidate hub genes were identified using the WGCNA package (v1.68) in R (Langfelder and Horvath, 2008). The automatic one step network construction was used for module detection, the power (soft threshold) was set to 34, minModuleSize to 30, maxBlockSize to 20000, mergeCutHeight to 0.20 and TOMType was unsigned. Finally, to identify candidate genes within relevant module eigengenes, we re-analyzed the genes contained in the correlated modules to select

genes with high module membership (MM, *i.e.* high connectivity) and high gene significance (GS, *i.e.* correlation between gene expression and traits).

3.6 DNA isolation, whole genome bisulfite sequencing and data analyses

Genomic DNA were extracted from embryo samples that were collected at four stages during seed maturation (S1, S2, S3 and S4) under standard and heat stress conditions by using NucleoSpin® Plant II kit following the protocol described in the manufacturers' instructions. DNA was pre-treated with RNaseA and concentration was measured using a Qubit™ Fluorometer (Invitrogen). DNA integrity was checked using an Agilent 2100 Bioanalyzer. DNA samples were sent to the BGI for library construction, bisulfite treatment using a ZYMO EZ DNA Methylation-Gold kit and paired-end sequencing using an Illumina Hiseq 2500 (PE150 60M) was performed by the Beijing Genomics Institute (BGI, https://www.bgi.com). Short reads (*i.e.* fastq files) received from BGI were first quality checked using FastQC algorithm and clean reads (*i.e.* Phred quality values >30) were mapped to the *Medicago truncatula* reference genome version 5 (Pecrix *et al.*, 2018, https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/) using Bismark software (Krueger and Andrews, 2011). After mapping, deduplication of sequences was performed, then, cytosine methylation sites were located and quantified using Bismark to generate methylation calling files. These methylation calling files were used to identify differentially-methylated regions (DMRs) using the 'bin' (or sliding window) method set at 100bp from the DMRCaller package available in R (Catoni *et al.*, 2018). Each context of methylation (*i.e.* CG, CHG or CHH) was considered independently.

3.7 Chromatin Immunoprecipitation (ChIP) of H3K27me3 histone marks and data analyses

ChIP experiments were performed on the embryo samples that were collected at the same four stages during maturation of seeds grown under standard and heat stress conditions. Formaldehyde cross-linking and extraction of chromatin were performed on about 150 mg of embryo powder. Sonication was carried out using a M220 Focused-ultra-sonicator (Covaris) for 15 minutes to obtain DNA fragments of approximately 500bp. Chromatin was pre-cleaned using Invitrogen Dynabeads™ Protein A and G before immunoprecipitation. Anti-trimethyl-

H3K27 (Cat. #17-622, Millipore) was used for chromatin Immunoprecipitation of the IP samples. Reverse cross-linkage of H3K27me3 proteins with associated DNA was performed and DNA was purified using the AMPure XP beads (A63881, Beckman Coulter), then quantified using a Qubit™ Fluorometer (Invitrogen). Immunoprecipitated DNA was amplified using the MicroPlex Library Preparation Kit v2 (Diagenode) then purified by AMPure XP beads (A63881, Beckman Coulter, Inc.). INPUT samples were used as negative controls to normalize the bias of aspecific immunoprecipitation and library construction, thus INPUT samples followed the same protocol without the immunoprecipitation step. All the prepared IP and INPUT libraries were sent to Beijing Genomics Institute (BGI, https://www.bgi.com) for sequencing using an Illumina Hiseq 2500 (PE150 60M). Short reads (*i.e.* fastq files) received from BGI were first quality checked using FastQC algorithm and clean reads (*i.e.*Phred quality values >30) were mapped to the *Medicago truncatula* reference genome version 5 (Pecrix *et al.*, 2018, https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/) using STAR mapper software (Dobin *et al.*, 2013). After mapping, deduplication of sequences was performed using Picard (mark duplicate function, GATK), then, peak calling between the input and IP samples was performed using MACS Broadpeak detection (Zhang *et al.*, 2008).

4. Supplementary materials

Supplementary Table S1: Transcriptomic analyses. ImpulseDE2 and DESeq2 results identifying differentially expressed genes at different developmental stages and growth conditions.

Supplementary Table S2: List of annotated candidate genes identified from the WGCN analysis regarding acquisition of seed weight and seed longevity.

Supplementary Table S3: List of 97 genes showing both differential expression and differentially methylated promoter regions.

Supplementary Table S4: Gene list corresponding to H3K27me3 genomic binding sites located in 1kb-promoter and coding sequence regions, which are common in control and stressed seeds (_common), specific to control (_spec20) and specific to heat stress (_spec26) conditions at the four seed developmental stages.

Supplementary Table S5: List of primers used to perform PCR experiments for *l1l* mutant genotyping.

CHAPTER 3: STUDY OF THE REGULATION OF SEED TRAITS IN OPTIMAL AND HEAT STRESS CONDITIONS USING NATURAL *MEDICAGO TRUNCATULA* ACCESSIONS AND GENOME-WIDE ASSOCIATION STUDIES

AIM: the aim of this second section is to use the natural genetic diversity present in the *Medicago truncatula* HapMap collection to identify loci/genes potentially involved in seed trait plasticity in response to heat stress using a genome-wide association study (GWAS) approach.

MAIN RESULTS:

- Comparison between single- and multi-locus models to perform genome-wide association studies, with an improved detection from multi-locus models.
- Identification of candidate genes potentially controlling seed size.
- Identification of candidate genes potentially controlling seed composition.
- Identification of candidate genes potentially controlling seed germination performances.
- Preliminary results from functional analysis showing that *MIEL1* gene acts as a regulator of germination plasticity of seeds in response to heat stress.

1. Introduction

Medicago truncatula as a model plant is used to study different aspects of legume biology, including developmental processes and response to (a)biotic stresses (for review Kang *et al.*, 2016). Natural *Medicago truncatula* accessions mainly distributed in the Mediterranean basin and originated from different geographical locations and diverse climate environments were collected to constitute a core collection of *Medicago truncatula* accessions (Ronfort *et al.*, 2006). Recently with the development of high-thought sequencing technologies and phenotyping platforms, genome-wide association study (GWAS) has been a powerful method to investigate the quantitative trait nucleotides (QTNs) controlling various traits. To this purpose, a relevant subset of these accessions was, subsequently, selected based on the structuration of the population, and 288 *Medicago* accessions were sequenced using next generation sequencing technologies in order to identify single nucleotide polymorphisms (SNPs) existing within this population (http://www.medicagohapmap.org/home/view). This resource of plant accessions with known polymorphisms, called the haplotype map (HapMap) population of *Medicago truncatula*, represents a valuable genetic resource in legume biology to identify candidate genes associated with the acquisition of agronomic traits (for review Mammadov *et al.*, 2012; Govindaraj *et al.*, 2015).

In this chapter, we took advantage of this HapMap population to first performed genome-wide association studies using 162 *Medicago* HapMap accession to identify candidate loci regulating seed size and seed composition from seed lots that were already produced in optimal conditions and were directly available for phenotyping. This first approach allowed us to compare classical single-locus mixed linear models such as EMMA (Hyun *et al.*, 2008) with more recent multi-locus models such as the fixed and random model circulating probability unification (FarmCPU) (Liu *et al.*, 2016) (section 2). In parallel, a second approach was to grow 200 HapMap accessions and produce mature seeds under both optimal and heat stress conditions to obtain putative candidate loci/genes involved in controlling seed size and seed vigor traits in optimal and stress conditions, as well as loci/genes controlling phenotypic plasticity of these seed performance traits in response to heat stress (section 3 and 4).

2. Genome-wide association study identified candidate genes for seed size and seed composition improvement in *M. truncatula*

(This section was published in journal of Scientific Report.)

scientific reports

Check for updates

OPEN

Genome-wide association study identified candidate genes for seed size and seed composition improvement in *M. truncatula*

Zhijuan Chen[1], Vanessa Lancon-Verdier[2,5], Christine Le Signor[3], Yi-Min She[2,6], Yun Kang[4] & Jerome Verdier[1,2]✉

Grain legumes are highly valuable plant species, as they produce seeds with high protein content. Increasing seed protein production and improving seed nutritional quality represent an agronomical challenge in order to promote plant protein consumption of a growing population. In this study, we used the genetic diversity, naturally present in *Medicago truncatula*, a model plant for legumes, to identify genes/loci regulating seed traits. Indeed, using sequencing data of 162 accessions from the *Medicago* HAPMAP collection, we performed genome-wide association study for 32 seed traits related to seed size and seed composition such as seed protein content/concentration, sulfur content/concentration. Using different GWAS and postGWAS methods, we identified 79 quantitative trait nucleotides (QTNs) as regulating seed size, 41 QTNs for seed composition related to nitrogen (i.e. storage protein) and sulfur (i.e. sulfur-containing amino acid) concentrations/contents. Furthermore, a strong positive correlation between seed size and protein content was revealed within the selected *Medicago* HAPMAP collection. In addition, several QTNs showed highly significant associations in different seed phenotypes for further functional validation studies, including one near an RNA-Binding Domain protein, which represents a valuable candidate as central regulator determining both seed size and composition. Finally, our findings in *M. truncatula* represent valuable resources to be exploitable in many legume crop species such as pea, common bean, and soybean due to its high synteny, which enable rapid transfer of these results into breeding programs and eventually help the improvement of legume grain production.

Legume seeds are an important source to provide human food and animal feed. The high contents in proteins and carbohydrates, as well as fibers and minerals in legumes are an essential component of human diets [1]. With the world population growing and the increasing need of plant proteins, producing highly nutritious seeds with high protein content, essential amino acids and minerals is in great demand.

Compared to grains, legume seeds have naturally high protein contents; however, they are deficient in sulfur-containing amino acids and have lower concentrations of certain dietary minerals such as Fe, Ca and Zn compared to animal proteins[2]. Increasing seed protein production and improving seed nutritional quality have been a challenge in the agronomic field.

The existing natural diversity of legume could help identify key molecular players in achieving these challenges by understanding its underlying molecular mechanisms and by identifying molecular markers. *Medicago truncatula* is a Mediterranean originated plant and has been a model plant of legumes from 1990[3,4]. Its genome was sequenced and has still been under development with a recent fifth release[5].

Several quantitative trait loci (QTL) analyses have been performed in *M. truncatula* to identify loci affecting seed protein and mineral compositions[6,7]. Nevertheless, QTL identification depends on mapping population

[1]Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, 49000 Angers, France. [2]Shanghai Center for Plant Stress Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China. [3]Agroecologie, AgroSup Dijon, INRAE, Université Bourgogne Franche Comte, 21000 Dijon, France. [4]Noble Research Institute, LLC, Ardmore, OK 73401, USA. [5]Present address: USC 1422 GRAPPE, INRAE, Ecole Supérieure d'Agricultures, SFR 4207 QUASAV, 55 rue Rabelais, 49100 Angers, France. [6]Present address: Centre for Biologics Evaluation, Biologics and Radiopharmaceutical Drugs Directorate, Health Canada, Ottawa, ON K1A 0K9, Canada. ✉email: jerome.verdier@inrae.fr

genetics of a few parents limited its use in exploratory genetic approach. Genome-wide association studies (GWAS) use a broad panel of natural accessions with high genetic diversity and could overcome QTL analysis limitations[8]. Nowadays, GWAS has become a useful approach to explore the genetics of natural accessions and agronomic traits. A *Medicago* HAPMAP collection of over 200 natural accessions has been developed, which contains several millions of single nucleotide polymorphisms (SNPs)[9]. This *Medicago* GWAS panel has been successfully employed to identify candidate loci/genes associated with various agronomic traits[4] such as seed protein composition[7].

In this study, we performed GWAS focusing on seed traits related to seed size and seed composition using 162 accessions from the *M. truncatula* HAPMAP collection. Moreover, we performed association studies using both single and multi-locus models as well as several postGWAS analyses in order to identify potential loci/genes that could be involved in seed nutritional qualities in *M. truncatula*.

Results

Phenotypic evaluation of seed traits among the HAPMAP seed collection. We evaluated the phenotypic variation of 162 *Medicago* accessions on 16 seed traits regarding seed size and composition, plus 16 additional traits related to seed mineral composition in a subset of 88 accessions. Seed size was determined by weight measurement, area, perimeter, length (called 'majellipse' for major axis of ellipse) and width (called 'minellipse' for minor axis of ellipse)[10]. Seed color variations (called CH1, CH2 and CH3) potentially reflected the secondary metabolite composition in the seed coat. Global seed composition was characterized including carbon, hydrogen, nitrogen and sulfur percentages (w/w) (called %C, %H, %N, %S). From these concentration values of nitrogen and sulfur, we estimated the nitrogen and sulfur contents per seed of each accession based on individual seed weights (traits called N Content and S Content and expressed in milligram per seed). Nitro-gen concentration/content is a good indicator of the global protein content in seed and is commonly used for total protein determination in food products. Indeed, a predefined coefficient factor, Jones Factor[11], is used to convert the nitrogen concentration into total protein content. This coefficient is 6.25, but might vary between species and plant tissues. We also calculated the ratio between carbon and nitrogen (C/N), which corresponds to a global seed composition estimation. Sulfur concentrations/contents were also characterized, which reflected high-quality storage proteins. Indeed, legume seeds generally have a low level of sulfur-containing amino acids,which were shown to be tightly regulated by plant sulfur status[12,13]. Finally, other minerals (i.e. macro- and micro-elements) were quantified in seeds from a subset of 88 accessions. Concentrations of macro- (P, K, Mg, Ca, Na) and micro- (Fe, Mn, Zn, Cu, Mo, Co, Ni, V) elements were determined in mature seeds. All phenotypic values for the analyzed accessions are provided in the Supplemental Table S1.

Phenotypic diversity and correlation between seed traits and Impact of geographical location. A wide range of phenotypic variation was observed among the different accessions tested (Supplementary Figure S1 and Supplemental Table S1) with a coefficient of variation (CV) ranging from 1% for the most stable traits such as carbon and hydrogen concentrations, to 84% for Fe concentration. Other seed traits showed a high variability such as seed weight, N content and S content with CVs around 20%. In general, seed mineral concentrations showed the highest phenotypic diversity with Fe, Zn and Na displaying higher CV values. All the phenotypic values and CVs are provided in Supplementary Table S1.

Due to the availability of geographical locations of each accession origin, we allocated different accessions to three geographical values (i.e. longitude, latitude, altimeter) and 19 bioclimatic values obtained from the WorldClim database (http://worldclim.org). These bioclimatic values (called BIO1 to BIO19) mainly represent temperature and rainfall values measured monthly, quarterly or annually (see details in Fig. 1 legend). A global correlation analysis was performed to identify correlations between seed phenotypic traits themselves and with their geographical and bioclimatic values (Fig. 1). Results showed that all seed traits related to seed size (i.e. weight, area, perimeter, minellipse and majellipse) were highly correlated (Pearson coefficient correlation, PCC > 0.9), which validated the accuracy of our measurements. Similar results were obtained for seed color values (i.e. PCC>0.85 for CH1, CH2, CH3).

Regarding seed content, we observed that nitrogen and sulfur contents were also highly correlated with seed size traits (PCC>0.89 for N content and 0.74 for S content), which suggested that variations in seed content were predominantly determined by seed size. Regarding mineral composition in seeds, we observed positive correlations between concentrations of some elements such as Ca, Mg, Fe, Cu and Na (PCC > 0.7) but also between the macro-elements P and K (PCC>0.75, Fig. 1).

With the addition of the geographical values, we observed a moderate positive correlation between accession longitudes and seed C/N ratio (see the legend in Fig. 1), which indicated that accessions collected from the East tended to have higher C/N ratio (i.e. less nitrogen). To explain this difference, we also observed moderate positive correlations (PCC > 0.35) between seed size, seed contents (N and S) and temperature (i.e. BIO 9, 10), and at the opposite moderate negative correlations (PCC < -0.3) between seed weight, N content and precipitations (i.e. BIO 14, 17, 18). The integration of the bioclimatic data suggested that temperature and precipitation played an important role in accession adaptability to final seed size determination, with outcome in sulfur and nitrogen contents.

Genome-wide association analysis of seed traits. In order to perform genome-wide association analysis, we first, used the Box-Cox procedure[14] to estimate the appropriate lambda to transform our phenotypic data and, therefore, validate the assumption of normality required when performing GWAS prediction. Out of the 32 measured seed phenotypes, 26 traits were normalized using respective lambdas to finally display a normal distribution according to Shapiro–Wilk test (Fig. 2, Supplementary Table S1 and Supplementary Figure S1).

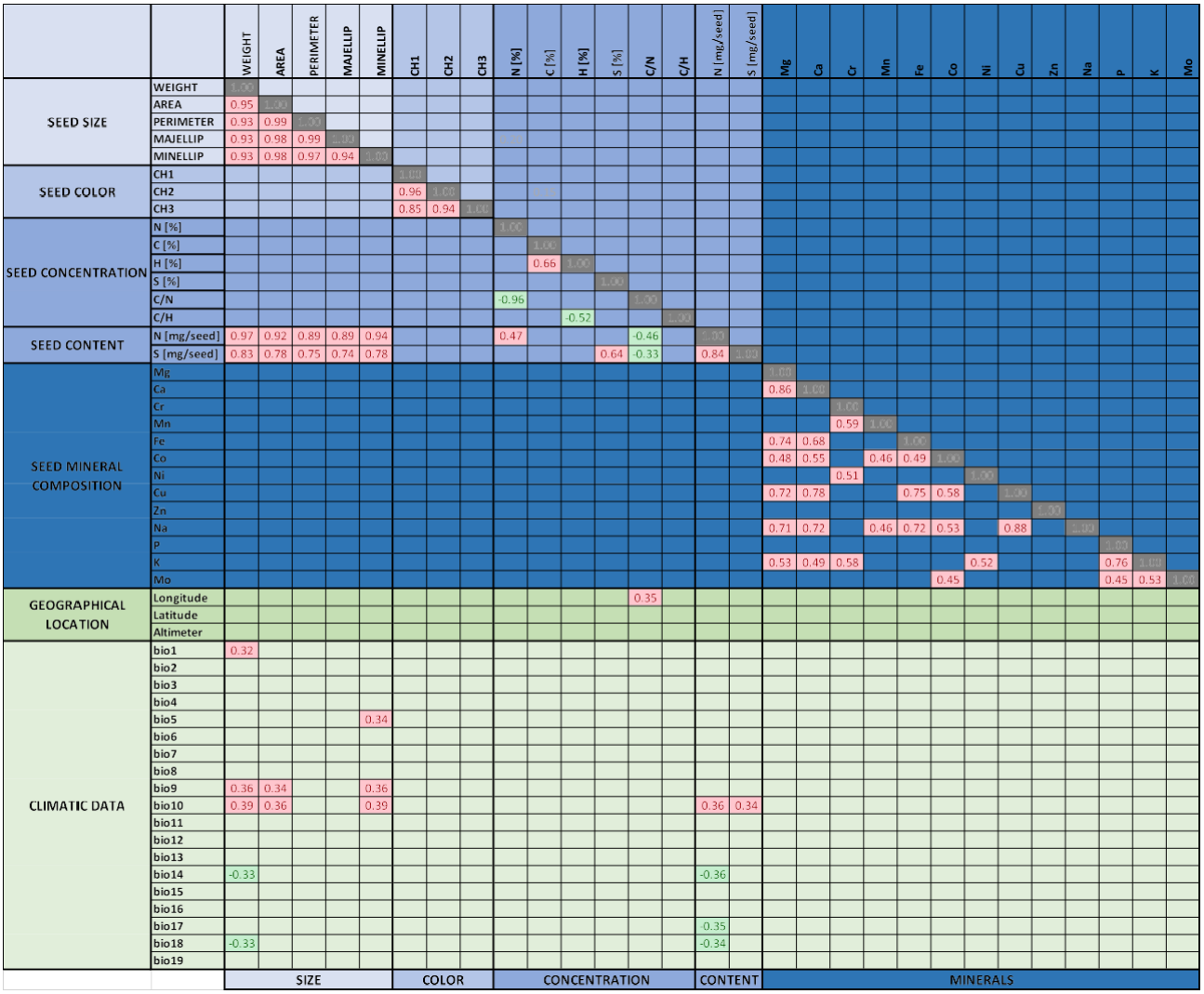


Figure 1. Correlation matrix between *Medicago* seed traits, and in relation to their geographical locations and climatic data. Only Pearson correlation coefficients (PCC) with adjusted p-values below 5% are indicated after BH procedure to control false discovery rate. Red color indicates PCC above 0.2 and green color indicatesPCC below − 0.2. Longitude is expressed in degrees with negative degrees representing west and positive degrees representing east. Latitude is also expressed in degrees with negative degrees representing south and positive degrees representing north. Climatic data are from WorldClim. *BIO1* annual mean temperature, *BIO2* mean diurnal range, *BIO3* isothermality, *BIO4* temperature seasonality, *BIO5* max temperature of warmest month, *BIO6* min temperature of coldest month, *BIO7* temperature annual range, *BIO8* mean temperature of wettest quarter, *BIO9* mean temperature of driest quarter, *BIO10* mean temperature of warmest quarter, *BIO11* mean temperature of coldest quarter, *BIO12* annual precipitation, *BIO13* precipitation of wettest month, *BIO14* precipitation of driest month, *BIO15* precipitation seasonality, *BIO16* precipitation of wettest quarter, *BIO17* precipitation of driest quarter, *BIO18* precipitation of warmest quarter, *BIO19* precipitation of coldest quarter.

However, six seed traits corresponding to the perimeter, CH1, %C, %H, C/H ratio and Arsenic (As) concentration were discarded from subsequent GWA analyses since, even after transformation, these traits did not reach normality.

In this study, two different models for genome-wide association predictions were applied to normalized phenotypes: a classical single-locus mixed linear model (EMMA[15]) with kinship and population structure as inputs, and a multi-locus model (FarmCPU[16]) with correction of population structure. When performing the multi-locus FarmCPU model, we observed QQ plots with a better fit between the expected and observed results following the expected null-hypothesis distribution of p-values (Supplementary Figure S2). These QQ plots reflected that most of the tested SNPs have no significant p-values, except for a few SNPs that have a strong and significant effect. Moreover, QQ plots obtained after performing the EMMA algorithm generally showed a curve corresponding to observed results below the theoretical curve (i.e. deflated curve), which suggested that this model was not appropriate for this association study. Regarding the Manhattan plots obtained from different

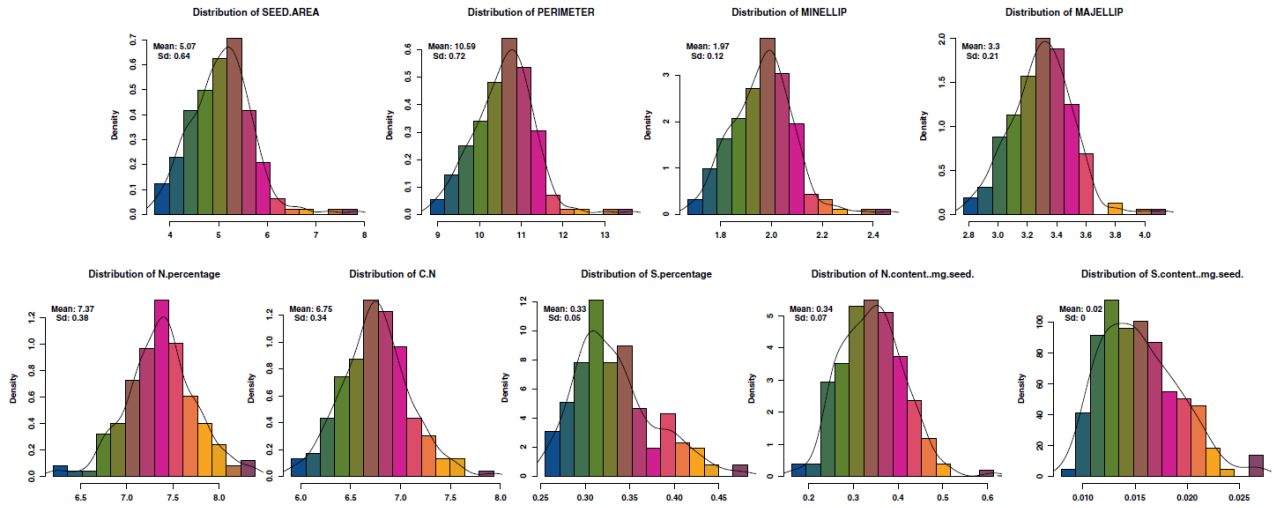


Figure 2. Distribution histograms of seed size and composition phenotypes in different Medicago accessions. Corresponding distribution curves are indicated on histograms. Different x-axes represent the corresponding values of the phenotypes.

models, we also observed differences between EMMA and FarmCPU (Supplementary Figure S2). In general, we obtained less background noise with FarmCPU, with more precise location and lower p-values of SNPs than the ones obtained from Mixed Linear Model (MLM), especially when statistical analysis showed highly significant SNPs. Manhattan plots obtained from MLM displayed broader "peaks" made of multiple significant SNPs (i.e. SNP clusters). Overall, we note that most of the highest significant SNPs were identified in both methods but FarmCPU provided more power detection and accuracy to identify quantitative trait nucleotides (QTNs) (Supplementary Figure S2). Therefore, we decided to focus on the multi-locus mixed model with FarmCPU in the subsequent analyses. All results (Manhattan and QQ plots) obtained from FarmCPU in this study are provided as Supplementary figures S3-S7. Moreover, gwas files directly readable on any genome browsers such as the web-accessible JBrowse[17] or desktop genome viewer such Integrative Genome Viewer (IGV[18]) are also provided as Supplemental Tables S2-S5.

As previously described, we observed two contrasting situations regarding association studies and their resulting Manhattan plots: identification of highly significant QTNs with clear genomic location and identification of clusters of SNPs indicating associated loci. As preliminary results of these analyses, we clearly identified highly significant QTNs associated with seed size (Supplementary Figure S3) and seed composition (Supplementary Figure S4) present on several chromosomes. For instance, we observed five, six, four and six QTNs highly associated respectively with seed area, seed length, seed width and seed weight with a $-\log_{10}$(p-value) > 10 (i.e. p-value < 10^{-10}). Regarding seed color (Supplementary Figure S5) and seed mineral concentrations (Supplementary Figure S6, S7), QTN p-values were significantly lower and nearer to background noise, which allowed only identification of specific genomic regions (i.e. SNP clusters), rather than highly significant individual QTNs.

To identify relevant QTNs, we combined association results from highly correlated seed traits. For instance, we combined FarmCPU results from weight, area, majellipse and minellipse (Fig. 3a) and identified common QTNs between seed size traits such as MtrunA17Chr4_56801315 on Chromosome 4. Interestingly, this QTN showed high p-values with all four seed size traits (10^{-18}, 10^{-25}, 10^{-21}, 10^{-10} with respective area, majellipse, minellipse and weight), suggesting a reliable QTN regulating seed size. This QTN is located within the genomic sequence encoding for a protein containing an RNA binding motif (gene ID MtrunA17Chr4g0065741). Another potentially reliable QTN (MtrunA17Chr1_35506650) was identified from three different seed size phenotypes with highly significant p-values of 10^{-9}, 10^{-8}, 10^{-19} for area, minellipse and weight, respectively. This QTN located on chromosome 1, closely related to a genomic sequence encoding a WD40-LIKE transcription factor (gene ID MtrunA17Chr1g0185101).

Similarly, we compared association studies between sulfur content and sulfur concentration to identify four major QTNs shared between these two traits with low p-values (Fig. 3b). MtrunA17Chr1_31627600 on chromosome 1, located within the coding sequence of the EXPORTIN5 protein (MtrunA17Chr1g0180461) closely related to Arabidopsis HASTY1 protein, which was shown to act as a nucleocytoplasmic transporter involved in the nuclear export of small RNAs[19]. MtrunA17Chr4_32623172 in chromosome 4, located in a chromosomic region rich in transposable elements. MtrunA17Chr5_8051955 present in chromosome 5 and is close to a gene encoding a salicylate methyltransferase (SAMT, MtrunA17Chr5g0404631), which catalyzes the methylation of salicylic acid with S-adenosyl-L-methionine to form methyl salicylate (MeSA), mainly in response to stress[20]. MtrunA17Chr8_48959923 on chromosome 8, located in the promoter region of a gene encoding a histidine kinase (MtrunA17Chr8g0392301).

Regarding nitrogen composition, we compared association studies between nitrogen concentration, nitrogen content and CN ratio in seeds (Fig. 3c). Following this experiment, it was more difficult to identify clear QTNs such as the N concentration and the CN ratio result showed more genomic regions that individual and distinct

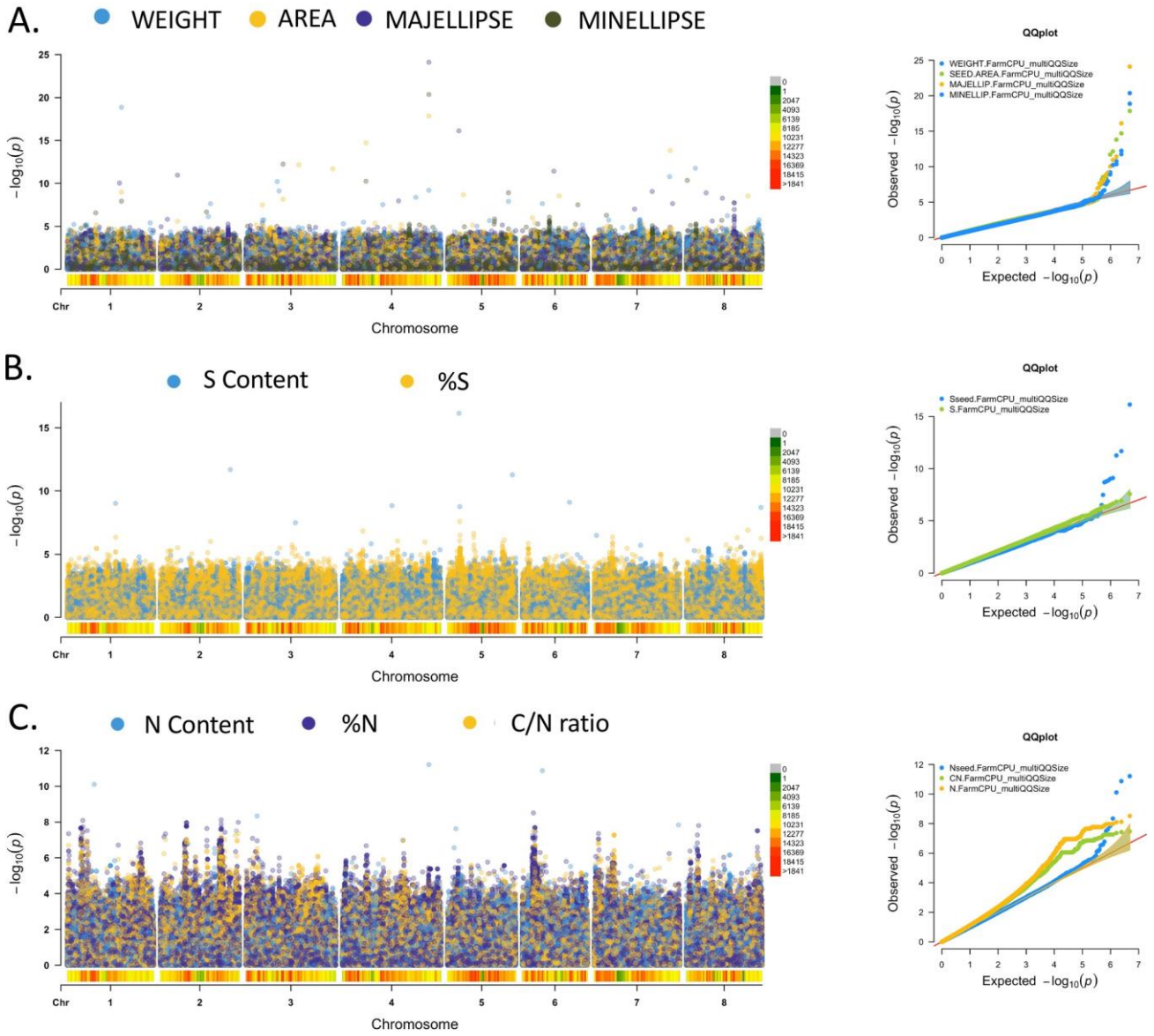


Figure 3. Genome-wide association studies of the *Medicago* seed traits with Manhattan plots and QQ plots obtained from FarmCPU. **(A)** Combination of association studies regarding seed size (weight, area, majellipse, minellipse). **(B)** Combination of association studies regarding seed sulfur content (mg/seed) and sulfur concentration (%, w/w). **(C)** Combination of association studies regarding seed protein content (nitrogen content (mg/seed); nitrogen concentration (%, w/w); carbon/nitrogen ratio).

QTNs associated with these phenotypes. However, it appeared that regions mainly located on chromosomes 1, 2, 6 and 8 showed strong associations between seed nitrogen composition and different accession polymorphisms, which suggested that these regions could play a role in seed nitrogen composition. Moreover, some particular QTNs were highly relevant for further analyses and indicated in Table 1. For instance, first, we identified a highly significant QTN (MtrunA17Chr6_7310002) associated with both protein concentration and C/N ratio, which is closely located to a genomic sequence encoding a putative amino acid transporter (MtrunA17Chr2g0333321). Second, we also identified a highly significant p-value for the QTN MtrunA17Chr4, which was already identified in the four seed size traits, in the N content association study. This result was predictable due to the high PCC between seed size and nitrogen content, which suggested that this QTN could be a regulator of both traits, making this QTN a potentially interesting candidate to improve concomitantly seed size and seed protein content.

Regarding seed color and seed mineral concentrations, several loci were identified by combining results from CH2 and CH3 and from all macro- and micro-element concentrations. However, no major QTNs (i.e.p-values $> 10^{-10}$) and precise location of SNP clusters were identified. This absence of highly significant QTNs

	GWAS (FarmCPU)				LD and putative causal gene(s) (PLINK)		Expression (RNA-seq) TPM							Annotations
	Traits	Chromosome	SNP ID (Chr_position)	P-Value	Number of potential SNP in LD according to PLINK (including QTN)	Associated gene(s)	Pod	Blade	Flower	Nodule	Root	Root_all	Shoot_all	Description
Seed size	Area	4	MtrunA17Chr4_56799264	1.40E−18	1	MtrunA17Chr4g0065741	1.12	1.42	1.22	0.60	1.62	3.89	11.25	RNA-binding (RRM RBD RNP motif) family
	Area	4	MtrunA17Chr4_15564603	2.00E−15	1									
	Area	7	MtrunA17Chr7_49921035	1.53E−14	1	MtrunA17Chr7g0267081	5.19	1.67	6.29	4.57	4.19	8.56	7.49	Probable CCR4-associated factor 1 homolog 11
	Area	3	MtrunA17Chr3_34669807	6.93E−13	1	MtrunA17Chr3g0112751	2.92	0.77	7.64	8.62	11.33	36.22	12.29	Hypothetical protein MTR_3g069670
	Area	3	MtrunA17Chr3_57001194	1.96E−12	1	MtrunA17Chr3g0143421								
	Area					MtrunA17Chr3g0143431	0.00	0.00	0.05	1.85	4.26	52.45	2.45	Syringolide-induced 14-1-1
	Area					MtrunA17Chr3g0143441	11.60	2.33	11.12	25.66	16.91	76.05	25.55	24-methylenesterol C-methyltransferase 2
	Majellip	4	MtrunA17Chr4_56801315	7.74E−25	1	MtrunA17Chr4g0065741	1.12	1.42	1.22	0.60	1.62	3.89	11.25	RNA-binding (RRM RBD RNP motif) family
	Majellip					MtrunA17Chr4g0065751	3.05	6.48	10.23	5.50	11.43	31.59	16.96	DUF21 domain-containing At4g14240-like
	Majellip	5	MtrunA17Chr5_7453281	7.83E−17	1									
	Majellip	6	MtrunA17Chr6_20728994	3.78E−12	1									
	Majellip	2	MtrunA17Chr2_11271424	1.10E−11	1	MtrunA17Chr2g0292281	0.00	0.00	0.00	0.00	0.00	0.45	0.11	Peroxidase family
	Majellip					MtrunA17Chr2g0292291	0.00	0.03	0.00	1.23	76.96	602.07	0.00	Peroxidase family
	Majellip	1	MtrunA17Chr1_34226006	9.20E−11	1	MtrunA17Chr1g0183471	2.04	0.56	8.05	5.46	4.05	3.49	2.62	Hypothetical protein MTR_1g069640
	Majellip					MtrunA17Chr1g0183481	2.44	1.39	4.30	6.04	3.33	0.00	0.00	unknown
	Minellip	4	MtrunA17Chr4_56799264	4.31E−21	1	MtrunA17Chr4g0065741	1.12	1.42	1.22	0.60	1.62	3.89	11.25	RNA-binding (RRM RBD RNP motif) family
	Minellip	3	MtrunA17Chr3_24325917	5.75E−13	1	MtrunA17Chr3g0099571	0.00	0.00	0.04	0.78	4.94	8.94	0.06	Disease resistance (CC-NBS-LRR class) family
	Minellip					MtrunA17Chr3g0099581	0.06	0.00	0.00	0.75	2.95	7.38	0.27	Probable disease resistance At4g27220
	Minellip					MtrunA17Chr3g0099591	0.86	3.27	1.95	0.08	0.33	1.03	14.96	Disease resistance (CC-NBS-LRR class) family
	Minellip	8	MtrunA17Chr8_628023	4.46E−11	1	MtrunA17Chr8g0334971	0.00	0.00	0.00	0.00	0.00	0.00	0.00	DUF247 domain
	Minellip	4	MtrunA17Chr4_15564603	5.65E−11	1									
	Minellip	5	MtrunA17Chr5_39987332	1.23E−09	1	MtrunA17Chr5g0441651	4.40	2.36	22.68	16.31	11.53	33.46	14.26	Dihydropyrimidinase
	Minellip					MtrunA17Chr5g0441661	0.58	0.00	39.34	0.00	0.00	0.00	0.00	RPM1-interacting 4 (RIN4) family
	Minellip					MtrunA17Chr5g0441671	11.02	4.34	26.25	30.95	42.28	163.92	77.69	Splicing factor 3B subunit 6
	Minellip					MtrunA17Chr5g0441681								
	Minellip					MtrunA17Chr5g0441691	0.58	0.66	7.45	0.18	0.51	0.57	2.74	Calcium-dependent kinase 17
	Minellip					MtrunA17Chr5g1024447								
	Weight	1	MtrunA17Chr1_35506650	1.33E−19	1	MtrunA17Chr1g0185101	2.79	2.49	4.33	12.71	12.22	22.03	4.59	BEACH domain-containing lvsA
	Weight	8	MtrunA17Chr8_5914802	1.74E−12	1									
	Weight	7	MtrunA17Chr7_49559389	1.66E−11	1	MtrunA17Chr7g0266521	202.23	7.53	100.55	140.68	41.92	39.45	24.23	Transmembrane protein, putative
	Weight					MtrunA17Chr7g0266531	70.18	1.94	27.86	52.11	14.60	17.34	10.28	Transmembrane protein, putative
	Weight					MtrunA17Chr7g0266541	43.90	4.20	9.31	30.74	0.00	0.35	0.00	Hypothetical protein MtrDRAFT_AC150442g27v2
	Weight					MtrunA17Chr7g0266551	0.00	0.00	0.00	0.85	0.00	0.41	0.00	Hypothetical protein MTR_7g104915
	Weight					MtrunA17Chr7g0266561	9.53	3.25	13.63	10.84	14.31	30.07	19.35	Probable small nuclear ribonucleo G
	Weight	3	MtrunA17Chr3_20500592	6.27E−11	1									
	Weight	4	MtrunA17Chr4_56799264	6.39E−10	1	MtrunA17Chr4g0065741	1.12	1.42	1.22	0.60	1.62	3.89	11.25	RNA-binding (RRM RBD RNP motif) family
Seed composition	S content	5	MtrunA17Chr5_7518926	7.21E−17	1	MtrunA17Chr5g0403771	0.02	1.48	0.59	0.16	0.18	2.86	1.68	Vicilin-like antimicrobial peptides 2-2
	S content					MtrunA17Chr5g0403781	5.98	11.36	8.86	2.58	2.27	8.34	94.20	Probable phosphatase 2C 80
	S content	2	MtrunA17Chr2_45988522	2.09E−12	1	MtrunA17Chr2g0326151	0.21	0.38	1.99	0.54	0.13	0.50	0.70	Pre-mRNA-processing-splicing factor 8
	S content					MtrunA17Chr2g0326161	0.00	0.00	0.00	0.00	0.00	0.42	0.00	Allergen gly M Bd 28 kDa
	S content	5	MtrunA17Chr5_42555471	5.47E−12	1	MtrunA17Chr5g0445531	0.00	0.00	52.16	0.00	0.00	0.00	0.08	Cytochrome P450 family 71
	S content					MtrunA17Chr5g0445541	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Cytochrome P450 family 71
	S content					MtrunA17Chr5g0445551	0.00	5.07	0.00	4.75	0.00	0.00	0.00	Cytochrome P450 family 71
	S content					MtrunA17Chr5g0445561	0.00	0.00	0.00	0.00	0.00	0.17	0.00	Zinc C3HC4 type (RING finger)
	S content	6	MtrunA17Chr6_30934715	8.09E−10	1									
	S content	1	MtrunA17Chr1_31627600	9.84E−10	1	MtrunA17Chr1g0180461	0.00	0.00	0.00	0.00	0.00	0.00	0.00	HASTY 1
	S content					MtrunA17Chr1g0180471	0.00	0.20	0.00	0.00	0.00	0.00	0.64	HASTY 1
	S content					MtrunA17Chr1g0180481	0.07	0.12	0.00	0.00	0.00	0.00	1.09	HASTY 1
	%S	5	MtrunA17Chr5_8051955	2.76E−08	3	MtrunA17Chr5g0404631	0.00	0.04	57.28	0.24	0.26	5.92	0.23	Salicylate O-methyltransferase
	%S					MtrunA17Chr5g0404641	0.00	0.00	0.11	0.83	0.00	0.83	0.33	Heavy-metal-associated domain
	N content	4	MtrunA17Chr4_56799264	6.14E−12	1	MtrunA17Chr4g0065741	1.12	1.42	1.22	0.60	1.62	3.89	11.25	RNA-binding (RRM RBD RNP motif) family
	N content	6	MtrunA17Chr6_13176638	1.33E−11	1									

Continued

	GWAS (FarmCPU)				LD and putative causal gene(s) (PLINK)		Expression (RNA-seq) TPM							Annotations
	Traits	Chromosome	SNP ID (Chr_position)	P-Value	Number of potential SNP in LD according to PLINK (including QTN)	Associated gene(s)	Pod	Blade	Flower	Nodule	Root	Root_all	Shoot_all	Description
	N content	1	MtrunA17Chr1_17611945	7.77E− 11	1	MtrunA17Chr1g0168711	34.54	36.56	59.95	28.95	53.04	113.94	152.11	3-isopropylmalate dehydratase large subunit-like
	N content	3	MtrunA17Chr3_7317464	4.52E− 09	1	MtrunA17Chr3g0085931	0.03	0.00	0.00	2.64	3.37	8.18	0.06	NBS-LRR type disease resistance
	N content					MtrunA17Chr3g0085941	0.00	0.00	0.00	0.00	0.09	0.00	0.00	Cytochrome C biogenesis ccsA
	N content	7	MtrunA17Chr7_55339972	1.45E− 08	1	MtrunA17Chr7g0275391	0.97	4.94	2.71	0.45	1.11	3.07	12.83	Copper-transporting ATPase chloroplastic-like isoform X1
	%N	6	MtrunA17Chr6_7310002	3.07E− 09	1	MtrunA17Chr6g0457641	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Cytochrome C biogenesis ccsA
	%N					MtrunA17Chr6g0457651	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Transmembrane protein, putative
	%N					MtrunA17Chr6g0457661	2.52	1.51	5.22	5.25	3.84	21.16	13.28	Zinc transporter 5-like isoform X2
	%N					MtrunA17Chr6g0457671	4.90	2.09	9.93	6.56	8.10	15.08	9.78	Zinc transporter 5
	%N	2	MtrunA17Chr2_39997412	7.52E− 09	27	MtrunA17Chr2g0318301	4.50	1.66	5.76	11.51	8.44	25.57	15.99	Receptor kinase THESEUS 1
	%N					MtrunA17Chr2g0318311	0.00	0.00	0.07	0.00	1.92	4.68	0.05	Root cap late embryogenesis
	%N					MtrunA17Chr2g0318321	0.00	0.80	1.10	0.00	0.00	0.00	0.00	RNA polymerase beta partial (chloroplast)
	%N					MtrunA17Chr2g0318331	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Receptor kinase THESEUS 1
	%N					MtrunA17Chr2g0318341	0.00	0.06	0.00	0.15	67.70	88.60	0.00	Root cap late embryogenesis
	%N					MtrunA17Chr2g0318351	28.50	12.81	54.94	39.90	53.20	110.12	69.26	Hypothetical protein MTR_2g080270
	%N					MtrunA17Chr2g0318361	0.00	0.00	0.00	0.00	0.00	0.83	0.00	Hypothetical protein MTR_2g080280
	%N					MtrunA17Chr2g0318371	4.61	28.32	12.88	1.56	6.66	7.97	66.65	Transmembrane protein, putative
	%N					MtrunA17Chr2g0318381	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Pentatricopeptide repeat-containing At1g20230-like
	%N	1	MtrunA17Chr1_10142836	8.00E− 09	4	MtrunA17Chr1g0159441	7.70	8.39	21.83	25.52	29.94	65.42	22.63	Transcriptional corepressor SEUSS
	%N					MtrunA17Chr1g0159451	2.40	3.50	7.74	6.58	10.84	19.75	9.76	Small RNA degrading nuclease 5
	%N	2	MtrunA17Chr2_51147031	1.03E− 08	1	MtrunA17Chr2g0333321	1.01	0.71	3.17	3.07	2.20	3.40	2.02	Probable sodium-coupled neutral amino acid transporter 6
	%N	2	MtrunA17Chr2_17222898	1.07E− 08	4	MtrunA17Chr2g0299211								
	%N					MtrunA17Chr2g0299221	0.00	0.00	0.00	1.87	0.00	0.00	0.00	Little zipper
	%N					MtrunA17Chr2g0299231	0.00	0.00	0.00	0.20	0.00	0.00	0.00	Hypothetical protein MTR_2g039220
	%N					MtrunA17Chr2g0299241	0.00	0.00	0.31	9.33	0.63	0.00	0.00	Nodule-specific Glycine Rich Peptide
	%N					MtrunA17Chr2g0299251	0.00	0.00	0.00	0.30	0.00	0.00	0.00	NA
	%N					MtrunA17Chr2g0299261	0.08	0.00	0.34	274.55	0.23	0.05	0.00	Nodule-specific Glycine Rich Peptide
	%N					MtrunA17Chr2g0299271	41.21	10.74	67.19	90.53	54.34	111.62	71.08	WD-40 repeat-containing MSI4
	CN ratio	6	MtrunA17Chr6_7310002	3.36E− 08	1	MtrunA17Chr6g0457641	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Cytochrome C biogenesis ccsA
	CN ratio					MtrunA17Chr6g0457651	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Transmembrane protein, putative
	CN ratio					MtrunA17Chr6g0457661	2.52	1.51	5.22	5.25	3.84	21.16	13.28	Zinc transporter 5-like isoform X2
	CN ratio					MtrunA17Chr6g0457671	4.90	2.09	9.93	6.56	8.10	15.08	9.78	Zinc transporter 5
	CN ratio	1	MtrunA17Chr1_10142836	3.75E− 08	4	MtrunA17Chr1g0159441	7.70	8.39	21.83	25.52	29.94	65.42	22.63	Transcriptional corepressor SEUSS
	CN ratio					MtrunA17Chr1g0159451	2.40	3.50	7.74	6.58	10.84	19.75	9.76	Small RNA degrading nuclease 5
	CN ratio	2	MtrunA17Chr2_51147031	4.35E− 08	1	MtrunA17Chr2g0333321	1.01	0.71	3.17	3.07	2.20	3.40	2.02	Probable sodium-coupled neutral amino acid transporter 6
	CN ratio	7	MtrunA17Chr7_13129833	5.48E− 08	30	MtrunA17Chr7g0227971	0.00	0.18	0.00	0.28	0.00	0.40	0.92	Nucleoporin GLE1
	CN ratio					MtrunA17Chr7g0227981	4.76	3.46	18.54	13.53	13.39	18.60	19.11	N-terminal glutamine amidohydrolase
	CN ratio					MtrunA17Chr7g0227991	0.00	0.00	0.18	0.00	0.00	0.00	0.00	Subtilisin-like serine endopeptidase family
	CN ratio	2	MtrunA17Chr2_49391628	5.58E− 08	1	MtrunA17Chr2g0330811	0.55	1.00	2.15	1.25	0.77	3.27	8.91	Chloroplastic group IIA intron splicing facilitator chloroplastic isoform X1
	CN ratio					MtrunA17Chr2g0330821	0.45	0.72	1.86	1.74	3.10	6.44	2.60	Heat shock transcription factor A8
	CN ratio					MtrunA17Chr2g0330831	0.00	0.00	8.21	6.70	3.44	2.69	1.71	NA
	CN ratio					MtrunA17Chr2g0330841	0.33	0.21	0.83	1.34	0.74	5.49	2.23	Heat stress transcription factor A-5-like

Table 1. Top five QTNs significantly associated with different seed size traits (i.e. weight, area, majellipse, minellipse) and seed compositions (S content, N content, %S, %N and C/N ratio). SNP/QTN names, positions and p-values are indicated from FarmCPU. Numbers of potential associated SNP(s) and putative causal genes are indicated from PLINK analysis. Gene expression in major *Medicago* plant organs, as well as tentative gene annotations are indicated. A more exhaustive list of highly significant QTNs related to all seed traits is provided as Supplementary Table S6, and complete lists of SNPs and their associated p-values are provided as Supplementary Tables S2 to S5.

ID	Description	p value	q value	Count
Size				
GO:0005689	U12-type spliceosomal complex	0.0004	0.0267	2
Composition				
GO:0045735	Nutrient reservoir activity	0.0000	0.0004	5
GO:0033609	Oxalate metabolic process	0.0000	0.0004	4
GO:0046564	Oxalate decarboxylase activity	0.0000	0.0004	4
GO:0030145	Manganese ion binding	0.0001	0.0011	4
GO:0015171	Amino acid transmembrane transporter activity	0.0002	0.0019	4
GO:0003333	Amino acid transmembrane transport	0.0002	0.0023	4
Color				
GO:0080043	Quercetin 3-*O*-glucosyltransferase activity	0.0003	0.0054	4
GO:0080044	Quercetin 7-*O*-glucosyltransferase activity	0.0003	0.0054	4
GO:0052696	Flavonoid glucuronidation	0.0004	0.0054	4

Table 2. Enrichment analysis of Gene Ontology (GO) terms on putative causal genes regulating different seed traits (i.e. size, composition and color). Enrichment p-values from hypergeometrical tests and q-values from Bonferroni corrections are indicated, as well as the number of genes annotated (count). Results were generated with R package "ClusterProfiler".

regarding seed mineral concentrations could be explained by the small population size used in this specific analysis (i.e. subset of 88 accessions).

PostGWAS analyses to identify putative causal genes. To shorten the list of candidate QTNs, we used p-value threshold of 10^{-7} when association studies displayed high SNP power detection such as seed size and seed composition phenotypes, and a p-value threshold of 10^{-5} when association analyses displayed low SNP power detection such as seed color and seed mineral concentrations. Then, the linkage disequilibrium (LD) was considered to identify putative causal genes associated with selected QTNs. Considering that in the *Medicago* HAPMAP collection, the average LD decay was determined around 15kb[21], we performed genome-wide correlations between selected SNPs present within this genomic range (i.e. ± 15 kb from QTNs) using PLINK[22]. A threshold correlation of 0.7 was used to identify SNPs potentially in LD within these genomic regions. From this analysis, we established a list of SNPs correlated to the selected QTNs due to LD and therefore potential causal genes. From this list, we revealed 56 putative causal genes related to the 34 QTNs with highly significant p-values that are potentially involved in seed size determination, 123 putative causal genes related to the 56 QTNs potentially involved in seed composition, 90 putative causal genes related to the 45 QTNs potentially involved in seed color and 906 putative causal genes related to the 597 QTNs potentially involved in seed mineral composition (Table 1 and Supplementary Table S1). Due to the relatively low number of ecotypes used for the QTN identification related to seed nutritional composition, which might affect the statistical accuracy of the study, we decided to provide these results as supplementary data but we will not analyze them further.

In order to identify functional classes that could be involved in regulating these different seed phenotypes, we performed over-representation gene ontology (GO) analyses with corresponding lists of putative causal genes for each phenotype (Table 2). Interestingly, we observed that list of putative causal genes regulating seed size were enriched in GO terms related to the U12-type spliceosomal complex (GO:0005689). Similarly, using list of putative causal genes regulating seed protein content/concentration, we observed enrichment of genes with GO terms referring to nutrient reservoir activity (GO:0045735), amino acid transport (GO:0015171, GO:0003333) and oxalate metabolic pathway (GO:0033609, GO:0046564), which are all functional classes closely related to biosynthesis or transport of amino acids[23]. From putative genes regulating the seed color, we revealed that the GO terms referring to flavonoid biosynthesis were enriched (i.e. GO:0080043, GO:0080044, GO:0052696), and it has been shown that, indeed, flavonoid composition/concentration is closely related to seed coat color[24]. Finally, we observed enrichment of the GO term related to the protein amino acid autophosphorylation (GO:0046777) concerning genes potentially regulating mineral concentrations, which was less intuitive and presumably has indirect relations.

In order to identify potential specific regulator of seed traits, we also focused on seed expression specificity and compared list of genes specifically expressed in seeds and pods with our list of candidate causal genes related to seed traits. Expression analysis in different *Medicago* plant organs was performed using publicly available information. To compare with our data, we mapped these reads to the *Medicago* genome version 5[5] and quantified transcript expression using the Salmon pipeline[25]. Out of 44,473 transcripts in the *Medicago* genome (v5). 375 were identified as specifically or preferentially expressed in pods/seeds (Supplementary Table S7). After combining a list of seed-specific genes and our list of putative causal genes from GWA studies, we revealed two seed-specific genes potentially regulating seed nitrogen concentration: a zinc-finger transcription factor (MtrunA17Chr7g0217321) and a CAAT-Binding Transcription factor (CBF, MtrunA17Chr2g0318461), and eight seed-specific genes potentially regulating various mineral concentrations in seeds (Supplementary Table S6).

Discussion

Improving seed protein content in *M. truncatula* seeds by increasing seed size. Grain legumes play a key role in providing plant proteins for food and feed. Therefore, understanding how to increase seed protein content and to produce storage proteins with high nutritional values (i.e. containing essential amino acid and sulfur-containing amino acids) represents a technological breakthrough that has to be yet overcome. In this study, we observed significant genetic variabilities regarding seed traits such as size, nitrogen content (i.e. storage protein content) and sulfur content (i.e. sulfur-containing amino acid content), which makes the *Medicago* HAPMAP collection a great tool to improve these agronomical traits. Interestingly, our correlation matrix between these different seed traits within the Hapmap population revealed a strong correlation (PCC>0.9) between seed size and protein content (Fig. 1), which suggested that increasing seed protein content could be directly achieved by increasing seed size. This hypothesis could, first, be confirmed by identification of colocalized QTLs of seed size and seed protein content in garden pea[26], soybean[27], Common Bean[28] and cowpea[29]. In parallel, even if several genetic studies already highlighted genes controlling seed size, which generally act via regulation of mitotic activity in embryo and endosperm, such as *SBT1.1*[30] and *DASH*[31]in *M. truncatula,* but also via regulation of cell elongation in endosperm and seed coat such as *ZHOUPI*[32] and *TTG2*[33] in *A. thaliana* (for review[34]). The hypothesis that increasing seed size would increase protein content is difficult to validate from literature because mutant lines displaying larger seeds were not tested for their protein contents and inversely, mutant lines affected in protein content were not tested for seed size. One exception is the gene *AP2* in Arabidopsis, which produced larger seeds in mutant plants combined with an increase in protein and fatty acid content[35], which validate our hypothesis. Finally, numerous correlation analyses between seed size and protein content have been conducted on cereals and legumes but no general trend was observed. Indeed, even if several studies concluded about clear positive correlations between seed size and seed protein content in pigeon pea[36], soybean[37] and this study in *Medicago*, many others did not, suggesting genotype-environment effects. As mentioned earlier, these results are undoubtedly dependent on plant genetic background, favorable growth conditions and optimal agricultural practices. Indeed, in our study, we revealed that the geographical and bioclimatic origins of *Medicago* accessions played an important role in plant adaptation with correlations between seed size, seed content, temperature and precipitation during the reproductive phase (Table 1). These accessions showed a phenotypic adaptability to produce larger and higher seed protein content. Moreover, the variations of these traits within the same genetic backgrounds are also to consider as abiotic stress is known to affect proper seed development in *Medicago*[38]. Finally, one essential aspect to validate this positive correlation between seed size and protein content is the non-limiting nitrogen supply, which could be achieved via intensive nitrogen fertilization or via nitrogen fixation in legumes, which is still active during seed filling. In this study, we highlighted genes/loci potentially involved in seed size, but also in both seed size and seed protein content, which could potentially improve simultaneously seed nutritional values and agronomical performances, as it is already well documented that larger seeds tend to improve germination vigor and plantlet establishment (for review[39]).

Efficiency of GWAS and post-GWAS algorithms. In the past 10 years due to the rapid developmentof genome sequencing technologies and phenotypic capacities, numerous genome-wide association studies (GWAS) have been performed in many species. This powerful tool is becoming a standard in forward genetic study to identify genes/loci controlling various traits. Its rapid development has been accompanied by the development of mainly two association study methodologies: classical single-locus GWAS methods based on General Linear Model (GLM) and Mixed Linear Model (MLM) (e.g. EMMA[15]; SUPER[40]), and recently developed multi-locus GWAS methods such as MLMM[41], FASTmrEMMA[42] and FarmCPU[43]. In the single-locus method, statistical tests are performed one locus at each time, whereas multi-locus methods consider the information of all loci simultaneously and consequently do not require false discovery rate correction, leading to higher QTN detection power[44]. In our study, we compared a single-locus method, EMMA, and a multi-locus method, FarmCPU, and we had two observations. (i) When association studies revealed highly significant candidate QTNs, FarmCPU (i.e. multi-locus method) resulted in more significant QTNs with lower p-values and more precise chromosome positions. Indeed, EMMA (i.e. the single-locus method) showed higher QTN p-values, closer to the background noise, which led to the identification of loci represented by broader "peaks" containing multiple significant SNPs (i.e. SNP clusters) in Manhattan plots, therefore more difficult to precisely locate on chromosomes (Figure S2). However, even if FarmCPU identified more significant QTNs with more precise locations, most of the highly significant QTNs were observed using both methods. (Figure S2A-B). (ii) When association studies did not reveal significant QTNs, single and multi-locus methods performed similarly (Figure S2C). In conclusion, from our study, it appeared that FarmCPU, the multi-locus method, globally performed better than the single-locus method, which explains why we focused on this method to identify candidate QTNs. Better performances of GWAS multi-locus models have also been observed in several other studies such as in Xu et al.[45] related to starch properties in maize, Jaiswal et al.[46] related to agronomic traits in wheat, and Li et al.[47] related to fiber quality in Cotton, rendering these methods attractive for association studies.

Potential regulation of seed size and protein content via RNA regulation. In order to determine reliable QTNs and mine for causal candidate genes controlling seed size and composition, we performed postGWAS analyses. First, we considered a 15 kb LD decay (r2 > 0.7), as determined in *Medicago* hapmap collection[21], to identify associated SNPs due to LD. Then, depending on the association results, we used different approaches to refine candidate gene selection: combination of association results from correlated phenotypes to identify putative causal genes, use of over-representation analysis to identify key functional classes regulating phenotypes, and integration of transcriptomics.

Regarding seed size, we mined two highly significant QTNs associated with multiple seed size phenotypes by combining GWAS results of weight, area, majellipse and minellipse. First, MtrunA17Chr1_35506650, a QTN detected in three association studies (i.e. weight, minellipse and area), is near a gene encoding a WD40/BEACH domain protein (MtrunA17Chr1g0185101) (Table 1 and Supplemental Table S6). A potential ortholog of this gene in Arabidopsis, called SPIRRIG (SPI, AT1G03060), has been shown to be involved in cell morphogenesis via interaction with processing bodies (i.e. p-bodies)[48], which is known to regulate mRNA processing during development or stress (for review[49]). In Arabidopsis, *spi* mutant lines displayed many developmental defects[50] including reduced seed coat mucilage and plant growth impairment under salt stress[51]. Interestingly, the second QTN (MtrunA17Chr4_56801315) detected in all four association studies related to seed size was closely related with a gene encoding an RNA-binding domain (RBD, MtrunA17Chr4g0065741), which is also a gene involved in the regulation of RNA. RDB proteins belong to a large protein family, which are known to determine RNA fate from synthesis to degradation. Few of them have been functionally characterized and depending on their RNA targets, they could play tissue- and developmental stage-specific roles[52]. For instance, one of RDB protein family functionally characterized in Arabidopsis seed development is *SUPPRESSOR OF ABI3* (*SUA*, AT3G54230), which controls alternative splicing of the ABI3, a master regulator of seed development and maturation[53]. This QTN identified from several seed size association studies was also detected in association with the seed nitrogen content (Table 2), which indicated the important role of this gene in regulating both seed size and protein content.

This role of RNA processing/regulation to regulate seed size was further highlighted by the over-representation analysis of all highly significant QTNs associated with seed size, which revealed that the "U12-type spliceosomal complex" class was over-represented. This complex is part of the minor spliceosome, which plays a crucial role in splicing regulation of the rare U12 introns. It has been shown in Arabidopsis that homozygote mutant lines impaired in the U12 spliceosome complex displayed premature embryo abortion, whereas heterozygote mutants were defective for seed maturation, indicating an essential role of this complex during embryonic development[54]. Moreover, proper splicing and alternative splicing have been shown to be crucial in normal embryo formation (for review[55]) and embryo development, which is a key stage in controlling the final seed size.

Methods

***Medicago* plant accession and growth.** Accessions from the HapMap germplasm collection were requested from the dedicated website (http://www.Medicagohapmap.org/hapmap/germplasm). Around 200 accessions were grown in growth chambers (20 °C/18 °C, 16 h photoperiod at 200 mmol m^{-2} s^{-1}) until maturity. Mature seeds of 162 accessions were collected in sufficient quantity to perform different phenotyping experiments.

Seed size and color determination. Individual seed weights of 162 accessions were estimated by weight- ing 50 seeds in triplicate using a precision balance at an accuracy ± 0.1 mg and displayed as mg per seed. To complete seed size phenotyping, image analyses were performed on 150 seeds of each accession using GrainS- can software[10] to automatically measure individual seed areas (i.e. pixel number, called "area"), seed perimeters ("perimeter"), seed lengths ("majellip") and seed widths ("minellip"). These seed size parameters were averaged for each of the 162 accessions and used for the subsequent analyses. Image analysis also allowed us to determine seed color values using GrainScan, which measured three color channels (i.e. CH1, CH2, CH3) from raw RGB values, reflecting seed coat pigmentation.

Seed composition analysis with elemental CHNS analyzer (162 accessions). Seed composi- tion was characterized using a CHNS elemental analyser, which measured the percentage (w/w) of carbon (C), hydrogen (H), nitrogen (N) and sulfur (S). Mature seeds were ground in liquid nitrogen and dried in an oven at 90 °C for 48 h. Then, triplicates of approximately 5 mg of powder were analyzed using an Elementar Vario Micro cube analyzer (Germany) using flash combustion of the sample based on the "Dumas" method. Concentrations of C, H, N, S were determined by the Elementar Vario software based on exact seed weights. From which, carbon–nitrogen ratios (C/N ratio) were calculated to provide an accurate overview of the global seed composition. Nitrogen and sulfur contents per seed for each accession (i.e. N content, S content) were calculated using average seed weights of each lot.

Macro- and micro-element concentrations. A subset of 88 accessions was analyzed to determine ele-mental concentrations for P, K, Mg, Ca, Na, Fe, Mn, Zn, Cu, Mo, V, Co, Ni, Ti, As, Cr using Induced Coupled Plasma-Mass Spectrometry (ICP-MS, Perkin Elmer model NexION 300D). Seed powders were dried in a heat- ing oven at 75 °C for overnight. Approximately 5 mg of seed powder were accurately weighed and transferred to a glass container with 3 ml of concentrated nitric acid (HNO_3). After digestion for 15 min at 200 °C, deionized water was added to adjust the final volume to 10.0 ml and samples were injected into the ICP-MS for measure- ment. A blank sample containing 5% HNO_3 was used for background subtraction. Concentrations (i.e. ppb or mg/L) of each element were calculated based on an internal standard mix (Perkin Elmer, ref. 9301721) and nor- malized according to a weight normalization procedure using the NexION software (Perkin Elmer).

Correlation analysis. Correlation matrix was performed on averages of phenotype values. Each pairwise comparison was performed using Pearson correlation calculated using the complete pairwise correlation of the 'corr.test' function from the R package 'psych'. P-values were adjusted using Benjamini-Hochberg (BH) to control false discovery rate and statistical significance threshold was set below 5% of adjusted p-values.

Phenotype normality distributions. All traits were checked and transformed to reach normality as it is required to perform genome wide association studies. Box Cox algorithm[14] was used to determine the appropriate transformation for each trait, and each trait was transformed separately according to the most suitable lambda values given by the Box Cox function implemented in the R package MASS[56]. After transformation, Shapiro–Wilk tests[57] were performed to validate the normality and traits that did not reach normality were discarded of following GWAS analyses. Supplementary Table S1 provides seed trait values before and after Box Cox transformation, respective lambda values for each trait and corresponding p-values of the Shapiro–Wilk test after transformation.

Genome-wide association studies and post-GWAS analyses. Single nucleotide polymorphisms (SNP) data were obtained by whole genome sequencing of the 262 *Medicago* accessions from the *M. trunca-tula* Hapmap project[9]. From the 6 million SNPs originally identified in *Medicago* genome version 4, 4,852,061SNPs were successfully mapped to the fifth version of the *Medicago* genome (Mtv5[5]) and were used for subse- quent analyses. The population structure and the kinship matrix used in this study were the same as previously described in Bonhomme et al.[58] and le Signor et al.[7], respectively. Two models were used to perform GWAS: (1) a classical single locus method using a mixed linear model called EMMA (Efficient Mixed-Model Association[15] with the kindship matrix and the population structure as inputs; (2) a multi-locus model called FarmCPU (Fixed and random model Circulating Probability Unification[16]) with correction of the population structure, both with a statistical test p-value threshold of 1%. The Manhattan and quantile–quantile (QQ) plots were plotted using the R package rMVP (https://github.com/xiaolei-lab/rMVP). PostGWAS analysis was performed to correct for the linkage disequilibrium (LD) using PLINK algorithm[22] with the "clump" function and the following options: clumb-kb-radius of 15, which represents the genomic range (in kilobases) to identify SNP in LD and clump-r2 of 0.7, which represents the r-squared threshold to identify correlation between SNPs. All GWAS result files were transformed into gwas files (Supplementary Tables S2 to S5) readable in web-application JBrowse[17] con- taining the *M. truncatula* genome version 5 such as https://Medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/ or in personal desktop genome viewer such as the freely available Integrative Genome Viewer (IGV[18], http://software.broadinstitute.org/software/igv/). Over-representation analyses (ORA) of candidate genes were performed using ClusterProfiler package available in R using hypergeometrical test (p-values) with a Bonferroni correction (q-values)[59].

RNA-seq analysis in major plant organs. Expression of *Medicago* transcripts in major plant organswas determined from existing experiments. Sequenced short reads (i.e. raw fastq files) were downloaded from the Sequencing Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) from different experiments and dif- ferent *Medicago* plant organs: nodule (SRX099057), seed pod (including seeds, SRX099058), 4-week blade (SRX099059), flower (SRX099061), 4-week root (SRX099062), all root system (SRX2943065, SRX2943064, SRX2943063) and all shoot system (SRX2943062, SRX2943058). Raw read files were mapped against the *Medicago* transcriptome version 5 (https://Medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/) and quantified as counts using Salmon algorithm[25]. Counts were normalized to corresponding library sizes (equivalent to count per million, CPM) then length of transcripts (Transcript per million, TPM) and displayed as TPM in our study.

Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Received: 15 July 2020; Accepted: 19 January 2021
Published online: 19 February 2021

References

1. Barman, A., M. Marak, C., Mitra Barman, R. & S. Sangma, C. Nutraceutical properties of legume seeds and their impact on human health. *Legume Seed Nutraceut. Res.* https://doi.org/10.5772/intechopen.78799 (2019).
2. Grusak, M. A. Enhancing mineral content in plant food products. *J. Am. Coll. Nutr.* **21**, 178S-183S (2002).
3. Barker, D. *et al. Medicago truncatula*, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. *Plant Mol. Biol. Rep.* **8**, 40–49 (1990).
4. Bandyopadhyay, K., Verdier, J. & Kang, Y. The model legume *Medicago truncatula*: Past, present, and future. in *Plant Bio- technology: Progress in Genomic Era* (eds. Khurana, S. M. P. & Gaur, R. K.) 109–130 (Springer, Singapore, 2019). https://doi. org/10.1007/978-981-13-8499-8_5
5. Pecrix, Y. *et al.* Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat. Plants* **4**, 1017–1025 (2018).
6. Sankaran, R. P., Huguet, T. & Grusak, M. A. Identification of QTL a V ecting seed mineral concentrations and content in the model legume *Medicago truncatula*. *Theor. Appl. Genet.* https://doi.org/10.1007/s00122-009-1033-2 (2009).
7. Le Signor, C. *et al.* Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. *New Phytol.* **214**, 1597–1613 (2017).
8. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **9**, 29 (2013).
9. Stanton-Geddes, J. *et al.* Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS ONE* **8**, 1–9 (2013).
10. Whan, A. P. *et al.* GrainScan: A low cost, fast method for grain size and colour measurements. *Plant Methods* **10**, 23 (2014).
11. Jones, D. B. Factors for converting percentages of nitrogen in foods and feeds into percentages of protein. *Br. Food J.* https://doi.org/10.1108/eb011242 (1932).
12. Zhao, F. J., Bilsborrow, P. E., Evans, E. J. & McGrath, S. P. Nitrogen to sulphur ratio in rapeseed and in rapeseed protein and its use in diagnosing sulphur deficiency. *J. Plant Nutr.* https://doi.org/10.1080/01904169709365273 (1997).
13. Dubousset, L., Etienne, P. & Avice, J. C. Is the remobilization of S and N reserves for seed filling of winter oilseed rape modulated by sulphate restrictions occurring at different growth stages?. *J. Exp. Bot.* https://doi.org/10.1093/jxb/erq233 (2010).

14. Box, G. E. P. & Cox, D. R. An analysis of transformations. *J. R. Stat. Soc. Ser. B* https://doi.org/10.1111/j.2517-6161.1964.tb00553.x (1964).
15. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).
16. Liu, X., Huang, M., Fan, B., Buckler, E. S. & Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genet.* **12**, e1005767 (2016).
17. Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. & Holmes, I. H. JBrowse: A next-generation genome browser. *Genome Res.* https://doi.org/10.1101/gr.094607.109 (2009).
18. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): High-performance genomics data visualiza-tion and exploration. *Brief. Bioinform.* https://doi.org/10.1093/bib/bbs017 (2013).
19. Merkle, T. Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell Rep.* https://doi.org/10.1007/s00299-010-0928-3 (2011).
20. Koo, Y. J. *et al.* Overexpression of salicylic acid carboxyl methyltransferase reduces salicylic acid-mediated pathogen resistance in *Arabidopsis thaliana*. *Plant Mol. Biol.* https://doi.org/10.1007/s11103-006-9123-x (2007).
21. Branca, A. *et al.* Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E864–E870 (2011).
22. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* https://doi.org/10.1086/519795 (2007).
23. Yang, J., Fu, M., Ji, C., Huang, Y. & Wu, Y. Maize oxalyl-coa decarboxylase1 degrades oxalate and affects the seed metabolome and nutritional quality[open]. *Plant Cell* https://doi.org/10.1105/tpc.18.00266 (2018).
24. Lepiniec, L. *et al.* Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **57**, 405–430 (2006).
25. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* https://doi.org/10.1038/nmeth.4197 (2017).
26. Bordat, A. *et al.* Translational genomics in legumes allowed placing in silico 5460 unigenes on the pea functional map and identi- fied candidate genes in *Pisum sativum* L. *G3 (Bethesda).* **1**, 93–103 (2011).
27. Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M. & Sams, C. E. quantitative trait loci for seed protein and oil concentra-tion, and seed size in soybean. *Crop Sci.* **45**, 2015–2022 (2005).
28. Johnson, W. C. *et al.* Association of a seed weight factor with the phaseolin seed storage protein locus across geno-types, environ-ments, and genomes in *Phaseolus-Vigna* spp. *J. Agric. Genomics* **2** (1996).
29. Lucas, M. R. *et al.* Association studies and legume synteny reveal haplotypes determining seed size in *Vigna unguiculata*. *Front. Plant Sci.* **4** (2013).
30. D'Erfurth, I. *et al.* A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytol.* **196**, 738–751 (2012).
31. Noguero, M. *et al.* DASH transcription factor impacts *Medicago truncatula* seed size by its action on embryo morphogenesis and auxin homeostasis. *Plant J.* **81**, 453–466 (2015).
32. Yang, S. *et al.* The endosperm-specific ZHOUPI gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. *Development* **135**, 3501–3509 (2008).
33. Garcia, D., Fitz Gerald, J. N. & Berger, F. Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* **17**, 52–60 (2005).
34. Orozco-Arroyo, G., Paolo, D., Ezquer, I. & Colombo, L. Networks controlling seed size in Arabidopsis. *Plant Reprod.* **28**, 17–32 (2015).
35. Ohto, M.-A., Fischer, R. L., Goldberg, R. B., Nakamura, K. & Harada, J. J. Control of seed mass by APETALA2. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3123–3128 (2005).
36. Saxena, K. B., Faris, D. G., Singh, U. & Kumar, R. V. Relationship between seed size and protein content in newly developed highprotein lines of pigeonpea. *Plant Foods Hum. Nutr.* **36**, 335–340 (1987).
37. Poeta, F., Borrás, L. & Rotundo, J. L. Variation in seed protein concentration and seed size affects soybean crop growth and development. *Crop Sci.* **56**, 3196–3208 (2016).
38. Righetti, K. *et al.* Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *Plant Cell* **27**, tpc.15.00632 (2015).
39. Ambika, S., Manonmani, V. & Somasundaram, G. Review on effect of seed size on seedling vigour and seed yield. *Res. J. Seed Sci.* **7**, 31–38 (2014).
40. Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. A SUPER powerful method for genome wide association study. *PLoS ONE.* **9**, e107684 (2014).
41. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* https://doi.org/10.1038/ng.2314 (2012).
42. Wen, Y. J. *et al.* Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* https://doi.org/10.1093/bib/bbw145 (2018).
43. Liu, X., Huang, M., Fan, B., Buckler, E. S. & Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* https://doi.org/10.1371/journal.pgen.1005767 (2016).
44. Zhang, Y.-M., Jia, Z. & Dunwell, J. M. Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* **10**, 1–6 (2019).
45. Xu, Y. *et al.* Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant Sci.* **9**, 1–10 (2018).
46. Jaiswal, V. *et al.* Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agro- nomic traits in common wheat (*T. aestivum* L.). *PLoS One* **11**, 1–25 (2016).
47. Li, C., Fu, Y., Sun, R., Wang, Y. & Wang, Q. Single-locus and multi-locus genome-wide association studies in the genetic dissectionof fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Front. Plant Sci.* **9**, 1–16 (2018).
48. Steffens, A., Bräutigam, A., Jakoby, M. & Hülskamp, M. The beach domain protein spirrig is essential for arabidopsis salt stress toler- ance and functions as a regulator of transcript stabilization and localization. *PLoS Biol.* https://doi.org/10.1371/journal.pbio.10021 88 (2015).
49. Maldonado-Bonilla, L. D. Composition and function of P bodies in *Arabidopsis thaliana*. *Front. Plant Sci.* **5**, 1–11 (2014).
50. Saedler, R., Jakoby, M., Marin, B., Galiana-Jaime, E. & Hülskamp, M. The cell morphogenesis gene SPIRRIG in Arabidopsis encodes a WD/BEACH domain protein. *Plant J.* https://doi.org/10.1111/j.1365-313X.2009.03900.x (2009).
51. Steffens, A., Jakoby, M. & Hülskamp, M. Physical, functional and genetic interactions between the beach domain protein spirrig and lip5 and skd1 and its role in endosomal trafficking to the vacuole in arabidopsis. *Front. Plant Sci.* **8**, 1–13 (2017).
52. Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S. & Gehring, C. The RNA-binding protein repertoire of *Arabidopsis thaliana*. *Sci. Rep.* **6**, 1–13 (2016).
53. Sugliani, M., Brambilla, V., Clerkx, E. J. M., Koornneef, M. & Soppe, W. J. J. The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in *Arabidopsis*. *Plant Cell* **22**, 1936–1946 (2010).
54. Kim, W. Y. *et al.* The Arabidopsis U12-type spliceosomal protein U11/U12-31K is involved in U12 intron splicing via RNA chap- erone activity and affects plant development. *Plant Cell* **22**, 3951–3962 (2010).

55. Szakonyi, D. & Duque, P. Alternative splicing as a regulator of early plant development. *Front. Plant Sci.* **9**, 1–9 (2018).
56. Ripley, B. *et al.* Package 'MASS' (Version 7.3-51.4). *Cran-R Proj.* (2019).
57. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* https://doi.org/10.2307/23337 09 (1965).
58. Bonhomme, M. *et al.* High-density genome-wide association mapping implicates an F-box encoding gene in *Medicago truncatula* resistance to *Aphanomyces euteiches*. *New Phytol.* **201**, 1328–1342 (2014).
59. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* https://doi.org/10.1089/omi.2011.0118 (2012).

Acknowledgements

Seeds of HAPMAP accessions used in this study were obtained from the *Medicago* HAPMAP germplasm resource center (http://www.Medicagohapmap.org/hapmap/germplasm), and from the Genetic Improvement and Adap-tation of Mediterranean and tropical plants unit (AGAP, INRA Montpellier, JM Prosperi). This research was conducted in the framework of the regional programme "Objectif Végétal, Research, Education and Innova- tion in Pays de la Loire", supported by the French Region Pays de la Loire, Angers Loire Métropole and the European Regional Development Fund. This study was also supported by the China Scholarship Council (CSC No. 201704910863) from the Ministry of Education of P.R. China. Finally, authors would like to thank Jaiana Malabarba for manuscript proofreading.

Author contributions

Z.C., V.L. and J.V. performed experiments. Z.C., V.L., Y.S., C.L.S., Y.K. and J.V. analysed data. Z.C., Y.K., C.L.S. and J.V. performed statistical analyses. Z.C., Y.K. and J.V. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-83581-7.

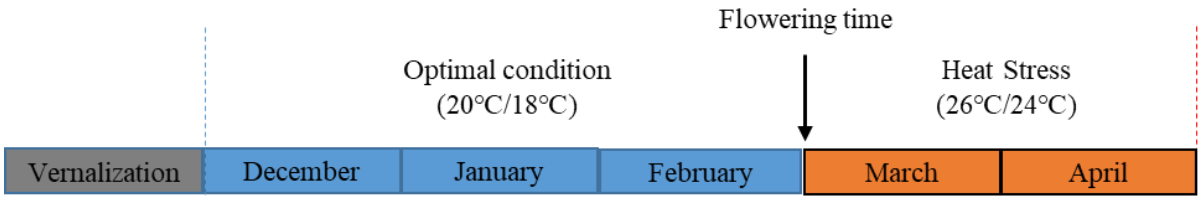
Correspondence and requests for materials should be addressed to J.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3. Production of mature seeds from the 200 *Medicago* HapMap accessions under optimal and heat stress conditions

In the previous section, mature seeds of 162 accessions were already produced in climate room from the Shanghai Center for Plant Stress Biology (PSC) using standard conditions. In the following study, 200 natural accessions selected from *Medicago truncatula* HapMap collection were grown in the greenhouse facility of the 'Institut de Recherche en Horticulture et Semences' in Angers in both optimal and heat stress conditions but using the same light intensity, photoperiod. Triplicates of the 200 accessions were grown in control conditions using minimal temperature of 20°C day /18°C night as part of the ANR REGULEG Project (no. ANR-15-CE20-0001). As part of my PhD project, triplicates of the 200 accessions were gown in heat stress conditions. To produce heat-stressed mature seeds, we first followed the same procedure as the control plants with a seedling treatment by vernalization (*i.e.* extended cold period for 2 weeks at 6°C) to reduce flowering time differences between accessions, then seedlings were moved in greenhouse under optimal conditions (as control plants) (20°C day /18°C night) with 16 hours light photoperiod until the flowering stage (*i.e.* apparition of five flowers). At the flowering stages, plants were moved to a nearby greenhouse room with same photoperiod and light conditions but with minimal temperatures of 26°C day /24°C night until seed maturity, corresponding to heat stress conditions for *Medicago*. During the heat stress treatment, even if greenhouse conditions were controlled in term of humidity, light and minimal temperature, warmer winter and spring than expected generated a stronger stress than planned due to high external temperatures (Figure 3.1). According to the temperature records in the dedicated greenhouse room, the average temperature in March and April were about 27°C during days with maximum at 32°C. These high temperatures occurring at flowering time strongly impacted the final seed yield. Eventually, mature seeds of 199 accessions were collected from heat stress conditions, but severe loss of yield was observed for most of the accessions. In consequence the limited seed number obtained from each plant did not allow us to perform all the intended physiological analyses.



	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
Avg	18.9	17.9	17.5	15.1	22.3	19.3	26.9	23.7	26.9	24
Min	14.9	14.4	11.5	11.5	17.7	16.9	22.3	21.3	21	22.6
Max	23.9	20.4	26.4	18.5	29.1	24.8	32.6	27.0	38.2	26.9

Figure 3.1 Records of temperature during production of *Medicago truncatula* HapMap population from heat stress conditions in greenhouse. The phase indicated in blue is the plant vegetative growth phase under optimal condition (20°C/18°C), while the phase indicated in orange is the heat treatment (26°C/24°C) occurring during seed development phase. The flowering time represents appearance of five flowers for plants of most accessions.

4. Genome-wide association studies of seed performance traits in response to heat stress in *Medicago truncatula* reveal *MIEL1* as a regulator of seed germination plasticity

(This section was summarized in a manuscript submitted to Frontiers in Plant Science, currently accepted.)

Zhijuan Chen, Joseph Ly Vu, Benoit Ly Vu, Julia Buitink, Olivier Leprince, Jerome Verdier*

Institut Agro, Univ Angers, INRAE, IRHS, SFR 4207 QuaSaV, 49000 Angers, France

*** For correspondence: jerome.verdier@inrae.fr**

Key words: GWAS, *Medicago truncatula*, heat stress, seed germination, plasticity

Abstract

Legume seeds are important nutrition source to provide proteins, minerals and vitamins for human and animal diets and represent a keystone for food security. With the climate change and global warming, production of grain legumes faces new challenges with respect to seed vigour traits, that allow fast and homogenous establishment of the crop in a wide range of environments. These seed performance traits are regulated during seed maturation and are under the strong influence of the maternal environment. In this study, we used 200 natural *Medicago truncatula* accessions, a model species of legumes, which were grown in optimal and under a moderate heat stress (26°C) during seed development and maturation. This moderate stress applied at flowering onwards impacted seed weight, germination capacity and seed longevity. Genome-wide association studies (GWAS) were performed to identify putative loci or genes that are involved in regulating seed traits and their plasticity in response to heat stress. We identified numerous significant QTNs and potential candidate genes involved in

regulating these traits under heat stress by using postGWAS analyses combined with transcriptomic data. Out of them, *MtMIEL1*, a RING-type zinc finger family gene, was shown to be highly associated with germination speed in heat-stressed seeds. In *Medicago*, we highlighted that *MtMIEL1* was transcriptionally regulated in heat-stressed seed production, and that its expression profile was associated with germination speed in different *Medicago* accessions. Finally, a loss-of-function analysis of the *Arabidopsis MIEL1* ortholog revealed its role as regulator of germination plasticity of seeds in response to heat stress.

Introduction

Legume is an economically important crop family including many plant species such as soybean, pea, common bean and chickpea. *Medicago truncatula* is a model plant of legumes originating from Mediterranean region (Barker *et al.*, 1990), which has been intensively studied for legume research. Grain legumes provide abundant proteins, minerals and other nutrients for human and animal diets and play vital role for global food security. However, climate change threatens crop production by causing reduced yield and loss of product quality. In the context of global warming, legume seed production suffers from environmental stresses, including heat stress, and legume crops need to be improved towards a higher phenotypic plasticity (Vadez *et al.*, 2012; Scheelbeek *et al.*, 2018). Indeed, while the local adaptation of a genotype is genetically determined under certain environmental conditions (Tognetti *et al.*, 2019), the phenotypic plasticity is the ability to generate different phenotypes according to the environment (Valladares *et al.*, 2006). This variation is created by interplay of genetic and environmental factors. Understanding of the genetic basis of local adaptation and phenotype plasticity is highly relevant in our current climate change context. Heat stress affects the proper development of female and male gametophytes, leading to impaired double fertilization and decreased seed number (reviewed in Liu *et al.*, 2019b). Also heat stress during early embryogenesis was shown to reduce grain yield in soybean and mungbean (Siebers *et al.*, 2015; Patriyawaty *et al.*, 2018). During seed development, maturation was shown to affect seed vigor. Seed vigor is a composite term that includes homogeneous and rapid germination and seedling establishment under a range of contrasted environmental (*i.e.* stress) conditions (Finch-Savage and Bassel, 2016). The capacity to avoid deterioration during storage (*i.e.* longevity) is also seen as a vigor trait as it directly impacts first speed of germination then viability upon sowing (Leprince *et al.*, 2017). In *M. truncatula* (Verdier *et al.*, 2013; Righetti *et al.*, 2015), the

different vigor traits are acquired sequentially, from seed filling until the late phase of seed maturation (reviewed in Leprince *et al.*, 2017). So far, genetic determinants of seed vigor in *Medicago* have been explored, mostly by QTL identification using several populations of recombinant inbred lines resulting from crosses between contrasting accessions (Vandecasteele *et al.*, 2011). These studies led to the identification of several key regulatory genes of the late maturation phase such as *MtABI5* (Zinsmeister *et al.*, 2016) and *MtHSFA9* (Zinsmeister *et al.*, 2020a). In *Medicago* (Righetti *et al.*, 2015), like many other species (Finch-Savage and Bassel, 2016; Penfield and MacGregor, 2017), seed vigor is also drastically affected by environmental conditions during seed development. This highly plastic response from the offpring to environment is considered as a bet-hedging strategy to ensure dissemination of the species. In this respect, one of the most studied germination vigor traits is dormancy (for review Penfield and MacGregor, 2017). In legume seeds, such as *M. truncatula* seeds, we distinguish two types of dormancy, which are the physical and physiological dormancies (according to definition of Baskin and Baskin, 2004). Physical dormancy is mainly controlled by the seed coat permeability, which prevents seed imbibition. However, physiological dormancy is regulated by the embryo and endosperm molecular signals, via the ratio of abscisic acid (ABA), acting as germination repressor and gibberellic acid (GA), allowing germination. For example, in *M. truncatula*, a slight increase in the seed coat properties regulating seed imbibition and physical dormancy was observed when plants were grown in 35°C/ 15°C compared to 25°C/ 15°C conditions (Renzi *et al.*, 2020). While a wide spread of germination via a decrease in germination speed or a delay of germination until favourable conditions is advantageous for wild species dissemination, it is not a desirable trait for crops. Furthermore, the plastic response of the germination of seeds produced under environmental conditions is also dependent on complex GxE interactions of the regulation of physiological dormancy involving zygotic and maternal tissues (Penfield and MacGregor, 2017; Awan *et al.*, 2018; Geshnizjani *et al.*, 2019; Chen *et al.*, 2020; Renzi *et al.*, 2020), and the dynamic balance between ABA and GA is poorly understood and likely to be species-dependent (Penfield and MacGregor, 2017; Chen *et al.*, 2020).

In recent years, genome-wide association study (GWAS) has been widely performed for the association mapping between genetics and agronomic traits in order to identify causal loci using population of natural accessions. Many new statistical models to compute the association mapping have been developed from initially single-locus analyses to recent multi-locus analyses including the fixed and random model circulating probability unification (FarmCPU)

(Liu *et al.*, 2016), which improved the statistic power to control false positives and reduce computing time (for review Tibbs Cortes *et al.*, 2021). In *Medicago truncatula*, a haplotype map (HapMap) population was selected based on their geographical origins and genomic diversity and resequenced using next-generation sequencing technologies to identify single nucleotide polymorphisms (SNP) (Stanton-Geddes *et al.*, 2013). The *Medicago* HapMap population, finally, comprises 226 natural accessions characterized by 4.8 million SNP. This collection has been used to study different aspects of *Medicago* biology such as different abiotic stress on vegetative part with salt stress (Kang *et al.*, 2019) and drought stress (Kang *et al.*, 2015), but also more specifically to seeds with seed nutritional content (Chen *et al.*, 2021a) and physical seed dormancy (Renzi *et al.*, 2020).

In this study, we used the *Medicago* HapMap collection to identify putative causal genes/loci associated with the plasticity of germination performance traits of seeds produced under heat stress conditions. We performed genome-wide association studies of seed weight, seed longevity and seed germination speed and homogeneity using 200 accessions from the *Medicago truncatula* HapMap collection via FarmCPU algorithm. PostGWAS analyses and RNA-seq data were used to refine our candidate gene lists related to different seed traits. A candidate gene, *MtMIEL1*, involved in the germination plasticity of seeds produced under heat stress was identified in *M. truncatula* and functionally validated in *A. thaliana*.

Materials and methods

Medicago population and plant growth conditions

From the *Medicago truncatula* HapMap project (http://www.medicagoHapMap.org/HapMap/germplasm), 200 accessions were selected and grown in six replicates in the greenhouse, where light intensity, photoperiod and minimal temperature were controlled. All plants were first produced under optimal conditions at 20°C/18°C day/night with 16 hours light photoperiod until the flowering stage as described in Vandecasteele *et al.* (2011). After apparition of five flowers, triplicate of plants for each accessions were maintained under these optimal conditions and the other triplicates were grown under heat stress conditions with 26°C/24°C day/night temperature but with the same light conditions. Mature seeds from 199 accessions were collected at pod abscission and further dried at 20°C in 44% relative humidity (RH). Seeds were stored hermetically at room temperature before use.

Phenotyping seed traits

The individual seed weights of 199 HapMap accessions were calculated from the average of total seed weight per plant. Number of mature seeds per plant were counted using a seed counter (Pfeuffer model Contador) and total seed weights were determined using a precision balance. For each accession and replicate, the average individual seed weight was calculated by dividing total seed weight per plant by the seed number per plant, providing an accurate estimate of the individual seed weight. Before longevity and any germination experiments, seeds were first scarified to avoid artefacts due to physical dormancy. Logistics and greenhouse conditions obliged us to optimize the number of accessions to allow assessment of germination and longevity traits. We used 151 HapMap accessions (from seed produced in optimal and stress conditions) that were stored at 35°C 60% RH for 204 days. This period corresponded to the average of ageing time when 50% of seeds still germinated (also called P50) for all accessions grown in control conditions. Triplicates of 50 seeds were retrieved from 204 days of storage and imbibed at 20°C in the dark and the percentage of viable seeds was calculated as percentage of germinated seeds after six days of imbibition in 9ml water in 14.5cm petri dishes containing a Whatman filter paper. 112 *Medicago* HapMap accessions were used to assay germination. Triplicates of 50 seeds were imbibed in 5ml of water in 5cm Petri dish containing one Whatman No1 filter paper at 15°C in the dark. Germinated seeds and speed of germination were monitored automatically for control seed lot using the phenotyping platform PHENOTIC (SFR QUASAV, Angers) (Benoit *et al.,* 2014) and manually for the stressed-seed lot by counting germinated seeds (*i.e.* protruding radicles> 1 mm) every four hours. Germination speed was calculated from the sigmoidal regression of each accession as the averaged time to reach 50% germination (T50). Germination homogeneity was calculated as the time difference between 80% (T80) and 20% (T20) germination (*i.e.* T80-T20). Finally, the phenotypic plasticity index of all seed traits was calculated based on the following formula: PLAS= (T_St – T_Ct)/T_Ct, where T_St is the mean value of the trait under heat stress conditions and T_Ct is the mean trait value under control conditions.

Correlation analysis

Correlations between traits were analyzed using the 'rcorr' function of the 'Hmisc' package (v4.4-1, Harrell Jr *et al.* 2020) in R. A global correlation matrix was performed using Pearson correlation coefficient and we selected a p-value threshold of 0.05 for statistical significance.

Normalization of phenotypic data

In order to carry out the genome-wide association studies, we checked and transformed, when necessary, our phenotypic data to reach distribution normality. The Shapiro-Wilk test was performed to test the distribution states of all phenotypic traits and phenotypic data were transformed to normal distributions using Box-Cox power transformation procedure (Box and Cox, 1964) using adapted lambda values calculated for each trait. The Shapiro-Wilk tests and the Box-Cox transformations were carried out using the 'MASS' package (v7.3-51.5, Venables and Ripley, 2002) available in R.

Genome-wide association analysis

Identification of single nucleotide polymorphisms (SNP) was obtained by whole-genome sequencing of the *Medicago* HapMap accessions selected in the *M. truncatula* HapMap project (Stanton-Geddes *et al.*, 2013). Using the *Medicago* genome version 5 (Mtv5, Pecrix *et al.*, 2018), more than 4.8 million SNP locations were identified and genotyped in the HapMap accessions. This 4.8 million SNP genotypic dataset was used in combination with the HapMap population structure (described in Bonhomme *et al.* 2014) and the normalized phenotypic dataset regarding seed performances. The multi-locus model FarmCPU (Fixed and random model Circulating Probability Unification, Liu *et al.*, 2016) was used to perform association analyses as described in Chen *et al.* (2021b) with p-value threshold set to 1%. The Quantile-quantile (QQ) and Manhattan plots were generated by FarmCPU package available in R.

Post-GWAS analyses

The PLINK algorithm (Purcell *et al.*, 2007) was used to identify correlated SNP and to correct for the linkage disequilibrium (LD) using the "clump" function. The following options of PLINK were used: 'clump-kb 30' and 'clump-r2 0.7', which represent the range of analyzed genomic region (± 30kb) and the R-squared threshold (0.7) to identify correlated SNP.

Enrichment analyses of Mapman functional classes of putative causal genes related to different seed performance traits were performed using the Clusterprofiler package (Yu *et al.*, 2012) using hypergeometric test with Benjamini-Hochberg correction (q-values) and available in R. Mapman functional classes were obtained from *Medicago* annotated proteins using Mercator v.4 (Schwacke *et al.*, 2019).

Transcriptomic data

The expression data of *Medicago truncatula* during seed development under optimal and heat stress conditions were obtained from Chen *et al.* (2021a) and raw data were stored on NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE160725). The differentially expressed genes (DEG) between optimal and heat stress conditions at the different seed developmental stages were identified using ImpulseDE2 (Fischer *et al.*, 2018) for embryo and endosperm and DEseq2 (Love *et al.*, 2014) for seed coat. DEG threshold was set as adjusted p-values below 5% following the Benjamini-Hochberg procedure to control false discovery rate (FDR) as described in Chen *et al.* (2021a).

RNA extraction and qRT-PCR

Total RNAs were extracted from two replicates of about 30 dry mature seeds, 24h-imbibed and 48h-imbibed (10°C) seeds of *Medicago* reference genotype A17 that were produced in optimal (20°C/18°C, 16-h photoperiod) and heat stress (26°C/24°C, 16-h photoperiod) conditions. Simultaneously RNA extractions were also performed on dry mature seeds in triplicates of four natural *Medicago* HapMap accessions (*i.e.* HM170, HM185, HM279 and HM314) produced under heat stress condition (26°C/24°C, 16-h photoperiod). HapMap genotypes were chosen based on their germination speed, with two belonging to the slowest germination set and two belonging to the fastest germination set. All RNA extractions were performed using Macherey-Nagel NucleoSpin® RNA Plant and Fungi kit following the Alfalfa seeds protocol described in the manufacturers' instructions. Total RNA were quantified using a Nanodrop spectrophotometer ND-1000 (NanoDrop Technologies), then treated with RNase-free DNase I (Thermo Fisher Scientific Inc.). Reverse transcriptions were performed using the iScript™ RT Supermix (Bio-Rad Laboratories, Inc.) from 1µg of DNAse-treated RNA. cDNA were

quantified with SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad Laboratories, Inc.) using a CFX96 Touch quantitative Real-Time PCR (qRT-PCR) Detection System (Bio-Rad Laboratories). The primers that were used for qRT-PCR are provided in Table S7. *MtMIEL1* primers were designed on Primer 3 website (https://bioinfo.ut.ee/primer3/). *MtTCTP* was used as reference gene (Verdier *et al.*, 2008; Zinsmeister *et al.*, 2020a). The relative expression levels were normalized according to the 2^-ΔCt method.

Arabidopsis T-DNA insertional mutants and seed germination assays

The T-DNA insertional *miel1* mutant (Salk_041369) from a Columbia-0 (Col0) background was obtained from the NASC germplasm collection. The primers used for isolation of T-DNA homozygote mutants were generated from the T-DNA Primer Design website (http://signal.salk.edu/tdnaprimers.2.html) and are provided in Table S7. *Arabidopsis thaliana* plants (Col0 and *miel* mutants) were grown under standard conditions (20°C/18°C, 16-h photoperiod) in growth chamber. At flowering time, half of plants were kept at control condition (20°C/18°C, 16-h photoperiod) and half were moved under heat stress condition (28°C/26°C, 16-h photoperiod). Mature seeds produced in both conditions were harvested and dried for three days at 44% relative humidity and 20°C. The dry seeds were stored at -20°C before germination test. Three biological replicates of about 100 seeds obtained from three independent *miel1* and wild-type (Col0) plants were used for germination assays. Freshly harvested seeds were imbibed in 1ml water in 3cm petri dishes containing Whatman No1 filter paper at 20°C with 16-h photoperiod. To release dormancy, freshly harvested seeds were stratified at 4°C for 72 hours in dark then transferred to 20°C with 16-h photoperiod for germination.

Results

Assessing seed performances in response to heat stress in *Medicago* HapMap collection

To evaluate the impact of heat stress on seed yield and vigor, 200 *Medicago truncatula* accessions from the HapMap collection were grown in triplicate in optimal (20°C/18°C) and supra-optimal temperature (*i.e.* 26°C/24°C) conditions by applying a constant but moderate

heat stress from flowering until pod abscission. After harvest and moisture content equilibration at 44% RH, we observed a significant decrease in seed yield from plants grown in the heat stress conditions. Across the 200 *Medicago* accessions, the average pod and seed numbers per plant in optimal conditions were 117 and 658 respectively, in contrast to 20 pods and 104 seeds in average from plants grown under heat stress (Figure 1). In addition, an overall 35% decrease in seed weight was observed in accessions produced in heat stress conditions (Figure 1, phenotypes named WEIGHT_C for optimal and WEIGHT_H for heat stress conditions).

The limited seed number produced under heat stress conditions for some accessions directly impacted the number of accessions available for phenotyping seed performance. For instance, HM059 accession did not produce enough seeds and was discarded. Regarding seed germination and longevity performances, we phenotypically characterized seed longevity and seed germination of mature seeds produced from 151 and 112 accessions, respectively. First, regarding seed longevity, we observed an overall 32% decrease in seed survival percentage after 204 days of artificial ageing across all accessions when seeds were produced under heat stress (Figure 1, phenotypes named G204DA_C for optimal and G204DA_H for heat stress conditions). Second, related to seed germination, we observed that about 100% of seeds germinated six days after imbibition, no matter if they were produced in optimal or stress conditions. To assess the impact of heat stress on seed vigor, we extracted two germination characteristics: the germination speed (T50, corresponding to the time to reach 50% of germination) and the germination homogeneity (T80T20, duration between 80% and 20% of germination). Germination speed and homogeneity across the population were positively impacted by the heat stress during seed production. Indeed, seeds produced in heat stress conditions displayed an overall tendency to germinate faster and more homogeneously than those produced in optimal conditions (Figure 1, phenotypes named T50_C and T80T20_C for optimal and T50_H and T80T20_H for heat stress conditions).

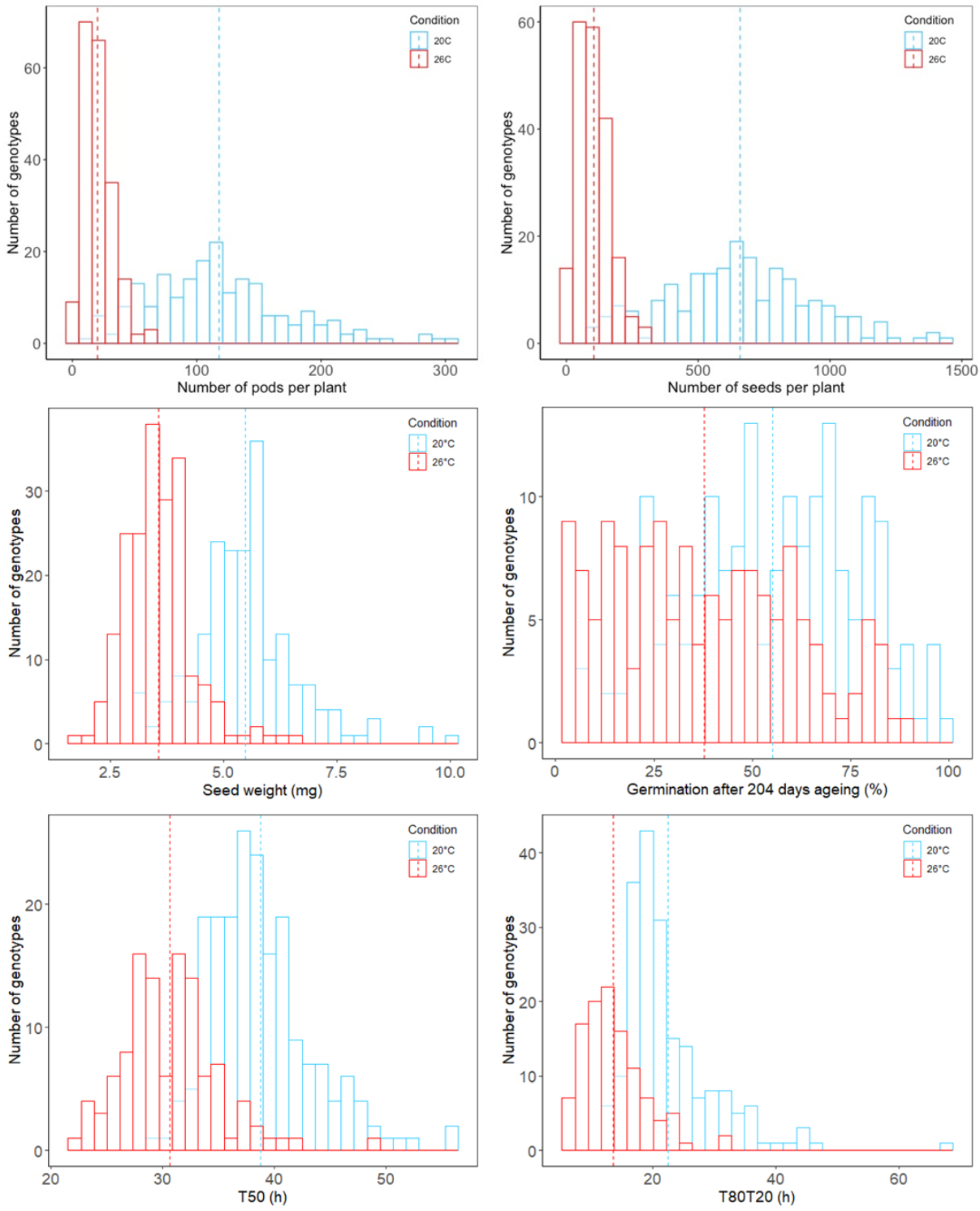


Figure 1. Distribution histograms of analyzed phenotypic data regarding seed traits across the *M. truncatula* HapMap accessions and grown under optimal (blue) and heat stress (red) conditions. Average values across the entire HapMap population are represented in dotted lines.

Even if the overall tendency from all different accessions displayed an increase in seed germination performances and decrease in seed weight and longevity, it is noteworthy that the individual tendency of each accessions is more contrasted with some that did not follow the overall tendency. This reflected a high phenotypic plasticity within the HapMap population regarding these traits (Figure S1, Table S1). To assess phenotypic plasticity (PL) of each accession, we calculated the plasticity index of seed traits obtained in the two contrasted seed production conditions. These plasticity indexes reflected the ability of each accession to produce different phenotypes according to the maternal environment (Table S1, phenotypes named WEIGHT_PL, G204DA_PL, T50_PL and T80T20_PL).

To determine if seed performance traits measured in different growth conditions were correlated, we performed a correlation analyses among them using Pearson coefficient correlation (Table 1). First, we observed a strong positive correlation (0.79) between weight of seeds produced in optimal and heat stress conditions, suggesting that seed weight is genetically determined in HapMap accessions by the same set of genes in both conditions. Moreover, we observed a weak positive correlation (0.2) between seed weight and speed of germination for seeds produced under control conditions. The correlation was also found for seeds produced under heat stress but was much stronger (0.48). Many studies have documented that seed size is correlated with germination performance, with larger seeds exhibiting better seedling survival rate due to more seed reserve accumulated during seed filling to supply embryo with sufficient energy during germination (reviewed in Finch-Savage and Bassel, 2016). However, the plasticity response of both traits were not correlated, suggesting that there exist different processes regulating seed filling and acquisition of germination performance in response to heat stress. Finally, we observed positive correlations between germination phenotypes measured during heat stress and plasticity indexes, which suggested that mechanisms controlling germination of seeds produced under heat stress could be similar to those controlling their plasticity. Surprisingly, any significant correlations were identified between geographical location of accessions and seed longevity/germination.

Table 1. Correlation matrix between all *Medicago* seed traits and climatic data. Pearson correlation coefficients above 0.2 with a p-value below 5% are indicated in red color. Pearson correlation coefficients above -0.2 with a p-value below 5% are indicated in green color. Longitude is indicated by degrees from west to east with negative

and positive values respectively. Latitude is also indicated by degrees from south to north with negative and positive values respectively.

TRAITS		WEIGHT_C	WEIGHT_H	WEIGHT_PL	T50_C	T50_H	T50_PL	T80T20_C	T80T20_H	T80T20_PL	G204A_C	G204A_H	G204A_PL
SEED WEIGHT	WEIGHT_C												
	WEIGHT_H	0.79											
	WEIGHT_PL	-0.35	0.29										
SEED GERMINATION SPEED	T50_C	0.20	0.16	-0.03									
	T50_H	0.41	0.48	0.11	0.16								
	T50_PL	0.12	0.19	0.09	-0.52	0.75							
SEED GERMINATION HOMOGENEITY	T80T20_C	0.09	0.14	0.10	0.69	0.18	-0.29						
	T80T20_H	0.17	0.26	0.17	-0.01	0.83	0.75	0.04					
	T80T20_PL	0.02	0.12	0.16	-0.34	0.64	0.79	-0.47	0.82				
SEED LONGEVITY	G204A_C	-0.06	-0.14	-0.12	-0.06	-0.24	-0.18	-0.06	-0.34	-0.23			
	G204A_H	-0.12	-0.11	0.04	-0.01	-0.53	-0.42	0.03	-0.52	-0.39	0.38		
	G204A_PL	-0.07	0.05	0.21	0.08	-0.40	-0.38	0.09	-0.36	-0.31	-0.33	0.58	
GEOGRAPHICAL LOCATION	Longitude	-0.06	-0.04	0.05	-0.11	0.07	0.14	-0.11	0.07	0.11	-0.16	-0.22	-0.12
	Latitude	-0.03	-0.09	-0.10	-0.05	0.03	0.05	-0.08	-0.04	-0.04	-0.14	-0.06	-0.08
	Altitude	-0.20	-0.27	-0.14	-0.01	-0.13	0.01	0.06	-0.04	0.04	0.24	0.10	0.00

Genome-wide association analyses of different seed traits in response to optimal and heat stress conditions and identification of putative causal genes

Following phenotypic characterization of HapMap accessions, we used the Box-Cox procedure (Box and Cox, 1964) to transform our phenotypic data that did not display normal distributions. Appropriate lambda values were estimated and sed to normalize our phenotypic data in order to validate the assumption of normality required to perform genome-wide association analyses. After this normalization step, the Shapiro-Wilk test was performed for each phenotype to verify that our phenotypic data reached the normal distribution (Figure S2). All lambda values, Shapiro-Wilk p-values and normalized phenotypic data are available in Table S1. However, three phenotypic traits did not pass the Shapiro-Wilk test (*i.e.* WEIGHT_C, G204DA_C and G204DA_H) but were conserved in subsequent analyses as they displayed acceptable fit to normal distribution based on their distribution histograms and their bell curves (Figure S2). Genome-wide association studies were performed on the 12 transformed seed phenotypic data using the Fixed and random model Circulating Probability Unification algorithm (FarmCPU, Liu *et al.*, 2016) combined with the *Medicago* HapMap population structure as covariable and the *Medicago* HapMap SNP genotypic dataset (described in Bonhomme *et al.*, 2014 and available at http://www.medicagoHapMap.org). From these association studies, we identified

sets of single nucleotide polymorphisms, called quantitative trait nucleotides (QTNs), statistically associated with the different seed traits. Manhattan and QQ (quantile-quantile) plots related to seed germination performances are provided in Figure 2 and those related to seed weight and longevity are provided in Figure S3. To facilitate visualization of these results, we generated 'gwas' files that contain all the statistical results of all the SNPs with respect to different seed traits, which allow visualization of significant QTNs on genome viewers such Integrative Genome Viewer (IGV, Thorvaldsdóttir *et al.*, 2013) or JBrowse (Skinner *et al.*, 2009) (provided as Table S2, S3, S4 and S5).

From these GWAS, we identified highly significant QTNs (p-values below 10^{-7}) associated with seed weight obtained from optimal (9 QTNs) and heat stress (20 QTNs) growing conditions, as well as 2 QTNs potentially involved in plasticity (Figure S3 and Table S6). Among these QTNs, one of them, MtrunA17Chr8_49244112, was identified to correlate with seed weight from optimal (WEIGHT_C) and heat stress (WEIGHT_H) conditions. Similarly, highly significant QTNs (p-values below 10^{-7}) were identified for seed germination traits: 2 and 1 QTNs regarding germination speed of seeds from control and heat stress conditions, respectively and 3 and 1 QTNs regarding germination homogeneity of seeds from control and heat stress conditions. Moreover, 2 QTNs were identified for plasticity of germination speed. We also observed common QTNs between germination traits located on chromosome 2: MtrunA17Chr2_6710478 common between T50_H and T50_PL and MtrunA17Chr2_18061650 common between T50_PL and T80T20_PL (Figure 2). Finally, regarding seed longevity, one highly significant QTN (p-values below 10^{-7}) was identified from seeds produced in heat stress conditions and five as correlated to longevity plasticity (Figure S3).

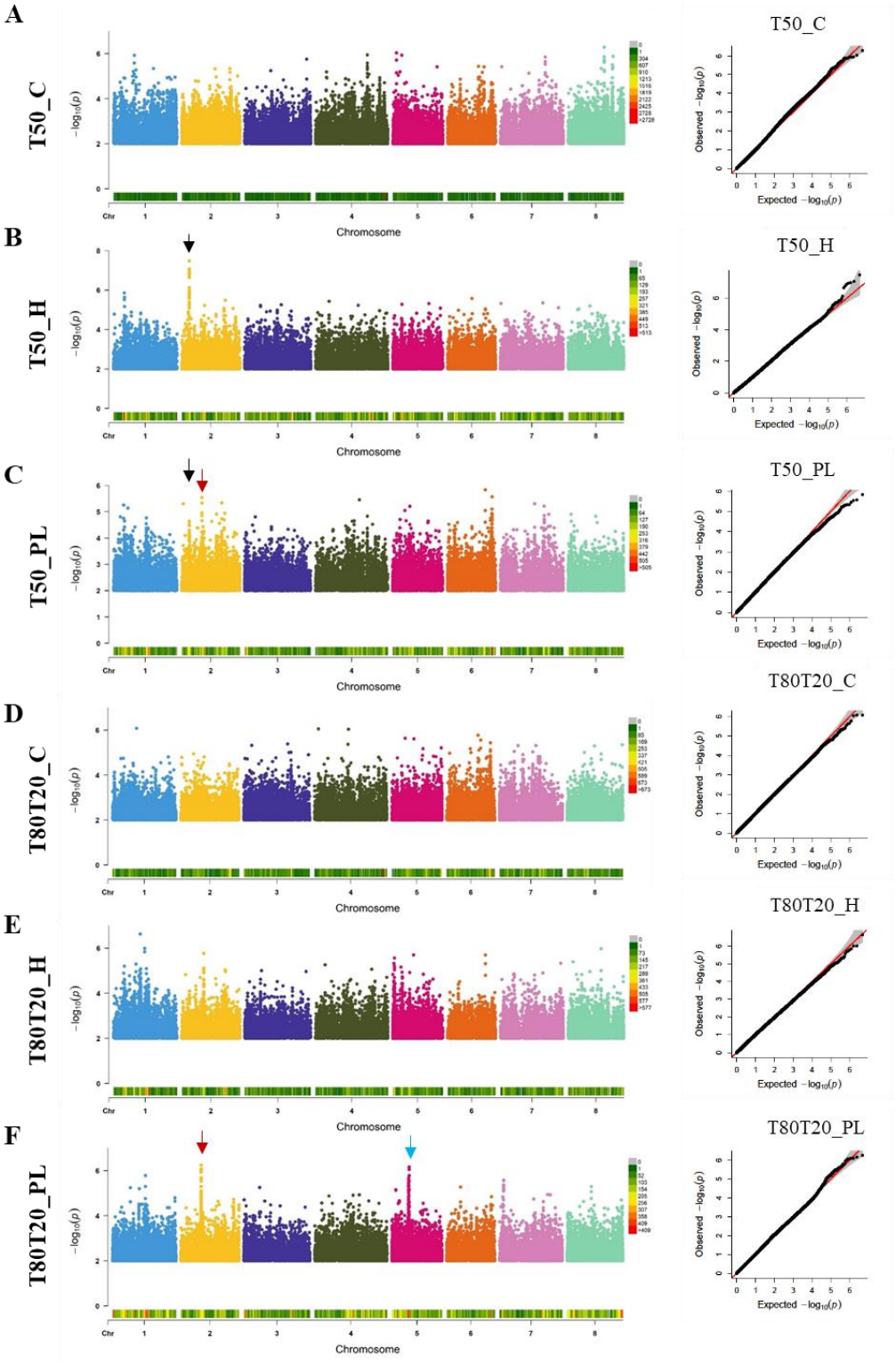


Figure 2. Manhattan plots and the corresponding Q-Q plots from GWAS results regarding seed germination speed (T50) (A, B and C) and germination homogeneity (T80T20) (D, E and F). Black arrows indicate the common QTN associated with both T50_H and T50_PL corresponding to MtrunA17_Chr2g0286331 gene. Red arrows indicate the common QTN associated with both T50_PL and T80T20_PL corresponding to MtrunA17_Chr2g0300261 gene. Blue arrow indicates the highly significant QTNs located in chromosome 5.

In order to precisely pinpoint putative causal genes associated with significant QTNs, we identified from all surrounding SNPs located around QTNs, which ones showed high correlations due to linkage disequilibrium (LD) and could be linked to the phenotype. Using PLINK algorithm (Purcell *et al.*, 2007), we performed genome-wide correlations of significant QTNs (p-values$<10^{-5}$) with surrounding correlated SNPs with the threshold of 0.7 ($r^2>0.7$) and located in a range of ± 30kb, corresponding to 2-fold the average LD decay in the HapMap population (Branca *et al.*, 2011). As a result, we identified 120 putative causal genes related to the 73 QTNs for seed weight, 106 putative causal genes related to the 59 QTNs for seed longevity (G204DA), 132 putative causal genes related to the 74 QTNs for germination speed (T50) and 109 putative causal genes related to the 63 QTNs for germination homogeneity (T80T20) (Table S6). From these lists of candidate genes identified, we performed gene set enrichment analyses (GSEA) of functional classes to determine which processes could be involved in the regulation of the different seed traits (Figure 3). Interestingly, we identified significant enrichments of functional classes related to 'HMG-CoA reductase' for germination speed; 'secondary metabolism and chalcone synthase', 'subtilases', 'hormone metabolism of auxin and cytokinin' for germination homogeneity; 'RNA regulation of transcription' for seed weight; 'RNA regulation of transcription.general transcription', 'DNA synthesis/chromatin structure', 'secondary metabolism.phenylpropanoids' and 'O-methyltransferases' for seed longevity.

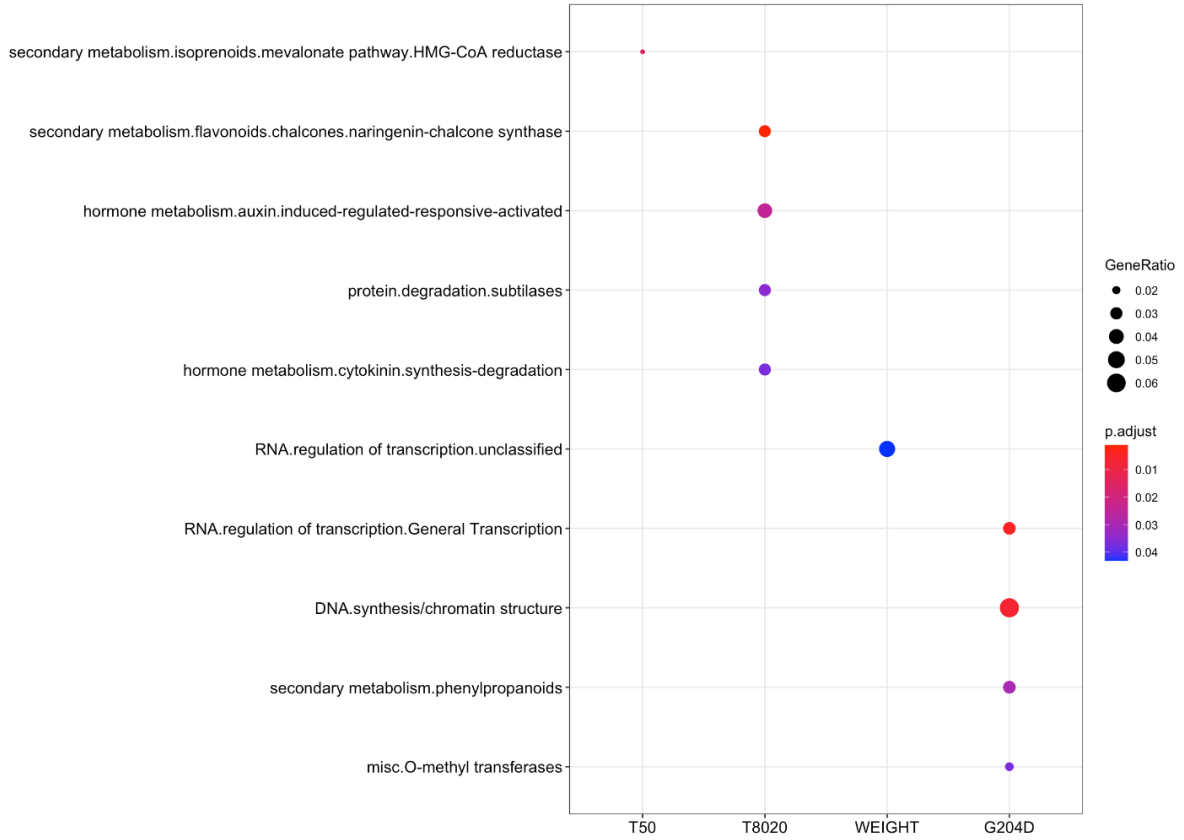


Figure 3. Gene set enrichment analysis (GSEA) of candidate gene lists obtained from GWAS with different seed traits: germination speed (T50), germination homogeneity (T80T20), seed weight (WEIGHT) and longevity (G204D). Clusterprofiler was used to perform a hypergeometric test using the Mapman functional terms, and the p-values were converted to false discovery rate (FDR) p.adjust-values as shown in colors, the red color being more significant than the blue color. The size of the dot represents the gene ratio between total gene number annotated in functional classes and the number of these genes present in your input list.

To reduce these gene lists and refine the identification of putative causal genes, we combined these dataset with gene annotations from *Medicago* Genome Version 5 (Pecrix *et al.*, 2018), transcriptomic data related to expression specificity in *M. truncatula* seeds (Chen *et al.*, 2021a) and transcriptomic data during maturation of seeds developed both in optimal and heat stress conditions (Chen *et al.*, 2021b; Table S6). In consequence, we highlighted some candidate genes related to seed weight such as MtrunA17_Chr8g0392741, encoding a phosphatidylethanolamine-binding protein, homologous to *MOTHER OF FT* (At1g18100, *MFT*) in *Arabidopsis*, which showed a highly significant association with WEIGHT_C (p-value = 8.10^{-20}) and WEIGHT_H (p-value =4.10^{-17}) and a strong differential expression during

M. truncatula seed development between optimal and heat stress production (Figure S3, indicated by red arrows and Table S6). Another example of candidate gene regulated to seed weight during heat stress conditions was MtrunA17_Chr5g0403261 gene, closely related to *Arabidopsis DA1* gene (AT1G19270), known to regulate plant organ size, including seed size (Li *et al.*, 2008) and also differentially expressed in the embryo between optimal and heat stress production. This gene did not display a single QTN with a high p-value but rather 9 significant QTNs (with p-values > 10^{-5}). (Figure S3, indicated by a green arrow). In the candidate gene list related to longevity of seeds from heat stress conditions and plasticity of longevity (Table S6), we found MtrunA17Chr5g0432251 gene, encoding an O-Methyltransferase associated with several significant QTNs (6 QTNs with p-values<10^{-5}) (Figure S3, indicated by a blue arrow).

In the subsequent part of this study, we decided to focus on candidate causal genes involved in germination speed/homogeneity. In chromosome 2 (Figure 2, indicated with black arrows), many QTNs associated with T50_H and T50_PL (> 20 QTNs with p-values<10^{-5}) were found in the MtrunA17_Chr2g0286331 gene, a member of a RING finger family containing a zinc-finger binding motif and the ortholog of *MYB30-INTERACTING E3 LIGASE 1* (*MIEL1*, At5g18650) in *Arabidopsis*. AtMIEL1 is a RING-type E3 ligase that plays a role in the proteosome pathway as a regulator of plant defense against bacteria (Marino *et al.*, 2013) and ABA (Lee and Seo, 2016). In our study, *MtMIEL1* also showed a differential expression in endosperm between seeds produced under optimal and heat stress conditions, making a good candidate gene for further analyses. In chromosome 2 (Figure 2, indicated with red arrows), we identified another genomic interval displaying many QTNs identified in both T50_PL and T80T20_PL, which were more difficult to precisely relate to a specific gene sequence. These QTNs were spread on three closely located genes: MtrunA17_Chr2g0300271 encoding a nodule glycin-rich peptide, MtrunA17_Chr2g0300291 encoding a DEAD-box ATP-dependent RNA helicase and MtrunA17_Chr2g0300261 encoding a NF-YA3 transcription factor. However, our transcriptome data showed that only the *NF-YA3* transcription factor exhibited differential expression between seeds produced under optimal and heat stress conditions (Table S6), which suggested that this potential pioneer gene could represent an interesting candidate regulator of plasticity of both germination speed and homogeneity. On chromosome 5 (Figure 2, indicated with a blue arrow), we also identified many QTNs associated with T80T20_PL and located in a genomic interval containing six closely located genes. Combination with our

transcriptomic data allowed us to refine this list of putative causal genes to four candidates, as four of them displayed differential expression during seed development produced in optimal and heat stress conditions, but did not allow us to more precisely predict the causal gene.

Functional validation of *MIEL1* as regulator of germination plasticity of seeds produced under heat stress

To further investigate the role of the candidate gene *MtMIEL1* (MtrunA17_Chr2g0286331) in the regulation of germination speed, we analyzed its expression profile in seeds of contrasting *M. truncatula* HapMap accessions showing slow and fast germination. We selected seeds HM170 and HM279 accessions as fast-germinating genotypes (*i.e.* T50 around 23h) and HM185 and HM314 accessions as slow-germinating genotypes (*i.e.* T50 at 37h and 40h, respectively) (Figure 4A). We extracted mRNA from their mature seeds produced under heat stress conditions. Fast-germinating genotypes displayed higher *MtMIEL1* relative transcript contents compared to slow-germinating genotypes (Figure 4B). Next, we assessed whether *MtMIEL1* would participate to the plasticity of germination speed in response to heat stress during seed production of the *M. truncatula* reference genotype A17. Germination speed of seeds produced at 26°C was significantly slower than seeds produced at 20°C (Figure 4C). Next, we performed transcript profiling of *MtMIEL1* at 0h and 24h of imbibition at 10°C in the dark (*i.e.* prior to radicle emergence). Figure 4D showed that at both times of imbibition, *MtMIEL1* transcripts were significantly higher in faster germinating seeds, compared to slower germinating seeds, consistent with the observations made on the four contrasting genotypes. From these results, we concluded that *MtMIEL1* was transcriptionally regulated following heat stress during seed production resulting in different transcripts levels during imbibition that are associated with speed of germination.

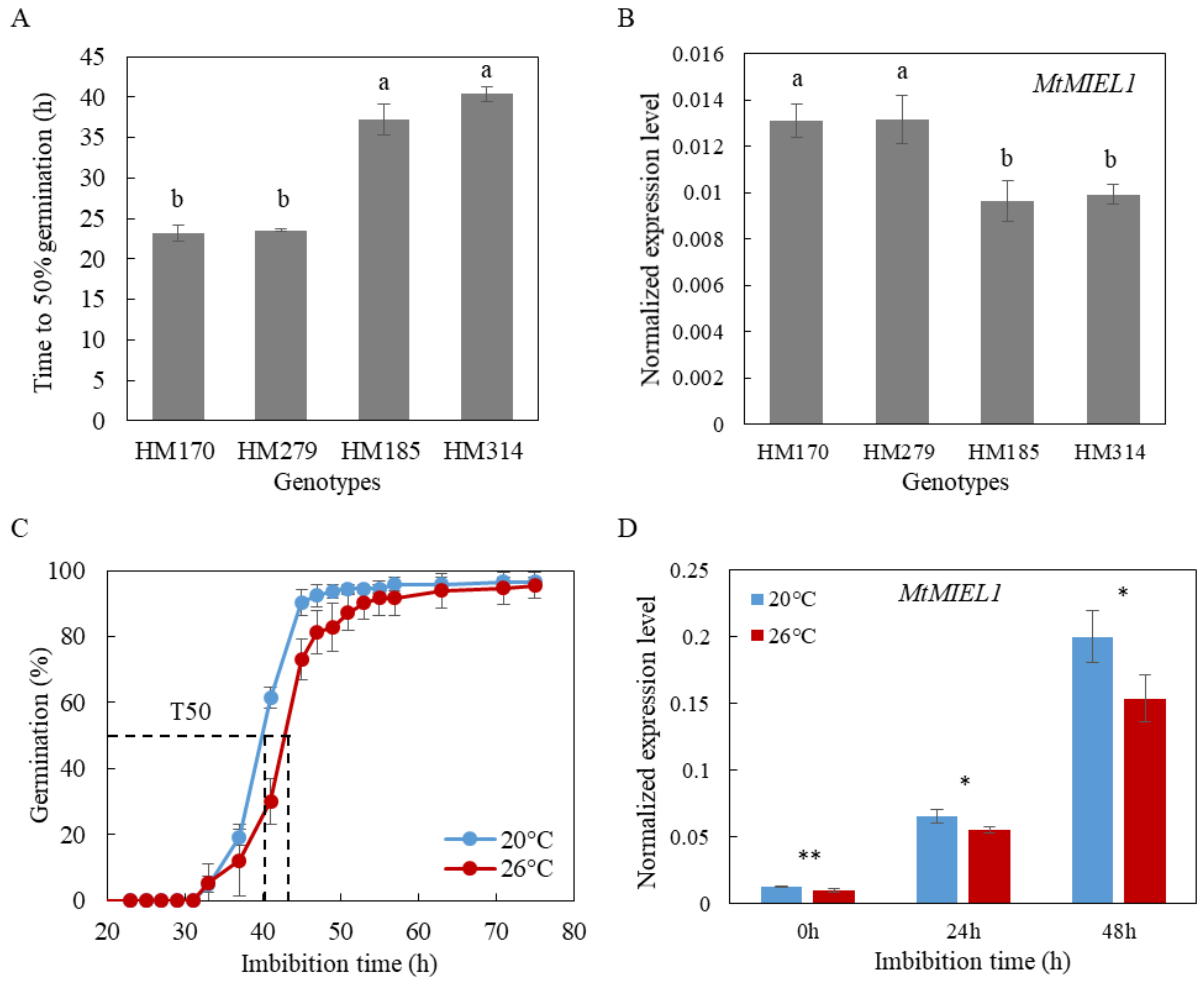


Figure 4. Characterization of seed germination and *MtMIEL1* expression in *M. truncatula* seeds. (A) Germination speed (T50) at 15°C of seeds produced under heat stress condition (26°C) of four natural *M. truncatula* HapMap accessions. (B) Expression level of *MtMIEL1* in dry mature seeds produced under heat stress condition (26°C) of the four natural *M. truncatula* HapMap accessions. Different letters indicate significant difference with each others (P<0.05) identifying by ANOVA and Tukey HSD test. (C) Germination curves at 10°C of *M. truncatula* reference genotype A17 seeds produced at 20°C and 26°C. Time to reach 50% germination (T50) of 20°C and 26°C seeds is indicated by the dash lines. (D) Expression levels of *MtMIEL1* in dry mature seeds (0h), 24-h imbibed seeds (24h) and 48-h imbibed seeds (48h) which were produced at 20°C and 26°C. *, 0.01<p-value<0.05; **, 0.001<p-value<0.01.

To validate the role of *MIEL1* in the regulation of germination speed and germination plasticity of seeds produced under heat stress conditions, we analyzed the ortholog of *MtMIEL1* in *Arabidopsis* by characterizing the germination kinetics from seeds produced at 20°C (control) and 28°C (heat stress) of the homozygote *miel1* mutants. Germination speed of freshly harvested wild-type seeds (Col0) produced at 28°C was much slower than that of seeds

produced at 20°C (Figure 5A, B). In contrast, *miel1* mutant seeds germinated at the same speed regardless of the temperature experienced by the seeds during development, indicating that mutants had lost their plasticity. Moreover, we observed that the germination curves of *miel1* mutants were similar to that of wild-type seeds produced under heat stress (Figure 5A). Next, we repeated this experiment in non-dormant seeds obtained after a 72h stratification treatment at 4°C to release dormancy. Wild-type and *miel1* mutant seeds displayed similar germination kinetics regardless of the production temperature (Figure 5C, D). These results obtained in *Arabidopsis* highlighted a new role of *MIEL1* as regulator of germination speed in response to heat stress during seed development.

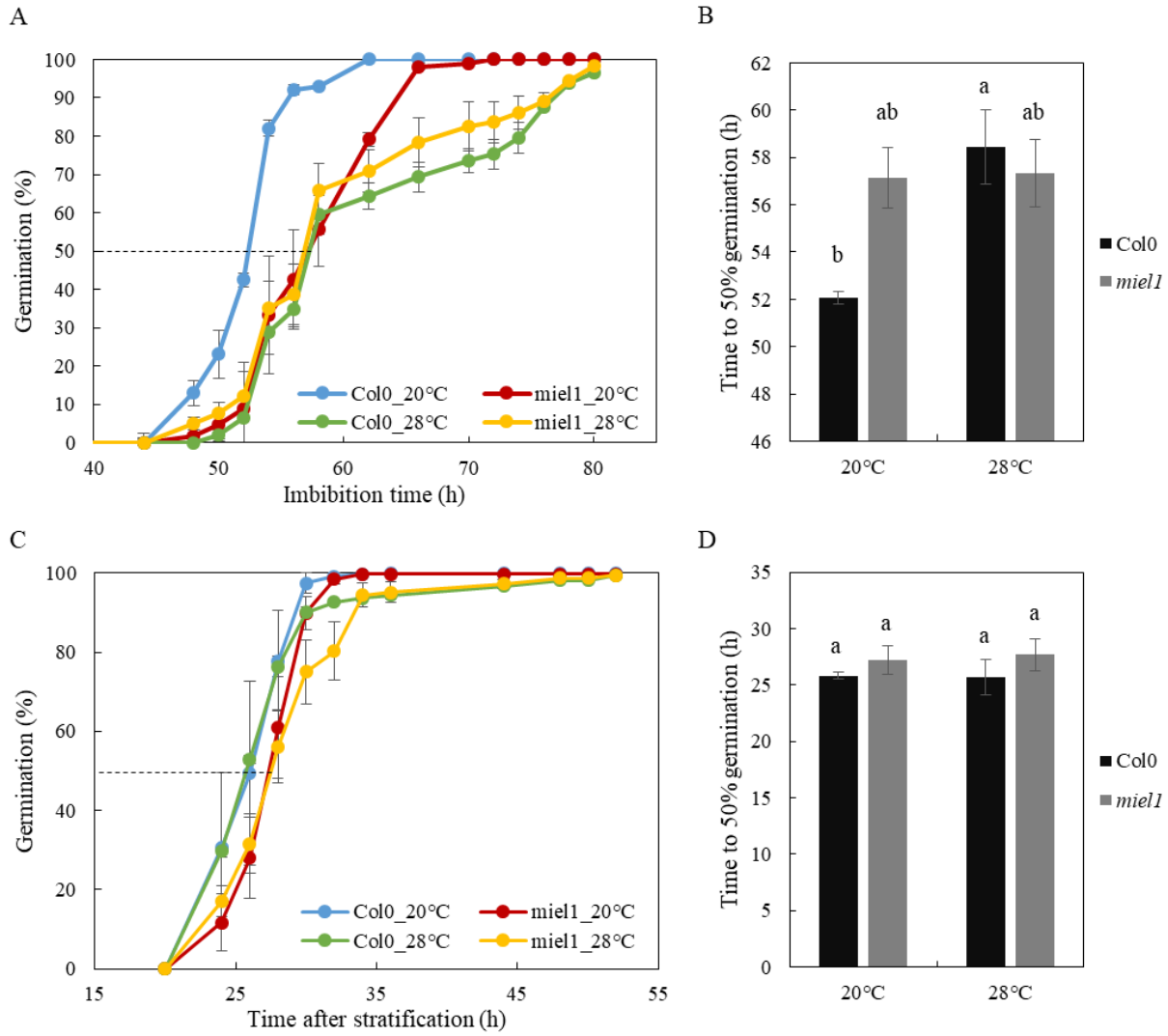


Figure 5. Seed germination of *Arabidopsis thaliana* wild-type (Col-0) and *miel1* T-DNA insertional mutant produced in optimal (20°C) and heat stress (28°C) conditions. (A-B) Germination curves and germination speed (T50) of freshly (dormant) harvested seeds of Col0 and *miel1* mutant grown in optimal (20°C) and heat stress (28°C) conditions. (C-D) Germination curves and germination speed (T50) of mature seeds from Col0 and *miel1* mutant after 72 hours of stratification at 4°C to release dormancy. Error bars represent standard errors of the mean.

Different letters in B and D indicate significant differences between samples (P<0.05) identifying by ANOVA and Tukey HSD test.

Discussion

Use of natural population and GWAS to decipher molecular mechanisms associated to seed traits

Natural variation within plant species causes phenotypic variations due to mutations generated by the evolutionary process. These natural variations are valuable resources to elucidate the molecular basis of phenotypic differences related to plant adaptation to distinct natural environments. In crops, phenotypic differences have been largely exploited in association genetic studies for QTL detection. Due to the development of sequencing technologies, many HapMap collections have been developed using the natural variations present in wild species, permitting genome-wide association studies to become a powerful approach to correlate genotype to phenotype. In our study, we fully benefited from the *M. truncatula* HapMap collection with the help of postGWAS and transcriptome analyses to understand how developing seeds cope with heat stress and modulate their germination response. This work extends previous studies showing that the temperature cues perceived by the mother plant are transmitted to their offspring (Penfield and MacGregor, 2017). Here, we characterized the genetic architecture that govern the plasticity response of *Medicago* and *Arabidopsis* seeds. We obtained a reasonable list of candidate genes potentially involved in regulating different seed traits and the GSEA from these candidate gene lists showed high relevance regarding the expected functional classes controlling the analyzed traits. For instance, candidate gene lists related to germination speed and germination homogeneity showed an enrichment in genes functionally annotated as involved in 'secondary metabolites-flavonoids-chalcone synthase', 'auxin and cytokinin hormone metabolisms' and 'subtilases'. The link between flavonoids and the plasticity response of germination is consistent with the sensitivity of the seed coat to temperature cues during development, which modulates the germination behavior (Penfield and MacGregor, 2017) and it has been largely documented that chalcone synthase, the central enzyme of the flavonoid pathway, showing up-regulation during the first 2-3 days of germination plays a role (Kubasek *et al.*, 1992). Other studies confirmed the role of this pathway during germination using loss-of-function mutants of genes involved in flavonoid

regulation such as *TRANSPARENT TESTA GLABRA* (*TTG*), which displayed more efficient germination than wild-type seeds (Koornneef, 1981). Similar results were observed in different *TRANSPARENT TESTA* mutants (for review Shirley, 1998). Second, the roles of auxin and cytokinin hormone metabolisms in germination performances were described in literature. Indeed, even if auxin is not necessary for seed germination, it has been reported that IAA accumulated in the cotyledons of mature seeds (Epstein *et al.*, 1986; Bialek and Cohen, 1989) influences seed germination with interplay of ABA (Brady *et al.*, 2003). This interplay was shown to be via miR160, which inhibits auxin related gene expression during germination resulting in modulating ABA sensitivity during germination (Liu *et al.*, 2007). Similar to auxin, cytokinin and cytokinin responses factors play a role of enhancer of seed germination when seeds were produced under stress (Khan and Ungar, 1997; Atici *et al.*, 2005; Peleg and Blumwald, 2011). Moreover, cytokinin was also demonstrated to play a role in the transition between dry seed and seedling in concert with ABA via *ABI5* gene regulation (Wang *et al.*, 2011). Finally, enrichment of 'subtilases' functional class in these candidate gene lists could be explained by the need of these proteases, which are highly active at very early stages of seed imbibition, regarding their role in the remobilization of storage proteins during seedling growth, as observed in barley (Galotta *et al.*, 2019). It was not surprising either to find enrichment of the 'HMG-CoA reductase' class in the candidate gene list of germination speed. This central enzyme of the mevalonate and therefore isoprenoid pathway acts upstream to produce many important molecules such as secondary metabolites or hormones (*e.g.* ABA, GA and cytokinin). However, despite its central and upstream position, a study reported that an inhibitor of HMG reductase (*i.e.* one step even before the HMG-CoA reductase) retarded seed germination (Liao *et al.*, 2014). In conclusion, by using the candidate gene lists obtained from the GWAS and GSEA, we could retrieve molecular mechanisms already described in the literature as directly or indirectly involved in studied seed traits, which makes GWA studies a reliable tool in exploratory analysis to decipher molecular processes controlling traits.

Furthermore, using GWAS and postGWAS analyses in combination with adequate transcriptomic data allowed us to identify solid candidate genes potentially regulating the different seed traits. For instance, an ortholog of the Arabidopsis *DA1* gene was identified as candidate regulator of seed weight in *M. truncatula* (MtrunA17_Chr5g0403261, Table S6). This gene *DA1* (AT1G19270) has already been demonstrated to be a regulator of seed and organ size in *Arabidopsis* (Li *et al.*, 2008). From this list of potentially reliable candidate genes, we also identified two of them strongly associated with seed germination performances, a

NUCLEAR FACTOR Y SUBUNIT A3 (*ATHAP2C/NF-YA3*, MtrunA17_Chr2g0300261) and a RING-type zinc finger family gene (MtrunA17_Chr2g0286331), potential ortholog of Arabidopsis *MIEL1* gene, that we called *MtMIEL1*.

MIEL1, a novel regulator of germination plasticity of seed produced under heat stress

The *MYB30-Interacting E3 Ligase1* (*MIEL1*) is an *Arabidopsis* RING type E3 ubiquitin ligase, which was identified to interact with and ubiquitinate MYB30, leading to MYB30 degradation via proteosome pathway. It was first discovered as a regulator of plant defense response to bacteria as MYB30 was known to trigger hypersensitive response in the inoculated zone to restrict bacterial growth (Marino *et al.*, 2013). More recently, it was showed to be involved in the protein turnover of another MYB protein, MYB96, a regulator of ABA signaling in seeds (Lee and Seo, 2016). It was reported that *miel1* mutants were hypersensitive to ABA compared to wild-type seeds, with *miel1* seeds that germinated 1.5-fold slower in the presence of 1µM ABA compared to wild types (Lee and Seo, 2016). In contrast, without ABA treatment, they did not observe any difference in germination of *miel1* mutants at 20°C. This result is similar to our observation using stratified (*i.e.* non dormant) *miel1* seeds, we did not observe any significant change in germination (Figure 5C-D). In contrast, we observed that in dormant seeds (*i.e.* with higher residual ABA content), *miel1* mutant seeds germinated significantly slower than wild-type seeds (Figure 5B), confirming the ABA hypersensitivity phenotype of the *miel1* seeds. In our study, we also observed a decrease of *MtMIEL1* expression in dry mature seeds with the two *M. truncatula* HapMap accessions displaying slow germination compared to the two fast-germinating accessions (Figure 4A, B) and a lower *MtMIEL1* expression level during germination of *M. truncatula* A17 seeds produced under heat stress, which germinated slower, with respect to seeds produced in optimal conditions (Figure 4C-D). Finally, in our study, we analyzed the impact of *miel1* mutation on germination kinetics of seeds produced under optimal and heat stress conditions. We found that *miel1* and wild-type seeds germinated at the same rate no matter of the environmental conditions of seed production (Figure 5 A-B). Our results strongly suggested that *MIEL1* plays a role in the germination plasticity of seeds produced under heat stress.

ACKNOWLEDGEMENTS

Seeds of HapMap accessions used in this study were obtained from the *Medicago* HapMap germplasm resource center (http://www.*Medicago*hapmap.org/hapmap/germplasm), and from the Genetic Improvement and Adaptation of Mediterranean and tropical plants unit (AGAP, INRA Montpellier, JM Prosperi). We thank Marie-Helene Wagner, Didier Demilly, Valérie Blouin and Jean Louis Queyreix (GEVES, Angers, France) and the phenotyping platform PHENOTIC Semences et Plantes (Phenome, Biogenouest, SFR 4207 QUASAV, Angers). This research was conducted in the framework of the regional programme "Objectif Végétal, Research, Education and Innovation in Pays de la Loire", supported by the French Region Pays de la Loire, Angers Loire Métropole and the European Regional Development Fund. This study was also supported by the ANR grant REGULEG (no. ANR-15-CE20-0001) and the China Scholarship Council (CSC No. 201704910863) from the Ministry of Education of P.R. China.

AUTHOR CONTRIBUTIONS

ZC, JLV, BLV, JB and JV performed experiments. ZC, JB, OL and JV analyzed data. ZC, OL and JV wrote the manuscript. All authors reviewed the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

SUPPLEMENTARY MATERIALS

Figure S1. Graphical representations of phenotypic changes in seed traits of individual HapMap accessions between the two seed production conditions.

Figure S2. Distribution histograms before (A) and after (B) the Box-Cox procedure to normalize phenotypic data of seed traits. Corresponding distribution curves are indicated on histograms. Traits are indicated on the x-axis and title of each histogram.

Figure S3. Manhattan plots and the corresponding Q-Q plots from GWAS results regarding seed weight (A, B and C) and seed longevity (D, E and F). Red, green and blue arrows indicate QTNs discussed in the manuscript.

Table S1. Raw phenotypic values, p-values of Shapiro-Wilk test and lambda values to transform phenotypic data. Lambda values were calculated from the Box-Cox algorithm to obtain normal distribution. P-values of Shapiro-Wilk test were calculated on phenotypic values before and after Box-Cox transformation.

Table S2. gwas result file related to seed weight (WEIGHT) containing p-values of all SNPs obtained from GWAS results.

Table S3. gwas result file related to seed germination speed (T50) containing p-values of all SNPs obtained from GWAS results.

Table S4. gwas result file related to seed germination homogeneity (T80T20) containing p-values of all SNPs obtained from GWAS results.

Table S5. gwas result file related to seed longevity (G204DA) containing p-values of all SNPs obtained from GWAS results.

Table S6. List of highly significant QTNs (p-value<10^{-5}) related to all seed traits. QTNs names, positions and p-values are indicated from FarmCPU. Numbers of potential associated SNP(s) and putative causal genes are indicated from PLINK analysis. Differentially expressed genes and gene annotations are indicated in the table.

Table S7. List of different primers used to perform PCR and qRT-PCR experiments.

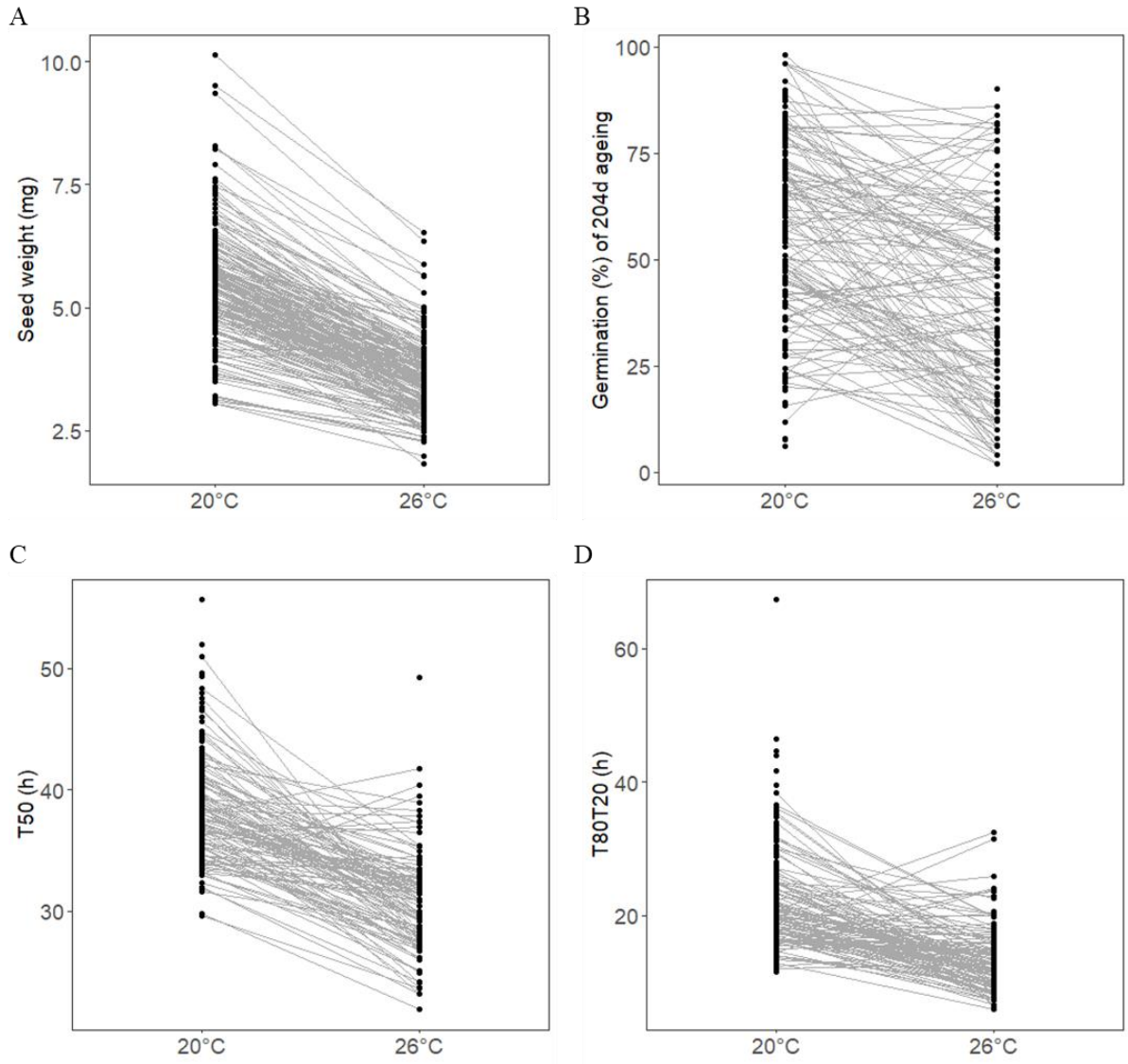


Figure S1: Graphical representations of phenotypic changes in seed traits of individual HapMap accessions between the two seed production conditions.

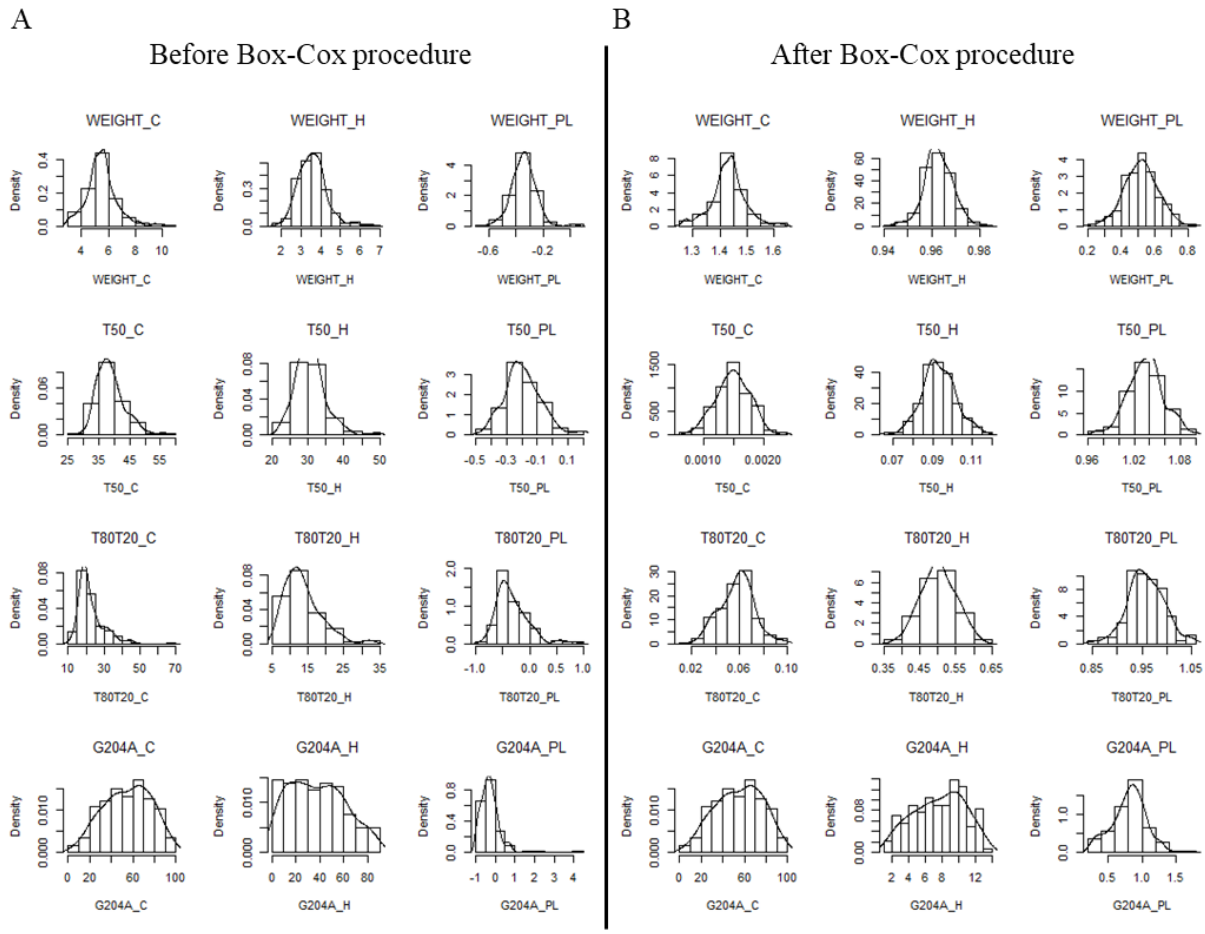


Figure S2: Distribution histograms before (A) and after (B) the Box-Cox procedure to normalize phenotypic data of seed traits. Corresponding distribution curves are indicated on histograms. Traits are indicated on the x-axis and title of each histogram.

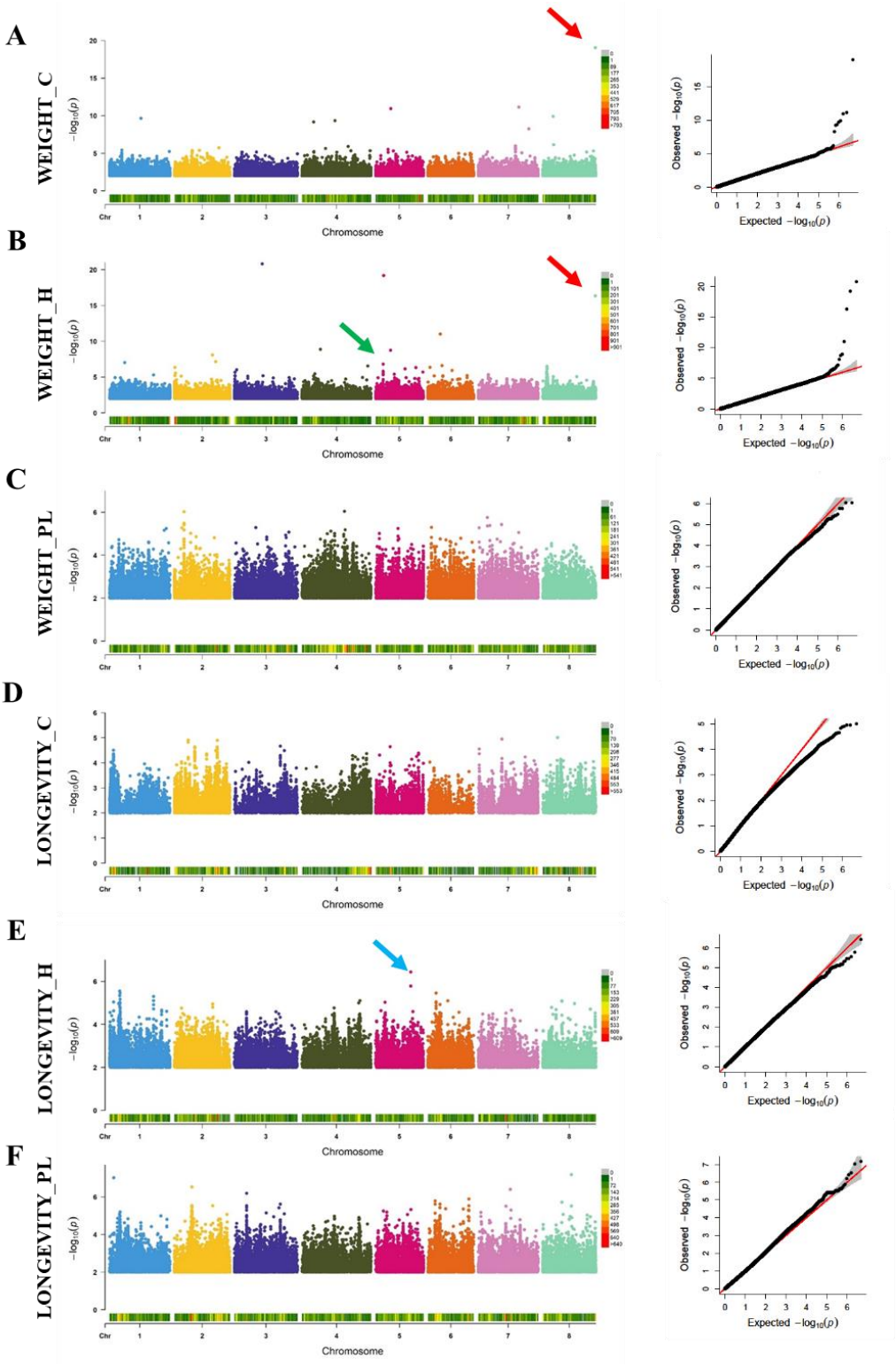


Figure S3. Manhattan plots and the corresponding Q-Q plots from GWAS results regarding seed weight (A, B and C) and seed longevity (D, E and F). Red, green and blue arrows indicate QTNs discussed in the manuscript.

CHAPTER 4: CONCLUSIONS AND PERSPECTIVES

This three-years thesis project aimed to explore seed molecular processes affected during heat stress in isolated seed tissues to identify candidate genes potentially regulating the stress response in *Medicago truncatula* seed tissues. The main goal was achieved with an exhaustive characterization of transcriptome changes in all three seed tissues along the seed development, plus a characterization of the methylome and chromatin dynamics in embryo following heat stress. Moreover, several candidate genes were identified and preliminary mutant analyses were initiated validating our choices.

One of the main achievements of this PhD work was to develop this map of (epi)genetic regulation due to heat stress response. Even if not entirely exploited at the end of these three years, this knowledge represents a valuable tool for the SEED team current/future projects. For instance, this knowledge has been already used in other projects such as the ANR DESWITCH (2020-2023) or the RFI epiDT (2020-2021) projects. It will become even more valuable in the coming years because this thesis project was the initial "seed stress" project but now several other projects are ongoing related to heat stress in different species but also to different (a)biotic stresses in *Medicago truncatula*. All together, these projects should provide, in the mid-term perspective, an accurate overview of how seeds sense, respond and adapt to stresses during their development. One of ongoing goal related to this thesis work is to valorise these data and make them accessible to the scientific community. In this perspective, we are currently collaborating with LIPM (Toulouse) to develop a Medicago Gene Expression Atlas webserver based on RNA-seq data, which already contains all the transcriptomic data generated in this thesis and offers many visualization and analysis tools to share and fully exploit these data. Regarding results from the GWAS approaches and the epigenome dynamics (*i.e.* methylome and H3K27me3), we already uploaded all the data in our in-house Jbrowse (*i.e.* genome browser), which will be publicly available soon after publication of these results.

The second main achievement of this project was to identify relevant candidate genes to have a better understanding of how seeds sense, respond and adapt to heat stress. Even if the data integration of different parts of this thesis work has not been completed on a timely basis, we identified candidate gene lists in each of the sections and these choices appeared to be

relevant based on preliminary results from mutant analyses (*e.g. MtMIEL1* and *HAP3/MtL1L*). Of course, the initial plan to integrate all OMICS data in order to refine the list of candidate gene is still ongoing with some solutions, which have already been tested but not discussed such as CAMOCO that allowed us to integrate transcriptomics and GWAS data (Schaefer *et al.*, 2018). To date, two solid candidate genes are under investigation, *MtMIEL1 (*MtrunA17_Chr2g0286331) and *HAP3/MtL1L (*MtrunA17_Chr4g0076381), and functional characterization of these two genes will continue. Moreover, now we have homozygote mutant lines for more candidates such as *MtABI5* (MtrunA17_Chr7g0266211), *MtDASH* (MtrunA17_Chr2g0282441) and *MtHAP2* (MtrunA17_Chr2g0300261). Even if *MtDASH* and *MtABI5* are already published genes known to be involved in seed developmental processes, their regulation by (heat) stresses is interesting to elucidate, in contrast regarding *MtHAP2*, no report has yet been done, therefore it could represent an interesting candidate in heat stress response but also in seed maturation process.

REFERENCES

Adams, S., Vinkenoog, R., Spielman, M., Dickinson, H. G., and Scott, R. J. (2000). Parent-of-origin effects on seed development in Arabidopsis thaliana require DNA methylation. *Development* 127, 2493–2502.

Ahmad, A., Zhang, Y., and Cao, X. F. (2010). Decoding the epigenetic language of plant development. *Mol. Plant* 3, 719–728.

Ali-Rachedi, S., Bouinot, D., Wagner, M. H., Bonnet, M., Sotta, B., Grappin, P., et al. (2004). Changes in endogenous abscisic acid levels during dormancy release and maintenance of mature seeds: Studies with the Cape Verde Islands ecotype, the dormant model of Arabidopsis thaliana. *Planta* 219, 479–488.

Almoguera, C., Prieto-Dapena, P., Díaz-Martín, J., Espinosa, J. M., Carranco, R., and Jordano, J. (2009). The HaDREB2 transcription factor enhances basal thermotolerance and longevity of seeds through functional interaction with HaHSFA9. *BMC Plant Biol.* 9, 75.

Almoguera, C., Rojas, A., Díaz-Martín, J., Prieto-Dapena, P., Carranco, R., and Jordano, J. (2002). A seed-specific heat-shock transcription factor involved in developmental regulation during embryogenesis in sunflower. *J. Biol. Chem.* 277, 43866–43872.

Alpert, P. (2005). The limits and frontiers of desiccation-tolerant life. *Integr. Comp. Biol.* 45, 685–695.

Atici, Ö., Ağar, G., and Battal, P. (2005). Changes in phytohormone contents in chickpea seeds germinating under lead or zinc stress. *Biol. Plant.* 49, 215–222.

Audran, C., Liotenberg, S., Gonneau, M., North, H., Frey, A., Tap-Waksman, K., et al. (2001). Localisation and expression of zeaxanthin epoxidase mRNA in Arabidopsis in response to drought stress and during seed development. *Aust. J. Plant Physiol.* 28, 1161–1173.

Awan, S., Footitt, S., and Finch-Savage, W. E. (2018). Interaction of maternal environment and allelic differences in seed vigour genes determines seed performance in Brassica oleracea. *Plant J.* 94, 1098–1108.

Bailly, C., Benamar, A., Corbineau, F., and Côme, D. (1996). Changes in malondialdehyde content and in superoxide dismutase, catalase and glutathione reductase activities in sunflower seeds as related to deterioration during accelerated aging. *Physiol. Plant.* 97,

104–110.

Bandyopadhyay, K., Uluçay, O., Şakiroğlu, M., Udvardi, M. K., and Verdier, J. (2016). Analysis of large seeds from three different Medicago truncatula ecotypes reveals a potential role of hormonal balance in final size determination of legume grains. *Int. J. Mol. Sci.* 17, 1–13.

Banerjee, A., and Roychoudhury, A. (2016). Group II late embryogenesis abundant (LEA) proteins: structural and functional aspects in plant abiotic stress. *Plant Growth Regul.* 79, 1–17.

Baniwal, S. K., Kwan, Y. C., Scharf, K. D., and Nover, L. (2007). Role of heat stress transcription factor HsfA5 as specific repressor of HsfA4. *J. Biol. Chem.* 282, 3605–3613.

Barker, D. G., Bianchi, S., Blondon, F., Dattée, Y., Duc, G., Essad, S., et al. (1990). Medicago truncatula, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. *Plant Mol. Biol. Report.* 8, 40–49.

Bartee, L., Malagnac, F., and Bender, J. (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes Dev.* 15, 1753–1758.

Barthole, G., To, A., Marchive, C., Brunaud, V., Soubigou-Taconnat, L., Berger, N., et al. (2014). MYB118 represses endosperm maturation in seeds of Arabidopsis. *Plant Cell* 26, 3519–3537.

Baskin, J. M., and Baskin, C. C. (2004). A classification system for seed dormancy. *Seed Sci. Res.* 14, 1–16.

Baud, S., Mendoza, M. S., To, A., Harscoët, E., Lepiniec, L., and Dubreucq, B. (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *Plant J.* 50, 825–838.

Baxter, A., Mittler, R., and Suzuki, N. (2014). ROS as key players in plant stress signalling. *J. Exp. Bot.* 65, 1229–1240.

Beaudoin, N., Serizet, C., Gosti, F., and Giraudat, J. (2000). Interactions between abscisic acid and ethylene signaling cascades. *Plant Cell* 12, 1103–1115.

Begcy, K., Sandhu, J., and Walia, H. (2018). Transient heat stress during early seed

development primes germination and seedling establishment in rice. *Front. Plant Sci.* 871.

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci. U. S. A.* 110.

Benoit, L., Rousseau, D., Belin, É., Demilly, D., and Chapeau-Blondeau, F. (2014). Simulation of image acquisition in machine vision dedicated to seedling elongation to validate image processing root segmentation algorithms. *Comput. Electron. Agric.* 104, 84–92.

Bentsink, L., Jowett, J., Hanhart, C. J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17042–17047.

Bernareggi, G., Carbognani, M., Petraglia, A., and Mondoni, A. (2015). Climate warming could increase seed longevity of alpine snowbed plants. *Alp. Bot.* 125, 69–78.

Bethke, P. C., Gubler, F., Jacobsen, J. V., and Jones, R. L. (2004). Dormancy of Arabidopsis seeds and barley grains can be broken by nitric oxide. *Planta* 219, 847–855.

Bethke, P. C., Libourel, I. G. L., and Jones, R. L. (2006). Nitric oxide reduces seed dormancy in Arabidopsis. *J. Exp. Bot.* 57, 517–526.

Bewley, J. D., and Black, M. (1994). *SEEDS.Physiology of development and germination. second edition*.

Bewley, J. D., Bradford, K. J., Hilhorst, H. W. M., and Nonogaki, H. (2013). *Seeds: Physiology of development, germination and dormancy, 3rd edition*. New York, NY: Springer New York.

Bharti, K., Von Koskull-Döring, P., Bharti, S., Kumar, P., Tintschl-Körbitzer, A., Treuter, E., et al. (2004). Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription coactivator with a histone-like motif interacting with the plant CREB binding protein ortholog HAC1. *Plant Cell* 16, 1521–1535.

Bialek, K., and Cohen, J. D. (1989). Free and conjugated lndole-3-acetic acid in developing bean seeds. *Plant Physiol.* 91, 775–779.

Bies-Ethève, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., et al. (2008). Inventory, evolution and expression profiling diversity of the LEA (late

embryogenesis abundant) protein gene family in Arabidopsis thaliana. *Plant Mol. Biol.* 67, 107–124.

Black, M., Corbineau, F., Gee, H., and Côme, D. (1999). Water content, raffinose, and dehydrins in the induction of desiccation tolerance in immature wheat embryos. *Plant Physiol.* 120, 463–471.

Boccaccini, A., Lorrai, R., Ruta, V., Frey, A., Mercey-Boutet, S., Marion-Poll, A., et al. (2016). The DAG1 transcription factor negatively regulates the seed-to-seedling transition in Arabidopsis acting on ABA and GA levels. *BMC Plant Biol.* 16, 198.

Bokszczanin, K. L., and Fragkostefanakis, S. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Front. Plant Sci.* 4.

Bonhomme, M., André, O., Badis, Y., Ronfort, J., Burgarella, C., Chantret, N., et al. (2014). High-density genome-wide association mapping implicates an F-box encoding gene in Medicago truncatula resistance to Aphanomyces euteiches. *New Phytol.* 201, 1328–1342.

Boudet, J., Buitink, J., Hoekstra, F. A., Rogniaux, H., Larré, C., Satour, P., et al. (2006). 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 Physiol.* 140, 1418–1436.

Boulard, C., Fatihi, A., Lepiniec, L., and Dubreucq, B. (2017). Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1860, 1069–1078.

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., et al. (2017). DNA methylation dynamics during early plant life. *Genome Biol.* 18, 179.

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb Repressive Complex 2 Controls the Embryo-to-Seedling Phase Transition. *PLoS Genet.* 7, e1002014.

Box, G. E. P., and Cox, D. R. (1964). An Analysis of Transformations. *J. R. Stat. Soc. Ser. B* 26, 211–243.

Bradford, K. J., and Nonogaki, H. (2007). *Seed Development, Dormancy and Germination*.

Brady, S. M., Sarkar, S. F., Bonetta, D., and McCourt, P. (2003). The ABSCISIC ACID

INSENSITIVE 3 (ABI3) gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in Arabidopsis. *Plant J.* 34, 67–75.

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., et al. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1–7.

Braybrook, S. A., and Harada, J. J. (2008). LECs go crazy in embryo development. *Trends Plant Sci.* 13, 624–630.

Buitink, J., and Leprince, O. (2004). Glass formation in plant anhydrobiotes: Survival in the dry state. *Cryobiology* 48, 215–228.

Busch, W., Wunderlich, M., and Schöffl, F. (2005). Identification of novel heat shock factor-dependent genes and biochemical pathways in Arabidopsis thaliana. *Plant J.* 41, 1–14.

Catoni, M., Tsang, J. M. F., Greco, A. P., and Zabet, N. R. (2018). DMRcaller: a versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. *Nucleic Acids Res.* 46.

Cernac, A., and Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *Plant J.* 40, 575–585.

Charng, Y. Y., Liu, H. C., Liu, N. Y., Chi, W. T., Wang, C. N., Chang, S. H., et al. (2007). A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. *Plant Physiol.* 143, 251–262.

Chatelain, E., Hundertmark, M., Leprince, O., Gall, S. Le, Satour, P., Deligny-Penninck, S., et al. (2012). Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant, Cell Environ.* 35, 1440–1455.

Chaudhury, A. M., Koltunow, A., Payne, T., Luo, M., Tucker, M. R., Dennis, E. S., et al. (2001). Control of early seed development. *Annu. Rev. Cell Dev. Biol.* 17, 677–699.

Chebrolu, K. K., Fritschi, F. B., Ye, S., Krishnan, H. B., Smith, J. R., and Gillman, J. D. (2016). Impact of heat stress during seed development on soybean seed metabolome. *Metabolomics* 12, 1–14.

Chen, F., Zhou, W., Yin, H., Luo, X., Chen, W., Liu, X., et al. (2020). Shading of the mother plant during seed development promotes subsequent seed germination in soybean. *J. Exp. Bot.* 71, 2072–2084.

Chen, Z., Lancon-Verdier, V., Le Signor, C., She, Y.-M., Kang, Y., and Verdier, J. (2021a). Genome-wide association study identified candidate genes for seed size and seed composition improvement in M. truncatula. *Sci. Rep.* 11, 4224.

Chen, Z., Ly Vu, B., Leprince, O., and Verdier, J. (2021b). RNA sequencing data for heat stress response in isolated medicago truncatula seed tissues. *Data Br.* 35, 106726.

Chinnusamy, V., and Zhu, J. K. (2009). Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12, 133–139.

Cuming, A. C. (1999). "LEA Proteins," in *Seed Proteins*, 753–780.

D'Erfurth, I., Cosson, V., Eschstruth, A., Lucas, H., Kondorosi, A., and Ratet, P. (2003). Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume Medicago truncatula. *Plant J.* 34, 95–106.

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

Dekkers, B. J. W., He, H., Hanson, J., Willems, L. A. J., Jamar, D. C. L., Cueff, G., et al. (2016). The Arabidopsis Delay of Germination 1 gene affects Abscisic Acid Insensitive 5 (ABI5) expression and genetically interacts with ABI3 during Arabidopsis seed development. *Plant J.* 85, 451–465.

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

Deng, W., Buzas, D. M., Ying, H., Robertson, M., Taylor, J., Peacock, W., et al. (2013). Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics* 14, 593.

Deshpande, S. S. (1992). Food Legumes in Human Nutrition: A Personal Perspective. *Crit. Rev. Food Sci. Nutr.* 32, 333–363.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Ellis, R. H., Hong, T. D., and Jackson, M. T. (1993). Seed production environment, time of harvest, and the potential longevity of seeds of three cultivars of rice (Oryza sativa L.). *Ann. Bot.* 72, 583–590.

Epstein, E., Sagee, O., Cohen, J. D., and Garty, J. (1986). Endogenous auxin and ethylene in the lichen Ramalina duriaei1. *Plant Physiol.* 82, 1122–1125.

Fatihi, A., Boulard, C., Bouyer, D., Baud, S., Dubreucq, B., and Lepiniec, L. (2016). Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. *Plant Sci.* 250, 198–204.

Figueiredo, D. D., and Köhler, C. (2018). Auxin: A molecular trigger of seed development. *Genes Dev.* 32, 479–490.

Finch-Savage, W. E., and Bassel, G. W. (2016). Seed vigour and crop establishment: Extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591.

Finch-Savage, W. E., and Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. *New Phytol.* 171, 501–523.

Finkelstein, R. R., Gampala, S. S. L., and Rock, C. D. (2002). Abscisic Acid Signaling in Seeds and Seedlings. *Plant Cell* 14, S15–S45.

Finkelstein, R. R., Li Wang, M., Lynch, T. J., Rao, S., and Goodman, H. M. (1998). The arabidopsis abscisic acid response locus ABI4 encodes an APETALA2 domain protein. *Plant Cell* 10, 1043–1054.

Finkelstein, R. R., and Lynch, T. J. (2000). The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. *Plant Cell* 12, 599–609.

Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C. (2008). Molecular aspects of seed dormancy. *Annu. Rev. Plant Biol.* 59, 387–415.

Fischer, D. S., Theis, F. J., and Yosef, N. (2018). Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res.* 46, 1–10.

Frey, A., Audran, C., Marin, E., Sotta, B., and Marion-Poll, A. (1999). Engineering seed dormancy by the modification of zeaxanthin epoxidase gene expression. *Plant Mol. Biol.*

39, 1267–1274.

Galau, G. A., Bijaisoradat, N., and Hughes, D. W. (1987). Accumulation kinetics of cotton late embryogenesis-abundant mRNAs and storage protein mRNAs: Coordinate regulation during embryogenesis and the role of abscisic acid. *Dev. Biol.* 123, 198–212.

Galotta, M. F., Pugliese, P., Gutiérrez-Boem, F. H., Veliz, C. G., Criado, M. V., Caputo, C., et al. (2019). Subtilase activity and gene expression during germination and seedling growth in barley. *Plant Physiol. Biochem.* 139, 197–206.

Garcia, D., Gerald, J. N. F., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in arabidopsis. *Plant Cell* 17, 52–60.

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jürgens, G., and Berger, F. (2003). Arabidopsis haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol.* 131, 1661–1670.

Geshnizjani, N., Sarikhani Khorami, S., Willems, L. A. J., Snoek, B. L., Hilhorst, H. W. M., and Ligterink, W. (2019). The interaction between genotype and maternal nutritional environments affects tomato seed and seedling quality. *J. Exp. Bot.* 70, 2905–2918.

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

Gnesutta, N., Saad, D., Chaves-Sanjuan, A., Mantovani, R., and Nardini, M. (2017). Crystal Structure of the Arabidopsis thaliana L1L/NF-YC3 Histone-fold Dimer Reveals Specificities of the LEC1 Family of NF-Y Subunits in Plants. *Mol. Plant* 10, 645–648.

Goldy, C., Pedroza-Garcia, J.-A., Breakfield, N., Cools, T., Vena, R., Benfey, P. N., et al. (2021). The Arabidopsis GRAS-type SCL28 transcription factor controls the mitotic cell cycle and division plane orientation. *Proc. Natl. Acad. Sci.* 118, e2005256118.

Gong, M., Van der Luit, A. H., Knight, M. R., and Trewavas, A. J. (1998). Heat-shock-induced changes in intracellular Ca2+ level in tobacco seedlings in relation to thermotolerance. *Plant Physiol.* 116, 429–437.

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet. Res. Int.* 2015.

Guilioni, L., Wery, J., and Tardieu, F. (1997). Heat stress-induced abortion of buds and flowers in pea: Is sensitivity linked to organ age or to relations between reproductive organs? *Ann. Bot.* 80, 159–168.

Harrell Jr., F. E. (2006). Hmisc: Harrell Miscellaneous. *R Packag. version 3.0-12*, 1–397. Available at: https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf.

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14, 9643–9684.

He, H., De Souza Vidigal, D., Basten Snoek, L., Schnabel, S., Nijveen, H., Hilhorst, H., et al. (2014). Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *J. Exp. Bot.* 65, 6603–6615.

Huang, X., Lu, Z., Wang, X., Ouyang, Y., Chen, W., Xie, K., et al. (2016). Imprinted gene OsFIE1 modulates rice seed development by influencing nutrient metabolism and modifying genome H3K27me3. *Plant J.* 87, 305–317.

Huh, J. H., Bauer, M. J., Hsieh, T. F., and Fischer, R. (2007). Endosperm gene imprinting and seed development. *Curr. Opin. Genet. Dev.* 17, 480–485.

Hundertmark, M., and Hincha, D. K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics* 9, 118.

Hyun, M. K., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723.

Ingouff, M., Haseloff, J., and Berger, F. (2005). Polycomb group genes control developmental timing of endosperm. *Plant J.* 42, 663–674.

Jean finnegan, E., and Dennis, E. S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. *Nucleic Acids Res.* 21, 2383–2388.

Jo, L., Pelletier, J. M., and Harada, J. J. (2019). Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *J. Integr. Plant Biol.* 61, 564–580.

Johnson, C. S., Kolevski, B., and Smyth, D. R. (2002). Transparent Testa Glabra2, a trichome

and seed coat development gene of arabidopsis, encodes a WRKY transcription factor. *Plant Cell* 14, 1359–1375.

Kafadar, K., Koehler, J. R., Venables, W. N., and Ripley, B. D. (1999). Modern Applied Statistics with S-Plus. *Am. Stat.* 53, 86.

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

Kang, X., Li, W., Zhou, Y., and Ni, M. (2013). A WRKY Transcription Factor Recruits the SYG1-Like Protein SHB1 to Activate Gene Expression and Seed Cavity Enlargement. *PLoS Genet.* 9, e1003347.

Kang, Y., Li, M., Sinharoy, S., and Verdier, J. (2016). A snapshot of functional genetic studies in Medicago truncatula. *Front. Plant Sci.* 7.

Kang, Y., Sakiroglu, M., Krom, N., Stanton-Geddes, J., Wang, M., Lee, Y. C., et al. (2015). Genome-wide association of drought-related and biomass traits with HapMap SNPs in Medicago truncatula. *Plant, Cell Environ.* 38, 1997–2011.

Kang, Y., Torres-Jerez, I., An, Z., Greve, V., Huhman, D., Krom, N., et al. (2019). Genome-wide association analysis of salinity responsive traits in Medicago truncatula. *Plant Cell Environ.* 42, 1513–1531.

Keith, K., Kraml, M., Dengler, N. G., and McCourt, P. (1994). fusca3: A heterochronic mutation affecting late embryo development in Arabidopsis. *Plant Cell* 6, 589–600.

Khan, M. A., and Ungar, I. A. (1997). Effects of light, salinity, and thermoperiod on the seed germination of halophytes. *Can. J. Bot.* 75, 835–841.

Kim, S. Y., Lee, J., Eshed-Williams, L., Zilberman, D., and Sung, Z. R. (2012). EMF1 and PRC2 Cooperate to Repress Key Regulators of Arabidopsis Development. *PLoS Genet.* 8, e1002512.

Koornneef, M. (1981). The complex syndrome of TTG mutants. *Arab. Inf. Serv.* 18, 45–51.

Kotak, S., Vierling, E., Bäumlein, H., and Von Koskull-Dörlng, P. (2007). A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. *Plant Cell* 19, 182–195.

Krueger, F., and Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572.

Kubasek, W. L., Shirley, B. W., McKillop, A., Goodman, H. M., Briggs, W., and Ausubel, F. M. (1992). Regulation of flavonoid biosynthetic genes in germinating Arabidopsis seedlings. *Plant Cell* 4, 1229–1236.

Kumar, M., Busch, W., Birke, H., Kemmerling, B., Nürnberger, T., and Schöffl, F. (2009). Heat shock factors HsfB1 and HsfB2b are involved in the regulation of Pdf1.2 expression and pathogen resistance in Arabidopsis. *Mol. Plant* 2, 152–165.

Kwong, R. W., Bui, A. Q., Lee, H., Kwong, L. W., Fischer, R. L., Goldberg, R. B., et al. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15, 5–18.

Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9.

Lau, S., Slane, D., Herud, O., Kong, J., and Jürgens, G. (2012). Early embryogenesis in flowering plants: Setting up the basic body pattern. *Annu. Rev. Plant Biol.* 63, 483–506.

Lee, H. G., and Seo, P. J. (2016). The Arabidopsis MIEL1 E3 ligase negatively regulates ABA signalling by promoting protein turnover of MYB96. *Nat. Commun.* 7.

Lefebvre, V., North, H., Frey, A., Sotta, B., Seo, M., Okamoto, M., et al. (2006). Functional analysis of Arabidopsis NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J.* 45, 309–319.

Lemontey, C., Mousset-Déclas, C., Munier-Jolain, N., and Boutin, J. P. (2000). Maternal genotype influences pea seed size by controlling both mitotic activity during early embryogenesis and final endoreduplication level/cotyledon cell size in mature seed. *J. Exp. Bot.* 51, 167–175.

Lepiniec, L., Devic, M., Roscoe, T. J., Bouyer, D., Zhou, D. X., Boulard, C., et al. (2018). Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reprod.* 31, 291–307.

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

Li, J., Nie, X., Tan, J. L. H., and Berger, F. (2013). Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15479–15484.

Li, N., and Li, Y. (2015). Maternal control of seed size in plants. *J. Exp. Bot.* 66, 1087–1097.

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32.

Li, N., Xu, R., and Li, Y. (2019). Molecular Networks of Seed Size Control in Plants. *Annu. Rev. Plant Biol.* 70, 435–463.

Li, Y., Cheng, R., Spokas, K. A., Palmer, A. A., and Borevitz, J. O. (2014). Genetic variation for life history sensitivity to seasonal warming in Arabidopsis thaliana. *Genetics* 196, 569–577.

Li, Y., Zheng, L., Corke, F., Smith, C., and Bevan, M. W. (2008). Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. *Genes Dev.* 22, 1331–1336.

Liao, P., Wang, H., Wang, M., Hsiao, A.-S., Bach, T. J., and Chye, M.-L. (2014). Transgenic Tobacco Overexpressing Brassica juncea HMG-CoA Synthase 1 Shows Increased Plant Growth, Pod Size and Seed Yield. *PLoS One* 9, e98264.

Lin, J. Y., Le, B. H., Chen, M., Henry, K. F., Hur, J., Hsieh, T. F., et al. (2017). Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9730–E9739.

Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., et al. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science (80-.).* 292, 2077–2080.

Liu, H. C., Liao, H. T., and Charng, Y. Y. (2011). The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. *Plant, Cell Environ.* 34, 738–751.

Liu, J., Feng, L., Gu, X., Deng, X., Qiu, Q., Li, Q., et al. (2019a). An H3K27me3 demethylase-HSFA2 regulatory loop orchestrates transgenerational thermomemory in Arabidopsis. *Cell Res.* 29, 379–390.

Liu, J., Feng, L., Li, J., and He, Z. (2015). Genetic and epigenetic control of plant heat

responses. *Front. Plant Sci.* 6, 1–21.

Liu, P. P., Montgomery, T. A., Fahlgren, N., Kasschau, K. D., Nonogaki, H., and Carrington, J. C. (2007). Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. *Plant J.* 52, 133–146.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., et al. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell* 10, 1391–1406.

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genet.* 12, e1005767.

Liu, Y., Li, J., Zhu, Y., Jones, A., Rose, R. J., and Song, Y. (2019b). Heat Stress in Legume Seed Setting: Effects, Causes, and Future Prospects. *Front. Plant Sci.* 10.

Liu, Y., Ye, N., Liu, R., Chen, M., and Zhang, J. (2010). H2O2 mediates the regulation of ABA catabolism and GA biosynthesis in Arabidopsis seed dormancy and germination. *J. Exp. Bot.* 61, 2979–2990.

Lobell, D. B., Schlenker, W., and Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science (80-.).* 333, 616–620.

Lohmann, C., Eggers-Schumacher, G., Wunderlich, M., and Schöffl, F. (2004). Two different heat shock transcription factors regulate immediate early expression of stress genes in Arabidopsis. *Mol. Genet. Genomics* 271, 11–21.

Long, S. P., and Ort, D. R. (2010). More than taking the heat: Crops and global change. *Curr. Opin. Plant Biol.* 13, 240–247.

Lotan, T., Ohto, M. A., Matsudaira Yee, K., West, M. A. L., Lo, R., Kwong, R. W., et al. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93, 1195–1205.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15.

Luerßen, H., Kirik, V., Herrmann, P., and Miséra, S. (1998). FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. *Plant J.* 15, 755–764.

Lukens, L. N., and Zhan, S. (2007). The plant genome's methylation status and response to stress: implications for plant improvement. *Curr. Opin. Plant Biol.* 10, 317–322.

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

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., et al. (2006). Different polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep.* 7, 947–952.

Malik, M. K., Slovin, J. P., Hwang, C. H., and Zimmerman, J. L. (1999). Modified expression of a carrot small heat shock protein gene, Hsp17.7, results in increased or decreased thermotolerance. *Plant J.* 20, 89–99.

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP Markers and Their Impact on Plant Breeding. *Int. J. Plant Genomics*, 1–11.

Manfre, A. J., Lanni, L. M., and Marcotte, W. R. (2006). The Arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiol.* 140, 140–149.

Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Hugueney, P., et al. (1996). Molecular identification of zeaxanthin epoxidase of Nicotiana plumbaginifolia, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of Arabidopsis thaliana. *EMBO J.* 15, 2331–2342.

Marino, D., Froidure, S., Canonne, J., Ben Khaled, S., Khafif, M., Pouzet, C., et al. (2013). Arabidopsis ubiquitin ligase MIEL1 mediates degradation of the transcription factor MYB30 weakening plant defence. *Nat. Commun.* 4, 1476.

McKevith, B. (2004). Nutritional aspects of cereals. *Nutr. Bull.* 29, 111–142.

Meinke, D. W., Franzmann, L. H., Nickle, T. C., and Yeung, E. C. (1994). Leafy cotyledon mutants of Arabidopsis. *Plant Cell* 6, 1049–1064.

Mishra, S. K., Tripp, J., Winkelhaus, S., Tschiersch, B., Theres, K., Nover, L., et al. (2002). In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes Dev.* 16, 1555–1567.

Molitor, A. M., Bu, Z., Yu, Y., and Shen, W.-H. (2014). Arabidopsis AL PHD-PRC1 Complexes Promote Seed Germination through H3K4me3-to-H3K27me3 Chromatin State Switch in Repression of Seed Developmental Genes. *PLoS Genet.* 10, e1004091.

Müller, K., Bouyer, D., Schnittger, A., and Kermode, A. R. (2012). Evolutionarily Conserved Histone Methylation Dynamics during Seed Life-Cycle Transitions. *PLoS One* 7, e51532.

Murakami, T., Matsuba, S., Funatsuki, H., Kawaguchi, K., Saruyama, H., Tanida, M., et al. (2004). Over-expression of a small heat shock protein, sHSP17.7, confers both heat tolerance and UV-B resistance to rice plants. *Mol. Breed.* 13, 165–175.

Nakabayashi, K., Bartsch, M., Xiang, Y., Miatton, E., Pellengahr, S., Yano, R., et al. (2012). The time required for dormancy release in arabidopsis is determined by DELAY OF GERMINATION1 protein levels in freshly harvested seeds. *Plant Cell* 24, 2826–2838.

Nakajima, S., Ito, H., Tanaka, R., and Tanaka, A. (2012). Chlorophyll b reductase plays an essential role in maturation and storability of Arabidopsis seeds. *Plant Physiol.* 160, 261–273.

Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., and Shigeoka, S. (2006). Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant J.* 48, 535–547.

Noguero, M., Atif, R. M., Ochatt, S., and Thompson, R. D. (2013). The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Sci.* 209, 32–45.

Noguero, M., Le Signor, C., Vernoud, V., Bandyopadhyay, K., Sanchez, M., Fu, C., et al. (2015). DASH transcription factor impacts Medicago truncatula seed size by its action on embryo morphogenesis and auxin homeostasis. *Plant J.* 81, 453–466.

Norman, S. M., Bennett, R. D., Poling, S. M., Maier, V. P., and Nelson, M. D. (1986). Paclobutrazol Inhibits Abscisic Acid Biosynthesis in Cercospora rosicola . *Plant Physiol.* 80, 122–125.

Nover, L., Bharti, K., Döring, P., Mishra, S. K., Ganguli, A., and Scharf, K. D. (2001).

Arabidopsis and the heat stress transcription factor world: How many heat stress transcription factors do we need? *Cell Stress Chaperones* 6, 177–189.

Ogé, L., Bourdais, G., Bove, J., BorisCollet, Godin, B., Granier, F., et al. (2008). Protein repair L-Isoaspartyl methyltransferasel is involved in both seed longevity and germination vigor in arabidopsis. *Plant Cell* 20, 3022–3037.

Ohama, N., Sato, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2017). Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends Plant Sci.* 22, 53–65.

Okamoto, M., Kuwahara, A., Seo, M., Kushiro, T., Asami, T., Hirai, N., et al. (2006). CYP707A1 and CYP707A2, which encode abscisic acid 8′-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. *Plant Physiol.* 141, 97–107.

Oliver, M. J., Farrant, J. M., Hilhorst, H. W. M., Mundree, S., Williams, B., and Bewley, J. D. (2020). Desiccation Tolerance: Avoiding Cellular Damage during Drying and Rehydration. *Annu. Rev. Plant Biol.* 71, 435–460.

Olszewski, N., Sun, T., and Gubler, F. (2002). Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant Cell* 14 Suppl, S61-80.

Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garciarrubio, A., and Covarrubias, A. A. (2010). Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in arabidopsis. *Plant Physiol.* 154, 373–390.

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

Papi, M., Sabatini, S., Bouchez, D., Camilleri, C., Costantino, P., and Vittorioso, P. (2000). Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. *Genes Dev.* 14, 28–33.

Patriyawaty, N. R., Rachaputi, R. C. N., and George, D. (2018). Physiological mechanisms underpinning tolerance to high temperature stress during reproductive phase in mungbean (Vigna radiata (L.) Wilczek). *Environ. Exp. Bot.* 150, 188–197.

Pecrix, Y., Staton, S. E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., et al. (2018). Whole-genome landscape of Medicago truncatula symbiotic genes. *Nat. Plants* 4, 1017–1025.

Peleg, Z., and Blumwald, E. (2011). Hormone balance and abiotic stress tolerance in crop plants. *Curr. Opin. Plant Biol.* 14, 290–295.

Pellizzaro, A., Neveu, M., Lalanne, D., Ly Vu, B., Kanno, Y., Seo, M., et al. (2020). A role for auxin signaling in the acquisition of longevity during seed maturation. *New Phytol.* 225, 284–296.

Penfield, S., and MacGregor, D. R. (2017). Effects of environmental variation during seed production on seed dormancy and germination. *J. Exp. Bot.* 68, 819–825.

Peng, F. Y., and Weselake, R. J. (2013). Genome-wide identification and analysis of the B3 superfamily of transcription factors in Brassicaceae and major crop plants. *Theor. Appl. Genet.* 126, 1305–1319.

Pereira Lima, J. J., Buitink, J., Lalanne, D., Rossi, R. F., Pelletier, S., da Silva, E. A. A., et al. (2017). Molecular characterization of the acquisition of longevity during seed maturation in soybean. *PLoS One* 12, e0180282.

Pratt, W. B., and Toft, D. O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol. Med.* 228, 111–133.

Prieto-Dapena, P., Castaño, R., Almoguera, C., and Jordano, J. (2006). Improved resistance to controlled deterioration in transgenic seeds. *Plant Physiol.* 142, 1102–1112.

Prieto-Dapena, P., Castaño, R., Almoguera, C., and Jordano, J. (2008). The ectopic overexpression of a seed-specific transcription factor, HaHSFA9, confers tolerance to severe dehydration in vegetative organs. *Plant J.* 54, 1004–1014.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.

Qu, A. L., Ding, Y. F., Jiang, Q., and Zhu, C. (2013). Molecular mechanisms of the plant heat stress response. *Biochem. Biophys. Res. Commun.* 432, 203–207.

Renzi, J. P., Duchoslav, M., Brus, J., Hradilová, I., Pechanec, V., Václavek, T., et al. (2020).

Physical Dormancy Release in Medicago truncatula Seeds Is Related to Environmental Variations. *Plants* 9, 503.

Righetti, K., Vu, J. L., Pelletier, S., Vu, B. L., Glaab, E., Lalanne, D., et al. (2015). Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *Plant Cell* 27, 2692–2708.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., and Mittler, R. (2004). When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. *Plant Physiol.* 134, 1683–1696.

Robert, H. S. (2019). Molecular Communication for Coordinated Seed and Fruit Development: What Can We Learn from Auxin and Sugars? *Int. J. Mol. Sci.* 20, 936.

Robert, H. S., Park, C., Gutièrrez, C. L., Wójcikowska, B., Pěnčík, A., Novák, O., et al. (2018). Maternal auxin supply contributes to early embryo patterning in Arabidopsis. *Nat. Plants* 4, 548–553.

Ronfort, J., Bataillon, T., Santoni, S., Delalande, M., David, J. L., and Prosperi, J. M. (2006). Microsatellite diversity and broad scale geographic structure in a model legume: Building a set of nested core collection for studying naturally occurring variation in Medicago truncatula. *BMC Plant Biol.* 6.

Roscoe, T. T., Guilleminot, J., Bessoule, J. J., Berger, F., and Devic, M. (2015). Complementation of seed maturation phenotypes by ectopic expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in arabidopsis. *Plant Cell Physiol.* 56, 1215–1228.

Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., et al. (2006). Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell* 18, 1292–1309.

Salvi, P., Saxena, S. C., Petla, B. P., Kamble, N. U., Kaur, H., Verma, P., et al. (2016). Differentially expressed galactinol synthase(s) in chickpea are implicated in seed vigor and longevity by limiting the age induced ROS accumulation. *Sci. Rep.* 6, 35088.

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

Arabidopsis. *Plant J.* 54, 608–620.

Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L. (2005). LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. *FEBS Lett.* 579, 4666–4670.

Schaefer, R. J., Michno, J. M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., et al. (2018). Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell* 30, 2922–2942.

Scheelbeek, P. F. D., Bird, F. A., Tuomisto, H. L., Green, R., Harris, F. B., Joy, E. J. M., et al. (2018). Effect of environmental changes on vegetable and legume yields and nutritional quality. *Proc. Natl. Acad. Sci. U. S. A.* 115, 6804–6809.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.

Schramm, F., Larkindale, J., Kiehlmann, E., Ganguli, A., Englich, G., Vierling, E., et al. (2008). A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of Arabidopsis. *Plant J.* 53, 264–274.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins. *Cell* 128, 735–745.

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol. Plant* 12, 879–892.

Scott, R. J., Spielman, M., Bailey, J., and Dickinson, H. G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. *Development* 125, 3329–3341.

Sehgal, A., Sita, K., Siddique, K. H. M., Kumar, R., Bhogireddy, S., Varshney, R. K., et al. (2018). Drought or/and heat-stress effects on seed filling in food crops: Impacts on functional biochemistry, seed yields, and nutritional quality. *Front. Plant Sci.* 871.

Seo, M., Hanada, A., Kuwahara, A., Endo, A., Okamoto, M., Yamauchi, Y., et al. (2006). Regulation of hormone metabolism in Arabidopsis seeds: Phytochrome regulation of abscisic acid metabolism and abscisic acid regulation of gibberellin metabolism. *Plant J.* 48, 354–366.

Seo, M., and Koshiba, T. (2002). Complex regulation of ABA biosynthesis in plants. *Trends Plant Sci.* 7, 41–48.

Shirley, B. W. (1998). Flavonoids in seeds and grains: Physiological function, agronomic importance and the genetics of biosynthesis. *Seed Sci. Res.* 8, 415–422.

Shu, K., Zhang, H., Wang, S., Chen, M., Wu, Y., Tang, S., et al. (2013). ABI4 Regulates Primary Seed Dormancy by Regulating the Biogenesis of Abscisic Acid and Gibberellins in Arabidopsis. *PLoS Genet.* 9, e1003577.

Siebers, M. H., Yendrek, C. R., Drag, D., Locke, A. M., Rios Acosta, L., Leakey, A. D. B., et al. (2015). Heat waves imposed during early pod development in soybean (Glycine max) cause significant yield loss despite a rapid recovery from oxidative stress. *Glob. Chang. Biol.* 21, 3114–3125.

Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J., and Holmes, I. H. (2009). JBrowse: A next-generation genome browser. *Genome Res.* 19, 1630–1638.

Skubacz, A., Daszkowska-Golec, A., and Szarejko, I. (2016). The role and regulation of ABI5 (ABA-insensitive 5) in plant development, abiotic stress responses and phytohormone crosstalk. *Front. Plant Sci.* 7.

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., et al. (2013). Candidate Genes and Genetic Architecture of Symbiotic and Agronomic Traits Revealed by Whole-Genome, Sequence-Based Association Genetics in Medicago truncatula. *PLoS One* 8, e65688.

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

Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., et al. (2008). Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3151–3156.

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., et al. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11806–11811.

Sun, W., Bernard, C., Van Cotte, B. De, Van Montagu, M., and Verbruggen, N. (2001). At-HSP17.6A, encoding a small heat-shock protein in Arabidopsis, can enhance

osmotolerance upon overexpression. *Plant J.* 27, 407–415.

Sundaresan, V. (2005). Control of seed size in plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17887–17888.

Suzuki, N., and Mittler, R. (2006). Reactive oxygen species and temperature stresses: A delicate balance between signaling and destruction. *Physiol. Plant.* 126, 45–51.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., et al. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. *Plant J.* 54, 335–347.

Tan, B. C., Joseph, L. M., Deng, W. T., Liu, L., Li, Q. B., Cline, K., et al. (2003). Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. *Plant J.* 35, 44–56.

Tao, Z., Shen, L., Gu, X., Wang, Y., Yu, H., and He, Y. (2017). Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature* 551, 124–128.

Tejedor-Cano, J., Prieto-Dapena, P., Almoguera, C., Carranco, R., Hiratsu, K., Ohme-Takagi, M., et al. (2010). Loss of function of the HSFA9 seed longevity program. *Plant, Cell Environ.* 33, 1408–1417.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192.

Tibbs Cortes, L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome*, 1–17.

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

Tognetti, P. M., Mazia, N., and Ibáñez, G. (2019). Seed local adaptation and seedling plasticity account for Gleditsia triacanthos tree invasion across biomes. *Ann. Bot.* 124, 307–318.

Vadez, V., Berger, J. D., Warkentin, T., Asseng, S., Ratnakumar, P., Rao, K. P. C., et al. (2012). Adaptation of grain legumes to climate change: A review. *Agron. Sustain. Dev.* 32, 31–44.

Valladares, F., Sanchez-Gomez, D., and Zavala, M. A. (2006). Quantitative estimation of phenotypic plasticity: Bridging the gap between the evolutionary concept and its ecological applications. *J. Ecol.* 94, 1103–1116.

Van Zanten, M., Koini, M. A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., et al. (2011). Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20219–20224.

Vandecasteele, C., Teulat-Merah, B., Morère-Le Paven, M. C., Leprince, O., Ly Vu, B., Viau, L., et al. (2011). Quantitative trait loci analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in Medicago truncatula. *Plant, Cell Environ.* 34, 1473–1487.

Verdier, J., Kakar, K., Gallardo, K., Le Signor, C., Aubert, G., Schlereth, A., et al. (2008). Gene expression profiling of M. truncatula transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol. Biol.* 67, 567–580.

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., et al. (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of medicago truncatula seeds. *Plant Physiol.* 163, 757–774.

Verdier, J., Leprince, O., and Buitink, J. (2019). " A physiological perspective of late maturation processes and establishment of seed quality in Medicago truncatula seeds ," in *The Model Legume Medicago truncatula* (Hoboken, NJ, USA: John Wiley & Sons, Inc.), 44–54.

Verdier, J., and Thompson, R. D. (2008). Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* 49, 1263–1271.

Verma, P., Kaur, H., Petla, B. P., Rao, V., Saxena, S. C., and Majee, M. (2013). PROTEIN L-ISOASPARTYL METHYLTRANSFERASE2 Is Differentially expressed in chickpea and enhances seed vigor and longevity by reducing abnormal isoaspartyl accumulation predominantly in seed nuclear proteins. *Plant Physiol.* 161, 1141–1157.

Vierling, E. (1991). The roles of heat shock proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42, 579–620.

Wahid, A., Gelani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: An

overview. *Environ. Exp. Bot.* 61, 199–223.

Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A., Berger, F., et al. (2010). The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. *Plant J.* 63, 670–679.

Wang, Y., Li, L., Ye, T., Zhao, S., Liu, Z., Feng, Y. Q., et al. (2011). Cytokinin antagonizes ABA suppression to seed germination of Arabidopsis by downregulating ABI5 expression. *Plant J.* 68, 249–261.

Waters, E. R., Lee, G. J., and Vierling, E. (1996). Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* 47, 325–338.

West, M. A. L., Yee, K. M., Danao, J., Zimmerman, J. L., Fischer, R. L., Goldberg, R. B., et al. (1994). LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. *Plant Cell* 6, 1731–1745.

Windels, D., Dang, T. T., Chen, Z., and Verdier, J. (2020). Snapshot of epigenetic regulation in legumes. *Legum. Sci.*

Xiao, W., Brown, R. C., Lemmon, B. E., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (2006). Regulation of seed size by hypomethylation of maternal and paternal genomes. *Plant Physiol.* 142, 1160–1168.

Yamanouchi, U., Yano, M., Lin, H., Ashikari, M., and Yamada, K. (2002). A rice spotted leaf gene, Spl7, encodes a heat stress transcription factor protein. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7530–7535.

Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y., and Yamaguchi, S. (2004). Activation of Gibberellin Biosynthesis and Response Pathways by Low Temperature during Imbibition of Arabidopsis thaliana Seeds. *Plant Cell* 16, 367–378.

Yan, B., Lv, Y., Zhao, C., and Wang, X. (2020). Knowing when to silence: roles of polycomb-group proteins in sam maintenance, root development, and developmental phase transition. *Int. J. Mol. Sci.* 21, 1–19.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL-and AtBMI1-Mediated H2Aub initiate the switch from embryonic to postgerminative growth in arabidopsis. *Curr. Biol.* 23, 1324–1329.

Yang, X., Tong, A., Yan, B., and Wang, X. (2017). Governing the silencing state of chromatin: The roles of polycomb repressive complex 1 in arabidopsis. *Plant Cell Physiol.* 58, 198–206.

Yoshida, T., Ohama, N., Nakajima, J., Kidokoro, S., Mizoi, J., Nakashima, K., et al. (2011). Arabidopsis HsfA1 transcription factors function as the main positive regulators in heat shock-responsive gene expression. *Mol. Genet. Genomics* 286, 321–332.

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524.

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* 16, 284–287.

Zhang, B., Li, C., Li, Y., and Yu, H. (2020). Mobile TERMINAL FLOWER1 determines seed size in Arabidopsis. *Nat. Plants* 6, 1146–1157.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9.

Zheng, B., and Chen, X. (2011). Dynamics of histone H3 lysine 27 trimethylation in plant development. *Curr. Opin. Plant Biol.* 14, 123–129.

Zhiguo, E., Li, T., Zhang, H., Liu, Z., Deng, H., Sharma, S., et al. (2018). A group of nuclear factor y transcription factors are sub-functionalized during endosperm development in monocots. *J. Exp. Bot.* 69, 2495–2510.

Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., et al. (2013). Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in arabidopsis under heat stress. *Plant Cell* 25, 2925–2943.

Zhou, Y., Zhang, X., Kang, X., Zhao, X., Zhang, X., and Ni, M. (2009). Short Hypocotyl Under Blue1 associates with Miniseed3 and Haiku2 promoters in vivo to regulate arabidopsis seed development. *Plant Cell* 21, 106–117.

Zinsmeister, J., Berriri, S., Basso, D. P., Ly-Vu, B., Dang, T. T., Lalanne, D., et al. (2020a). The seed-specific heat shock factor A9 regulates the depth of dormancy in Medicago truncatula seeds via ABA signalling. *Plant Cell Environ.* 43, 2508–2522.

Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Ly Vu, B., et al. (2016). ABI5 is a regulator of seed maturation and longevity in legumes. *Plant Cell* 28, 2735–2754.

Zinsmeister, J., Leprince, O., and Buitink, J. (2020b). Molecular and environmental factors regulating seed longevity. *Biochem. J.* 477, 305–323.

DOCTORAT ECOLOGIE
BRETAGNE GEOSCIENCES
LOIRE AGRONOMIE ALIMENTATION

l'institut Agro
agriculture • alimentation • environnement

AGRO CAMPUS OUEST

Titre : Impacts du stress thermique sur l'acquisition des caractères de qualité des semences chez *Medicago truncatula*

Mots clés : Maturation des graines, Stress thermique, Qualité des semences, Transcriptomique, Epigénétique, GWAS

Résumé : Les graines de légumineuses sont une source de nourriture importante pour garantir la sécurité alimentaire mondiale. Cependant, le réchauffement climatique met en danger les productions agricoles en affectant le rendement et la qualité des semences. Chez *M. truncatula*, une plante modèle des légumineuses, la maturation des graines est fortement affectée par les conditions environnementales, conduisant à l'altération de la qualité des graines. Dans cette étude, nous proposons d'explorer les processus moléculaires des semences aboutissant à la modification de la qualité des graines suite au stress chaleur pour identifier des gènes candidats régulant la réponse au stress et la plasticité phénotypique des caractères de qualité de semences.
Deux approches ont été développées pour identifier les gènes candidats : (i) décrire les régulations (épi)génétiques intervenant lors du développement des tissus des graines en réponse au stress thermique dans le génotype de référence *M. truncatula* A17, et (ii) exploiter la diversité naturelle présente dans la population HapMap de *M. truncatula* en utilisant des approches d'association pangénomique. Un ensemble de gènes candidats potentiellement régulant la réponse au stress thermique et la plasticité des graines a été identifié. Parmi eux, deux gènes ont été sélectionnés pour être caractérisés fonctionnellement et sur la base de résultats préliminaires, *MtHAP3* pourrait avoir un rôle dans la reprogrammation de la maturation des graines en réponse au stress thermique ; et *MtMIEL1* pourrait agir en régulateur de la plasticité de germination des graines en réponse à un stress thermique.

Title : Deciphering the impacts of heat stress on the acquisition of seed quality traits in *Medicago truncatula*

Keywords : Seed maturation, Heat stress, Seed quality, Transcriptomics, Epigenetics, GWAS

Abstract : Legume seeds represent a crucial source of human food and animal feed to ensure global food security. However, global warming and climate change endanger agricultural productions by impacting seed yield and seed quality. In *Medicago truncatula*, a model plant for legumes, the timing of seed maturation is greatly affected by adverse environmental growing conditions, leading to alterations of final seed traits. In this study, we intend to explore and describe seed molecular processes underlying the alteration of seed traits in response to heat stress to identify candidate genes regulating stress response and phenotypic plasticity of seed quality traits.
Two complementary approaches were implemented to identify candidate genes: (i) deciphering the (epi)genetic regulations occurring in developing seed tissues in response to heat stress in the reference genotype A17, and (ii) exploiting the natural diversity of the *Medicago truncatula* HapMap population using genome-wide association study approaches. As a result, we observed an intense molecular reprogramming of seed maturation processes via the alteration of the main regulators of seed development. A set of candidate genes potentially regulating heat stress response and plasticity in seeds was identified. Among them, two genes were selected for functional characterizations and based on preliminary results, *MtHAP3* showed a potential role in the reprogramming of early seed maturation determining seed size and seed longevity in response to heat stress; and *MtMIEL1* acted as a regulator of seed germination plasticity in response to heat stress.