



**HAL**  
open science

# Déterminisme génétique de la plasticité de la composition protéique des graines de légumineuses vis-à-vis de l'environnement : rôle du métabolisme du soufre

Kévin Cartelier

► **To cite this version:**

Kévin Cartelier. Déterminisme génétique de la plasticité de la composition protéique des graines de légumineuses vis-à-vis de l'environnement : rôle du métabolisme du soufre. Sciences agricoles. Université Bourgogne Franche-Comté, 2021. Français. NNT : 2021UBFCK012 . tel-04047141

**HAL Id: tel-04047141**

**<https://theses.hal.science/tel-04047141>**

Submitted on 27 Mar 2023

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

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

## **THESE DE DOCTORAT**

**DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE  
PREPAREE A INRAE-UMR 1347 Agroécologie, Pôle GEAPSI – Equipe FILEAS**

Ecole Doctorale n°554

Environnements - Santé

Doctorat de spécialité Sciences agronomiques

Par

**Kévin CARTELIER**

**DETERMINISME GENETIQUE DE LA PLASTICITE DE LA  
COMPOSITION PROTEIQUE DES GRAINES DE  
LEGUMINEUSES VIS-A-VIS DE L'ENVIRONNEMENT :  
ROLE DU METABOLISME DU SOUFRE**

Date de soutenance : 19 mars 2021

Composition du jury :

M. Jacques Le Gouis	Directeur de Recherche INRAE	Président du jury
M. Olivier Loudet	Directeur de Recherche INRAE	Rapporteur
M. Olivier Leprince	Professeur AgroCampus Ouest	Rapporteur
Mme Julia Buitink	Directrice de Recherche INRAE	Examinatrice
Mme Christine Le Signor	Ingénieure de Recherche INRAE	Co-directrice de thèse
Mme Karine Gallardo	Directrice de Recherche INRAE	Co-directrice de thèse

## Remerciements

Je tiens tout d'abord à remercier les membres de mon jury : les deux rapporteurs, Olivier Leprince et Olivier Loudet ainsi que les deux examinateurs, Julia Buitink et Jacques Le Gouis, pour avoir accepté d'étudier le condensé de ces 3 années de recherche. Je tiens aussi à remercier les membres de mon comité de pilotage de thèse : Judith Burstin, Nathalie Nesi et Jérôme Verdier pour leurs conseils bénéfiques à certains choix stratégiques du déroulement de la thèse.

Mes pensées vont ensuite tout naturellement à mes deux co-directrices, Christine Le Signor et Karine Gallardo, je tiens aussi à vous exprimer ma plus grande gratitude pour vos relectures poussées qui m'ont aidé à donner le meilleur de moi-même en matière de rédaction ainsi que pour vos conseils avisés qui m'ont permis de pousser mes réflexions toujours plus loin. Je vous remercie aussi d'avoir toujours été là pour me soutenir durant 3 années (et même un peu plus) malgré les aléas (#pandémie, #confinements, #télétravail). J'espère, à travers ce manuscrit de thèse, avoir réussi à raconter une histoire intéressante à partir de toutes les données produites comme vous m'avez souvent conseillé de le faire.

Je tiens à remercier toutes les personnes de l'équipe FILEAS qui m'ont aidé lors des récoltes de pois, dont certains géotypes mesuraient plus de 4 mètres, ce qui n'en facilitait pas le transport. Je remercie Delphine Aimé pour son aide lors des analyses protéomiques (du broyage des graines jusqu'à l'analyse des gels), ainsi que lors de la préparation des solutions nutritives. Je remercie Myriam Sanchez pour la caractérisation de la position de l'insert Tnt1 du mutant HMT3. Je remercie Vanessa Vernoud pour m'avoir permis d'utiliser les données du GeneAtlas lors du développement de la graine de pois. Je la remercie, ainsi que Richard Thompson, d'avoir accepté de relire la publi medicago et de nous avoir fourni des commentaires qui nous ont été utiles. Je remercie aussi mon stagiaire Augustin pour son investissement lors de son stage et pour avoir CLIQé un nombre de fois non négligeable. Mes remerciements vont aussi à Aurélien pour sa débrouillardise, tu as su attacher les plantes avec ingéniosité ainsi qu'à Mathilde (dite Strass) qui n'est pas restée très longtemps dans notre équipe, qui nous a aidé à l'attachage des plantes et avec qui le feeling est tout de suite passé. Je tiens aussi à remercier les MOO, Justine, Corentin et Lilou qui sont venus effectuer le broyage et/ou le comptage des graines. Merci aussi à Thomas Badie pour son aide sur les analyses statistiques des résultats de la manip pois.

Je remercie l'ensemble des personnes de l'équipe ECP pour leur aide, en particulier Judith Burstin pour m'avoir donné accès au génome du pois avant même la publication des séquences.

Je remercie aussi Judith Burstin ainsi que Virginie Bourion pour m'avoir permis d'utiliser les données de génotypage de la collection de pois obtenues dans le cadre de IA-PEAMUST et ANR-GRASP. Je remercie Jonathan Kreplak pour les très nombreuses fois où il m'a aidé que ce soit pour la prise en main de ce génome ou pour des problèmes informatiques. Merci aussi à Myriam Huart et à Jennifer Regulier qui ont analysé par NIRS les poudres de graines de pois.

Je remercie aussi Pascal Joly de l'AGAP, Laëticia Ly Vu, Daniel Sochard, Benoit Ly Vu et Daniel Beucher et la plateforme phénotypique de l'IRHS ainsi que leurs équipes pour nous avoir fourni les graines des plantes de *Medicago* soumises aux stress. Je suis reconnaissant envers les serristes pour leur engagement dans les manips, en particulier Karine Palavioux et Damien Gironde qui se sont toujours démenés pour trouver des solutions aux problèmes que nous avons pu rencontrer. Je remercie aussi Christophe Salon qui a autorisé la poursuite du développement des plantes de *Medicago* durant le premier confinement, ainsi que Franck Zenk et Julien Martinet qui ont soigneusement suivi ces plantes pendant cette période. Leur contribution m'a permis d'acquérir les données menant à la validation du mutant *HMT3*.

Je tiens aussi à remercier Momo, Mégane, Amandine, Coraline, Ludivine, Valentin, Aurélie, Adrien, Aude, Mickaël et Thibault avec qui j'ai pu passer de bons moments dans et hors du bâtiment Mendel. Je remercie aussi Dīnia et Ingvild pour nos repas du midi, moments privilégiés des potins quotidiens. Je remercie aussi les membres organisateurs du Beer-time, Damien et Séverin, vous m'avez permis de rencontrer plein de belles personnes. Je tiens à remercier Estelle, Chulie et Gayane qui ont toujours été là quand j'en avais besoin. Je remercie aussi toutes les autres personnes que j'ai pu rencontrer, Caro et Rico, Inès, Gwladys, Auxence et en général beaucoup de personnes du bâtiment Costes, qui ont su, de par leur dynamisme, instaurer une ambiance conviviale et énergique au sein de l'INRA mais aussi hors de ses murs.

Je remercie Adrien, Audrey et Quentin pour leurs conversations toujours aussi enrichissantes sur les petits moments du quotidien.

Je remercie aussi mes parents pour leur soutien psychologique.

Je remercie belle maman de m'avoir accepté durant le premier confinement.

Je tiens enfin à remercier tout particulièrement Maïana (dite la bichette), pour m'avoir soutenu et (surtout) supporté tout au long de cette intense fin de thèse.

## Valorisation des travaux de thèse

Genetic determinants of seed protein plasticity in response to the environment in *Medicago truncatula*. Cartelier et al., 2021. The Plant Journal

## Liste des présentations issues des travaux de thèse

K. Cartelier, K. Gallardo, C. Le Signor. « Optimiser et stabiliser la composition protéique des graines de légumineuses ». Journée des doctorants de l'UMR Agroécologie, Dijon, France, 15 mai 2018.

**Poster**

K. Cartelier, K. Gallardo, C. Le Signor. « Optimiser et stabiliser la composition protéique des graines de légumineuses ». Forum des Jeunes Chercheurs de l'école doctorale Environnement-Santé, Besançon, France, 14 et 15 juin 2018. **Poster**

K. Cartelier, K. Gallardo, C. Le Signor. « Optimizing and stabilizing protein composition of legume seeds ». Journée du département INRA de Biologie et Amélioration des Plantes, Paris, France, 20 et 21 septembre 2018. **Poster**

K. Cartelier, K. Gallardo, C. Le Signor. « Optimiser et stabiliser la composition protéique des graines de légumineuses ». 2<sup>ème</sup> rencontres Francophones sur les Légumineuses, Toulouse, France, 17 et 18 octobre 2018. **Poster**, Prix du meilleur poster

K. Cartelier, K. Gallardo, C. Le Signor. « Elucidating the genetic determinism of the plasticity of seed proteins in response to the environment using *Medicago truncatula* ». International Conference on Legume Genetics and Genomics 2019, Dijon, France, 13 au 17 mai 2019. **Communication orale**

K. Cartelier, K. Gallardo, C. Le Signor. « Identification des gènes contrôlant la plasticité de la composition protéique vis-à-vis de l'environnement ». Colloque GRAINES 2019, Angers, France, 21 au 23 mai 2019. **Communication orale**

K. Cartelier, K. Gallardo, C. Le Signor. « Identification des déterminismes génétiques de la plasticité de la composition protéique des graines ». Forum des Jeunes Chercheurs de l'école doctorale Environnement-Santé, Dijon, France, 13 et 14 juin 2019. **Communication orale**

# Table des matières

Liste des abréviations et sigles.....	1
Liste des figures .....	4
Liste des tableaux.....	7
Liste des tableaux supplémentaires.....	8
Liste des annexes .....	9
Introduction générale.....	10
CHAPITRE I :.....	15
Synthèse bibliographique .....	15
1. Atout des légumineuses et ressources génétiques et génomiques .....	16
1.1 Avantages agroécologiques et nutritionnels .....	16
1.2 Le pois.....	16
1.3 <i>Medicago truncatula</i> , espèce modèle des légumineuses .....	20
2. Développement de la graine et accumulation des réserves .....	24
2.1 Développement de la graine.....	24
2.2 Composition des graines de légumineuses.....	25
2.3 Synthèse des protéines de réserve.....	27
2.4 Synthèse des acides aminés soufrés .....	28
2.5 Stratégies visant à modifier le contenu en acides aminés des graines .....	33
3. Variations environnementales : impact sur le rendement et la composition protéique des graines de légumineuses .....	35
3.1 Impact des variations environnementales sur le développement des plantes .....	35
3.2 Les stress nutritionnels : le cas d'une carence en soufre.....	36
4. Déterminismes génétiques du rendement et de la composition protéique des graines de légumineuses .....	38
4.1 Plasticité vis à vis des fluctuations environnementales .....	38
4.2 Etude du déterminisme génétique du rendement et de la composition protéique des graines de légumineuses.....	40
CHAPITRE II :.....	43
Etude du déterminisme génétique de la plasticité de la composition protéique des graines de <i>M. truncatula</i> en réponse à des variations environnementales.....	43
1. Introduction .....	48
2. Results .....	50
2.1 Genotype and environment effects on seed protein composition in <i>M. truncatula</i> ....	50
2.2 Phenotypic plasticity of major seed proteins across environments.....	51

2.3 Identifying loci and candidate genes for seed globulin plasticity.....	52
2.4 Identification of biological processes that are likely to contribute to seed globulin plasticity.....	54
2.5 Abiotic stress regulation of candidate genes for globulin plasticity.....	56
2.6 <i>HMT3</i> is preferentially expressed in seeds and contributes to globulin plasticity.....	56
2.7 Candidate regulators of seed protein plasticity.....	58
3. Discussion.....	59
3.1 Met recycling is a key contributor to globulin composition establishment and plasticity.....	59
3.2 Regulators were identified that might mitigate the effect of environmental stresses on seed globulin composition.....	61
3.3 The maintenance of DNA integrity is likely to play a role in plasticity of seed protein accumulation.....	62
4. Materials and Methods.....	63
4.1 Seed production for the ecotype collection in <i>M. truncatula</i> .....	63
4.2 Seed protein separation and quantification.....	64
4.3 Identification of mature seed proteins.....	64
4.4 Calculation of plasticity indices.....	65
4.5 Core collection genotyping data.....	65
4.6 Procedure for GWAS and gene selection.....	66
4.7 Growing and genotyping <i>hmt3</i> insertion mutants.....	66
CHAPITRE III : .....	68
Etude de la variabilité de réponse à une carence en soufre au sein d'une collection de pois et du déterminisme génétique de cette réponse.....	68
1. Application d'une carence en soufre et caractérisation des conditions.....	71
2. Variabilité et déterminisme génétique de la plasticité de la composition protéique des graines de pois vis-à-vis d'une carence en soufre.....	74
2.1 Effet du génotype et d'une carence en soufre sur la composition protéique des graines.....	74
2.2 Plasticité de la composition protéique des graines de pois.....	79
2.3 Exploration par GWAS des régions du génome contrôlant la composition protéique des graines de pois et sa plasticité vis-à-vis de l'environnement.....	81
2.4 Etablissement d'un répertoire de gènes candidats pour la composition protéique des graines de pois et sa régulation sous contrainte soufrée.....	83
2.5 Gènes candidats pour moduler la composition protéique des graines dans des environnements pauvres en soufre.....	86
3. Variabilité et contrôle génétique des composantes du rendement au sein d'une collection de pois sous contrainte soufrée.....	97

3.1 Effets de la carence en soufre sur les composantes du rendement .....	97
3.2 Variabilité génétique pour la réponse à la carence en soufre.....	101
3.3 Analyses GWAS des composantes du rendement .....	107
4. Matériels et Méthodes .....	114
4.1 Dispositif expérimental et production des génotypes de pois .....	114
4.2 Séparation des protéines de réserve des graines et quantification .....	116
4.3 Identification et annotation des protéines de réserve des graines .....	117
4.4 Calcul de l'indice de plasticité .....	117
4.5 Génotypage et structuration de la core collection de pois .....	118
4.6 Analyses de variance, ACP et CAH.....	119
4.7 Procédure GWAS et sélection de gènes candidats.....	120
4.8 Données transcriptomiques exploitées pour l'interprétation des résultats.....	121
CHAPITRE IV : .....	123
Conclusion générale et perspectives.....	123
1. Synténie entre les hotspots de composition protéique détectés chez les deux espèces ..	124
2. Gènes de plasticité communs aux deux espèces.....	127
3. Le métabolisme du soufre est un point de contrôle majeur de la plasticité de la composition protéique chez les deux espèces.....	128
4. Gènes candidats liés à la plasticité des composantes du rendement en réponse à une carence en soufre chez le pois. ....	130
5. Perspectives de validation des gènes candidats .....	130
Références .....	132
ANNEXES .....	154

## Liste des abréviations et sigles

**AAP6** : Amino acid permease 6  
**ACP** : Analyse en composantes principales  
**ADI** : Acireductone dioxygenases  
**AdoHcy** : *S*-Adenosyl-1-homocysteine  
**AdoMet** : *S*-adenosylmethionine  
**Af** : Afila  
**AMS** : Architecture and MultiStress  
**ANOVA** : Analyse de variance  
**AP2-EREBP** : Apetala2-Ethylene-Responsive Element Binding Proteins  
**APK** : Adenosine-5'-phosphosulfate kinase  
**APR** : APS réductase  
**APR2** : ADENOSINE 5'-PHOSPHOSULFATE REDUCTASE2  
**APRT** : Adenine phosphoribosyltransferase  
**APS** : Adenylyl-sulfate  
**ARD** : Acidoreductone dioxygenase  
**Arg** : Arginine  
**ASD** : Aspartate-semialdehyde dehydrogenase  
**Asp** : Aspartate  
**Asp-P** : L-4-aspartyl phosphate  
**AspSA** : Asp semialdehyde  
**BASS6** : BILE ACID SODIUM SYMPORTER 6  
**BAT** : BIDIRECTIONAL AMINO ACID TRANSPORTER  
**BAT5** : BILE ACID TRANSPORTER 5,  
**BH** : Benjamini-Hochberg  
**CAH** : Classification Ascendante Hiérarchique  
**CBL** : Cystathionine beta lyase  
**CGS** : Cystathionine  $\gamma$ -synthase  
**conv** : Convicilines  
**COP** : Coat protein complexes  
**CPS** : Carbamoyl-phosphate synthase (glutamine-hydrolysing)  
**CRB** : Centre de Ressources Biologiques  
**CRK** : Cysteine-rich Receptor-like protein Kinase  
**CV** : Coefficient de variation  
**CXIP1** : (H<sup>+</sup>)/Ca<sup>2+</sup>) exchanger CAX-interacting protein-1  
**Cys** : Cysteine  
**CysC1** :  $\beta$ -Cyano-alanine synthase  
**DAP-AT** : LL-diaminopimelate aminotransferase  
**DHKMP** : 12-dihydro-3-keto-5-methylthiopentene  
**DIN10** : Dark inducible 10  
**DREB** : Dehydration-responsive element binding  
**FAO** : Food and Agriculture Organization of the United Nations  
**FW** : Bayesian Finlay-Wilkinson regression  
**G6PD** : Glucose-6-phosphate dehydrogenase  
**GATT** : General Agreement on Tariffs and Trade

**GEMMA** : Genome-wide Efficient Mixed Model Association  
**GGCT** : Gamma-glutamyl cyclotransférase  
**Gln** : Glutamine  
**Glu** : Glutamate  
**GO** : Gene ontology  
**GRaSP** : Genetics of Rhizobia Selection by Pea  
**GSH** : Glutathion  
**GSSG** : Disulfure de glutathion  
**Guanine-7-MTase** : Guanine-7-méthyltransférase  
**GWAS** : Genome Wide Association Study  
**HET-P** : 4-méthyl-5-β-hydroxyéthylthiazole  
**HMP-PP** : 4-amino-2-méthyl-5-hydroxyméthylpyrimidine diphosphate  
**HMT** : Homocystéine S-méthyltransférase  
**hTrm4MTase** : tRNA (cytosine34-C5)-méthyltransférase  
**JAP** : Jours après pollinisation  
**KMTB** : α-keto γ-méthyl thiobutyrate  
**LD** : Linkage disequilibrium  
**LEC1** : Leafy Cotyledon 1  
**MAF** : Minimum allele frequency  
**MATE** : Multidrug And Toxic Compound Extrusion  
**Met** : Méthionine  
**Met-GSL** : Glucosinolates dérivés de la méthionine  
**MMT** : Méthionine S-méthyltransférase  
**MRI** : Méthylthioribose-1-phosphate isomerase  
**MSA1** : More Sulfur accumulation 1  
**MTA** : 5'-méthylthioadénosine  
**MTase** : Méthyltransférase  
**mTERF** : Mitochondrial Transcription Termination Factor  
**MTI** : 5'-méthylthioribose-1-phosphate isomérase  
**MTO2** : Methionine over-accumulator 2  
**MTOB** : 4-méthylthio-2-oxobutanoate  
**MTR** : 5'-méthylthioribose  
**MTR-P** : 5'-méthylthioribose-1-phosphate  
**NAD+** : Nicotinamide adénine dinucléotide  
**nb\_nodes\_repro** : Nombre de nœuds reproducteurs  
**nb\_nodes\_veg** : Nombre de nœuds végétatifs  
**nbgo** : Nombre de gousses  
**nbgr** : Nombre de graines  
**nbgrgo** : Nombre de graines par gousse  
**NF-YA** : Nuclear factor Y  
**NIRS** : Near-infrared spectroscopy  
**NIT4** : Nitrilase 4  
**NMN** : Nicotinamide mononucléotide  
**NMNAT** : NMN adenylyltransférase  
**NWS** : Normalized Weighted Spectra count

**OAS** : O-acétylséine  
**OASTL** : O-acétylséine (thiol)-lyase  
**OPH** : O-phosphohomoséine  
**Orc2** : Origin recognition complex second largest subunit 2  
**PA** : Poids sec des parties aériennes végétatives  
**PAPS** : 3'-Phosphoadénosine-5'-phosphosulfate  
**PARP** : Poly(ADP-ribose) polymerase  
**PCA** : Principal component analysis  
**pds1go** : Poids d'une gousse  
**pds1gr** : Poids d'une graine  
**pdsgo** : Poids total de gousses  
**pdsgr** : Poids total de graines  
**PG** : Plasticity genes  
**PMG** : Poids de mille grains  
**PQL** : Protein Quantity Loci  
**PRMT** : Protein arginine methyltransferase  
**PSG** : Plasticity specific genes  
**PSV** : Protein storage vacuole  
**RE** : Réticulum endoplasmique  
**RFOs** : Raffinose family oligosaccharides  
**ROS** : Reactive oxygen species  
**SAT** : Sérine acétyltransférase  
**SBP1** : Selenium-Binding Protein 1  
**SDI** : Sulfur Deficiency Induced  
**Ser** : Sérine  
**SHMT** : Sérine hydroxyméthyltransférase  
**Sip1** : Seed imbibition protein  
**SIR** : Sulfite réductase  
**SK11** : SHAGGY-related kinase  
**SMM** : S-méthylméthionine  
**SNARE** : Soluble N-ethylmaleimide-sensitive factor Attachment protein REceptor  
**SNP** : Single Nucleotide Polymorphism  
**SQD2** : Sulfoquinovosyltransférase 2  
**SQDG** : Sulfoquinovosyldiacylglycérol  
**SST** : Somme totale des carrés  
**SULTR** : Sulfate transporter  
**TAIR** : The Arabidopsis Information Resource  
**TF** : Transcription factor  
**thiE** : Thiamine phosphate synthase  
**TILLING** : Targeting Induced Local Lesions in Genomes  
**TMP** : Thiamine monophosphate  
**tprot** : Teneur protéique des graines  
**TTG1** : TRANSPARENT TESTA GLABRA1  
**Vic** : Vicilines

## Liste des figures

**Figure 1 :** Production des protéagineux en France.

**Figure 2 :** Répartition géographique des surface cultivées de pois en France en 2019.

**Figure 3 :** Schéma des principaux stades du cycle de développement du pois.

**Figure 4 :** Caractéristiques morphologiques des graines de pois.

**Figure 5 :** Diagramme de la synténie entre le pois et *M. truncatula*.

**Figure 6 :** Répartition géographique des individus de *M. truncatula* présents dans notre étude.

**Figure 7 :** Quelques caractéristiques morphologiques de *M. truncatula*.

**Figure 8 :** Phases de développement de la graine de *M. truncatula*.

**Figure 9 :** Impact des conditions environnementales sur la durée des phases de développement de la graine de *M. truncatula*.

**Figure 10 :** Schéma des différentes étapes de maturation des légumine dans les cotylédons.

**Figure 11 :** Modèle de transport et de mise en réserve des globulines.

**Figure 12 :** Le métabolisme du soufre chez les plantes.

**Figure 13 :** Effet de la modification de l'activité de la glutathion réductase sur l'assimilation du soufre.

**Figure 14 :** Cycle biogéochimique du soufre dans la nature.

**Figure 15 :** Exemple de normes de réactions.

**Figure 16 :** Illustration de la méthode de régression de Finlay et Wilkinson.

**Figure 17 :** Carte génétique des 27 metaQTL (mQTL) détectés pour des composantes du rendement chez le pois.

**Figure 18 :** Variabilité de la composition protéique des graines de 200 écotypes de *Medicago truncatula* cultivés dans quatre environnements.

**Figure 19 :** Plasticité de l'abondance relative des globulines selon les environnements.

**Figure 20 :** Position des hotspots de GWAS sur la v5 de la carte physique de *Medicago truncatula*.

**Figure 21 :** Gènes associés à la plasticité pour les métabolismes du soufre et de l'azote positionnés dans leurs voies respectives.

**Figure 22 :** Composition en globulines et plasticité des graines d'un mutant *hmt3* comparées aux graines de plantes sauvages.

**Figure 23 :** Date de floraison pour l'ensemble des géotypes de pois.

**Figure 24 :** Températures moyennes journalières durant le développement des géotypes de pois.

**Figure 25 :** Evolution des valeurs de SPAD suite à l'application d'une carence en soufre chez le pois.

**Figure 26 :** Variabilité de la composition protéique des graines des géotypes de pois cultivés en condition contrôle et en condition de carence en soufre appliquée deux semaines avant floraison.

**Figure 27 :** Répartition des individus en fonction de leur composition protéique selon l'année et l'application d'une carence en soufre.

**Figure 28 :** Dendrogramme réalisé à partir des géotypes analysés les deux années pour la composition protéique des graines de pois.

**Figure 29 :** Heatmap des corrélations entre les contenus relatifs des protéines de réserve mesurés chaque année dans chaque condition.

**Figure 30 :** Distribution des géotypes utilisés dans l'analyse en fonction de leur indice  $\beta$  pour les dix bandes protéiques et le ratio 7S/11S.

**Figure 31 :** Nombre de SNPs associés à la composition protéique des graines de pois et sa plasticité, et nombre de gènes correspondants.

**Figure 32 :** Position des principaux hotspots issus des analyses GWAS sur la carte génomique du pois.

**Figure 33 :** Nombre de gènes de pois communs aux différentes conditions pour chaque classe protéique et principales ontologies présentes.

**Figure 34 :** Gènes associés ou connectés aux métabolismes du soufre et de l'azote.

**Figure 35 :** Distribution des géotypes de pois selon les composantes du rendement.

**Figure 36 :** Corrélations entre variables des parties végétatives du pois et celles des composantes du rendement en condition contrôle et de carence en soufre.

**Figure 37 :** Structuration de la population de pois en fonction des composantes du rendement.

**Figure 38 :** Analyse en composante principale mettant en évidence la distribution des géotypes utilisés dans l'analyse en fonction de leur indice  $\beta$  pour les variables de rendement.

**Figure 39 :** Heatmap des corrélations entre les variables de rendement et leur indice de plasticité associé.

**Figure 40 :** Performance de 6 géotypes pour le poids d'une graine de pois en fonction des valeurs environnementales estimées.

**Figure 41 :** Nombre de SNPs associés aux composantes du rendement et à la teneur protéique en  $S^+$ , en  $S^-$  et à l'indice de plasticité  $\beta$ , et nombre de gènes sélectionnés.

**Figure 42 :** Position des principaux hotspots issus des analyses GWAS du rendement sur la carte génomique du pois.

**Figure 43 :** Gène identifié pour la plasticité du poids d'une graine en réponse à une carence en soufre.

**Figure 44 :** Effet de la mise en cofacteur de *Le* sur une analyse GWAS.

**Figure 45 :** Impact de la mise en cofacteur d'un gène majeur à effet pléiotropique sur les résultats de GWAS.

**Figure 46 :** Plan d'expérimentation en serre pour la culture des génotypes de pois.

**Figure 47 :** Principaux hotspots associés à la composition protéique des graines communs entre le pois et *M. truncatula*.

**Figure 48 :** Hotspots majeurs associés à la composition protéique des graines conservés entre le pois et *M. truncatula*.

## Liste des tableaux

**Tableau 1 :** La régulation de l'assimilation du sulfate chez les plantes.

**Tableau 2 :** Les différentes approches utilisées pour augmenter le contenu en méthionine dans les graines.

**Tableau 3 :** Analyse d'enrichissement en ontologies pour la liste de gènes identifiés par GWAS pour la plasticité des globulines.

**Tableau 4 :** Descriptif de la collection selon les 5 groupes génétiques.

**Tableau 5 :** Variabilité de la composition protéique des graines de pois à l'échelle de la population.

**Tableau 6 :** Analyse de l'effet génotypique et environnemental sur la composition protéique des graines de pois.

**Tableau 7 :** Plasticité des bandes protéiques dans les graines de pois.

**Tableau 8 :** Gènes candidats associés au métabolisme du soufre.

**Tableau 9 :** Effet d'une carence en soufre sur des variables d'architecture et de rendement du pois.

**Tableau 10 :** Analyse de l'effet génotypique et environnemental sur des composantes du rendement et des variables liées aux parties végétatives aériennes chez le pois.

**Tableau 11 :** Indice de plasticité de variables de composantes du rendement du pois en réponse à une carence en soufre deux années.

**Tableau 12 :** Gènes associés à la plasticité de la composition protéique des graines communs entre l'analyse menée chez le pois et *M. truncatula*.

## Liste des tableaux supplémentaires

**Tableau S1 :** Liste des 200 écotypes de *M. truncatula* utilisés dans l'étude.

**Tableau S2 :** Protéines identifiées par LC-MS/MS dans les bandes issues des gels d'électrophorèse 1D de graines de *M. truncatula*.

**Tableau S3 :** Termes d'ontologies enrichis dans la liste des gènes identifiés par GWAS pour la composition protéique des graines dans chacun des quatre environnements.

**Tableau S4 :** Liste des gènes candidats pour la plasticité en globulines et leur régulation en réponse à des stress abiotiques.

**Tableau S5 :** Gènes liés à la plasticité portant des SNPs dans leur séquence induisant un changement d'acide aminé, un changement dans l'épissage ou l'apparition d'un codon stop.

**Tableau S6 :** Gènes de plasticité associés aux voies métaboliques du soufre et de l'aspartate, incluant le métabolisme de la méthionine.

**Tableau S7 :** Analyse d'enrichissement dans les différents jeux de données post-GWAS en gènes préférentiellement exprimés dans les graines ou régulés en réponse à des stress abiotiques.

**Tableau S8 :** Gènes de plasticité préférentiellement exprimés dans les graines et leur régulation dans différentes conditions de stress.

**Tableau S9 :** Gènes de plasticité annotés comme régulateurs et leur réponse à différentes conditions de stress.

**Tableau S10 :** Liste des gènes candidats pour la composition protéique des graines chez le pois en  $S^+$ , en  $S^-$  et en  $\beta$ .

**Tableau S11 :** Liste des gènes porteurs de SNPs dans leur séquence induisant un changement d'acide aminé, un changement dans l'épissage ou l'apparition d'un codon stop prématurément.

**Tableau S12 :** Liste des gènes candidats pour la plasticité de la composition protéique des graines de pois en réponse à une carence en soufre.

**Tableau S13 :** Liste des gènes annotés "régulateurs" candidats pour la composition protéique des graines en  $S^+$ , en  $S^-$  et en  $\beta$ .

**Tableau S14 :** Liste des gènes associés au transport et candidats pour la composition protéique des graines en  $S^-$  et en  $\beta$ .

**Tableau S15 :** Liste des gènes associés aux modifications post-traductionnelle des protéines et candidats pour la composition protéique des graines en  $S^-$  et en  $\beta$ .

**Tableau S16 :** Liste des gènes associés aux variables de composantes du rendement de la graine et de la teneur protéique.

## Liste des annexes

**Annexe 1 :** Résultats des analyses GWAS menées sur la plasticité de la composition protéique des graines chez *M. truncatula*.

**Annexe 2 :** Les familles de facteurs de transcription les plus représentées dans la liste de gènes identifiés pour la plasticité de la composition protéique des graines de *M. truncatula*.

**Annexe 3 :** Transformation des variables permettant l'homoscédasticité des résidus.

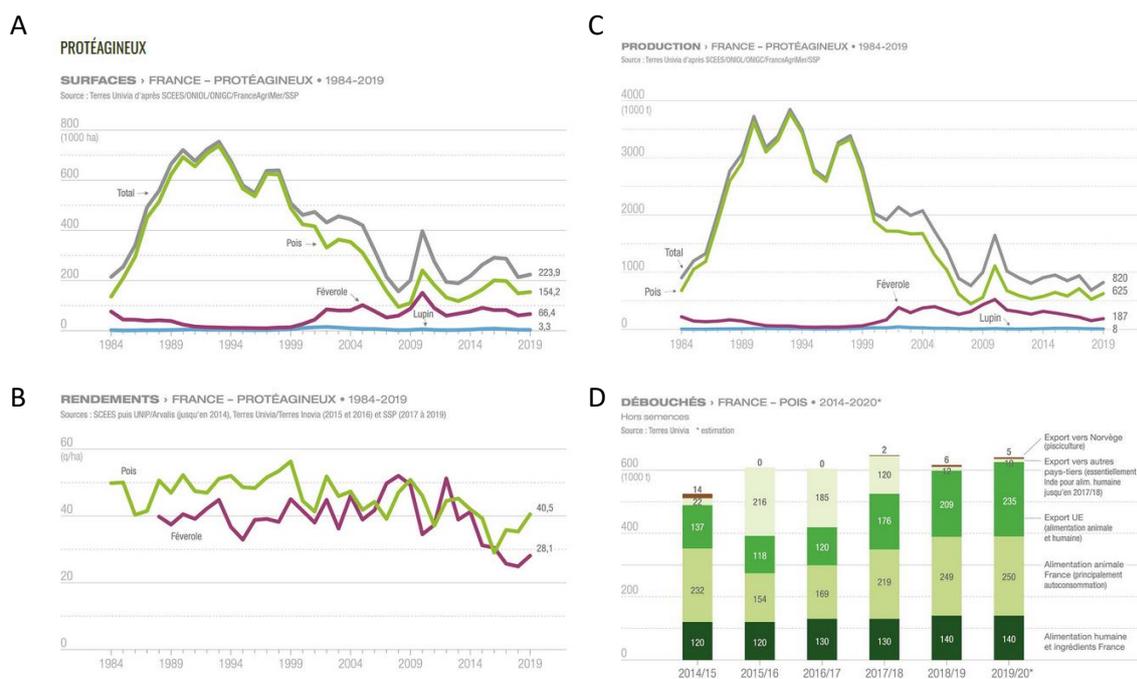
**Annexe 4 :** Manhattan plots associés à la composition protéique des graines de pois ainsi qu'à sa plasticité.

**Annexe 5 :** Manhattan plots associés aux variables de rendement ainsi qu'à leur plasticité.

# Introduction générale

D'un point de vue botanique, les légumineuses appartiennent à la famille des fabacées qui contient environ 765 genres et 19500 espèces. On trouve dans cette famille les légumineuses à graines utilisées directement pour l'alimentation telles que le soja, le haricot, la fève, le pois, la lentille, l'arachide et le pois chiche et aussi des légumineuses fourragères comme le trèfle, la luzerne et le sainfoin ou des plantes ornementales telles que l'acacia ou la glycine. Les légumineuses sont extrêmement bénéfiques en agriculture dans la mesure où elles sont capables de fixer l'azote atmosphérique grâce à des symbioses racinaires avec des bactéries rhizosphériques fixatrices d'azote. Autrefois envisagées uniquement comme des féculents de par leur forte teneur en glucides, les légumineuses cultivées pour leurs graines sont désormais considérées à juste titre comme des aliments source de protéines (des protéagineux). Pour faire écho à ce changement, la FAO (Food and Agriculture Organization of the United Nations) a déclaré 2016 comme étant l'année internationale des légumineuses ([www.legumese.com/2016-annee-internationale-des-legumineuses](http://www.legumese.com/2016-annee-internationale-des-legumineuses)). Les objectifs étaient de faire prendre conscience à la population de l'importance de la valeur nutritionnelle et agroécologique des légumineuses. L'augmentation de la consommation mondiale en produits carnés, en particulier celle exponentielle des pays comme la Chine et l'Inde induit une demande de plus en plus forte en protéines végétales pour l'alimentation des animaux d'élevage. Par conséquent, les légumineuses sont devenues des denrées stratégiques dans les échanges agricoles mondiaux. La production insuffisante pour couvrir les besoins en Europe et plus particulièrement en France nous place dans une situation de dépendance des importations en provenance du continent américain, principalement des USA, du Brésil et de l'Argentine. De nombreux projets et initiatives tels que le "Plan protéines végétales pour la France 2014-2020" et le projet H2020-EUCLEG 2017-2021 (coord. Bernadette Julier, INRA Lusignan) ont été mis en place avec pour objectif de diminuer les importations en protéines végétales et ainsi réduire la dépendance vis-à-vis des pays exportateurs. Le déficit Européen en production de protéagineux est d'origine historique. Les accords commerciaux négociés dans les années 60 dans le cadre du GATT (General Agreement on Tariffs and Trade) ont eu pour conséquence une répartition mondiale des différentes catégories de productions agricoles. Le continent américain s'est axé sur la production de protéagineux et l'Europe s'est focalisée sur celles des céréales et des animaux d'élevage. Un embargo américain sur les exportations de soja en 1973, dû à de mauvaises récoltes, a induit une prise de conscience du risque lié à la dépendance de l'Europe en protéagineux provenant du continent américain. Un plan pour promouvoir la culture des protéagineux a alors été mis en place et a permis d'atteindre une autosuffisance de 40% en protéines végétales dans les années 80. Les accords de Blair House signés en 1992 limitent

cependant la surface de soja cultivé en Europe. Les légumineuses sont peu cultivées en Europe. Au-delà des accords d'échanges internationaux, les revenus inférieurs à ceux pouvant être générés par d'autres cultures comme le colza ou le blé sont une entrave majeure. En ce qui concerne les légumineuses fourragères, la production a été divisée par 10 en 60 ans, passant de 3 millions à 278 000 ha. Cela s'explique en partie par l'augmentation de l'apport d'engrais azotés, limitant l'intérêt des légumineuses en tant qu'espèces fixatrices d'azote mais aussi par l'importation de tourteaux de soja pour l'alimentation animale. L'engouement pour la culture de légumineuses est par ailleurs freiné par la fluctuation des rendements. En décembre 2020, le gouvernement français a mis en place une stratégie nationale en faveur du développement des protéines végétales ([agriculture.gouv.fr](http://agriculture.gouv.fr)). L'objectif à court terme est une augmentation de 40% (+400 000 ha) des surfaces cultivées avec des légumineuses à l'horizon 2023. L'objectif à long terme est un doublement de ces surfaces avant 2030, pour atteindre 8% de la surface agricole utile, correspondant à 2 millions d'hectares. En France, le pois est le principal facteur de déclin de la production des protéagineux, avec une diminution des surfaces de 700 000 ha à 200 000 ha en 30 ans (Figure 1A).



**Figure 1 : Production des protéagineux en France.** (A) Evolution des surfaces cultivées en France entre 1984 et 2019 (en 1000 ha). (B) Évolution des rendements de pois et de féverole (en q/ha). (C) Poids de protéagineux produits (en 1000 t). (D) Principaux débouchés du pois en France (d'après Terres Univia).

Les rendements du pois sont très instables (Figure 1B), principalement à cause d'une sensibilité aux stress biotiques (champignons : *Aphanomyces*, ascochytose ; insectes : bruches, pucerons) et abiotiques (stress hydrique, thermique, nutritionnels). Ces différents stress peuvent, en plus de leurs effets sur le rendement, impacter la teneur ainsi que la composition protéique des graines, induisant des changements sur la valeur nutritionnelle et donc marchande. Avec le dérèglement climatique, les événements d'une intensité extrême (sécheresse, épisodes caniculaires) seront de plus en plus fréquents. La nécessité de pouvoir stabiliser la composition protéique des graines face à ces perturbations est donc primordiale. Les légumineuses sont riches en protéines. Parmi les acides aminés essentiels qu'elles possèdent, la lysine est présente en forte proportion, ce qui en fait un bon complément des céréales où cet acide aminé est peu représenté (Perrot, 1995). Cependant, la faible proportion en acides aminés soufrés (2%) et en tryptophane (4%) dans les protéines de réserve des graines de légumineuses est un frein à leur utilisation, ces acides aminés étant essentiels pour les humains et les animaux d'élevage. Si l'on souhaite promouvoir les protéines issues des graines de légumineuses, il est donc nécessaire d'une part d'améliorer l'équilibre nutritionnel en acides aminés en augmentant la proportion d'acides aminés soufrés et de tryptophane mais aussi de stabiliser cet équilibre face aux stress biotiques et abiotiques. Chez les légumineuses, les globulines représentent environ 70 à 80% des protéines de réserve des graines. Parmi les globulines, on trouve les globulines 7S et les globulines 11S, classées en fonction de leur coefficient de sédimentation (Osborne, 1909). Les globulines 11S sont relativement plus riches en acides aminés soufrés que les globulines 7S et sont donc recherchées d'un point de vue nutritionnel. En fonction des géotypes et des conditions environnementales, les proportions entre globulines 7S et globulines 11S peuvent varier, et par conséquent les valeurs nutritionnelles et techno-fonctionnelles des poudres issues de ces graines seront elles aussi différentes.

Dans ce contexte, l'objectif de ma thèse était d'identifier, par une approche de génétique d'association (GWAS), des gènes candidats pour optimiser ou stabiliser la composition protéique des graines de légumineuses. Le manuscrit de thèse est structuré en quatre chapitres.

Le **Chapitre I** est une synthèse bibliographique des connaissances générales utiles à la compréhension des travaux de thèse. Le **Chapitre II** correspond à l'étude du déterminisme génétique de la plasticité de la composition protéique des graines en réponse à des variations environnementales en utilisant la variabilité naturelle disponible chez la légumineuse modèle *Medicago truncatula*. Le **Chapitre III** porte sur la réponse d'une collection d'écotypes de pois

à une carence en soufre. Ce troisième chapitre est structuré en deux parties. La première est consacrée à l'étude du déterminisme génétique de la composition protéique des graines et de sa réponse à une carence en soufre. La deuxième traite les composantes du rendement et la teneur en protéine des graines jusqu'à l'identification des régions du génome et des gènes potentiellement impliqués dans leur variabilité en réponse à une carence en soufre. Le **Chapitre IV** est une conclusion générale des travaux et apporte des pistes de recherche faisant suite aux résultats de la thèse.

## CHAPITRE I :

### Synthèse bibliographique

# 1. Atout des légumineuses et ressources génétiques et génomiques

## 1.1 Avantages agroécologiques et nutritionnels

Les légumineuses ont la capacité d'assimiler l'azote atmosphérique grâce à la symbiose avec des bactéries fixatrices d'azote du genre *Rhizobium* formant des nodosités au niveau des racines (Barker et al., 1990). Ces nodosités se développent en particulier dans des conditions où la disponibilité en azote dans le sol est limitée. Cette spécificité en fait des espèces d'intérêt en rotation dans les cultures permettant de diminuer les apports d'engrais azotés, coûteux pour l'agriculteur et polluants pour l'environnement. Par ailleurs, les graines de légumineuses sont capables d'accumuler des quantités importantes de protéines dans leurs graines (20 à 40% selon les espèces) même en l'absence d'engrais azotés. Elles sont donc à la fois d'intérêt nutritionnel et agroécologique. Les légumineuses sont riches en lysine, un acide aminé essentiel, mais pauvres en acides aminés soufrés, dont la méthionine qui est un autre acide aminé essentiel. A l'inverse, les céréales sont pauvres en lysine mais plus riches en acides aminés soufrés. Les graines de légumineuses sont donc de bons compléments aux céréales dans l'alimentation humaine et des animaux d'élevage. Il y a actuellement un regain d'intérêt pour les protéines végétales en alimentation humaine, ce qui stimule les recherches visant à développer des produits attractifs à base de légumineuses visant à mieux équilibrer les apports en protéines d'origine animale et végétale (Chardigny and Walrand, 2016) . Au-delà de l'apport protéique, la consommation de légumineuses a de nombreux effets bénéfiques sur la santé, contribuant à minimiser les risques de diabète, cardiovasculaires, d'hypertension artérielle, et d'obésité (Rémond and Walrand, 2017).

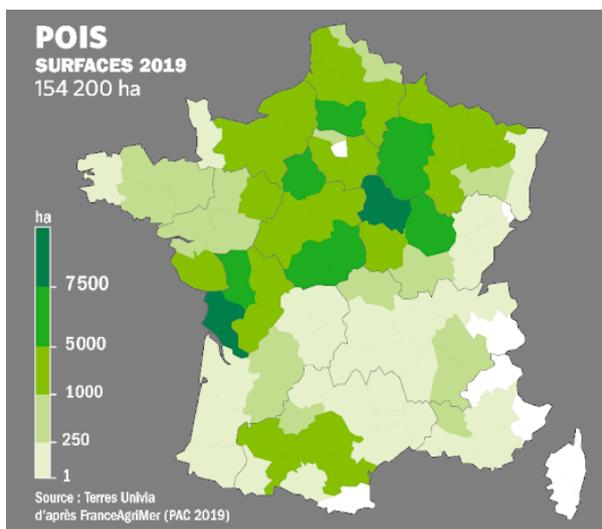
## 1.2 Le pois

### 1.2.a Production, culture du pois protéagineux et débouchés

Le pois (*Pisum sativum*) est la principale légumineuse à graines cultivée en Europe (ex: 154 000 ha en France en 2019, Figure 1). Il s'agit d'une espèce diploïde ( $2n = 14$ ) autogame dont la taille du génome est d'environ 4.3 Gb. Sa domestication a eu lieu au Proche-Orient à partir de la préhistoire. Il peut être consommé avant maturité (petit pois) ou à maturité après décorticage (pois cassé). La mode du "petit pois" a été initiée par Louis XIV puis s'est étendue dans de nombreux pays. Il s'est ensuite démocratisé au XXème siècle avec le "pois de conserve". Les protéines de graines de pois récoltées à maturité sont généralement utilisées

comme ingrédients alimentaires sous forme de farine, de concentré de protéines ou d'isolats protéiques. Le pois est principalement utilisé pour l'alimentation des animaux d'élevage mais on note une évolution à la hausse de son utilisation en alimentation humaine (Figure 1D). Il présente un intérêt croissant pour l'industrie alimentaire, notamment en raison de la faible allergénicité de ses protéines et de l'absence de gluten (Barac et al., 2010).

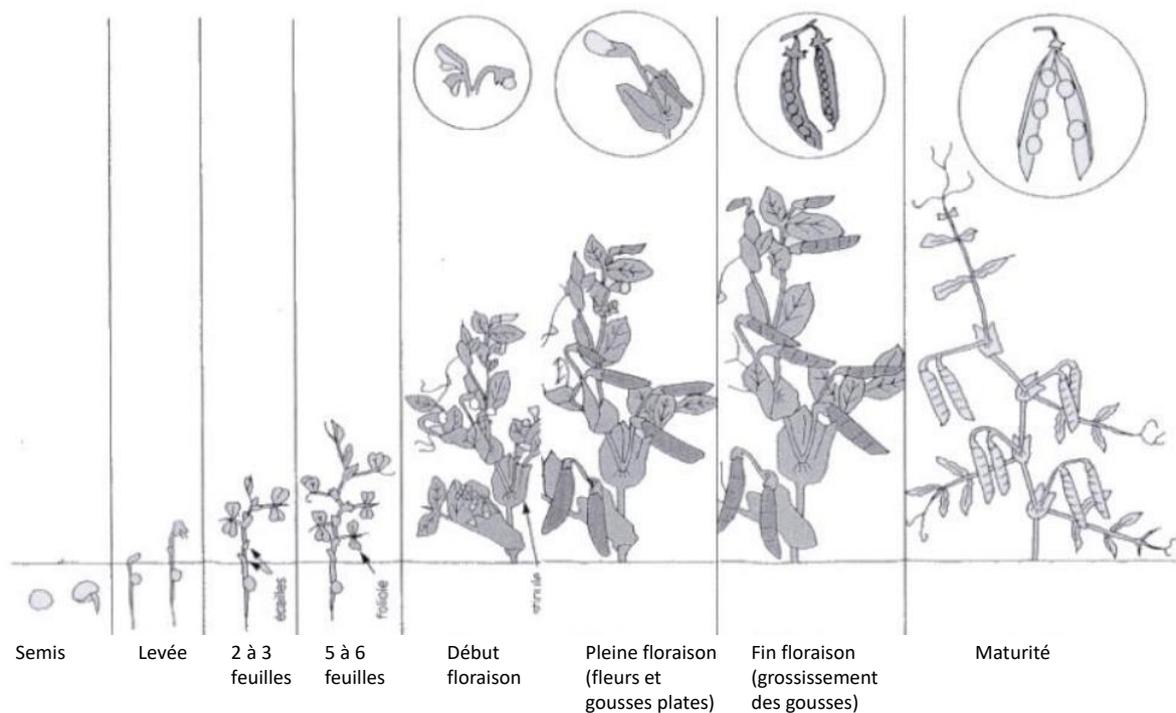
La culture du pois est une culture à cycle court d'origine méditerranéenne (terresunivia.fr). A l'origine, il est semé en début d'hiver (novembre) et récolté en fin de printemps (pois d'hiver), il a un cycle de culture long (environ 240 jours). Ce cycle long permet d'avoir un rendement élevé et d'éviter les stress de fin de cycle (hydrique et thermique). Il existe deux catégories de pois d'hiver, les variétés "hr" (high response) indifférentes à la photopériode et les variétés Hr, sensibles à la photopériode. Lorsque les variétés sont sensibles à la photopériode, cela permet d'éviter une floraison qui serait trop précoce ce qui réduit les risques de gel à floraison. Le pois a par la suite été adapté pour être cultivé dans le nord de l'Europe, avec un semis en fin d'hiver (entre février et mars) et être récolté pendant l'été (pois de printemps), son cycle de développement, plus court, est alors compris entre 90 et 140 jours. En France, les pois d'hiver et de printemps sont tous deux cultivés. Cependant, avec l'augmentation des températures au fil des ans, la culture du pois d'hiver tend à se développer. Le pois est principalement produit dans 6 régions : la Nouvelle-Aquitaine, la Bourgogne-Franche-Comté, le Centre-Val de Loire, le Grand-Est, les Hauts-de-France et l'Occitanie (Figure 2). La production de pois protéagineux en 2019 était de 625 kt, ce qui correspond à un rendement moyen de 40.5 q/ha de graines de pois dont 250 000 tonnes sont exportées (Figure 1B-C).



**Figure 2 : Répartition géographique des surface cultivées de pois en France en 2019 (d'après Terres Univia)**

### 1.2.b Cycle de développement du pois

Le pois est une plante annuelle sans dormance ce qui signifie qu'elle peut être semée sans que les graines aient besoin de vernalisation. Son développement se déroule en différentes phases décrites dans la Figure 3. Durant la période végétative (jusqu'à floraison), la plante accumule des composés dans ses tissus qui seront remobilisés vers les organes puits, dont les graines, lors de la période reproductive associée au développement et au remplissage des graines (Schiltz et al., 2005).



**Figure 3 : Schéma des principaux stades du cycle de développement du pois (d'après Boyeldieu, 1999).**

La phase végétative est constituée de la germination, de la levée ainsi que de la mise en place des racines, de la tige et des feuilles des nœuds végétatifs. A titre d'exemple, chez le génotype Caméor utilisé comme référence pour le séquençage du génome du pois (Kreplak et al., 2019), cette phase végétative dure six semaines. La phase reproductive quant à elle débute dès la mise en place des nœuds reproducteurs et l'apparition des fleurs, ces dernières apparaissent au fur et à mesure. On trouve donc des gousses à des stades de maturité différents sur la même plante, les plus matures au niveau basal et les moins matures au niveau apical. Lorsque les gousses sont mûres, elles deviennent déhiscents et libèrent leurs graines plus ou moins spontanément selon les génotypes.

### 1.2.c Morphologie du pois

Le pois est une plante grimpante herbacée annuelle. Les variétés cultivées en France sont pour la très grande majorité de type indéterminé. Leur taille peut varier de quelques dizaines de centimètres à plusieurs mètres pour certains génotypes fourragers. La partie aérienne est composée d'une tige principale avec plusieurs ramifications. Après la germination de la graine, la tige principale produit deux feuilles primordiales sous forme d'écailles puis est ensuite composée de nœuds produisant des feuilles ou des ramifications. La tige principale ainsi que les ramifications sont composées d'un empilement de phytomères et se terminent par un méristème caulinaire. Chaque phytomère est composé de trois parties : un entre-nœud, une feuille et un méristème axillaire. Les feuilles de pois sont constituées de deux stipules situées à la base du pétiole puis de folioles et de vrilles. Chez certaines variétés, toutes les folioles sont transformées en vrilles, ces variétés sont nommées "afila". Ce caractère a été sélectionné afin de rendre plus facile la culture du pois et permet aux plantes de s'accrocher les unes aux autres, limitant le phénomène de verse, qui, en créant des conditions denses en fourrage, favorise le développement de pathogènes. Il existe aussi des variétés "acacia" où toutes les vrilles sont transformées en folioles. Les fruits du pois sont des gousses allongées contenant les graines, ces dernières ayant des tailles, des formes (graines lisses ou ridées) et des couleurs différentes en fonction des génotypes (Santos et al., 2019) (Figure 4).



**Figure 4 : Caractéristiques morphologiques des graines de pois.** Les différentes catégories de graines de pois en fonction de leur couleur, leur rugosité et leur forme (d'après Santos et al., 2019)

### 1.2.d Ressources génétiques disponibles chez le pois

Les quatre plus grandes collections génétiques (germplasm collections) de pois en 2015 étaient par ordre décroissant : le CRB protéagineux de l'INRA de Dijon avec 8839 accessions, l'Australian Temperate Field Crop Collection avec 7432 accessions, l'US Department of Agriculture avec 6827 accessions et le Vavilov Research Institute of Plant Industry avec 6790

accessions (Warkentin et al., 2015). Ces collections sont un pilier important et stratégique pour la mise en évidence des déterminismes génétiques de caractères agronomiques (<http://florilege.arcad-project.org/fr/crb/proteagineux/crb-proteagineux>). Les légumineuses sont difficiles à transformer génétiquement (Warkentin et al., 2015), avec de faibles taux de transformation et une régénération difficile des plantes. Cependant, la mise en place de ressources TILLING (Targeting Induced Local Lesions in Genomes) de génétique inverse, d'abord chez le cultivar de pois de printemps Caméor (Dalmais et al., 2008) puis récemment à partir d'une lignée élite de pois d'hiver (Ps336/11, IA-PeaMUST) offre un outil de validation fonctionnelle chez cette espèce.

Des ressources génétiques comme de nombreuses populations de lignées recombinantes sont également disponibles au CRB protéagineux de Dijon. Une carte génétique haute densité consensus de 12 populations de lignées recombinantes incluant plus de 15000 marqueurs est publiée et a permis d'établir des ponts pour le transfert d'informations entre espèces légumineuses dont *M. truncatula* (Tayeh et al., 2015). Un atlas de gènes chez le génotype de pois Caméor regroupant 46 099 contigs et 20<sup>1</sup> organes à différents stades de développement et conditions de nutrition azotée contrastées est disponible (Alves-Carvalho et al., 2015). La première version du séquençage du génome obtenu chez le cultivar Caméor a été récemment publiée (Kreplak et al., 2019). Elle est disponible via l'outil jbrowse ([https://urgi.versailles.inra.fr/jbrowse/gmod\\_jbrowse/?data=myData](https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData)) qui regroupe également les données transcriptomiques publiques (geneatlas et nouvelles ressources RNAseq). Plus récemment, dans le cadre des projets PeaMUST (Coord. Judith Burstin, équipe ECP) et GRASP (Coord. Virginie Bourion, équipe ECP), une ressource unique de données de polymorphisme sur plus de 300 lignées a été développée par la technique de capture d'exome (Coll. BIOGEMMA<sup>1</sup>). De cette initiative a également été développée une puce de technologie Axiom de densité 90000 SNP, rendant possible les approches GWAS sur de nombreuses populations chez le pois.

## 1.3 *Medicago truncatula*, espèce modèle des légumineuses

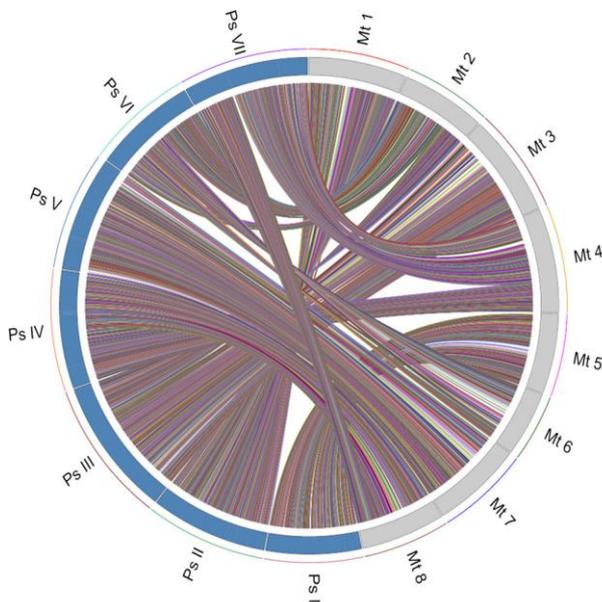
### 1.3.a Intérêts de l'espèce modèle *M. truncatula*

Afin de pouvoir élucider plus aisément les mécanismes moléculaires régissant certains caractères agronomiques, il est possible d'utiliser des organismes proches génétiquement des

---

<sup>1</sup> [https://www.peamust-project.fr/content/download/4616/45159/version/1/file/Booklet\\_G%C3%A9notypage\\_PeaMUST\\_VF.pdf](https://www.peamust-project.fr/content/download/4616/45159/version/1/file/Booklet_G%C3%A9notypage_PeaMUST_VF.pdf)

espèces d'intérêt mais avec une plus petite taille de génome. La plante modèle par excellence est *Arabidopsis thaliana*. Cependant, son inaptitude à former des nodosités en font un modèle limité pour étudier la biologie des légumineuses. Chez les fabacées, deux modèles ont émergé dans les années 1990, *M. truncatula* et *Lotus japonicus* (Cannon et al., 2006), pour l'étude des symbioses et leur génome a été séquencé dans les années 2000 (Pecrix et al., 2018; Tabata and Stougaard, 2014). L'espèce *M. truncatula* (aussi appelée luzerne tronquée), diploïde ( $2n = 16$ ), autogame, avec un génome de seulement 500 Mbp (Ané et al., 2008), est privilégiée pour les études visant à transférer les connaissances vers le pois car phylogénétiquement très proche de cette espèce (Kreplak et al., 2019; Tayeh et al., 2015) (Figure 5). De plus, la composition protéique des graines matures de *M. truncatula* ainsi que la cinétique d'accumulation des protéines de réserve lors du remplissage des graines sont très semblables à celle du pois (Djemel et al., 2005; Gallardo et al., 2008; Burstin J. Pers Comm.).



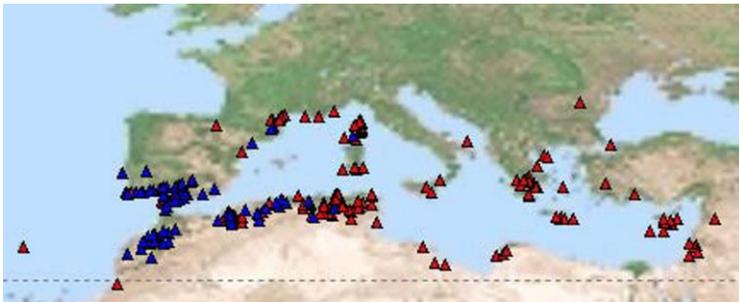
**Figure 5 : Diagramme de la synténie entre le pois (Ps) et *Medicago truncatula* (Mt).** Construit en utilisant le package Rcirco. En bleu sont les 7 groupes de liaison du pois et en gris sont les 8 chromosomes de *Medicago truncatula*. Figure extraite de Tayeh et al., 2015.

Les deux principaux génotypes de *M. truncatula* utilisés en recherche sont A17, issu du cultivar Jemalong (Cannon et al., 2006), lignée de référence pour le séquençage et les cartes génétiques, et R-108-1 issu de l'écotype 108-1 lignée de référence pour la transgénèse, présentant une bonne aptitude à la régénération (contrairement à A17) (Journet et al., 2001),

### 1.3.b Cycle de développement et morphologie de *M. truncatula*

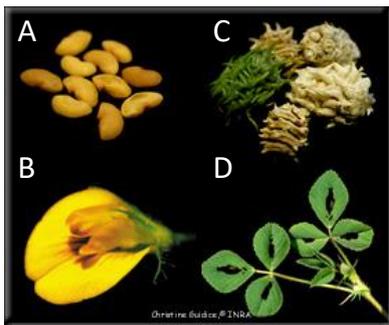
*M. truncatula* est une espèce annuelle fourragère originaire du pourtour du bassin méditerranéen (Ronfort et al., 2006) (Figure 6). Dans ce climat, sa croissance végétative a lieu entre la fin de l'automne et le début du printemps lorsque les températures sont plus froides et

que la photopériode est plus courte que le reste de l'année. La floraison a lieu au printemps et les graines sont matures entre la fin du printemps et le début de l'été. Le cycle de vie de *M. truncatula* est donc adapté pour éviter les stress hydriques pouvant survenir en été (Garmier et al., 2017). Le temps de génération de graine à graine est de 10 à 12 semaines. Sa qualité fourragère (Derkaoui and Romman, 1993) et son rendement (Zhu et al., 1998) étant similaires à ceux de la luzerne (*Medicago sativa*), elle est cultivée dans certains pays comme l'Australie et l'Afrique du sud (Brand et al., 1991).



**Figure 6 : Répartition géographique des individus de *M. truncatula* présents dans notre étude.** Les individus sont répartis selon les deux groupes génétiques (rouge et bleu) déduits d'après le polymorphisme des microsatellites (Figure extraite de Ronfort et al., 2006).

Il existe une forte variabilité génétique et phénotypique chez *M. truncatula* au niveau intra et interpopulation (Bataillon et al., 2006; Ronfort et al., 2006). Les plantes peuvent atteindre 40 cm de haut et produire des graines de petite taille en abondance. Les fleurs sont jaunes, les gousses sont petites, enroulées, épineuses et en forme de tonneau (Figure 7).



**Figure 7 : Quelques caractéristiques morphologiques de *M. truncatula*.** (A) Graines, (B) Fleur, (C) gousses, (D) Feuilles. Extrait de <http://www.medicagohapmap2.org/about>, photos de Christine Guidice

### 1.3.c Ressources génétiques disponibles chez *M. truncatula*

*M. truncatula* ayant été choisie comme espèce modèle des légumineuses (Branca et al., 2011; Gentzbittel et al., 2015), de nombreuses ressources (séquençage du génome, atlas d'expression de gènes) et de multiples outils (core collection, collections de mutants) ont été développés pour étudier les mécanismes moléculaires sous-jacents aux processus biologiques (Garmier et al., 2017). Le Centre de Ressources Biologiques (CRB) INRAE situé à Montpellier possède la deuxième plus grande collection mondiale représentant la diversité de *M. truncatula* avec 1 478 génotypes en 2018, derrière le South Australian Research and Development Institute (SARDI)

(Garmier et al., 2017) ([www1.montpellier.inra.fr/BRC-MTR/accueil.php](http://www1.montpellier.inra.fr/BRC-MTR/accueil.php)). Elle met à disposition des écotypes, des lignées pures, des populations de lignées recombinantes, des hybrides F1 ainsi que leur descendance.

Un consortium international a produit la première version de la séquence du génome de l'écotype A17 de *M. truncatula* (Mt3.5) comprenant 418Mb d'euchromatine en 2011 puis, une version 4 (Mt4) génome-entier (technique short-reads Illumina) a été mise à disposition en 2014 (Tang et al., 2014). De façon concomitante, un consortium élargi s'est engagé dans le reséquençage de 330 écotypes de *M. truncatula* (dont 196 issus du CRB de Montpellier) en utilisant la même technique « short reads Illumina ». Trente lignées ont été séquencées avec une profondeur de 20X, les autres lignées à une profondeur de 5X. La recherche de variants a été conduite sur l'échantillon 30X, permettant d'identifier 4 à 6 SNP/kb et de réaliser une première caractérisation du déséquilibre de liaison (LD) chez cette espèce. Les auteurs ont mis en évidence une décroissance du LD de moitié sur 3kb pour atteindre une valeur <0.3 sur ~5kb. Le reséquençage des 300 autres lignées à 5X a permis de générer plus de 6 millions de SNP, ouvrant la voie aux analyses d'association génome-entier (GWAS) sur des traits aussi variés que la floraison, la symbiose et la tolérance à différents stress biotique et abiotique (Bonhomme et al., 2014; Gentzittel et al., 2015). Récemment, le Laboratoire des Interactions Plantes-Microorganismes (LIPM Toulouse) a produit une nouvelle version de très haute qualité de l'assemblage du génome de la lignée A17 de Jemalong (Mt5.0, (Pecrix et al., 2018)) en utilisant les nouvelles techniques en long-read PacBio (100x) associées à des cartes optiques. L'intégralité des SNP produits dans le cadre du projet Hapmap (<http://www.medicagohapmap2.org/about>) a été repositionné sur cette nouvelle version, nous permettant aujourd'hui de réaliser directement des études d'association sur la version 5 du génome.

Des outils de validation fonctionnelle sont également disponibles chez cette espèce, comme une ressource TILLING chez l'écotype Jemalong-A17 (Le Signor et al., 2009) associé à une plateforme de criblage haut débit (<https://www.jic.ac.uk/research-impact/technology-research-platforms/reverse-genetics/>) et une collection de mutants d'insertion chez l'écotype R108 totalisant 12 000 lignées d'insertion et 300 000 insertions séquencées (Cheng et al., 2011).

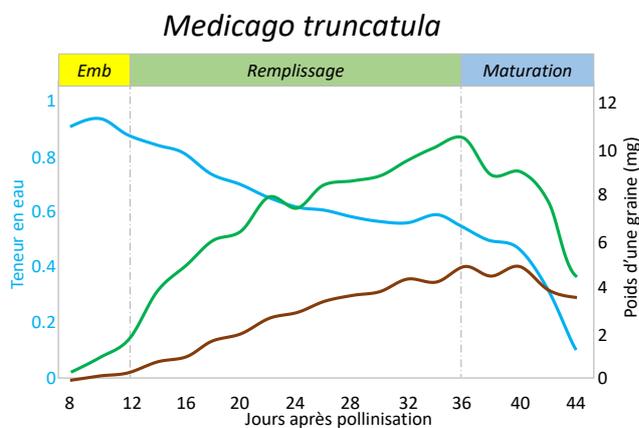
Un atlas d'expression des gènes a été développé par la « Samuel Roberts Noble Foundation » (USA) en utilisant la technique Affymetrix (puce de 50k, benedito,2008). De nombreux tissus et conditions de culture sont disponibles (<https://mtgea.noble.org/v3/>) à la fois sur A17 et R108.

Récemment, la technique RNAseq a permis de donner une vision plus exhaustive du transcriptome. Les données brutes de ces études sont libres d'accès sur <https://www.ncbi.nlm.nih.gov/sra> (>2600 échantillons déposés à ce jour).

## 2. Développement de la graine et accumulation des réserves

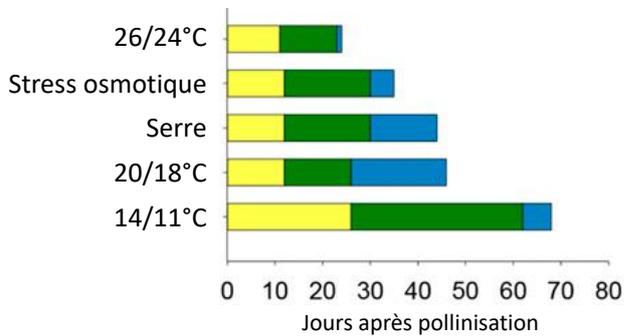
### 2.1 Développement de la graine

Le développement de la graine se déroule en trois grandes étapes : l'embryogenèse, le remplissage et la maturation. Suite à la floraison, la double fécondation du sac embryonnaire a lieu. La fécondation de la cellule centrale va former le zygote accessoire, qui constituera l'albumen. La fécondation de l'oosphère forme un zygote principal qui va ensuite se diviser de manière intensive et acquérir une polarité pour constituer l'embryon. Cette étape correspond à l'embryogenèse. L'embryon et l'albumen sont entourés par les téguments qui sont les tissus maternels. Le stade de transition entre l'embryogenèse et le remplissage correspond à l'arrêt des divisions cellulaires et à l'initiation de l'expansion cellulaire. Dans des conditions standards en chambre climatique, ce shift vers le remplissage a lieu environ 12 jours après pollinisation (JAP) (Gallardo et al., 2003). Pendant la phase de remplissage (de 12 à 36 JAP), l'embryon accumule des composés de réserve (lipides, protéines, glucides) et la teneur en eau décroît progressivement. A la fin du développement, une étape de dessiccation intense, associée à la diminution des activités enzymatiques a lieu (phase de maturation). La graine rentre alors dans un état de quiescence métabolique. Son métabolisme sera réactivé lors de la germination.



**Figure 8 : Phases de développement de la graine de *Medicago truncatula*.** Les trois grandes phases du développement de la graine sont l'embryogenèse (Emb), le remplissage et la maturation. La courbe bleue correspond à la teneur en eau d'une graine (1 = 100%), la courbe verte correspond au poids frais d'une graine et la courbe marron correspond au poids sec d'une graine. D'après Gallardo et al., 2003.

Il est à noter que la durée de chacune de ces phases varie en fonction des conditions environnementales. Par exemple, chez *M. truncatula*, il a été montré que des températures froides prolongent les phases d'embryogenèse et de remplissage tandis que des températures élevées ou un stress osmotique raccourcissent la durée du développement des graines (Righetti et al., 2015) (Figure 9).

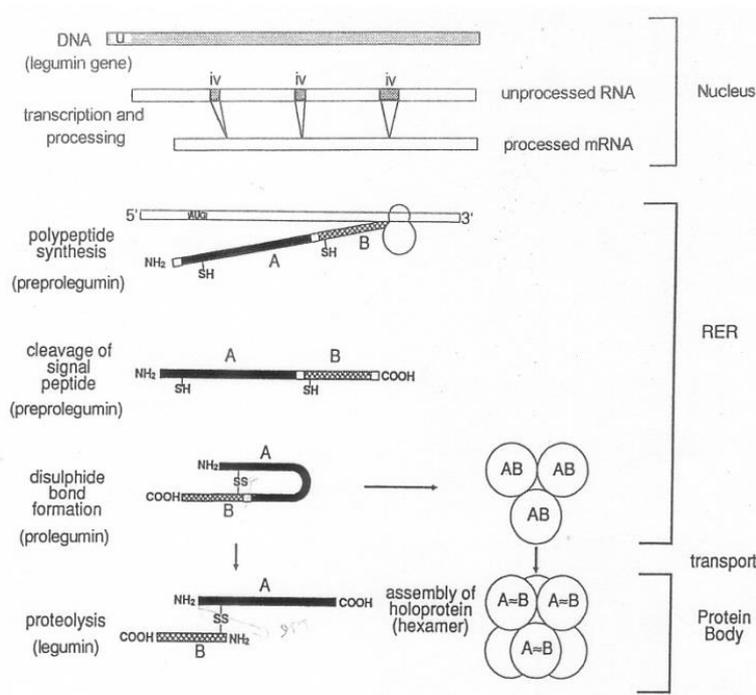


**Figure 9 : Impact des conditions environnementales sur la durée des phases de développement de la graine de *M. truncatula*.** Jaune : embryogenèse ; vert : remplissage ; bleu : maturation ; **26/24°C** : 26°C de jour, 24°C de nuit, fortes températures ; **Stress osmotique** : 20° de jour, 18°C de nuit, potentiel hydrique du sol de -0.1 MPa ; **Serre** : conditions de température et de luminosité variables ; **20/18°C** : 20°C de jour, 18°C de nuit, considérée comme la condition standard de culture ; **14/11°C** : 14°C de jour, 11°C de nuit. Données extraites de Righetti *et al.*, 2015.

## 2.2 Composition des graines de légumineuses

A maturité, la composition des graines varie selon les espèces. Par exemple, les graines des protéagineux comme le pois ou la féverole se distinguent des graines oléoprotéagineuses par leur faible teneur en lipides. Les graines de pois sont caractérisées par une forte teneur en amidon (50%) et en protéines (23 à 28%), et ne contiennent que très peu de lipides (<2%) (Burstin et al., 2011). Les graines de *M. truncatula* ne présentent que des traces d'amidon, sont riches en protéines (32 à 42%) et contiennent 8-10% de lipides et 6-10% de glucides (Djemel et al., 2005). Chez les légumineuses, les protéines de réserve s'accumulent au sein de l'embryon dans des vacuoles de stockage appelées corps protéiques. La composition protéique est propre à chaque espèce et à chaque génotype. Chez les légumineuses, on trouve deux principales classes de protéines de réserve, les albumines (2S), minoritaires (20-30% des protéines totales), et les globulines (7S et 11S) qui représentent 70-80% de l'azote total des graines (Perrot, 1995). Les albumines de pois (PA1 et PA2) sont solubles dans l'eau et s'accumulent dans les graines sous forme de dimères avec respectivement des monomères de 6 kDa (PA1) et 25 kDa (PA2). Comparativement aux globulines, elles ont une teneur plus élevée en lysine et en acides aminés soufrés que sont la cystéine et la méthionine. Les globulines sont insolubles dans l'eau, elles sont séparées en deux fractions, 7S et 11S, en fonction de leur coefficient de sédimentation après centrifugation (Osborne, 1909). Dans les graines matures, les globulines 7S sont constituées des vicilines et des convicilines. Les vicilines sont des trimères de 150 à 200 kDa

composés de sous-unités dérivées de clivages protéolytiques à des sites spécifiques. Les convicilines sont des tétramères de 210 à 290 kDa ne subissant pas de clivages protéolytiques. Les globulines 11S, nommées légumines, sont des hexamères de 350-400 kDa. Pendant leur maturation, les formes précurseurs des légumines subissent un clivage protéolytique (coupure entre Asn-Gly, N↓GLEET) libérant deux sous-unités (acide, A ou alpha ; basique, B ou bêta) qui restent reliées par un pont disulfure (Figure 10, extrait de (Bewley and Black, 1994)). Les globulines 11S sont plus riches en acides aminés soufrés que les globulines 7S. La composition en globulines détermine donc la qualité nutritionnelle des graines mais aussi les propriétés techno-fonctionnelles des protéines. A titre d'exemple, les ponts disulfures sont importants pour concevoir des gels et des émulsions à base de protéines de pois (Sun and Arntfield, 2012). Il est possible de quantifier les différentes fractions protéiques par chromatographie échangeuse d'ions ainsi que par électrophorèse mono- ou bi-dimensionnelle. L'électrophorèse bi-dimensionnelle a permis d'obtenir des cartes de référence du protéome des graines chez *M. truncatula* (Gallardo et al., 2003; Le Signor et al., 2017), avec 96 spots correspondant à des globulines. Une carte de référence du protéome des graines de pois existe aussi (Bourgeois et al., 2009) sur laquelle, 104 polypeptides sur les 156 annotés sont des globulines. L'électrophorèse mono-dimensionnelle a aussi été utilisée pour analyser la composition en globulines des graines de *M. truncatula* (Le Signor et al., 2017) et de pois (Henriet et al., 2019). L'avantage de cette approche est qu'elle permet d'analyser la composition protéique d'une grande série d'échantillons, et d'estimer à partir de ces données le ratio entre les globulines 7S et 11S.



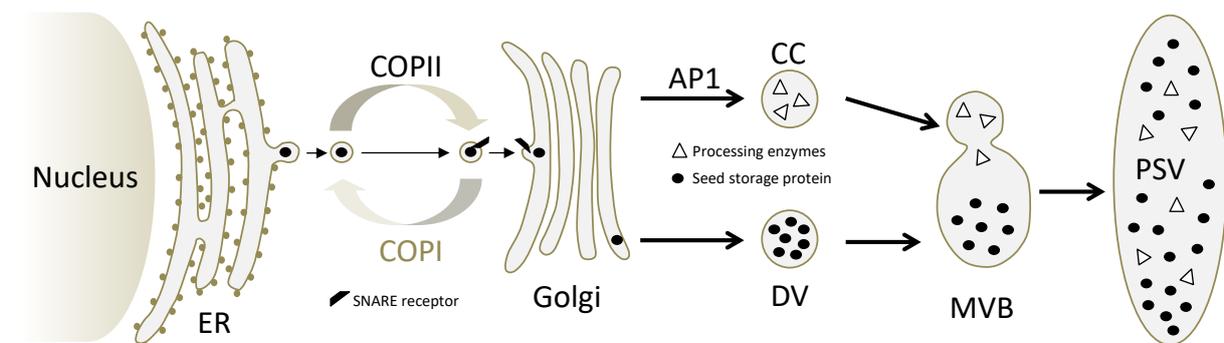
**Figure 10 : Schéma des différentes étapes de maturation des légumine dans les cotylédons.**  
A: chaîne acide; B: chaîne basique.  
(Figure extraite de Bewley *et al.*, 1994)

## 2.3 Synthèse des protéines de réserve

La synthèse des réserves de la graine fait intervenir un jeu complexe de facteurs de transcription dont certains ont été identifiés en utilisant la plante modèle *Arabidopsis thaliana*. Les facteurs de transcription NF-YB Leafy Cotyledon1 (LEC1) et à domaine B3 (LEC2 ; abscisic acid insensitive3, ABI3 ; et FUSCA3, FUS3) ont été montrés comme régulant la synthèse des réserves lipidiques et protéiques (Reidt *et al.*, 2000). Des facteurs de transcription plus spécifiquement impliqués dans l'accumulation de certains composés ont aussi été identifiés, comme les BZIP#10, #25, #53, #67 et MYC#2, #3, #4 qui contrôlent l'accumulation des protéines de réserve (Boulard *et al.*, 2017; Gao, 2016). Chez le maïs, les BZIP tels que OPAQUE2 ou OHP1 (Pysh *et al.*, 1993) contrôlent la synthèse des protéines de réserve de type zéïnes, et chez les légumineuses comme le pois, ABI5 contrôle la synthèse des globulines 7S (Le Signor *et al.*, 2017; Pandurangan *et al.*, 2016; W-K Ng and Hall, 2008). Il faut noter à titre d'exemple que l'homologue fonctionnel d'AtABI3 chez le maïs (VP1), se lie au même complexe protéique impliquant le domaine AP2-ABI4 et le bZIP-ABI5, suggérant une bonne conservation des réseaux de régulation parmi les différentes espèces (Suzuki *et al.*, 2014, 2007, 2002).

Une fois synthétisées, les protéines de réserve sont transportées jusqu'aux vésicules de stockage selon différentes étapes mettant en jeu de nombreux événements moléculaires (Figure 11, et

références citées, les différentes étapes de cette figure sont extraites de (Robinson et al., 2005; Vitale and Hinz, 2005; Xiang et al., 2013)). Elles pénètrent dans un premier temps le réticulum endoplasmique (RE), où les globulines 7S et 11S trimérisent (formation des ponts disulfure pour les 11S). Elles sont ensuite transportées vers l'appareil de Golgi par l'intermédiaire de vésicules COP (« coat protein complexes », COPI et COPII). La fusion entre ces vésicules et l'appareil de Golgi est médiée par des protéines SNARE (« Soluble N-ethylmaleimide-sensitive factor Attachment protein REceptor »). Localisées à la surface des membranes des vésicules COP et de l'appareil de Golgi, ces protéines SNARE permettent la reconnaissance entre ces deux compartiments. Dans le RE et l'appareil de Golgi, les globulines subissent d'importantes modifications post-traductionnelles (ex. phosphorylations, glycosylations). Suite à leur translocation dans l'appareil de Golgi, les protéines de réserve sont transportées vers les vacuoles de stockage (PSV) par l'intermédiaire de vésicules denses. Dans les vacuoles de stockage, elles terminent leur maturation grâce à des clivages protéolytiques (ex. cystéine endopeptidase), permettant notamment de cliver les globulines 11S (Asn-Gly, N↓GLEET) qui s'assemblent alors en hexamères (Hara-Nishimura et al., 1995).

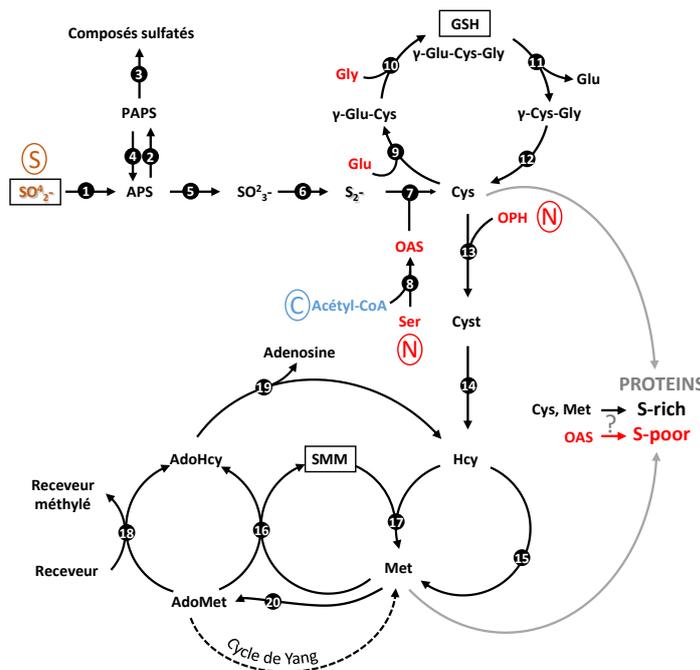


**Figure 11 : Modèle de transport et de mise en réserve des globulines.** Les différentes étapes menant à l'accumulation de protéines de réserve sont extraites de Vitale & Hinz (2005), Robinson et al. (2005) et Xiang et al. (2013). ER, réticulum endoplasmique ; COPI et COPII, coat protein complexes I et II ; SNARE, Soluble N-ethylmaleimide-sensitive factor Attachment protein REceptor ; AP1, adaptor protein 1 ; CC, clathrin-coated vesicles ; DV, dense vesicles ; MVB, multivesicular bodies ; PSV, protein storage vacuole.

## 2.4 Synthèse des acides aminés soufrés

Il a été montré chez le soja que les graines immatures sont le site principal de réduction du sulfate et que la majorité des acides aminés est synthétisée dans la graine plutôt qu'importés d'autres tissus *via* le phloème (Sexton and Shibles, 1999). Lorsque le sulfate arrive dans la graine, il est stocké dans les vacuoles ou métabolisé (Aarabi et al., 2020). Lors de sa métabolisation, il est tout d'abord réduit en adénosine 5'-phosphosulfate (APS) en présence

d'ATP lors d'une réaction catalysée par l'ATP sulfurylase (Figure 12, modifié d'après Zuber et al., 2009).



**Figure 12 : Le métabolisme du soufre (S) chez les plantes.** Modifié d'après Zuber et al. (2009). Les liens avec les métabolismes du carbone (C) et de l'azote (N) sont respectivement en bleu et rouge. Les molécules soufrées transportées entre les différents organes des plantes sont encadrées.

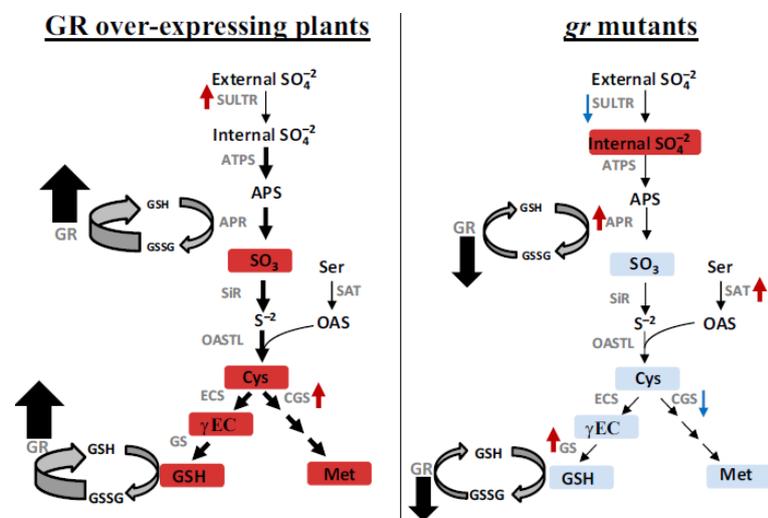
**Métabolites :** AdoHcy, S-adénosylhomocystéine ; ATP, adénosyl triphosphate ; APS, adényl sulfate ; Cys, cystéine ; Cyst, Cystathionine ; Glu, glutamate ; Hcy, homocystéine ; Ser, Sérine ; OAS, O-acétylsérine ; OPH, O-phosphohomocystéine ; PAPS, 3'-Phosphoadénosine 5'-phosphosulfate. **Enzymes :** 1, ATP-sulfurylase ; 2, APS-kinase ; 3, sulfotransférase ; 4, 5-bisphosphate nucléotidase ; 5, APS-réductase ; 6, sulfite réductase ; 7, Cys synthase ; 8, sérine O-acétyltransférase ; 9, γ-Glu-Cys synthétase ; 10, GSH synthase ; 11, γ-Glu-transpeptidase ; 12, Membrane alanil aminopeptidase ; 13, Cystathionine γ-synthase (CGS) ; 14, Cyst β-lyase ; 15, Méthionine synthase ; 16, Met S-méthyltransférase ; 17, Homocystéine S-méthyltransférase ; 18, Méthyltransférases ; 19, AdoHcy hydrolase ; 20, AdoMet synthétase.

**Dans la voie secondaire du métabolisme du soufre,** l'APS peut servir à former du 3'-Phosphoadénosine-5'-phosphosulfate (PAPS) grâce à l'APS kinase (APK). Le PAPS a un rôle dans la signalisation cellulaire, dans la réponse au stress oxydatif ainsi que dans les mécanismes de défense en servant de source de sulfate pour la sulfatation de différentes molécules du métabolisme secondaire telles que les glucosinolates, les brassinostéroïdes, les flavonoïdes ou les phytosulfokines (Anjum et al., 2015; Hirschmann et al., 2014).

**Dans la voie primaire du métabolisme du soufre,** l'APS est réduite en sulfite grâce à l'APS réductase (APR). Cette étape permet d'orienter le métabolisme du soufre vers la synthèse de cystéine et de méthionine. Chez Arabidopsis, APR2 est un déterminant majeur de l'utilisation du sulfate dans les parties aériennes, et son inactivation conduit à une suraccumulation de sulfate dans ces tissus (Loudet et al., 2007). Il est à noter que le PAPS peut aussi être réduit en sulfite lors d'une réaction catalysée par la PAPS réductase. Les sulfites sont réduits en sulfure par une ferredoxin-dépendant sulfite réductase (SIR), puis la synthèse de cystéine est catalysée par le complexe cystéine synthase composé de deux enzymes : la sérine acétyltransférase (SAT), et l'O-acétylsérine (thiol)-lyase (OASTL). La SAT permet la synthèse d'O-acétylsérine (OAS) à partir d'acétyl-CoA (dérivé du métabolisme carboné) et de sérine (dérivée du métabolisme azoté) et l'OASTL permet de produire la cystéine à partir d'OAS et de sulfure. La formation du complexe cystéine synthase joue un rôle important dans la régulation de l'activité

de ces deux enzymes. En effet, l'OASTL n'est active que sous forme libre, tandis que l'activité de la SAT dépend de son association avec l'OASTL (Droux et al., 1998; Wirtz and Hell, 2007). Il a aussi été montré que les ions sulfides stimulent la formation du complexe cystéine synthase, tandis que l'OAS favorise sa dissociation (Lewandowska and Sirko, 2008). Ces régulations permettent d'orienter les flux d'azote (serine) et de carbone (acetyl-CoA) vers la synthèse de cystéine lorsque les ions sulfides sont disponibles.

**La cystéine permet la synthèse du glutathion (GSH)** qui joue un rôle central dans les mécanismes antioxydants en contrôlant la quantité des formes actives de l'oxygène (Zagorchev et al., 2013). Lorsque la plante est soumise à un stress oxydatif, la demande en soufre sous forme réduite augmente (Jacob and Anwar, 2008). Il a été montré que des lignées transgéniques avec une plus grande quantité de cystéine et de glutathion ont une résistance accrue au stress oxydatif ainsi qu'au cadmium (Blaszczyk et al., 1999; Domínguez-Solís et al., 2004; Noji et al., 2001). De plus, le glutathion sert à la réduction de l'APS en sulfite via l'APS réductase (APR, Figure 13). Il est donc un acteur clé de la voie d'assimilation du soufre. Des plantes transgéniques surexprimant le gène codant la glutathion réductase, qui permet de redonner du GSH à partir de sa forme oxydée (GSSG), présentent des niveaux de cystéine et de méthionine plus élevés alors que ces composés sont trouvés en quantités plus faibles chez des mutants de glutathion réductase au profit du sulfate (Figure 13). De plus, en condition de carence en soufre modérée, les plantes transgéniques ont une meilleure croissance que les plantes témoins alors que les mutants sont plus sensibles au manque de soufre (Cohen et al., 2019).



**Figure 13 : Effet de la modification de l'activité de la glutathion réductase (GR) sur l'assimilation du soufre.** Les plantes d'Arabidopsis surexpriment cette enzyme (gauche) ou sont affectés dans la biosynthèse de cette enzyme (mutants *gr*, droite). Les métabolites sont surlignés en rouge s'ils sont suraccumulés et en bleu s'ils sont sous-accumulés. Les flèches à côté des enzymes sont en rouge en cas de surexpression et en bleu en cas de sous-expression (Figure extraite de Cohen et al., 2019).

Dans le contexte du développement de la graine, il a été montré qu'un certain nombre de protéines antioxydantes pouvaient avoir un rôle dans la diminution des dommages cellulaires,

notamment en conditions de carence en soufre (Henriet et al., sous presse). Parmi ces protéines, la glutathione *S*-transférase est impliquée dans la conjugaison du glutathion avec une grande variété de composés électrophiles, notamment toxiques, permettant de détoxifier les cellules ou d'atténuer l'effet d'un stress oxydatif.

**La biosynthèse *de novo* de méthionine** se fait en trois étapes. La première est la formation de cystathionine à partir de cystéine et d'O-phosphohomosérine (OPH, dérivé du métabolisme azoté) grâce à la cystathionine gamma-synthase (CGS). La deuxième étape est la synthèse d'homocystéine par la cystathionine beta-lyase (CBL) et la dernière est la synthèse de méthionine à partir d'homocystéine lors d'une réaction catalysée par la méthionine synthase (S. Ravanel et al., 1998). La méthionine joue un rôle majeur dans le métabolisme puisqu'elle est le premier acide aminé de toutes les protéines. Elle est aussi à l'origine de la synthèse de *S*-adénosylméthionine (AdoMet) qui est le précurseur de l'éthylène, des polyamines, et donneur de méthyl pour de nombreuses réactions de méthylation (Figure 12). La méthionine peut, avec l'AdoMet, former de la *S*-methylmethionine (SMM) lors d'une réaction catalysée par la méthionine *S*-méthyltransférase (MMT). La SMM peut ensuite être utilisée comme donneur de méthyl pour la synthèse de méthionine à partir d'homocystéine *via* l'homocystéine *S*-methyltransférase (HMT). La SMM contribue au transport du soufre dans la plante, des feuilles et des tiges jusqu'aux graines, ainsi qu'au contrôle de la concentration en AdoMet (Ranocha et al., 2001). La méthionine peut aussi être recyclée à partir d'AdoMet via le cycle de Yang, aussi appelé « methionine salvage pathway » (Sauter et al., 2013) (Figure 12).

Les différents mécanismes de régulation de l'assimilation du soufre ont récemment été recensés par (Aarabi et al., 2020). Ces régulations peuvent avoir lieu au niveau de la signalisation et de la détection (Ball et al., 2004; Dong et al., 2019, 2017; Hara-Nishimura et al., 1995; Hartmann, 2004; Hopkins et al., 2005; Hubberten et al., 2012; Kankipati et al., 2015; Lappartient et al., 1999; Maruyama-Nakashita et al., 2004b; Vauclare et al., 2002; Zhang et al., 2014), de la régulation transcriptionnelle (Aarabi et al., 2016; Bielecka et al., 2015; Falkenberg et al., 2008; Henríquez-Valencia et al., 2018; Huang et al., 2016; Koprivova et al., 2014; Lee et al., 2011, 2007; Maruyama-Nakashita et al., 2015, 2006, 2005, 2004b, 2004a; Mugford et al., 2009; Sakuraba and Yanagisawa, 2018; Wawrzyńska et al., 2015, 2010; Wawrzyńska and Sirko, 2016, 2014; Yatusевич et al., 2010) et de la régulation post transcriptionnelle (Allen et al., 2005; Berkowitz et al., 2002; Francois et al., 2006; Jones-Rhoades and Bartel, 2004; Kawashima et al., 2011, 2009; Liang et al., 2010; Shibagaki and Grossman, 2010; Wirtz et al., 2010; Yoshimoto et al., 2007). Ces mécanismes, décrits dans le tableau 1, font intervenir des

molécules signal comme l'OAS, le GSH, et des récepteurs kinases à répétitions riches en leucine (CLAVATA3 (CLV3)/EMBRYO SURROUNDING REGION (CLE) peptide and CLAVATA1 (CLV1)). Ils mettent aussi en jeu des protéines régulant la transcription, telles que MORE SULFUR ACCUMULATION 1 (MSA1) et SULFUR DEFICIENCY INDUCED (SDI). Chez *Arabidopsis*, SDI régule négativement la biosynthèse des glucosinolates en condition de carence en soufre (Aarabi et al., 2016). Cette régulation est susceptible de réorienter le métabolisme du soufre vers la biosynthèse des acides aminés soufrés. Quant à MSA1, il a été montré comme étant impliqué dans le contrôle épigénétique de l'homéostasie du soufre (Huang et al., 2019, 2016). Il module la méthylation, et par conséquent l'expression, de plusieurs gènes régulés en réponse à une carence en soufre, tels que *APR3*, *ATPS4*, et les gènes codant les transporteurs de sulfate *SULTR1;1* et *SULTR1;2*. Il a été suggéré que MSA1 contrôle la méthylation de l'ADN dans le noyau en maintenant la quantité d'AdoMet nécessaire aux processus de méthylation (Huang et al., 2016).

**Tableau 1 : La régulation de l'assimilation du sulfate chez les plantes** (d'après Aarabi et al., 2020)

Régulation de l'assimilation du sulfate	Références
<i>Molécules de signalisation et de détection</i>	
Le GSH a des effets d'inhibition par rétroaction sur l'absorption de sulfate et sur l'APR	Lappartient et al., 1999 ; Vauclare et al., 2002 ; Hartmann et al., 2004 ; Ball et al., 2004
La cytokinine agit comme un signal négatif pour la régulation des transporteurs de sulfate dans les racines	Maruyama-Nakashita et al., 2004
L'OAS agit comme une molécule de signalisation de la carence en soufre, induisant l'expression des gènes régulés positivement par cette carence	Hirai et al., 2003 ; Hubberten et al., 2012 ; Hopkins et al., 2005
<i>SULTR1;2</i> a un rôle de régulation de part son implication dans la détection de la concentration en sulfate	Zhang et al., 2014
Les transporteurs <i>Sul1</i> et <i>Sul2</i> de la levure ont été identifiés comme les premiers détecteurs du sulfate extracellulaire chez les eukaryotes	Kankipati et al., 2015
Un nouveau mécanisme de détection des précurseurs de la cystéine via TOR a été identifié chez les plantes	Dong et al., 2017
Les leucine-rich repeat receptor kinase <i>CLV3</i> , <i>CLE</i> et <i>CLV1</i> ont un rôle de régulation du développement des racines latérales en fonction de la disponibilité en soufre chez <i>Arabidopsis</i>	Dong et al., 2019
<i>Régulation transcriptionnelle</i>	
Le récepteur de cytokinine <i>CRE1</i> affecte la transcription de <i>SULTR1;2</i> en condition de carence en soufre	Maruyama-Nakashita et al., 2004
Une protéine phosphatase inconnue est requise pour la régulation positive de <i>SULTR1;1</i> en condition de carence en soufre	Maruyama-Nakashita et al., 2004 (2)
L'élément cis <i>SURE</i> (sulfur-responsive) a été identifié comme promoteur de <i>SULTR1;2</i>	Maruyama-Nakashita et al., (2005)
<i>SLIM1</i> a été identifié comme étant un régulateur de nombreux gènes sensibles à la carence en soufre	Maruyama-Nakashita et al., 2006 review dans Wawrzyńska and Sirko, 2014 et Marayuma-Nakashita et al., 2017
Etude de l'interaction entre la signalisation de l'auxine et l'homéostasie en soufre via des TFs responsables de la signalisation de l'auxine : <i>IAA13</i> , <i>IAA28</i> et <i>ARF-2</i>	Falkenberg et al., 2008
<i>SLIMA</i> se lie à la séquence consensus 20-t, appelée UPE-box, au niveau du promoteur de nombreux gènes sensibles à la carence en soufre	Wawrzyńska et al., 2010
Des TFs MYB, contrôlant les GSLs aliphatiques et indoliques ont été identifiés comme des régulateurs positifs de l'APR, de l'ATPS et de SIR	Yatusevich et al., 2010 ; Mugford et al., 2009
La régulation contrôlée par <i>HY5</i> a été identifiée comme synchronisant l'assimilation d'azote et de soufre en réponse à la lumière	Lee et al., 2007 ; Lee et al., 2011 ; review dans Sakuraba et al., 2018
Phytochrome and Flowering Time ( <i>PFT1</i> ) a été proposé comme un composant du circuit de régulation de l'assimilation du sulfate	Koprivova et al., 2014

21 TF répondant à la carence en soufre ont été identifiés : parmi les candidats il y avait MADS, HAT14, MYB31 et MYB45	Bielecka et al., 2015
SURE21 a été identifié sur la région non traduite 3' de SULTR2;1 il est responsable de l'activation de ce gène dans les racines	Maruyama-Nakashita et al., 2015
EIN3 se lie à SLIM1 et interfère avec le signal de carence en soufre, suggérant un possible lien de régulation entre l'éthylène et la voie du soufre	Wawrzyńska and Sirko, 2016 ; review dans Wawrzyńska et al., 2015
SDI1 et SDI2 ont été identifiés comme des répresseurs de GSL de par leur action directe avec MYB28 et de repression de la transcription des gènes de la voie de biosynthèse des GSLs	Aarabi et al., 2016
MSA1 (SHM7) a été identifié comme un régulateur de la synthèse d'AdoMet et de méthylation de l'ADN (potentiel régulateur de l'homéostasie en soufre par l'augmentation de la capacité d'absorption du sulfate	Huang et al., 2016
Identification de nouveaux modules géniques sous-jacents à la réponse en soufre chez <i>Arabidopsis</i>	Henríquez-Valencia et al., 2018
<b>Régulation post transcriptionnelle</b>	
La régulation post-transcriptionnelle a été proposée dans l'expression en réponse au soufre de SULTR1;1 et de SULTR1;2 dans les racines d' <i>Arabidopsis</i>	Yoshimoto et al., 2007
En condition de carence en soufre, SLIM1 induit l'expression du miRNA mi395, qui vise ATPS1, 3 et 4 ainsi que SULTR2;1, augmentant le transfert du soufre de racines vers les parties aériennes	Jones-Rhoades and Bartel, 2004 ; Allen et al., 2005 ; Kawashima et al., 2009 ; Liang et al., 2010 ; Kawashima et al., 2011
Identification d'une régulation de SERAT et d'OASTL par l'OAS et le sulfide	Berkowitz et al., 2002 ; Francois et al., 2006 ; Wirtz et al., 2010
La réduction de l'absorption du sulfate dans la levure via l'interaction de l'OASTL avec le domaine STAS de SULTR1;2 a été décrite	Shibagaki and Grossman, 2010

## 2.5 Stratégies visant à modifier le contenu en acides aminés des graines

Quatre stratégies différentes ont été utilisées pour augmenter le contenu en méthionine dans les graines (Amir et al., 2019) (Tableau 2). La première a consisté à augmenter l'expression, par transgénèse, de gènes codant des protéines de réserve riches en méthionine (ex. albumines 2S) (Altenbach et al., 1989; Chiaiese, 2004; Dinkins et al., 2001; Kim and Krishnan, 2004; Lee et al., 2003; Li et al., 2005; Molvig et al., 1997; Y. Zhang et al., 2014). La teneur en méthionine des graines a ainsi été augmentée chez différentes espèces, dont le lupin (Molvig et al., 1997), mais au dépend de certaines molécules de défense comme le glutathion (Tabe and Droux, 2002). La deuxième stratégie a consisté à diminuer la synthèse des protéines de réserve pauvres en méthionine pour favoriser l'accumulation des protéines de réserve riches en méthionine (Lai and Messing, 2002; Newell et al., 2014; Pandurangan et al., 2015; Taylor et al., 2008). Cette approche a permis d'augmenter le contenu en méthionine incorporée dans les protéines de 10-13% (chez le haricot) à 56% (chez le maïs). La troisième stratégie a consisté à manipuler des enzymes impliquées dans la biosynthèse, le catabolisme, ou le recyclage de la méthionine (Cohen et al., 2014, 2016, 2017; Hanafy et al., 2013; Karchi et al., 1993; W.-S. Kim et al., 2012; Matityahu et al., 2013, 2019; Nguyen et al., 2012; Song et al., 2013; Xiang et al., 2018). Pour exemple, des lignées transgéniques surexprimant une forme de CGS insensible au rétrocontrôle négatif par la méthionine produisent des graines dont le contenu en méthionine est augmenté de 60% (Cohen et al., 2014). Comme autre exemple de rétrocontrôle, le tryptophane inhibe sa

propre voie de biosynthèse en régulant négativement la première enzyme de cette voie, l'antranilate synthase (Ufaz and Galili, 2008). Des lignées transgéniques transformées avec des allèles d'antranilate synthase insensibles au rétrocontrôle par le tryptophane chez *Arabidopsis* ont permis d'augmenter respectivement la synthèse de méthionine (Cohen et al., 2014) et de tryptophane dans les graines (Inaba et al., 2007; Ishimoto et al., 2010; Ufaz and Galili, 2008). Cependant, si les acides aminés soufrés s'accumulent sous forme libre dans la graine, ils peuvent induire une toxicité et diminuer la vigueur germinative des graines (Shen et al., 2002). Il est donc important de favoriser leur incorporation dans les protéines de réserve. C'est pourquoi la quatrième stratégie a consisté à augmenter la synthèse de méthionine tout en augmentant la synthèse des protéines de réserve riches en méthionine (Demidov et al., 2003; Planta et al., 2017; Tabe et al., 2010). Une augmentation de 57% du contenu en méthionine des grains de maïs a été obtenue en surexprimant une PAPS réductase d'*Escherichia Coli* (enzyme impliquée dans la formation de sulfite à partir de PAPS) suivi d'une introgression du transgène dans des fonds génétiques d'intérêt agronomique (Planta et al., 2017).

La variabilité génétique naturelle ou induite par TILLING peut être exploitée pour améliorer l'équilibre en acides aminés des graines. Chez le pois, des mutants TILLING du gène *ABI5* produisent des graines pauvres en vicilines sans modification significative de la teneur en protéines (Le Signor et al., 2017). Comme les vicilines sont pauvres en acides aminés soufrés, ces modifications peuvent être nutritionnellement avantageuses.

**Tableau 2 : Les différentes approches utilisées pour augmenter le contenu en méthionine dans les graines.** D'après Amir et al., 2019.

		Teneur en méthionine dans les graines transgéniques					
		Manipulation			Méthionine soluble	Méthionine incorporée dans les SSP	Références
	Espèce	Gène ciblé	Type d'expression	Méthionine soluble	Méthionine incorporée dans les SSP	Références	
STRATEGIE 1	<b>Régulation positive des SSP riches en méthionine</b>						
	Tabac	2S BNA	Spécifique graine	NR	31% fois plus	Altenbach et al. (1989)	
	Soja	2S 11 kDa $\delta$ -zein	Spécifique graine	NR	75% <sup>b</sup> fois plus	Kim and Krishnan (2004)	
	Lupin	2S SSA	Spécifique graine	NR	90% fois plus	Molvig et al. (1997)	
	Pois chiche	2S SSA	Spécifique graine	NR	90% fois plus	Chiaiese et al. (2004)	
	Riz	2S SA	Spécifique graine	NR	75% fois plus	Lee et al. (2003)	
	Soja	15 kDa $\beta$ -zein	Spécifique graine	NR	20% fois plus	Dinkins et al. (2001)	
	Soja	27 kDa $\gamma$ -zein	Spécifique graine	NR	18% fois plus	Li et al. (2005)	
	Soja	Synthetic MB-16	Spécifique graine	NR	16% fois plus	Zhang et al. (2014)	
	Haricot	2S BNA	Constitutive	NR	23% fois plus	Aragão et al. (1999)	
STRATEGIE 2	<b>Suppression des SSP pauvres en méthionine</b>						
	Haricot	Phaseolin	Mutant	27%	10% fois plus	Taylor et al. (2008)	
	Haricot	Phaseolin	Mutant	NR	13% fois plus	Pandurangan et al. (2015)	
	Maïs	27 kDa $\gamma$ -zein	Mutant	NR	56% fois plus	Newell et al. (2014)	
	Maïs	10 kDa $\delta$ -zein	Mutant	NR	37% fois plus	Lai and Messing (2002)	

Augmentation synthèse de méthionine						
STRATEGIE 3	Soja, haricot azuki	mto1-1 CGS	Spécifique graine	2x plus	NS	Hanafy et al. (2013)
	Tabac	AtD-CGS	Spécifique graine	NS	60% fois plus	Matityahu et al. (2013)
	Soja	AtD-CGS	Spécifique graine	91% ; 3.2x plus	2.3 fois plus ; 3 fois plus	Song et al. (2013) ; Cohen et al. (2016)
	Arabidopsis	AtD-CGS	Spécifique graine	6x plus	60% fois plus	Cohen et al. (2014)
	Tabac	EcAK	Spécifique graine	3x plus	6.5% fois plus	Karchi et al. (1993)
	Arabidopsis	AtHMT3	Spécifique graine	13%	33% fois plus	Cohen et al (2017)
	Tabac	AtD-CGS+EcAK	Spécifique graine	2x plus	97% fois plus	Matityahu et al. (2019)
	Maïs	AtSAT	Spécifique graine	4x plus	4 fois plus	Xiang et al. (2018)
	Riz	EcSAT	Constitutive	1.4x plus	4.8 fois plus	Nguyen et al. (2012)
	Soja	OASTL	Constitutive	NR	17% fois plus	Kim et al. (2012)
Régulation positive des SSP riches en méthionine couplée à l'augmentation synthèse en méthionine						
STRATEGIE 4	Vesce de Narbonne	2S BNA+EcAK	Spécifique graine	NR	2 fois plus	Demidov et al. (2003)
	Lupin bleu	AtSAT+BNA	Spécifique graine	NR	2 fois plus	Tabe et al. (2010)
	Maïs	EcAPR+10 kDa $\delta$ -zein	Spécifique feuilles	NR	57% fois plus	Planta et al. (2017)

\* dans la fraction protéique soluble dans l'alcool, mais potentiellement non significatif comparé au contenu total en méthionine des graines du témoin sauvage. 2S BNA : 2S Brazil nut albumin ; 2S sA : 2S sesame albumin ; 2S SSA : 2S sunflower seed albumin; AtCGS : Arabidopsis cystathionine- $\gamma$ -synthase ; AtD-CGS : feedback-insensitive form of the Arabidopsis cystathionine- $\gamma$ -synthase ; AtHMT3 : Arabidopsis homocysteine S-methyltransferase 3 ; BNA : Brazil nut albumin ; EcAK : *Escherichia coli* feedback-insensitive aspartate kinase ; EcAPR : *Escherichia coli* 3'-phosphoadenosine-5'-phosphosulfate reductase ; EcSAT : *Escherichia coli* serine acetyl transferase ; OASTL : O-acetyl(thiol)lyase ; NR : non reporté; NS : non significatif.

## 3. Variations environnementales : impact sur le rendement et la composition protéique des graines de légumineuses

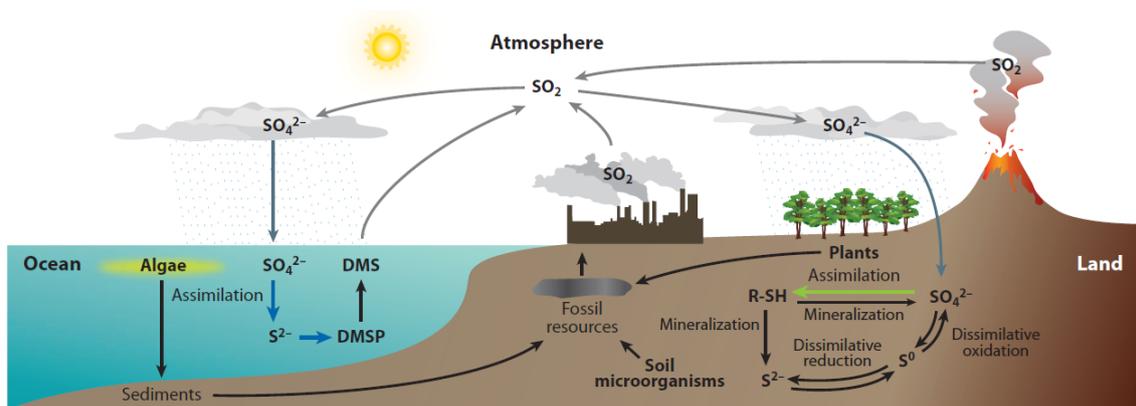
### 3.1 Impact des variations environnementales sur le développement des plantes

Les plantes sont soumises à de nombreux stress d'origine biotique et abiotique pouvant impacter leur développement et réduire le rendement et la qualité des produits. Le pois est par exemple très sensible aux fortes températures et à la sécheresse, particulièrement durant la phase reproductive (Bueckert et al., 2015; Guilioni, 1997; New et al., 1994; Tafesse et al., 2020). Ces stress induisent un avortement des graines et réduisent le nombre de nœuds reproducteurs, affectant la production de gousses et le nombre de graines par gousse. La teneur en protéines varie fortement en réponse aux stress abiotiques : elle augmente en réponse aux fortes températures et diminue avec l'augmentation des précipitations (Karjalainen and Kortet, 1987). Les variations environnementales influencent aussi la composition protéique des graines. Par exemple, une augmentation du ratio globulines 7S/11S a été observée dans les

graines de pois produites une année de forte chaleur (Bourgeois et al., 2009). Cette augmentation de la proportion de globulines 7S peut s'expliquer par des périodes de remplissage et de maturation plus courtes (Figure 9), limitant la possibilité d'accumulation des légumineuses qui, dans des conditions standards, continuent d'être synthétisées tardivement lors du développement de la graine. Il a aussi été montré que des facteurs agronomiques tels que la composition des sols pouvaient influencer le ratio 7S/11S (Mertens et al., 2012).

### 3.2 Les stress nutritionnels : le cas d'une carence en soufre

La nutrition soufrée joue un rôle prépondérant dans le maintien des rendements et de la qualité des graines de légumineuses (Henriet et al., 2019). Le soufre est un élément naturellement présent dans la nature (ex. océans, marais, volcans). Sa présence à des taux non négligeables dans l'atmosphère sous forme de dioxyde de soufre est majoritairement d'origine anthropique. En effet, le dioxyde de soufre a été rejeté par les usines de façon massive dans l'atmosphère à l'ère industrielle (Figure 14). Afin de limiter l'impact de ces émissions sur l'environnement et sur la santé, des réglementations antipollution ont été mises en place (Mcgrath et al., 2003), ce qui a réduit les retombées de soufre atmosphérique sur les sols.



**Figure 14 : Cycle biogéochimique du soufre dans la nature.** Le sulfate est assimilé par les organismes photosynthétiques : les plantes (flèche verte) et les algues (flèche bleue). Les microorganismes du sol utilisent le soufre dans des réactions dissimilatives (réduction des sulfates en milieu anaérobie). Le soufre sous forme volatile peut être rejeté dans l'atmosphère, principalement par les volcans et les activités humaines. Le soufre sous forme volatile est ensuite oxydé en sulfate dans l'atmosphère (Figure extraite de Takahashi et al., 2011).

Dans les sols, le soufre subit des étapes d'oxydation et de réduction via les micro-organismes pour former du sulfate assimilable par les plantes (Takahashi et al., 2011). Le sulfate est la principale source de soufre pour les plantes bien que le soufre puisse aussi être absorbé au niveau des stomates sous forme de gaz (sulfure d'hydrogène) (Calderwood and Kopriva, 2014). La diminution des retombées de soufre atmosphérique et l'épuisement des réserves de soufre dans les sols induisent l'apparition de symptômes de carence en soufre chez les plantes. Ces

symptômes incluent un jaunissement des feuilles, une carence en soufre induisant une diminution de la production de chlorophylle (Hoefgen and Nikiforova, 2007).

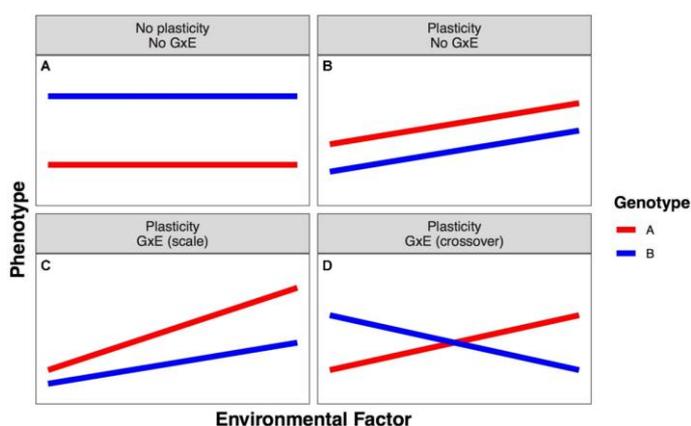
Une carence en soufre peut avoir un effet plus ou moins marqué sur le développement des plantes et des graines en fonction du stade auquel elle survient. Chez *M. truncatula*, une carence en soufre appliquée à un stade végétatif diminue le rendement, l'allocation d'azote et de carbone vers les graines, la quantité d'oligosaccharides dans les graines matures et la vigueur germinative (Zuber et al., 2013). Si la carence est appliquée lors de la période reproductive, l'impact sur le rendement et l'allocation des nutriments est moindre ; les graines sont néanmoins affectées dans leur composition protéique et vigueur germinative (Zuber et al., 2013). Une carence en soufre a pour conséquence un changement extrême de la composition protéiques des graines de pois et de *M. truncatula* avec une forte augmentation de la proportion des vicilines (très pauvres en acides aminés soufrés) et une forte diminution de la quantité de légumine (relativement plus riches en acides aminés soufrés) (Chandler et al., 1983; Henriët et al., 2019; Zuber et al., 2013). Des études conduites chez le pois ont montré que la diminution de l'accumulation des légumine en réponse à une carence en soufre est régulée transcriptionnellement (Chandler et al., 1984). La littérature suggère que ce rétrocontrôle implique deux métabolites soufrés, la méthionine et l'OAS, qui agissent comme des signaux régulateurs de l'expression des gènes codant les légumine (Kim et al., 1999; Tabe et al., 2010) (Figure 12).

Il a été montré que le stress soufré, en association avec d'autres types de stress, pouvait avoir des impacts différents sur la plante comparativement à un stress soufré seul. Chez le colza, un stress thermique associé à une carence en soufre augmente la remobilisation du soufre vers les graines mettant en évidence que la remobilisation du soufre est un processus stimulé par un stress thermique. Le stress thermique permet aussi de limiter l'impact de la carence en soufre sur la composition protéique en permettant d'avoir un ratio protéines pauvres en soufre/protéines riches en soufre plus élevé qu'en condition de carence en soufre seule (Brunel-Muguet et al., 2015). Chez le pois, le même phénomène est observé en condition de stress hydrique couplé à un stress soufré, un stress hydrique modéré permettant de limiter l'impact du déficit en soufre sur l'accumulation des globulines 11S (riches en soufre) (Henriët et al., 2019).

# 4. Déterminismes génétiques du rendement et de la composition protéique des graines de légumineuses

## 4.1 Plasticité vis à vis des fluctuations environnementales

La plasticité est définie comme la gamme de variation des phénotypes que peut prendre un génotype en réponse à l'environnement (Bradshaw, 1965). En tant qu'organismes sessiles, les plantes doivent pouvoir résister aux variations environnementales auxquelles elles sont soumises, que ce soit, par exemple, des variations de température, de disponibilité en eau, en sels minéraux ou des caractéristiques de sol comme la salinité. Ces mécanismes se mettent en place différemment en fonction des génotypes et des conditions environnementales et génèrent des variations dans les réponses phénotypiques qui peuvent entraîner, chez les plantes agronomiques, une instabilité de rendement ou de qualité des produits. Il est donc important de pouvoir caractériser et quantifier ces variations de plasticité. La variation phénotypique observée se décompose en variance génétique ( $VarG$ ), environnementale ( $VarE$ ), d'interaction génotype\*environnement ( $VarG*E$ ) et d'une variance d'erreur. La variance d'interaction  $VarG*E$ , estimable par analyse de variance, représente la variation dans les réponses des génotypes aux différents effets de l'environnement. La contribution de cette interaction  $GxE$  se visualise au travers du tracé des normes de réaction des phénotypes en fonction de l'environnement (Figure 15). Les génotypes plastiques présentent une norme de réaction avec une pente différente de zéro (Figure 15 B, C, D) mais des variations de plasticité existent entre génotypes seulement en présence d'interaction  $GxE$  (Figure 15 C et D, d'après (Kusmec et al., 2018)) où les génotypes répondent différemment à la variation environnementale.

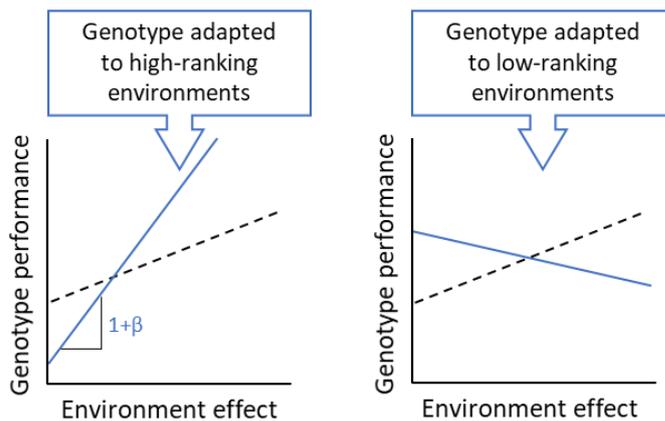


**Figure 15 : Exemple de normes de réactions.** (A) Les génotypes expriment le même phénotype quel que soit l'environnement ; pas de plasticité. (B) Les génotypes expriment la même variation plastique en fonction de l'environnement ; pas d'interaction ( $GxE$ ) ; la plasticité existe mais n'est pas variable. (C) et (D) Les deux génotypes expriment différentes réponses plastiques en fonction de l'augmentation de l'indice environnemental. Présence d'interaction  $GxE$  qui induit une réponse graduelle des génotypes (C, scaled  $GxE$ ) ou une réponse inversée (D, crossover  $GxE$ ). D'après Kusmec et al., 2018.

Les normes de réaction peuvent présenter des formes polynomiales variées mais, en présence d'un petit nombre d'environnements, il est courant de les modéliser de façon linéaire comme l'ont proposé Finlay et Wilkinson (FW) (Finlay and Wilkinson, 1963) (Figure 16):

$$y_{ijk} = \mu + g_i + h_j + \beta h_j + \varepsilon_{ijk},$$

où  $y_{ijk}$  est la valeur du phénotype pour le  $k^{\text{ième}}$  replicat du  $i^{\text{ième}}$  génotype observé dans le  $j^{\text{ième}}$  environnement,  $g_i$  est l'effet moyen du  $i^{\text{ième}}$  génotypes,  $h_j$  est l'effet moyen du  $j^{\text{ième}}$  environnement, et  $\varepsilon_{ijk}$  est le terme d'erreur.



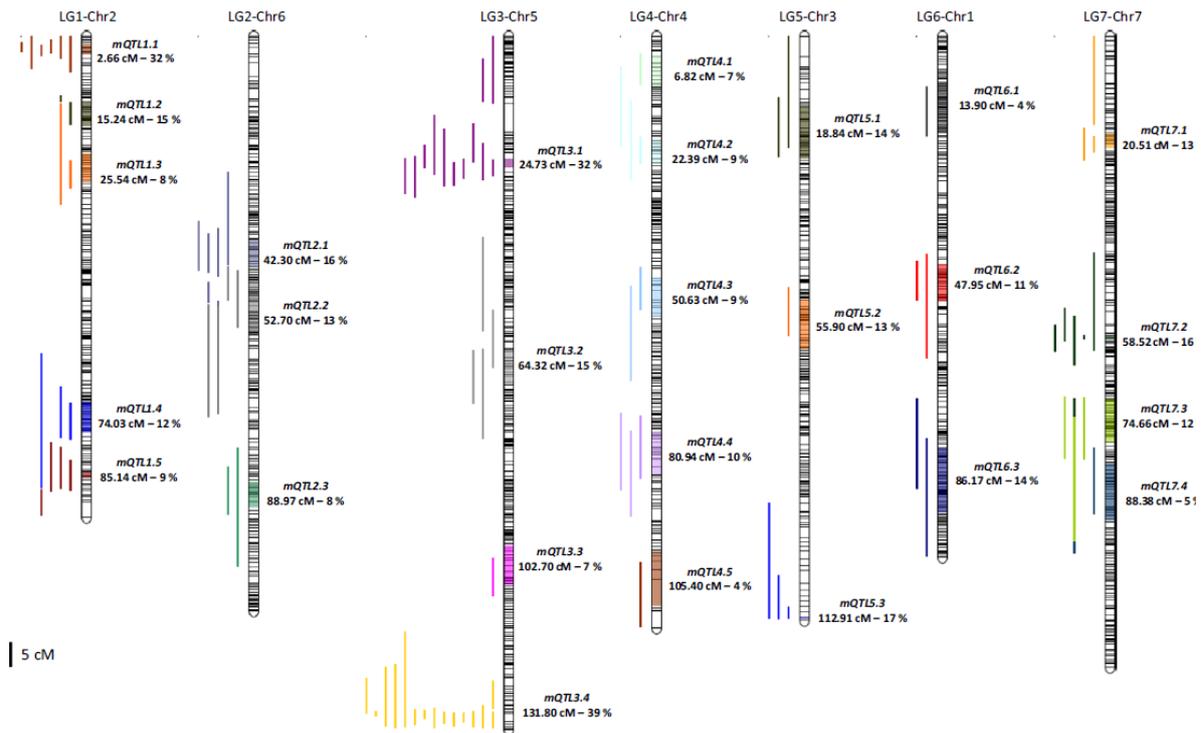
**Figure 16 : Illustration de la méthode de régression de Finlay et Wilkinson (1963).** La pente de la droite de régression représentée sur le schéma est  $(1+\beta)$ .

Cette méthode de régression linéaire utilise la valeur moyenne des phénotypes observés dans chaque environnement comme indices environnementaux à partir desquels le phénotype de chaque individu est régressé. La pente de cette régression linéaire,  $\beta$ , caractérise la plasticité du génotype, les valeurs positives indiquent les génotypes valorisant bien les milieux favorables (à fort indice environnemental), les valeurs négatives les génotypes les mieux adaptées à des milieux défavorables (à faible indice environnemental) et les valeurs proches de zéro les génotypes moyennement adaptés à tous les milieux. Cet indice caractérise donc un génotype à l'autre et peut être un outil de mesure de la variabilité génétique associée à la plasticité (Kusmec et al., 2017). Deux visions existent sur les bases génétiques de la plasticité phénotypique. Selon (Via, 1993), des gènes sensitifs exprimés spécifiquement dans certains environnements détermineraient la valeur du phénotype dans ces environnements et seraient indirectement responsables de la plasticité phénotypique. Selon (Pigliucci, 1996), la plasticité serait contrôlée par des gènes de plasticité (des gènes régulateurs) et pourrait être une cible de sélection. De plus, ce concept suppose une potentielle indépendance entre la valeur moyenne du phénotype et sa sensibilité à l'environnement.

Depuis, de nombreuses études ont mis en évidence les bases génétiques de la plasticité du rendement de plusieurs espèces telles que l'orge, le maïs ou la tomate en utilisant l'indice de plasticité de Finlay et Wilkinson. Lacaze et al. (2009) ont étudié la variation des pentes de régression de Finlay et Wilkinson sur deux populations de lignées recombinantes d'orge et dans 7 à 22 environnements. Ils ont mis en évidence des QTLs de plasticité colocalisant toujours avec des QTLs de rendement ou de poids de graines dans un environnement, avançant l'hypothèse de la présence de gènes sensitifs en accord avec le modèle de Via (1993). Ces auteurs concluent que le type de modèle de plasticité (gènes sensitifs *vs* gènes régulateurs) est largement dépendant du nombre d'environnements étudiés, la mise en évidence de gènes spécifiques à la plasticité serait limitée aux dispositifs expérimentant peu d'environnements. Or, de nombreuses études ont mis en évidence la coexistence des deux modèles dans leurs études en exploitant la variabilité de plus de 12 environnements, comme (Kraakman et al., 2004) sur le rendement de l'orge dans 15 environnements, (Diouf et al., 2020) sur les caractères de rendement de la tomate sur 12 environnements ou (Li et al., 2019) sur le poids de la graine du maïs sur 12 environnements. Cette dernière étude est plus précisément une analyse génétique de la plasticité de Finlay Wilkinson par approche « genome wide » (GWAS) regroupant 1972 RILs et 43252 SNP : 30% seulement des QTLs détectés sont communs au phénotype et à la plasticité laissant une place prépondérante aux gènes spécifiques de la plasticité.

## 4.2 Etude du déterminisme génétique du rendement et de la composition protéique des graines de légumineuses

De nombreuses études génétiques se sont intéressées à la recherche de loci impliqués dans le contrôle du rendement et de la qualité de la graine chez le pois (Bourgeois et al., 2011; Burstin et al., 2007; Ferrari et al., 2016; Gali et al., 2018; Klein et al., 2014; Krajewski et al., 2012; Moreau et al., 2018; Tar'an et al., 2004; Timmerman-Vaughan et al., 2005). La plupart de ces analyses se sont basées sur une ou un petit nombre de population biparentale. Récemment, une méta-analyse de données de QTL obtenus à partir de 9 populations de lignées recombinantes interconnectées a mis en évidence 27 métaQTL (Klein et al., 2020) dont 9 sont liés aux variations de nombre de graines, 8 au poids de graines, 17 au poids d'une graine et 17 à la teneur en protéines confirmant la complexité de l'architecture génétique de ces traits (Figure 17).



**Figure 17 : Carte génétique des 27 metaQTL (mQTL) détectés pour des composantes du rendement chez le pois.** Les variables étudiées sont le poids de graines (SW), le nombre de graines (SN), le PMG (TSW) et la teneur protéique des graines (SPC). La position(cM) et le maximum de variance phénotypique expliquée pour chaque mQTL sont indiqués à droite des groupes de liaison. Les intervalles de confiance sont représentés par la couleur sur le groupe de liaison. (Figure extraite de Klein et al., 2020)

L'accès aux technologies de génotypage haut débit a permis de développer les approches GWAS chez le pois. Récemment, une analyse d'association sur une population de 135 accessions caractérisées avec 16877 SNP (Gali et al., 2019) a mis en évidence des loci consensus de rendement et traits de qualité des graines identifiés sur plusieurs lieux et années d'évaluation (comme par exemple Chr1LG6\_57305683 et Chr1LG6\_366513463 pour le poids de graines, et Chr3LG5\_194530376 pour la teneur en protéines). Ce dernier colocalise avec le mQTL3.1 de (Klein et al., 2020), alors que le premier locus colocalise avec le metaQTL de nombre de graines mQTL6.1. Les écarts entre les deux études s'expliquent de par la base génétique plus ou moins large des populations utilisées et aussi le relativement faible nombre d'individus impliqués dans l'analyse GWAS.

Partant du constat que nombre de QTLs de rendement ou de teneur en protéines des graines colocalisent avec des gènes de développement, (Burstin et al., 2007) ont suggéré que ces QTLs correspondent à des gènes impliqués soit dans le développement d'organes sources pour le rendement soit dans la force de puits des graines pour la teneur en protéine. Par exemple, les gènes contrôlant l'architecture des plantes comme *Le* qui détermine la longueur des entre-

nœuds (et donc la hauteur de la plante, cartographié en bas du LG3Chr5) ou *Af* (Afila, bas du LG1Chr2) qui contrôle le changement des folioles en vrilles augmentent le développement d'organes sources alors que les gènes *rugosus* qui modifient la teneur en amidon des graines (phénotype de graine ridée) augmentent la force de puits des graines.

Toutes ces études ont permis d'identifier des listes de gènes candidats situés sous ces zones ainsi que des marqueurs moléculaires utilisables en sélection.

Concernant la composition protéique des graines, une approche PQL (Protein Quantity Loci) (Damerval et al., 1994) sur une population de lignée recombinantes issues du croisement de deux génotypes présentant de forts contrastes de composition protéique (VavD265 accumule plus de viciline alors que Caméor accumule plus de convicilines et de légumine) a permis d'identifier les régions du génome associées à la composition protéique des graines chez le pois (Bourgeois et al., 2011). Ces travaux ont mis en évidence des régions du génome régulant la synthèse ainsi que le clivage des polypeptides. Les cartes consensus pois-*M. truncatula* ont mis en évidence la conservation d'environ la moitié des PQL entre ces deux espèces, suggérant qu'il existe des mécanismes communs de régulation de mise en réserve des globulines (Le Signor et al., 2017). Dans cette même étude, des données protéomiques mono et bi-dimensionnelles disponibles pour 175 écotypes de *M. truncatula* ont permis de mener des analyses GWAS sur ces données de composition protéiques des graines. Une liste de gènes candidats impliqués dans la transcription, le remodelage de la chromatine, les modifications post-traductionnelles, le transport et l'adressage des globulines vers les vacuoles de stockage, a été établi pour la première fois chez les légumineuses. La recherche des profils d'expression dans la graine des gènes présents sous les régions PQL (Benedito et al., 2008) a permis de créer une carte consensus génétique-expressionnelle chez *M. truncatula* composée de 15 495 gènes avec leurs données d'expression.

## CHAPITRE II :

Etude du déterminisme génétique de la plasticité de la composition protéique des graines de *M. truncatula* en réponse à des variations environnementales.

## **Objectifs**

---

Stabiliser la valeur nutritionnelle des graines face aux contraintes environnementales est l'un des principaux défis à relever pour promouvoir le développement de la culture des légumineuses. Pour atteindre cet objectif, nous avons mené une étude génétique de la plasticité de la composition protéique sur une large collection de génotypes cultivée dans quatre environnements contrastés chez la légumineuse modèle *M. truncatula*. Les connaissances acquises sur les gènes candidats ainsi identifiés pourront par la suite être transférées à une espèce agronomique et utilisées en sélection afin d'obtenir des variétés améliorées dans la stabilité de leur composition protéique.

## **Stratégie**

---

Les travaux de ce chapitre s'inscrivent dans le cadre du projet ANR-REGULEG (Coord. Julia Buitink, IRHS Angers) dont l'objectif est d'étudier la régulation de l'adaptation de la graine à son environnement pour produire des variétés mieux adaptées aux changements climatiques. Les différentes études de ce projet portent sur deux expérimentations basées à l'IRHS d'Angers et à l'AGAP de Montpellier, lieux dans lesquels 200 écotypes de *M. truncatula* ont été cultivés dans deux conditions contrastées : présence vs absence de stress hydrique à Angers et semis d'hiver vs semis de printemps à Montpellier. Les compositions protéiques des graines pour chacun des écotypes dans chacun des quatre environnements ont été analysées par électrophorèse monodimensionnelle suivie de la quantification relative de chacune des bandes détectées. Ces données ont ensuite permis de calculer un indice de plasticité pour chacune des bandes protéiques par la méthode de Finlay et Wilkinson (1963) et d'évaluer sa variabilité dans notre collection. Des analyses de génétique d'association (GWAS) menées sur les données obtenues pour chacun des environnements ainsi que pour la plasticité ont permis d'établir une liste de gènes candidats associés à la variabilité observée et pouvant potentiellement faire varier la composition protéique des graines en réponse à des stress environnementaux.

## **Résultats majeurs**

---

La liste de gènes trouvée pour les indices de plasticité est enrichie en facteurs de transcription, en gènes impliqués dans la réparation de l'ADN et dans la transduction du signal ainsi que des gènes du métabolisme du soufre et de l'aspartate, la plupart étant dérégulés en conditions de stress abiotiques. Nous avons focalisé sur les gènes du métabolisme du soufre et de l'aspartate car ces voies métaboliques mènent à la synthèse des acides aminés essentiels comme la méthionine et la lysine et que l'équilibre entre ces acides aminés est un point crucial à améliorer chez les légumineuses. En effet, les graines des légumineuses sont riches en lysine et pauvres

en méthionine. Par l'étude d'un mutant, nous avons validé l'implication du gène *HMT*, conduisant à la régénération de la méthionine à partir de *S-methylmethionine*, dans le contrôle de la plasticité de la composition protéique des graines chez *M. truncatula*. L'analyse globale de ces gènes a mis en évidence la contribution majeure des voies de recyclage de la méthionine dans l'établissement de la composition protéique et de sa plasticité. Ces données fournissent une ressource unique de gènes qui peuvent être ciblés pour atténuer les impacts négatifs des stress environnementaux sur la qualité des protéines de graines.

*Heading: Seed protein plasticity in M. truncatula*

## **Elucidating the Genetic Determinism of Seed Protein Plasticity in Response to the Environment using *Medicago truncatula***

Kevin Cartelier<sup>1</sup>, Delphine Aimé<sup>1</sup>, Joseph Ly Vu<sup>2</sup>, Lucie Combes-Soia<sup>3</sup>, Valérie Labas<sup>3</sup>, Jean-Marie Proserpi<sup>4</sup>, Julia Buitink<sup>2</sup>, Karine Gallardo<sup>1\*</sup> and Christine Le Signor<sup>1\*</sup>

<sup>1</sup>Agroécologie, AgroSup Dijon, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), Université Bourgogne Franche-Comté, 21000 Dijon, France; <sup>2</sup>Institut de recherche en horticulture et semences (IRHS), INRAE, Agrocampus-Ouest, Université d'Angers, SFR 4207 QuaSaV, 49071 Beaucouzé, France; <sup>3</sup>Physiologie de la Reproduction et des Comportements (PRC) UMR85, INRAE, CNRS, Université de Tours, IFCE, Nouzilly, France; <sup>4</sup>Genetic Improvement and Adaptation of Mediterranean and Tropical Plants (AGAP), INRAE, Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Montpellier Supagro, 34060 Montpellier, France.

\*Co-corresponding authors, contributed equally to this work

*Christine Le Signor; Tel: + (00)33-380-693-147; christine.le-signor@inrae.fr*

*Karine Gallardo; Tel: + (00)33-380-693-391; karine.gallardo-guerrero@inrae.fr*

## Summary

Because extreme environmental events are expected to increase with climate change, identifying candidate genes for stabilizing the protein composition of legume seeds or optimizing it in a given environment becomes increasingly important. To elucidate the genetic determinants of seed protein plasticity, major seed proteins from 200 ecotypes of *M. truncatula* grown in four contrasted environments were quantified after one-dimensional electrophoresis. The plasticity index of these proteins was recorded for each genotype as the slope of Finlay and Wilkinson's regression and used for Genome Wide Association Studies, enabling the identification of candidate genes determining this plasticity. The candidate list was enriched in genes related to transcription, DNA repair, and signal transduction, many of them being stress-responsive. The other over-represented genes were related to sulfur and aspartate family pathways leading to the synthesis of the nutritionally essential amino acids methionine and lysine. By replacing these genes in metabolic pathways, and using a *M. truncatula* mutant impaired in regenerating methionine from *S*-methylmethionine, we discovered that methionine recycling pathways are major contributors to globulin composition establishment and plasticity. These data provide a unique resource of genes that can be targeted to mitigate negative impacts of environmental stresses on seed proteins.

## Significance statement

In the context of climate change, legumes that produce high content protein seeds without nitrogen fertilizer became very attractive for human and livestock nutrition and optimizing and stabilizing their seed protein yield and composition is of crucial importance. By exploiting through GWAS natural genetic variability in *Medicago truncatula*, methionine recycling pathways appeared to be major determinants of the plastic response of seed protein composition.

**Keywords:** Genome-wide association studies (GWAS), plasticity, seed, storage proteins, legumes, methionine recycling

# 1. Introduction

The Fabaceae is a large family of legume plants that includes key crops such as soybean (*Glycine max*), pea (*Pisum sativum*), lentils (*Lens culinaris*) and common bean (*Phaseolus vulgaris*) (Graham and Vance, 2003). These plants produce protein-rich seeds in the absence of nitrogen fertilization due to the ability of their root systems to establish symbiosis with nitrogen-fixing rhizobia. They are thus good candidates to meet the challenge of increasing European autonomy in vegetable proteins while reducing nitrogen fertilizer use. Legume seed proteins are rich in lysine but poor in tryptophan and sulfur amino acids (cysteine (Cys) and methionine (Met)), making them complementary to cereals in both food and feed (Perrot, 1995). The major determinants of seed protein content in legumes are the globulins, which account for up to 70-80% of total seed proteins. The globulins are separated in two fractions (7S and 11S) according to their sedimentation coefficients after ultracentrifugation (Osborne, 1909). In mature seeds, the 11S globulins, named legumins, are hexamers of 350-400 kDa made of acidic and basic polypeptides linked by a disulfide bond. The 7S globulins are composed of vicilins and convicilins, which accumulate as trimeric forms of 150-180 kDa and 210-280 kDa, respectively. According to their amino acid composition and structure, the globulins may have different nutritional and functional properties (Créviu-Gabriel et al., 1999; Duranti and Gius, 1997; Rubio et al., 2014). For instance, 7S globulins are poor in the sulfur amino acids Met and Cys compared to 11S globulins.

There is a large genetic variability for seed protein content and composition in legumes such as pea and soybean (Bourgeois et al., 2011; Vollmann et al., 2000). The genotypic effect on seed protein content and composition is preponderant over the environmental effect (Bourgeois et al., 2009; Vollmann et al., 2000). However, a three-year field trial on pea with heat/water stress periods demonstrated that significant environmental effects could be identified, such as a higher 7S/11S globulin ratio in the seeds (Bourgeois et al., 2009). Moreover, the balance between the accumulation of the 7S and 11S globulins strongly depends on sulfur nutrition, as shown in soybean, pea and the model legume species *Medicago truncatula* (Gayler and Sykes, 1985; Henriot et al., 2019; Sexton et al., 1998; Zuber et al., 2013). Because extreme environmental events are expected to increase with climate change, identifying candidate genes for stabilizing seed globulin composition or optimizing it in a specific environment is of crucial importance. In studies of the genetic determinism of seed protein composition in both pea (Bourgeois et al., 2011) and *M. truncatula* (Le Signor et al., 2017), candidate loci important for globulin synthesis

and packaging were identified. However, no studies have investigated, so far, the genetic determinism of the plasticity of seed globulin composition in response to the environment.

The plasticity is the ability of a genotype to modify its phenotype in response to environmental changes (Bradshaw, 2006). As described by (Des Marais and Juenger, 2010), plants can use three different strategies to cope with environmental changes : a general strategy where a genotype optimizes its phenotype to adapt to several environments, a specialization strategy consisting in producing a phenotype enabling survival in a few environments, and an adaptive plastic response where a genotype produces an appropriate phenotype in a particular environment. Considering these differences, genetic approaches for crop improvement aim to discriminate genomic regions specifically implicated in the plasticity of the phenotype and those controlling both the plasticity and the phenotypic values in particular environments (Des Marais et al., 2013). The plasticity can differ between genotypes from the same species due to genotype by environment interactions (Des Marais et al., 2013; El-Soda et al., 2014). The plasticity of a genotype can be recorded as the slope of a linear regression of its phenotypes across environments on an environmental index (mean over all genotypes in each environment) (Finlay and Wilkinson, 1963). Recently, a Bayesian Finlay-Wilkinson (FW) regression method was proposed, allowing simultaneous estimation of environmental and genotypic parameters and the incorporation of genomic information as kinship matrix to better estimate the plasticity index (Lian and de los Campos, 2016). Investigations of the genetic determinism of plasticity have been carried out for several traits related to grain production or quality, such as grain yield or protein content in barley (Lacaze et al., 2009), and kernel weight, size or volume in maize (Kusmec et al., 2017; Li et al., 2019). Some loci were found to be associated with both the phenotype and its plasticity (Lacaze et al., 2009; Li et al., 2019), while others were specifically identified from plasticity indices (Kusmec et al., 2017).

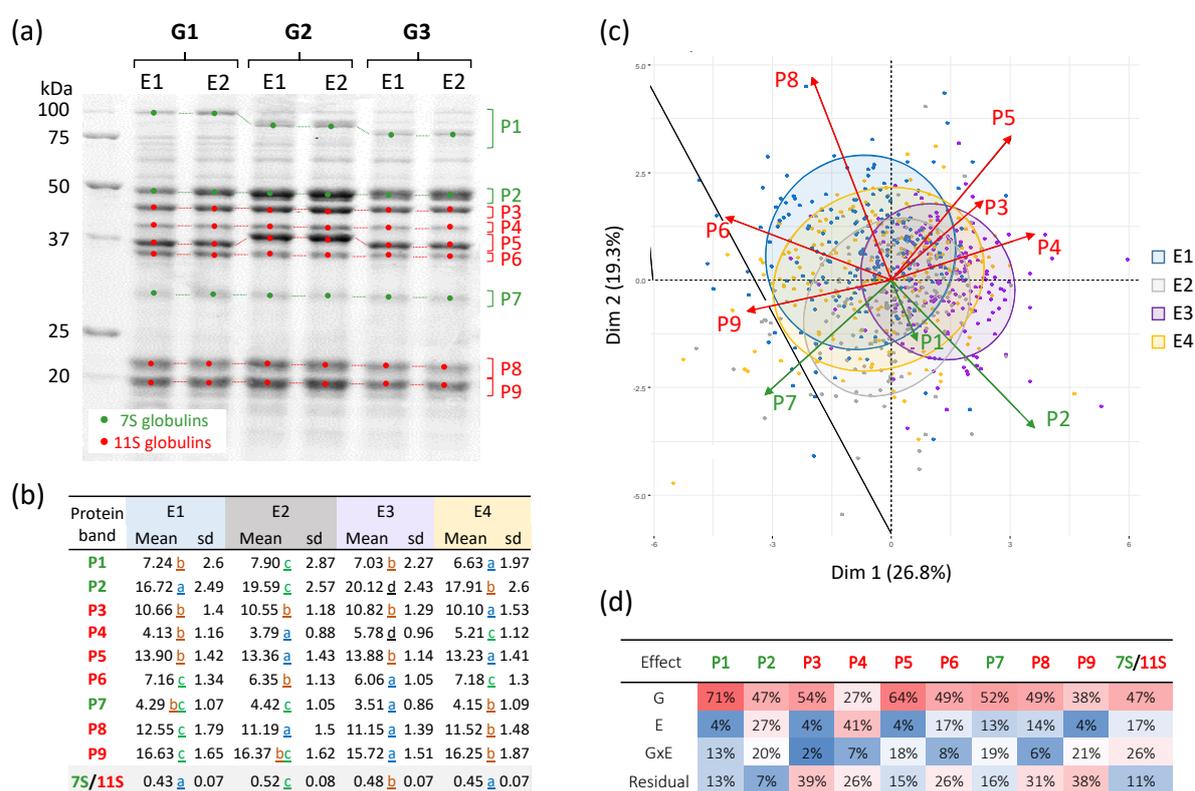
The present study investigated the genetic determinism of seed protein plasticity in response to the environment by exploiting the large genetic variability that exists in *Medicago truncatula* for seed globulin composition (Le Signor et al., 2017). Genome Wide Association Studies (GWAS) were conducted on proteotyping data of seeds from 200 ecotypes grown in four environments to identify genomic regions and genes controlling the dynamic plasticity of seed protein composition. Genes related to the interconnected sulfur and aspartate (Asp) family pathways were identified as candidates for governing this plasticity, providing new directions towards stabilizing the protein composition of legume seeds in a changing climate. One of these

directions relates to Met recycling that we demonstrate here to be a key control point of seed globulin composition and plasticity.

## 2. Results

### 2.1 Genotype and environment effects on seed protein composition in *M. truncatula*

Mass spectrometry data and previously obtained 1-DE maps (Le Signor et al., 2017) allowed annotating nine major protein bands, with their allelic variants, in the protein profile of *M. truncatula* seeds. Three of which mainly contained 7S globulins (P1, P2 and P7), the remaining six 11S globulins/legumins (P3, P4, P5, P6, P8 and P9) (Figure 18a and Table S2).



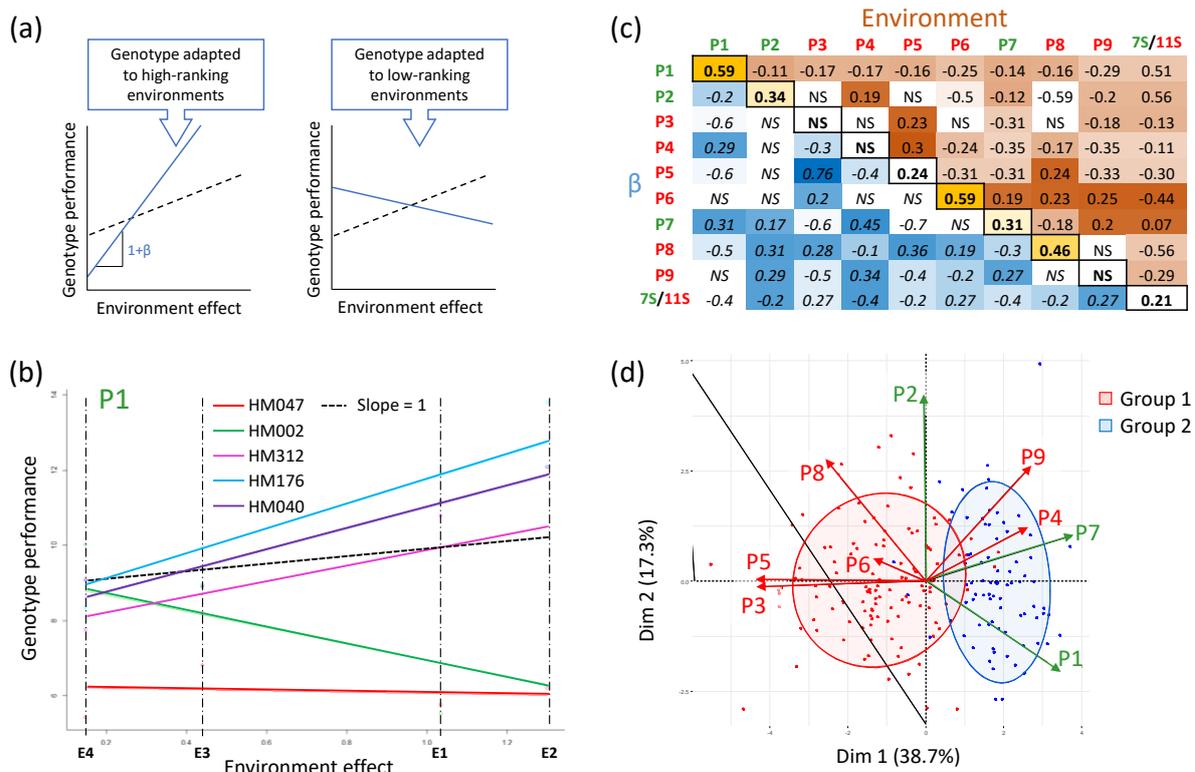
**Figure 18: Variability of seed protein composition in 200 *M. truncatula* ecotypes grown in four environments.** (a) example of seed protein profiles obtained by 1-DE for 3 genotypes in two environments. (b) Mean values and standard deviation (sd) of relative protein abundance per environment (Tukey's multiple comparison test is indicated with colored letters) (c) Principal component analysis of the 200 genotypes based on relative protein abundance for the nine bands (7S globulins, green arrows; 11S globulins, red arrows) grouped by environment and encircled by 75% confidence ellipse (E1, blue; E2, grey; E3, purple; E4, orange). (d) Relative percentage of total variance explained by G, E and GXE of a two-way random effects ANOVA (lmer function in R) calculated from the abundance of each of the nine protein bands and for the 7S/11S globulin ratio. G, genotype; E, environment.

The molecular weight and sequence of each 7S globulin allowed us to distinguish convicilins (P1) from vicilins (P2 & P7). The normalized quantitative values (*i.e* relative protein abundance) of each protein band for 200 *M. truncatula* genotypes grown in four environments (E1 to E4) were obtained. Details about variations in protein band abundance (mean and standard deviation) are provided in Figure 18b. PCA on mean genotype values for the nine 1-DE bands separated the individuals according to the relative abundance of 7S and 11S globulins along axis 2 (Figure 18c). An analysis of variance revealed significant genotype (G), environmental (E) as well as GXE interaction effects for each of the nine protein bands. The genotype effect accounted for on average 50% of total variance whereas environmental and GXE effects explained, respectively, 14% and 12% of total variance (Figure 18d). The highest percentage of variance due to a GXE effect (26%) was found for the 7S/11S ratio, reflecting a strong impact of GXE interaction on globulin composition.

## 2.2 Phenotypic plasticity of major seed proteins across environments

The FW method allowed to assess the plasticity of each genotype by comparing, for each band and for the 7S/11S globulin ratio, the range of changes across environments (Figure 19a).

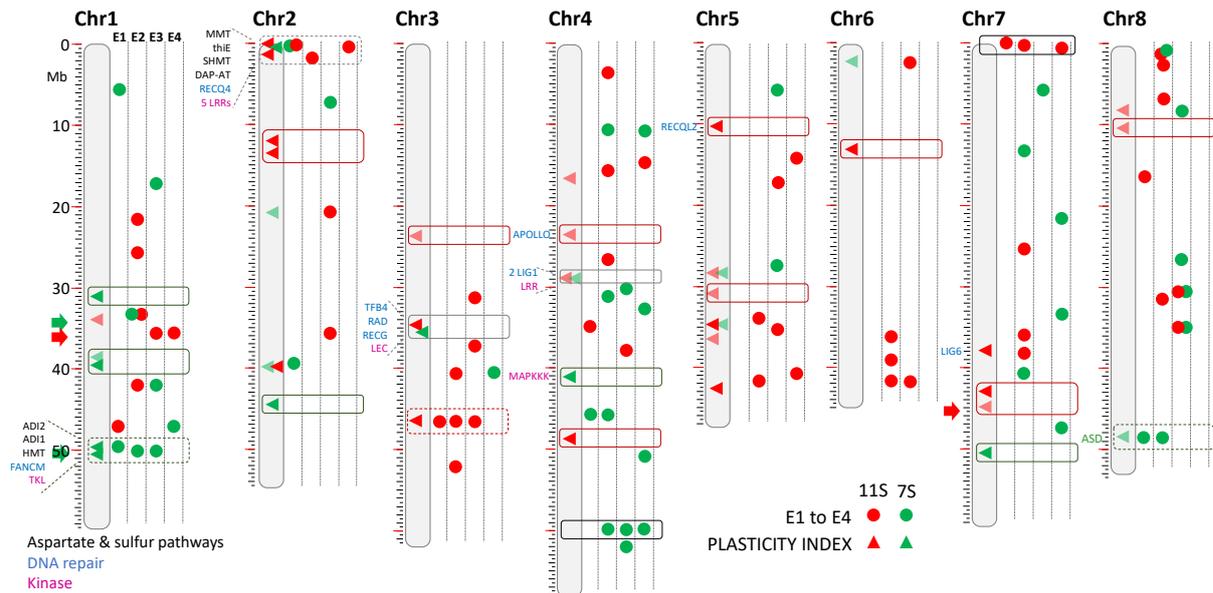
Taking protein band P1 as an example, a plot of regression values of genotype performance *vs* environmental effects for five genotypes (Figure 19b) highlighted large variation in the slope of the line (corresponding to  $1+\beta$ ); HM047 being stable across all environments, HM002 performing best in low ranking environments E3 and E4, whereas the three other genotypes performed best in higher ranking environments E1 and E2. The mean correlations between protein band abundance across environments (Figure 19c, upper triangular matrix) differed from correlation values between plasticity indices (Figure 19c, lower triangular matrix). Moreover, correlations between protein abundance and the plasticity index (Figure 19c, diagonal values of matrix) were low or not significant for most of the bands, suggesting the existence of plasticity-specific genetic determinants. PCA from the plasticity index of the nine protein bands separated well the individuals according to the genetic structure of the population (Figure 19d) along axis 1, which explains 38.7% of the variation. The two genetic groups were characterized by accessions originating from different geographical distributions: Mediterranean west for Group 1 *vs* Mediterranean center/east for Group 2 (Ronfort et al., 2006). This suggests that plasticity of seed protein composition is likely to be linked to evolutionary adaptation to environmental fluctuations. Consequently, integrating the population structure in the GWAS model appeared here essential to reveal the genetic determinism of seed protein plasticity.



**Figure 19: Plasticity of relative globulin abundance across environments.** (a) Schematic representation of the  $\beta$  meaning (modified from Finlay & Wilkinson, 1963), the dashed line corresponds to the behaviour of an average genotype ( $\beta+1 = 1$ ), in blue the behaviour of a given genotype specifically adapted to favourable environments (left plot) and a genotype specifically adapted to unfavourable environments (right plot). The slope of the regression line corresponds to  $1+\beta$ . (b) FW regression for the relative content of the P1 protein band displayed for five genotypes. The horizontal axis displays the estimated environmental effect and the vertical axis shows the relative content of P1 for each genotype. (c) Correlation matrix between plasticity indices of protein bands (lower triangular), between protein bands (upper triangular) and between protein bands and indices (diagonal) (NS, non-significant correlations,  $P>0.05$ ). (d) Principal component analysis showing distributions of the 200 genotypes according to their  $\beta$  coefficients for the nine protein bands (7S globulins, green; 11S globulins, red). Repartition of the genotypes is colored according to their genetic group based on SNPs (group 1, red; group 2, blue; see procedure for GWAS section in Materials & Methods).

## 2.3 Identifying loci and candidate genes for seed globulin plasticity

Protein band abundancies in the four environments and their plasticity indices were used to identify, by GWAS, genomic regions potentially involved in their genetic variability and/or plasticity. The plasticity index of the 7S/11S ratio did not yield results due to a bimodal distribution which did not fit the linear GWAS model. Distribution of the significant SNPs along the eight chromosomes of *M. truncatula* highlighted hotspot regions associated with variations in protein band abundance in one or more environments as well as hotspots associated with plasticity indices (Figure 20).



**Figure 20: Position of GWAS hotspots on the v5 physical map (in Mbp) of *Medicago truncatula*.** The hotspots identified by GWAS for protein composition (green: 7S globulins; red: 11S globulins) in the four environments (E1 to E4, a dot corresponds to more than 20 selected SNPs, as described in the Materials and Methods, in a 500 kbp window) and for protein composition plasticity (light colored triangle:  $\geq 5$  SNPs in a 500 kbp window; dark colored triangle:  $\geq 10$  SNPs in a 500 kbp window). Encircled hotspots are specific to plasticity or specific to environments (minimum three environments). The dashed lines encircled the hotspots identified in both plasticity and per environment analyses. Genes from GO enrichment analysis (kinases in black, genes related to Asp and sulfur metabolisms in blue or DNA repair processes in purple, Table 3) that colocalized with hotspots for plasticity index are indicated (abbreviations in Table S4, ‘Gene Name’ column).

The large dispersion of hotspots throughout the genetic map reflects a complex genetic determinism of seed protein composition, which is likely to be mainly dependent on the environment since only five hotspots were shared between at least three environments. As expected, many hotspots linked to seed protein composition in one or several environments colocalized with those previously identified in *M. truncatula* (Le Signor et al., 2017). Significantly, plasticity-specific hotspots were identified for 7S globulins (5 hotspots), 11S globulins (9 hotspots), and for both 7S and 11S globulins (2 hotspots, chromosomes 3 and 4). This study also highlighted hotspots for both plasticity and protein band abundance in one or several environments as exemplified by a region close to 50Mb on chromosome 1 or a region at the top of chromosome 2.

Next, we compared the lists of genes retrieved by GWAS for each environment (E1: 507 genes, E2: 645 genes, E3: 552 genes and E4: 545 genes; ontology enrichment by environment available in Table S3). Of these, 240 genes were identified in at least two environments, only 10 were common to three environments, and none was common to the four environments. The low number of genes shared between environments shows the complexity of the genetic bases of seed globulin plasticity. A total of 781 SNP were retrieved that corresponded to 556 genes

that we refer to as plasticity-linked candidate genes (PG, listed in Table S4). Of these PG, 392 were only identified from plasticity indices (*i.e.* not found in gene lists by environment) which we therefore refer to as plasticity-specific candidate genes (PSG). Inspection of allelic variants in the PG and PSG lists revealed 41 and 26 genes, respectively, bearing SNPs in the coding sequence leading to an amino acid change or a premature stop codon (Table S5).

## 2.4 Identification of biological processes that are likely to contribute to seed globulin plasticity

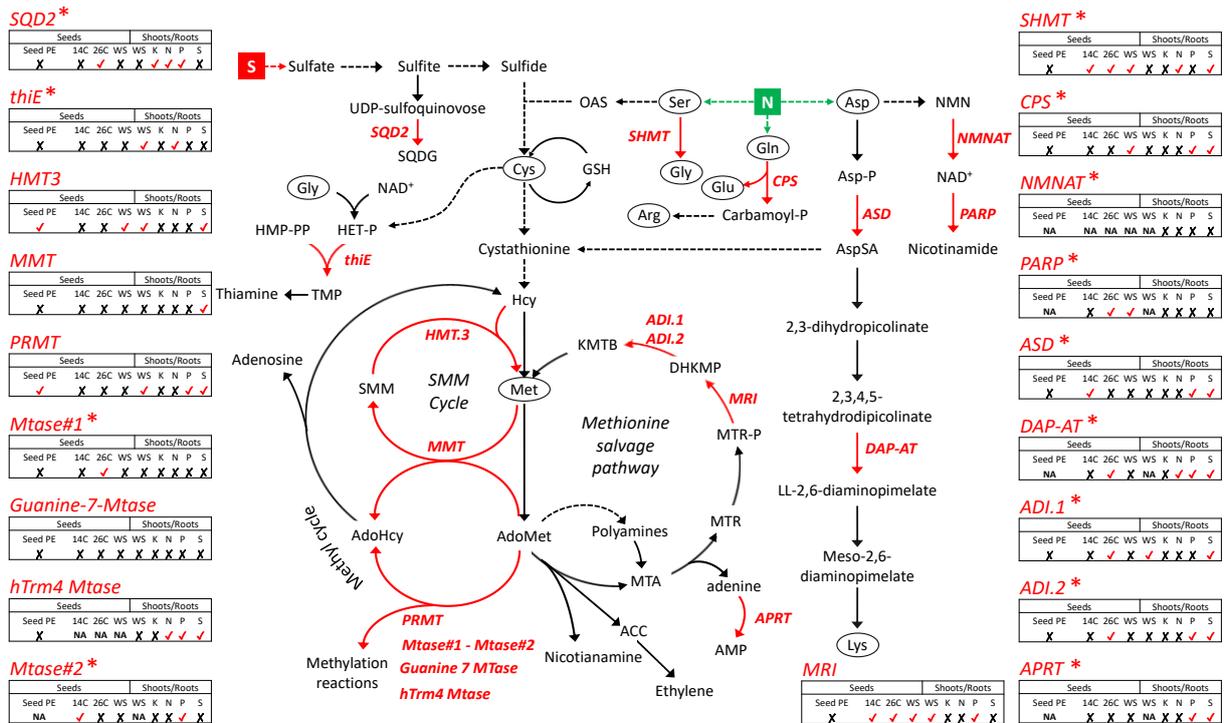
To provide insights into biological processes that might contribute to globulin plasticity, we performed a gene ontology (GO) enrichment analysis from the PG list. This revealed six GO terms over-represented in this dataset (Table 3).

**Table 3: GO terms enriched in the list of genes (556) identified by GWAS for globulin plasticity.**

GO group	GO.ID	Term	Significant/ annotated	Expected	Classic
I	GO:0007178	Transmembrane receptor protein serine/threonine kinase signaling	13/274	3.57	7.00E-05
II	GO:0009067	Aspartate family amino acid biosynthetic process	5/55	0.72	0.00075
III	GO:0071407	Cellular response to organic cyclic compound	5/59	0.77	0.00103
IV	GO:0044272	Sulfur compound biosynthetic process	7/132	1.72	0.00177
V	GO:0006555	Methionine metabolic process	5/74	0.97	0.00283
VI	GO:0006281	DNA repair	11/324	4.23	0.00372

For each GO term, the number of genes with the GO term (significant), the number of annotated genes in the genome (annotated), the expected number of genes, and the associated *P*-value (Fisher's test, threshold of  $P < 0.005$ ), are given.

Among the most represented were transmembrane receptor protein serine/threonine kinase signaling pathway (GO I, 13 genes) and DNA repair (GO VI, 11 genes). Interestingly, three kinases and seven DNA repair genes, including helicases and ligases, were in plasticity-specific hotspots (Figure 20). The three other most represented GO terms (II, IV, and V) were Asp family pathway, sulfur and Met metabolisms. A total of 19 genes related to these metabolisms were identified in the PG list (13 in PSG list) (Table S6). By replacing these genes in their corresponding pathways, we could highlight specific pathways as being potentially involved in globulin plasticity (Figure 21).



**Figure 21: Plasticity-linked genes for sulfur and nitrogen metabolisms mapped to their respective pathways.** Metabolic pathways connecting nitrogen metabolism to sulfur metabolism are shown. The sulfur amino acid Met can be synthesized *de novo* from cysteine or recycled through the Yang (Met salvage pathway), SMM and methyl cycles. Asp is not only the precursor for Met synthesis but also for lysine synthesis. These pathways, and those leading to the accumulation of defense compounds such as thiamine, nicotianamine, and polyamines, are shown. Red arrows indicate that the corresponding gene was identified by GWAS for globulin plasticity and asterisk indicate that the gene was identified from plasticity indices only (PSG list). For each gene, red ticks indicate a preferential expression in seeds (seed PE), a regulation in response to abiotic stresses in seeds (at 14°C, 26°C or in response to water stress, WS) and/or in shoots or roots (WS, nutrient deficiencies). **Metabolites.** AdoHcy: S-Adenosyl-l-homocysteine, AdoMet: S-adenosylmethionine, Arg: arginine, AspSA: aspartate semialdehyde, Asp-P: L-4-aspartyl phosphate, Cys: Cysteine, DHKMP: 1,2-dihydro-3-keto-5-methylthiopentene, Gln: glutamine, Glu: glutamate, G(5')pppPur: a 5'-(5'-triphosphoguanosine), GSH: Glutathione, HET-P: 4-methyl-5-β-hydroxyethylthiazole, HMP-PP: 4-amino-2-methyl-5-hydroxymethylpyrimidine diphosphate, Met: methionine, 5,10-Methylene-THF: 5,10-methylenetetrahydrofolate, m7G(5')pppR: a 5'-(N7-methyl 5'-triphosphoguanosine), MTA: 5'-methylthioadenosine, MTR: 5'-methylthioribose, MTR-P: 5'-methylthioribose-1-phosphate, NAD<sup>+</sup>: Nicotinamide adenine dinucleotide, NMN: Nicotinamide mononucleotide, OAS: O-acetylserine, Ser: Serine, SMM: S-methyl-l-methionine, KMTB : 2-keto-4-(thiomethyl) butyrate. **Enzymes.** ADI, acireductone dioxygenase, APRT: adenine phosphoribosyltransferase, ARD: acidoreductone dioxygenase, ASD: aspartate-semialdehyde dehydrogenase, CPS: carbamoyl-phosphate synthase (glutamine-hydrolysing), DAP-AT: LL-diaminopimelate aminotransferase, Guanine-7-MTase: guanine-7-methyltransferase, HMT: homocysteine S-methyltransferase, hTrm4 MTase: tRNA (cytosine34-C5)-methyltransferase, MMT: methionine S-methyltransferase, MTase: methyltransferase, MRI: methylthioribose-1-phosphate isomerase, MTI: 5'-methylthioribose-1-phosphate isomerase, NMNAT: NMN adenylyltransferase, PARP: poly(ADP-ribose) polymerase, PRMT: protein arginine methyltransferase, SHMT: glycine hydroxymethyltransferase; serine aldolase, SQDG: Sulfoquinovosyldiacylglycerol, SQD2: sulfoquinovosyltransferase 2, thiE: thiamine phosphate synthase.

In particular, 11 PG in Figure 21 were related to the recycling of Met. Five encoded S-adenosylmethionine (AdoMet)-dependent methyltransferases that act in the methyl cycle regenerating homocysteine to yield Met. Four encode proteins related to the Met salvage pathway, which synthesizes Met from 5'-methylthioadenosine (MTA) released as a by-product during the synthesis of ethylene, polyamines or nicotianamine (Sauter et al., 2013). These four genes corresponded to the housekeeping enzyme adenine phosphoribosyltransferase (APRT)

that converts adenine in AMP, methylthioribose-1-phosphate isomerase (MRI) that synthesizes 1,2-dihydro-3-keto-5-methylthiopentene (DHKMP), and acireductone dioxygenases (ADI.1 & ADI.2) that synthesizes 2-keto-4-(thiomethyl) butyrate (KMTB), the keto-acid precursor of Met, from DHKMP. The remaining two encode enzymes of the S-methylmethionine (SMM) cycle. One corresponded to methionine S-methyltransferase (MMT, Chr2g0276741) that synthesizes SMM, and the other (Chr1g0205081) to homocysteine S-methyltransferase (*HMT3*, Arabidopsis homolog: At3g22740), which converts SMM into Met (Bourgis et al., 1999). By studying the chromosomal position of these genes, we observed that *HMT3* and *MMT* perfectly colocalized with GWAS hotspots for both plasticity indices and globulin abundance in several environments (Figure 20). Significantly, *MMT* includes one SNP (1459C>A) inducing an amino acid change (Gln487Lys) (Table S5). These results point to a major role of Met recycling in controlling both globulin composition and plasticity.

## 2.5 Abiotic stress regulation of candidate genes for globulin plasticity

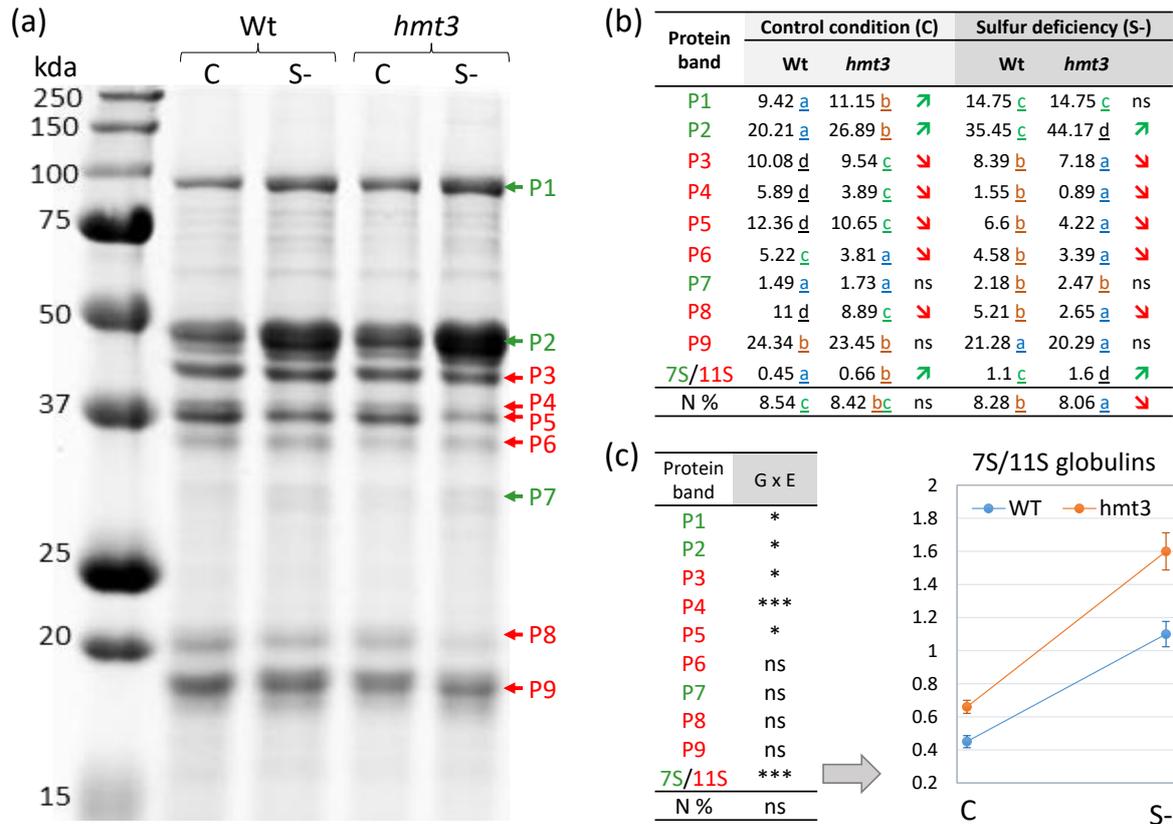
To study the regulation of the PG or PSG in seeds in response to abiotic stresses, we have used the transcriptomics data previously obtained from *M. truncatula* seeds collected during early (stage I), mid (stage II), and late (stage III, abscission and maturity) stages of seed development under water stress, high temperature (26°C), or low temperature (14°C) (Righetti et al., 2015). This revealed that both PG and PSG lists were significantly enriched in genes regulated by water stress (17% for PG; 19% for PSG) at the end of seed development (*i.e.* abscission stage) (Table S7). Regarding sulfur, Met and Asp metabolisms depicted in Figure 21, 12 genes related to these pathways were regulated in seeds in response to stresses, especially water stress and/or high temperature (Table S6). Because plasticity-linked genes might not necessarily exert their function within the seed, we next investigated their regulation in response to abiotic stresses in shoots and/or roots of *M. truncatula* subjected to water stress (Zhang et al., 2014) or nutrient deficiencies (N, P, K and S) (de Bang et al., 2017). Both PG and PSG lists were significantly enriched in genes regulated by nitrogen (~20%) or sulfur (~30%) deficiencies, and the PSG list was also enriched in genes regulated by water stress (49%). Of the 19 genes related to sulfur, Met, and Asp metabolisms, 15 were differentially expressed in response to water stress or nutrient deficiencies, especially sulfur deficiency (Figure 21 and Table S6).

## 2.6 *HMT3* is preferentially expressed in seeds and contributes to globulin plasticity

We next searched for PG preferentially expressed in seeds using the *M. truncatula* gene expression atlas (Benedito et al., 2008). Interestingly, the PSG dataset was significantly

enriched in genes preferentially expressed in seeds (16%, Table S7). Of these 37 PSG, listed in Table S8, several are good candidates to play a role in globulin synthesis or processing since they were annotated as transcription factors (six genes) or peptidases. It was noteworthy that 12 PG that were preferentially expressed in seeds, including eight PSG, were also regulated by abiotic stresses in seeds (Table S8), thus highlighting stress-responsive genes that might act in seeds to control globulin plasticity. Five of them were preferentially expressed during storage protein synthesis, occurring from 16 to 24 days after pollination (Benedito et al., 2008), suggesting they are excellent candidates for contributing to globulin plasticity within the seed. One of these (Chr3g0114701) encodes a protein with an EamA (named after the O-acetylserine/cysteine export gene in *E. coli*) domain. Its best Arabidopsis homolog, AT4G08290, is a UmamiT (Usually Multiple Acids Move In and out Transporter) facilitator controlling amino acid export in developing Arabidopsis seeds (Müller et al., 2015). Another PG corresponds to *HMT3* that is a good candidate for regenerating Met within the seed for seed globulin accumulation and plasticity.

To investigate the potential relationship between *HMT3* and these traits, plants derived from seeds of three *Tnt1* insertion lines for *HMT3* were genotyped and homozygous *hmt3* mutants could be recovered from seeds of the NF7660 insertion line. To modulate seed globulin composition, we applied sulfur deficiency (*versus* sulfur sufficiency) that leads to major changes in the 7S/11S globulin ratio in *M. truncatula* seeds (Zuber et al., 2013). Major proteins in wild type and *hmt3* seeds were quantified after 1-DE (Figure 22a). As expected, the relative abundance of the sulfur-rich globulins (11S) decreased in response to sulfur deficiency in both wild-type and *hmt3* seeds (Figure 22b). In addition, *hmt3* seeds accumulated less 11S globulins than wild-type seeds under both control and sulfur-deficient conditions. The higher accumulation of sulfur-poor globulins (7S) in *hmt3* seeds compared to wild-type was sufficient to maintain nitrogen content in the control condition but not under sulfur deficiency (Figure 22b). Significant GXE effects were also identified for five major protein bands and for the 7S/11S globulin ratio, which displayed a more plastic response in *hmt3* seeds compared to wild-type seeds (Figure 22c).



**Figure 22: Globulin composition and plasticity in *hmt3* seeds compared to wild-type seeds.** (a) Typical 1-DE profile obtained from *M. truncatula* seeds of the wild type and *hmt3* mutant grown in control (C, sulfur sufficiency) and (S-) sulfur-deficient conditions (7S globulins, green; 11S globulins, red). (b) This table shows the mean relative abundance of each protein band in each condition, and of the 7S to 11S ratio and seed nitrogen content. Different letters indicate significant variations (Tukey test, N = 8). (c) G x E interaction significance (\*, P-value < 0.05; \*\*\*, P-value < 0.001) are shown along with differences in the plasticity of the 7S to 11S ratio between the wild-type and *hmt3* mutant.

## 2.7 Candidate regulators of seed protein plasticity

To identify transcriptional regulators of seed protein plasticity, all transcription factors (TF) and chromatin remodeling elements were identified in the PG and PSG lists (Table S9). An enrichment test, considering all genes of the Mt5 genome annotated as TF and chromatin remodelers, revealed that both PG and PSG lists were enriched in regulators (15% and 16% respectively, Table S7). The most highly represented TF families in the PG dataset were *mTERF* (Mitochondrial Transcription Termination Factor) and *AP2-EREBP* (Apetala2-Ethylene-Responsive Element Binding Proteins) (Annexe 2). An inspection of the expression profile of these genes across plant tissues enabled the identification of six genes preferentially expressed in seeds (Table S9), of which one (Chr1g0190851) encodes a GTL2-like TF whose expression increased during seed filling up to late maturation, indicating a concomitant role of this TF with storage protein synthesis and deposition. It is worth noting that most of the regulators in Table

S9 were not preferentially expressed in seeds (76/82), suggesting that regulatory events at the level of the mother plant might contribute to seed globulin plasticity. To get insight into their response to abiotic stresses, we used the transcriptomics data from (Zhang et al., 2014) and (de Bang et al., 2017). This revealed 22 regulatory genes whose expression is induced by drought in shoots and/or roots (Table S9). Of the five genes up-regulated by drought in both tissues was a homeobox-WOX (Chr8g0346251), whose best ortholog in Arabidopsis, HB-7 (At2g46680), is regulated in an ABA-dependent manner and may have a role in signal transduction pathways, mediating drought response (Wang et al., 2015).

### 3. Discussion

Increasing climate fluctuations, such as drought, combined with the need to increase the use of plant proteins in human nutrition underline the importance of identifying genes that can be targeted to stabilize the protein composition of legume seeds or to optimize it in a particular environment. Using a core-collection of 200 *M. truncatula* ecotypes grown under four environmental conditions, we first demonstrated that there was a high genetic variability for seed globulin plasticity, scored using the FW method, and then, we investigated its genetic determinism by GWAS. This is the first time to our knowledge that GWAS has been applied to identify loci determining the plasticity of seed protein composition. The results showed that some regions controlling seed globulin composition in one or several environments also control globulin plasticity (Figure 20). However, other regions or genes were only identified from plasticity indices (PSG), suggesting the existence of plasticity-specific determinants. Here, we provide details about all genes identified from globulin plasticity indices (PG), including the number of significant SNPs and publicly available transcriptomics data describing their regulation during seed development or in response to abiotic stresses (Table S4). Some of the metabolic pathways or genes that are relevant candidates for controlling globulin plasticity are discussed hereafter.

#### 3.1 Met recycling is a key contributor to globulin composition establishment and plasticity

The PG list was significantly enriched in GO terms related to the sulfur and Asp amino acid family pathways (Table 3), both leading to the synthesis of the nutritionally essential amino acid Met. These GO terms were not enriched in gene lists for globulin composition per environment (Table S3), and of the 19 PG related to these metabolisms, 13 were only identified from plasticity indices (Figure 21). This implies a prominent role for these pathways in globulin

plasticity. By inspecting the expression of these PG across plant tissues and in response to stresses, we noticed that most of them (17/19) were not preferentially expressed in seeds but regulated by abiotic stresses in developing seeds, shoots or roots (15/19 genes in Figure 21). There have been significant advances towards understanding the role of the sulfur and Asp family pathways in plant stress responses (Batista-Silva et al., 2019; Galili, 2011; Khan et al., 2008). However, their relationship with seed protein plasticity has not previously been reported. One PSG encodes a key enzyme of the Asp family pathway, Asp-semialdehyde dehydrogenase (ASD), which produces Asp semialdehyde (AspSA) acting at the branch point between lysine and methionine metabolisms (Stéphane Ravanel et al., 1998). The identification of another PSG as corresponding to LL-diaminopimelate aminotransferase (DAP-AT), which catalyzes a subsequent step of lysine biosynthesis, suggests that regulation of this biosynthetic pathway may influence globulin plasticity. Intriguingly, there were no PG related to *de novo* Met biosynthesis. Instead, we found 11 PG in pathways regenerating Met from other metabolites, indicating that the plasticity of seed proteins towards environmental changes may rely on regulation of Met recycling rather than on its *de novo* biosynthesis. Four PG were related to the Met salvage pathway, also called the Yang cycle (Yang and Hoffman, 1984) which regenerates Met from 5'-methylthioadenosine (*MTA*) produced during the synthesis of ethylene and polyamines, two hormones well known for their role in the plant response to environmental stimuli (Booker and DeLong, 2015; Gill and Tuteja, 2010). Genes of the Met salvage pathway were up-regulated in response to water stress in roots (*MRI* and *ADI.1*), shoots (*MRI*) and late developing seeds (*MRI*) (Table S6), supporting a role for this salvage pathway in different tissues under stress conditions. The recycling of Met *via* this pathway can prevent the depletion of AdoMet (Rzewuski et al., 2007), which is a methyl donor for numerous methylation reactions and the precursor for the synthesis of ethylene and polyamines. It can also allow for the synthesis of SMM, which plays a major role in the phloem transport of sulfur from leaves to seeds (Bourgis et al., 1999; Tan et al., 2010). Interestingly, two PG corresponded to *MMT* (*At5g49810* homolog), which converts Met and AdoMet into SMM, and *HMT3*, which releases Met from SMM and homocysteine. In particular, *HMT3* was found to be highly expressed in seeds of *Arabidopsis* (Cohen et al., 2017) and *M. truncatula* (Figure 21 and table S4), suggesting it to be a good candidate for converting SMM into Met within the seed, as previously proposed (Bourgis et al., 1999; Gallardo et al., 2007). Taken altogether, these data suggest that sulfur supply to seeds for globulin plasticity relies on the capacity of the plant to recycle Met from other tissues in response to environmental stimuli. They also raised the question as to which extent seed sulfur loading in the form of SMM could contribute to globulin plasticity.

By characterizing seed protein composition of a *Tnt1*- insertion mutant of *M. truncatula* for *HMT3* (Figure 22), we showed that *HMT3* is a key contributor towards sulfur-rich globulin accumulation in legume seeds, the 7S/11S ratio being higher in *hmt3* seeds. These observations complement well (Cohen et al., 2017) who showed an increase of protein-bound Met in *Arabidopsis* mutant seeds overexpressing *HMT3*. Moreover, by studying GXE effects, we showed that *HMT3* acts as a plasticity gene by contributing to stabilizing the 7S to 11S ratio under contrasting sulfur supply (Figure 22c). These data not only validated our LD-based post-GWAS method used for gene selection, but also demonstrated that Met recycling in the form of SMM is involved in globulin composition and in its plastic response to sulfur deficiency.

### 3.2 Regulators were identified that might mitigate the effect of environmental stresses on seed globulin composition

The enrichment of the PG dataset in genes for transmembrane receptor protein serine/threonine kinases (Table 3) and transcription factors (Table S7) demonstrated the importance of regulatory processes in responses to environmental variations. Genes regulating transcription and/or signal transduction are key members of the gene regulatory network underlying plastic responses of plants (Nicotra et al., 2010; Shao et al., 2007), but their relationship with seed globulin plasticity has remained unexplored until now. Among the 13 kinases identified here, six belong to the large group of LRR-RLKs with roles in plant development and stress responses, and three belong to the Lectin receptor kinase family generally described to have roles in plant immunity (Liu et al., 2017; Wang and Bouwmeester, 2017). In the PG list, the kinase gene which was the most responsive to abiotic stresses is a NPK1-like MAPKKK (Chr5g0430501, Table S4). When constitutively expressed, NPK1 enhances tolerance to multiple stresses in transgenic tobacco, acting as a key regulator of an oxidative stress signaling pathway, and is thus considered as a good candidate for engineering multiple stress tolerance in plants (Kovtun et al., 2000). Because none of the kinases in the PG list was preferentially expressed in seeds, we propose that these regulatory proteins may indirectly influence seed globulin plasticity by increasing the plant's ability to cope with adverse environmental conditions. This is also likely to be the case for the majority of the PG-TF, which were not preferentially expressed in seeds. Interestingly, in the *AP2/EREBP* gene family overrepresented in the PG data set (Annexe 2), five genes carry a dehydration-responsive element binding (DREB) domain (Chr6g0473741, Chr6g0473941, Chr6g0473961, Chr6g0473971, Chr7g0257911) known to be involved in the regulation of drought and cold responses (Wang et al., 2003). The second TF family overrepresented in the PG dataset (Annexe 2), the mTERF

family, has more recently been characterized in plants and shown to be required for accurate expression of organelle genomes (Quesada, 2016). It is worth noting that loss of function Arabidopsis mutants for two *mTERF* genes (*SHOT1* (Kim et al., 2012) and *SOLDAT10* (Meskauskiene et al., 2009)) provide thermotolerance and light stress acclimation, respectively, in *A. thaliana*. Authors considered them as good targets for improving tolerance to abiotic stress in crops.

Of the six TFs from the PG list that were preferentially expressed in seeds, one (Chr1g0190851) merits attention since it is expressed at seed filling stages, suggesting it to be a candidate for playing a direct role in seed globulin plasticity. Its closest Arabidopsis homolog, AtGT2L (At5g28300), as other GT factors from Arabidopsis or soybean, is involved in plant response to abiotic stresses (Xi et al., 2012), but its role during seed development has not been investigated yet. The finding that one GT factor recognizes the GT-box present in the promoter region of the seed storage protein 2S3 and thus compete with the binding of transcriptional activators (Gao et al., 2009) further suggests that GT factors could directly influence seed protein plasticity.

### 3.3 The maintenance of DNA integrity is likely to play a role in plasticity of seed protein accumulation

DNA repair might contribute to seed globulin plasticity owing to the overrepresentation of this biological process in the PG dataset (Table 3). Nine DNA repair-related genes, out of 11, were specifically identified from plasticity indices and six of them were deregulated by abiotic stresses in seeds and/or other plant tissues (Table S4). DNA damages could occur during seed development, storage period and germination since seed dehydration and rehydration phases in combination with prolonged periods in a dry quiescent state are associated with high levels of oxidative stress (Waterworth et al., 2019). Environmental factors, like drought or ionizing radiation, are also known to challenge DNA integrity. Role of these genes during germination and in seed longevity is well documented (Waterworth et al., 2019). For instance, three ligases identified as PG (Chr7g0251371, Chr4g0028151, Chr4g0028151) share homologies with AtLIG6, which is an important determinant of seed longevity in Arabidopsis (Waterworth et al., 2010). Maximum expression in seeds of DNA repair PG are late embryogenesis and end of seed filling (Table S4). These stages, which respectively mark the switch towards seed filling and dehydration, were emphasized by (Parreira et al., 2018) as requiring a tight control of DNA integrity. The DNA repair-related PG identified in the present study are thus good candidates

to control genome integrity at these key stages determining the onset and end of storage protein synthesis.

In conclusion, we provide here a unique resource of genes associated with the plastic response of seed proteins to environmental changes that can be targeted for improving and stabilizing the nutritional value of legume seeds. In particular, genes with roles in Met metabolism, DNA repair, transcription and signal transduction were enriched in the dataset, thus highlighting these processes as important contributors to globulin plasticity. There was no preferential expression in seeds for a large proportion of plasticity-linked genes, many of them being regulated by abiotic stresses in other tissues. This indicates that seed protein plasticity mainly relies on the capacity of the plant to combat abiotic stresses, and presumably maintain nutrient uptake, remobilization, and loading in seeds. Our results revealed the contribution of Met recycling to this plastic response, and suggest that SMM is likely to be a major transport form of sulfur mitigating the effect of environmental changes on the accumulation of sulfur-rich seed proteins. These results advance our understanding of the complex processes involved in the plastic response of seed globulins to environmental changes and pave the way for future researches especially in relation to methionine recycling pathways.

## 4. Materials and Methods

### 4.1 Seed production for the ecotype collection in *M. truncatula*

A population of 200 ecotypes from the *M. truncatula* HapMap project ([www.medicagohapmap.org](http://www.medicagohapmap.org)) was used in this study (Table S1). Seeds were obtained in insect-proof glasshouses at IRHS (Angers, France) or AGAP (Montpellier, France) under four environments (E): **E1**: standard growing condition (IRHS), sowing in November, 3L pots (1 plant per pot) with a mixture of soil and sand (70:30%), 20°C day/18°C night, 16h photoperiod, (mean flowering date: March 3, 2016; harvest in June); **E2**: same conditions as E1 except that water stress was applied from flowering to maturity by maintaining a soil water potential at  $-0.1$  MPa; **E3**: winter sowing in November (AGAP) in insect-proof glasshouses, which were neither heated nor lit, in 0.5 m<sup>3</sup> box pallets filled with a mixture of soil and sand, 5 plants/box pallet (mean flowering date: March 2, 2016); **E4**: sowing in February (AGAP) in 5L pots (1 plant per pot) with the same substrate as E3 (mean flowering date: Apr. 20, 2016), reduced duration of the seed filling period compared to E3. Mature seeds from E3 and E4 were harvested at the same period (Jul. 2016). All plants from E1 to E4 were drip-irrigated with water enriched with a nitrate containing solution. In each environment, the accessions were grown in complete

random design in triplicate and the seeds from the three replicates were pooled by environment (~800 seeds lots were analyzed).

## 4.2 Seed protein separation and quantification

For each accession and environment between 30 and 50 seeds were crushed in liquid nitrogen. Storage proteins were extracted from the seed powder in a Tris-HCl lysis buffer (Schroeder, 1982). For each sample, 2 $\mu$ L of crude protein extracts mixed with 8  $\mu$ l of loading buffer (5X = TrisHCl 0.5M, SDS 4%, glycerol 20% and bromophenol blue 1mg) were separated by one-dimensional electrophoresis (1-DE) using the XCell4 *Surelock*<sup>TM</sup> Midi-Cell system (Life Technology) as described in (Henriet et al., 2019). The protein extracts (200 per environment) were processed in two batches, E1 & E2 and E3 & E4, respectively. For each batch and genotype, extracts from two environments were loaded in adjacent wells on the same gel. Experimental design included 20% of replicates (2 genotypes/gel). Five runs per batch were performed and preliminary ANOVA showed that run or tank effects were not significant for major bands (data not shown). Gels were stained with Coomassie Brilliant Blue R-250 (Mathesius et al., 2001) and scanned using the Odyssey infrared imaging system (LI-COR). Protein band detection and quantification were performed using the CLIQS software (<http://www.totallab.com>). The molecular weight of each protein band was calculated using a low-range protein ladder (Bio-Rad). In each well, the normalized volume of each band was calculated as the ratio of each band volume to total band volume after “rolling ball” background reduction (radius 300). The quantitative data (*i.e.* relative protein abundancies) were subjected to principal component analysis (PCA) and to an analysis of variance (ANOVA) followed by a Tukey test for multiple comparisons (R packages FactoMineR, factoextra, car).

## 4.3 Identification of mature seed proteins

Protein bands were annotated in comparison with previously established protein maps of mature seed proteins in *M. truncatula* (Le Signor et al., 2017) and/or from mass spectrometry data obtained for eight protein bands in the present study (Supplementary Table S2). Major protein bands were excised from the seed protein profile of five genotypes to identify protein variants (e.g. four variants of a convicilin, 7S globulin of 78 kDa, identified for P1). Protein digestion and peptide extraction were performed according to (Labas et al., 2015). The peptides were analyzed by nanoflow liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) on a LTQ Orbitrap Velos mass spectrometer coupled to an UltimateR 3000 RSLC Ultra High Pressure Liquid Chromatographer (ThermoFisher Scientific, Illkirch-Graffenstaden, France)

controlled by Chromeleon (v6.8 SR11; Dionex). Peptide separation and data acquisition were done according to (Labas et al., 2015). MS/MS ion searches were done using Mascot (v2.3.2, Matrix Science), with the acceptability for the parent ions set to 5 ppm and 0.8 Da for daughter-fragment ion matches. Peptide and protein identifications were validated using Scaffold (v4.8.3, Proteome Software) with the criteria detailed in Supplementary Table S2. Major proteins in each band were identified using the Normalized Weighted Spectra (NWS) count (Supplementary Table S2).

#### 4.4 Calculation of plasticity indices

For each protein band, a plasticity index was calculated using the bayesian FW regression method implemented in the R package (Lian and de los Campos, 2016; R Core Team, 2014). Bayesian FW method jointly estimates genotype and environmental parameters of the equation:  $y_{ijk} = \mu + g_i + h_j + \beta_i h_j + \varepsilon_{ijk}$  where  $y_{ijk}$  is the phenotypic record of the  $k^{\text{th}}$  replicate of the  $i^{\text{th}}$  variety observed in the  $j^{\text{th}}$  environment,  $g_i$  is the main effect of  $i^{\text{th}}$  variety,  $h_j$  is the main effect of the  $j^{\text{th}}$  environment, and  $\varepsilon_{ijk}$  is the error term. When the equation is reduced to  $\mu + g_i + (1 + \beta_i) h_j + \varepsilon_{ijk}$ , the term  $(1 + \beta_i)$  is the measure of a genotype's linear response to environment. All effects are considered as random and with normal distributions:  $h \sim (0, I\sigma^2_h)$ ,  $g \sim (0, K\sigma^2_g)$ ,  $\beta \sim (0, K\sigma^2_\beta)$ , where I is the identity matrix, and K the kinship matrix (marker-derived relationship matrix, same kinship matrix as for GWAS). In this method, the plasticity index is the slope ( $\beta$ ). For each genotype,  $\beta$  plasticity indices were calculated from the quantitative values of each of the nine protein bands in the four environments.

#### 4.5 Core collection genotyping data

The Single Nucleotide Polymorphism (SNP) data (obtained by Illumina sequencing) were from the *M. truncatula* HapMap project (Stanton-Geddes et al., 2013). Mt5.0-based data were downloaded from <https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/> (Pecrix et al., 2018). Raw genotypic data scored 16 515 723 SNPs that passed the first quality control filtering. These data were then filtered for our set of 200 genotypes with a minimum count of 84 non-missing data and a minimum allele frequency (MAF) threshold of 3%, excluding multiallelism. The imputation was done using the BEAGLE4.0 software as described (Browning and Browning, 2007) (except for the number of iterations = 10 and gprobs = T). The imputed file was filtered on the genotype probability scored by BEAGLE to be superior or equal to 0.7. The final genotype file of 5.626.623 SNP and 200 genotypes scored less than 1% of missing values.

## 4.6 Procedure for GWAS and gene selection

GWAS was performed using the Genome-wide Efficient Mixed Model Association (GEMMA) algorithm software v.0.94 (Zhou and Stephens, 2014). This algorithm tested through a standard linear mixed model for marker association with a single phenotype accounting for population structure (Q) and kinship matrix (K). K was calculated with all intergenic markers following the GEMMA procedure (centered relatedness matrix was computed). To infer population structure, we computed PCA on a random sample of 50000 intergenic SNPs with the R MVP package (Robin and Hankin, 2019), and kept the first PCA axis as (Q) so that the population separated in two groups according to (Ronfort et al., 2006). *P*-values for each SNP were calculated from a likelihood ratio test. The  $-\log_{10}P$ -values were exported to generate Manhattan and Q-Q plots using the MVP package of R. Manhattan, and Q-Q plots obtained for globulin plasticity are provided in Annexe 1. The *P*-values were adjusted using Benjamini-Hochberg (BH) multiple test (Benjamini and Hochberg, 1995). The position of SNPs inside or near the genes were found in context files downloaded from [medicago.toulouse.inra.fr/r5.0-ANR](https://medicago.toulouse.inra.fr/r5.0-ANR). To select candidate genes, a first list was made with the most significant SNPs (*P*-value lower than  $10^{-6}$  corresponding to a mean BH threshold of 0.1). Then all SNPs in linkage disequilibrium (LD) with the SNPs present in the list ( $r^2$  higher than 0.6 and *P*-value lower than  $10^{-5}$  with a maximum distance of 20 kb) were retained using LD-based clumping in plink v1.9. From this extended SNP list, candidate genes were retrieved at a physical distance of 5 kbp around each significant SNP. By using this method, the number of genes increased by 15% compared to the single BH  $10^{-6}$  threshold. Gene Ontology enrichment analyses were performed using the topGO R package 2.36.0. (Alexa A and Rahnenfuhrer J, 2016) and GO annotations available at <https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/>.

## 4.7 Growing and genotyping *hmt3* insertion mutants

To investigate the role of Homocysteine methyltransferase 3 (HMT3) in globulin plasticity, three Tnt1 insertion lines of the *M. truncatula* ecotype R108 (NF7660, NF16472 and NF1932) were obtained from Noble Research Institute in Oklahoma (<https://medicago-mutant.noble.org>), (Cheng et al., 2017). Plants from line NF1932 did not yield seeds and no Tnt insertion in *HMT3* was detected in plants from NF16472. Plants from NF7660 carried an insertion in exon 6 of *HMT3* (NF7660\_high5). The presence of the insertion was verified by PCR using primers: amplification of 355pb for wild-type *HMT3* allele with R108-HMT3-F1 (GTACTGCTCCTAGATTCATCCATGG) and R108-HMT3-R1 (CACCACCCATGTGTTGTTCTCAGC), and amplification of 700bp for *hmt3* allele with

TNT1F (ACAGTGCTACCTCCTCTGGATG) and R108-HMT3-R1 CACCACCCATGTGTTGTTCTCAGC). The insertion site was precisely mapped on the genomic sequence at nucleotide 2996 by Sanger sequencing of amplicon (TNT1F & R108-HMT3-R1). Seeds from the NF7660 line were sown in 1L pots containing a perlite/sand mixture (3/1, v/v) in two blocks with a randomized design, each block including 16 wild-type and 16 *hmt3* mutant plants (glasshouse at INRAE Dijon). Seedlings were irrigated with a nitrate and sulfur-containing solution, as described in Zuber et al., (2013). At flowering, half of the plants (one block) were subjected to sulfur deficiency until maturity as described (Zuber et al., 2013). At harvest, a sample of 50 mature seeds from each plant was used for protein extraction and 1-DE as described above. The quantitative data were analyzed by ANOVA followed by a Tukey test.

### **Acknowledgements**

We sincerely thank Franck Zenk and Julien Martinet from the greenhouse team of INRAE Dijon for monitoring *hmt3* plant growth during the confinement period due to COVID-19 pandemic. We thank Pascal Joly from AGAP, Laëticia Ly Vu, Daniel Sochard, Benoit Ly Vu and Daniel Beucher and the Phenotic platform of the IRHS for help with seed production. We thank Myriam Sanchez for characterizing the position of the Tnt1 insert in the *hmt3* mutant. The *M. truncatula* mutants utilized in this research project, which are jointly owned by the Centre National de la Recherche Scientifique, were obtained from Noble Research Institute and were created through research funded, in part, by the National Science Foundation (grant NSF-0703285). We sincerely thank Judith Burstin, Nathalie Nesi and Jérôme Verdier, members of the PhD committee of Kévin Cartelier, for fruitful discussions and helpful suggestions. We are grateful to Vanessa Vernoud and Richard Thompson for critical reading of the manuscript and helpful comments. This work was supported by grants from the French National Research Agency (REGULEG project, ANR-15-CE20-0001). The PhD fellowship of Kévin Cartelier was funded by the INRAE BAP department and the Regional Council of Burgundy.

## CHAPITRE III :

Etude de la variabilité de réponse à une carence en soufre au sein d'une collection de pois et du déterminisme génétique de cette réponse.

## **Objectifs**

---

Les travaux de ce chapitre s'inscrivent dans le cadre du projet H2020 EUCLEG (Coord. Bernadette Julier, INRAE Lusignan) dont l'une des vocations est d'améliorer la stabilité du rendement et la qualité protéique des graines de légumineuses. L'objectif de nos travaux conduits chez le pois était d'identifier les gènes sous-jacents à la plasticité de différents traits de composantes de rendement ainsi que de qualité des graines en réponse à une carence en soufre.

## **Stratégie**

---

La stratégie a consisté à utiliser la variabilité existante au sein du genre *Pisum* pour identifier par génétique d'association (GWAS) des variants alléliques corrélés à la variabilité des composantes de rendement et de qualité des graines. Une expérimentation a été conduite sur deux années en utilisant un grand nombre de génotypes issus des collections AMS (projet PeaMUST, Coord. Judith Burstin, équipe ECP) et GRaSP (Coord. Virginie Bourion, équipe ECP) : 304 génotypes en 2018 et 198 en 2019. L'effectif réduit en 2019 est néanmoins représentatif de la diversité présente dans les collections. Une carence en soufre a été appliquée à partir d'un stade mi végétatif à la moitié des plantes, l'autre moitié étant maintenue en condition de nutrition optimale.

## **Résultats majeurs de la partie centrée sur la composition protéique des graines**

---

Dans une première partie, nous avons analysé les variations de composition protéique des graines après séparation par électrophorèse monodimensionnelle des protéines majeures des graines de tous les génotypes récoltés. Ces données ont mené à l'acquisition de connaissances sur la diversité de la composition protéique des graines présente au sein de la collection en condition optimale et carencée en soufre. Une classification hiérarchique des génotypes a révélé des différences de réponse à la carence en fonction des génotypes. Considérant les deux années de production et les deux conditions de nutrition comme quatre environnements, nous avons calculé des indices de plasticité associés à chaque bande protéique, en utilisant la méthode de régression de Finlay and Wilkinson, (1963). La variabilité présente pour les bandes protéiques dans les différents environnements ainsi que pour leur plasticité vis-à-vis de la carence en soufre nous a permis de mener des analyses GWAS et d'établir une carte des principaux hotspots détectés que nous avons comparé aux précédentes études menées chez le pois. La liste de gènes liés aux SNP significatifs identifiés par GWAS s'est révélée enrichie en gènes associés au métabolisme du soufre, au transport des molécules ainsi qu'aux modifications post-

traductionnelles des protéines. De nombreux gènes candidats potentiellement impliqués dans la réponse de la composition protéique à la carence soufrée sont discutés.

### **Résultats majeurs de la partie centrée sur les composantes du rendement**

---

Dans une seconde partie, nous avons exploité les mesures des composantes de rendement acquises sur l'ensemble des plantes des deux expérimentations, comme le nombre et le poids des gousses et des graines, le poids d'une graine et d'une gousse ainsi que le nombre de graines par gousse. Nous avons analysé l'impact de la carence en soufre sur ces variables ainsi que les différences de comportement des génotypes, sur la base de ces variables, en réponse à la carence. Nous avons calculé, de la même façon que précédemment, les indices de plasticité associés à ces variables en réponse à une carence en soufre et réalisé des analyses GWAS sur ces indices ainsi que sur les variables mesurées dans les quatre environnements. La synthèse de ces analyses sous forme de hotspots nous a guidé vers l'identification, sous ces pics majeurs, de gènes pertinents pour le contrôle de ces variables en réponse à la carence en soufre. L'exploitation de l'ensemble des données ouvre des perspectives vers l'identification de gènes candidats limitant l'impact négatif d'une carence en soufre sur les composantes du rendement chez le pois.

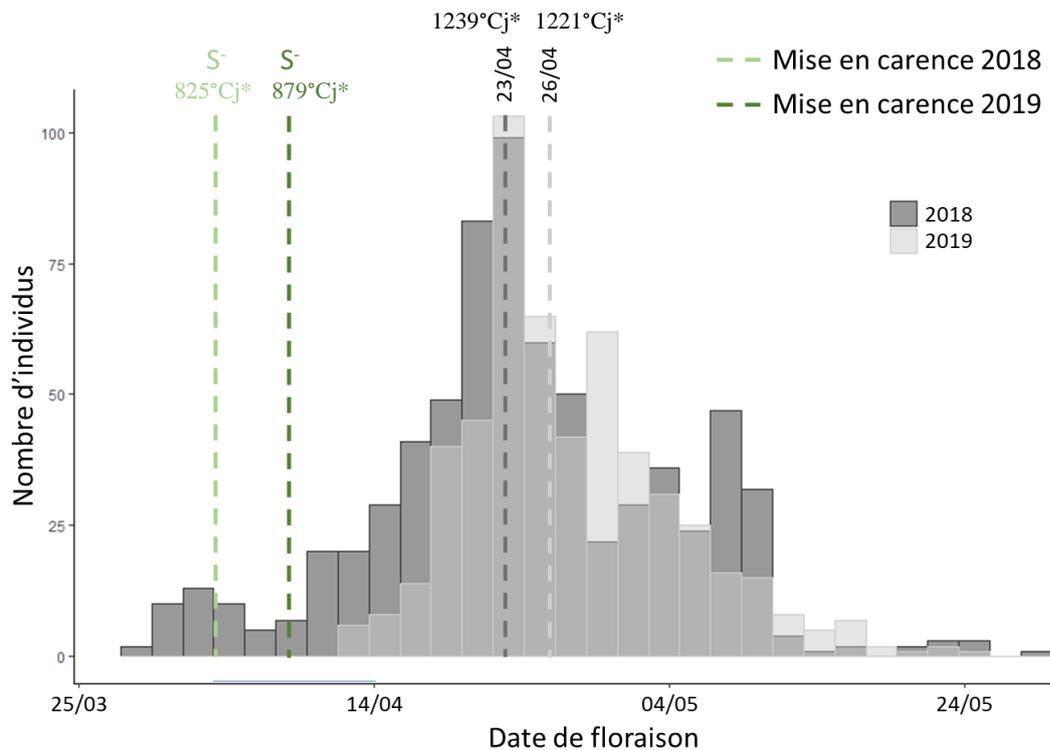
# 1. Application d'une carence en soufre et caractérisation des conditions

La population de pois choisie pour étudier l'effet d'une carence en soufre est constituée de 304 géotypes provenant des collections AMS et GRASP. Cette population est structurée en 5 groupes génétiques (cf. matériels et méthodes) constitués des différents types variétaux, culturaux, d'espèces et de sous-espèces. Ces groupes présentent certaines caractéristiques distinctes (Tableau 4). On remarque à titre d'exemple que les groupes génétiques 1 et 5 sont composés de géotypes appartenant respectivement aux *P. sativum ssp elatius* et *P. sativum ssp abyssinicum* pour le groupe 1 et aux *P. fulvum* pour le groupe 5. On observe aussi que le groupe 4 est majoritairement composé de pois d'hiver.

**Tableau 4 : Descriptif de la collection selon les 5 groupes génétiques.**

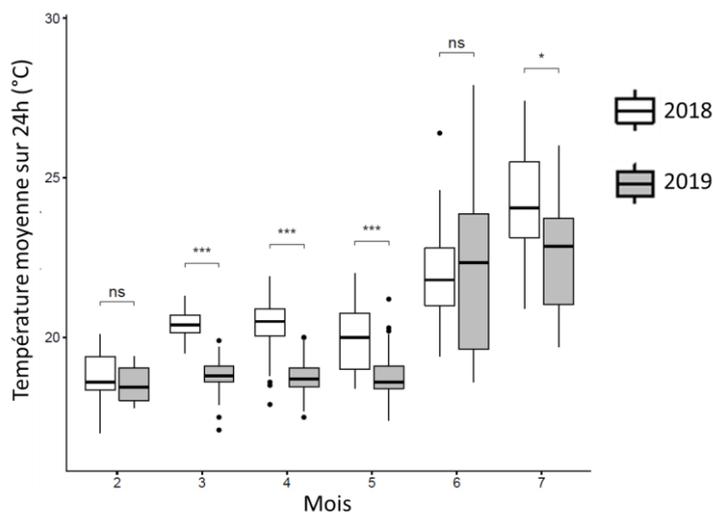
		Groupe génétique				
Caractéristique		G1	G2	G3	G4	G5
Variété	PROTEAGINEUX	0%	34%	0%	17%	0%
	MANGETOUT	0%	2%	0%	0%	0%
	POTAGER	0%	23%	8%	4%	0%
	FOURRAGER	4%	9%	13%	54%	0%
	NA	96%	32%	79%	25%	100%
Culture	PRINTEMPS	79%	66%	67%	12%	45%
	HIVER	0%	18%	4%	81%	0%
	NA	21%	16%	29%	8%	55%
Statut	cultivar	0%	35%	4%	19%	0%
	breeding	0%	14%	0%	33%	0%
	landrace	29%	15%	63%	19%	0%
	wild	57%	12%	29%	2%	100%
	NA	14%	24%	4%	27%	0%
Nom botanique	<i>P. sativum ssp sativum</i>	0%	95%	88%	98%	0%
	<i>P. sativum ssp elatius</i>	75%	4%	4%	0%	0%
	<i>P. sativum ssp abyssinicum</i>	25%	0%	0%	0%	0%
	<i>P. sativum ssp humile</i>	0%	1%	4%	0%	0%
	<i>P. sativum ssp hibernicum</i>	0%	0%	0%	2%	0%
	<i>P. sativum ssp asiaticum</i>	0%	0%	4%	0%	0%
	<i>P. sativum ssp transcaasicum</i>	0%	1%	0%	0%	0%
	<i>P. fulvum</i>	0%	0%	0%	0%	100%

Nous avons appliqué une carence en soufre aux géotypes de pois à partir d'un stade mi végétatif, qui pour le géotype Caméor, correspond au stade de formation du 6<sup>ème</sup> nœud. Appliquée à ce stade, une carence en soufre affecte significativement les composantes de rendement et de composition protéique des graines (Henriet et al., 2019). Pour les deux expériences (2018 et 2019), le jour de mise en carence correspond à ~ 850 degrés jour depuis le 1<sup>er</sup> semis (825 °Cj en 2018 et 879 °Cj en 2019, Figure 23).



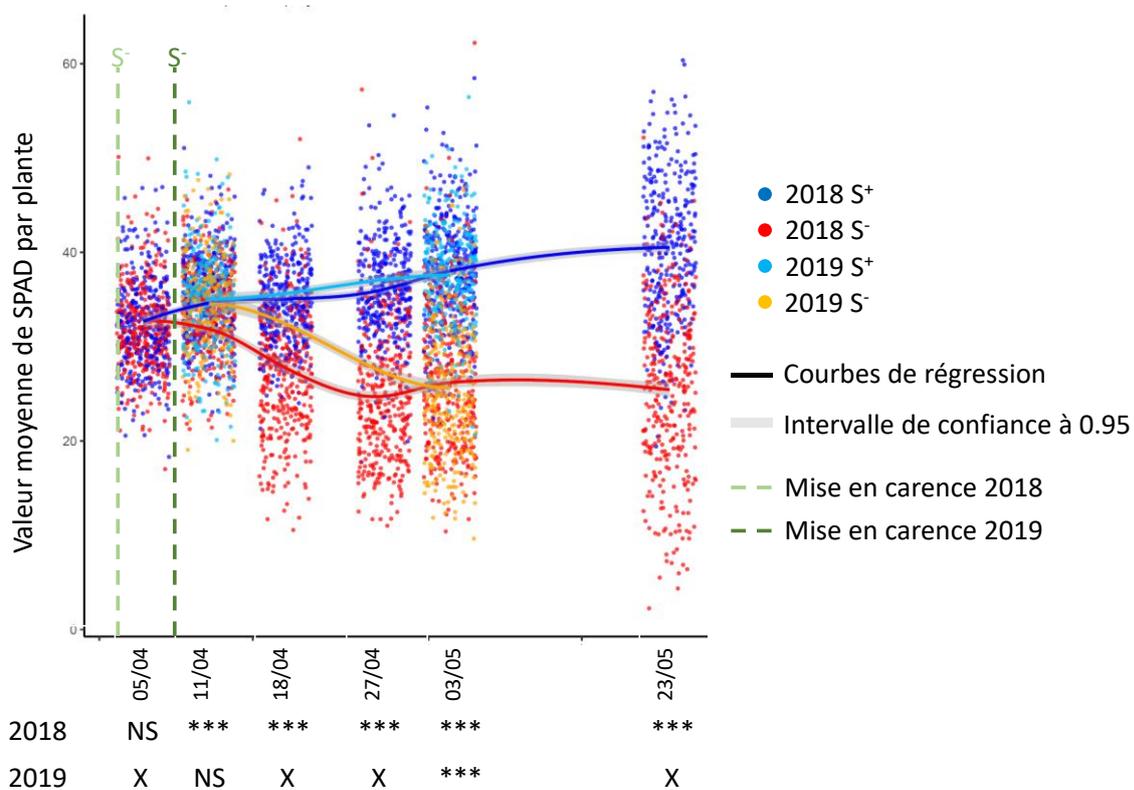
**Figure 23 : Date de floraison pour l'ensemble des génotypes de pois.** Gris foncé, 2018 ; Gris clair, 2019. Les dates médianes de floraison par année sont indiquées en pointillés. Les lignes verticales en pointillés vert clair et vert foncé correspondent respectivement à la date de mise en carence en 2018 et 2019. \* degrés jour calculés depuis la date du 1<sup>er</sup> semis.

Les plantes ont fleuri, en moyenne, environ 20 jours après l'application de la carence (~ 1230 °Cj depuis le 1<sup>er</sup> semis) mais avec une grande amplitude liée à la difficulté de synchroniser parfaitement la floraison de génotypes très contrastés. La floraison plus précoce en 2018 par rapport à 2019 peut s'expliquer par des températures moyennes journalières plus élevées au début du développement (mars et avril) en 2018 (Figure 24).



**Figure 24 : Températures moyennes journalières durant le développement des génotypes de pois.** Température moyenne sur 24h mesurée à l'intérieur de la serre. En février, seules les températures mesurées à partir du 22 sont prises en compte. Boxplots blancs, 2018 ; boxplots gris, 2019 ; Test de comparaison des moyennes entre les deux années : ns, non significatif ; \*, p-value<0.05 ; \*\*\*, p-value<0.001

Il est à noter que les températures moyennes journalières ont varié fortement entre mai et juin, en particulier en 2019 où un épisode de fortes chaleurs a été observé entre le 24 et le 30 juin (les températures moyennes journalières étaient toutes supérieures à 25°C). De tels épisodes de températures extrêmes survenant en fin de cycle peuvent avoir des conséquences sur le développement des plantes et le déroulement de la phase de remplissage/maturation des graines (Figure 8). Afin de visualiser les premiers symptômes caractéristiques d'un manque de soufre, la teneur en chlorophylle des feuilles des deux derniers nœuds formés a été mesurée pour l'ensemble des écotypes en 2018 et 2019 avec un appareil SPAD (Figure 25).



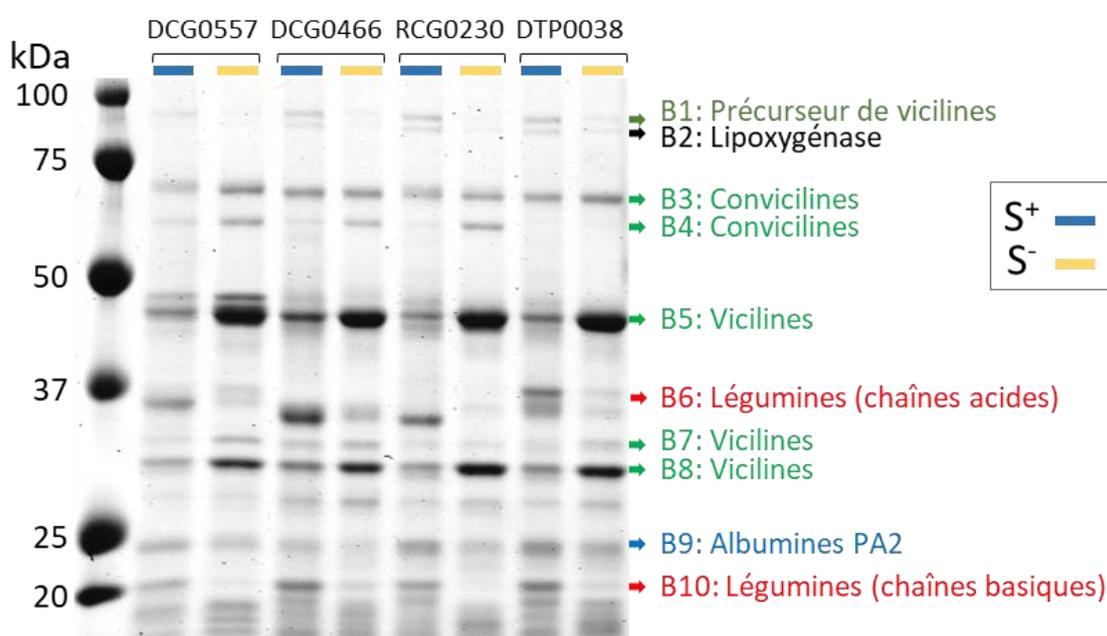
**Figure 25 : Evolution des valeurs de SPAD suite à l'application d'une carence en soufre chez le pois.** Chaque point correspond à la moyenne des valeurs de SPAD sur les deux derniers nœuds végétatifs. Bleu foncé, 2018 S<sup>+</sup> ; rouge, 2018 S<sup>-</sup> ; bleu clair, 2019 S<sup>+</sup> ; orange, 2019 S<sup>-</sup>. Les courbes de régression correspondantes sont de la même couleur. Les lignes verticales en pointillés vert clair et vert foncé correspondent respectivement à la date de mise en carence en 2018 et 2019. La significativité entre les écarts de la condition S<sup>+</sup> et la condition S<sup>-</sup> est indiquée dans le tableau sous le graphique, \*\*\*, p-value < 0.001 ; NS, non significatif, des croix sont placées lorsque la mesure n'a pas été faite ce jour.

Une diminution des valeurs moyennes de teneur en chlorophylle a été observée à partir d'une semaine après mise en carence et s'est accentuée les trois semaines suivantes pour atteindre un plateau à mi-floraison (50% des plantes ont fleuri le 23 avril en 2018 et le 27 avril en 2019). Cette diminution de la teneur en chlorophylle a la même intensité en 2018 et 2019 au regard de la date de mise en carence.

## 2. Variabilité et déterminisme génétique de la plasticité de la composition protéique des graines de pois vis-à-vis d'une carence en soufre.

### 2.1 Effet du génotype et d'une carence en soufre sur la composition protéique des graines

Afin d'étudier l'effet du génotype et d'une carence en soufre sur la composition protéique des graines de pois, les protéines des graines matures de 304 génotypes de pois produites en 2018 et 198 produites en 2019 dans deux conditions (apport suffisant de soufre = condition contrôle, et carence en soufre appliquée à partir du stade mi-végétatif) ont été séparées par électrophorèse mono-dimensionnelle. Un exemple de gel pour 4 génotypes est présenté Figure 26.



**Figure 26 : Variabilité de la composition protéique des graines des génotypes de pois cultivés en condition contrôle et en condition de carence en soufre appliquée deux semaines avant floraison.** Exemple de profil protéique de graines obtenu par électrophorèse monodimensionnelle pour 4 génotypes en condition contrôle et en condition de carence en soufre 2 semaines avant floraison en 2018 (condition contrôle, bleu ; condition de carence en soufre, jaune). Les protéines majoritaires identifiées par spectrométrie de masse sous chaque bande sont indiquées (en vert : globulines 7S, en rouge : légumine 11S, en bleu : albumines, et en noir : lipoxygénases)

Les abondances relatives de chacune de ces 10 bandes protéiques pour l'ensemble des génotypes étudiés les deux années en condition contrôle et en condition de carence en soufre ont été obtenues par analyse quantitative des gels 1-DE (cf. matériels et méthodes). Parmi les

10 bandes protéiques majeures détectées, cinq correspondent à des globulines 7S (B3 & B4 = convicilines (conv), B5 = vicilines de 47 kDa (vic47), B7 et B8 = sous-unités de vicilines (vic)), deux correspondent à des globulines 11S (B6 et B10, légumine), et une à l'albumine PA2 (B9). Les deux autres bandes protéiques de haut poids moléculaire ont été détectées et identifiées comme correspondant à une forme de précurseur de viciline (B1) et à une lipoxygénase (B2). Les moyennes et les écart-types des abondances relatives de ces bandes ainsi que l'effet de la carence en soufre sur chacune d'entre elles sont présentés Tableau 5.

**Tableau 5 : Variabilité de la composition protéique des graines de pois à l'échelle de la population.**

Bande protéique	S+						S-						Effet carence en S	
	2018			2019			2018			2019			2018	2019
	Moyenne	T	SD	Moyenne	T	SD	Moyenne	T	SD	Moyenne	T	SD		
Précurseur de viciline B1	2.87	c	1.23	3.75	d	1.30	0.67	a	0.56	1.05	b	0.78	-77%	-72%
Lipoxygénase B2	1.64	b	0.75	1.85	c	0.86	0.54	a	0.35	0.59	a	0.55	-67%	-68%
B3	9.43	b	3.1	10.86	c	2.79	8.01	a	4.21	10.10	b	3.10	-15%	-7%
B4	2.06	a	1.69	3.59	b	1.95	3.95	b	2.42	6.18	c	2.90	92%	72%
Vicilines B5	19.29	b	6.55	16.38	a	4.20	47.87	d	12.74	40.30	c	9.87	148%	146%
B7	2.82	a	1.08	3.52	c	1.23	2.96	b	1.37	3.91	d	1.43	5%	11%
B8	12.69	a	4.23	13.64	a	3.41	23.56	b	8.5	26.11	c	6.97	86%	91%
Légumine B6	30.17	b	6.89	30.93	b	4.76	6.41	a	5.22	6.04	a	3.02	-79%	-80%
B10	9.13	d	2.78	5.93	c	1.76	1.4	b	1.73	0.59	a	0.59	-85%	-90%
Albumine B9	9.9	b	3.51	9.55	b	2.55	4.63	a	2.65	5.14	a	2.05	-53%	-46%
7S/11S	1.27	a	0.49	1.36	a	0.40	23.396	c	35.77	17.46	b	11.02	1742%	1181%

Moyennes et écarts types de l'abondance relative pour chaque bande protéique par condition et par année et impact (en pourcentage par rapport à la condition contrôle) de la carence en soufre sur ces variables pour 2018 et 2019, les résultats des tests de Tuckey sont indiqués par des lettres (de a à d).

On observe une forte augmentation de la proportion des globulines 7S B4, B5 et B8 et une forte diminution de la proportion des légumine B6 et B10 ainsi que de l'albumine PA2 (B9) en condition de carence en soufre, et ce pour les deux années. Ces résultats concordent avec les données de la littérature montrant qu'une carence en soufre affecte l'accumulation des protéines de réserve les plus riches en soufre dans les graines (Blagrove et al., 1976; Chandler et al., 1984; Henriet et al., 2019; Spencer et al., 1990; Zuber et al., 2013). Comme précédemment observé par (Henriet et al., 2019), l'abondance relative de la bande B3, correspondant à une conviciline, n'augmente pas en condition de carence en soufre.

Une analyse de variance en modèle aléatoire sur les deux années montre des effets génotype (G), environnement (E), et GxE significatifs pour chacune des 10 bandes protéiques et pour le ratio 7S/11S (Tableau 6, partie gauche).

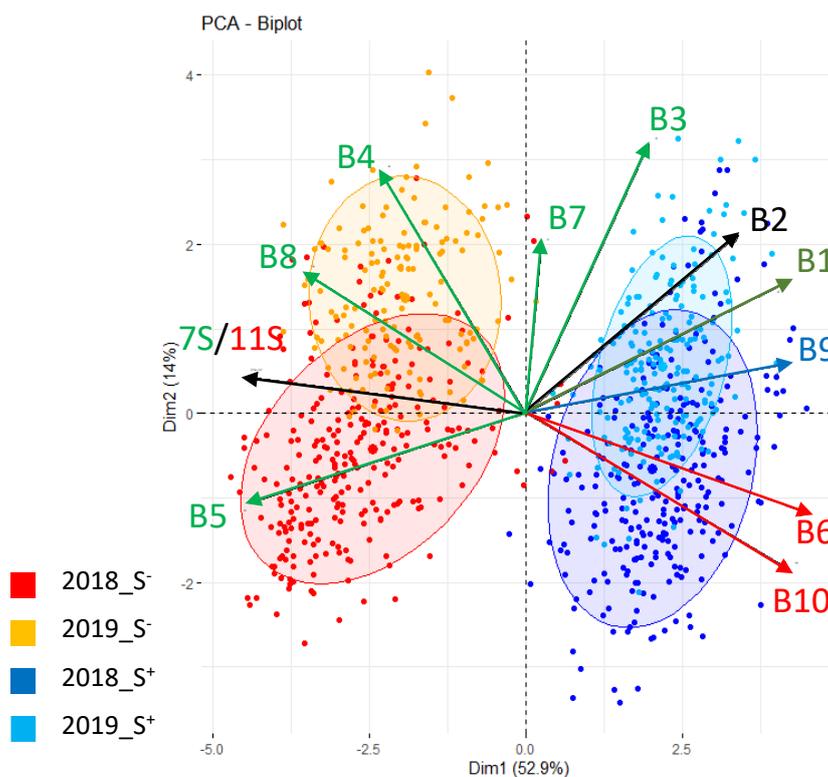
**Tableau 6 : Analyse de l'effet génotypique et environnemental sur la composition protéique des graines.**

	Année*Cuve			Condition			Effets aléatoires					2018_S+			2019_S+			2018_S-			2019_S-		
	SS	ddl	F	SS	ddl	F	VarG	Var GxE	VarRes	%Var G	%Var GxE	VarG	VarRes	h²									
B1	142.63	15	22.52	594.33	1	1407.78	0.28	0.34	0.42	0.36	0.43	1.35	0.18	0.88	0.11	0.18	0.37	1.02	0.70	0.59	0.35	0.31	0.53
B2	18.20	15	7.53	149.52	1	928.37	0.13	0.12	0.16	0.41	0.38	0.48	0.08	0.86	0.05	0.07	0.43	0.58	0.17	0.77	0.21	0.10	0.69
B3	1221.22	15	22.89	120.96	1	34.02	5.77	2.73	3.56	0.58	0.27	7.26	2.95	0.71	14.12	4.65	0.75	6.58	1.17	0.85	7.92	2.29	0.78
B4	1105.07	15	33.48	953.14	1	433.20	2.57	0.47	2.20	0.53	0.10	1.17	2.07	0.36	2.58	4.62	0.36	3.41	0.49	0.87	7.27	1.60	0.82
B5	7156.89	15	16.79	69531.00	1	2446.20	30.83	25.81	28.42	0.47	0.39	33.19	5.39	0.86	118.41	45.84	0.72	14.32	3.82	0.79	78.13	19.76	0.80
B6	490.15	15	2.81	53187.12	1	4580.88	7.88	11.75	11.61	0.35	0.52	40.84	5.07	0.89	11.61	19.13	0.38	18.54	5.51	0.77	2.79	6.97	0.29
B7	226.57	15	31.75	5.68	1	11.94	0.73	0.43	0.48	0.56	0.33	0.92	0.26	0.78	1.51	0.35	0.81	1.31	0.15	0.90	1.72	0.30	0.85
B8	1344.81	15	12.37	11075.16	1	1527.96	24.54	8.24	7.25	0.73	0.24	15.34	2.52	0.86	63.03	10.93	0.85	9.82	1.31	0.88	39.71	7.66	0.84
B9	387.65	15	8.54	2798.47	1	924.27	2.85	2.14	3.03	0.44	0.33	10.21	2.55	0.80	3.00	4.11	0.42	5.91	0.78	0.88	2.94	1.23	0.70
B10	1465.74	15	43.85	5461.13	1	2450.90	0.70	1.52	2.23	0.21	0.46	6.05	1.84	0.77	1.13	1.94	0.37	2.50	0.78	0.76	0.11	0.11	0.49
log(7S/11S)	10.81	15	4.42	700.03	1	4294.44	0.06	0.11	0.16	0.23	0.46	0.18	0.05	0.78	246.29	136.89	0.64	0.13	0.04	0.77	45.85	37.10	0.55

Number of obs: 1451, groups: Accession:Condition, 574; Accession, 290

Partie gauche de la table : ANOVA (fonction lmer) à effets fixes (année\*cuve, environnement) et aléatoires (génotype et interaction génotype\*environnement). Un environnement est la combinaison année, condition de nutrition soufrée. Les sommes de carrés de type III (SS), degrés de liberté (ddl) et statistiques de Fisher (F) sont indiqués pour les effets fixes. Les estimations des variances sont indiquées pour les effets aléatoires ; toutes sont significativement différentes de zéro. La part des variance génétique (%VarG) et d'interaction (%VarGxE) dans la variation totale (VarRes + VarG + ¼VarGxE) est indiquée. Partie droite de la table : ANOVA (fonction lmer) à effets fixe (cuve) et aléatoire (génotype) pour chaque environnement. Les héritabilités sont estimées par le ratio VarG/(VarG+VarRes). Pour chaque analyse, le ratio 7S/11S a dû être transformé (logarithme) pour procéder à l'estimation des variances (en raison de la singularité de la matrice des effets fixes).

Un environnement est la combinaison d'une année et d'une condition de nutrition : 2018\_S<sup>+</sup>, 2018\_S<sup>-</sup>, 2019\_S<sup>+</sup>, 2019\_S<sup>-</sup>). Dans notre dispositif, les deux effets année et cuve sont confondus et donc l'effet année n'est pas directement testable. L'effet cuve\*année est également toujours significatif mais contribue peu à la variation totale (entre 1 et 10% de la somme des carrés totale d'une anova avec tous les effets fixes). Il est le plus marqué pour les bandes 3, 4 et 6 et justifie de travailler par la suite sur les moyennes ajustées pour cet effet. La part de la variance d'interaction GxE dans la variation totale est la plus élevée (~0.4-0.5) pour les légumineuses B6 et B10, le ratio 7S/11S et le précurseur des vicilines B1 et la plus faible (~0.1-0.3) pour les convicilines (B3, B4), l'albumine (B9) et les vicilines de faible poids moléculaires (B7 et B8). L'estimation des variances génétique et résiduelle par année et condition de nutrition à l'aide d'un modèle d'ANOVA à effets aléatoires (Tableau 6, partie droite), montre que l'héritabilité moyenne des caractères est plus élevée en condition de nutrition soufrée suffisante qu'en condition de carence soufrée (0.79 vs 0.62). De façon générale, l'héritabilité est plus importante pour les bandes majeures que pour les bandes mineures dont le volume est plus délicat à estimer. L'Analyse en Composantes Principales (ACP) sur les moyennes ajustées des valeurs par génotype pour les 10 bandes protéiques ainsi que pour le ratio globulines 7S/11S sépare les individus sur l'axe 1 en fonction de leur composition protéique opposant les légumineuses et l'albumine aux vicilines (Figure 27).

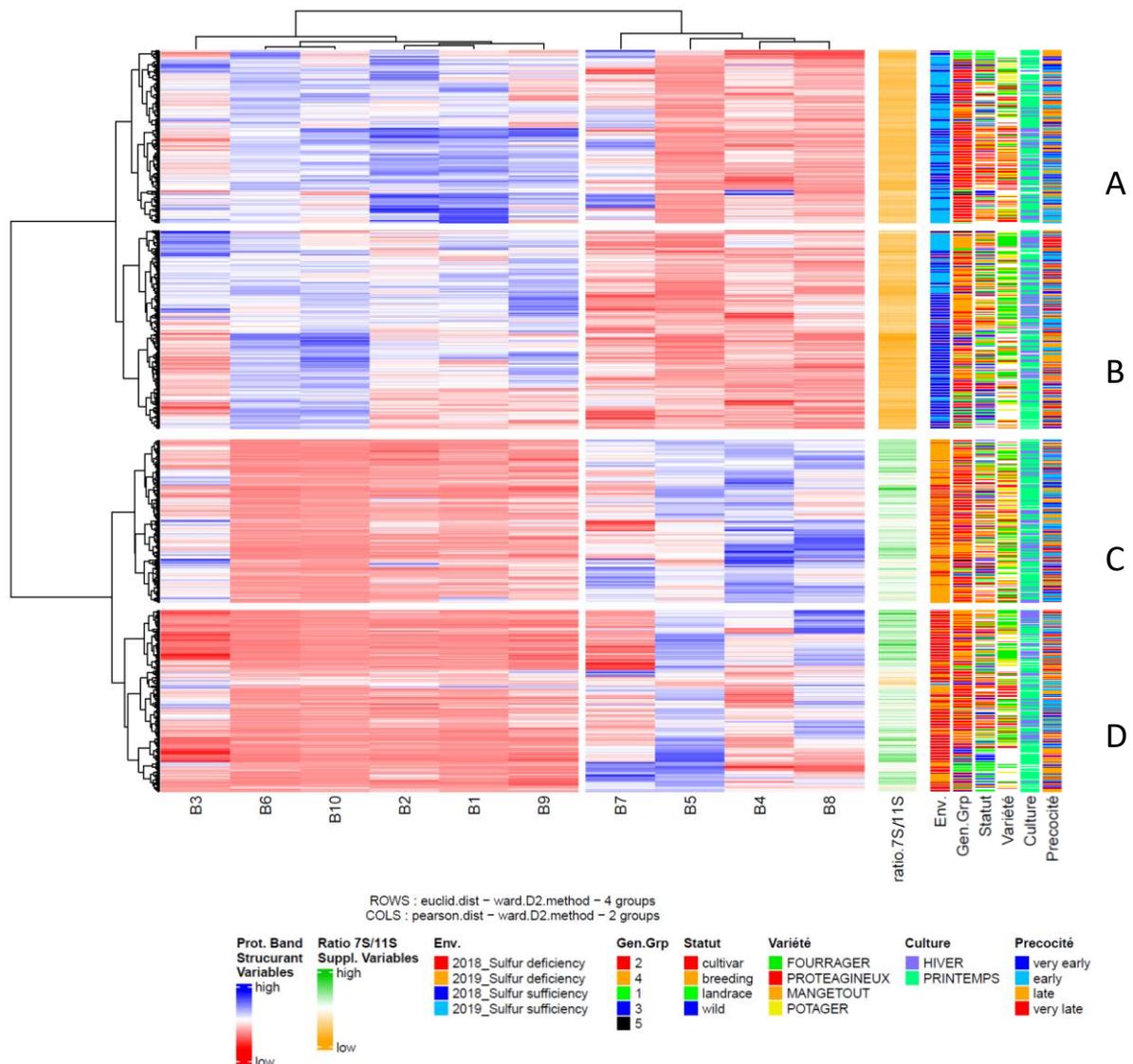


**Figure 27 : Répartition des individus en fonction de leur composition protéique selon l'année et l'application d'une carence en soufre.** Analyse en composantes principales des génotypes pour lesquels l'analyse a été effectuée en 2018 et/ou 2019 pour les 10 bandes protéiques ainsi que pour le ratio 7S/11S (globulines 7S, flèches vertes ; globulines 11S, flèches rouges ; albumine, flèche bleue ; précurseur de viciline, flèche vert foncé ; lipoxygénase, noir, ratio 7S/11S, noir).

L'axe deux n'explique que 14% de la variation et sépare les individus des deux années d'expérience. L'existence de cet effet année était suggéré par la significativité de l'effet année\*cuve, et pourrait s'expliquer par des variations de température observées dans la serre entre les deux années avec un épisode de canicule fin juin plus prononcé en 2019 (figure 24).

Pour la composition protéique, ce serait les bandes protéiques de type 7S (convicilines et bande 7) qui seraient plus abondantes en 2019, quelle que soit la condition de nutrition. Ce qui pourrait s'expliquer par l'existence en 2019 d'un double stress thermique et de carence en soufre sur la période de remplissage, cela a pu potentiellement atténuer l'effet de la carence sur le ratio 7S/11S.

Un dendrogramme de l'ensemble des données de composition protéique (Figure 28) divise celles obtenues en condition de nutrition optimale (groupes A et B) et celles obtenues en condition de carence (groupes C et D), cette séparation est basée sur des niveaux contrastés de deux groupes de variables : l'un comprenant les globulines 11S, l'albumine B9 et les bandes B1 à B3, l'autre comprenant les globulines 7S. Cette séparation s'accompagne d'une variation très nette du ratio 7S/11S.



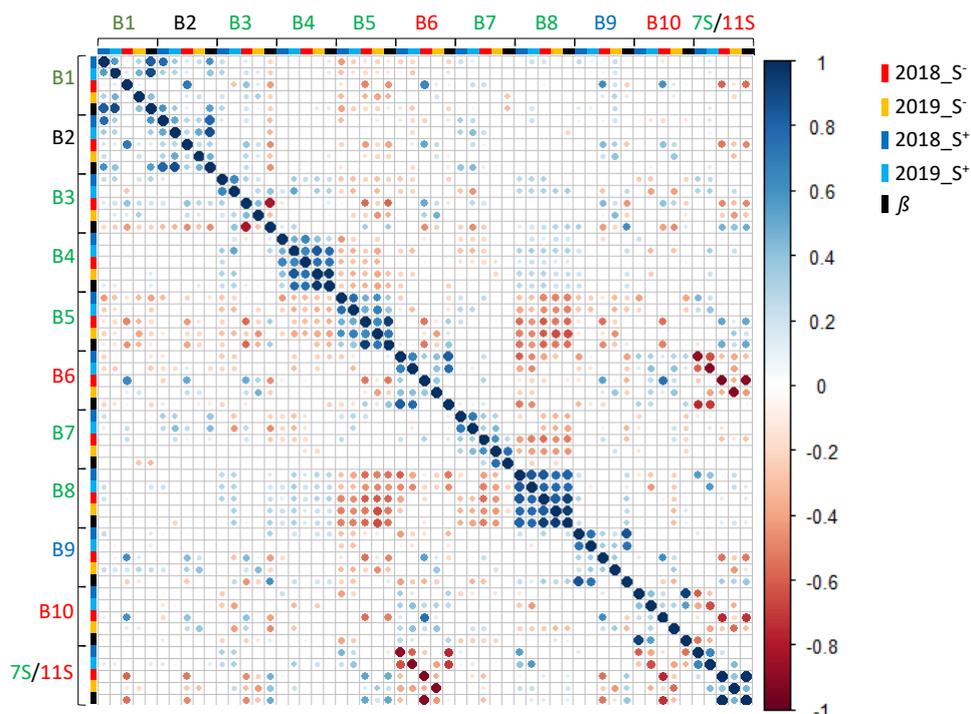
**Figure 28 : Dendrogramme réalisé à partir des génotypes analysés les deux années pour la composition protéique des graines de pois (870 profils protéiques).** Les moyennes ajustées pour l'effet cuve ont été standardisées. Quatre groupes d'individus ont été retenus par l'algorithme k-means et la classification ascendante hiérarchique a été effectuée à partir d'une matrice de distance euclidienne et les individus regroupés par la méthode de Ward. Les 10 variables analysées (B1 à B10) ont été regroupées en deux groupes sur la base de leurs corrélations. Le ratio 7S/11S a été ajouté en donnée supplémentaire ainsi que l'appartenance des observations à un environnement (2018, 2019, S<sup>+</sup>, S<sup>-</sup>), un groupe génétique (DAPC), un statut génétique, un type variétal, cultural ou un niveau de précocité. Les niveaux de couleurs du dendrogramme symbolisent le niveau de la variable (élevée en bleu et faible en rouge), pour le ratio 7S/11S, les valeurs élevées sont vertes et les faibles sont jaunes.

En condition S<sup>+</sup>, les groupes A et B se distinguent par le niveau des bandes de viciline B7 et de lipoxgénase B2, toutes deux fortes dans le groupe A composé principalement de pois protéagineux et faibles dans le groupe B composé principalement de pois fourragers. En condition S<sup>-</sup>, les groupes C et D se séparent sur le niveau des bandes B3 (9.7% pour le groupe 3 vs 7.1% pour le groupe 4) et B4 (6.8% vs 3.1%), toutes deux des convicilines et également sur le niveau de la viciline B5 (40.8% vs 50.8%). Les écarts type des moyennes sont similaires pour les deux groupes. Les convicilines sont plus présentes dans le groupe C où la bande B5

est en moyenne de moindre importance alors que le niveau de la bande B5 est plus élevé dans le groupe D où les convicilines sont d'un niveau plus faible. L'impact de la carence sur l'ensemble de la collection permet de séparer les génotypes en fonction du niveau respectif en globulines 7S vicilines *versus* 7S convicilines. Contrairement à la séparation des groupes A et B en condition S<sup>+</sup> clairement basée sur le type variétal (protéagineux, fourrager), la séparation des groupes C et D n'est pas visiblement liée à des caractéristiques variétales, de type de culture ou de groupe génétique. Seule l'aptitude des génotypes à accumuler différemment les globulines 7S semble les distinguer en condition de carence soufrée. Le fait que la variance d'interaction GXE soit forte pour la bande 5 et faible pour les convicilines (bandes 3 et 4) suggère également des réponses différentielles à la carence pour ces bandes.

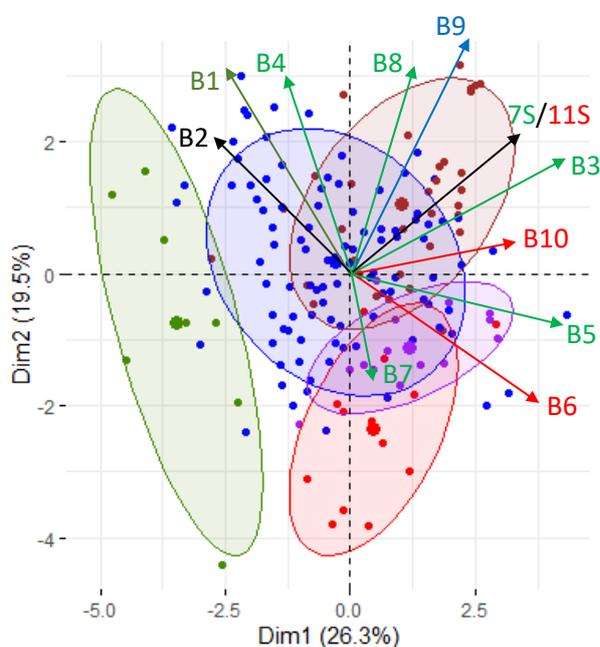
## 2.2 Plasticité de la composition protéique des graines de pois

La méthode de Finlay et Wilkinson (FW) a été utilisée pour déterminer l'indice de plasticité  $\beta$  de chaque bande protéique et du ratio 7S/11S en prenant en compte les données obtenues les deux années dans les deux conditions de nutrition soufrée, soit quatre environnements. Une heatmap des corrélations entre les  $\beta$  calculés et les quantités relatives des bandes protéiques dans chaque environnement (2018\_S<sup>+</sup>, 2018\_S<sup>-</sup>, 2019\_S<sup>+</sup>, 2019\_S<sup>-</sup>) est présentée Figure 29.



**Figure 29 : Heatmap des corrélations entre les contenus relatifs des protéines de réserve mesurés chaque année dans chaque condition (2018\_S<sup>-</sup>, rouge ; 2019\_S<sup>-</sup>, orange ; 2018\_S<sup>+</sup>, bleu foncé ; 2019\_S<sup>+</sup>, bleu clair) et leur indice de plasticité  $\beta$  (en noir). Les données pour le ratio 7S/11S sont également ajoutées. Seules les corrélations significatives au seuil 0.05 sont présentées, après ajustement de Benjamini Hochberg pour comparaisons multiples.**

On observe des corrélations différentes selon les conditions et les bandes protéiques. Pour B6-leg, B9-PA2 et B10-leg, préférentiellement accumulées lorsque la nutrition soufrée est optimale (Tableau 6), les  $\beta$  sont positivement corrélés aux quantités relatives des bandes dans les environnements  $S^+$ , tandis que pour B5-vic47 et B7-vic, les  $\beta$  sont positivement et fortement corrélés aux quantités relatives de ces bandes dans les environnements  $S^-$ , là où elles sont les plus abondantes. Les  $\beta$  calculés pour B4-conv et B8-vic sont positivement et fortement corrélés à l'abondance relative des bandes protéiques dans au moins trois environnements. La tendance observée est que les plus fortes corrélations entre l'indice de plasticité et la valeur des bandes protéiques dans un ou plusieurs environnements donnés sont liées à une plus forte variance génétique pour la bande protéique de cet ou ces environnements par rapport aux autres environnements ( $r^2 > 0.8$  entre plasticité et environnements  $S^-$  2019 pour B4 et 2018 et 2019 pour B8, voir tableau 7). L'indice  $\beta$  du ratio 7S/11S est corrélé à sa valeur dans l'environnement 2018\_ $S^-$  en accord avec une très forte variance de ce ratio dans cet environnement. Les indices beta des différentes bandes sont peu corrélés entre eux (corrélation maximale observée entre B5-vic47 et B6-leg de 0.4), ce qui justifie de faire des analyses génétiques sur chacun des indices de plasticité calculés. Une ACP a été réalisée à partir des indices de plasticité des 10 bandes protéiques et du ratio 7S/11S en tenant compte de la répartition des individus au sein des cinq groupes génétiques qui structurent la collection (Figures 30).



**Figure 30 : Distribution des génotypes utilisés dans l'analyse en fonction de leur indice  $\beta$  pour les dix bandes protéiques et le ratio 7S/11S.** Analyse en composante principale montrant la distribution des génotypes utilisés dans l'analyse en fonction de leur indice  $\beta$  pour les dix bandes protéiques et le ratio 7S/11S (globulines 7S, flèches vertes ; globulines 11S, flèches rouges ; albumine, flèche bleue ; précurseur de viciline, flèche vert foncé ; lipoxygénase, noir, ratio 7S/11S, noir). La répartition des génotypes est colorée avec des ellipses contenant au moins 75% des individus en fonction de leur appartenance aux groupes génétiques (G1 à G5).

L'ACP sépare les individus en fonction de leur groupe génétique à la fois sur l'axe 1 (26.3% de la variation) et sur l'axe 2 (19.5% de la variation). Une comparaison des moyennes de plasticité des bandes protéiques pour chacun des 5 groupes génétiques a été réalisée (Tableau 8).

**Tableau 7 : Plasticité des bandes protéiques.**

Groupe	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	7S/11S
1	-0.31 a	-0.37 a	0.01 ab	-0.39 a	0.29 bc	0.22 cd	0.35 b	-0.34 a	-0.25 a	-0.15 b	-0.54 a
2	0.08 c	0.14 b	-0.1 a	0.03 b	0.09 bc	0.07 b	-0.12 a	0.04 b	-0.03 a	0.07 c	-0.16 a
3	-0.21 ab	-0.16 a	0.45 bc	-0.1 ab	0.34 c	0.25 d	0.17 ab	0 ab	-0.27 a	0.19 c	-0.21 a
4	-0.05 bc	-0.18 a	0.83 c	0.17 b	0.07 bc	0.06 bc	0.15 b	0.37 c	0.48 b	0.06 c	0.01 a
5	0.05 bc	-0.06 ab	-0.83 a	0.03 ab	-0.58 a	-0.36 a	0.37 ab	-0.37 a	-0.24 a	-0.94 a	-0.69 a

Moyennes des valeurs des indices de plasticité  $\beta$  pour chacune des bandes et le ratio 7S/11S pour chacun des groupes génétiques. Un test de comparaison de moyennes par la méthode de Tukey est indiqué sous forme de lettres.

Le groupe d'individus qui répond le plus différenciellement des autres pour la plasticité est le groupe 5 (uniquement composé de *Pisum fulvum*) (voir la distribution des différentes sous-espèces par groupe Tableau 5). En effet, ce groupe d'individus présente des valeurs extrêmes et négatives de plasticité pour les bandes protéiques majeures correspondant aux vicilines (B5 et B3) et légumine (B6 et B10) comparativement aux autres groupes, ce qui lui confère, au niveau phénotypique, une faible augmentation de la teneur en viciline assortie d'une faible diminution des légumine soit le meilleur maintien de la composition protéique de la collection en condition de carence.

## 2.3 Exploration par GWAS des régions du génome contrôlant la composition protéique des graines de pois et sa plasticité vis-à-vis de l'environnement

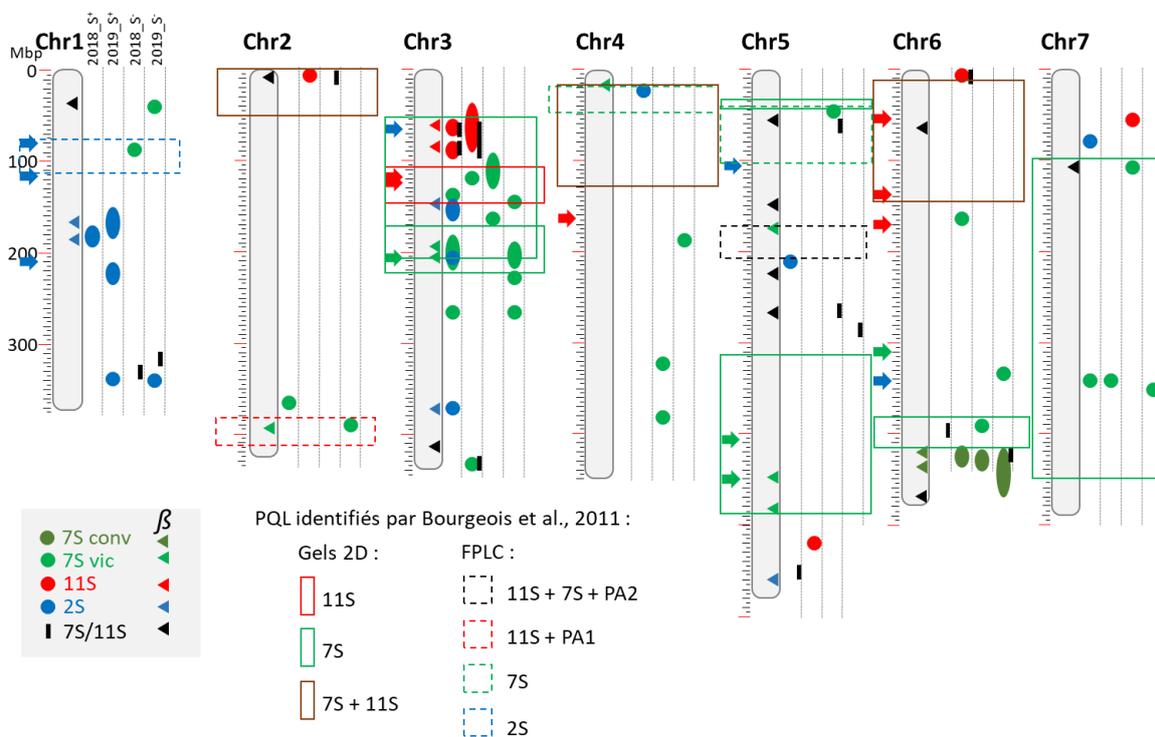
Les abondances relatives et indices de plasticité des globulines (B3 à B8, B10 et ratio 7S/11S) et de l'albumine PA2 (B9) ont été utilisés pour identifier, par GWAS, les régions génomiques impliquées dans leur variabilité au sein de la population. Le nombre total de SNP retenus (cf. matériels et méthodes) pour chaque bande protéique dans chaque environnement est présenté Figure 31.

Des SNPs associés à des variations de quantité relative des protéines ont été identifiés pour chaque bande dans chaque environnement, et des SNPs associés à des variations de la plasticité ( $\beta$ ) ont aussi été identifiés pour chaque bande protéique à l'exception de B5-vic47. La distribution des SNPs sur les 7 chromosomes du pois a permis de mettre en évidence des

hotspots associés à la variation d'abondance de chacune des protéines d'intérêt dans un ou plusieurs environnements ainsi que des hotspots associés aux indices de plasticité (Figure 32).

(a) Nombre de SNPs identifiés							(b) Nombre de gènes identifiés								
Prot.	S <sup>+</sup>			S <sup>-</sup>			$\beta$	Prot.	S <sup>+</sup>			S <sup>-</sup>			$\beta$
	2018	2019	Communs	2018	2019	Communs			2018	2019	Communs	2018	2019	Communs	
B1	177	194	29	51	17	0	137	B1	100	88	21	35	12	0	58
B2	572	461	361	124	2	0	468	B2	260	189	158	90	7	0	204
B3	4	1	0	19	4	0	4	B3	5	2	0	15	7	0	9
B4	21	123	0	122	418	69	337	B4	35	89	0	120	211	69	161
B5	5	3	0	38	19	0	0	B5	13	7	0	27	18	0	0
B7	10	32	0	2	23	0	107	B7	13	38	0	2	37	0	88
B8	752	116	37	485	336	209	162	B8	286	141	39	201	201	128	104
B6	58	355	51	6	15	0	3	B6	26	225	14	15	24	0	8
B10	122	201	65	13	1	0	150	B10	62	122	51	21	4	0	69
B9	235	244	82	3	16	0	207	B9	196	184	94	4	25	0	112
Total	1639	1295	493	832	883	278	1348	Total	812	816	318	506	544	197	705
7S/11S	225	363	154	81	94	1	371	7S/11S	126	199	66	78	115	5	504

**Figure 31 : Nombre de SNPs associés à la composition protéique des graines de pois et sa plasticité, et nombre de gènes correspondants.** Bilan des analyses GWAS réalisées sur les 10 bandes protéiques et le ratio 7S/11S année par année et dans les deux conditions ainsi que sur les indices de plasticité (a) Nombre de SNP identifiés par approche GWAS au seuil  $10^{-6}$  auxquels ont été ajoutés les SNP en déséquilibre de liaison avec ceux-ci ayant une P-value comprise entre  $10^{-6}$  et  $10^{-5}$  (cf. matériels et méthodes). (b) Nombre de gènes identifiés dans le voisinage de chacun des SNP (5kb).

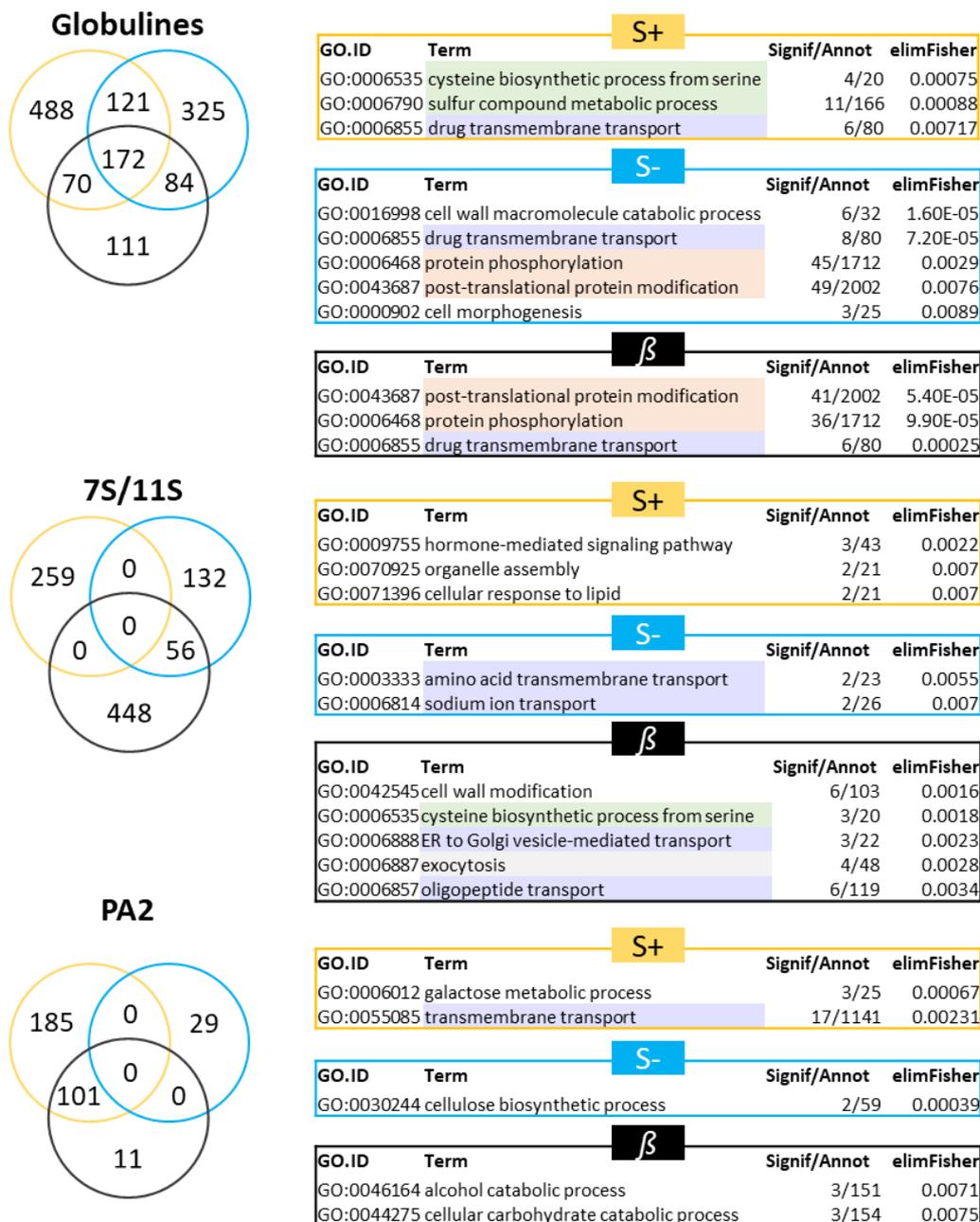


**Figure 32 : Position des principaux hotspots issus des analyses GWAS sur la carte génomique du pois (en Mbp).** Les principaux hotspots identifiés par GWAS pour les globulines et l'albumine (globulines 7S, vert ; globulines 11S, rouge ; albumine, bleu ; ratio 7S/11S, noir) en condition témoin (S<sup>+</sup>), en condition de carence en soufre (S<sup>-</sup>) et pour l'indice de plasticité  $\beta$ . Un hotspot est défini lorsqu'il y a plus de 5 SNPs sélectionnés pour 5 Mb. Les flèches correspondent aux gènes de structure connus associés à la composition protéique de la graine. La position des PQL détectés par Bourgeois et al. (2011) pour la composition protéique des graines est indiquée pour chaque type de protéine de réserve quantifié soit après électrophorèse bi-dimensionnelle (gel 2D), soit après FPLC (chromatographie liquide de protéine rapide).

Les hotspots ont été définis comme la présence d'au moins 5 SNPs sélectionnés par notre méthode dans un intervalle de 5 Mb. La distribution des hotspots met en évidence un déterminisme génétique complexe de la composition protéique des graines, qui varie selon l'environnement. Certains hotspots colocalisent entre environnements et avec la plasticité ainsi qu'avec des gènes de structure codant les protéines de réserve, en particulier sur le chromosome 3. Il est important de noter que certains hotspots liés à la plasticité sont aussi identifiés dans un ou plusieurs environnements tels que le hotspot à 200 Mbp sur le chromosome 3 et celui présent à 420 Mbp sur le chromosome 6, ce dernier hotspot étant spécifique aux convicilines 7S. Les hotspots des globulines semblent être spécifiques soit des vicilines voire des convicilines, soit des légumineuses, probablement en raison d'un déterminisme génétique différent. En utilisant les marqueurs ponts entre les cartes physiques et génétiques du pois (Tayeh et al., 2015) il a été possible de comparer la position des hotspots à celle des PQL (Protein Quantity Loci) précédemment détectés pour la composition protéique des graines en condition normale de nutrition soufrée (Bourgeois et al., 2011). Cette comparaison met par exemple en évidence des colocalisations entre hotspots et PQL de vicilines sur les chromosomes 3 et 5 (Figure 32). La région à ~ 200 Mbp sur le chromosome 3 est particulièrement intéressante car elle apparaît spécifique des vicilines et identifiée à partir des données PQL, de plasticité et par environnement (2018\_ S<sup>+</sup> et 2019\_ S<sup>-</sup>).

## 2.4 Etablissement d'un répertoire de gènes candidats pour la composition protéique des graines de pois et sa régulation sous contrainte soufrée

Ces données ont permis d'établir une liste de gènes candidats par environnement et par variable. Le nombre de gènes identifiés pour chaque variable est présenté Figure 31. Les résultats obtenus pour les globulines, le ratio 7S/11S et l'albumine PA2, prenant en compte l'ensemble des gènes détectés au moins une des deux années, sont présentés sous forme de diagramme de Venn dans la Figure 33. Ils sont aussi listés dans le Tableau S10, où ils sont ordonnés selon leur position chromosomique en lien avec la figure 32. Cette analyse a mis en évidence 851 gènes associés aux globulines en condition contrôle, 702 gènes en condition de carence en soufre et 437 associés à la plasticité, dont 172 gènes en commun entre les deux conditions et l'indice de plasticité  $\beta$  (Figure 33).



**Figure 33 : Nombre de gènes de pois communs aux différentes conditions pour chaque classe protéique et principales ontologies présentes.** Diagrammes de Venn représentant les gènes associés aux globulines, à l'albumine et au ratio 7S/11S en 2018 et/ou 2019 en condition contrôle, en condition de carence en soufre et pour l'indice de plasticité  $\beta$ . Les principales ontologies présentes dans les listes de gènes pour chaque condition sont indiquées en encadré. Les familles d'ontologies proches présentes dans plusieurs conditions sont surlignées en couleur en vert pour celles liées au métabolisme du soufre, en violet pour celles liées aux mécanismes de transport, en orange pour celles liées aux modifications post-traductionnelles).

Pour le ratio 7S/11S, 259 gènes ont été identifiés en condition contrôle uniquement et 188 gènes en condition de carence en soufre, dont 56 sont partagés avec le  $\beta$  (554 gènes détectés). En ce qui concerne l'albumine PA2, 286 gènes sont liés à des variations de composition protéique dans la condition contrôle, 29 dans la condition de carence en soufre uniquement et 112 sont liés à des variations du  $\beta$ , la plupart (101/111) étant communs avec la condition contrôle.

L'ensemble des informations sur les SNPs dont la variation a pour effet un changement d'acide aminé pouvant impacter la fonction de la protéine de manière modérée à élevée est disponible dans le Tableau S11. La liste de gènes identifiés par GWAS a été croisée avec des données récemment obtenues par transcriptomique (Tableau S10) afin d'étudier leur expression lors du développement de la graine (Vernoud et al., données non publiées) et en réponse à une carence en soufre dans les graines (Henriet et al., 2019) et les feuilles des nœuds végétatifs (Gallardo et al., données non publiées) et reproducteurs (Henriet et al., données non publiées).

Le répertoire de gènes candidats issus de GWAS contient 211 régulateurs (188 homologues à des facteurs de transcription et 23 associés au remodelage de la chromatine) que nous avons extrait dans le Tableau S12. En triant les gènes sur la base de leur *P*-value issue de GWAS et de leur profil d'expression lors du développement de la graine, nous avons mis en évidence 17 gènes régulateurs préférentiellement exprimés pendant le remplissage des graines dans le top20 des plus faibles « *P*-values » (surlignés en jaune dans le Tableau S13). Un de ces gènes (Psat5g165120), identifié pour B4-conviciline en condition S<sup>+</sup>, est homologue aux sous-unités NF-YA (Nuclear factor Y, sous-unité A). Ces sous-unités interagissent avec l'ADN, sous forme de complexe avec d'autres sous-unités (NF-YB et NF-YC), par reconnaissance du motif CCAAT. Des facteurs de la famille NF-Y sont connus pour réguler le remplissage des graines. C'est le cas du facteur de transcription Leafy Cotyledon 1 (LEC 1), de type NF-YB, qui contrôle l'expression des gènes codant les protéines de réserve chez Arabidopsis en interagissant avec d'autres facteurs de transcription (Kagaya et al., 2005). Ces données suggèrent que Psat5g165120 est un bon candidat pour réguler la synthèse des protéines de réserve dans les graines de pois. Parmi les autres régulateurs dans le top20 des plus faibles « *P*-values » obtenues en condition S<sup>-</sup> et/ou pour le  $\beta$  sont 12 régulateurs exprimés pendant le remplissage des graines, donc susceptibles de contrôler ce processus en réponse à un manque de soufre. D'autres gènes dans le top20 des plus faibles « *P*-values » au profil d'expression plus tardif lors du développement de la graine (cluster maturation, gènes surlignés en vert dans le tableau S13) pourraient jouer un rôle répresseur de l'accumulation des protéines de réserve. C'est le cas de Psat3g107600 dont le plus proche homologue chez Arabidopsis (At2g46020, BRAHMA chromatin-remodeling ATPase) réprime dans les feuilles l'expression des gènes codant les protéines de réserve spécifiques des graines (Tang et al., 2008). Si ce facteur de remodelage de la chromatine contrôle l'arrêt de la synthèse des protéines de réserve aux stades tardifs du développement de la graine reste à déterminer.

## 2.5 Gènes candidats pour moduler la composition protéique des graines dans des environnements pauvres en soufre

Afin d'identifier les processus biologiques impliqués dans la composition protéique des graines de pois en condition contrôle, de carence en soufre et pour le  $\beta$ , des tests d'enrichissement en ontologies ont été effectués à partir des différentes listes de gènes. Certaines ontologies sur-représentées (Figure 33, seuil de  $P$ -value à 0.01) sont identifiées dans différentes conditions et catégories (globulines, 7S/11S et albumine PA2) comme les ontologies **de transport, de modifications post-traductionnelles des protéines et le métabolisme du soufre** (en surbrillance dans la Figure 33). Dans chacune de ces classes fonctionnelles, nous avons mis en évidence des gènes susceptibles de jouer un rôle important en réponse au manque de soufre, c'est-à-dire identifiés à partir du  $\beta$  et/ou en condition de carence en soufre. Nous mettons ci-après l'accent sur ces gènes (Tableaux 8, S14 et S15, cf. tableau S12 pour la liste complète des gènes après exclusion de ceux seulement identifiés en  $S^+$  ou en  $S^+$  et  $\beta$ ).

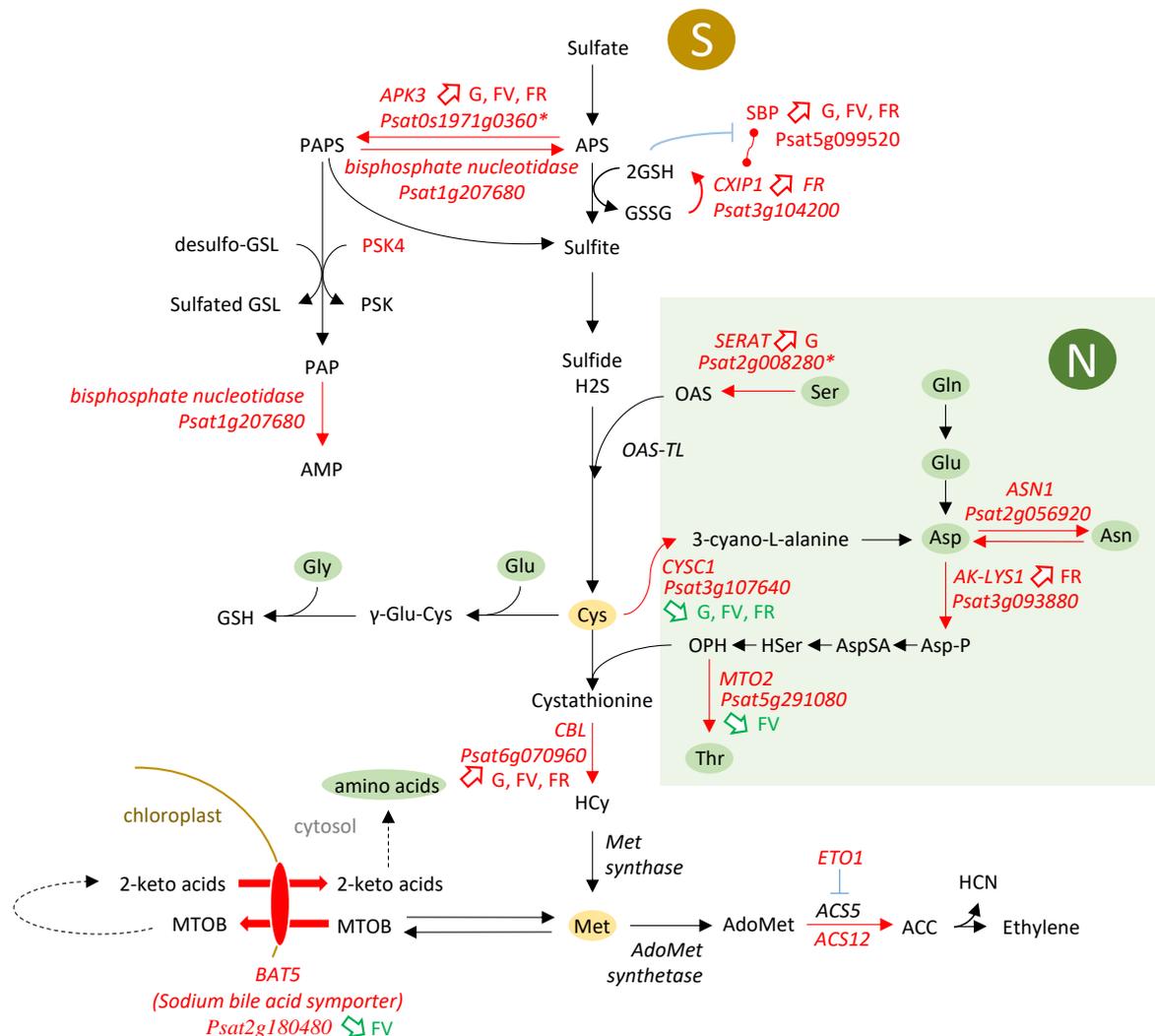
### 2.5.a Gènes candidats associés au métabolisme du soufre

L'ontologie « biosynthèse de cystéine à partir de sérine » étant enrichie dans la liste de gènes identifiés par GWAS pour le  $\beta$  du ratio 7S/11S (Figure 33), les gènes associés au métabolisme du soufre ont été sélectionnés (Tableau 8) et replacés dans les voies correspondantes (Figure 34).

**Tableau 8 : Gènes candidats associés au métabolisme du soufre.**

Nom	Chr	Pos (Mbp)	S <sup>+</sup>	S <sup>-</sup>	β	Orthologue Arabidopsis	Orthologue description	Ratio S- sur S+ d'expression en réponse à une carence en soufre (log2)						
								Graines 9		Feuilles noeuds repro			Feuilles noeuds vég	
								JAP	T.0	T.2	T.5	T.9	T.12	
<b>Cluster embryogenèse</b>														
Psat6g223200	chr6LG2	440	✓	✓	✓	AT2G01140.1 FBA					-0.33			
Psat5g099520	chr5LG3	180	✓			AT4G14040.1 SBP	0.62	2.08	1.3	2.19	1.44	1.36	1.3	
Psat6g070960	chr6LG2	89	✓			AT3G57050.1 CBL	0.6	1.36	0.96	1.04	1.01	0.78	0.76	
Psat2g008280*	chr2LG1	7		✓		AT3G13110.1 SERAT2	0.52	0.07	0.99	0.25	0.16	-0.37	1.61	
Psat2g008320	chr2LG1	7		✓		AT3G13110.1 SERAT2				-0.84	-0.74		-1.01	-0.77
Psat5g224760	chr5LG3	450	✓			AT1G30400.1 MRP1								0.67
<b>Cluster transition</b>														
Psat3g107640	chr3LG5	212	✓	✓		AT3G61440.1 CVSC1	-0.27	-1.3		-1.23	-0.65	-0.71	-1.18	
Psat0s1971g0360*	scaffold0197	0	✓	✓		AT3G03900.1 APK3	0.8	2.44	1.92	2.72	1.89	1.99	1.97	
Psat1g062160	chr1LG6	97	✓			AT5G15860.1 PCME			-0.25					
Psat1g207680	chr1LG6	358	✓			AT5G54390.1 HL (PAP phosphatase)								
Psat5g157400	chr5LG3	288	✓			AT3G51770.2 ETO1						0.59		
<b>Cluster remplissage</b>														
Psat6g214400	chr6LG2	428	✓			AT3G49780.1 PSK4								
<b>Cluster maturation</b>														
Psat3g093880	chr3LG5	191	✓	✓		AT5G13280.1 AK-LYS1		0.4						
Psat2g056920	chr2LG1	100	✓			AT3G47340.1 ASN1								
Psat5g291080	chr5LG3	558	✓			AT4G29840.1 MTO2								
<b>Cluster Emb_Mat</b>														
Psat3g104200	chr3LG5	205	✓	✓		AT3G54900.1 CXIP1			0.31					
<b>Non différentiellement exprimés lors du développement des graines</b>														
Psat2g180640	chr2LG1	420	✓			AT5G51690.1 ACS12								
Psat2g008240	chr2LG1	7		✓		AT3G13110.1 SERAT2								

Seuls les gènes identifiés en  $S^-$  et en  $\beta$  sont listés (les gènes identifiés uniquement en  $\beta$  et en  $S^+$  sont exclus). Des astérisques ont été ajoutés à la suite des gènes lorsqu'ils possèdent une variation allélique au sein de la population induisant une modification d'acide aminé : une étoile pour un effet modéré et deux étoiles pour un effet fort. Les coches marrons correspondent aux gènes identifiés pour les globulines, les coches bleues correspondent aux gènes identifiés pour l'albumine et les coches sont soulignées lorsque les gènes sont identifiés pour le ratio 7S/11S.



**Figure 34 : Gènes candidats associés ou connectés aux métabolismes du soufre (S) et de l'azote (N).** Les gènes de pois identifiés par GWAS dans notre analyse sont indiqués en rouge. Les flèches indiquent le sens de variation de l'expression du gène en condition de carence en soufre dans les graines à 9 DAP (G), dans les feuilles des nœuds végétatifs (FV) et dans les feuilles des nœuds reproducteurs (FR) (plus de détails sur les données transcriptomiques utilisées dans le matériels et méthodes). En vert, les acides aminés non sulfurés, en jaune les acides aminés sulfurés. Une étoile à côté du nom du gène indique que celui-ci possède une variation d'allèle induisant un changement d'acide aminé. APK3, adenylyl-sulfate kinase 3 ; SBP, selenium binding protein ; CXIP1, CAX interacting protein 1 ; PSK4, phytosulfokine 4 precursor ; SERAT, serine O-acetyltransferase ; ASN1, glutamine-dépendant asparagine synthetase ; AK-LYS1, aspartate kinase 1 ; CYSC1, cystéine synthase C1 ; MTO2, méthionine over-accumulator 2 ; CBL, cystathionine beta-lyase ; BAT5, bile acid transporter 5 ; ETO1, ethylene overproducer 1 ; ACS12, 1-amino-cyclopropane-1-carboxylate synthase 12.

Ce travail a permis d'identifier des points de contrôle métaboliques susceptibles de contribuer à réguler les flux de soufre et d'azote vers la synthèse de cystéine et de méthionine. Trois de ces gènes, en cluster sur le chromosome 2 (Psat2g008240, Psat2g008280 et Psat2g008320), ont été identifiés pour le  $\beta$  du ratio 7S/11S. Ils correspondent à des isoformes de serine O-acétyltransférase, une enzyme impliquée dans la formation d'O-acétylsérine (OAS) à partir de sérine. L'OAS est une molécule clé au carrefour entre les métabolismes du soufre et de l'azote car elle est synthétisée à partir de sérine et utilisée pour produire la cystéine. Le gène

Psat2g008240 n'est pas exprimé différentiellement lors du développement de la graine alors que les gènes Psat2g008280 et Psat2g008320 sont exprimés préférentiellement durant l'embryogenèse. La régulation positive de Psat2g008280 en réponse à une carence en soufre dans les graines au stade de transition vers le remplissage, et l'identification d'un variant allélique induisant un changement d'acide aminé (Ile-Thr en position 1025), suggèrent que ce gène est un bon candidat pour contribuer à contrôler la disponibilité en OAS et ainsi la plasticité de la composition en globulines des graines. Un autre gène identifié par GWAS pour le  $\beta$  du ratio 7S/11S (Psat5g291080) est homologue au gène d'Arabidopsis *MTO2* (Methionine over-accumulator 2, AT4G29840) impliqué dans la synthèse de thréonine à partir de l'O-phosphohomoserine. Ce gène a été nommé ainsi car son inactivation oriente les voies du métabolisme de l'aspartate vers la synthèse de méthionine, conduisant à une sur-accumulation de méthionine (Kusano et al., 2010). Le gène de Psat5g291080 s'exprime préférentiellement lors de la maturation des graines de pois, suggérant qu'il puisse contribuer à réguler le flux d'azote entre les métabolismes de la thréonine (précurseur de l'isoleucine) et de la méthionine à ces stades.

Deux autres gènes présentés dans la figure 34 sont susceptibles de contribuer à réguler le métabolisme des acides aminés dérivés de l'aspartate. L'un correspond à une aspartate kinase (Psat3g093880, AK-LYS1) identifiée à partir des données obtenues pour les globulines en S<sup>-</sup> et  $\beta$ , et l'autre à une asparagine synthétase (Psat2g056920) identifiée pour le  $\beta$  du ratio 7S/11S. Un autre résultat intéressant est l'identification du gène Psat6g070960 à partir des données de ratio 7S/11S en condition S<sup>-</sup>. Ce gène code l'enzyme cystathionine beta-lyase qui catalyse la synthèse d'homocystéine, le précurseur direct de la méthionine, à partir de cystathionine. Son expression augmente en réponse à une carence en soufre dans les graines et les feuilles de pois, suggérant que la biosynthèse *de novo* de méthionine soit activée dans cette condition. Cette enzyme pourrait donc contribuer à fournir de la méthionine pour la synthèse des protéines dans les graines de pois produites dans des environnements pauvres en soufre. Un gène de pois (Psat3g107640) codant une cystéine synthase homologue à CysC1 (At3g61440) a aussi été identifié par GWAS pour les globulines en S<sup>-</sup> et S<sup>+</sup>. Chez Arabidopsis, CysC1 (aussi appelé  $\beta$ -Cyano-alanine synthase) est principalement impliquée dans la réaction de synthèse de 3-Cyano-L-alanine, un intermédiaire de la voie de détoxification des cyanides, à partir de cystéine (Most and Papenbrock, 2015). Ce composé intermédiaire est connecté aux métabolismes du soufre et de l'azote car il peut être utilisé pour la synthèse d'aspartate et d'asparagine *via* NIT4 (nitrilase 4). Ce gène est régulé négativement dans les feuilles et les graines en réponse à une carence en

soufre, probablement pour privilégier l'utilisation de la cystéine pour la synthèse de méthionine ou pour son incorporation dans les protéines.

D'autres gènes identifiés par GWAS sont susceptibles de jouer un rôle plus indirect dans le contrôle du métabolisme du soufre. C'est le cas d'une glutaredoxine (Psat3g104200) dont l'homologue chez *Arabidopsis* est CXIP1 H<sup>+</sup>/Ca<sup>2+</sup> exchanger CAX-interacting protein-1, At3g54900). Cette enzyme a une activité de réduction du disulfide de glutathion (GSSG) en glutathion (GSH) mais aussi d'activation du transporteur de calcium CAX1. Il a été montré qu'un déficit en soufre déclenche la mobilisation du Ca<sup>2+</sup> dans le cytosol, permettant de lever au moins partiellement l'inhibition de la sérine acétyltransférase cytosolique, impliquée dans la synthèse d'O-acétylsérine, par la cystéine (Saito, 2000). Il en résulte une augmentation de l'activité de la sérine acétyltransférase et par conséquent de la production d'O-acétylsérine, le précurseur de la synthèse de cystéine. Il serait donc intéressant d'étudier le lien possible entre l'homologue de CXIP1 identifié chez le pois et la régulation par le calcium de la biosynthèse de cystéine. Par ailleurs, de par son rôle dans la réduction du GSSG en GSH, CXIP1 est susceptible de participer aux réactions d'oxydoréduction dépendantes du GSH (Cheng et al., 2006), telle que la voie de réduction de l'APS en sulfite *via* l'APS réductase (Figure 34). Il est à noter que l'expression de l'homologue de *CXIP1* chez le pois augmente en réponse à une carence en soufre dans les feuilles des nœuds reproducteurs, suggérant un rôle en lien avec le manque de soufre pour la synthèse de GSH. Il a aussi été montré que *CXIP1* interagit avec la protéine SBP1 (selenium-Binding Protein 1) pour répondre aux variations de potentiel redox (Valassakis, 2019; Valassakis et al., 2018). Une SBP détectée dans notre étude (Psat5g099520) par GWAS en condition S<sup>-</sup> (Tableau 8) est le plus proche homologue chez le pois de *SBP1* et *SBP2* chez *Arabidopsis*, et c'est le seul gène codant une SBP identifié par Blast dans le génome du pois. Chez *Arabidopsis*, l'expression de *SBP1* est fortement liée à la demande cellulaire en soufre et sa surexpression diminue la sensibilité aux stress oxydatifs (Hugouvieux et al., 2009). La synthèse de *SBP1* est inhibée lorsque les plantes sont alimentées avec du GSH et stimulée lorsque la synthèse de GSH est réduite, mettant en évidence un rôle inhibiteur du GSH sur l'induction de SBP (Hugouvieux et al., 2009). Ces résultats suggèrent un rôle antioxydant de *SBP1* dans des conditions de carence en soufre où la synthèse des composés soufrés tels que le glutathion est limitée. Le gène Psat5g099520 voit son expression augmenter dans les feuilles ainsi que dans les graines de pois en réponse à une carence en soufre, suggérant un rôle similaire chez le pois. Il est important de souligner que le glutathion est une des formes de soufre transportées dans le phloème jusqu'aux graines (Wongkaew et al., 2018). Il serait donc

intéressant d'étudier la possibilité de *SBPI* à prendre le relai d'antioxydant si le flux de glutathion vers les graines est activé par manque de soufre.

Deux gènes associés aux voies secondaires du métabolisme du soufre ont aussi été identifiés. Le premier gène code une adenosine-5'-phosphosulfate kinase (*APK*) (Psat0s1971g0360) connue pour transformer l'APS (adénylyl-sulfate) en PAPS (adénosine 3'-phosphate 5'-phosphosulfate), un donneur de soufre pour la sulfatation de divers composés tels que les flavonoïdes ou les phytosulfokines (Chan et al., 2019). Le gène Psat0s1971g0360 présente une variation allélique changeant une Thr en Pro (acide aminé 17) et a été identifié par GWAS pour les globulines en conditions S<sup>+</sup> et S<sup>-</sup>. Contrairement à ce qui a été observé pour les *APK* chez *Arabidopsis* (Chan et al., 2019), ce gène *APK* de pois est régulé positivement en réponse à une carence en soufre dans les feuilles et les graines. Ainsi, chez le pois, cette voie secondaire du métabolisme du soufre pourrait être activée dans ces tissus en réponse à un manque de soufre. Le PAPS ainsi formé via l'*APK* aurait aussi un rôle de stockage de l'APS visant à redonner de l'APS dans des conditions où le métabolisme primaire orienté vers la synthèse de cystéine et de méthionine doit être réactivé (Droux, 2004). De manière intéressante, un gène potentiellement impliqué dans la reconversion du PAPS en APS (Hell, 1997) a été identifié par GWAS pour les globulines et l'albumine PA2 en condition S<sup>-</sup> (Psat1g207680, Tableau 8). Ce gène code une bisphosphate nucléotidase, aussi impliquée dans la conversion du PAP en AMP par homologie avec *Arabidopsis* (At5g54390). Il a été montré que l'inactivation d'une telle phosphatase chez les plantes induit des modifications de l'expression des gènes impliqués dans les voies du métabolisme primaire du soufre de manière similaire à l'effet d'une carence en soufre (Lee et al., 2012). Psat1g207680 n'est pas régulé par une carence en soufre mais présente, comme l'*APK*, une expression préférentielle au stade de transition entre l'embryogenèse et le remplissage des graines. Les deux gènes identifiés par GWAS et impliqués dans l'interconversion APS/PAPS sont donc susceptibles de jouer un rôle clé dans l'orientation des flux de soufre soit en direction du métabolisme secondaire (*APK*), soit en direction du métabolisme primaire (bisphosphate nucléotidase) au sein même des graines.

### 2.5.b Gènes candidats pour le transport de molécules

Les ontologies liées au transport de molécules sont enrichies dans les listes de gènes obtenues pour les deux conditions de nutrition soufrée et pour la plasticité et ceci tant pour les globulines, le ratio 7S/11S que pour l'albumine PA2. Au total, 94 gènes codant des transporteurs ont été identifiés par GWAS et classés selon leur profil d'expression lors du développement de la graine de pois (cf. clusters d'expression dans le Tableau S14).

## Transport d'acides aminés

L'ontologie “**amino acid transmembrane transport**” est significativement surreprésentée pour le ratio 7S/11S en condition de carence en soufre. Les trois gènes (Psat1g207760, Psat3g194360, Psat5g303720) identifiés par GWAS à partir du ratio 7S/11S sont exprimés aux stades précoces du développement de la graine (embryogenèse - transition) et respectivement homologues à At1g47670, At3g30390, et At1g80510 dont le rôle exact reste à déterminer. Au total, 13 gènes codant des transporteurs d'acides aminés ont été identifiés par GWAS (Tableau S14) dont 10 sont distribués en clusters sur le génome : cluster I avec trois gènes en tandem sur le chromosome 1 (Psat1g064280, Psat1g064480, Psat1g064560), cluster II avec cinq gènes en tandem sur le chromosome 6 (Psat6g211120, Psat6g211600, Psat6g212840, Psat6g213240 et Psat6g213400), et cluster III avec 2 gènes en tandem sur ce même chromosome (Psat6g223720 et Psat6g223760).

Tous ces gènes sont impliqués dans le contrôle des convicilines (bandes 3 et 4 et leurs plasticités). Le plus proche homologue chez *Arabidopsis* des trois gènes du cluster I est At3g30390, un transporteur putatif d'acides aminés, tandis que le plus proche homologue des cinq gènes du cluster II est *AAP6* (amino acid permease 6), un transporteur de haute affinité capable de transporter l'aspartate et le tryptophane (Fischer et al., 2002) et potentiellement impliqué dans l'absorption des acides aminés à partir du xylème (Okumoto et al., 2002). Il convient de souligner que deux gènes de pois homologues à *AAP6* présentent des variations alléliques induisant une modification modérée d'acide aminé : un changement d'une Lys en Glu (acide aminé 345) pour Psat6g212840 et un changement de Gly en Ser (acide aminé 82) pour Psat6g211600. La position de ces gènes dans un hotspot de SNPs en bas du chromosome 6 (~ 420 Mpb, Figure 32) associés à l'accumulation des vicilines en S<sup>+</sup>, S<sup>-</sup> et  $\beta$ , et leur expression préférentielle durant l'embryogenèse (Tableau S14), suggèrent que ces transporteurs puissent contribuer à préparer le pool d'acides aminés nécessaire à l'accumulation des convicilines lors du remplissage, particulièrement dans des environnements carencés en soufre. Un de ces gènes, Psat6g213240, est préférentiellement exprimé dans les graines de pois à 12 JAP comparé aux autres tissus de la plante, suggérant que ce soit un bon candidat pour transporter les acides aminés à l'intérieur même de la graine (indice graine calculé à partir des données du Gene Atlas du pois d'Alves-Carvalho *et al.*, 2015, Table S10).

Les deux gènes du cluster III colocalisent aussi avec un hotspot de SNPs pour les variables de vicilines en bas du chromosome 6 (~ 440 Mpb, Figure 32). Ils sont homologues à *BATI*

(BIDIRECTIONAL AMINO ACID TRANSPORTER, At2g01170) décrit chez *Arabidopsis* comme pouvant transporter l'alanine, l'arginine, le glutamate et la lysine (Dündar and Bush, 2009). De plus, *BATI* s'exprime dans les tissus vasculaires (Choi et al., 2013), suggérant un rôle dans l'export d'acides aminés du phloème vers les tissus puits tels que les graines. L'expression des deux gènes de pois est régulée positivement dans les feuilles des nœuds végétatifs et/ou reproducteurs en réponse à une carence en soufre (Tableau S14), suggérant que ces transporteurs puissent contribuer à remobiliser le pool d'acides aminés dans ces conditions. De plus, Psat6g223720 est préférentiellement exprimé lors du remplissage des graines, et Psat6g223760 au stade de transition entre l'embryogenèse et le remplissage. Ils sont donc de bons candidats pour jouer un rôle à la fois dans la remobilisation du pool d'acides aminés au sein des feuilles et dans le transport des acides aminés dans la graine.

#### Transport de molécules en lien avec le métabolisme du soufre

La seconde ontologie identifiée pour le ratio 7S/11S en condition de carence en soufre, “**sodium transport**” (Figure 33), contient deux gènes en tandem sur le chromosome 2 (Psat2g180440 et Psat2g180480, Tableau S12). Leur plus proche homologue chez *Arabidopsis* est *BASS6* (BILE ACID SODIUM SYMPORTER 6, At4g22840), impliqué dans le transport de glycolate à partir des chloroplastes et dans la photorespiration (South et al., 2017). Les deux protéines de pois présentent aussi une forte homologie (~80%) avec *BAT5* (BILE ACID TRANSPORTER 5, aussi nommé *BASS5*, At4g12030). D'ailleurs, un blast à partir de la protéine Psat2g180480 dans TAIR (The Arabidopsis Information Resource) indique qu'il s'agit du plus proche homologue de *BAT5*. Ce transporteur, inclus dans la Figure 34 sur le métabolisme du soufre, permet l'import dans les chloroplastes du 4-methylthio-2-oxobutanoate (MTOB) produit à partir de méthionine (Gigolashvili et al., 2009). Dans les plastes, le MTOB peut être utilisé pour produire des 2-keto acides qui sont ensuite exportés dans le cytosol *via* *BAT5*. Dans le cytosol, les 2-keto acides peuvent être utilisés pour la synthèse des glucosinolates dérivés de la méthionine (Met-GSL), comme c'est le cas chez *Arabidopsis* (Gigolashvili et al., 2009; Sawada et al., 2009) et/ou pour la synthèse des acides aminés (Tashiro et al., 2015). Chez les légumineuses, qui ne produisent pas de glucosinolates, les 2-keto acides formés et transportés grâce à *BAT5* pourraient avoir un rôle prépondérant dans la synthèse des acides aminés, notamment dans les graines qui ne reçoivent du phloème que de l'asparagine et de la glutamine. L'homologue de *BAT5* chez le pois (Psat2g180480) présente une variation allélique induisant un changement modéré d'acide aminé, une Pro en Ser (acide aminé 326), et est exprimé préférentiellement durant la maturation des graines (23-29 JAP). L'ensemble de ces données fait de ce transporteur

un bon candidat pour réguler l'équilibre en acides aminés aux stades d'accumulation massive des protéines de réserve. Ce gène est aussi exprimé dans les feuilles des nœuds végétatifs (en sénescence) où il est négativement régulé par une carence en soufre. Cette régulation est probablement favorable à la régénération de méthionine dans les feuilles, le MTOB non transporté dans les plastes pouvant servir à régénérer de la méthionine *via* le cycle de Yang (« methionine salvage pathway »). En accord avec cette hypothèse, l'inactivation de *BAT5* chez *Arabidopsis* augmente le contenu en méthionine des feuilles (Sawada et al., 2009).

### Transporteurs MATE (Multidrug And Toxic Compound Extrusion)

On observe que l'ontologie “**drug transmembrane transport**” est sur-représentée dans la liste de gènes candidats pour les globulines en condition contrôle, carencée en soufre et pour leur indice de plasticité, avec respectivement 6, 3 et 6 gènes (Figure 33, Tableau S14). La majorité des gènes appartenant à cette ontologie code des protéines MATE décrites chez *Arabidopsis* comme transportant différents types de molécules (ex. métabolites secondaires, hormones, acides organiques) (Kusakizako et al., 2020). Ces transporteurs sont connus pour leur fonction détoxifiante, notamment en réponse aux stress biotiques et abiotiques, et leur contribution aux processus développementaux (Upadhyay et al., 2019). Dans le contexte du développement de la graine, un transporteur MATE nommé *BIG EMBRYO1*, a été montré comme coordonnant la croissance de l'embryon et de l'albumen (Suzuki et al., 2015). Deux des gènes *MATE* identifiés chez le pois possèdent des variations alléliques induisant un changement modéré d'acide aminé et semblent être de bons candidats pour jouer un rôle lors du développement des graines. Le premier, Psat3g097080 (SNP modifiant Phe en Leu, acide aminé 336), est préférentiellement exprimé lors du remplissage des graines. Le second, Psat6g221160 (SNP modifiant Ile en Val, acide aminé 223), est préférentiellement exprimé lors de la maturation des graines, positivement régulé en réponse à une carence en soufre au stade de transition entre l'embryogenèse et le remplissage (9 JAP, Tableau S12), et préférentiellement exprimé dans la graine (Tableau S10). Ces deux gènes sont localisés dans des hotspots spécifiques aux vicilines à ~195 Mpb sur le chromosome 3 (Psat3g097080) et aux convicilines ~438 Mpb sur le chromosome 6 (Psat6g221160). Psat3g097080 co-localise (à 195 Mpb sur le chromosome 3) aussi avec des PQL de vicilines précédemment identifiés chez le pois (Bourgeois *et al.*, 2011). Ces gènes ont été identifiés par GWAS à partir des données obtenues en conditions S<sup>+</sup>, S<sup>-</sup> et pour le  $\beta$ , suggérant qu'ils puissent jouer un rôle, qui reste à déterminer, dans l'ajustement de l'abondance des vicilines et convicilines, notamment dans les environnements pauvres en soufre.

## Trafic vésiculaire

Seize gènes appartenant aux ontologies de transport des protéines depuis leur synthèse dans le réticulum endoplasmique (RE) jusqu'aux vésicules de stockage (“**ER to Golgi vesicle-mediated transport**”, “**vesicle-mediated transport**”, “**exocytosis**”) ont été recensés dans notre liste de gènes candidats GWAS (Tableau S14). Quatorze ont été identifiés pour la plasticité du ratio 7S/11S et 11 sont différentiellement exprimés dans la graine en cours de développement.

Trois gènes identifiés pour la plasticité du ratio 7S/11S et potentiellement impliqués dans le transport des protéines du réticulum endoplasmique vers l'appareil de Golgi sont des candidats intéressants pour le contrôle de la mise en réserve des protéines. Le premier gène, Psat1g037840, est l'homologue de At1g51740 (*AtSYP81* ou *AtUfe1*), une protéine SNARE du RE ayant pour cible MAG2 (At3g47700) dont le mutant chez *A. thaliana* est impacté dans l'adressage des protéines de réserve à l'appareil de Golgi. Chez ce mutant *mag2*, les précurseurs des protéines de réserve 2S et 12S restent suraccumulés sous forme d'agrégats dans la lumière du réticulum endoplasmique (Li et al., 2006), le complexe SNARE\_MAG étant essentiel au transport des protéines de réserve vers l'appareil de Golgi. Les graines matures de *mag2* contiennent une proportion anormalement élevée de ces précurseurs non processés. Chez le pois, le gène *SYP81* est exprimé fortement tout au long du développement de la graine avec un pic en phase de maturation. Les deux autres gènes, Psat7g259720 et Psat5g304720, codent des récepteurs de type SNARE11. Psat7g259720 est exprimé préférentiellement durant la période de transition vers le remplissage tandis que Psat5g304720 est préférentiellement exprimé durant la période de maturation et de plus régulé positivement dans les feuilles des nœuds reproducteurs en réponse à une carence en soufre. Ces deux gènes sont homologues au gène Medtr4g006570 identifié chez *M. truncatula* par GWAS comme étant associé à la composition protéique des graines (Le Signor et al., 2017). Nos résultats confortent l'hypothèse selon laquelle ces protéines SNARE11 pourraient être impliquées dans le transport des globulines nouvellement synthétisées du RE vers l'appareil de Golgi chez les légumineuses et suggèrent qu'elles jouent également un rôle dans la plasticité de la composition protéique en réponse à l'environnement.

D'autres gènes sont impliqués dans le transfert des protéines post-Golgi avec, pour leurs homologues chez *A. thaliana*, un rôle dans le développement des organes reproducteurs comme Psat3g100360, protéine du complexe TRAPP (Garcia et al., 2020) ou Psat4g068520, protéine Sec1/munc18-like, homologue de *KEULE* (At1g12360) impliqué dans la cytokinèse lors de

l'embryogenèse (Wu et al., 2013). Bien que ces gènes soient préférentiellement exprimés pendant le remplissage de la graine, leur implication dans le contrôle de la composition protéique reste à démontrer. A l'opposé, le gène Psat7g191520, homologue de *VPS41* (At1g08190) codant pour une protéine clathrine recrutée par l'adaptateur AP-3 dans le système d'excrétion post-golgi (Asensio et al., 2013), a été mis en évidence par GWAS chez *M. truncatula* (Medtr7g111870) comme potentiellement impliqué dans le transport vésiculaire des enzymes nécessaires au processing des protéines de réserve (Le Signor et al., 2017). Ce gène a été identifié dans notre expérience en condition de nutrition carencée en soufre ; il est préférentiellement exprimé pendant la phase de remplissage des graines et surexprimé dans les graines carencées en soufre en début de remplissage. Existe t'il en condition de carence en soufre, un besoin accru d'enzymes de protéolyse au moment du remplissage de la graine pour maintenir la composition protéique de la graine ?

### 2.5.c Gènes candidats associés aux modifications post-traductionnelles des protéines

Deux ontologies, “**protein phosphorylation**” et “**post-translational protein modification**” sont sur-représentées seulement en condition de carence en soufre et en  $\beta$  pour les globulines. Les gènes liés à ces ontologies sont présentés dans le Tableau S15.

Plusieurs de ces gènes candidats sont susceptibles de contribuer à moduler la composition en globulines des graines dans des environnements carencés en soufre puisqu'ils ont été identifiés pour la condition  $S^-$  et/ou pour le  $\beta$ . Le gène Psat3g018280 a retenu notre attention. Il a été identifié par GWAS à partir des variations d'abondance des globulines et du ratio 7S/11S en condition  $S^+$  et pour les globulines en condition  $S^-$ , et présente une variation allélique induisant un changement d'acide aminé à effet modéré, de Phe en Leu (acide aminé 311). Son plus proche homologue chez *Arabidopsis* (At5g26751) code une « SHAGGY-related kinase » (*SK11*) qui contrôle la répartition du carbone entre les puits zygotiques et maternels dans les graines en régulant la phosphorylation de TRANSPARENT TESTA GLABRA1 (*TTG1*) (Li et al., 2018). Autre fait intéressant, *TTG1* code un facteur de transcription à domaine WD40 connu pour réguler négativement l'accumulation des protéines de réserve des graines chez *Arabidopsis* (Chen et al., 2015). Bien que Psat3g018280 fasse partie du cluster de gènes préférentiellement exprimés pendant l'embryogenèse des graines de pois, son niveau d'expression reste élevé jusqu'à 23 JAP (Tableau S12), stade d'accumulation massive des protéines de réserve. Ces résultats suggèrent que Psat3g018280 puisse contribuer aux mécanismes de régulation post-traductionnels nécessaires à l'accumulation des protéines de réserve dans la graine de pois en condition  $S^+$  et  $S^-$ . Par ailleurs, il a été montré que *SK11* est induit par les stress abiotiques chez

Arabidopsis, et active la glucose-6-phosphate dehydrogenase (G6PD) essentielle au maintien de l'équilibre redox (Dal Santo et al., 2012). Ce gène est régulé positivement en condition de carence en soufre dans les feuilles des nœuds végétatifs et reproducteurs, suggérant qu'il puisse aussi contribuer aux mécanismes de tolérance vis-à-vis d'une carence en soufre.

Un autre gène candidat pour le maintien de l'état redox est Psat6g211680 dont le plus proche homologue chez Arabidopsis (AT4G05200, CRK25) appartient à la famille des CRK (Cysteine-rich Receptor-like protein Kinase) (Tableau S15). En effet, les CRKs sont connues comme participant à la perception et signalisation des formes actives de l'oxygène et systèmes redox (Bourdais et al., 2015). L'expression de Psat6g211680 est régulée positivement en réponse à une carence en soufre dans les feuilles des nœuds végétatifs et reproducteurs, ainsi que dans les graines, suggérant un rôle dans le maintien du statut redox au sein de ces tissus. Les formes actives de l'oxygène (ROS) jouent un rôle central dans la régulation du développement et de la maturation des graines, mais leur concentration doit être rigoureusement contrôlée. En effet, si l'équilibre redox n'est pas maintenu et la production de ROS augmente, il en résulte un stress oxydatif défavorable aux processus développementaux (Bailly and Kranner, 2011). La surexpression d'enzymes ayant une action d'élimination des ROS peut avoir un effet bénéfique sur la résistance aux stress, à l'image de la surexpression d'une superoxyde dismutase qui permettrait d'avoir une meilleure résistance au stress hydrique (McKersie et al., 1996). Il a été montré que des gènes codant pour des enzymes ayant une action antioxydante sont régulés positivement en réponse à une carence en soufre (Higashi et al., 2006; Nikiforova et al., 2003). Ce serait une conséquence de la faible accumulation de glutathion (GSH) dans des conditions de carence en soufre, ce qui augmente le stress oxydatif (Nikiforova et al., 2003). Des données protéomiques obtenues au laboratoire, dans le cadre de la thèse de Charlotte Henriët, ont mis en évidence un réseau de protéines antioxydantes susceptibles de jouer un rôle clé lors du développement de la graine de pois. L'abondance de ces protéines antioxydantes augmente significativement en réponse à une carence en soufre, soulignant l'importance des systèmes anti-oxydants pour le développement des graines dans ces conditions (Henriët et al., 2021). Les gènes codant les protéines antioxydantes identifiées par Henriët et al. (2021) ne sont pas dans la liste de gènes issus de GWAS, suggérant que ces protéines ne régulent pas de façon majeure les mécanismes qui sous-tendent l'accumulation des protéines de réserve.

### 3. Variabilité et contrôle génétique des composantes du rendement au sein d'une collection de pois sous contrainte soufrée.

#### 3.1 Effets de la carence en soufre sur les composantes du rendement

L'effet de la carence en soufre et du génotype sur la variabilité des composantes du rendement au sein de la collection d'écotypes de pois a été étudié pour les récoltes 2018 et 2019. Ces composantes sont le nombre de gousses (nbgo), le poids total de gousses (pdsgo), le poids d'une gousse (pds1go), le nombre de graines (nbgr), le poids total de graines (pdsgr), le poids d'une graine (pds1gr) et le nombre de graines par gousse (nbgrgo). Des variables liées aux parties végétatives ont également été prises en compte : la hauteur à la récolte (nommée hauteur), le poids sec des parties aériennes végétatives (PA), le nombre de nœuds végétatifs (nb\_nodes\_veg) et reproducteurs (nb\_nodes\_repro). D'autre part, la teneur protéique (tprot) des graines a été acquise par la méthode NIRS (Near-infrared spectroscopy) sur les poudres issues des graines de la récolte 2018. Les moyennes et les écart-types pour chacune de ces variables ainsi que l'effet de la carence en soufre exprimée en pourcentage de la condition contrôle sont présentés par année dans le Tableau 9.

On observe que toutes les variables analysées sont significativement affectées par la carence en soufre, excepté le nombre de nœuds végétatifs car ceux-ci ont été formés avant la mise en carence. Parmi les variables affectées par la carence en soufre, seule la teneur en protéines augmente ; toutes les autres diminuent, et ce avec des intensités différentes. Ces résultats sont cohérents avec ceux de la littérature qui montrent un effet négatif de la carence en soufre sur des variables de rendement et positif sur la teneur en protéines des graines (Henriet et al., 2019b; Inal et al., 2003).

Une analyse de variance (ANOVA) en modèle mixte réalisée sur l'ensemble des données a permis d'évaluer les effets fixes environnement (E, combinaison année et condition de nutrition), année\*Bloc, année\*Sous-bloc fixes et les effets aléatoires génotype (G) et interaction GxE pour toutes les variables liées aux parties végétatives et aux composantes de rendement (Tableau 10, partie gauche). L'analyse des résidus de cette analyse de variance a mis en évidence qu'une transformation des variables était nécessaire pour assurer la normalité, l'indépendance et l'homoscédasticité des résidus (Annexe 3).

**Tableau 9 : Effet d'une carence en soufre sur des variables d'architecture et de rendement du pois.**

Architecture de la plante	S <sup>+</sup>				S <sup>-</sup>				Effet carence en soufre	
	2018		2019		2018		2019		2018	2019
	Moyenne	T SD	Moyenne	T SD	Moyenne	T SD	Moyenne	T SD		
Hauteur	237.52	d 90.57	211.85	c 94.81	177.54	b 83.59	154.09	a 90.77	-0.25	-0.27
Parties végétatives	78.08	c 67.62	47.72	b 44.44	29.6	a 27.95	30.3	a 38.18	-0.62	-0.37
Nombre de nœuds végétatifs	21.47	b 8.46	18.23	a 8.11	21.1	b 7.69	18.31	a 8.14	-0.02	0.004
Nombre de nœuds reproducteurs	13.51	d 5.93	11.57	c 5.89	8.19	b 4.98	6.35	a 3.33	-0.39	-0.45
<b>Rendement</b>										
Nombre de gousses (nbgo)	119.45	d 95.35	84.44	c 77.71	36.51	b 33.09	35.3	a 46.97	-0.69	-0.58
Poids de gousses (pdsgo)	83.61	d 52.37	57.58	c 37.16	19.03	b 13.73	20.43	a 21.91	-0.77	-0.65
Poids d'une gousse (pds1go)	0.84	b 0.34	0.87	b 0.39	0.66	a 0.32	0.69	a 0.35	-0.21	-0.21
Nombre de graines (nbgr)	475.3	d 394.1	327.22	c 291.24	113.9	b 107.93	114.74	a 155.85	-0.76	-0.65
Poids de graines (pdsgr)	68.63	d 44.05	48.14	c 31.79	14.42	b 10.83	15.3	a 17.5	-0.79	-0.68
Poids d'une graine (pds1gr)	0.18	b 0.08	0.19	b 0.09	0.17	a 0.08	0.16	a 0.08	-0.06	-0.16
Nombre de graines par gousse (nbgrgo)	3.93	b 0.94	3.91	b 0.87	3.15	a 0.99	3.18	a 0.96	-0.20	-0.19
<b>Caractéristiques de la graine</b>										
Teneur protéique (tprot)	22.23	a 2.42			30.31	b 2.39			0.36	

Moyennes et écarts types de chacune des variables par condition et par année et effet de la carence en soufre exprimée en pourcentage par rapport à la condition S<sup>+</sup> de l'année correspondante. Les résultats des tests de Tuckey réalisés à la suite d'une ANOVA (modèle ~G + E + G\*E par année) sont indiqués par des lettres allant de a à d. L'effet négatif de la carence en soufre (exprimé par le ratio (S<sup>+</sup> - S<sup>-</sup>)/S<sup>+</sup>) est d'autant plus fort que la case est rouge.

**Tableau 10 : Analyse de l'effet génotypique et environnemental sur des composantes du rendement et des variables liées aux parties végétatives aériennes.**

	Anova modèle mixte sur l'ensemble des données												Anova modèle mixte par année et condition													
	Environnement			Année*Bloc			Année*Sous-bloc			Effets aléatoires			2018_S+			2019_S+			2018_S-			2019_S-				
	SS	ddl	F	SS	ddl	F	SS	ddl	F	VarG	VarGxE	VarRes	%VarG	%VarGxE	VarG	VarRes	h <sup>2</sup>	VarG	VarRes	h <sup>2</sup>	VarG	VarRes	h <sup>2</sup>	VarG	VarRes	h <sup>2</sup>
√(Hauteur_totale)	753.78	3	259.26**	25.72	5	5.31**	10.53	6	1.81 <sup>ns</sup>	8.62	0.59	0.97	89%	6%	6.96	1.39	0.83	6.94	1.51	0.82	8.84	0.61	0.94	9.49	1.23	0.88
√(nb_nodes_veg)	0.59	3	0.91 <sup>ns</sup>	1.33	5	1.23 <sup>ns</sup>	3.56	6	2.74*	0.59	0.02 <sup>ns</sup>	0.22	73%	2%	0.61	0.24	0.71	0.53	0.28	0.65	0.55	0.17	0.76	0.63	0.21	0.75
log(nb_nodes_repro)	68.42	3	126.79**	1.68	5	1.87 <sup>ns</sup>	2.91	6	2.70*	0.06	0.07	0.18	22%	27%	0.08	0.16	0.32	0.05	0.20	0.20	0.13	0.22	0.38	0.21	0.15	0.58
log(PA)	206.63	3	218.65**	5.65	5	3.59**	3.13	6	1.66 <sup>ns</sup>	0.87	0.06	0.32	72%	5%	0.68	0.26	0.72	0.59	0.38	0.60	0.87	0.22	0.80	0.98	0.66	0.60
log(nbgo)	259.09	3	293.03**	2.13	5	1.45 <sup>ns</sup>	3.33	6	1.88 <sup>ns</sup>	0.51	0.12	0.29	61%	14%	0.55	0.24	0.69	0.50	0.33	0.60	0.58	0.22	0.73	0.67	0.47	0.59
log(pdsgo)	379.28	3	344.08**	4.60	5	2.50*	3.95	6	1.79 <sup>ns</sup>	0.26	0.17	0.37	39%	26%	0.37	0.26	0.59	0.24	0.35	0.41	0.32	0.30	0.52	0.76	0.57	0.57
log(pds1go)	15.32	3	95.19**	0.86	5	3.21**	0.23	6	0.71 <sup>ns</sup>	0.20	0.02	0.05	77%	8%	0.17	0.04	0.79	0.25	0.04	0.88	0.24	0.05	0.82	0.35	0.06	0.85
log(nbgr)	387.84	3	345.53**	3.38	5	1.81 <sup>ns</sup>	3.24	6	1.44 <sup>ns</sup>	0.54	0.14	0.37	57%	15%	0.67	0.22	0.75	0.53	0.37	0.59	0.60	0.33	0.64	0.76	0.59	0.56
log(pdsgr)	445.19	3	361.81**	4.45	5	2.17 <sup>ns</sup>	4.16	6	1.69 <sup>ns</sup>	0.26	0.21	0.41	36%	29%	0.38	0.25	0.60	0.25	0.36	0.41	0.36	0.35	0.51	0.93	0.65	0.59
√(pds1gr)	0.11	3	64.11**	0.01	5	3.50**	0.00	6	1.38 <sup>ns</sup>	0.01	0.00	0.00	92%	8%	0.01	0.00	0.95	0.01	0.00	0.95	0.01	0.00	0.98	0.01	0.00	0.93
log(nbgrgo)	12.69	3	97.30**	0.42	5	1.92 <sup>ns</sup>	0.49	6	1.90 <sup>ns</sup>	0.03	0.01	0.04	42%	13%	0.05	0.01	0.77	0.04	0.02	0.65	0.04	0.07	0.35	0.08	0.07	0.54

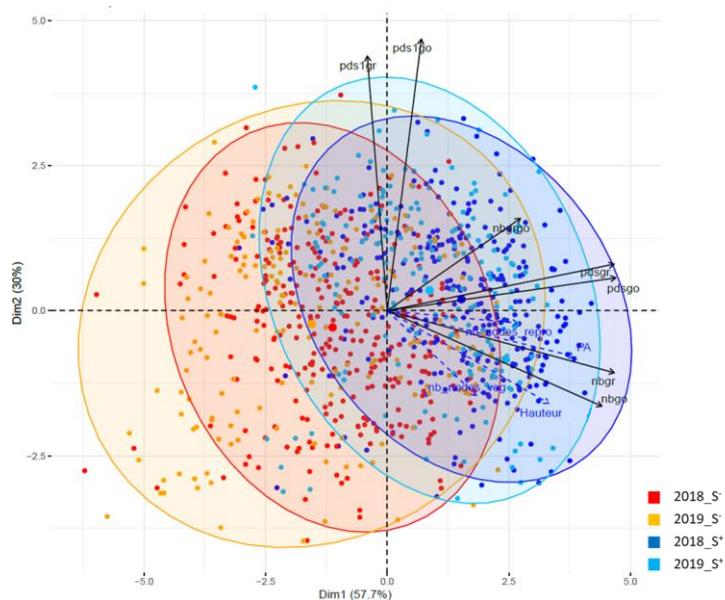
\* P-values entre 0.01 et 0.05 %, \*\* P-value < 0.01, ns : P-value > 0.05  
Number of obs: 1260, groups: Accessionenv, 962; Accession, 300

Partie gauche de la table : ANOVA (fonction lmer) à effets fixes (environnement, année\*Bloc, année\*Sous-bloc) et aléatoires (génotype et interaction génotype\*environnement). Un environnement est la combinaison année et condition de nutrition soufrée. Les sommes de carrés de type III (SS), degrés de liberté (ddl) et statistiques de Fisher (F) sont indiqués pour les effets fixes. Les estimations des variances sont indiquées pour les effets aléatoires ; toutes sont significativement différentes de zéro à l'exception de la variance GxE pour le nombre de noeuds végétatifs. La part des variance génétique (%VarG) et d'interaction (%VarGxE) dans la variation totale (VarRes + VarG + ¼VarGxE) est indiquée. Partie droite de la table : ANOVA (fonction lmer) à effets fixe (cuve) et aléatoire (génotype) pour chaque environnement. Les héritabilités sont estimées par le ratio VarG/(VarG+VarRes). Pour chaque analyse les variables ont été transformées pour satisfaire aux hypothèses de l'ANOVA, soit racine carrée soit logarithme (comme indiqué dans la première colonne).

Une transformation racine carrée a été retenue pour la hauteur, et pour le poids d'une graine et une transformation en logarithme pour le poids sec des parties végétatives, le nombre de nœuds reproducteurs, le nombre de gousses, le poids total de gousses, le poids d'une gousse, le nombre de graines, le poids total de graines et le nombre de graines par gousse. L'effet environnement (année\*condition de nutrition) est significatif pour toutes les variables excepté le nombre de nœuds végétatifs. Parmi les effets du dispositif expérimental, l'effet bloc par année est significatif pour cinq variables seulement (hauteur, PA, podsgo, pds1go et pds1gr) et l'effet

sous-bloc par année est significatif seulement pour les nombres de nœuds. Pour les analyses multivariées et génétiques ultérieures, les moyennes des variables ont été ajustées pour les effets dispositif significatifs. Les effets génétique et d'interaction GxE sont significatifs pour toutes les variables (les variances sont significativement différentes de zéro) exceptée l'effet GxE pour le nombre de nœuds végétatifs.

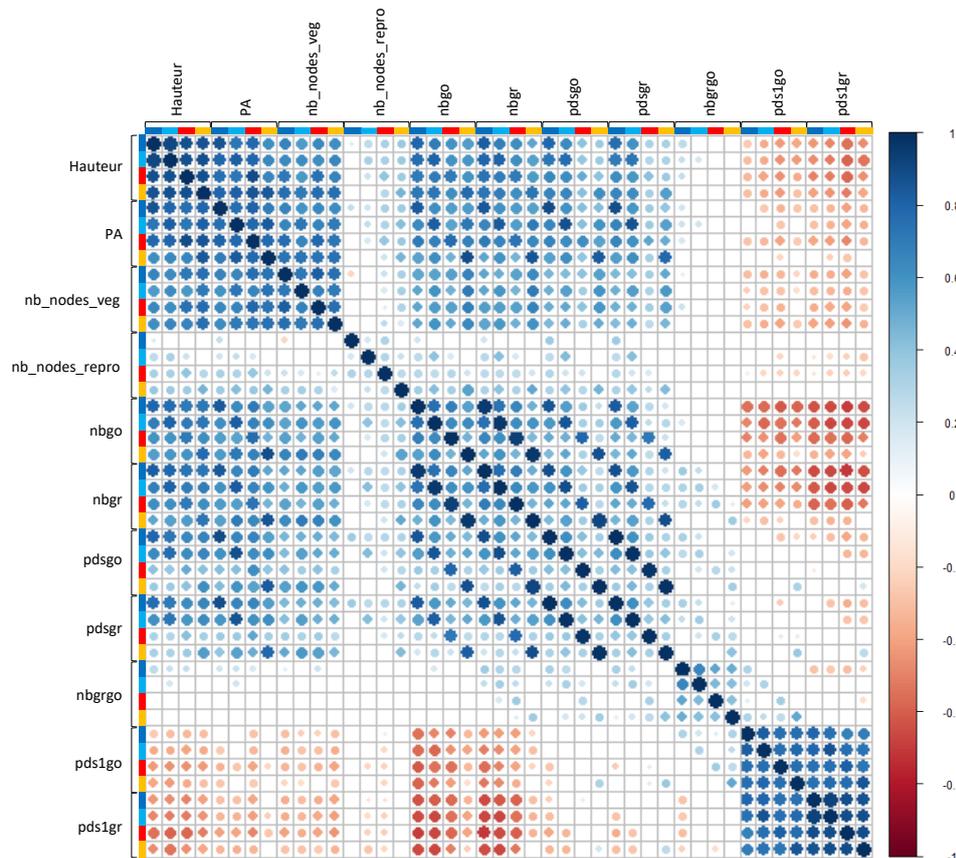
L'ACP sur les moyennes ajustées de chacune des variables de rendement transformées par génotype sépare les individus sur l'axe 1 en fonction du poids et du nombre des graines et de gousses et sur l'axe 2 en fonction du poids d'une graine et du poids d'une gousse (Figure 35).



**Figure 35 : Distribution des génotypes de pois selon les composantes du rendement.** Analyse en composante principale des génotypes pour les composantes de rendement. Les variables ont été transformées (racine carrée pour la hauteur, le poids d'une graine et nombre nœuds végétatifs et logarithme pour toutes les autres), 2018 S<sup>-</sup> ; orange, 2019 S<sup>-</sup> ; bleu foncé, 2018 S<sup>+</sup> ; bleu clair, 2019 S<sup>+</sup>. pds1gr, poids d'une graine ; pds1go, poids d'une gousse ; nbgrgo, nbre de graines par gousse ; pdsgr, poids de graines ; pdsgrgo, poids de gousses ; nbgr, nombre de graines ; nbgo, nombre de gousses. Sont indiquées en bleues les variables liées aux parties végétatives de la plante, nb\_nodes\_veg, nombre de nœuds végétatifs ; nb\_nodes\_repro, nombre de nœuds reproducteurs ; PA, parties végétatives ; Hauteur, hauteur de la plante à la récolte.

On observe que l'axe 1 discrimine bien les conditions S<sup>+</sup> et S<sup>-</sup> mais ne sépare pas distinctement les années, ce qui suggère que l'effet condition est prépondérant sur l'effet l'année. Ceci est confirmé par la décomposition de la somme de carrés totale (SST) d'une analyse de variance en effets fixes testant les effets année (>3% SS totale), condition (2 à 30% SS totale), G et G\*E sur les génotypes communs aux deux années.

Une heatmap des corrélations entre les variables d'architecture de la plante et les variables de rendement dans chacun des environnements (2018 S<sup>+</sup>, 2018 S<sup>-</sup>, 2019 S<sup>+</sup>, 2019 S<sup>-</sup>) est présentée figure 36.



**Figure 36 : Corrélations entre variables des parties végétatives du pois et celles des composantes du rendement en condition contrôle et de carence en soufre.** Heatmap des corrélations entre les variables d'architecture de la plante et les variables de rendement chaque année et dans chaque condition (2018 S<sup>+</sup>, bleu foncé ; 2019 S<sup>+</sup>, bleu clair ; 2018 S<sup>-</sup>, rouge ; 2019 S<sup>-</sup>, orange). pds1gr, poids d'une graine ; pds1go, poids d'une gousse ; nbgrgo, nbre de graines par gousse ; pdsgr, poids de graines ; pdsgo, poids de gousses ; nbgr, nombre de graines ; nbgo, nombre de gousses ; nb\_nodes\_veg, nombre de nœuds végétatifs ; nb\_nodes\_repro, nombre de nœuds reproducteurs ; PA, poids sec des parties aériennes végétatives. Les corrélations non significatives au seuil de p-value ajustée par Benjamni-Hochberg à 0.05 sont en blanc.

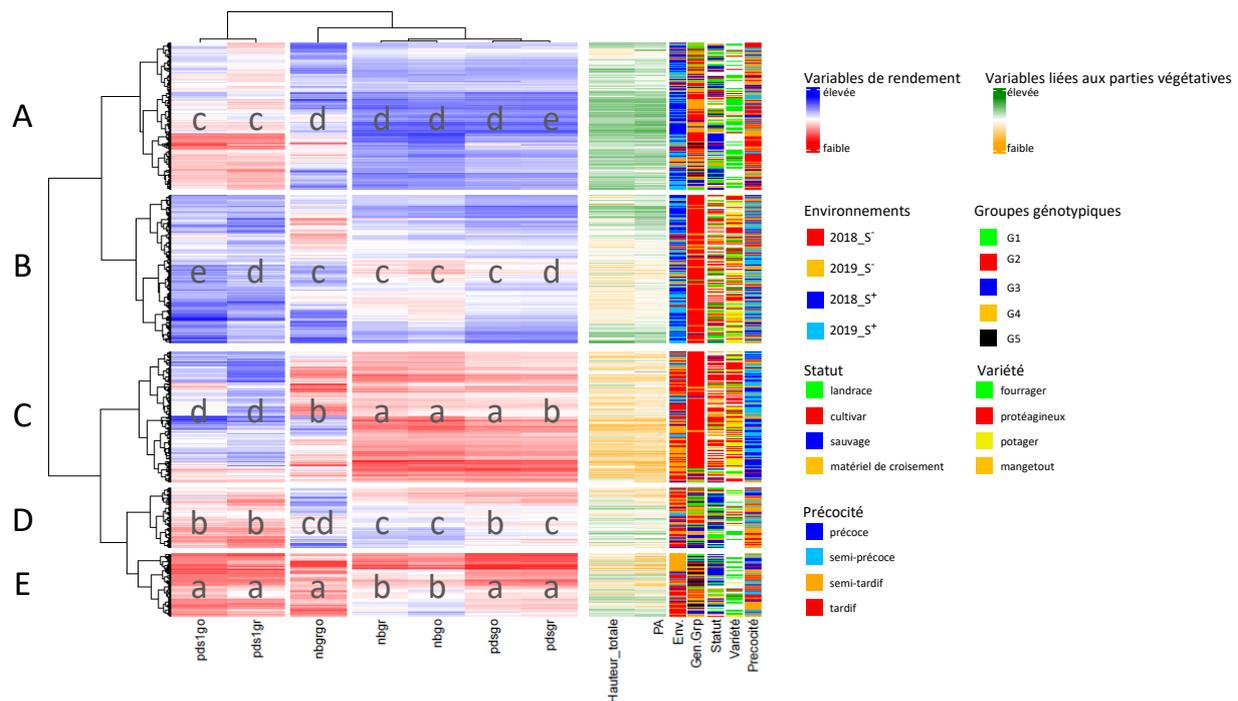
On observe que pour tous les environnements, les variables phénotypiques correspondant à la hauteur, au poids sec des parties végétatives ainsi qu'au nombre de nœuds végétatifs sont très corrélées entre elles alors que le nombre de nœuds reproducteurs ne l'est pas (il est très peu corrélé à l'ensemble des variables, incluant les composantes du rendement). Les composantes du rendement forment trois groupes de corrélations. Le premier est composé du nombre de gousses, du nombre de graines, du poids total de gousses et du poids total de graines. Le deuxième est composé uniquement du nombre de graines par gousse et le troisième est constitué du poids d'une gousse et du poids d'une graine. On observe que le premier groupe est corrélé positivement avec les variables liées aux parties végétatives alors que le troisième l'est plutôt négativement. De plus, on observe que le poids d'une gousse et le poids d'une graine sont négativement corrélés au nombre de gousses et au nombre de graines. On peut en déduire que de manière générale, pour toutes les conditions, plus les plantes sont grandes et développées, plus la quantité de gousses et de graines produites est élevée mais plus les gousses et les graines

sont petites. Ces données concordent avec celles de Burstin et al., 2007 suggérant l'existence de déterminismes génétiques communs entre les caractères d'architecture de la plante comme la hauteur et les composantes du rendement chez le pois.

## 3.2 Variabilité génétique pour la réponse à la carence en soufre

### 3.2.a Effet d'une carence en soufre sur la typologie de la collection

Afin de pouvoir comprendre comment la collection de pois est structurée en fonction des composantes du rendement en condition  $S^+$  et en condition  $S^-$ , un dendrogramme a été élaboré. Il a mis en évidence une structuration de la collection en cinq groupes distincts (A à E dans la Figure 37).



**Figure 37 : Structuration de la population de pois en fonction des composantes du rendement.** Dendrogramme construit à partir des variables de rendements avec  $k = 5$ . Plus les variables sont en bleu et plus leur valeur est élevée, plus les variables sont en rouge et plus leur valeur est faible. Les variables d'architecture de la plante sont illustratives, plus elles sont en vert et plus leur valeur est élevée et plus elles sont en orange et plus leur valeur est faible. Les résultats des tests de comparaison de moyenne sont indiqués sur le graphique par les lettres allant de a à e. Les différents environnements, les groupes génotypiques le statut, le type de variété, la culture et la précocité des plantes sont indiqués. pds1gr, poids d'une graine ; pds1go, poids d'une gousse ; nbgrgo, nbre de graines par gousse ; pdsgr, poids de graines ; pdsgr, poids de gousses ; nbgr, nombre de graines ; nbgo, nombre de gousses ; nb\_nodes\_veg, nombre de nœuds végétatifs ; nb\_nodes\_repro, nombre de nœuds reproducteurs ; PA, poids sec des parties aériennes végétatives.

Les groupes A et B sont très majoritairement constitués de plantes cultivées en condition  $S^+$  alors que les groupes C, D et E sont très majoritairement constitués de plantes cultivées en condition  $S^-$ . Les groupes B et C sont enrichis en *P. sativum ssp sativum*, le groupe D est enrichi

en en *P. sativum ssp abyssinicum*, les groupes A, D et E sont enrichis en *P. sativum ssp elatius* et le groupe E est en plus enrichi en *P. fulvum*.

Le groupe A est enrichi en pois fourragers cultivés en condition S<sup>+</sup>, représentés principalement par des génotypes sauvages et tardifs. Il est caractérisé par un nombre et un poids total de gousses et de graines élevés et par un nombre élevé de graines par gousse et un poids d'une graine et d'une gousse faibles. Les plantes présentes dans ce groupe ont une hauteur totale ainsi qu'un poids sec des parties végétatives très élevé.

Le groupe B est enrichi en variétés potagères, de pois protéagineux cultivés en condition S<sup>+</sup>, représenté majoritairement par des cultivars. Il est constitué de génotypes produisant des graines et des gousses de grosse taille en quantité modérée à élevée. Les plantes présentes dans ce groupe ont un poids sec des parties végétatives et une hauteur modérée voire faible par rapport au groupe A.

Le groupe C est similaire au groupe B pour la condition S<sup>-</sup>, enrichi en pois protéagineux représenté majoritairement par des cultivars et en lignées en cours de sélection. Les génotypes correspondants sont décrits comme précoces et semi-précoces. Il est caractérisé par un poids d'une graine et d'une gousse élevé. Il est aussi caractérisé par un nombre et un poids total de gousses faibles et un nombre de graines par gousse faible. Les plantes présentes dans ce groupe sont caractérisées par une hauteur et un poids sec de parties aériennes faibles.

Le groupe D est enrichi en pois fourragers, en variétés sauvages et « landrace » semi tardives, tous cultivés en condition S<sup>-</sup>. Il est caractérisé par un poids d'une gousse et un poids d'une graine relativement faibles et un nombre de graines par gousse élevé. Il s'agit de plantes ayant une hauteur élevée en comparaison du groupe C.

Le groupe E est enrichi en pois fourragers correspondant à des génotypes de type sauvages en condition S<sup>-</sup>. Il se distingue du groupe D par un nombre de graines par gousse faible.

On peut donc en déduire que malgré la carence en soufre, les génotypes de type protéagineux et potager restent ceux qui ont le poids d'une graine et le poids d'une gousse le plus élevé et que les génotypes fourragers restent ceux avec le nombre de graines le plus élevé. Néanmoins les génotypes fourragers sont scindés en deux groupes (D et E) en S<sup>-</sup>, sur la base du nombre de graines par gousse, alors qu'ils ne le sont pas en S<sup>+</sup> (A). Le dendrogramme met donc en évidence une différence de réponse à la carence en soufre très marquée parmi les génotypes de type fourrager et basée sur le nombre de graines par gousse.

### 3.2.b Estimation de l'héritabilité et des variance génétique et d'interaction GxE

L'héritabilité correspond au pourcentage de variance phénotypique expliqué par le génotype. Les calculs ont été faits au niveau de chaque condition, pour chaque dispositif, 2018 et 2019. L'héritabilité moyenne pour les variables de rendement est relativement stable d'un environnement à un autre : 0.73 en S<sup>+</sup> en 2018, 0.65 en S<sup>-</sup> en 2018, 0.64 en S<sup>+</sup> en 2019 et 0.66 en S<sup>-</sup> en 2019 (Table 10, partie droite). L'héritabilité est très différente en fonction des variables avec de très fortes valeurs pour la hauteur, le pds1gr et pds1go (0.8, 0.9) et des valeurs moyennes pour les autres composantes du rendement (0.53 pour le poids total de graines).

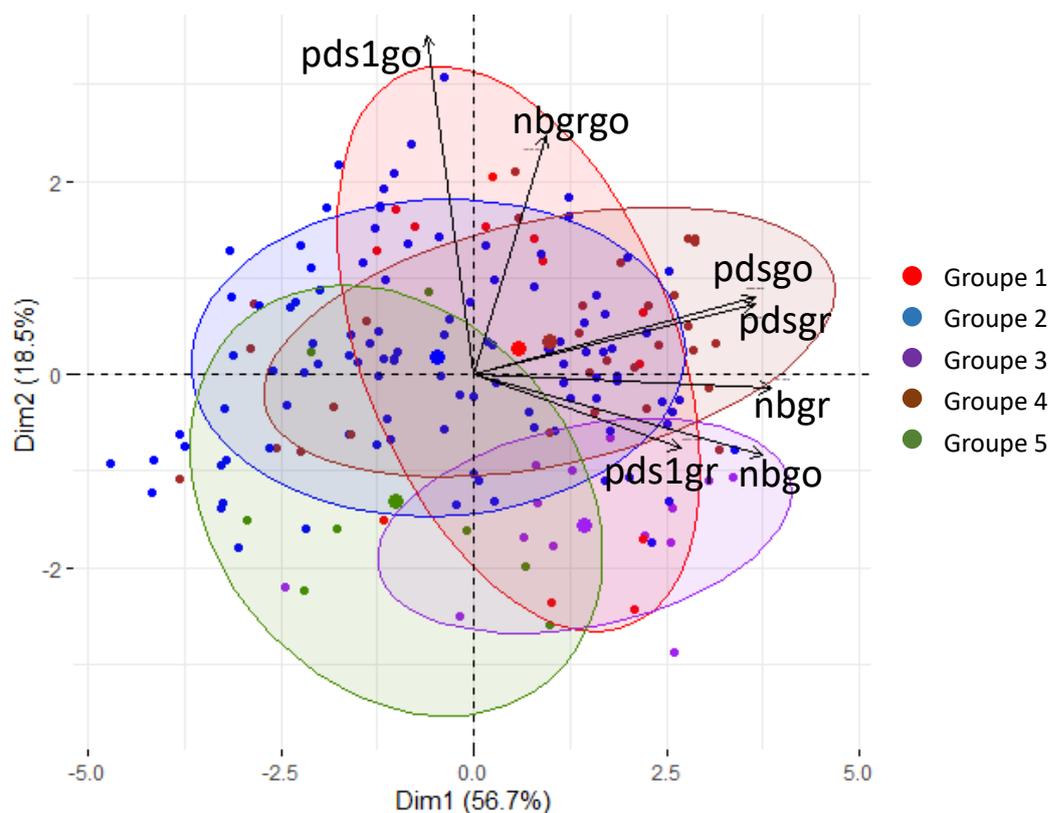
Les variances génétiques d'interaction ont été estimées au niveau du dispositif global (tableau 10). De la même façon que pour l'héritabilité, on observe deux groupes de variables. D'une part les variables phénotypiques (hormis nb\_nodes\_repro) et pds1go, pds1gr pour lesquelles la variance génétique explique près de ~80% de la variance totale alors que la variance d'interaction explique moins de 10% du total. D'autre part, les autres variables de composantes du rendement (nombre et poids de graines et gousses) ont une participation de la variance génétique à la variance totale de l'ordre 40-60 % compensée par une forte proportion de la variance d'interaction GxE (15 à 30%). La variable nb\_nodes\_repro est la moins héritable du dispositif comme le confirme la faible part de la variance génétique à la variance totale. L'ensemble de ces résultats met en avant l'importance de s'intéresser à la plasticité des variables de composantes de rendement en réponse à une carence en soufre.

### 3.2.c Plasticité des variables de rendement en réponse à une carence en soufre

Nous avons calculé les indices de plasticité  $\beta$  avec la méthode de Finlay Wilkinson de la même manière que pour la composition protéique, à partir des moyennes phénotypiques ajustées pour les effets dispositifs par environnement et sur la base des génotypes communs aux 4 environnements soit 173 génotypes.

#### Plasticité en fonction de la structure de la population

Une ACP a été réalisée à partir des indices de plasticité des variables de rendement en indiquant la répartition des individus dans les cinq groupes génétiques qui structurent la collection (Figures 38).

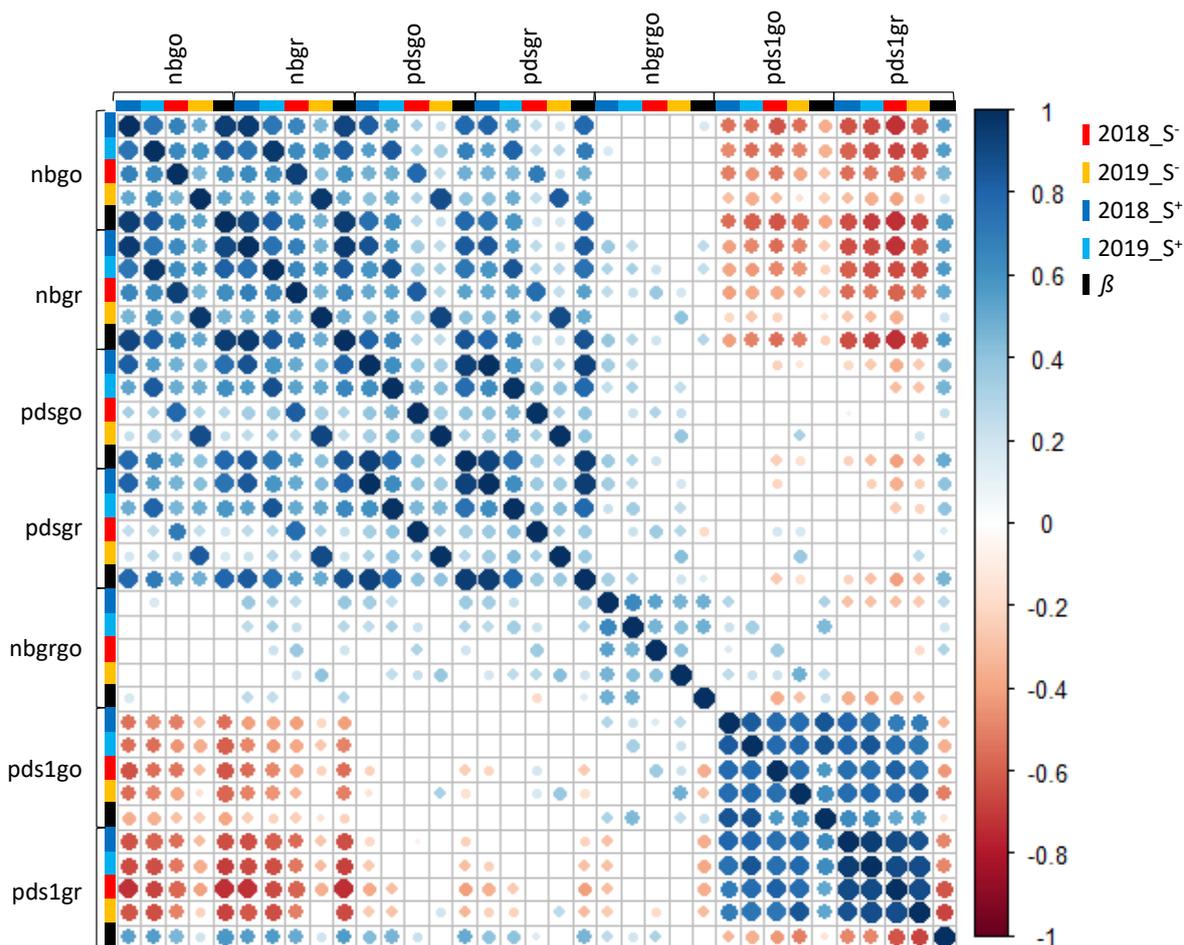


**Figure 38 : Analyse en composante principale mettant en évidence la distribution des génotypes utilisés dans l'analyse en fonction de leur indice  $\beta$  pour les variables de rendement.** La répartition des génotypes est colorée avec des ellipses contenant au moins 75% des individus en fonction de leur groupe génétique (G1, rouge ; G2, bleu ; G3, violet ; G4, marron ; G5, vert).

L'ACP sépare les individus en fonction de la plasticité de leurs composantes de rendement à la fois sur l'axe 1 (56.7% de la variation) et sur l'axe 2 (18.5% de la variation). L'axe 1 est principalement expliqué par la plasticité du nombre et du poids total des graines et des gousses et l'axe 2 est expliqué par le poids d'une gousse. Les génotypes ayant la plasticité la plus élevée sur l'axe 1 sont ceux appartenant aux groupes 3 et 4 et les génotypes ayant la plasticité la plus faible sur l'axe 2 sont ceux appartenant aux groupes 3 et 5.

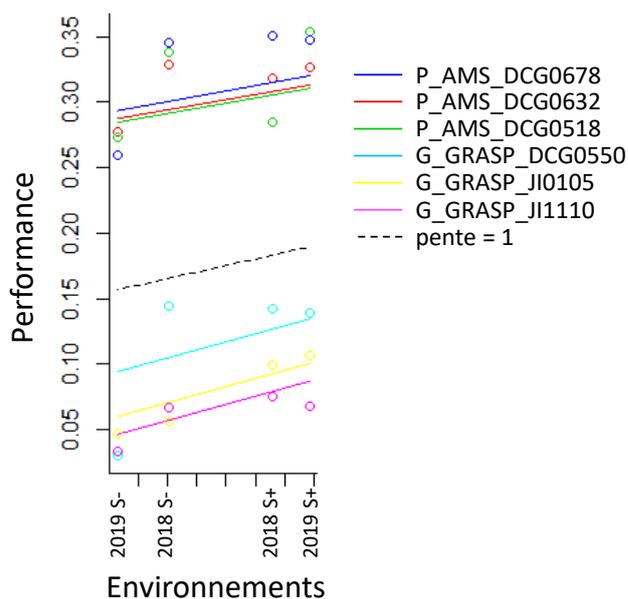
#### Corrélation $\beta$ et variables

Une analyse des corrélations entre les  $\beta$  des variables de rendement (Figure 39) a mis en évidence que les  $\beta$  du nombre et du poids total de gousses et de graines ainsi que le  $\beta$  du poids d'une gousse sont plus corrélés aux valeurs respectives de ces variables en condition  $S^+$  qu'en condition  $S^-$ . Toutes les variables de nombre de graines par gousse ne sont que très faiblement corrélées avec les valeurs en condition  $S^+$ .



**Figure 39 : Heatmap des corrélations entre les variables de rendement et leur indice de plasticité associé.** Bleu foncé, 2018 S+ ; bleu clair, 2019 S+ ; rouge, 2018 S- ; orange, 2019 S- ; noir,  $\beta$ . Les corrélations non significatives au seuil de p-value ajustée à 0.05 sont en blanc.

Un résultat particulièrement intéressant est une corrélation négative entre le  $\beta$  du poids d'une graine avec les valeurs cette variable en condition  $S^+$  et en condition  $S^-$ . Ce résultat est dû au fait que les graines de plus petit poids sont celles dont celui-ci varie le plus en réponse à la carence soufrée. Un exemple est montré avec 6 génotypes figure 40, où 3 génotypes ont un poids d'une graine élevé (valeurs de  $1+\beta$  faibles soient  $<1$ ) et 3 génotypes avec un poids d'une graine faible (valeurs de  $1+\beta$  fortes soient  $>1$ ). De manière intéressante, on a constaté que les génotypes ayant les  $\beta$  les plus négatifs pour le poids d'une graine ont une augmentation de la moyenne du poids d'une graine la plus forte en condition de carence en soufre et qu'il s'agit uniquement de *P. sativum sativum* et plus particulièrement de variétés classées parmi les protéagineux. Les génotypes ayant le poids d'une graine le plus négativement impacté par la carence en soufre ( $\beta > 0$ ) sont les *P. sativum ssp abyssinicum* dont certains sont classés parmi les variétés fourragères.



**Figure 40 : Performance de 6 géotypes pour le poids d'une graine de pois en fonction des valeurs environnementales estimées.** Chaque couleur représente un géotype différent. Les droites sont les droites de régression  $1+\beta$  et les cercles sont les moyennes de chaque géotype dans chacun des environnements. L'axe horizontal correspond aux effets environnementaux et l'axe vertical correspond à la performance des géotypes pour le poids d'une graine. La pente correspond à  $1+\beta$ . La ligne en pointillé correspond à la valeur moyenne de régression de tous les géotypes.

Les variations de  $\beta$  sont différentes en fonction des groupes génétiques, ces différences sont présentées Tableau 11.

Un des résultats les plus intéressants est que le groupe génétique 2 semble être moins impacté que les autres groupes génétiques pour presque toutes les composantes du rendement, ce qui est cohérent avec les résultats observés sur l'ACP (Figure 38). Les autres groupes génétiques sont impactés fortement, soit positivement soit négativement, sur au moins une des variables de plasticité. Par exemple, le groupe génétique G1 est très plastique pour le poids d'une graine et le nombre de graines par gousse mais ne l'est pas pour le poids total de graines alors que G3 est très plastique pour le nombre de gousses et le nombre de graines mais ne l'est pas pour le poids d'une gousse et le nombre de graines par gousse (Tableau 11). Ces résultats mettent en avant l'existence potentielle de stratégies différentes entre groupes génétiques pour le maintien du rendement dans des conditions de carence en soufre.

**Tableau 11 : Indice de plasticité de variables de composantes du rendement du pois en réponse à une carence en soufre deux années.**

Groupe génétique	nbgo		pdsgo		pds1go		nbgr		pdsgr		pds1gr		nbgrgo	
	Moyenne	T	Moyenne	T	Moyenne	T	Moyenne	T	Moyenne	T	Moyenne	T	Moyenne	T
G1	0.87	ab	0.11	ab	0.13	b	2.09	ab	-0.04	ab	0.38	c	0.21	c
G2	0.09	a	0.15	a	0.07	b	0.94	a	0.13	a	-0.06	a	-0.05	ab
G3	1.13	b	0.31	ab	-0.22	a	2.74	ab	0.26	ab	0.19	b	-0.28	a
G4	0.88	b	0.63	b	-0.01	b	3.25	b	0.61	b	-0.02	a	0.21	c
G5	0.59	ab	-0.59	a	-0.48	a	1.05	ab	-0.62	a	-0.14	a	0.19	bc

Moyennes des  $\beta$  des variables de composantes du rendement. Les résultats des tests de Tuckey sont indiqués par des lettres allant de a à c. L'effet négatif de la carence en soufre est d'autant plus fort que la case est rouge ( $\beta$  plus élevé dû au classement des environnements).

### 3.3 Analyses GWAS des composantes du rendement

#### 3.3.a Identification de hotspots

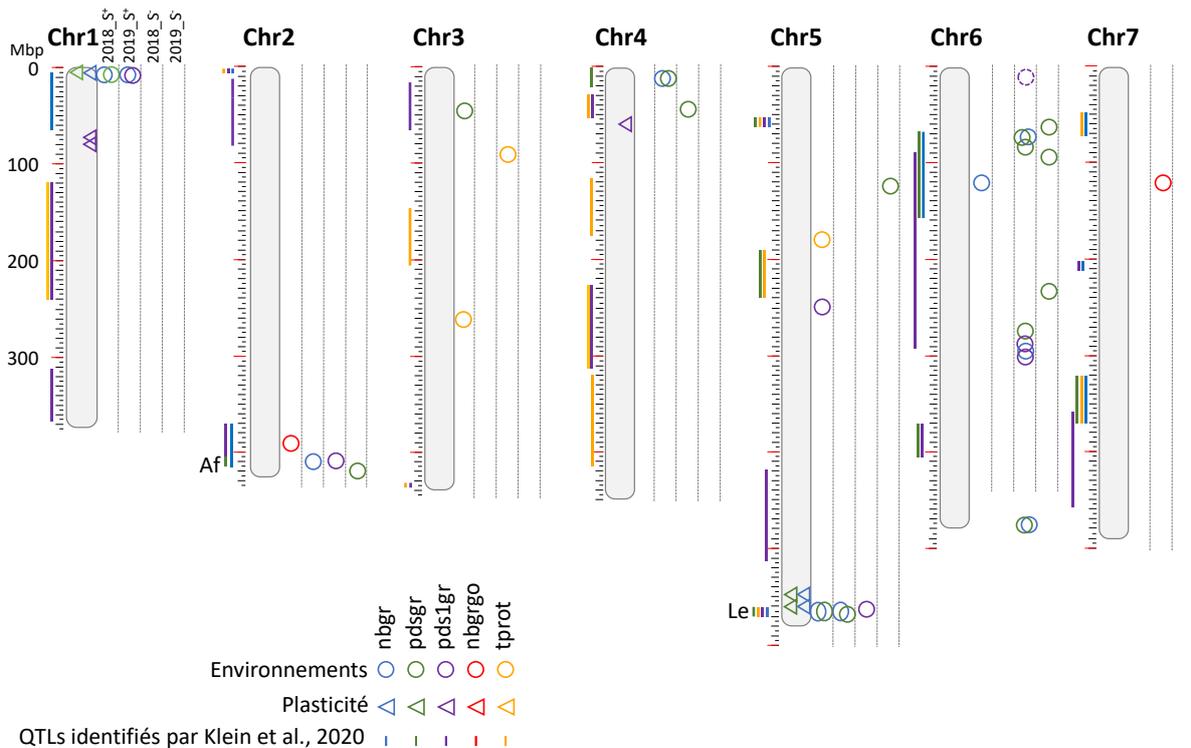
La forte corrélation entre les variables associées aux gousses et celles associées aux graines justifie l'intérêt de mener les analyses de génétique d'association uniquement sur les variables associées aux graines. Le nombre de graines, le poids total de graines, le poids d'une graine, le nombre de graines par gousse ainsi que la teneur protéique ont permis d'identifier par GWAS les régions du génome potentiellement impliquées dans leur variabilité parmi les génotypes de la population étudiée. Le nombre de SNPs retenus (méthode identique à celle utilisée pour la composition protéique) pour chacune des variables de rendement analysées et pour la teneur protéique est présenté figure 41.

(a) Nombre de SNPs identifiés							(b) Nombre de gènes identifiés								
Variable	S <sup>+</sup>			S <sup>-</sup>			$\beta$	Variable	S <sup>+</sup>			S <sup>-</sup>			$\beta$
	2018	2019	Communs	2018	2019	Communs			2018	2019	Communs	2018	2019	Communs	
nbgr	280	95	47	55	5	0	134	nbgr	172	81	48	27	14	0	81
pdsgr	164	33	25	101	57	2	77	pdsgr	106	31	25	89	23	1	64
pds1gr	36	11	2	76	3	0	77	pds1gr	25	16	2	61	4	0	36
nbgrgo	26	6	0	3	3	0	2	nbgrgo	18	3	0	10	8	0	9
tprot	34			35				tprot	30			25			

**Figure 41 : Nombre de SNPs associés aux composantes du rendement et à la teneur protéique en S<sup>+</sup>, en S<sup>-</sup> et à l'indice de plasticité  $\beta$ , et nombre de gènes sélectionnés.** Les SNPs (a) et les gènes (b) communs aux analyses faites sur les données de 2018 et de 2019 dans chaque condition sont indiqués. nbgr, nombre total de graines ; pdsgr, poids total de graines ; pds1gr, poids d'une graine ; nbgrgo, nombre de graines par gousse ; tprot, teneur protéique. La GWAS sur la teneur protéique a été faite uniquement sur les graines issues de la récolte de 2018.

Des SNPs ont été identifiés pour chacune des variables analysées, les deux années, pour la condition contrôle et la condition de carence en soufre ainsi que pour le  $\beta$  (Annexe 5). La répartition de l'ensemble de ces SNPs sur les chromosomes du pois a mené à l'identification de hotspots associés à la variation de chacune de ces variables dans un seul ou plusieurs environnements ainsi que des hotspots associés à l'indice de plasticité  $\beta$  (Figure 42).

De la même manière que pour les hotspots associés à la composition protéique, les hotspots ont été définis par la présence d'un minimum de 5 SNPs sélectionnés dans un intervalle de 5 Mb. La répartition de ces hotspots a mis en évidence un déterminisme génétique complexe des variables de rendement, pouvant varier en fonction de l'année et de la condition.



**Figure 42 : Position des principaux hotspots issus des analyses GWAS du rendement sur la carte génomique du pois (en Mbp).** Les principaux hotspots identifiés par GWAS pour les variables de rendement en condition  $S^+$ ,  $S^-$  et pour l'indice de plasticité  $\beta$ . Un hotspot est défini lorsqu'il y a plus de 5 SNPs sélectionnés pour 5 Mb. Les ronds correspondent aux identifications par environnement et les triangles correspondent aux identifications pour la plasticité. Bleu, nombre de graines ; Vert, poids de graines ; Violet, poids d'une graine ; Rouge, nombre de graines par gousse ; Orange, teneur protéique. Le rond en pointillés correspond au hotspot supplémentaire détecté lors de la mise en cofacteur de *Le*. La teneur protéine n'a été estimée qu'en 2018. A gauche des chromosomes sont indiqués les QTL identifiés par Klein et al., 2020, le même code couleur que pour les ronds est appliqué.

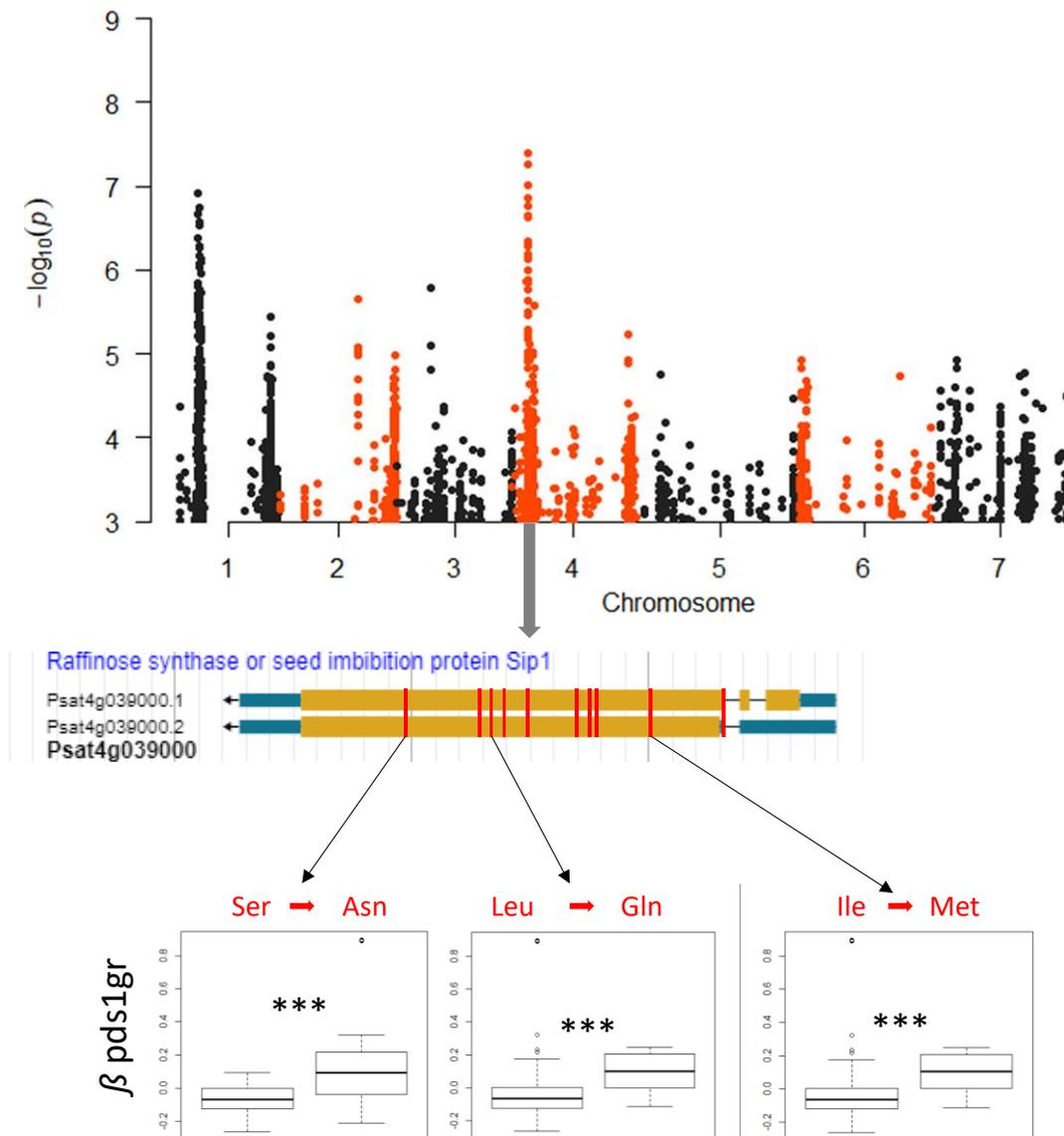
Certains hotspots associés à différents environnements ainsi qu'à l'indice de plasticité  $\beta$  se trouvent à la même position, comme par exemple celui situé à 5 Mb sur le chromosome 1, celui situé à 400 Mb sur le chromosome 2 et celui situé à 565 Mb sur le chromosome 5. Ces deux dernières régions contiennent des gènes connus pour impacter le rendement, respectivement *Af* et *Le*. Le gène *Af* (*Afila*) induit la transformation des folioles en vrilles et le gène *Le* est connu pour réguler la longueur des entre-nœuds et par conséquent la hauteur de la plante. Les hotspots associés au nombre de graines, au poids total de graines et au poids d'une graine se trouvent souvent très proches, et parfois à la même position (5 Mb sur le chromosome 1, 410 Mb sur le chromosome 2, 10 Mb sur le chromosome 4, 565 Mb sur le chromosome 5, 80 et 300 Mb sur le chromosome 6). Les hotspots associés au nombre de graines par gousse et à la teneur protéique des graines semblent quant à eux plus isolés et donc plus spécifiques. Il est à noter que pour le nombre de graines par gousse, des hotspots ont été uniquement identifiés pour la condition  $S^+$ . En comparant ces résultats aux QTL de rendement identifiés par Klein et al., (2020), on observe des régions similairement associées aux composantes du rendement dans

les deux études. Ces régions sont celles contenant les gènes *Af* et *Le* qui sont identifiées très nettement dans l'analyse QTL, et d'autres régions comme celle à 10 Mb sur le chromosome 4 et celle à 80 Mb sur le chromosome 6 (Figure 42). Le fait que certains hotspots identifiés pour les composantes du rendement soient situés au niveau de gènes majeurs ayant un effet pléiotropique coïncide avec la littérature, ces gènes étant susceptibles de contribuer à déterminer le statut azoté des plantes (Burstin et al., 2007; Klein et al., 2020; Weeden, 2007). De plus il a été montré par des analyses QTL que ces gènes pouvaient être associés à la teneur protéique des graines ainsi qu'au rendement, ce qui suggère un ajustement de la mise à disposition de l'azote entre les organes sources et les organes puits dépendamment des conditions environnementales dans lesquelles la plante se développe (Burstin et al., 2007).

### 3.3.b Identification de gènes candidats sous les hotspots

Parmi les gènes les plus pertinents détectés sous ces hotspots, on trouve un gène codant une phosphoadénosine phosphosulfate (PAPS) réductase (Psat3g041880). Ce gène est identifié sous un hotspot (à 90 Mb sur le chromosome 3) associé à la variation de teneur en protéines des graines en condition de carence en soufre en 2018. Lors du développement de la graine, Psat3g041880 est préférentiellement exprimé au stade de transition vers le remplissage (Tableau S10). Son plus proche orthologue chez *Arabidopsis* est At5g03430. Il a pour rôle de réduire le PAPS en sulfite et ainsi de remobiliser le soufre. Une augmentation de la teneur en protéines riches en soufre a été mise en évidence dans des grains de maïs surexprimant, au niveau foliaire, une PAPS réductase d'*Escherichia coli* (Planta et al., 2017) (Tableau 8). On peut donc supposer que ce gène puisse contribuer à moduler l'accumulation des protéines de réserve lors d'une carence en cet élément afin de maintenir, voire d'augmenter la teneur en protéines (Tableau 11). Cette modulation pourrait se faire en remobilisant le soufre présent dans le PAPS et en le rendant disponible pour la synthèse des acides aminés soufrés nécessaires à l'accumulation des protéines de réserve, y compris des globulines 7S qui s'accumulent massivement en condition de carence en soufre.

En haut du chromosome 4, sous le hotspot de plasticité du poids d'une graine (Figure 42) se trouve le gène Psat4g039000 annoté comme étant une raffinose synthase. Ce gène est aussi significativement associé aux variations du poids d'une graine en 2019 en condition de carence en soufre (Tableau S10). Ce gène présente 10 SNPs, tous détectés dans notre analyse, dont trois possèdent des variants induisant un changement d'acide aminé susceptible d'affecter la fonction de la protéine: sérine en asparagine (SNP#1), leucine en glutamine (SNP#3) et isoleucine en méthionine (SNP#9) (Figure 43).



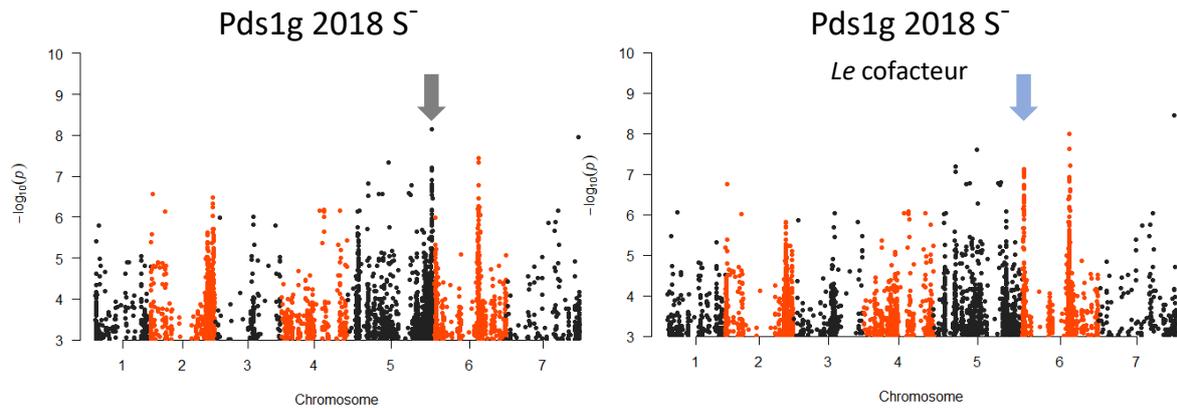
**Figure 43 : Gène identifié pour la plasticité du poids d'une graine en réponse à une carence en soufre.** Le gène de la raffinose synthase se trouve sous un hotspot associé à la plasticité du poids d'une graine en condition de carence en soufre. La figure sous le Manhattan plot provient du browser de la v1a du génome du pois qui est disponible sur [urgi.versailles.inra.fr](http://urgi.versailles.inra.fr) sur lequel ont été indiqués les SNPs détectés dans l'analyse GWAS ainsi que l'effet de leur variant sur le changement d'acide aminé. L'effet du changement d'acide aminé sur la plasticité du poids d'une graine ainsi que la significativité de l'effet sont indiqués. \*\*\*, p-value < 0.001.

Ce gène est particulièrement exprimé durant le remplissage des graines et est régulé négativement dans les feuilles des nœuds reproducteurs en condition de carence en soufre (Tableau S10). Son plus proche orthologue chez *Arabidopsis*, At5g20250, code une protéine nommée DIN10 (Dark inducible 10) ou Raffinose synthase 6, impliquée dans la synthèse du raffinose, un des principaux sucres solubles présents dans les graines, à partir de saccharose. Il a été montré que des plantes subissant un stress hydrique, qui a pour conséquence une accélération de la maturation des graines, produisent des graines avec une teneur plus élevée en raffinose (Obendorf and Górecki, 2012). Le raffinose aurait un rôle protecteur en empêchant la

cristallisation du saccharose dans des conditions de stress (Sun and Leopold, 1997). De plus, il a été proposé que ce gène soit régulé par ABI5 pour atténuer les stress oxydatifs à l'intérieur et à l'extérieur du chloroplaste durant la maturation des graines (Zinsmeister et al., 2016). On peut donc émettre l'hypothèse qu'en condition de carence en soufre, le raffinose joue aussi un rôle protecteur dans les graines, probablement pour limiter les dégâts liés au stress oxydatif (des protéines antioxydantes sont induites dans les graines en réponse à une carence en soufre, Henriot et al., 2021). Il a été montré par Zuber et al. (2013) que les graines de *M. truncatula* produites en condition de carence en soufre appliquée à floraison présentent une augmentation de plus de 30% du poids d'une graine et ont la capacité de maintenir la quantité d'oligosaccharides de la famille du raffinose. A l'inverse, lorsque *M. truncatula* est privé de soufre plus précocement (stade végétatif), les graines produites sont de petite taille et accumulent de faibles quantités d'oligosaccharides. Bien que le lien direct entre la taille de la graine et la quantité de raffinose n'est pas encore établi, nos résultats suggèrent que le gène Psat4g039000 puisse contribuer à la plasticité de ce caractère.

### 3.3.c Optimisation de la détection de gènes candidats en utilisant *Le* comme cofacteur

Afin d'éliminer l'effet majeur et pléiotropique du gène d'architecture *Le* dans l'analyse GWAS des composantes du rendement, nous l'avons ajouté comme cofacteur dans nos analyses (sous la forme d'un SNP polymorphe et présent dans *Le*) pour toutes les variables pour lesquelles ce gène a été identifié. Les variables correspondantes sont le nombre de graines en S<sup>+</sup> les deux années et leur  $\beta$ , le poids total de graines en S<sup>+</sup> les deux années et leur  $\beta$ , le poids d'une graine en S<sup>-</sup> en 2018. Le résultat sur les analyses GWAS est visible sur les Manhattan plots où l'on distingue la disparition du pic à 565 Mb sur le chromosome 5 lorsque *Le* est mis en cofacteur (Figure 44). De plus, on remarque, pour certaines variables comme le poids d'une graine en 2018 en condition de carence en soufre ou le poids total de graines en 2019 en condition contrôle, que de nouveaux SNPs deviennent significatifs au seuil  $10^{-6}$ . Il en résulte que lors de la sélection des SNPs on a d'une part la perte d'un grand nombre de SNPs en déséquilibre de liaison avec celui présent dans *Le* et mis en cofacteur et d'autre part le gain d'autres SNPs dont l'effet pouvait être masqué par la présence de *Le*. Ce gain de nouveaux gènes est particulièrement important pour le poids d'une graine en 2018 en condition de carence en soufre où 40 nouveaux gènes ont été identifiés après la mise en cofacteur de *Le*. Le figure 45 résume la liste des SNPs (a) et des gènes (b) sélectionnés dans les analyses GWAS classiques et avec la mise en cofacteur de *Le* pour toutes les variables dont la GWAS menait à l'identification de ce gène.



**Figure 44 : Effet de la mise en cofacteur de *Le* sur une analyse GWAS.** Manhatten plots issus des analyses GWAS sur le poids d'une graine en 2018 en condition de carence en soufre. La flèche grisée correspond à la position de *Le*. La flèche bleue signale un pic qui voit son  $-\log_{10}(p)$  augmenter lorsque *Le* est mis en cofacteur.

(a) Nombre de SNPs identifiés						(b) Nombre de Gènes identifiés						
Variable	Condition	Année	GWAS			Communs	Variable	Condition	Année	GWAS		
			Classique	GWAS gène <i>Le</i> cofacteur	Communs					Classique	GWAS gène <i>Le</i> cofacteur	Communs
nbgr	S <sup>+</sup>	2018	280	5	0	nbgr	S <sup>+</sup>	2018	172	12	0	
nbgr	S <sup>+</sup>	2019	95	1	0	nbgr	S <sup>+</sup>	2019	81	2	0	
nbgr	$\beta$	$\beta$	134	5	0	nbgr	$\beta$	$\beta$	81	9	0	
pdsgr	S <sup>+</sup>	2018	164	0	0	pdsgr	S <sup>+</sup>	2018	106	0	0	
pdsgr	S <sup>+</sup>	2019	33	3	1	pdsgr	S <sup>+</sup>	2019	31	8	2	
pdsgr	$\beta$	$\beta$	77	3	0	pdsgr	$\beta$	$\beta$	64	5	0	
pds1gr	S <sup>-</sup>	2018	76	84	32	pds1gr	S <sup>-</sup>	2018	61	64	24	

**Figure 45 : Impact de la mise en cofacteur d'un gène majeur à effet pléiotropique sur les résultats de GWAS.** Les SNPs (a) et les gènes (b) communs aux deux analyses sont indiqués. nbgr, nombre de graines ; pdsgr, poids total de graines ; pds1gr, poids d'une graine. Seules les variables pour lesquelles *Le* était identifié en GWAS classique ont été analysées.

La mise en cofacteur du gène *Le* a permis de faire ressortir un pic à 9 Mb sur le chromosome 6 pour le poids d'une graine en condition de carence en soufre en 2018 (Figure 44).

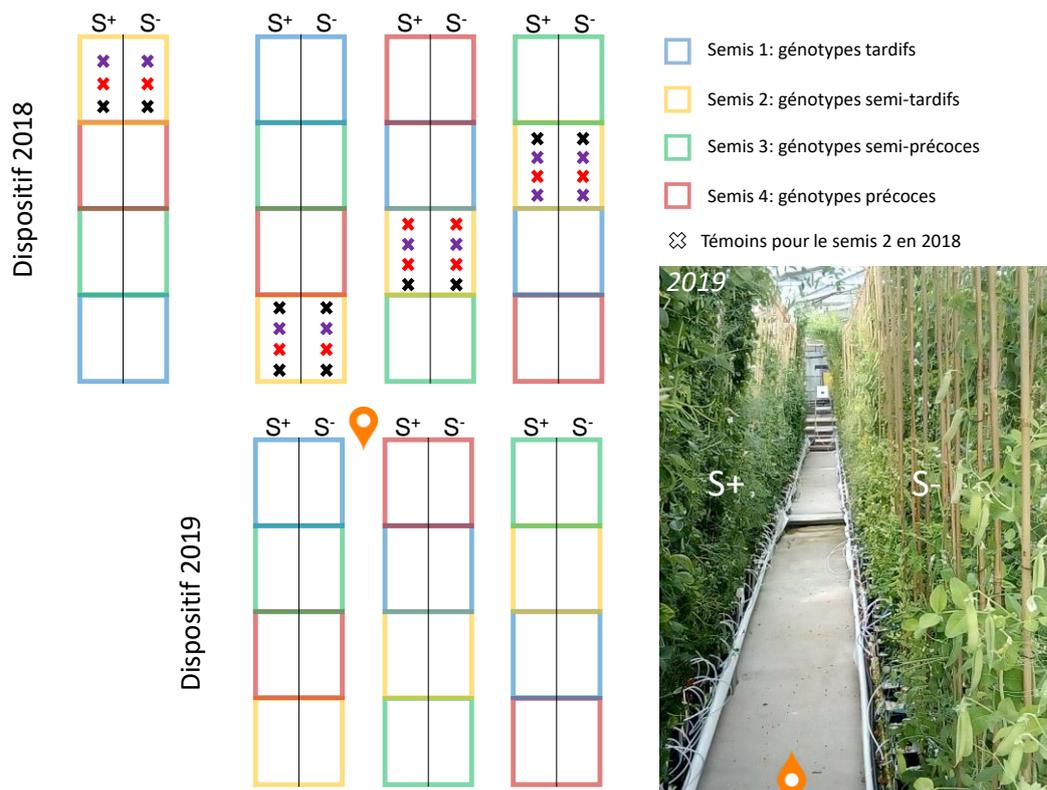
On retrouve sous ce pic un gène nommé *Orc2* (origin recognition complex second largest subunit 2) (Psat6g012600). Il est préférentiellement exprimé lors de l'embryogenèse de la graine et est régulé négativement dans les feuilles des nœuds végétatifs en réponse à une carence en soufre. Son plus proche orthologue chez *Arabidopsis* est At2g37560, et serait impliqué dans la cellularisation de l'albumen lors du développement de la graine chez *Arabidopsis* (Collinge et al., 2004). Sa régulation en condition de carence en soufre pourrait par conséquent impacter le développement voire l'expansion de l'albumen et donc le poids d'une graine. Cette analyse souligne l'importance de placer en cofacteur dans l'analyse GWAS des gènes majeurs connus pour avoir un effet pléiotropique sur le trait étudié et situé sous des hotspots. Les retirer de l'analyse permet en effet de révéler de nouveaux gènes dont les effets sont mineurs.

En conclusion, l'analyse des composantes du rendement a mis en évidence l'impact négatif de la carence en soufre sur ces variables, répétable sur les deux années, et a permis d'identifier 3 zones de hotspots communes à plusieurs de ces variables ainsi qu'à leur plasticité pour deux d'entre eux (Figure 42). Ces régions correspondent, dans deux cas sur trois, à l'emplacement de gènes majeurs d'architecture (*Le* pour la longueur des entrenœuds et *Af* pour la transformation des folioles en vrille) ayant des effets pléiotropiques sur le rendement. L'intégration de ces gènes majeurs dans le modèle GWAS comme covariable nous a permis d'affiner la détection de gènes candidats en révélant de nouveaux SNP significatifs. De nombreux autres hotspots détectés pour une ou deux variables ont été placés sur la carte du pois. Ils colocalisent souvent avec des régions chromosomiques porteuses de QTLs des composantes de rendement identifiés chez cette espèce (Klein et al., 2020). Nous avons proposé trois gènes candidats qui nous ont paru pertinents pour expliquer la variation de ces caractères. La caractérisation des phénotypes de mutants pour ces gènes en condition carencée en soufre permettra d'étudier le lien entre ces gènes et la plasticité des composantes du rendement, afin de proposer des pistes d'amélioration et de stabilisation du rendement dans des environnements appauvris en soufre.

## 4. Matériels et Méthodes

### 4.1 Dispositif expérimental et production des géotypes de pois

Pour tester l'effet d'une carence en soufre sur le rendement et la composition protéique des graines de pois 304 géotypes provenant des collections des projets PEAMUST (collection AMS : Architecture and MultiStress) et ANR-GRaSP (Genetics of Rhizobia Selection by Pea) ont été semés en 2018. L'expérience a été répétée en 2019 en utilisant 198 géotypes, ces derniers ayant été sélectionnés pour être représentatifs de la diversité génétique de la collection (variétés cultivées : pois de printemps, d'hiver, fourragers, protéagineux, potagers ; accessions sauvages, landraces). Compte tenu de la grande diversité génétique associée à la précocité, 4 semis ont été réalisés de manière échelonnée à une semaine d'intervalle (22 février, 1 mars, 8 mars et 15 mars pour 2018 ; 21 février, 28 février, 7 mars et 14 mars pour 2019), en se basant sur les dates de floraison mesurées en 2016 (données non présentées). Les graines ont été mises à germer sur boîte de pétri durant 5 jours dans un germinateur Fitoclima S600 (Aralab, Rio de Mouro, Portugal) à 20°C à l'obscurité. Pour améliorer le taux de germination, les graines des *P. fulvum* et des *P. sativum ssp elatius* ont été scarifiées. Puis, une graine germée de chaque géotype a été repiquée dans des pots de 2 L contenant un mélange de perlite/sable (3/1 en volume) en serre "insect-proof" avec une température régulée à 19 °C le jour et à 15 °C la nuit, et une photopériode de 16h obtenue grâce à un complément de lumière artificielle (250  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) apporté par des lampes à sodium (MACS 400W ; Mazda, Dijon, France). Les plantes ont été irriguées automatiquement avec une solution nutritive riche en nitrate et en soufre identique à celle utilisée par (Henriet et al., 2019), 4 fois par jour, afin de maintenir une teneur en eau du substrat optimale. Au stade mi-végétatif, deux semaines avant la floraison de Caméor (carence appliquée au stade de 6 nœuds), le 3 avril en 2018 et le 8 avril en 2019), une carence en soufre ( $\text{S}^-$ ) a été appliquée à la moitié des plantes par fertilisation avec une solution nutritive sans traces de soufre, une carence appliquée à ce stade étant connue pour affecter la composition protéique des graines (Henriet et al., 2019). Avant la mise en carence, les pots ont été rincés deux fois avec 2l d'eau osmosée et une fois avec 2l de solution  $\text{S}^-$  de façon à minimiser les reliquats en soufre présents dans le substrat. Le dispositif expérimental mis en place était un plan d'expérimentation à une répétition par géotype avec 12 géotypes contrôles répétés 5 fois en 2018 (4 fois en 2019) et répartis en 4 blocs en 2018 (3 en 2019) et 4 sous-blocs (un par date de semis) randomisés. Pour chaque géotype, la plante alimentée avec une solution optimale ( $\text{S}^+$ ) faisait face à celle à laquelle la carence en soufre a été appliquée (Figure 46).



**Figure 46 : Plan d'expérience en serre pour la culture des géotypes de pois.** Le dispositif est un plan en blocs et sous-blocs incomplets. 4 semis ont été effectués à une semaine de décalage afin que les géotypes ayant une précocité variée fleurissent à la même période. Durant le semis 1 (bleu) les géotypes tardifs ont été semés, pour le semis 2 (orange), ce sont les semi-tardifs, pour le semis 3 (en vert), ce sont les semi-précoces et pour le semis 4 (en marron) ce sont les précoces qui ont été semés. Les croix correspondent aux géotypes témoins semi-tardifs de l'année 2018. A l'intérieur d'un bloc, les sous-blocs ont été attribués aléatoirement chacun à une des 4 dates de semis réalisées. Des témoins (12 en 2018 et 9 en 2019) répétés 5 fois en 2018 et 4 fois en 2019 ont été répartis selon leur précocité dans chacun des blocs et sous-blocs.

Les témoins ont été choisis de façon à représenter équitablement les 5 groupes génétiques décrits chez le pois à partir des données de polymorphisme de séquence ((Kreplak et al., 2019), voir section structuration population). Chaque année, pour la mise en place du dispositif, l'attribution des dates de semis aux blocs et sous-blocs a été randomisée ainsi que l'attribution des géotypes aux pots dans les blocs et sous-blocs. Avant la mise en carence, la hauteur de chaque plante a été mesurée et une étiquette a été placée au niveau du dernier nœud végétatif formé. Au moment de la mise en carence puis toutes les semaines pendant un mois, des mesures de teneur en chlorophylle des feuilles ont été réalisées avec un appareil de type SPAD (SPAD-502 chlorophyll meter, Minolta) sur les deux derniers nœuds formés, afin de visualiser les premiers symptômes de carence des géotypes. Pendant la période de floraison, un passage journalier a été fait pour noter les plantes fleuries. A la récolte, chaque plante a été mesurée pour la hauteur totale, la hauteur depuis la mise en carence, le nombre de nœuds végétatifs, le nombre de nœuds reproducteurs et le nombre de gousses. Des commentaires liés à l'état sanitaire des plantes (présence d'oïdium ou de fusariose) ont été notés et ont permis, lors de

l'analyse statistique, de pointer des données aberrantes. Les parties aériennes et racinaires de chaque plante ont été ensachées individuellement et passées à l'étuve 48h à 80°C. Post-récolte, les tissus récoltés, parties aériennes et racinaires, séchés ont été pesés, les gousses et graines comptées et pesées.

## 4.2 Séparation des protéines de réserve des graines et quantification

Pour chaque génotype dans chaque condition, entre 3 grammes de graines ou au moins 20 graines (selon le PMG, poids de mille grains) ont été broyées avec un broyeur (ZM100 Retsch grinder) et une grille de 0.2 mm pour obtenir une poudre fine. La teneur en protéine des graines a été prédite sur les poudres par analyse en proche infra-rouge (appareil NIRS 6500 (Foss), protocole et calibration décrits par (Burstin et al., 2007)). Les protéines de réserve ont été extraites de cette poudre avec un tampon Tris-HCl (Schroeder, 1982). Les protéines ont été séparées par électrophorèse mono-dimensionnelle (1-DE) en conditions dénaturantes avec le système XCell4 Surelock™ Midi-Cell (Life Technology) comme décrit dans (Henriet et al., 2019). Pour chaque échantillon, 2 µL d'extrait protéique, mélangés à 8 µL de tampon de charge (5X = TrisHCl 0.5M, SDS 4%, glycérol 20% et 1 mg de bleu de bromophenol), a été déposé sur un gel de concentration (4% d'acrylamide : 4.1 mM Tris-HCl pH 8.8, 40% (v/v) acrylamide/bisacrylamide (30%/0.8%), 1% (g/v) SDS, 0.05% (g/v) ammonium persulfate, 0.05% (v/v) TEMED) puis la séparation des protéines s'est faite sur un gel de résolution à 11% d'acrylamide (4.1 mM Tris-HCl pH 8.8, 37% (v/v) acrylamide/bisacrylamide (30%/0.8%), 1% (g/v) SDS, 0.05% (g/v) ammonium persulfate, 0.05% (v/v) TEMED). Les extraits protéiques (276 en 2018 ; 198 en 2019) ont été traités en deux lots, correspondant aux années de récolte. Un total de 8 cuves d'électrophorèse 1D a été réalisée en 2018 et 6 cuves en 2019 : une cuve contient 4 gels de chacun 24 puits et correspond aux 90 génotypes présents sur un demi-chêneau, S<sup>+</sup> et S<sup>-</sup> additionné d'un extrait témoin répété sur chacun des gels. Pour chaque année et génotype, les extraits correspondant à la condition témoin (S<sup>+</sup>) et la condition carence en soufre (S<sup>-</sup>) ont été déposés dans des puits adjacents sur le même gel. Un marqueur de taille coloré (prestained blue standard 10–250 kD, Bio-Rad) a été ajouté en bordures de chacun des gels. Les gels ont été colorés avec du bleu de Coomassie (Bio-safe Coomassie G-250 stain de Biorad) selon le protocole du fournisseur et scannés en utilisant le système d'imagerie infrarouge Odyssey (LI-COR) avec une intensité de 7.5 et une résolution de 84 µm. La détection et la quantification des bandes protéiques ont été effectuées avec le logiciel CLIQS (<http://www.totallab.com>). Le poids moléculaire de chaque protéine a été estimé en utilisant le marqueur de taille inséré dans les gels. Dans chaque puits, le volume normalisé de chaque bande

a été calculé comme le rapport du volume de cette bande au volume total des bandes détectées après réduction du bruit de fond avec la méthode “rolling ball” (rayon de 300). Les données quantitatives (correspondant aux abondances relatives des protéines) ont été soumises aux analyses statistiques.

### 4.3 Identification et annotation des protéines de réserve des graines

Les bandes protéiques ont été annotées en comparant les profils protéiques aux cartes 1-DE précédemment établies au laboratoire pour les graines du génotype Caméor (Henriet et al., 2019). Pour obtenir ces cartes, les principales bandes protéiques ont été digérées par la trypsine, les peptides analysés par chromatographie liquide couplée à la spectrométrie de masse, et les protéines identifiées par interrogation de la base de données de séquences obtenues par RNAseq chez Caméor (Alves-Carvalho et al., 2015). L’ensemble des bandes protéiques étudiées dans cette expérimentation a été annoté au regard de cette carte 1D de référence. Pour chacune de nos 10 bandes protéiques, est indiquée entre parenthèses la correspondance avec les bandes décrites par (Henriet et al., 2019) (Figure 26) : B1 (#1), B2 (#1), B3 (#2), B4 (#3), B5 (#7), B6 (#9 et #10), B7 (#11), B8 (#12 et #13), B9 (#16 et #18) et B10 (#19).

### 4.4 Calcul de l’indice de plasticité

Pour chacune des bandes protéiques et des variables de rendement non transformées, un indice de plasticité a été calculé en utilisant la méthode de régression bayésienne de (Finlay and Wilkinson, 1963) (FW) modifiée par (Lian and de los Campos, 2016) et implémentée dans un package R (R Core Team, FW package). Cette méthode estime conjointement les paramètres génotypiques et environnementaux de l’équation suivante :  $y_{ijk} = \mu + g_i + h_j + \beta_i h_j + \varepsilon_{ijk}$  dans laquelle  $y_{ijk}$  est la valeur phénotypique du réplicat  $k^{\text{th}}$  de la variété  $i^{\text{th}}$  observée dans l’environnement  $j^{\text{th}}$ ,  $g_i$  est l’effet moyen de la variété  $i^{\text{th}}$ ,  $h_j$  est l’effet moyen de l’environnement  $j^{\text{th}}$ , et  $\varepsilon_{ijk}$  est le terme associé à l’erreur estimée. Si on réduit l’équation à  $\mu + g_i + (1 + \beta_i) h_j + \varepsilon_{ijk}$ , le terme  $(1 + \beta_i)$  est la mesure associée à la réponse linéaire d’un génotype à l’environnement. Tous les effets sont considérés comme aléatoires et avec une distribution de type normal :  $h \sim (0, I s^2_h)$ ,  $g \sim (0, K s^2_g)$ ,  $\beta \sim (0, K s^2_\beta)$ , où  $I$  est la matrice d’identité, et  $K$  la matrice d’apparentement (obtenue à partir de SNPs, elle est identique à celle utilisée en GWAS). Dans cette méthode, l’indice de plasticité est la pente de la droite  $(1 + \beta)$  de réponse du génotype à l’environnement. Pour chacun des 168 génotypes pour lesquels on disposait de données de composition protéique dans les deux années, en condition  $S^+$  et en condition  $S^-$ , cet indice de plasticité a été calculé à partir des valeurs quantitatives de chacune des 10 bandes protéiques et

des variables de rendement dans les quatre environnements : 2018\_S<sup>+</sup>, 2018\_S<sup>-</sup>, 2019\_S<sup>+</sup>, 2019\_S<sup>-</sup>. Seuls les génotypes présents en 2018 et en 2019 dans les deux conditions S<sup>+</sup> et S<sup>-</sup> ont été pris en compte pour le calcul des  $\beta$  (168 génotypes pour la composition protéique et 184 pour les variables de rendement).

## 4.5 Génotypage et structuration de la core collection de pois

Les données de génotypage de la collection de pois utilisée proviennent de l'équipe ECP (Espèce Cible Pois, dirigée par Judith Burstin). Dans le cadre des projet PIA-PEAMUST et ANR-GRASP, l'équipe ECP a utilisé la technique de capture d'exome pour identifier un grand nombre de SNP. Cette technique consiste en un sous-échantillonnage ciblé de l'ADN pour limiter le séquençage aux régions d'intérêt (Ng et al., 2009). Dans le cas d'une capture d'exome, les exons constituant l'ensemble des parties codantes des gènes sont capturés puis séquencés. Les séquences, après application de différents filtres de qualité, sont alignées sur une référence afin de détecter les variations nucléotidiques dans les régions concernées. Les données des projets PEAPOL et GENOPEA ont été utilisées afin de créer une banque d'ADNc de 57000 séquences de référence qui a servi au design des sondes utilisées pour la réaction de capture. Les séquences obtenues après capture d'exome des ADN des différentes accessions ont été alignées sur la version préliminaire de la séquence du génome (Leveugle et al., PeaMust, booklet disponible sur [www.peamust-project.fr/content/download/4616/45159/version/1/file/Booklet\\_G%C3%A9notypage\\_PeaMUST\\_VF.pdf](http://www.peamust-project.fr/content/download/4616/45159/version/1/file/Booklet_G%C3%A9notypage_PeaMUST_VF.pdf)). Les données ont ensuite été transférées sur la version finale du génome (v1). Cette approche a permis de mettre en évidence plus de 3,92 millions de SNP de haute qualité chez les 304 individus de la collection utilisée dans le cadre de la thèse. A cette matrice de départ, nous avons appliqué deux filtres : seuil de 20% maximum de données manquantes et fréquence allélique minimum (MAF) de 1% pour retenir 1,66 millions de SNPs. A titre indicatif, un filtre de la MAF à 5% permettait de ne retenir que 743060 SNP. Pour l'expérimentation 2019 avec seulement 198 génotypes présents, le nombre de SNP polymorphes est de 1,54 millions. La matrice filtrée de 304 génotypes et 1.666742 M SNP a été imputée avec le logiciel Beagle v4.1 (Browning and Browning, 2007) en utilisant les options par défaut (window=50000) hormis Gprobs=T, pour estimer les probabilités à posteriori des allèles aux génotypes imputés. La matrice de sortie est en format dose d'allèle par génotype (valeurs comprises entre 0 et 1) et on ne garde comme données imputées que les valeurs qui ont un Gprobs  $\geq$  0.7. Les autres valeurs sont mises en données manquantes dans la matrice finale ; au final le nombre de données manquantes de la matrice est de 407 137 soit 0.08 % des données totales et 229 780 SNP (~14%)

ont au moins une donnée manquante. La matrice de Kinship a été calculée à partir de cette matrice « génotype » imputée avec le programme Gemma v.0.94 (Zhou and Stephens, 2014) en retenant la version standardisée (-gk 2) fournie par l'outil. Contrairement à l'espèce *Medicago truncatula*, où nous avons retenu la matrice centrée (-gk1), la matrice standardisée nous a paru plus adaptée pour l'espèce pois, dans la mesure où la fréquence d'allèles rares est beaucoup plus élevée dans cette espèce et que ces allèles rares sont susceptibles d'avoir des effets forts. La présence d'allèles rares est due à la présence de pois de type sauvage très éloignée des génotypes cultivés.

La population a été structurée à l'aide de la méthode DAPC (discriminant analysis of principal components) développée par (Jombart et al., 2010) et implémentée dans le package R « adegenet ». Dans un premier temps, pour optimiser le temps de calcul, nous avons réduit la taille de la matrice de génotypes non imputée, pour ne retenir que les SNP présentant un LD <0.5 sur une fenêtre glissante de 50 SNP soient 333586 SNPs (LD-based pruning sous plink v1.9 –indep 50 5 2) et les chromosomes 1 à 7 (élimination des scaffolds non affectés à une groupe de liaison). Le nombre optimal de clusters est estimé (fonction « find.clusters ») par la méthode euclidienne des k-means. La méthode pour évaluer le « k » optimal est basée sur le calcul statistique du BIC (Bayesian Information Criterion) à partir d'une ACP réalisée sur la matrice génotype (avec 600 composantes retenues). Le nombre de clusters optimal retenu est 5 et 200 composantes d'ACP (critères BIC et a-score respectivement). Ensuite une analyse discriminante est réalisée sur ces 200 composantes (fonction « dapc ») en considérant ce nombre optimal de clusters. Finalement, les contributions (« posterior ») des 304 individus aux 4 premiers axes de l'analyse DAPC sont retenues pour constituer la matrice de structure de la population.

## 4.6 Analyses de variance, ACP et CAH

Les données quantitatives brutes ont été soumises d'abord à une analyse de variance en modèle mixte, en appliquant les corrections de sommes de carrés de type III pour les effets fixes (environnement ou année\* nutrition soufrée et dispositif) et en estimant les variances pour les effets génotype et d'interaction génotype\*environnement (packages R lme4, lmerTest). Les comparaisons des moyennes ont été faites avec un test de Tukey (package CAR). Les analyses multivariées (ACP, Classification Ascendante Hiérarchique, CAH) ont été réalisées sur les moyennes ajustées des génotypes par environnement pour les effets dispositif (bloc et sous-bloc ou cuve) lorsque ceux-ci étaient significatifs (package Emmeans, FactoMineR, factoextra). Pour chaque dispositif, les héritabilités de toutes nos variables ont été estimées à l'aide d'un

modèle mixte avec seulement l'effet génétique aléatoire par condition de nutrition. Lorsque les estimations de variance n'ont pas été possible, le déséquilibre du dispositif conduisant à une matrice singulière, nous avons transformé les données (transformation log ou racine carrée).

#### 4.7 Procédure GWAS et sélection de gènes candidats

Les moyennes ajustées pour toutes les variables analysées en GWAS ont été calculées en ajustant par rapport au dispositif expérimental adapté (package emmeans, R). Pour les variables de morphologie, teneur en protéine et de rendement, un ajustement pour les effets blocs et sous-blocs a été appliqué, lorsque ces effets étaient significatifs dans le modèle. Pour les variables de composition protéique, un ajustement pour l'effet cuve a été appliqué lorsque cet effet était significatif dans le modèle.

Les analyses GWAS ont été effectuées en utilisant l'algorithme du logiciel GEMMA (Genome-wide Efficient Mixed Model Association) v.0.94 (Zhou and Stephens, 2014). Cet algorithme permet de tester *via* un modèle linéaire mixte standard l'association de marqueurs génétiques (ici des SNPs) avec un phénotype en prenant en compte la structure de la population (Q) et la matrice d'apparentement (K). Les P-values de chaque SNP ont été calculées à partir d'un test de likelihood ratio. Les  $-\log_{10}$  des P-values ont été utilisées pour des Manhattan plots en utilisant le package ggplot de R, tous les Manhattans plots sont présentés en Annexe 4. Pour sélectionner les gènes candidats, une première liste de SNPs a été obtenue en sélectionnant les SNPs les plus significatifs, avec une P-value inférieure à  $10^{-6}$  correspondant à une limite de P-value moyenne ajustée par la méthode de Benjamini Hochberg (BH) de 0.1. Une seconde liste de SNPs a été sélectionnée en prenant en compte les SNPs en déséquilibre de liaison (LD) avec les SNPs de la première liste ( $r^2$  plus élevé que 0.6, avec une P-value au maximum de  $10^{-5}$  à une distance maximale de 200 kb, le LD étant généralement très faible au-delà de cette distance (Sjol et al., 2017)), ils ont été retenus en utilisant le clumping LD-based de plink v1.9. A partir de l'ensemble des SNPs de ces deux listes, des gènes candidats ont été sélectionnés à une distance maximale de 50 kb de chaque côté de ces SNPs en se basant sur la version1 (V1) du génome du pois (Kreplak et al., 2019). Les analyses d'enrichissement en ontologies ont été effectuées en utilisant le package R topGO v 2.36.0 (Alexa A and Rahnenfuhrer J, 2016) et les annotations GO du pois disponibles au sein de notre unité (J. Kreplak, comm pers). Les annotations des gènes de pois ainsi que les annotations des variants, indiquant en particulier l'effet de la variation sur la protéine ont été fourni par l'équipe de Judith Burstin (Jonathan Kreplak).

## 4.8 Données transcriptomiques exploitées pour l'interprétation des résultats

Afin d'étudier l'expression des gènes au cours du remplissage de la graine de pois, nous avons utilisé les données d'expression RNA-seq obtenues dans l'équipe par Vanessa Vernoud dans le cadre du projet ANR REGULEG. Ces données ont été acquises en utilisant le génotype Caméor selon le protocole brièvement décrit ci-après. Les plantes ont été cultivées en condition de nutrition optimale jusqu'à floraison puis la moitié des plantes a été placée sous stress hydrique dans les conditions décrites par (Henriet et al., 2019), pendant 4 semaines. Des graines récoltées sur le second nœud florifère ont été récoltées à 8, 12, 16, 19, 23 et 29 jours après pollinisation (JAP), en condition bien arrosée et sous stress hydrique. Trois répétitions biologiques par stade et conditions ont été réalisées. Les données de séquençage RNAseq (HiSeq300, 2x125bp) ont été normalisées et analysées avec le logiciel DeSeq2 (version 1.22.2). Tous les stades de développement ont été comparés entre eux pour obtenir une liste de Gènes Différentiellement Exprimés (DEG) entre au moins deux stades de développement. Au total, 24681 gènes sont classés comme étant DEG ( $p\text{-adjust} < 0.05$ ). Cette liste de gènes DEG a été utilisée pour faire un clustering hiérarchique ( $n=5$ ) visualisé sous forme de heatmap grâce au package pheatmap de R (v1.0.12) en utilisant la méthode "complete". La caractérisation des stades en termes de teneur en eau, poids sec et d'expression des gènes codant les protéines de réserve a permis d'identifier le stade 12 JAP comme le stade de transition entre l'embryogenèse et le remplissage des graines (Vanessa Vernoud, comm. pers.).

Afin d'étudier la régulation des gènes en réponse à une carence en soufre, des données transcriptomiques précédemment obtenues par l'équipe ont été utilisées. Celles présentées pour les graines et les feuilles du 2<sup>ème</sup> nœud reproducteur du génotype Caméor ont été obtenues par RNA-seq. Les conditions de culture sont détaillées dans (Henriet et al., 2019) (carence en soufre appliquée 3 semaines après semis, floraison 2 semaines après mise en carence). Les graines ont été récoltées au stade 9 JAP sur le second nœud florifère (3 à 4 réplicats biologiques). Les feuilles ont été récoltées au niveau du second nœud reproducteur à T0 (date de pollinisation), et T2, T5, T9 et T12 correspondant à 2, 5, 9, et 12 JAP (3 à 4 réplicats biologiques). A chaque stade, l'expression différentielle des gènes entre les conditions S<sup>-</sup> et S<sup>+</sup> a été obtenue en utilisant le package DESeq2 R (version 1.12.3) (Love et al., 2014). Les gènes ont été considérés comme différentiellement exprimés lorsque la valeur  $P$  ajustée (FDR, False Discovery rate) était  $< 0,05$ . Les données obtenues pour les gènes codant les protéines de réserve ont permis d'observer que

9 JAP correspond, dans ces conditions expérimentales, au début de la synthèse des protéines de réserve.

Quant aux données transcriptomiques présentées pour les feuilles des nœuds végétatifs (stade senescence, 27 jours après floraison), celles-ci ont été obtenues en utilisant des puces NimbleGen (Gallardo *et al.*, données non publiées obtenues dans le cadre du projet FUI SERAPIS). Les plantes du génotype Caméor ont été cultivées dans des conditions similaires à nos expérimentations et la carence en soufre appliquée à floraison. Les feuilles ont été prélevées sur les trois derniers nœuds végétatifs (4 réplicats biologiques). Les gènes ont été considérés comme différentiellement exprimés lorsque la valeur  $P$  ajustée (Bonferroni) était  $< 0,05$ . Les données d'expression différentielle présentées dans la thèse pour les graines et les feuilles sont exprimées en  $\log_2$  des ratios d'expression en  $S^-$  versus  $S^+$ .

## CHAPITRE IV :

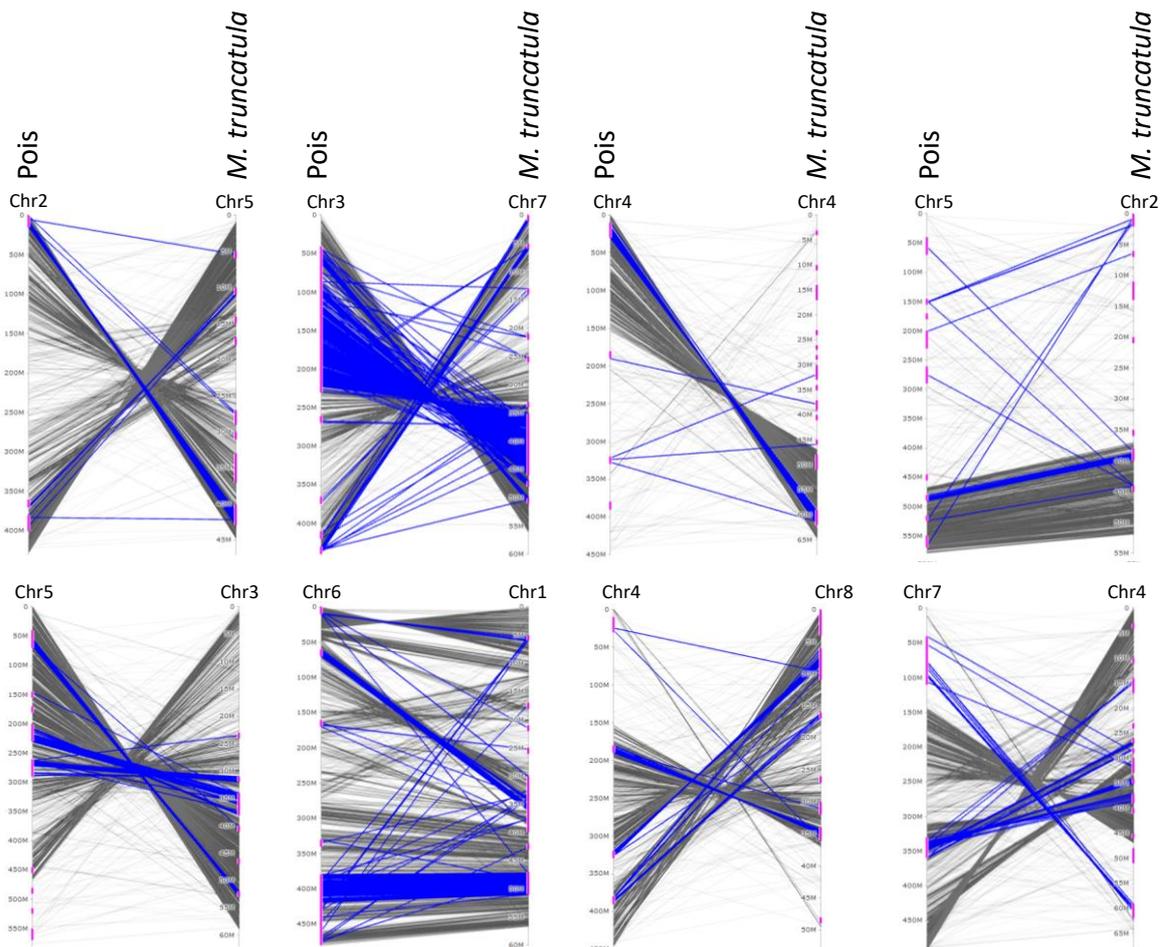
### Conclusion générale et perspectives

L'objectif de cette thèse était de révéler des gènes candidats susceptibles de moduler la composition protéique des graines de *M. truncatula* et de pois en réponse aux contraintes environnementales. Au cours de la thèse, des données sur la variabilité et le déterminisme génétique des composantes du rendement en réponse à une carence en soufre chez le pois ont aussi été obtenues. Dans cette conclusion, l'accent est mis sur la convergence entre hotspots détectés chez *M. truncatula* et le pois pour la composition protéique des graines ainsi qu'entre mécanismes moléculaires sous-jacents au déterminisme de sa plasticité chez ces deux espèces.

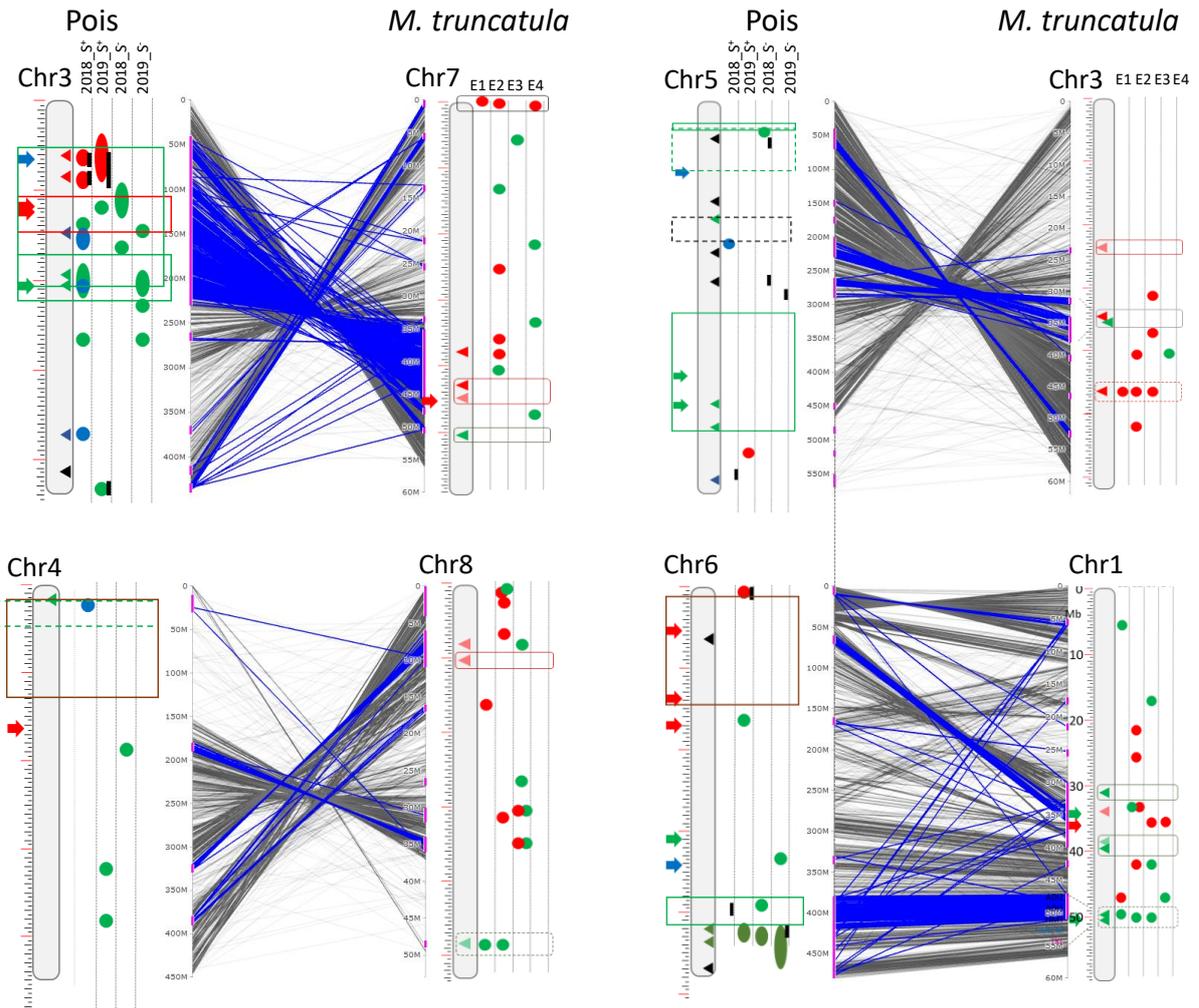
## 1. Synténie entre les hotspots de composition protéique détectés chez les deux espèces

Ce travail de thèse a mis en évidence une forte variabilité de la plasticité de la composition protéique des graines dans une collection d'écotypes de *M. truncatula* ainsi que dans une collection diversifiée de pois. Les analyses GWAS réalisées à partir de ces données nous ont permis de générer une ressource unique de gènes candidats potentiellement impliqués dans la plasticité de la composition protéique des graines de légumineuses en réponse à des environnements variés pour *M. truncatula* ainsi qu'à une carence en soufre pour le pois. Les loci et gènes identifiés sont soit spécifiques de la plasticité soit détectés à la fois pour la plasticité et pour un environnement donné. Les jeux de données obtenus pour chaque espèce ont été confrontés afin d'étudier la conservation des hotspots associés à la composition protéique des graines en général ainsi qu'à sa plasticité entre le pois et *M. truncatula* (Figure 47 et Figure 48). Dans certains cas (haut du chromosome 3 chez le pois et bas du chromosome 7 de *M. truncatula*, Figure 48), la conservation des hotspots entre les deux espèces est parfaite, avec la correspondance des hotspots de légumineuses, de vicilines et de leurs indices de plasticité chez les deux espèces. Notons que ces deux cas correspondent à des zones synténiques pour les gènes de structure des protéines de réserve. Dans la majorité des autres cas, on observe une synténie entre hotspots de vicilines chez une espèce *vs* légumineuses chez l'autre espèce ou entre plasticité *vs* par environnement (Figure 48). La convergence des déterminismes des vicilines et des légumineuses pourrait être liée à notre mesure relative des différentes bandes protéiques. Même si individuellement les bandes protéiques sont peu corrélées entre elles, la somme des vicilines est fortement corrélée négativement à la somme des légumineuses ( $r^2 = -0.98$  chez le pois en moyenne sur les deux années, et  $-0.87$  chez *M. truncatula* en moyenne sur les 4 environnements), ce qui pourrait expliquer des déterminismes communs aux deux groupes de protéines dans notre étude. Pour expliquer la convergence entre hotspots de plasticité et

déterminisme dans un environnement, nous distinguons les environnements que nous avons étudié chez les deux espèces. D'une part, à partir d'un petit nombre d'environnements très diversifiés chez *M. truncatula*, nous avons majoritairement identifié des gènes spécifiques de la plasticité, ce qui en accord avec l'hypothèse de Lacaze et al. (2009) selon laquelle les gènes spécifiques de la plasticité sont identifiés par des études menées sur un faible nombre d'environnements. D'autre part, dans l'étude menée sur le pois, les quatre environnements montrent deux à deux une proximité de conditions ( $S^+$  versus  $S^-$  sur les deux années) et pourraient donc être plus propices à la détection de gènes sensibles, dont l'expression est ici spécifiquement modulée par la carence en soufre.



**Figure 47 : Principaux hotspots associés à la composition protéique des graines communs entre le pois et *M. truncatula*.** Un trait gris ou bleu relie chaque gène du pois à son orthologue le plus proche chez *M. truncatula*. Le trait gris est celui par défaut. Il est bleu lorsque les gènes orthologues sont tous les deux situés sous un hotspot associé à la composition protéique identifié par GWAS.



**Figure 48 : Hotspots majeurs associés à la composition protéique des graines conservés entre le pois et *M. truncatula* ; zoom sur 4 chromosomes.** Un trait gris ou bleu relie chaque gène du pois à son orthologue le plus proche chez *M. truncatula*. Le trait gris est celui par défaut ; il est bleu lorsque les gènes orthologues sont tous les deux situés sous un hotspot associé à la composition protéique identifié par GWAS. Pour le pois, les principaux hotspots identifiés par GWAS pour les globulines et l'albumine (globulines 7S, vert ; globulines 11S, rouge ; albumine, bleu ; ratio 7S/11S, noir) en condition témoin ( $S^+$ ), en condition de carence en soufre ( $S^-$ ) et pour l'indice de plasticité  $\beta$  sont indiqués. Les flèches correspondent aux gènes de structure connus associés à la composition protéique de la graine. La position des PQL détectés par Bourgeois et al. (2011) pour la composition protéique des graines est indiquée en encadré pour chaque type de protéine de réserve quantifié soit après électrophorèse bi-dimensionnelle (gel 2D), soit après FPLC (chromatographie liquide de protéine rapide). Pour *M. truncatula*, les hotspots identifiés par GWAS pour la composition protéique des graines (vert : globulines 7S ; rouge : globulines 11S) dans les 4 environnements (E1 à E4) et pour la plasticité de la composition protéique des grains (triangles) sont indiqués. Les hotspots entourés d'un rectangle sont ceux spécifiques de la plasticité ou des environnements (minimum 3 environnements).

## 2. Gènes de plasticité communs aux deux espèces

L'intersection des listes de gènes obtenues pour la plasticité de la composition protéique chez *M. truncatula* et le pois nous permet de proposer des gènes candidats potentiellement importants pour le contrôle de ce caractère en réponse à des environnements variés (ex : stress hydrique, carence en soufre). Ces 13 gènes candidats de pois et leurs homologues chez *M. truncatula* sont listés dans le tableau 12. Cette liste de gènes met en évidence la possible implication de transporteurs MATE connus pour jouer un rôle détoxifiant face à des stress environnementaux (Upadhyay et al., 2019). L'un de ces gènes *MATE*, Psat3g097080, a été sélectionné dans le chapitre III (§3.2) comme gène candidat pertinent exprimé lors du remplissage des graines de pois. Quatre gènes dans le tableau 12 codent des régulateurs transcriptionnels, dont un facteur de transcription AP2 de type EREBP proposé dans le chapitre III comme gène candidat pour jouer un rôle dans les mécanismes de protection vis-à-vis des stress environnementaux, notamment hydriques (Wang et al., 2003). Cette liste contient également un gène codant une DNA ligase (Chr7g0251371) que nous avons présenté dans le chapitre II comme susceptible de participer au contrôle de l'intégrité du génome, *via* la réparation de l'ADN, dans la graine en développement.

**Tableau 12 : Gènes associés à la plasticité de la composition protéique des graines communs entre l'analyse menée chez le pois et *M. truncatula*.**

<i>P. sativum</i>	Description <i>P. sativum</i>	<i>M. truncatula</i>	Description <i>M. truncatula</i>	Orthologue At	Description <i>A. thaliana</i>
Psat1g103600	AP2 domain	Chr6g0473741*	Putative transcription factor AP2-EREBP family	AT5G51990.1	C-repeat-binding factor 4
Psat2g027920	MTERF	Chr5g0438661	Putative transcription regulator mTERF family	AT4G19650.1	Mitochondrial transcription termination factor family prot
Psat2g027960	MTERF	Chr5g0438661	Putative transcription regulator mTERF family	AT4G19650.1	Mitochondrial transcription termination factor family prot
Psat2g058480	K+ potassium transporter	Chr5g0429991	Putative potassium transporter	AT5G09400.1	K+ uptake permease 7
Psat2g058520	Alternative oxidase	Chr5g0430001	Alternative oxidase 3	AT5G64210.1	alternative oxidase 2
Psat3g096600	MatE	Chr7g0251331	Putative multi antimicrobial extrusion protein	AT1G71140.1	MATE efflux family protein
Psat3g096760	DNA ligase N terminus	Chr7g0251371*	Putative DNA ligase (ATP)	AT1G66730.1	DNA LIGASE 6
Psat3g097040	Zinc-finger of C2H2 type	Chr7g0251361	Putative transcription factor C2H2 family	AT1G60640.1	Stress response protein
Psat3g097080*	MatE	Chr7g0251331	Putative multi antimicrobial extrusion protein	AT2G04050.1	MATE efflux family protein
Psat3g097120	Secondary active transmembrane transporter activit	Chr7g0251331	Putative multi antimicrobial extrusion protein	AT1G71140.1	MATE efflux family protein
Psat3g097200	MatE	Chr7g0251331	Putative multi antimicrobial extrusion protein	AT1G71140.1	MATE efflux family protein
Psat6g206920	ERG2 and Sigma1 receptor like protein	Chr1g0205271	Putative ERG2 sigma1 receptor	AT1G05440.1	C-8 sterol isomerases
Psat6g207000	Best Hit with Mt4 : DUF1677 family protein	Chr1g0205251	hypothetical protein	AT3G22540.1	Protein of unknown function (DUF1677)

13 gènes identifiés chez le pois ont 10 orthologues chez *M. truncatula*. Les gènes avec un astérisque sont ceux discutés dans le manuscrit.

Il n'est pas surprenant que cette liste soit relativement courte, les stress environnementaux appliqués à *M. truncatula* étant de toute autre nature (stress hydrique et décalage de semis) que le stress soufré appliqué chez le pois. On peut donc supposer que la réponse de la composition protéique de la graine à ces stress met en jeu des mécanismes moléculaires différents. C'est d'ailleurs ce que nous avons observé concernant les voies du métabolisme du soufre, orientées

vers le recyclage du soufre dans les conditions environnementales appliquées chez *M. truncatula*, et vers la synthèse *de novo* des acides aminés soufrés en réponse à une carence en soufre appliquée chez le pois.

### 3. Le métabolisme du soufre est un point de contrôle majeur de la plasticité de la composition protéique chez les deux espèces.

De façon remarquable, les deux études de plasticité de la composition protéique réalisées sur deux espèces de légumineuses et dans des gammes d'environnements très différents ont mis en évidence l'implication majeure du métabolisme du soufre. Ce résultat offre des perspectives intéressantes visant à cibler les voies métaboliques identifiées afin d'augmenter et de stabiliser le contenu en acides aminés soufrés des graines des légumineuses. Cet enjeu est d'importance fondamentale pour promouvoir l'utilisation des graines de pois comme source de protéines en alimentation humaine. En effet, la méthionine est un acide aminé essentiel, c'est à dire indispensable au développement et devant être apporté par l'alimentation.

Concernant l'étude conduite chez *M. truncatula* (Chapitre II), la majorité des gènes identifiés sont impliqués dans la réponse au stress à l'échelle de la plante et non uniquement au niveau de la graine, qui est pourtant l'organe étudié. Ceci suggère une régulation de la plasticité de la composition protéique des graines à l'échelle de la plante entière. Ce travail de thèse a révélé que le recyclage de la méthionine synthétisée dans les tissus de la plante joue un rôle majeur dans la plasticité de la composition protéique des graines vis-à-vis des variations environnementales. Les gènes identifiés codent des enzymes appartenant aux cycles de la SMM et de Yang (Figure 21). Le premier cycle a pour rôle de synthétiser de la SMM à partir de méthionine grâce à la méthionine *S*-méthyltransférase (Figure 21) puis de resynthétiser de la méthionine à partir de SMM au niveau des graines grâce à HMT (Bourgis et al., 1999). La SMM jouerait alors un rôle majeur dans le transport du soufre des organes sources vers les graines comme suggéré par (Bourgis et al., 1999) et (Tan et al., 2010). Le second cycle, de Yang, permet de recycler la méthionine à partir du MTA, un sous-produit de la synthèse de l'éthylène (Figure 21). Nous avons validé, par l'analyse d'un mutant, le rôle clé d'une enzyme du cycle de la SMM, *HMT3*, dans le contrôle de la synthèse des légumineuses et dans la plasticité du ratio globulines 7S/11S en réponse à une carence en soufre. Ce résultat suggère qu'en

stimulant le transport de SMM vers les graines et sa conversion en méthionine dans les graines, il serait possible de maintenir le ratio de globulines 7S/11S dans des environnements contrastés.

L'étude de la plasticité de la composition protéique des graines de pois en réponse à une carence en soufre (Chapitre III) a mené à l'identification de gènes impliqués ou connectés aux voies de biosynthèse des acides aminés soufrés. Parmi ces gènes, on trouve la cystathionine beta-lyase (*CBL*) qui permet de former de la méthionine à partir de cystathionine. D'autres gènes codent des enzymes situées au carrefour entre les métabolismes azoté et soufré. C'est le cas de l'enzyme *SERAT* permettant de former l'OAS (précurseur de la cystéine) à partir de sérine, et de la thréonine synthase (gène nommé *MTO2* pour méthionine over-accumulator 2) qui permet de synthétiser de la thréonine à partir d'*O*-homophosphosérine, un précurseur de la cystathionine elle-même précurseur de la méthionine. Ces réactions sont des points de contrôle métabolique sur lesquels il est possible d'agir pour orienter les flux d'azote vers la synthèse de cystéine (*via SERAT*) et de méthionine (*via MTO*). Comme l'inactivation de *MTO* favorise la synthèse de méthionine (Kusano et al., 2010), il serait par exemple intéressant de rechercher des variants alléliques naturels ou induits par *TILLING* réduisant l'activité de cette enzyme chez le pois.

D'autres gènes identifiés par GWAS peuvent contrôler l'homéostasie des acides aminés soufrés. C'est le cas de *CYSCI*, qui permet la synthèse de 3-cyano-L-alanine à partir de cystéine et dont l'expression diminue dans les feuilles et les graines de pois en réponse à une carence en soufre. C'est aussi le cas de *BAT5* qui est impliqué dans le transport de produits dérivés de la méthionine du cytosol vers le chloroplaste et dont l'expression diminue dans les feuilles en réponse à une carence en soufre. Il serait intéressant de tester si l'inactivation de *BAT5* peut favoriser l'incorporation de méthionine dans les protéines de réserve. Dans de telles approches, il est nécessaire de prendre en compte l'existence de mécanismes de rétrocontrôle de la voie de biosynthèse des acides aminés soufrés par les métabolites de cette voie, comme par exemple le rétrocontrôle négatif de la synthèse de cystathionine par l'AdoMet (Chiba et al., 2003). Ces rétrocontrôles complexifient d'autant plus les stratégies visant à augmenter la teneur en méthionine des graines, soulignant l'importance de combiner une approche visant à modifier l'expression de gènes candidats et une approche visant à rechercher des formes d'autres gènes candidats insensibles à de tels rétrocontrôles.

## 4. Gènes candidats liés à la plasticité des composantes du rendement en réponse à une carence en soufre chez le pois.

Les expériences conduites en 2018 et 2019 sur la collection de génotypes de pois ont aussi permis d'acquérir une ressource de données sur la diversité de réponse des composantes du rendement à cette carence (Chapitre III, §3.3). L'un des principaux résultats qui ressort de cette étude est que même en condition de carence en soufre, les gènes majeurs ayant un impact sur le rendement sont des gènes d'architecture comme le gène *Le* qui contrôle la longueur des entrenœuds et donc la hauteur de la plante avec un fort effet pléiotropique sur les composantes du rendement. Les plus grandes plantes ayant une capacité photosynthétique plus élevée sont celles qui produisent le plus de gousses et de graines. Cependant, cette expérimentation a permis de révéler des gènes candidats associés aux variations des composantes du rendement en condition de carence en soufre. Par exemple, une raffinose synthase a été identifiée sous un hotspot majeur de poids d'une graine en condition de carence en soufre. Cette enzyme joue un rôle dans la synthèse du raffinose, un sucre soluble présent dans les graines dont le lien direct avec le contrôle du poids d'une graine n'est pas encore établi. La mise en cofacteur du gène *Le* a permis de révéler d'autres gènes comme le gène *Orc2* (origin recognition complex second largest subunit 2) qui pourrait avoir un rôle dans l'expansion de l'albumen et donc dans le poids d'une graine en condition de carence en soufre.

Par ailleurs, l'analyse de la teneur protéique des graines évaluée sur l'expérimentation 2018 a permis l'identification d'un gène candidat pour jouer un rôle dans le contrôle de ce caractère en condition de carence en soufre. Ce gène code une PAPS réductase impliquée dans la réduction du PAPS en sulfite, et donc susceptible de contribuer à remobiliser le soufre stocké sous forme de PAPS pour la synthèse des acides aminés soufrés. On peut donc supposer que l'efficacité des graines à maintenir des niveaux non négligeables de protéines de réserve riches en soufre est partiellement dépendante de l'efficacité de cette enzyme en condition de carence en soufre.

## 5. Perspectives de validation des gènes candidats

La complexité des caractères étudiés mettant en jeu de nombreux gènes porteurs d'allèles aux effets faibles sur le phénotype oriente les perspectives vers :

(1) L'exploitation d'une plus large variabilité génétique naturelle pour sélectionner des allèles avec des effets plus forts et utilisables en sélection. D'ailleurs la collection du CRB protéagineux de Dijon recense plus de 2000 accessions de pois qui pourraient être mis à disposition pour de telles études. Le travail consisterait à rechercher de tels allèles par séquençage, puis à phénotyper des géotypes porteurs des différents allèles. Il serait également envisageable d'effectuer le séquençage de gènes candidats pour lesquels aucun SNP n'est présent afin d'en valider la pertinence.

(2) La validation fonctionnelle de gènes. Les gènes candidats identifiés dans ces analyses pourront être validés à l'aide de mutants, comme par exemple des mutants TILLING, disponibles dans deux fonds génétiques (pois de printemps et pois d'hiver) chez cette espèce. Le principal avantage de cette collection est qu'elle est une grande source de variabilité induite pour l'ensemble des gènes du pois, offrant la possibilité d'identifier des allèles directement utilisables en sélection. Le TILLING ouvre aussi des perspectives d'étude de grandes séries alléliques pour des gènes essentiels du métabolisme (ex. thréonine synthase) dont une perte totale de fonction serait létale. L'inconvénient du TILLING est que la majorité des variants sont des allèles pertes de fonction, rendant difficile l'accès à un gain de fonction, et que leur impact sur le phénotype étudié dépend du fond génétique mutagénéisé. Une autre stratégie pour valider la fonction de ces gènes serait de faire appel à d'autres méthodes, comme la technique CRISPR-Cas9 qui permet de cibler plusieurs gènes d'intérêt. Cependant, cette technique se heurte au problème de régénération chez le pois, et des actions ont été conduites afin d'améliorer cette étape, notamment dans le cadre du projet REGENCROP ([symposium.inrae.fr/regencrop2018/Consortium-RegenCrop](http://symposium.inrae.fr/regencrop2018/Consortium-RegenCrop)).

# Références

- Aarabi, F., Kusajima, M., Tohge, T., Konishi, T., Gigolashvili, T., Takamune, M., Sasazaki, Y., Watanabe, M., Nakashita, H., Fernie, A.R., Saito, K., Takahashi, H., Hubberten, H.-M., Hoefgen, R., Maruyama-Nakashita, A., 2016. Sulfur deficiency–induced repressor proteins optimize glucosinolate biosynthesis in plants. *Sci. Adv.* 2, e1601087. <https://doi.org/10.1126/sciadv.1601087>
- Aarabi, F., Naake, T., Fernie, A.R., Hoefgen, R., 2020. Coordinating Sulfur Pools under Sulfate Deprivation. *Trends Plant Sci.* 25, 1227–1239. <https://doi.org/10.1016/j.tplants.2020.07.007>
- Alexa A and Rahnenfuhrer J, 2016. topGO: Enrichment Analysis for Gene Ontology. R package version 2.24.0. [WWW Document]. Bioconductor. URL <http://bioconductor.org/packages/topGO/> (accessed 10.1.16).
- Allen, E., Xie, Z., Gustafson, A.M., Carrington, J.C., 2005. microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell* 121, 207–221. <https://doi.org/10.1016/j.cell.2005.04.004>
- Altenbach, S.B., Pearson, K.W., Meeker, G., Staraci, L.C., Sun, S.S.M., 1989. Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. *Plant Mol. Biol.* 13, 513–522. <https://doi.org/10.1007/BF00027311>
- Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A.-L., Jacquin, F., Klein, A., Martin, C., Boucherot, K., Kreplak, J., da Silva, C., Moreau, S., Gamas, P., Wincker, P., Gouzy, J., Burstin, J., 2015. Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 84, 1–19. <https://doi.org/10.1111/tpj.12967>
- Amir, R., Cohen, H., Hacham, Y., 2019. Revisiting the attempts to fortify methionine content in plant seeds. *J. Exp. Bot.* 70, 4105–4114. <https://doi.org/10.1093/jxb/erz134>
- Ané, J.-M., Zhu, H., Frugoli, J., 2008. Recent Advances in *Medicago truncatula* Genomics. *Int. J. Plant Genomics* 2008, 1–11. <https://doi.org/10.1155/2008/256597>
- Anjum, N.A., Gill, R., Kaushik, M., Hasanuzzaman, M., Pereira, E., Ahmad, I., Tuteja, N., Gill, S.S., 2015. ATP-sulfurylase, sulfur-compounds, and plant stress tolerance. *Front. Plant Sci.* 6. <https://doi.org/10.3389/fpls.2015.00210>
- Asensio, C.S., Sirkis, D.W., Maas, J.W., Egami, K., To, T.-L., Brodsky, F.M., Shu, X., Cheng, Y., Edwards, R.H., 2013. Self-Assembly of VPS41 Promotes Sorting Required for Biogenesis of the Regulated Secretory Pathway. *Dev. Cell* 27, 425–437. <https://doi.org/10.1016/j.devcel.2013.10.007>
- Bailly, C., Kranner, I., 2011. Analyses of Reactive Oxygen Species and Antioxidants in Relation to Seed Longevity and Germination, in: Kermodé, A.R. (Ed.), *Seed Dormancy, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 343–367. [https://doi.org/10.1007/978-1-61779-231-1\\_20](https://doi.org/10.1007/978-1-61779-231-1_20)
- Ball, L., Accotto, G.-P., Bechtold, U., Creissen, G., Funck, D., Jimenez, A., Kular, B., Leyland, N., Mejia-Carranza, J., Reynolds, H., Karpinski, S., Mullineaux, P.M., 2004. Evidence for a Direct Link between Glutathione Biosynthesis and Stress Defense Gene Expression in Arabidopsis. *Plant Cell* 16, 2448–2462. <https://doi.org/10.1105/tpc.104.022608>
- Barac, M., Cabrilo, S., Pesic, M., Stanojevic, S., Zilic, S., Macej, O., Ristic, N., 2010. Profile and Functional Properties of Seed Proteins from Six Pea (*Pisum sativum*) Genotypes. *Int. J. Mol. Sci.* 11, 4973–4990. <https://doi.org/10.3390/ijms11124973>
- Barker, D.G., Bianchi, S., Blondon, F., Dattée, Y., Duc, G., Essad, S., Flament, P., Gallusci, P., Génier, G., Guy, P., Muel, X., Tourneur, J., Dénarié, J., Huguet, T., 1990. *Medicago truncatula*, a model plant for studying the molecular genetics of

- theRhizobium-legume symbiosis. *Plant Mol. Biol. Report.* 8, 40–49.  
<https://doi.org/10.1007/BF02668879>
- Batista-Silva, W., Heinemann, B., Rugen, N., Nunes-Nesi, A., Araújo, W.L., Braun, H.-P., Hildebrandt, T.M., 2019. The role of amino acid metabolism during abiotic stress release. *Plant Cell Environ.* 42, 1630–1644. <https://doi.org/10.1111/pce.13518>
- Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., Moreau, S., Niebel, A., Frickey, T., Weiller, G., He, J., Dai, X., Zhao, P.X., Tang, Y., Udvardi, M.K., 2008. A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* 55, 504–513. <https://doi.org/10.1111/j.1365-313X.2008.03519.x>
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berkowitz, O., Wirtz, M., Wolf, A., Kuhlmann, J., Hell, R., 2002. Use of Biomolecular Interaction Analysis to Elucidate the Regulatory Mechanism of the Cysteine Synthase Complex from *Arabidopsis thaliana*. *J. Biol. Chem.* 277, 30629–30634. <https://doi.org/10.1074/jbc.M111632200>
- Bewley, J.D., Black, M., 1994. Seeds: Germination, Structure, and Composition, in: *Seeds*. Springer US, Boston, MA, pp. 1–33. [https://doi.org/10.1007/978-1-4899-1002-8\\_1](https://doi.org/10.1007/978-1-4899-1002-8_1)
- Bielecka, M., Watanabe, M., Morcuende, R., Scheible, W.-R., Hawkesford, M.J., Hesse, H., Hoefgen, R., 2015. Transcriptome and metabolome analysis of plant sulfate starvation and resupply provides novel information on transcriptional regulation of metabolism associated with sulfur, nitrogen and phosphorus nutritional responses in *Arabidopsis*. *Front. Plant Sci.* 5. <https://doi.org/10.3389/fpls.2014.00805>
- Blagrove, R., Gillespie, J., Randall, P., 1976. Effect of Sulphur Supply on the Seed Globulin Composition of *Lupinus angustifolius*. *Funct. Plant Biol.* 3, 173. <https://doi.org/10.1071/PP9760173>
- Blaszczyk, A., Brodzik, R., Sirko, A., 1999. Increased resistance to oxidative stress in transgenic tobacco plants overexpressing bacterial serine acetyltransferase: Overexpression of bacterial SAT in tobacco. *Plant J.* 20, 237–243. <https://doi.org/10.1046/j.1365-313x.1999.00596.x>
- Bonhomme, M., André, O., Badis, Y., Ronfort, J., Burgarella, C., Chantret, N., Prospero, J.-M., Briskine, R., Mudge, J., Debéllé, F., Navier, H., Miteul, H., Hajri, A., Baranger, A., Tiffin, P., Dumas, B., Pilet-Nayel, M.-L., Young, N.D., Jacquet, C., 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. <https://doi.org/10.1111/nph.12611>
- Booker, M.A., DeLong, A., 2015. Producing the Ethylene Signal: Regulation and Diversification of Ethylene Biosynthetic Enzymes. *Plant Physiol.* 169, 42–50. <https://doi.org/10.1104/pp.15.00672>
- Boulard, C., Fatihi, A., Lepiniec, L., Dubreucq, B., 2017. Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1860, 1069–1078. <https://doi.org/10.1016/j.bbagrm.2017.08.008>
- Bourdais, G., Burdiak, P., Gauthier, A., Nitsch, L., Salojärvi, J., Rayapuram, C., Idänheimo, N., Hunter, K., Kimura, S., Merilo, E., Vaattovaara, A., Oracz, K., Kaufholdt, D., Pallon, A., Anggoro, D.T., Glów, D., Lowe, J., Zhou, J., Mohammadi, O., Puukko, T., Albert, A., Lang, H., Ernst, D., Kollist, H., Brosché, M., Durner, J., Borst, J.W., Collinge, D.B., Karpinski, S., Lyngkjær, M.F., Robatzek, S., Wrzaczek, M., Kangasjärvi, J., on behalf of the CRK Consortium, 2015. Large-Scale Phenomics

- Identifies Primary and Fine-Tuning Roles for CRKs in Responses Related to Oxidative Stress. *PLOS Genet.* 11, e1005373. <https://doi.org/10.1371/journal.pgen.1005373>
- Bourgeois, M., Jacquin, F., Cassecuelle, F., Savoies, V., Belghazi, M., Aubert, G., Quillien, L., Huart, M., Marget, P., Burstin, J., 2011. A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *PROTEOMICS* 11, 1581–1594. <https://doi.org/10.1002/pmic.201000687>
- Bourgeois, M., Jacquin, F., Savoies, V., Sommerer, N., Labas, V., Henry, C., Burstin, J., 2009. Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *PROTEOMICS* 9, 254–271. <https://doi.org/10.1002/pmic.200700903>
- Bourgis, F., Roje, S., Nuccio, M.L., Fisher, D.B., Tarczynski, M.C., Li, C., Herschbach, C., Rennenberg, H., Pimenta, M.J., Shen, T.-L., Gage, D.A., Hanson, A.D., 1999. S-Methylmethionine Plays a Major Role in Phloem Sulfur Transport and Is Synthesized by a Novel Type of Methyltransferase. *Plant Cell* 11, 1485–1497. <https://doi.org/10.1105/tpc.11.8.1485>
- Bradshaw, A.D., 2006. Unravelling phenotypic plasticity ? why should we bother? *New Phytol.* 170, 644–648. <https://doi.org/10.1111/j.1469-8137.2006.01761.x>
- Bradshaw, A.D., 1965. Evolutionary Significance of Phenotypic Plasticity in Plants, in: *Advances in Genetics*. Elsevier, pp. 115–155. [https://doi.org/10.1016/S0065-2660\(08\)60048-6](https://doi.org/10.1016/S0065-2660(08)60048-6)
- Branca, A., Paape, T.D., Zhou, P., Briskine, R., Farmer, A.D., Mudge, J., Bharti, A.K., Woodward, J.E., May, G.D., Gentzbittel, L., Ben, C., Denny, R., Sadowsky, M.J., Ronfort, J., Bataillon, T., Young, N.D., Tiffin, P., 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci.* 108, E864–E870. <https://doi.org/10.1073/pnas.1104032108>
- Brand, T.S., Cloete, S.W.P., de Villiers, T.T., Franck, F., 1991. Nutritive evaluation of *Medicago truncatula* (cv. jernalong) pasture for sheep.
- Browning, S.R., Browning, B.L., 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* 81, 1084–1097. <https://doi.org/10.1086/521987>
- Brunel-Muguet, S., D’Hooghe, P., BataillÃ©, M.-P., LarrÃ©, C., Kim, T.-H., Trouverie, J., Avice, J.-C., Etienne, P., DÃ¼rr, C., 2015. Heat stress during seed filling interferes with sulfur restriction on grain composition and seed germination in oilseed rape (*Brassica napus* L.). *Front. Plant Sci.* 6. <https://doi.org/10.3389/fpls.2015.00213>
- Bueckert, R.A., Wagenhoffer, S., Hnatowich, G., Warkentin, T.D., 2015. Effect of heat and precipitation on pea yield and reproductive performance in the field. *Can. J. Plant Sci.* 95, 629–639. <https://doi.org/10.4141/cjps-2014-342>
- Burstin, J., Gallardo, K., Mir, R.R., Varshney, R.K., Duc, G., 2011. Improving protein content and nutrition quality., in: Pratap, A., Kumar, J. (Eds.), *Biology and Breeding of Food Legumes*. CABI, Wallingford, pp. 314–328. <https://doi.org/10.1079/9781845937669.0314>
- Burstin, J., Marget, P., Huart, M., Moessner, A., Mangin, B., Duchene, C., Desprez, B., Munier-Jolain, N., Duc, G., 2007. Developmental Genes Have Pleiotropic Effects on Plant Morphology and Source Capacity, Eventually Impacting on Seed Protein Content and Productivity in Pea. *Plant Physiol.* 144, 768–781. <https://doi.org/10.1104/pp.107.096966>
- Plant Physiol.* 144, 768–781. <https://doi.org/10.1104/pp.107.096966>

- Calderwood, A., Kopriva, S., 2014. Hydrogen sulfide in plants: From dissipation of excess sulfur to signaling molecule. *Nitric Oxide* 41, 72–78.  
<https://doi.org/10.1016/j.niox.2014.02.005>
- Cannon, S.B., Sterck, L., Rombauts, S., Sato, S., Cheung, F., Gouzy, J., Wang, X., Mudge, J., Vasdewani, J., Schiex, T., Spannagl, M., Monaghan, E., Nicholson, C., Humphray, S.J., Schoof, H., Mayer, K.F.X., Rogers, J., Quetier, F., Oldroyd, G.E., Debelle, F., Cook, D.R., Retzel, E.F., Roe, B.A., Town, C.D., Tabata, S., Van de Peer, Y., Young, N.D., 2006. Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proc. Natl. Acad. Sci.* 103, 14959–14964.  
<https://doi.org/10.1073/pnas.0603228103>
- Chan, K.X., Phua, S.Y., Van Breusegem, F., 2019. Secondary sulfur metabolism in cellular signalling and oxidative stress responses. *J. Exp. Bot.* 70, 4237–4250.  
<https://doi.org/10.1093/jxb/erz119>
- Chandler, P.M., Higgins, T.J.V., Randall, P.J., Spencer, D., 1983. Regulation of Legumin Levels in Developing Pea Seeds under Conditions of Sulfur Deficiency: Rates of Legumin Synthesis and Levels of Legumin mRNA. *Plant Physiol.* 71, 47–54.  
<https://doi.org/10.1104/pp.71.1.47>
- Chandler, P.M., Spencer, D., Randall, P.J., Higgins, T.J.V., 1984. Influence of Sulfur Nutrition on Developmental Patterns of Some Major Pea Seed Proteins and Their mRNAs. *Plant Physiol.* 75, 651–657. <https://doi.org/10.1104/pp.75.3.651>
- Chardigny, J.-M., Walrand, S., 2016. Plant protein for food: opportunities and bottlenecks. *OCL* 23, D404. <https://doi.org/10.1051/ocl/2016019>
- Chen, M., Zhang, B., Li, C., Kulaveerasingam, H., Chew, F.T., Yu, H., 2015. *TRANSPARENT TESTA GLABRA1* Regulates the Accumulation of Seed Storage Reserves in *Arabidopsis*. *Plant Physiol.* 169, 391–402.  
<https://doi.org/10.1104/pp.15.00943>
- Cheng, N.-H., Liu, J.-Z., Brock, A., Nelson, R.S., Hirschi, K.D., 2006. AtGRXcp, an *Arabidopsis* Chloroplastic Glutaredoxin, Is Critical for Protection against Protein Oxidative Damage. *J. Biol. Chem.* 281, 26280–26288.  
<https://doi.org/10.1074/jbc.M601354200>
- Cheng, X., Krom, N., Zhang, S., Mysore, K.S., Udvardi, M., Wen, J., 2017. Enabling Reverse Genetics in *Medicago truncatula* Using High-Throughput Sequencing for Tnt1 Flanking Sequence Recovery. *Methods Mol. Biol.* Clifton NJ 1610, 25–37.  
[https://doi.org/10.1007/978-1-4939-7003-2\\_3](https://doi.org/10.1007/978-1-4939-7003-2_3)
- Cheng, X., Wen, J., Tadege, M., Ratet, P., Mysore, K.S., 2011. Reverse Genetics in *Medicago truncatula* Using Tnt1 Insertion Mutants, in: Pereira, A. (Ed.), *Plant Reverse Genetics, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 179–190.  
[https://doi.org/10.1007/978-1-60761-682-5\\_13](https://doi.org/10.1007/978-1-60761-682-5_13)
- Chiaiese, P., 2004. Sulphur and nitrogen nutrition influence the response of chickpea seeds to an added, transgenic sink for organic sulphur. *J. Exp. Bot.* 55, 1889–1901.  
<https://doi.org/10.1093/jxb/erh198>
- Chiba, Y., Sakurai, R., Yoshino, M., Ominato, K., Ishikawa, M., Onouchi, H., Naito, S., 2003. S-adenosyl-L-methionine is an effector in the posttranscriptional autoregulation of the cystathionine -synthase gene in *Arabidopsis*. *Proc. Natl. Acad. Sci.* 100, 10225–10230. <https://doi.org/10.1073/pnas.1831512100>
- Choi, S., Cho, Y., Kim, K., Matsui, M., Son, S.-H., Kim, S.-K., Fujioka, S., Hwang, I., 2013. BAT1, a putative acyltransferase, modulates brassinosteroid levels in *Arabidopsis*. *Plant J.* 73, 380–391. <https://doi.org/10.1111/tpj.12036>
- Cohen, A., Hacham, Y., Welfe, Y., Khatib, S., Avice, J.-C., Amir, R., 2019. Evidence of a significant role of glutathione reductase in the sulfur assimilation pathway. *Plant J.* 16.

- Cohen, H., Israeli, H., Matityahu, I., Amir, R., 2014. Seed-Specific Expression of a Feedback-Insensitive Form of CYSTATHIONINE- $\gamma$ -SYNTHASE in Arabidopsis Stimulates Metabolic and Transcriptomic Responses Associated with Desiccation Stress. *PLANT Physiol.* 166, 1575–1592. <https://doi.org/10.1104/pp.114.246058>
- Cohen, H., Salmon, A., Tietel, Z., Hacham, Y., Amir, R., 2017. The relative contribution of genes operating in the S-methylmethionine cycle to methionine metabolism in Arabidopsis seeds. *Plant Cell Rep.* 36, 731–743. <https://doi.org/10.1007/s00299-017-2124-1>
- Cohen, H., Shir, O.M., Yu, Y., Hou, W., Sun, S., Han, T., Amir, R., 2016. Genetic background and environmental conditions drive metabolic variation in wild type and transgenic soybean (*Glycine max*) seeds: Metabolic variation of wild type and transgenic soybean seeds. *Plant Cell Environ.* 39, 1805–1817. <https://doi.org/10.1111/pce.12748>
- Collinge, M.A., Spillane, C., Köhler, C., Gheyselinck, J., Grossniklaus, U., 2004. Genetic Interaction of an Origin Recognition Complex Subunit and the *Polycomb* Group Gene *MEDEA* during Seed Development. *Plant Cell* 16, 1035–1046. <https://doi.org/10.1105/tpc.019059>
- Créviu-Gabriel, I., Gomez, J., Caffin, J.-P., Carré, B., 1999. Comparison of pig and chicken pepsins for protein hydrolysis. *Reprod. Nutr. Dev.* 39, 443–454. <https://doi.org/10.1051/rnd:19990404>
- Dalmais, M., Schmidt, J., Le Signor, C., Moussy, F., Burstin, J., Savoie, V., Aubert, G., Brunaud, V., de Oliveira, Y., Guichard, C., Thompson, R., Bendahmane, A., 2008. UTILdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol.* 9, R43. <https://doi.org/10.1186/gb-2008-9-2-r43>
- Damerval, C., Josse, J.M., de Vienne, D., n.d. Quantitative Trait Loci Underlying Gene Product Variation: A Novel Perspective for Analyzing Regulation of Genome Expression 13.
- de Bang, T.C., Lay, K.S., Scheible, W.-R., Takahashi, H., 2017. Small peptide signaling pathways modulating macronutrient utilization in plants. *Curr. Opin. Plant Biol.* 39, 31–39. <https://doi.org/10.1016/j.pbi.2017.05.005>
- Demidov, D., Horstmann, C., Meixner, M., Pickardt, T., Saalbach, I., Galili, G., Müntz, K., 2003. [No title found]. *Mol. Breed.* 11, 187–201. <https://doi.org/10.1023/A:1022814506153>
- Derkaoui, M., Romman, L.L., 1993. Derkaoui, 1993.pdf.
- Des Marais, D.L., Hernandez, K.M., Juenger, T.E., 2013. Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annu. Rev. Ecol. Evol. Syst.* 44, 5–29. <https://doi.org/10.1146/annurev-ecolsys-110512-135806>
- Des Marais, D.L., Juenger, T.E., 2010. Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance: Plant abiotic stress tolerance. *Ann. N. Y. Acad. Sci.* 1206, 56–79. <https://doi.org/10.1111/j.1749-6632.2010.05703.x>
- Dinkins, R.D., Srinivasa Reddy, M.S., Meurer, C.A., Yan, B., Trick, H., Thibaud-Nissen, F., Finer, J.J., Parrott, W.A., Collins, G.B., 2001. Increased sulfur amino acids in soybean plants overexpressing the maize 15 kDa zein protein. *Vitro Cell. Dev. Biol. - Plant* 37, 742–747. <https://doi.org/10.1007/s11627-001-0123-x>
- Diouf, I., Derivot, L., Koussevitzky, S., Carretero, Y., Bitton, F., Moreau, L., Causse, M., 2020. Genetic basis of phenotypic plasticity and genotype  $\times$  environment interactions in a multi-parental tomato population. *J. Exp. Bot.* 71, 5365–5376. <https://doi.org/10.1093/jxb/eraa265>

- Djemel, N., Guedon, D., Lechevalier, A., Salon, C., Miquel, M., Prosperi, J.-M., Rochat, C., Boutin, J.-P., 2005. Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol. Biochem.* 43, 557–566. <https://doi.org/10.1016/j.plaphy.2005.04.005>
- Domínguez-Solís, J.R., López-Martín, M.C., Ager, F.J., Ynsa, M.D., Romero, L.C., Gotor, C., 2004. Increased cysteine availability is essential for cadmium tolerance and accumulation in *Arabidopsis thaliana*: Cadmium tolerance in *Arabidopsis*. *Plant Biotechnol. J.* 2, 469–476. <https://doi.org/10.1111/j.1467-7652.2004.00092.x>
- Dong, W., Wang, Y., Takahashi, H., 2019. CLE-CLAVATA1 Signaling Pathway Modulates Lateral Root Development under Sulfur Deficiency. *Plants* 8, 103. <https://doi.org/10.3390/plants8040103>
- Dong, Y., Silbermann, M., Speiser, A., Forieri, I., Linster, E., Poschet, G., Allboje Samami, A., Wanatabe, M., Sticht, C., Teleman, A.A., Deragon, J.-M., Saito, K., Hell, R., Wirtz, M., 2017. Sulfur availability regulates plant growth via glucose-TOR signaling. *Nat. Commun.* 8, 1174. <https://doi.org/10.1038/s41467-017-01224-w>
- Droux, M., 2004. Sulfur Assimilation and the Role of Sulfur in Plant Metabolism: A Survey. *Photosynth. Res.* 79, 331–348. <https://doi.org/10.1023/B:PRES.0000017196.95499.11>
- Droux, M., Ruffet, M.-L., Douce, R., Job, D., 1998. Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants. Structural and kinetic properties of the free and bound enzymes. *Eur. J. Biochem.* 255, 235–245. <https://doi.org/10.1046/j.1432-1327.1998.2550235.x>
- Dündar, E., Bush, D.R., 2009. BAT1, a bidirectional amino acid transporter in *Arabidopsis*. *Planta* 229, 1047–1056. <https://doi.org/10.1007/s00425-009-0892-8>
- Duranti, M., Gius, C., 1997. Legume seeds: protein content and nutritional value. *Field Crops Res.* 53, 31–45. [https://doi.org/10.1016/S0378-4290\(97\)00021-X](https://doi.org/10.1016/S0378-4290(97)00021-X)
- El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., Aarts, M.G.M., 2014. Genotype × environment interaction QTL mapping in plants: lessons from *Arabidopsis*. *Trends Plant Sci.* 19, 390–398. <https://doi.org/10.1016/j.tplants.2014.01.001>
- Falkenberg, B., Witt, I., Zanor, M.I., Steinhauser, D., Mueller-Roeber, B., Hesse, H., Hoefgen, R., 2008. Transcription factors relevant to auxin signalling coordinate broad-spectrum metabolic shifts including sulphur metabolism. *J. Exp. Bot.* 59, 2831–2846. <https://doi.org/10.1093/jxb/ern144>
- Ferrari, B., Romani, M., Aubert, G., Boucherot, K., Burstin, J., Pecetti, L., Huart-Naudet, M., Klein, A., Annicchiarico, P., 2016. Association of SNP markers with agronomic and quality traits of field pea in Italy. *Czech J. Genet. Plant Breed.* 52, 83–93. <https://doi.org/10.17221/22/2016-CJGPB>
- Finlay, K., Wilkinson, G., 1963. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* 14, 742. <https://doi.org/10.1071/AR9630742>
- Fischer, W., Loo, D.D.F., Koch, . Wolfgang, Ludewig, U., Boorer, K.J., Tegeder, M., Rentsch, D., Wright, E.M., Frommer, W.B., 2002. Low and high affinity amino acid H<sup>+</sup>-cotransporters for cellular import of neutral and charged amino acids. *Plant J.* 29, 717–731. <https://doi.org/10.1046/j.1365-313X.2002.01248.x>
- Francois, J.A., Kumaran, S., Jez, J.M., 2006. Structural Basis for Interaction of O - Acetylserine Sulfhydrylase and Serine Acetyltransferase in the *Arabidopsis* Cysteine Synthase Complex. *Plant Cell* 18, 3647–3655. <https://doi.org/10.1105/tpc.106.047316>
- Gali, K.K., Liu, Y., Sindhu, A., Diapari, M., Shunmugam, A.S.K., Arganosa, G., Daba, K., Caron, C., Lachagari, R.V.B., Tar'an, B., Warkentin, T.D., 2018. Construction of high-density linkage maps for mapping quantitative trait loci for multiple traits in field pea (*Pisum sativum* L.). *BMC Plant Biol.* 18, 172. <https://doi.org/10.1186/s12870-018-1368-4>

- Gali, K.K., Sackville, A., Tafesse, E.G., Lachagari, V.B.R., McPhee, K., Hybl, M., Mikić, A., Smýkal, P., McGee, R., Burstin, J., Domoney, C., Ellis, T.H.N., Tar'an, B., Warkentin, T.D., 2019. Genome-Wide Association Mapping for Agronomic and Seed Quality Traits of Field Pea (*Pisum sativum* L.). *Front. Plant Sci.* 10, 1538. <https://doi.org/10.3389/fpls.2019.01538>
- Galili, G., 2011. The aspartate-family pathway of plants. *Plant Signal. Behav.* 6, 192–195. <https://doi.org/10.4161/psb.6.2.14425>
- Gallardo, K., Firnhaber, C., Zuber, H., Hélicher, D., Belghazi, M., Henry, C., Küster, H., Thompson, R., 2007. A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol. Cell. Proteomics MCP* 6, 2165–2179. <https://doi.org/10.1074/mcp.M700171-MCP200>
- Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., Burstin, J., 2003. Proteomics of *Medicago truncatula* Seed Development Establishes the Time Frame of Diverse Metabolic Processes Related to Reserve Accumulation. *Plant Physiol.* 133, 664–682. <https://doi.org/10.1104/pp.103.025254>
- Gallardo, K., Signor, C.L., Duc, G., Thompson, R., Burstin, J., 2017. Quels leviers génétiques mobilisables afin d'améliorer l'équilibre en acides aminés des graines de légumineuses ? 16.
- Gallardo, K., Thompson, R., Burstin, J., 2008. Reserve accumulation in legume seeds. *C. R. Biol.* 331, 755–762. <https://doi.org/10.1016/j.crvi.2008.07.017>
- Gao, C., 2016. MYC2, MYC3, and MYC4 function redundantly in seed storage protein accumulation in *Arabidopsis*. *Plant Physiol. Biochem.* 8.
- Gao, M.-J., Lydiate, D.J., Li, X., Lui, H., Gjetvaj, B., Hegedus, D.D., Rozwadowski, K., 2009. Repression of Seed Maturation Genes by a Trihelix Transcriptional Repressor in *Arabidopsis* Seedlings. *Plant Cell* 21, 54–71. <https://doi.org/10.1105/tpc.108.061309>
- Garcia, V.J., Xu, S.-L., Ravikumar, R., Wang, W., Elliott, L., Gonzalez, E., Fesenko, M., Altmann, M., Brunschweiler, B., Falter-Braun, P., Moore, I., Burlingame, A., Assaad, F.F., Wang, Z.-Y., 2020. TRIPP Is a Plant-Specific Component of the *Arabidopsis* TRAPP II Membrane Trafficking Complex with Important Roles in Plant Development. *Plant Cell* 32, 2424–2443. <https://doi.org/10.1105/tpc.20.00044>
- Garmier, M., Gentzbittel, L., Wen, J., Mysore, K.S., Ratet, P., 2017. *Medicago truncatula* : Genetic and Genomic Resources. *Curr. Protoc. Plant Biol.* 2, 318–349. <https://doi.org/10.1002/cppb.20058>
- Gayler, K.R., Sykes, G.E., 1985. Effects of Nutritional Stress on the Storage Proteins of Soybeans. *Plant Physiol.* 78, 582–585. <https://doi.org/10.1104/pp.78.3.582>
- Gentzbittel, L., Andersen, S.U., Ben, C., Rickauer, M., Stougaard, J., Young, N.D., 2015. Naturally occurring diversity helps to reveal genes of adaptive importance in legumes. *Front. Plant Sci.* 6. <https://doi.org/10.3389/fpls.2015.00269>
- Gigolashvili, T. and Kopriva S., 2014. Transporters in plant sulfur metabolism. *Front. Plant Sci.* 16.
- Gigolashvili, T., Yatusевич, R., Rollwitz, I., Humphry, M., Gershenzon, J., Flügge, U.-I., 2009. The Plastidic Bile Acid Transporter 5 Is Required for the Biosynthesis of Methionine-Derived Glucosinolates in *Arabidopsis thaliana*. *Plant Cell* 21, 1813–1829. <https://doi.org/10.1105/tpc.109.066399>
- Gill, S.S., Tuteja, N., 2010. Polyamines and abiotic stress tolerance in plants. *Plant Signal. Behav.* 5, 26–33. <https://doi.org/10.4161/psb.5.1.10291>
- Graham, P.H., Vance, C.P., 2003. Legumes: Importance and Constraints to Greater Use. *Plant Physiol.* 131, 872–877. <https://doi.org/10.1104/pp.017004>

- Guilioni, L., 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. <https://doi.org/10.1006/anbo.1997.0425>
- Hanafy, M.S., Rahman, S.M., Nakamoto, Y., Fujiwara, T., Naito, S., Wakasa, K., Ishimoto, M., 2013. Differential response of methionine metabolism in two grain legumes, soybean and azuki bean, expressing a mutated form of Arabidopsis cystathionine  $\gamma$ -synthase. *J. Plant Physiol.* 170, 338–345. <https://doi.org/10.1016/j.jplph.2012.10.018>
- Hara-Nishimura, I., Shimada, T., Hiraiwa, N., Nishimura, M., 1995. Vacuolar Processing Enzyme Responsible for Maturation of Seed Proteins. *J. Plant Physiol.* 145, 632–640. [https://doi.org/10.1016/S0176-1617\(11\)81275-7](https://doi.org/10.1016/S0176-1617(11)81275-7)
- Hartmann, T., 2004. Regulation of sulphate assimilation by glutathione in poplars (*Populus tremulaxP. alba*) of wild type and overexpressing  $\gamma$ -glutamylcysteine synthetase in the cytosol. *J. Exp. Bot.* 55, 837–845. <https://doi.org/10.1093/jxb/erh094>
- Hell, R., 1997. Molecular physiology of plant sulfur metabolism. *Planta* 202, 138–148. <https://doi.org/10.1007/s004250050112>
- Henriet, C., Aimé, D., Térézol, M., Kilandamoko, A., Rossin, N., Combes-Soia, L., Labas, V., Serre, R.-F., Prudent, M., Kreplak, J., Vernoud, V., Gallardo, K., 2019. Water stress combined with sulfur deficiency in pea affects yield components but mitigates the effect of deficiency on seed globulin composition. *J. Exp. Bot.* 70, 4287–4304. <https://doi.org/10.1093/jxb/erz114>
- Henriet, C., Balliau, T., Aimé, D., Le Signor, C., Kreplak, J., Zivy, M., Gallardo, K., Vernourd, V., 2021. Proteomics provides insights into the antioxidant network underlying the response of developing pea seeds to sulfur deficiency and drought. *Sous presse*
- Henríquez-Valencia, C., Arenas-M, A., Medina, J., Canales, J., 2018. Integrative Transcriptomic Analysis Uncovers Novel Gene Modules That Underlie the Sulfate Response in *Arabidopsis thaliana*. *Front. Plant Sci.* 9, 470. <https://doi.org/10.3389/fpls.2018.00470>
- Higashi, Y., Hirai, M.Y., Fujiwara, T., Naito, S., Noji, M., Saito, K., 2006. Proteomic and transcriptomic analysis of *Arabidopsis* seeds: molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. *Plant J.* 48, 557–571. <https://doi.org/10.1111/j.1365-313X.2006.02900.x>
- Hirschmann, F., Krause, F., Papenbrock, J., 2014. The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. *Front. Plant Sci.* 5. <https://doi.org/10.3389/fpls.2014.00556>
- Hoefgen, R., Nikiforova, V.J., 2007. Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiol. Plant.* 132, 190–198. <https://doi.org/10.1111/j.1399-3054.2007.01012.x>
- Hopkins, L., Parmar, S., Błaszczuk, A., Hesse, H., Hoefgen, R., Hawkesford, M.J., 2005. *O*-Acetylserine and the Regulation of Expression of Genes Encoding Components for Sulfate Uptake and Assimilation in Potato. *Plant Physiol.* 138, 433–440. <https://doi.org/10.1104/pp.104.057521>
- Huang, X.-Y., Chao, D.-Y., Koprivova, A., Danku, J., Wirtz, M., Müller, S., Sandoval, F.J., Bauwe, H., Roje, S., Dilkes, B., Hell, R., Kopriva, S., Salt, D.E., 2016. Nuclear Localised MORE SULPHUR ACCUMULATION1 Epigenetically Regulates Sulphur Homeostasis in *Arabidopsis thaliana*. *PLOS Genet.* 12, e1006298. <https://doi.org/10.1371/journal.pgen.1006298>
- Huang, X.-Y., Li, M., Luo, R., Zhao, F.-J., Salt, D.E., 2019. Epigenetic regulation of sulfur homeostasis in plants. *J. Exp. Bot.* 70, 4171–4182. <https://doi.org/10.1093/jxb/erz218>

- Hubberten, H.-M., Klie, S., Caldana, C., Degenkolbe, T., Willmitzer, L., Hoefgen, R., 2012. Additional role of O-acetylserine as a sulfur status-independent regulator during plant growth: Regulatory role of O-acetylserine. *Plant J.* 70, 666–677. <https://doi.org/10.1111/j.1365-313X.2012.04905.x>
- Hugouvieux, V., Dutilleul, C., Jourdain, A., Reynaud, F., Lopez, V., Bourguignon, J., 2009. Arabidopsis Putative Selenium-Binding Protein1 Expression Is Tightly Linked to Cellular Sulfur Demand and Can Reduce Sensitivity to Stresses Requiring Glutathione for Tolerance1[W] 151, 14.
- Inaba, Y., Brotherton, J.E., Ulanov, A., Widholm, J.M., 2007. Expression of a feedback insensitive anthranilate synthase gene from tobacco increases free tryptophan in soybean plants. *Plant Cell Rep.* 26, 1763–1771. <https://doi.org/10.1007/s00299-007-0381-0>
- Inal, A., Günes, A., Alpaslan, M., Sait Adak, M., Taban, S., Eraslan, F., 2003. Diagnosis of Sulfur Deficiency and Effects of Sulfur on Yield and Yield Components of Wheat Grown in Central Anatolia, Turkey. *J. Plant Nutr.* 26, 1483–1498. <https://doi.org/10.1081/PLN-120021056>
- Ishimoto, M., Rahman, S.M., Hanafy, M.S., Khalafalla, M.M., El-Shemy, H.A., Nakamoto, Y., Kita, Y., Takanashi, K., Matsuda, F., Murano, Y., Funabashi, T., Miyagawa, H., Wakasa, K., 2010. Evaluation of amino acid content and nutritional quality of transgenic soybean seeds with high-level tryptophan accumulation. *Mol. Breed.* 25, 313–326. <https://doi.org/10.1007/s11032-009-9334-3>
- Jacob, C., Anwar, A., 2008. The chemistry behind redox regulation with a focus on sulphur redox systems. *Physiol. Plant.* 133, 469–480. <https://doi.org/10.1111/j.1399-3054.2008.01080.x>
- Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94. <https://doi.org/10.1186/1471-2156-11-94>
- Jones-Rhoades, M.W., Bartel, D.P., 2004. Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Mol. Cell* 14, 787–799. <https://doi.org/10.1016/j.molcel.2004.05.027>
- Journet, E.-P., Carreau, V., Gouzy, J., Thoquet, P., Rosenberg, C., Barker, D., Huguet, T., Denarie, J., Gamas, P., 2001. Génomique de la légumineuse modèle *Medicago truncatula* : état des lieux et perspectives. *Ol. Corps Gras Lipides* 8, 478–484. <https://doi.org/10.1051/ocl.2001.0478>
- Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T., 2005. LEAFY COTYLEDON1 Controls Seed Storage Protein Genes through Its Regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406. <https://doi.org/10.1093/pcp/pci048>
- Kang, Y., Sakiroglu, M., Krom, N., Stanton-Geddes, J., Wang, M., Lee, Y.-C., Young, N.D., Udvardi, M., 2015. Genome-wide association of drought-related and biomass traits with HapMap SNPs in *Medicago truncatula*: GWAS of drought-related traits in *Medicago truncatula*. *Plant Cell Environ.* 38, 1997–2011. <https://doi.org/10.1111/pce.12520>
- Kang, Y., Torres-Jerez, I., An, Z., Greve, V., Huhman, D., Krom, N., Cui, Y., Udvardi, M., 2019. Genome-wide association analysis of salinity responsive traits in *MEDICAGO TRUNCATULA*. *Plant Cell Environ.* 42, 1513–1531. <https://doi.org/10.1111/pce.13508>
- Kankipati, H.N., Rubio-Teixeira, M., Castermans, D., Diallinas, G., Thevelein, J.M., 2015. Sul1 and Sul2 Sulfate Transceptors Signal to Protein Kinase A upon Exit of Sulfur Starvation. *J. Biol. Chem.* 290, 10430–10446. <https://doi.org/10.1074/jbc.M114.629022>

- Karchi, H., Shaul, O., Galili, G., 1993. Seed-specific expression of a bacterial desensitized aspartate kinase increases the production of seed threonine and methionine in transgenic tobacco. *Plant J.* 3, 721–727. <https://doi.org/10.1111/j.1365-313X.1993.00721.x>
- Karjalainen, R., Kortet, S., 1987. Environmental and genetic variation in protein content of peas under northern growing conditions and breeding implications. *Agric. Food Sci.* 59, 1–9. <https://doi.org/10.23986/afsci.72238>
- Kawashima, C.G., Matthewman, C.A., Huang, S., Lee, B.-R., Yoshimoto, N., Koprivova, A., Rubio-Somoza, I., Todesco, M., Rathjen, T., Saito, K., Takahashi, H., Dalmay, T., Kopriva, S., 2011. Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in Arabidopsis: miR395 in the control of S assimilation. *Plant J.* 66, 863–876. <https://doi.org/10.1111/j.1365-313X.2011.04547.x>
- Kawashima, C.G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y.N., Saito, K., Takahashi, H., Dalmay, T., 2009. Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J.* 57, 313–321. <https://doi.org/10.1111/j.1365-313X.2008.03690.x>
- Khan, N.A., Singh, S., Umar, S. (Eds.), 2008. Sulfur Assimilation and Abiotic Stress in Plants. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-540-76326-0>
- Kim, H., Hirai, M.Y., Hayashi, H., Chino, M., Naito, S., Fujiwara, T., 1999. Role of O -acetyl- l -serine in the coordinated regulation of the expression of a soybean seed storage-protein gene by sulfur and nitrogen nutrition. *Planta* 209, 282–289. <https://doi.org/10.1007/s004250050634>
- Kim, M., Lee, U., Small, I., Francs-Small, C.C. des, Vierling, E., 2012. Mutations in an Arabidopsis Mitochondrial Transcription Termination Factor–Related Protein Enhance Thermotolerance in the Absence of the Major Molecular Chaperone HSP101. *Plant Cell* 24, 3349–3365. <https://doi.org/10.1105/tpc.112.101006>
- Kim, W.-S., Chronis, D., Juergens, M., Schroeder, A.C., Hyun, S.W., Jez, J.M., Krishnan, H.B., 2012. Transgenic soybean plants overexpressing O-acetylserine sulfhydrylase accumulate enhanced levels of cysteine and Bowman–Birk protease inhibitor in seeds. *Planta* 235, 13–23. <https://doi.org/10.1007/s00425-011-1487-8>
- Kim, W.-S., Krishnan, H.B., 2004. Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. *Plant Biotechnol. J.* 2, 199–210. <https://doi.org/10.1111/j.1467-7652.2004.00063.x>
- Klein, A., Houtin, H., Rond, C., Marget, P., Jacquin, F., Boucherot, K., Huart, M., Rivière, N., Boutet, G., Lejeune-Hénaut, I., Burstin, J., 2014. QTL analysis of frost damage in pea suggests different mechanisms involved in frost tolerance. *Theor. Appl. Genet.* 127, 1319–1330. <https://doi.org/10.1007/s00122-014-2299-6>
- Klein, A., Houtin, H., Rond-Coissieux, C., Naudet-Huart, M., Touratier, M., Marget, P., Burstin, J., 2020. Meta-analysis of QTL reveals the genetic control of yield-related traits and seed protein content in pea. *Sci. Rep.* 10, 15925. <https://doi.org/10.1038/s41598-020-72548-9>
- Koprivova, A., Calderwood, A., Lee, B.-R., Kopriva, S., 2014. Do PFT1 and HY5 interact in regulation of sulfate assimilation by light in Arabidopsis? *FEBS Lett.* 588, 1116–1121. <https://doi.org/10.1016/j.febslet.2014.02.031>
- Kovtun, Y., Chiu, W.-L., Tena, G., Sheen, J., 2000. Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl. Acad. Sci.* 97, 2940–2945. <https://doi.org/10.1073/pnas.97.6.2940>

- Kraakman, A.T.W., Niks, R.E., Van den Berg, P.M.M.M., Stam, P., Van Eeuwijk, F.A., 2004. Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. *Genetics* 168, 435–446. <https://doi.org/10.1534/genetics.104.026831>
- Krajewski, P., Bocianowski, J., Gawłowska, M., Kaczmarek, Z., Pniewski, T., Świącicki, W., Wolko, B., 2012. QTL for yield components and protein content: a multienvironment study of two pea (*Pisum sativum* L.) populations. *Euphytica* 183, 323–336. <https://doi.org/10.1007/s10681-011-0472-4>
- Kreplak, J., Madoui, M.-A., Cápál, P., Novák, P., Labadie, K., Aubert, G., Bayer, P.E., Gali, K.K., Syme, R.A., Main, D., Klein, A., Bérard, A., Vrbová, I., Fournier, C., d'Agata, L., Belser, C., Berrabah, W., Toegelová, H., Milec, Z., Vrána, J., Lee, H., Kougbéadjó, A., Térézol, M., Huneau, C., Turo, C.J., Mohellibi, N., Neumann, P., Falque, M., Gallardo, K., McGee, R., Tar'an, B., Bendahmane, A., Aury, J.-M., Batley, J., Le Paslier, M.-C., Ellis, N., Warkentin, T.D., Coyne, C.J., Salse, J., Edwards, D., Lichtenzweig, J., Macas, J., Doležel, J., Wincker, P., Burstin, J., 2019. A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* 51, 1411–1422. <https://doi.org/10.1038/s41588-019-0480-1>
- Kusakizako, T., Miyauchi, H., Ishitani, R., Nureki, O., 2020. Structural biology of the multidrug and toxic compound extrusion superfamily transporters. *Biochim. Biophys. Acta BBA - Biomembr.* 1862, 183154. <https://doi.org/10.1016/j.bbamem.2019.183154>
- Kusano, M., Fukushima, A., Redestig, H., Kobayashi, M., Otsuki, H., Onouchi, H., Naito, S., Hirai, M.Y., Saito, K., 2010. Comparative metabolomics charts the impact of genotype-dependent methionine accumulation in *Arabidopsis thaliana*. *Amino Acids* 39, 1013–1021. <https://doi.org/10.1007/s00726-010-0562-y>
- Kusmec, A., de Leon, N., Schnable, P.S., 2018. Harnessing Phenotypic Plasticity to Improve Maize Yields. *Front. Plant Sci.* 9, 1377. <https://doi.org/10.3389/fpls.2018.01377>
- Kusmec, A., Srinivasan, S., Nettleton, D., Schnable, P.S., 2017. Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nat. Plants* 3, 715–723. <https://doi.org/10.1038/s41477-017-0007-7>
- Labas, V., Grasseau, I., Cahier, K., Gargaros, A., Harichaux, G., Teixeira-Gomes, A.-P., Alves, S., Bourin, M., Gérard, N., Blesbois, E., 2015. Qualitative and quantitative peptidomic and proteomic approaches to phenotyping chicken semen. *J. Proteomics* 112, 313–335. <https://doi.org/10.1016/j.jprot.2014.07.024>
- Lacaze, X., Hayes, P.M., Korol, A., 2009. Genetics of phenotypic plasticity: QTL analysis in barley, *Hordeum vulgare*. *Heredity* 102, 163–173. <https://doi.org/10.1038/hdy.2008.76>
- Lai, J., Messing, J., 2002. Increasing maize seed methionine by mRNA stability. *Plant J.* 30, 395–402. <https://doi.org/10.1046/j.1365-313X.2001.01285.x>
- Lappartient, A.G., Vidmar, J.J., Leustek, T., Glass, A.D.M., Touraine, B., 1999. Inter-organ signaling in plants: regulation of ATP sulfurylase and sulfate transporter genes expression in roots mediated by phloem-translocated compound. *Plant J.* 18, 89–95. <https://doi.org/10.1046/j.1365-313X.1999.00416.x>
- Le Signor, C., Aimé, D., Bordat, A., Belghazi, M., Labas, V., Gouzy, J., Young, N.D., Prosperi, J.-M., Leprince, O., Thompson, R.D., Buitink, J., Burstin, J., Gallardo, K., 2017. 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. <https://doi.org/10.1111/nph.14500>
- Le Signor, C., Savoies, V., Aubert, G., Verdier, J., Nicolas, M., Pagny, G., Moussy, F., Sanchez, M., Baker, D., Clarke, J., Thompson, R., 2009. Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. *Plant Biotechnol. J.* 7, 430–441. <https://doi.org/10.1111/j.1467-7652.2009.00410.x>

- Lee, B.-R., Huseby, S., Koprivova, A., Chételat, A., Wirtz, M., Mugford, S.T., Navid, E., Brearley, C., Saha, S., Mithen, R., Hell, R., Farmer, E.E., Kopriva, S., 2012. Effects of *fou8/fry1* Mutation on Sulfur Metabolism: Is Decreased Internal Sulfate the Trigger of Sulfate Starvation Response? *PLoS ONE* 7, e39425. <https://doi.org/10.1371/journal.pone.0039425>
- Lee, B.-R., Koprivova, A., Kopriva, S., 2011. The key enzyme of sulfate assimilation, adenosine 5'-phosphosulfate reductase, is regulated by HY5 in *Arabidopsis*: HY5 in control of sulfate assimilation. *Plant J.* 67, 1042–1054. <https://doi.org/10.1111/j.1365-313X.2011.04656.x>
- Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., Deng, X.W., 2007. Analysis of Transcription Factor HY5 Genomic Binding Sites Revealed Its Hierarchical Role in Light Regulation of Development. *Plant Cell* 19, 731–749. <https://doi.org/10.1105/tpc.106.047688>
- Lee, T.T.T., Wang, M.M.C., Hou, R.C.W., Chen, L.-J., Su, R.-C., Wang, C.-S., Tzen, J.T.C., 2003. Enhanced Methionine and Cysteine Levels in Transgenic Rice Seeds by the Accumulation of Sesame 2S Albumin. *Biosci. Biotechnol. Biochem.* 67, 1699–1705. <https://doi.org/10.1271/bbb.67.1699>
- Lewandowska, M., Sirko, A., 2008. Recent advances in understanding plant response to sulfur-deficiency stress. *Acta Biochim. Pol.* 55, 457–471. [https://doi.org/10.18388/abp.2008\\_3051](https://doi.org/10.18388/abp.2008_3051)
- Li, C., Wu, X., Li, Yongxiang, Shi, Y., Song, Y., Zhang, D., Li, Yu, Wang, T., 2019. Genetic architecture of phenotypic means and plasticities of kernel size and weight in maize. *Theor. Appl. Genet.* 132, 3309–3320. <https://doi.org/10.1007/s00122-019-03426-w>
- Li, C., Zhang, B., Chen, B., Ji, L., Yu, H., 2018. Site-specific phosphorylation of TRANSPARENT TESTA GLABRA1 mediates carbon partitioning in *Arabidopsis* seeds. *Nat. Commun.* 9, 571. <https://doi.org/10.1038/s41467-018-03013-5>
- Li, L., Shimada, T., Takahashi, H., Ueda, H., Fukao, Y., Kondo, M., Nishimura, M., Hara-Nishimura, I., 2006. MAIGO2 Is Involved in Exit of Seed Storage Proteins from the Endoplasmic Reticulum in *Arabidopsis thaliana*. *Plant Cell* 18, 3535–3547. <https://doi.org/10.1105/tpc.106.046151>
- Li, Z., Meyer, S., Essig, J.S., Liu, Y., Schapaugh, M.A., Muthukrishnan, S., Hainline, B.E., Trick, H.N., 2005. High-level Expression of Maize  $\gamma$ -zein Protein in Transgenic Soybean (*Glycine max*). *Mol. Breed.* 16, 11–20. <https://doi.org/10.1007/s11032-004-7658-6>
- Lian, L., de los Campos, G., 2016. FW: An R Package for Finlay–Wilkinson Regression that Incorporates Genomic/Pedigree Information and Covariance Structures Between Environments. *G3Bethesda GenesGenomesGenetics* 6, 589–597. <https://doi.org/10.1534/g3.115.026328>
- Liang, G., Yang, F., Yu, D., 2010. MicroRNA395 mediates regulation of sulfate accumulation and allocation in *Arabidopsis thaliana*: miRNA395 and sulfate homeostasis. *Plant J.* no-no. <https://doi.org/10.1111/j.1365-313X.2010.04216.x>
- Liu, P.-L., Du, L., Huang, Y., Gao, S.-M., Yu, M., 2017. Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evol. Biol.* 17, 47. <https://doi.org/10.1186/s12862-017-0891-5>
- Loudet, O., Saliba-Colombani, V., Camilleri, C., Calenge, F., Gaudon, V., Koprivova, A., North, K.A., Kopriva, S., Daniel-Vedele, F., 2007. Natural variation for sulfate content in *Arabidopsis thaliana* is highly controlled by APR2. *Nat. Genet.* 39, 896–900. <https://doi.org/10.1038/ng2050>

- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., Takahashi, H., 2006. *Arabidopsis* SLIM1 Is a Central Transcriptional Regulator of Plant Sulfur Response and Metabolism. *Plant Cell* 18, 3235–3251. <https://doi.org/10.1105/tpc.106.046458>
- Maruyama-Nakashita, A., Nakamura, Y., Watanabe-Takahashi, A., Inoue, E., Yamaya, T., Takahashi, H., 2005. Identification of a novel cis-acting element conferring sulfur deficiency response in *Arabidopsis* roots: Sulfur-responsive cis-acting element of SULTR1;1. *Plant J.* 42, 305–314. <https://doi.org/10.1111/j.1365-313X.2005.02363.x>
- Maruyama-Nakashita, A., Nakamura, Y., Watanabe-Takahashi, A., Yamaya, T., Takahashi, H., 2004a. Induction of SULTR1;1 Sulfate Transporter in *Arabidopsis* Roots Involves Protein Phosphorylation/Dephosphorylation Circuit for Transcriptional Regulation. *Plant Cell Physiol.* 45, 340–345. <https://doi.org/10.1093/pcp/pch029>
- Maruyama-Nakashita, A., Nakamura, Y., Yamaya, T., Takahashi, H., 2004b. A novel regulatory pathway of sulfate uptake in *Arabidopsis* roots: implication of CRE1/WOL/AHK4-mediated cytokinin-dependent regulation. *Plant J.* 38, 779–789. <https://doi.org/10.1111/j.1365-313X.2004.02079.x>
- Maruyama-Nakashita, A., Watanabe-Takahashi, A., Inoue, E., Yamaya, T., Saito, K., Takahashi, H., 2015. Sulfur-Responsive Elements in the 3'-Nontranscribed Intergenic Region Are Essential for the Induction of *SULFATE TRANSPORTER 2;1* Gene Expression in *Arabidopsis* Roots under Sulfur Deficiency. *Plant Cell* 27, 1279–1296. <https://doi.org/10.1105/tpc.114.134908>
- Mathesius, U., Keijzers, G., Natera, S.H., Weinman, J.J., Djordjevic, M.A., Rolfe, B.G., 2001. Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics* 1, 1424–1440. [https://doi.org/10.1002/1615-9861\(200111\)1:11<1424::AID-PROT1424>3.0.CO;2-J](https://doi.org/10.1002/1615-9861(200111)1:11<1424::AID-PROT1424>3.0.CO;2-J)
- Matityahu, I., Godo, I., Hacham, Y., Amir, R., 2019. The level of threonine in tobacco seeds is limited by substrate availability, while the level of methionine is limited also by the activity of cystathionine  $\gamma$ -synthase. *Plant Sci.* 283, 195–201. <https://doi.org/10.1016/j.plantsci.2019.02.020>
- Matityahu, I., Godo, I., Hacham, Y., Amir, R., 2013. Tobacco seeds expressing feedback-insensitive cystathionine gamma-synthase exhibit elevated content of methionine and altered primary metabolic profile. *BMC Plant Biol.* 13, 206. <https://doi.org/10.1186/1471-2229-13-206>
- Mcgrath, S., Zhao, F.-J., Blake-Kalff, M.M., 2003. History and outlook for sulphur fertilizers in Europe.
- McKersie, B.D., Bowley, S.R., Harjanto, E., Leprince, O., 1996. Water-Deficit Tolerance and Field Performance of Transgenic Alfalfa Overexpressing Superoxide Dismutase. *Plant Physiol.* 111, 1177–1181. <https://doi.org/10.1104/pp.111.4.1177>
- Mertens, C., Dehon, L., Bourgeois, A., Verhaeghe-Cartryse, C., Blecker, C., 2012. Agronomical factors influencing the legumin/vicilin ratio in pea (*Pisum sativum* L.) seeds: Agronomical factors influencing legumin/vicilin ratio in pea seeds. *J. Sci. Food Agric.* 92, 1591–1596. <https://doi.org/10.1002/jsfa.4738>
- Meskauskiene, R., Würsch, M., Laloi, C., Vidi, P.-A., Coll, N.S., Kessler, F., Baruah, A., Kim, C., Apel, K., 2009. A mutation in the *Arabidopsis* mTERF-related plastid protein SOLDAT10 activates retrograde signaling and suppresses  $1O_2$ -induced cell death. *Plant J.* 60, 399–410. <https://doi.org/10.1111/j.1365-313X.2009.03965.x>

- Molvig, L., Tabe, L.M., Eggum, B.O., Moore, A.E., Craig, S., Spencer, D., Higgins, T.J.V., 1997. Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *Proc. Natl. Acad. Sci.* 94, 8393–8398. <https://doi.org/10.1073/pnas.94.16.8393>
- Moreau, C., Knox, M., Turner, L., Rayner, T., Thomas, J., Philpott, H., Belcher, S., Fox, K., Ellis, N., Domoney, C., 2018. Recombinant inbred lines derived from cultivars of pea for understanding the genetic basis of variation in breeders' traits. *Plant Genet. Resour. Charact. Util.* 16, 424–436. <https://doi.org/10.1017/S1479262118000345>
- Most, P., Papenbrock, J., 2015. Possible Roles of Plant Sulfurtransferases in Detoxification of Cyanide, Reactive Oxygen Species, Selected Heavy Metals and Arsenate. *Molecules* 20, 1410–1423. <https://doi.org/10.3390/molecules20011410>
- Mugford, S.G., Yoshimoto, N., Reichelt, M., Wirtz, M., Hill, L., Mugford, S.T., Nakazato, Y., Noji, M., Takahashi, H., Kramell, R., Gigolashvili, T., Flügge, U.-I., Wasternack, C., Gershenzon, J., Hell, R., Saito, K., Kopriva, S., 2009. Disruption of Adenosine-5'-Phosphosulfate Kinase in *Arabidopsis* Reduces Levels of Sulfated Secondary Metabolites. *Plant Cell* 21, 910–927. <https://doi.org/10.1105/tpc.109.065581>
- Müller, B., Fastner, A., Karmann, J., Mansch, V., Hoffmann, T., Schwab, W., Suter-Grotemeyer, M., Rentsch, D., Truernit, E., Ladwig, F., Bleckmann, A., Dresselhaus, T., Hammes, U.Z., 2015. Amino Acid Export in Developing *Arabidopsis* Seeds Depends on UmamiT Facilitators. *Curr. Biol.* 25, 3126–3131. <https://doi.org/10.1016/j.cub.2015.10.038>
- New, B., Duthion, C., Turc, O., 1994. Phenological Response of Pea to Water Stress during Reproductive Development. *Crop Sci.* 34, 141–146. <https://doi.org/10.2135/cropsci1994.0011183X003400010025x>
- Newell, M.A., Vogel, K.E., Adams, M., Aydin, N., Bodnar, A.L., Ali, M., Lauter, A.N., Scott, M., 2014. Genetic and biochemical differences in populations bred for extremes in maize grain methionine concentration. *BMC Plant Biol.* 14, 49. <https://doi.org/10.1186/1471-2229-14-49>
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., Bamshad, M., Nickerson, D.A., Shendure, J., 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276. <https://doi.org/10.1038/nature08250>
- Nguyen, H.C., Hoefgen, R., Hesse, H., 2012. Improving the nutritive value of rice seeds: elevation of cysteine and methionine contents in rice plants by ectopic expression of a bacterial serine acetyltransferase. *J. Exp. Bot.* 63, 5991–6001. <https://doi.org/10.1093/jxb/ers253>
- Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F., van Kleunen, M., 2010. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* 15, 684–692. <https://doi.org/10.1016/j.tplants.2010.09.008>
- Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H., Hoefgen, R., 2003. Transcriptome analysis of sulfur depletion in *Arabidopsis thaliana*: interlacing of biosynthetic pathways provides response specificity. *Plant J.* 18.
- Noji, M., Saito, M., Nakamura, M., Aono, M., Saji, H., Saito, K., 2001. Cysteine Synthase Overexpression in Tobacco Confers Tolerance to Sulfur-Containing Environmental Pollutants. *Plant Physiol.* 126, 973–980. <https://doi.org/10.1104/pp.126.3.973>
- Obendorf, R.L., Górecki, R.J., 2012. Soluble carbohydrates in legume seeds. *Seed Sci. Res.* 22, 219–242. <https://doi.org/10.1017/S0960258512000104>
- Okumoto, S., Schmidt, R., Tegeder, M., Fischer, W.N., Rentsch, D., Frommer, W.B., Koch, W., 2002. High Affinity Amino Acid Transporters Specifically Expressed in Xylem

- Parenchyma and Developing Seeds of *Arabidopsis*. *J. Biol. Chem.* 277, 45338–45346. <https://doi.org/10.1074/jbc.M207730200>
- Osborne, T.B., 1909. *The vegetable proteins*, by Thomas B. Osborne. Longmans, Green, London,.
- Pandurangan, S., Diapari, M., Yin, F., Munholland, S., Perry, G.E., Chapman, B.P., Huang, S., Sparvoli, F., Bollini, R., Crosby, W.L., Pauls, K.P., Marsolais, F., 2016. Genomic Analysis of Storage Protein Deficiency in Genetically Related Lines of Common Bean (*Phaseolus vulgaris*). *Front. Plant Sci.* 7. <https://doi.org/10.3389/fpls.2016.00389>
- Pandurangan, S., Sandercock, M., Beyaert, R., Conn, K.L., Hou, A., Marsolais, F., 2015. Differential response to sulfur nutrition of two common bean genotypes differing in storage protein composition. *Front. Plant Sci.* 6. <https://doi.org/10.3389/fpls.2015.00092>
- Parreira, J.R., Balestrazzi, A., Fevereiro, P., Araújo, S.D.S., 2018. Maintaining Genome Integrity during Seed Development in *Phaseolus vulgaris* L.: Evidence from a Transcriptomic Profiling Study. *Genes* 9, 463. <https://doi.org/10.3390/genes9100463>
- Pecrix, Y., Staton, S.E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., Blein, T., Jardinaud, M.-F., Latrasse, D., Zouine, M., Zahm, M., Kreplak, J., Mayjonade, B., Satgé, C., Perez, M., Cauet, S., Marande, W., Chantry-Darmon, C., Lopez-Roques, C., Bouchez, O., Bérard, A., Debelle, F., Muñoz, S., Bendahmane, A., Bergès, H., Niebel, A., Buitink, J., Frugier, F., Benhamed, M., Crespi, M., Gouzy, J., Gamas, P., 2018. Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat. Plants* 4, 1017–1025. <https://doi.org/10.1038/s41477-018-0286-7>
- Perrot, C., 1995. Les protéines de pois : de leur fonction dans la graine à leur utilisation en alimentation animale. *INRAE Prod. Anim.* 8, 151–164. <https://doi.org/10.20870/productions-animales.1995.8.3.4122>
- Pigliucci, M., 1996. How organisms respond to environmental changes: from phenotypes to molecules (and vice versa). *Trends Ecol. Evol.* 11, 168–173. [https://doi.org/10.1016/0169-5347\(96\)10008-2](https://doi.org/10.1016/0169-5347(96)10008-2)
- Planta, J., Xiang, X., Leustek, T., Messing, J., 2017. Engineering sulfur storage in maize seed proteins without apparent yield loss. *Proc. Natl. Acad. Sci.* 114, 11386–11391. <https://doi.org/10.1073/pnas.1714805114>
- Pysh, L.D., Aukerman, M.J., Schmidt, R.J., 1993. OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2. *Plant Cell* 5, 227–236. <https://doi.org/10.1105/tpc.5.2.227>
- Quesada, V., 2016. The roles of mitochondrial transcription termination factors (MTERFs) in plants. *Physiol. Plant.* 157, 389–399. <https://doi.org/10.1111/ppl.12416>
- R Core Team, 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Ranocha, P., McNeil, S.D., Ziemak, M.J., Li, C., Tarczynski, M.C., Hanson, A.D., 2001. The S-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity: S-methylmethionine cycle. *Plant J.* 25, 575–584. <https://doi.org/10.1046/j.1365-313x.2001.00988.x>
- Ravel, S., Gakiere, B., Job, D., Douce, R., 1998. The specific features of methionine biosynthesis and metabolism in plants. *Proc. Natl. Acad. Sci.* 95, 7805–7812. <https://doi.org/10.1073/pnas.95.13.7805>
- Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, andreas, Tewes, A., Ezcurra, I., Rask, L., Baumlein, H., 2000. Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J.* 21, 401–408. <https://doi.org/10.1046/j.1365-313x.2000.00686.x>

- Rémond, D., Walrand, S., 2017. Les graines de légumineuses: caractéristiques nutritionnelles et effets sur la santé. <https://doi.org/10.15454/1.5138524482202214E12>
- Righetti, K., Vu, J.L., Pelletier, S., Vu, B.L., Glaab, E., Lalanne, D., Pasha, A., Patel, R.V., Provart, N.J., Verdier, J., Leprince, O., Buitink, J., 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* tpc.15.00632. <https://doi.org/10.1105/tpc.15.00632>
- Robinson, D.G., Oliviusson, P., Hinz, G., 2005. Protein Sorting to the Storage Vacuoles of Plants: A Critical Appraisal. *Traffic* 6, 615–625. <https://doi.org/10.1111/j.1600-0854.2005.00303.x>
- Ronfort, J., Bataillon, T., Santoni, S., Delalande, M., David, J.L., 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, 28. <https://doi.org/10.1186/1471-2229-6-28>
- Rubio, L.A., Pérez, A., Ruiz, R., Guzmán, M.Á., Aranda-Olmedo, I., Clemente, A., 2014. Characterization of pea (*Pisum sativum*) seed protein fractions: Characterization of pea (*Pisum sativum*) seed protein fractions. *J. Sci. Food Agric.* 94, 280–287. <https://doi.org/10.1002/jsfa.6250>
- Rzewuski, G., Cornell, K.A., Rooney, L., Bürstenbinder, K., Wirtz, M., Hell, R., Sauter, M., 2007. OsMTN encodes a 5'-methylthioadenosine nucleosidase that is up-regulated during submergence-induced ethylene synthesis in rice (*Oryza sativa* L.). *J. Exp. Bot.* 58, 1505–1514. <https://doi.org/10.1093/jxb/erm014>
- Saito, K., 2000. Regulation of sulfate transport and synthesis of sulfur-containing amino acids. *Curr. Opin. Plant Biol.* 3, 188–195. [https://doi.org/10.1016/S1369-5266\(00\)00063-7](https://doi.org/10.1016/S1369-5266(00)00063-7)
- Sakuraba, Y., Yanagisawa, S., 2018. Light signalling-induced regulation of nutrient acquisition and utilisation in plants. *Semin. Cell Dev. Biol.* 83, 123–132. <https://doi.org/10.1016/j.semcdb.2017.12.014>
- Santo, S.D., Stamp, H., Rozhon, W., Heuck, A., Clausen, T., Jonak, C., n.d. Stress-Induced GSK3 Regulates the Redox Stress Response by Phosphorylating Glucose-6-Phosphate Dehydrogenase in ArabidopsisC W OA 14.
- Santos, C.S., Carbas, B., Castanho, A., Vasconcelos, M.W., Vaz Pato, M.C., Domoney, C., Brites, C., 2019. Variation in Pea (*Pisum sativum* L.) Seed Quality Traits Defined by Physicochemical Functional Properties. *Foods* 8, 570. <https://doi.org/10.3390/foods8110570>
- Sauter, M., Moffatt, B., Saechao, M.C., Hell, R., Wirtz, M., 2013. Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* 451, 145–154. <https://doi.org/10.1042/BJ20121744>
- Sawada, Y., Toyooka, K., Kuwahara, A., Sakata, A., Nagano, M., Saito, K., Hirai, M.Y., 2009. Arabidopsis Bile Acid:Sodium Symporter Family Protein 5 is Involved in Methionine-Derived Glucosinolate Biosynthesis. *Plant Cell Physiol.* 50, 1579–1586. <https://doi.org/10.1093/pcp/pcp110>
- Schiltz, S., Munier-Jolain, N., Jeudy, C., Burstin, J., Salon, C., 2005. Dynamics of Exogenous Nitrogen Partitioning and Nitrogen Remobilization from Vegetative Organs in Pea Revealed by <sup>15</sup>N in Vivo Labeling throughout Seed Filling. *Plant Physiol.* 137, 1463–1473. <https://doi.org/10.1104/pp.104.056713>
- Schroeder, H.E., 1982. Quantitative studies on the cotyledonary proteins in the genus *Pisum*. *J. Sci. Food Agric.* 33, 623–633. <https://doi.org/10.1002/jsfa.2740330707>

- Sexton, P.J., Paek, N.C., Shibles, R.M., 1998. Effects of nitrogen source and timing of sulfur deficiency on seed yield and expression of 11S and 7S seed storage proteins of soybean. *Field Crops Res.* 59, 1–8. [https://doi.org/10.1016/S0378-4290\(98\)00101-4](https://doi.org/10.1016/S0378-4290(98)00101-4)
- Sexton, P.J., Shibles, R.M., 1999. Activity of ATP Sulfurylase in Reproductive Soybean. *Crop Sci.* 39, 131–135. <https://doi.org/10.2135/cropsci1999.0011183X003900010021x>
- Shao, H.-B., Guo, Q.-J., Chu, L.-Y., Zhao, X.-N., Su, Z.-L., Hu, Y.-C., Cheng, J.-F., 2007. Understanding molecular mechanism of higher plant plasticity under abiotic stress. *Colloids Surf. B Biointerfaces* 54, 37–45. <https://doi.org/10.1016/j.colsurfb.2006.07.002>
- Shen, B., Li, C., Tarczynski, M.C., 2002. High free-methionine and decreased lignin content result from a mutation in the *Arabidopsis S*-adenosyl-L-methionine synthetase 3 gene. *Plant J.* 29, 371–380. <https://doi.org/10.1046/j.1365-313X.2002.01221.x>
- Shibagaki, N., Grossman, A.R., 2010. Binding of Cysteine Synthase to the STAS Domain of Sulfate Transporter and Its Regulatory Consequences. *J. Biol. Chem.* 285, 25094–25102. <https://doi.org/10.1074/jbc.M110.126888>
- Siol, M., Jacquin, F., Chabert-Martinello, M., Smýkal, P., Le Paslier, M.-C., Aubert, G., Burstin, J., 2017. Patterns of Genetic Structure and Linkage Disequilibrium in a Large Collection of Pea Germplasm. *G3 and Genomes Genetics* 7, 2461–2471. <https://doi.org/10.1534/g3.117.043471>
- Song, S., Hou, W., Godo, I., Wu, C., Yu, Y., Matiyahu, I., Hacham, Y., Sun, S., Han, T., Amir, R., 2013. Soybean seeds expressing feedback-insensitive cystathionine  $\gamma$ -synthase exhibit a higher content of methionine. *J. Exp. Bot.* 64, 1917–1926. <https://doi.org/10.1093/jxb/ert053>
- South, P.F., Walker, B.J., Cavanagh, A.P., Rolland, V., Badger, M., Ort, D.R., n.d. Bile Acid Sodium Symporter BASS6 Can Transport Glycolate and Is Involved in Photorespiratory Metabolism in 16.
- Spencer, D., Rerie, W., Randall, P., Higgins, T., 1990. The Regulation of Pea Seed Storage Protein Genes by Sulfur Stress. *Funct. Plant Biol.* 17, 355. <https://doi.org/10.1071/PP9900355>
- Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., Bharti, A.K., Farmer, A.D., Zhou, P., Denny, R., May, G.D., Erlandson, S., Yakub, M., Sugawara, M., Sadowsky, M.J., Young, N.D., Tiffin, P., 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. <https://doi.org/10.1371/journal.pone.0065688>
- Sun, W.Q., Leopold, A.C., 1997. Cytoplasmic Vitrification and Survival of Anhydrobiotic Organisms. *Comp. Biochem. Physiol. A Physiol.* 117, 327–333. [https://doi.org/10.1016/S0300-9629\(96\)00271-X](https://doi.org/10.1016/S0300-9629(96)00271-X)
- Sun, X.D., Arntfield, S.D., 2012. Molecular forces involved in heat-induced pea protein gelation: Effects of various reagents on the rheological properties of salt-extracted pea protein gels. *Food Hydrocoll.* 28, 325–332. <https://doi.org/10.1016/j.foodhyd.2011.12.014>
- Suzuki, M., Kao, C.-Y., Cocciolone, S., McCarty, D.R., 2002. Maize VP1 complements *Arabidopsis abi3* and confers a novel ABA/auxin interaction in roots: Functional conservation of VP1/ABI3. *Plant J.* 28, 409–418. <https://doi.org/10.1046/j.1365-313X.2001.01165.x>
- Suzuki, M., Sato, Y., Wu, S., Kang, B.-H., McCarty, D.R., 2015. Conserved Functions of the MATE Transporter BIG EMBRYO1 in Regulation of Lateral Organ Size and Initiation Rate. *Plant Cell* 27, 2288–2300. <https://doi.org/10.1105/tpc.15.00290>

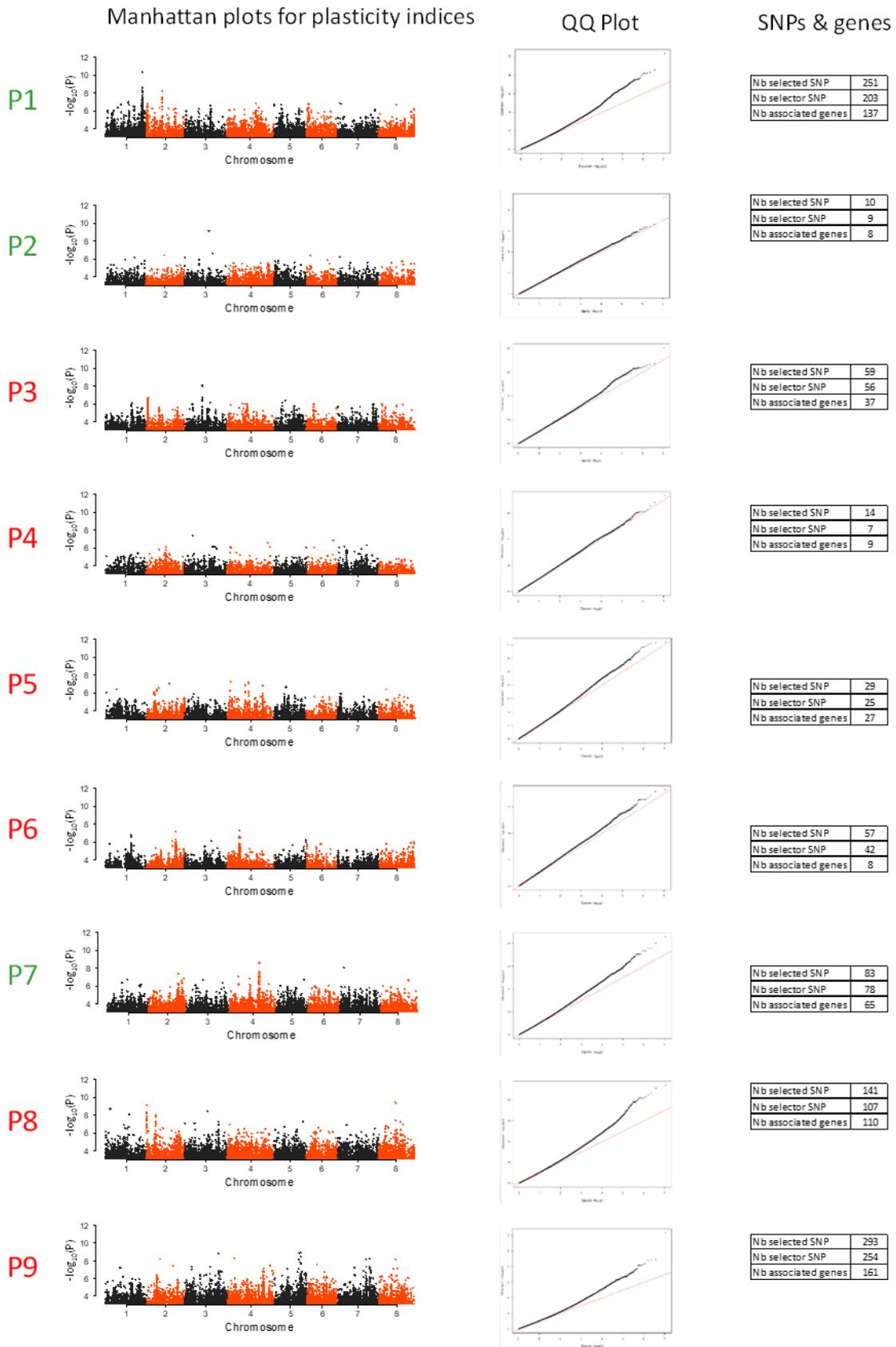
- Suzuki, M., Wang, H.H.-Y., McCarty, D.R., 2007. Repression of the *LEAFY COTYLEDON 1/B3* Regulatory Network in Plant Embryo Development by *VP1 / ABSCISIC ACID INSENSITIVE 3 - LIKE B3* Genes. *Plant Physiol.* 143, 902–911. <https://doi.org/10.1104/pp.106.092320>
- Suzuki, M., Wu, S., Li, Q., McCarty, D.R., 2014. Distinct functions of COAR and B3 domains of maize VP1 in induction of ectopic gene expression and plant developmental phenotypes in Arabidopsis. *Plant Mol. Biol.* 85, 179–191. <https://doi.org/10.1007/s11103-014-0177-x>
- Tabata, S., Stougaard, J. (Eds.), 2014. The Lotus japonicus Genome, Compendium of Plant Genomes. Springer Berlin Heidelberg, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-44270-8>
- Tabe, L., Wirtz, M., Molvig, L., Droux, M., Hell, R., 2010. Overexpression of serine acetyltransferase produced large increases in O-acetylserine and free cysteine in developing seeds of a grain legume. *J. Exp. Bot.* 61, 721–733. <https://doi.org/10.1093/jxb/erp338>
- Tabe, L.M., Droux, M., 2002. Limits to Sulfur Accumulation in Transgenic Lupin Seeds Expressing a Foreign Sulfur-Rich Protein. *Plant Physiol.* 128, 1137–1148. <https://doi.org/10.1104/pp.010935>
- Tafesse, E.G., Gali, K.K., Lachagari, V.B.R., Bueckert, R., Warkentin, T.D., 2020. Genome-Wide Association Mapping for Heat Stress Responsive Traits in Field Pea. *Int. J. Mol. Sci.* 21, 2043. <https://doi.org/10.3390/ijms21062043>
- Takahashi, H., Kopriva, S., Giordano, M., Saito, K., Hell, R., 2011. Sulfur Assimilation in Photosynthetic Organisms: Molecular Functions and Regulations of Transporters and Assimilatory Enzymes. *Annu. Rev. Plant Biol.* 62, 157–184. <https://doi.org/10.1146/annurev-arplant-042110-103921>
- Tan, Q., Zhang, L., Grant, J., Cooper, P., Tegeder, M., 2010. Increased Phloem Transport of S-Methylmethionine Positively Affects Sulfur and Nitrogen Metabolism and Seed Development in Pea Plants. *Plant Physiol.* 154, 1886–1896. <https://doi.org/10.1104/pp.110.166389>
- Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K.L., Yandell, M., Gundlach, H., Mayer, K.F., Schwartz, D.C., Town, C.D., 2014. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* 15, 312. <https://doi.org/10.1186/1471-2164-15-312>
- Tang, X., Hou, A., Babu, M., Nguyen, V., Hurtado, L., Lu, Q., Reyes, J.C., Wang, A., Keller, W.A., Harada, J.J., Tsang, E.W.T., Cui, Y., 2008. The Arabidopsis BRAHMA Chromatin-Remodeling ATPase Is Involved in Repression of Seed Maturation Genes in Leaves. *Plant Physiol.* 147, 1143–1157. <https://doi.org/10.1104/pp.108.121996>
- Tar'an, B., Warkentin, T., Somers, D.J., Miranda, D., Vandenberg, A., Blade, S., Bing, D., 2004. Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). *Euphytica* 136, 297–306. <https://doi.org/10.1023/B:EUPH.0000032721.03075.a0>
- Tashiro, Y., Rodriguez, G.M., Atsumi, S., 2015. 2-Keto acids based biosynthesis pathways for renewable fuels and chemicals. *J. Ind. Microbiol. Biotechnol.* 42, 361–373. <https://doi.org/10.1007/s10295-014-1547-8>
- Tayeh, N., Aluome, C., Falque, M., Jacquin, F., Klein, A., Chauveau, A., Bérard, A., Houtin, H., Rond, C., Kreplak, J., Boucherot, K., Martin, C., Baranger, A., Pilet-Nayel, M., Warkentin, T.D., Brunel, D., Marget, P., Le Paslier, M., Aubert, G., Burstin, J., 2015. Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map. *Plant J.* 84, 1257–1273. <https://doi.org/10.1111/tpj.13070>

- Taylor, M., Chapman, R., Beyaert, R., Hernández-Sebastià, C., Marsolais, F., 2008. Seed Storage Protein Deficiency Improves Sulfur Amino Acid Content in Common Bean (*Phaseolus vulgaris* L.): Redirection of Sulfur from  $\gamma$ -Glutamyl-*S*-methyl-cysteine. *J. Agric. Food Chem.* 56, 5647–5654. <https://doi.org/10.1021/jf800787y>
- Timmerman-Vaughan, G.M., Mills, A., Whitfield, C., Frew, T., Butler, R., Murray, S., Lakeman, M., McCallum, J., Russell, A., Wilson, D., 2005. Linkage Mapping of QTL for Seed Yield, Yield Components, and Developmental Traits in Pea. *Crop Sci.* 45, 1336–1344. <https://doi.org/10.2135/cropsci2004.0436>
- Ufaz, S., Galili, G., 2008. Improving the Content of Essential Amino Acids in Crop Plants: Goals and Opportunities. *Plant Physiol.* 147, 954–961. <https://doi.org/10.1104/pp.108.118091>
- Upadhyay, N., Kar, D., Mahajan, B.D., Nanda, S., Rahiman, R., Panchakshari, N., Bhagavatula, L., Datta, S., 2019. The multitasking abilities of MATE transporters in plants 14.
- Valassakis, C., 2019. Novel interactions of Selenium Binding Protein family with the PICOT containing proteins AtGRXS14 and AtGRXS16 in *Arabidopsis thaliana*. *Plant Sci.* 11.
- Valassakis, C., Livanos, P., Minopetrou, M., Haralampidis, K., Roussis, A., 2018. Promoter analysis and functional implications of the selenium binding protein (SBP) gene family in *Arabidopsis thaliana*. *J. Plant Physiol.* 224–225, 19–29. <https://doi.org/10.1016/j.jplph.2018.03.008>
- Vauclare, P., Kopriva, S., Fell, D., Suter, M., Sticher, L., von Ballmoos, P., Krähenbühl, U., den Camp, R.O., Brunold, C., 2002. Flux control of sulphate assimilation in *Arabidopsis thaliana*: adenosine 5'-phosphosulphate reductase is more susceptible than ATP sulphurylase to negative control by thiols. *Plant J.* 31, 729–740. <https://doi.org/10.1046/j.1365-313X.2002.01391.x>
- Via, S., 1993. Adaptive Phenotypic Plasticity: Target or By-Product of Selection in a Variable Environment? *Am. Nat.* 142, 352–365. <https://doi.org/10.1086/285542>
- Vitale, A., Hinz, G., 2005. Sorting of proteins to storage vacuoles: how many mechanisms? *Trends Plant Sci.* 10, 316–323. <https://doi.org/10.1016/j.tplants.2005.05.001>
- Vollmann, J., Fritz, C.N., Wagentristl, H., Ruckenbauer, P., 2000. Environmental and genetic variation of soybean seed protein content under Central European growing conditions. *J. Sci. Food Agric.* 80, 1300–1306. [https://doi.org/10.1002/1097-0010\(200007\)80:9<1300::AID-JSFA640>3.0.CO;2-I](https://doi.org/10.1002/1097-0010(200007)80:9<1300::AID-JSFA640>3.0.CO;2-I)
- Wang, H., Lin, J., Li, X.G., Chang, Y., 2015. Genome-wide identification of pear HD-Zip gene family and expression patterns under stress induced by drought, salinity, and pathogen. *Acta Physiol. Plant.* 37, 189. <https://doi.org/10.1007/s11738-015-1933-5>
- Wang, W., Vinocur, B., Altman, A., 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. <https://doi.org/10.1007/s00425-003-1105-5>
- Wang, Y., Bouwmeester, K., 2017. L-type lectin receptor kinases: New forces in plant immunity. *PLOS Pathog.* 13, e1006433. <https://doi.org/10.1371/journal.ppat.1006433>
- Warkentin, T.D., Smýkal, P., Coyne, C.J., Weeden, N., Domoney, C., Bing, D.-J., Leonforte, A., Xuxiao, Z., Dixit, G.P., Boros, L., McPhee, K.E., McGee, R.J., Burstin, J., Ellis, T.H.N., 2015. Pea, in: De Ron, A.M. (Ed.), *Grain Legumes, Handbook of Plant Breeding*. Springer New York, New York, NY, pp. 37–83. [https://doi.org/10.1007/978-1-4939-2797-5\\_2](https://doi.org/10.1007/978-1-4939-2797-5_2)
- Waterworth, W.M., Bray, C.M., West, C.E., 2019. Seeds and the Art of Genome Maintenance. *Front. Plant Sci.* 10, 706. <https://doi.org/10.3389/fpls.2019.00706>

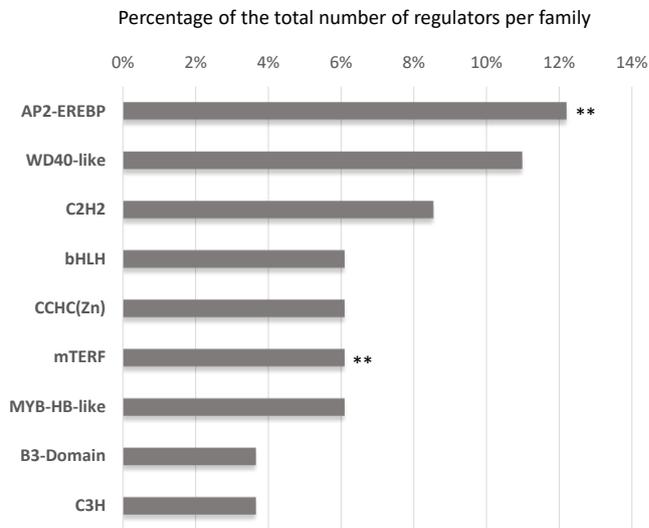
- Waterworth, W.M., Masnavi, G., Bhardwaj, R.M., Jiang, Q., Bray, C.M., West, C.E., 2010. A plant DNA ligase is an important determinant of seed longevity. *Plant J.* 63, 848–860. <https://doi.org/10.1111/j.1365-313X.2010.04285.x>
- Wawrzyńska, A., Lewandowska, M., Sirko, A., 2010. *Nicotiana tabacum* EIL2 directly regulates expression of at least one tobacco gene induced by sulphur starvation. *J. Exp. Bot.* 61, 889–900. <https://doi.org/10.1093/jxb/erp356>
- Wawrzyńska, A., Moniuszko, G., Sirko, A., 2015. Links Between Ethylene and Sulfur Nutrition—A Regulatory Interplay or Just Metabolite Association? *Front. Plant Sci.* 6. <https://doi.org/10.3389/fpls.2015.01053>
- Wawrzyńska, A., Sirko, A., 2016. EIN3 interferes with the sulfur deficiency signaling in *Arabidopsis thaliana* through direct interaction with the SLIM1 transcription factor. *Plant Sci.* 253, 50–57. <https://doi.org/10.1016/j.plantsci.2016.09.002>
- Wawrzyńska, A., Sirko, A., 2014. To control and to be controlled: understanding the *Arabidopsis* SLIM1 function in sulfur deficiency through comprehensive investigation of the EIL protein family. *Front. Plant Sci.* 5. <https://doi.org/10.3389/fpls.2014.00575>
- Weeden, N.F., 2007. Genetic Changes Accompanying the Domestication of *Pisum sativum*: Is there a Common Genetic Basis to the “Domestication Syndrome” for Legumes? *Ann. Bot.* 100, 1017–1025. <https://doi.org/10.1093/aob/mcm122>
- Wirtz, M., Birke, H., Heeg, C., Müller, C., Hosp, F., Throm, C., König, S., Feldman-Salit, A., Rippe, K., Petersen, G., Wade, R.C., Rybin, V., Scheffzek, K., Hell, R., 2010. Structure and Function of the Hetero-oligomeric Cysteine Synthase Complex in Plants. *J. Biol. Chem.* 285, 32810–32817. <https://doi.org/10.1074/jbc.M110.157446>
- Wirtz, M., Hell, R., 2007. Dominant-Negative Modification Reveals the Regulatory Function of the Multimeric Cysteine Synthase Protein Complex in Transgenic Tobacco. *Plant Cell* 19, 625–639. <https://doi.org/10.1105/tpc.106.043125>
- W-K Ng, D., Hall, T.C., 2008. PvALF and FUS3 activate expression from the phaseolin promoter by different mechanisms. *Plant Mol. Biol.* 66, 233–244. <https://doi.org/10.1007/s11103-007-9265-5>
- Wongkaew, A., Asayama, K., Kitaiwa, T., Nakamura, S.-I., Kojima, K., Stacey, G., Sekimoto, H., Yokoyama, T., Ohkama-Ohtsu, N., 2018. AtOPT6 Protein Functions in Long-Distance Transport of Glutathione in *Arabidopsis thaliana*. *Plant Cell Physiol.* <https://doi.org/10.1093/pcp/pcy074>
- Wu, J., Tan, X., Wu, C., Cao, K., Li, Y., Bao, Y., 2013. Regulation of Cytokinesis by Exocyst Subunit SEC6 and KEULE in *Arabidopsis thaliana*. *Mol. Plant* 6, 1863–1876. <https://doi.org/10.1093/mp/sst082>
- Xi, J., Qiu, Y., Du, L., Poovaiah, B.W., 2012. Plant-specific trihelix transcription factor AtGT2L interacts with calcium/calmodulin and responds to cold and salt stresses. *Plant Sci.* 185–186, 274–280. <https://doi.org/10.1016/j.plantsci.2011.11.013>
- Xiang, L., Etxeberria, E., den Ende, W., 2013. Vacuolar protein sorting mechanisms in plants. *FEBS J.* 280, 979–993. <https://doi.org/10.1111/febs.12092>
- Xiang, X., Wu, Y., Planta, J., Messing, J., Leustek, T., 2018. Overexpression of serine acetyltransferase in maize leaves increases seed-specific methionine-rich zeins. *Plant Biotechnol. J.* 16, 1057–1067. <https://doi.org/10.1111/pbi.12851>
- Yang, S.F., Hoffman, N.E., 1984. Ethylene Biosynthesis and its Regulation in Higher Plants. *Annu. Rev. Plant Physiol.* 35, 155–189. <https://doi.org/10.1146/annurev.pp.35.060184.001103>
- Yatusevich, R., Mugford, S.G., Matthewman, C., Gigolashvili, T., Frerigmann, H., Delaney, S., Koprivova, A., Flüge, U.-I., Kopriva, S., 2010. Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis*

- thaliana: Glucosinolate biosynthesis and sulfate assimilation*. *Plant J.* 62, 1–11.  
<https://doi.org/10.1111/j.1365-313X.2009.04118.x>
- Yoshimoto, N., Inoue, E., Watanabe-Takahashi, A., Saito, K., Takahashi, H., 2007. Posttranscriptional Regulation of High-Affinity Sulfate Transporters in Arabidopsis by Sulfur Nutrition. *Plant Physiol.* 145, 378–388. <https://doi.org/10.1104/pp.107.105742>
- Zagorchev, L., Seal, C., Kranner, I., Odjakova, M., 2013. A Central Role for Thiols in Plant Tolerance to Abiotic Stress. *Int. J. Mol. Sci.* 14, 7405–7432.  
<https://doi.org/10.3390/ijms14047405>
- Zhang, B., Pasini, R., Dan, H., Joshi, N., Zhao, Y., Leustek, T., Zheng, Z.-L., 2014. Aberrant gene expression in the Arabidopsis *SULTR1;2* mutants suggests a possible regulatory role for this sulfate transporter in response to sulfur nutrient status. *Plant J.* 77, 185–197. <https://doi.org/10.1111/tpj.12376>
- Zhang, J.-Y., Cruz De Carvalho, M.H., Torres-Jerez, I., Kang, Y., Allen, S.N., Huhman, D.V., Tang, Y., Murray, J., Sumner, L.W., Udvardi, M.K., 2014. Global reprogramming of transcription and metabolism in *Medicago truncatula* during progressive drought and after rewatering: Drought-stressed *Medicago truncatula*. *Plant Cell Environ.* 37, 2553–2576. <https://doi.org/10.1111/pce.12328>
- Zhang, Y., Schernthaner, J., Labbé, N., Hefford, M.A., Zhao, J., Simmonds, D.H., 2014. Improved protein quality in transgenic soybean expressing a de novo synthetic protein, MB-16. *Transgenic Res.* 23, 455–467. <https://doi.org/10.1007/s11248-013-9777-5>
- Zhou, X., Stephens, M., 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11, 407–409.  
<https://doi.org/10.1038/nmeth.2848>
- Zhu, Y., Sheaffer, C.C., Russelle, M.P., Vance, C.P., 1998. Dry Matter Accumulation and Dinitrogen Fixation of Annual Medicago Species. *Agron. J.* 90, 103–108.  
<https://doi.org/10.2134/agronj1998.00021962009000010019x>
- Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Vu, B.L., Dubois-Laurent, C., Geoffriau, E., Signor, C.L., Dalmais, M., Gutbrod, K., Dörmann, P., Gallardo, K., Bendahmane, A., Buitink, J., Leprince, O., 2016. ABI5 Is a Regulator of Seed Maturation and Longevity in Legumes. *Plant Cell* 28, 2735–2754.  
<https://doi.org/10.1105/tpc.16.00470>
- Zuber, H., Poignavent, G., Le Signor, C., Aimé, D., Vieren, E., Tadla, C., Lugan, R., Belghazi, M., Labas, V., Santoni, A.-L., Wipf, D., Buitink, J., Avice, J.-C., Salon, C., Gallardo, K., 2013c. Legume adaptation to sulfur deficiency revealed by comparing nutrient allocation and seed traits in *Medicago truncatula*. *Plant J.* 76, 982–996.  
<https://doi.org/10.1111/tpj.12350>

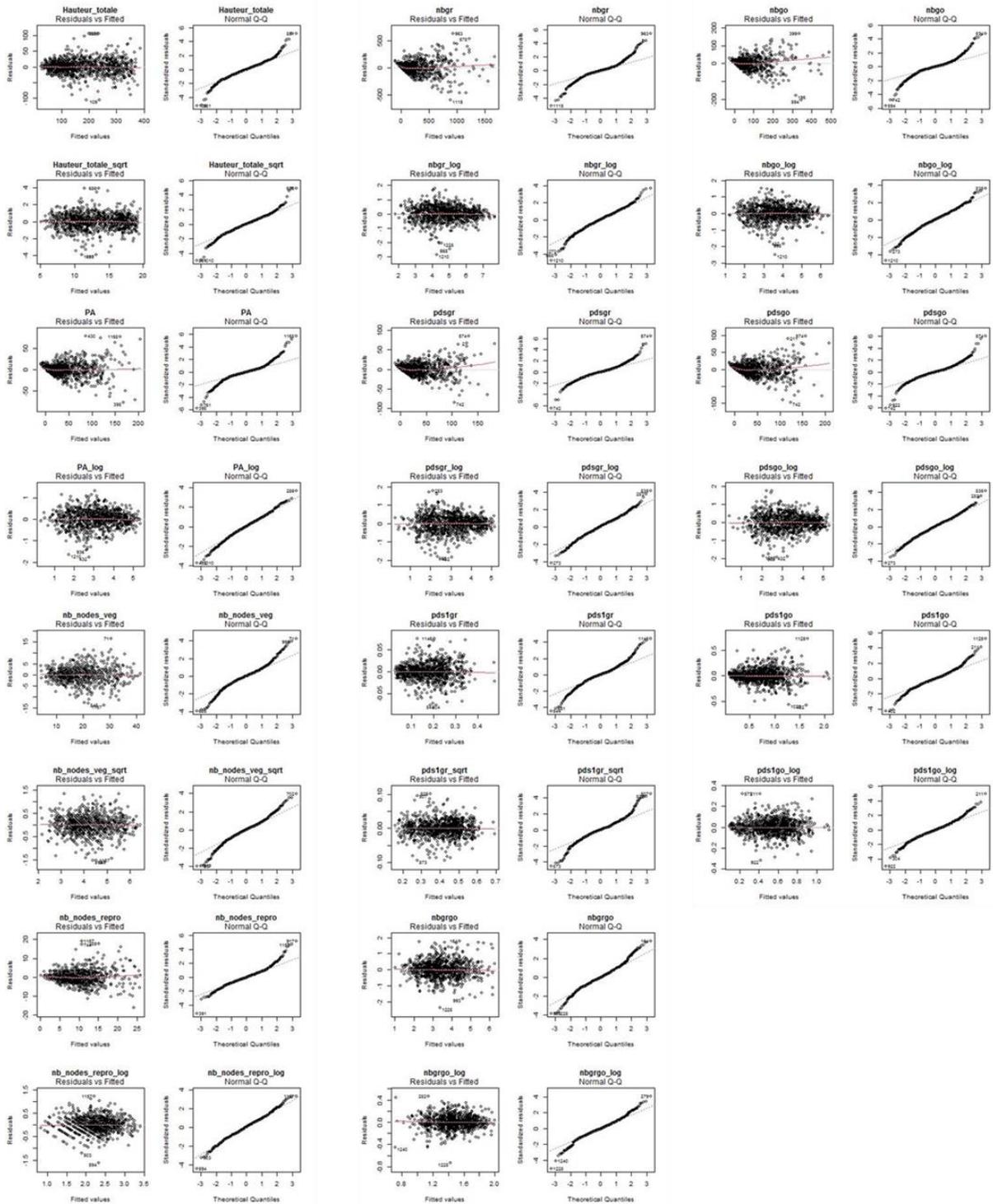
# ANNEXES



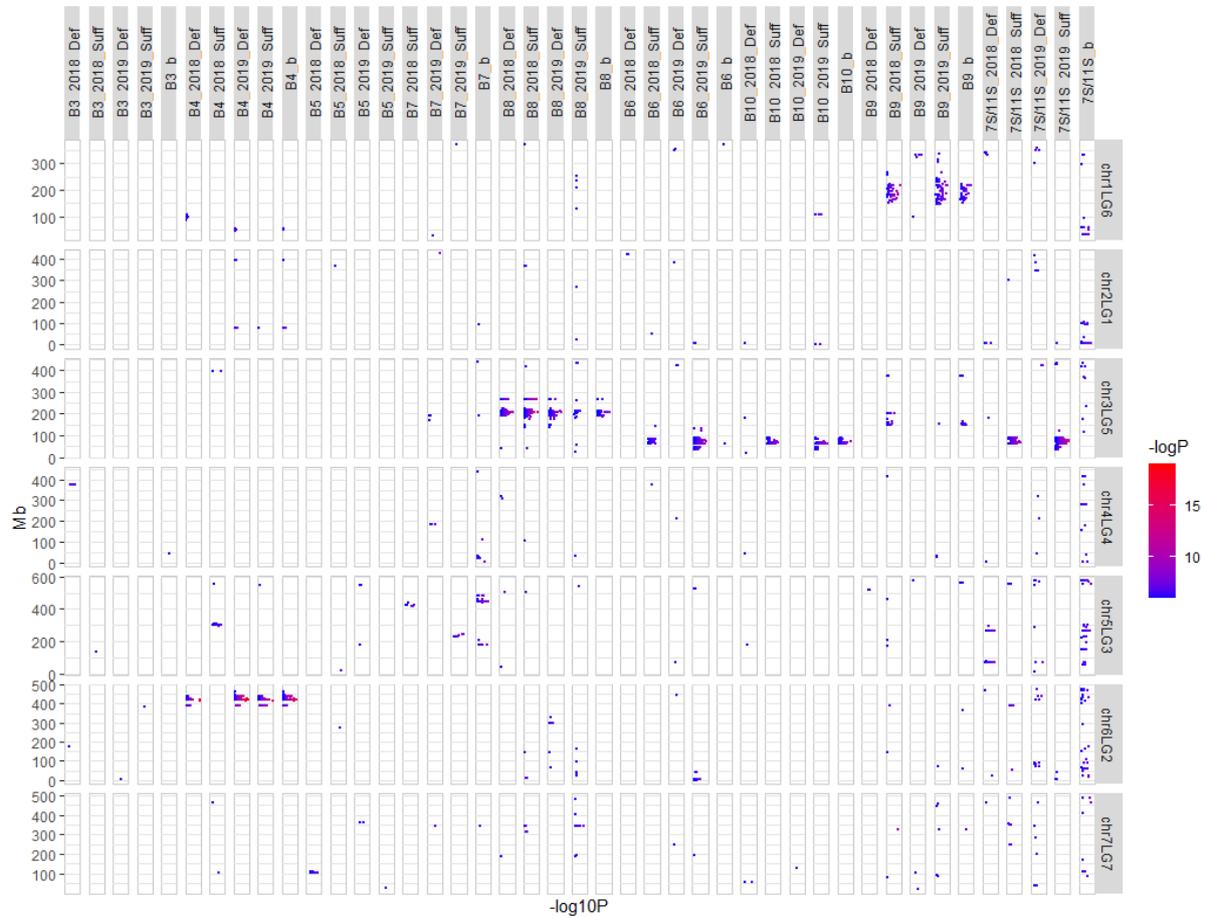
**Annexe 1: GWAS results of plasticity.** Manhattan plots, quantile-quantile (Q-Q) plots, and number of SNP & genes identified by GWAS from plasticity indices.



**Annexe 2 : The most highly represented TF families in the PG dataset.**



**Annexe 3 : Transformation des variables permettant l'homoscédasticité des résidus.**



**Annexe 4 : Manhattan plots associés à la composition protéique des graines de pois ainsi qu'à sa plasticité. Seuls les SNPs sélectionnés dans les analyses sont représentés.**



## DETERMINISME GENETIQUE DE LA PLASTICITE DE LA COMPOSITION PROTEIQUE DES GRAINES DE LEGUMINEUSES VIS-A-VIS DE L'ENVIRONNEMENT : ROLE DU METABOLISME DU SOUFRE

**Mots clés :** Genome-wide association studies (GWAS), protéines de réserve, pois, *M.truncatula*, globulines.

Le regain d'intérêt pour les protéines végétales stimule les recherches visant à développer des marqueurs d'aide à la sélection de variétés de légumineuses mieux adaptées aux besoins nutritionnels. Parmi les caractères à améliorer et à stabiliser est l'équilibre en acides aminés des graines, les acides aminés essentiels méthionine et tryptophane étant particulièrement sous-représentés dans les graines de légumineuses. Cette thèse est centrée sur la composition protéique des graines, un déterminant majeur de l'équilibre en acides aminés. L'objectif était d'explorer la variabilité génétique et environnementale de ce caractère et d'identifier des gènes potentiellement impliqués dans sa plasticité vis-à-vis de l'environnement.

Dans le premier volet de la thèse, la composition protéique des graines de 200 génotypes de *Medicago truncatula* cultivés dans quatre environnements contrastés (apport suffisant en eau, stress hydrique, et deux semis décalés) a été analysée par électrophorèse mono-dimensionnelle. Les indices de plasticité calculés pour chaque bande protéique ont permis d'identifier, par génétique d'association à l'échelle du génome (GWAS), des polymorphismes de séquence associés à des variations de plasticité des protéines de réserve majeures, les globulines 7S et 11S. La liste des gènes candidats est enrichie en gènes liés à la transcription, à la réparation de l'ADN et à la transduction du signal. D'autres gènes sur-représentés sont impliqués dans les métabolismes du soufre et de l'aspartate conduisant à la synthèse de méthionine et de lysine. En exploitant des données d'expression publiquement disponibles chez *Medicago truncatula*, et en replaçant ces gènes dans les voies métaboliques correspondantes, nous avons découvert l'importance des voies de recyclage de la méthionine dans le contrôle de la plasticité de la composition protéique des graines. Ce recyclage fait intervenir l'enzyme homocystéine S-méthyltransférase 3 impliquée dans la régénération de méthionine à partir de S-méthylméthionine, une forme de transport de méthionine dans le phloème. Cette section de la thèse a fait émerger de nouveaux gènes à cibler pour améliorer et stabiliser le contenu en acides aminés soufrés, ouvrant des perspectives de recherches translationnelles visant à améliorer la qualité des graines de protéagineux comme le pois.

Le second volet de la thèse avait pour objectifs d'explorer la variabilité de réponse d'une collection d'écotypes de pois à une carence en soufre, aux niveaux des composantes de rendement et de la composition protéique des graines, et d'identifier par GWAS des gènes potentiellement impliqués dans cette variabilité de réponse. Deux expériences ont été conduites deux années successives avec respectivement 304 et 198 génotypes développés dans deux conditions : apport suffisant ou carencé en soufre. L'effet année et l'effet de la carence en soufre étant significatifs sur les variables étudiées, des indices de plasticité reflétant l'amplitude de réponse vis-à-vis d'une carence en soufre ont été calculés pour chaque variable. Les résultats obtenus par GWAS ont mis en évidence des gènes candidats pour contrôler la composition protéique des graines dans les environnements pauvres en soufre. Plusieurs sont associés aux mécanismes de transport (ex. d'acides aminés) et de régulation post-traductionnelle, tandis que d'autres jouent un rôle dans le métabolisme du soufre ou sont au carrefour entre les métabolismes azotés et soufrés. Ce travail a permis de mettre en lumière des gènes de pois susceptibles de contrôler le statut azoté et soufré des graines, et ainsi de moduler l'équilibre entre les globulines 7S et 11S. La même approche GWAS menée au niveau des composantes de rendement a conduit à proposer des gènes candidats pertinents en particulier pour le contrôle du poids d'une graine.

L'ensemble des résultats obtenus offrent des perspectives d'amélioration variétale pour stabiliser le rendement et la composition protéique des graines de légumineuses dans des environnements fluctuants, notamment pour la disponibilité en soufre.

## GENETIC DETERMINISM OF THE PLASTICITY OF LEGUME SEED PROTEIN IN DIFFERENT ENVIRONMENTS: ROLE OF SULFUR METABOLISM

**Keywords:** Genome-wide association studies (GWAS), storage proteins, pea, *M. truncatula*, globulins.

The renewed interest in plant proteins has stimulated research aimed at developing markers to aid in the selection of legume varieties better adapted to nutritional needs. Among the traits to be improved and stabilized is the amino acid balance of seeds, the essential amino acids methionine and tryptophan being particularly under-represented in legume seeds. This thesis focuses on seed protein composition, which is a major determinant of the seed amino acid balance. The objective was to explore the genetic and environmental variability in this trait and to identify genes potentially involved in its plasticity when subjected to environmental stresses.

In the first part of the thesis, the protein composition of seeds from 200 *Medicago truncatula* genotypes cultivated in four contrasted environments (sufficient water supply, water stress, and two time-shifted sowing) was analyzed by one-dimensional electrophoresis. The plasticity indices calculated for each protein band made it possible to identify, by genome-wide association studies (GWAS), sequence polymorphisms associated with variations in the plasticity of the major storage proteins, the 7S and 11 globulins. The list of candidate genes is enriched in genes related to transcription, DNA repair and signal transduction. Other over-represented genes are involved in sulfur and aspartate metabolism leading to the synthesis of methionine and lysine. By exploiting publicly available expression datasets in *Medicago truncatula*, and by localizing these genes in the corresponding metabolic pathways, we discovered that methionine recycling pathways are likely to contribute to the plasticity of seed protein composition. This recycling involves the homocysteine *S*-methyltransferase 3 enzyme that regenerates methionine from *S*-methylmethionine, a long-distance transport form of methionine in the phloem. This part of the thesis has led to the identification of new genes to be targeted to improve and stabilize the amount of sulfur amino acids in seeds, opening prospects for translational research aimed at improving the quality of seeds in protein crops such as pea.

The second part of the thesis aimed at exploring the variability of response of a collection of pea ecotypes to sulfur deficiency, at the levels of yield components and of seed protein composition, and to identify by GWAS genes potentially involved in this variability of response. Two experiments were carried out in two successive years with respectively 304 and 198 genotypes developed under two conditions: sufficient or deficient in sulfur. Since both year effect and sulfur deficiency effect were significant on the variables, plasticity indices reflecting the amplitude of the sulfur deficiency response were calculated for each variable. The results obtained by GWAS have revealed candidate genes for controlling the protein composition of seeds in low sulfur environments. Several are associated with transport mechanisms (e.g. amino acids) and post-translational regulation, while others play a role in sulfur metabolism or in the cross-talk between nitrogen and sulfur metabolisms. This work has shed light on pea genes that might control the nitrogen and sulfur status of seeds, and thus modulate the balance between the 7S and 11S globulins. The same GWAS approach conducted on yield component traits has led to the proposal of relevant candidate genes in particular for seed weight control.

Altogether, these results offer prospects for varietal improvement to stabilize yield and the protein composition of legume seeds in fluctuating environments, in particular for the availability of sulfur.