



L'accumulation des protéines dans les graines de légumineuses

Karine Gallardo, Christine Le Signor et Judith Burstin
INRA, UMR 1347 Agroécologie Dijon, France

Manuscrit révisé le 6 mars 2017 - Publié le 23 mars 2017

Résumé : Le bénéfice agroécologique des légumineuses et le regain d'intérêt pour la consommation de protéines végétales stimulent les recherches cognitives et appliquées visant à optimiser la composition protéique des graines de légumineuses. Cet article présente une synthèse des connaissances sur le contrôle génétique et environnemental de l'accumulation des protéines majeures des graines de légumineuses, en vue de développer des variétés améliorées pour la qualité nutritionnelle des graines. Il montre également comment les nombreuses données génomiques et post-génomiques issues du haut-débit, et la recherche translationnelle visant à transférer les connaissances acquises entre espèces modèles et cultivées, permettent d'accélérer la découverte de gènes contrôlant l'accumulation des protéines de réserve pour ensuite moduler leur concentration ou leurs propriétés.

Mots-clés : Génétique d'association génome entier, graines, légumineuses, nutrition, protéines de réserve, recherche translationnelle.

Abstract : The agroecological benefits of legumes and the renewed interest in the consumption of vegetable proteins stimulate researches on cognitive and applied aspects of seed biology aiming at optimizing protein composition of legume seeds. This article presents the state of knowledge on the genetic and environmental control of the accumulation of major proteins found in legume seeds, with a view to developing varieties with improved nutritional seed quality. It also shows how high-throughput genomics and post-genomics data, along with translational research aiming at transferring knowledge between model and crop species, can accelerate the identification of genes controlling the accumulation of storage proteins to further modulate their accumulation or properties.

Keywords: Genetics, genome-wide association studies, legumes, nutrition, seeds, storage proteins, translational research.

Les graines de légumineuses accumulent des quantités importantes de protéines sans fertilisation azotée

Les graines de légumineuses protéagineuses que sont le pois (*Pisum sativum* L.), la féverole (*Vicia faba* L.) et les lupins (*Lupinus* sp.) contribuent à l'autonomie européenne en protéines végétales en raison de leur teneur élevée en protéines (23 à 40% selon les espèces, les génotypes et les conditions de milieu). Riches en lysine, elles sont majoritairement utilisées en alimentation animale et peuvent apporter un complément dans les rations à base de céréales plus pauvres en protéines et en cet acide aminé essentiel. Par ailleurs, leur valeur nutritionnelle et santé remet à la mode leur utilisation en nutrition humaine sous forme brute ou transformée.

Les légumineuses ont la remarquable capacité d'accumuler dans leurs graines des quantités importantes d'azote sous forme de protéines, même en l'absence de fertilisation azotée. Ceci est permis grâce aux symbioses racinaires qu'elles établissent avec des bactéries fixatrices d'azote de l'air et à leur capacité à remobiliser l'azote mis en réserve dans les organes végétatifs vers les graines.

La capacité des légumineuses à fixer symbiotiquement l'azote fait d'elles un véritable atout pour le développement de systèmes de cultures agroécologiques à bas niveau d'intrants. Pourtant, en France, seulement 2% des terres arables agricoles sont utilisées pour la production de légumineuses à graines (2015, source : Terres Univia, <http://www.terresunivia.fr/>). L'utilisation plus large de ces légumineuses permettrait d'enrichir naturellement les sols en azote et de réduire les émissions de gaz à effet de serre liées à l'utilisation d'engrais, tout en augmentant la production de protéines végétales pour l'alimentation humaine et animale. Pour répondre à ce double enjeu, il est nécessaire d'augmenter la consommation de graines de légumineuses (Santi, 2017) en proposant de nouveaux produits simplifiant leur utilisation (exemple des pâtes à base de protéines de légumineuses et des yaourts enrichis en protéines végétales) et en améliorant la qualité nutritionnelle des graines. Un caractère majeur de qualité nutritionnelle à améliorer est la proportion d'acides aminés soufrés (méthionine et cystéine) et de tryptophane, chacun représentant entre 2 et 4% de la teneur totale en acides aminés des graines de légumineuses européennes comme le pois et la féverole (Burstin *et al.*, 2011). Il est à noter que la méthionine et le tryptophane sont des acides aminés essentiels que l'homme et les animaux sont incapables de synthétiser et qu'ils doivent donc obligatoirement trouver dans leur alimentation. Cette amélioration doit en parallèle prendre en compte la digestibilité des fractions protéiques qui dépend de la structure des protéines et de la présence de facteurs antinutritionnels (e.g. inhibiteurs de la trypsine) dans les graines.

Les protéines des graines ont été classées en quatre groupes selon leur solubilité dans l'eau (albumines), les sels (globulines), l'alcool à 70% (les prolamines), ou les solutions acides ou alcaline (gluténines) (Osborne, 1924). Dans les graines matures de légumineuses, ce sont les globulines qui déterminent majoritairement la composition en acides aminés puisqu'elles représentent jusqu'à 70% des protéines totales. Communément appelées protéines de réserve, elles sont codées par une grande famille multigénique et ont été classées en deux sous-groupes selon leur coefficient de sédimentation en ultracentrifugation analytique : globulines 7S et 11S (Osborne, 1924). Leur nom varie entre les espèces. Par exemple, les globulines 7S sont nommées vicilines et convicilines chez le pois, phaséolines chez le haricot, et β -conglycinines chez le soja, et les 11S sont nommées légumines chez le pois et glycinines chez le soja. Chez le pois, les albumines, les vicilines et les légumines peuvent être séparées par chromatographie échangeuse d'ions, offrant la possibilité de comparer la quantité de chaque fraction entre géotypes ou en réponse aux variations environnementales (**Figure 1**). Les globulines 11S sont plus avantageuses d'un point de vue nutritionnel, étant plus riches en acides aminés soufrés que les 7S. Les protéines PA1 et PA2 appartenant à la famille des albumines contiennent d'avantage d'acides aminés soufrés que les globulines mais leur proportion dans les graines (<10%) ne laisse que très peu de marge d'amélioration par la voie génétique classique. Des progrès ont été obtenus par transgénèse *via* le transfert de gènes étrangers codant une albumine 2S riche en méthionine chez le soja (*Glycine max* L.) et le lupin (*Lupinus angustifolius* L.) (Altenbach *et al.*, 1989, Molvig *et al.*, 1997). Bien que ces plantes transgéniques produisent des graines plus riches en méthionine, les protéines 2S introduites ont généralement des propriétés allergéniques (Nordlee *et al.*, 1996, Pastorello *et al.*, 2001) et elles s'accumulent au détriment d'autres composés soufrés comme le glutathion bien connu pour son rôle antioxydant dans les processus de défense (Tabe and Droux, 2002).

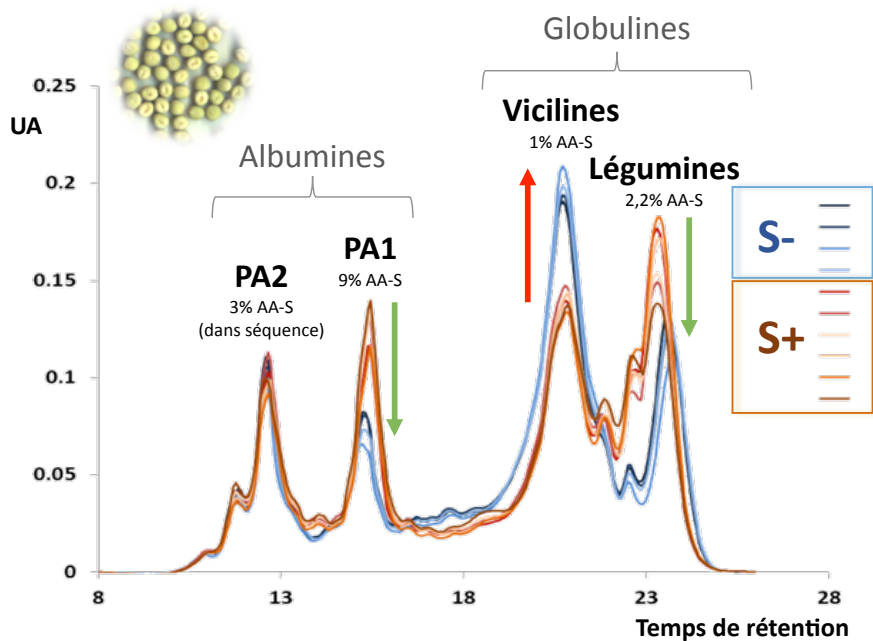


Figure 1. Effet d'une carence en soufre lors de la phase reproductive sur la composition protéique des graines de pois. Les graines sont issues de pois (génotype Cameor) cultivés en présence de soufre (S+) ou en l'absence de soufre dans le milieu pendant la phase reproductive (S-). Le profil a été obtenu par chromatographie liquide haute performance. Dans chaque groupe de protéines (albumines et globulines), la carence en soufre réduit l'accumulation des protéines les plus riches en acides aminés soufrés (AA-S) : légumine (globulines 11S) et albumine de pois PA1 (flèches vertes). Cette diminution est compensée par l'accumulation des protéines de réserve les plus pauvres en AA-S (flèche rouge): les vicilines (globulines 7S).

Au-delà de l'aspect nutritionnel, les protéines majeures des graines jouent un rôle lors de la croissance initiale (hétérotrophe) des plantules, puisqu'elles fournissent, sous l'action de protéases, les acides aminés nécessaires au métabolisme (Tan-Wilson and Wilson, 2012). Ainsi, les protéines accumulées lors du développement des graines déterminent non seulement la valeur nutritionnelle mais aussi le potentiel germinatif ou de levée dont l'optimisation contribuera à promouvoir l'utilisation des légumineuses locales, comme le pois ou la féverole, dans les cycles culturaux. Des interactions fortes entre équipes spécialistes des protéines de la graine et celles focalisées sur les aspects de germination ont permis l'élaboration de projets communs visant à étudier le lien entre la composition de la graine et sa vigueur germinative ou tolérance à la conservation (projets QualityLegSeed-ANR-06-GPLA-0008 et REGULEG-ANR-15-CE20-0001 ; Vandecasteele *et al.*, 2011 ; Zuber *et al.*, 2013).

Déterminisme de l'accumulation des protéines dans les graines

Les contraintes environnementales, telles que le stress hydrique ou la chaleur, ont généralement un fort impact sur le rendement et la taille des graines mais n'ont que très peu d'influence sur la concentration en protéines des graines (Prudent *et al.*, 2016 ; Salon *et al.*, 2011). C'est probablement en ajustant le nombre et la taille de leurs graines que les plantes maintiennent le flux d'azote vers les graines en développement, leur permettant ainsi d'accumuler des protéines pour les étapes ultérieures de germination et de croissance initiale. Toutefois, des variations significatives dans la composition protéique des graines, par exemple dans le rapport globulines 7S/11S, se produisent en réponse aux changements environnementaux qui peuvent influencer à la fois l'équilibre en acides aminés et la digestibilité (Bourgeois *et al.*, 2009). L'identification des gènes permettant de maintenir sous contrainte environnementale le niveau des protéines les plus avantageuses aux niveaux nutritionnel et technologique est un enjeu majeur. En permettant l'accumulation des protéines les plus riches en acides aminés soufrés, la nutrition soufrée joue un rôle prépondérant dans le maintien de la valeur nutritionnelle des graines (Chandler *et al.*, 1983 ; Figure 1), d'où l'importance de

prendre en compte cette nutrition lors de la conception de systèmes de cultures agroécologiques, et de s'intéresser aux mécanismes favorisant l'allocation de soufre vers les graines dans différents environnements. Des études ont révélé un effet bénéfique de la symbiose mycorhizienne à arbuscules sur la nutrition soufrée des légumineuses (Casieri *et al.*, 2012 ; Giovannetti *et al.*, 2014 ; Sieh *et al.*, 2013), suggérant que cette symbiose contribuerait à stabiliser la valeur nutritionnelle des graines. D'autres travaux ont montré la capacité qu'ont les légumineuses d'accumuler des quantités élevées de sulfate dans leurs vacuoles lorsque la nutrition soufrée est suffisante (Zuber *et al.*, 2013). L'établissement de ce puits de sulfate et sa remobilisation vers les graines sont probablement des leviers d'action pour contourner les problèmes de carence en soufre en fin de cycle lorsque les sols s'appauvrissent en nutriments et/ou en eau (Gallardo *et al.*, 2014).

Les mesures d'épuration des fumées industrielles ont conduit à une baisse des retombées de soufre atmosphérique sur les sols européens, diminuant la disponibilité en sulfate, précurseur de la synthèse des acides aminés soufrés et de nombreuses autres molécules nécessaires au développement des graines.

De nombreux travaux ont étudié le contrôle génétique de l'accumulation des protéines dans les graines de légumineuses. Chez le pois, il existe une variabilité de la composition protéique des graines entre géotypes suffisante pour permettre un large éventail d'applications alimentaires (Bourgeois *et al.*, 2009 ; Gueguen and Barbot, 1988 ; Tzitzikas *et al.*, 2006). La teneur en protéines varie aussi significativement entre géotypes de pois (de 18 à 32%, Gueguen and Lemarié, 1996), et pour un géotype donné, des variations dans le clivage des différents isoformes de vicilines (Bourgeois *et al.*, 2009) offrent des perspectives d'amélioration de la solubilité/digestibilité de la fraction viciline. La variabilité génétique a été exploitée pour identifier les régions des génomes du pois et de la légumineuse modèle *Medicago truncatula* (*M. truncatula*)¹ qui contrôlent l'accumulation des différentes globulines (Bourgeois *et al.*, 2011 ; Le Signor *et al.*, 2017) (**Figure 2**). Pour ce faire, l'approche PQL (Protein Quantity Loci), décrite ci-après, a été privilégiée. Les régions génomiques détectées chez le pois, peuvent être utilisées dans les programmes de sélection assistée par marqueurs pour optimiser la qualité des graines de variétés performantes du point de vue du rendement. Toutefois, la limite de cette approche est qu'elle ne donne pas directement accès aux gènes responsables (causaux) dont la découverte offre des perspectives directes d'amélioration *via* la recherche de formes (polymorphismes) favorables de ces gènes. Pour accélérer l'identification des gènes causaux dans ces régions génomiques, des méthodes dites de génotypage par séquençage ont été développées. Elles permettent d'enrichir les cartes génétiques en marqueurs (cartes à haute densité) et de préciser la région contenant le polymorphisme en cause, qui est ensuite comparée aux génomes de référence, c'est-à-dire entièrement séquencés, pour rechercher des gènes candidats. De tels outils sont maintenant disponibles chez le pois. Par exemple, une carte génétique dense en marqueurs a été développée (Tayeh *et al.*, 2015). Elle permet d'affiner la détection des régions génomiques contrôlant les caractères de qualité des graines, comme ceux présentés **Figure 2** pour la composition protéique.

Pour mettre en évidence les régions du génome contrôlant l'accumulation des différentes protéines de réserve, l'approche PQL développée par Damerval *et al.* (1994) a été appliquée aux graines d'une population de plantes issues d'un croisement entre deux géotypes contrastés pour la composition protéique de la graine. L'approche consiste à séparer les protéines des graines en gels d'électrophorèse bidimensionnelle, puis à utiliser les données quantitatives

¹ *Medicago truncatula* est une légumineuse fourragère choisie comme modèle des légumineuses (en 2000) de par son génome de petite taille (360 Mb) plus facile à séquencer que celui des espèces au génome de grande taille comme le pois (5000 Mb).

obtenues pour chaque protéine afin d'identifier (ou cartographier), via une recherche de QTL (Quantitative Trait Loci), les régions du génome responsables des variations de leur abondance (Figure 2).

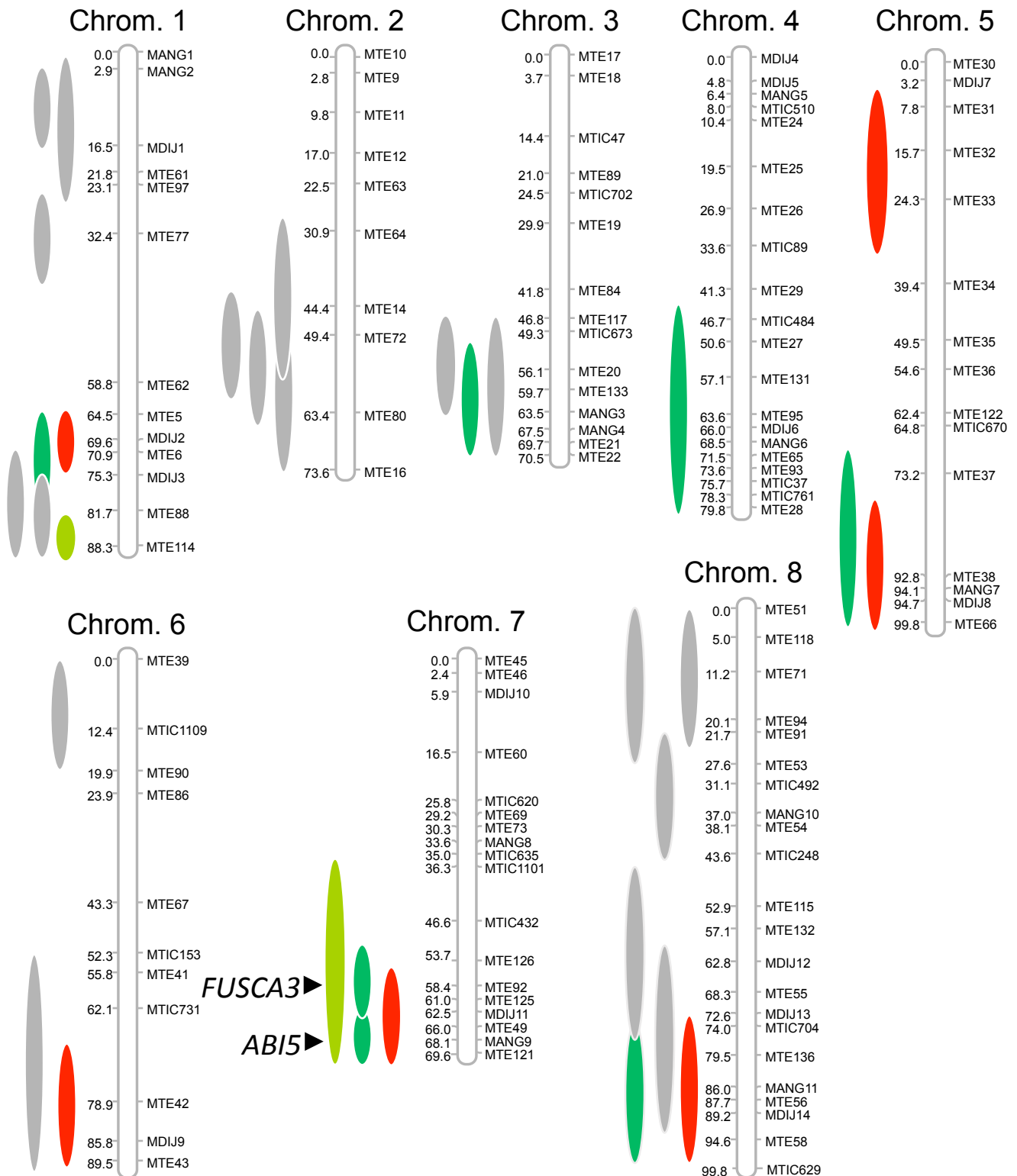


Figure 2. Régions PQL (Protein Quantity Loci) pour les globulines 7S et 11S chez *Medicago truncatula*. La carte génétique présentée a été obtenue à partir de la population de lignées recombinantes issue du croisement entre les génotypes parentaux Jemalong 6 et DZA 315.16 (carte initialement développée par Thierry Huguet, ENSAT Toulouse; puis complétée par Vandecasteele *et al.*, 2011). Les ovals montrent les régions de génome (groupes de PQLs) qui contrôlent l'accumulation des protéines de réserve majeures des graines (légumine, vicilines et convicilines). Ces régions PQLs sont réparties sur les 8 chromosomes (Chrom.) de *Medicago truncatula*. Les ovals colorés montrent les régions PQLs conservées avec le pois pour les vicilines (vert), convicilines (kaki) et légumine (rouge). Les ovals gris montrent les régions PQLs uniquement détectées chez *M. truncatula*. La position de deux gènes codant les facteurs de transcription *FUSCA3* et ABA-insensitive 5 (*ABI5*) est indiquée. Ces données sont extraites de l'article Le Signor *et al.*, 2017.

Apport des espèces modèles et transfert d'informations vers les légumineuses cultivées

Comme la qualité des graines s'établit lors de la phase de développement sur la plante mère, cette phase complexe a fait l'objet de nombreuses études visant à identifier les gènes impliqués (Thompson *et al.*, 2009). En utilisant des espèces modèles de légumineuses au génome séquencé, des avancées significatives dans la connaissance des gènes exprimés lors de l'accumulation des protéines de réserve ont été réalisées. Des jeux de données transcriptomiques et protéomiques obtenus lors du remplissage sont disponibles chez *M. truncatula* (Benedito *et al.*, 2008 ; Gallardo *et al.*, 2007), et de nouvelles méthodes d'analyses permettent d'étudier les liens entre les différent(e)s transcrits/protéines identifié(e)s par ces techniques. Par exemple, un réseau de co-expression a été développé à partir de données transcriptomiques à différents stades du développement de la graine de *M. truncatula* (Righetti *et al.*, 2015). Les connections entre les gènes du réseau ont permis d'identifier des régulateurs de la longévité des graines (Zinsmeister *et al.*, 2016) et offrent la possibilité de sélectionner des gènes régulateurs de la qualité nutritionnelle.

Des réseaux de co-expression permettent, *via* des études de corrélations entre les profils d'expression de gènes issus des analyses « omiques », d'identifier des cibles potentielles de facteurs de transcription et le type de régulation (positive ou négative selon le coefficient de corrélation).

Des approches visant à transférer les acquis entre espèces modèles et cultivées sont mises en place pour les caractères de qualité des graines. Le séquençage récent des génomes d'espèces d'intérêt agronomique devrait accélérer ces initiatives de recherche translationnelle, qu'elles soient ciblées sur des gènes ou prennent en compte l'expression du génome entier. Chez le pois, grâce aux techniques de séquençage par RNAseq², un atlas d'expression de gènes a été développé qui permet de rechercher les homologues des gènes identifiés chez le modèle *M. truncatula* (Alves-Carvalho *et al.*, 2015). On peut citer ici l'exemple de la protéase SBT1.1 (subtilase 1.1), initialement identifiée dans le protéome des graines de *M. truncatula*, dont le gène co-localise avec un QTL de taille de la graine chez cette espèce modèle. Le lien entre cette protéase et le contrôle de la taille de la graine a été validé chez le pois (D'Erfurth *et al.*, 2012), démontrant la faisabilité du transfert de données entre des espèces proches au niveau génétique mais qui diffèrent cependant dans leur architecture et la composition des graines : *M. truncatula* est une fourragère aux petites graines riches en protéines (~40%) alors que le pois produit des grosses graines à la fois riches en protéines (~23%) et en amidon (~50%). Ce constat encourage la mise en œuvre d'approches globales visant à comparer les régions génomiques contrôlant les caractères de qualité des graines entre légumineuses à graines et espèces modèles pour ensuite cibler des régions conservées et accélérer l'identification ou la validation des gènes causaux en exploitant les données (e.g. approches « omiques »), outils (e.g. transgénése) ou ressources (e.g. mutants) disponibles chez les modèles.

Comme exemple de recherche translationnelle, une comparaison entre le pois et *M. truncatula* des régions génomiques contrôlant l'accumulation des globulines (approche PQL) a été réalisée grâce au développement de marqueurs moléculaires reliant les génomes de ces deux espèces (Bordat *et al.*, 2011, Tayeh *et al.*, 2015). Les régions conservées entre ces espèces (**Figure 2**), suggérant des processus de régulation communs, ont fait l'objet d'une étude approfondie. Les données post-génomiques disponibles chez l'espèce modèle, comme l'atlas d'expression de gènes (Benedito *et al.*, 2008), ont permis d'identifier dans ces régions cinq gènes codant des

2 Le RNAseq (séquençage de l'ARN) est une méthode de séquençage aléatoire du transcriptome, basée sur des technologies à haut-débit permettant d'identifier et de quantifier l'ARN issu de la transcription du génome d'un tissu donné à un instant donné.

facteurs de régulation dont l'expression augmente de manière concomitante à l'accumulation des protéines de réserve (16-24 jours après la pollinisation). Deux de ces gènes sont localisés en bas du chromosome 7 (**Figure 2**). Le premier code le facteur de transcription FUSCA3 déjà décrit chez *Arabidopsis* pour son rôle dans l'accumulation des protéines de réserve (Kroj *et al.*, 2003), suggérant une fonction conservée entre différentes espèces qu'elles soient crucifères ou légumineuses. Le second code le facteur de transcription ABI5 impliqué dans la longévité des graines (Zinsmeister *et al.*, 2016). En caractérisant des plantes de pois porteuses de mutations dans le gène *ABI5*, son rôle dans le contrôle de la composition protéique de la graine de pois a été vérifié (Le Signor *et al.*, 2017). Les mutants produisent des graines pauvres en vicilines, corroborant des données obtenues chez le haricot qui montrent qu'ABI5 interagit avec la région promotrice des gènes codant les phaséolines (globulines 7S, Ng and Hall, 2008). Ce facteur de transcription aurait donc une fonction conservée entre légumineuses dans l'accumulation des globulines 7S. Le rapport légumine/viciline est augmenté dans les graines des mutants de pois pour *ABI5*, ce qui est avantageux du point de vue nutritionnel. Il n'est toutefois pas à négliger le second rôle d'ABI5 dans l'acquisition de la tolérance à la conservation (Zinsmeister *et al.*, 2016), nécessaire à la survie des graines.

Un pas de géant vers la connaissance des gènes contrôlant la composition protéique de la graine

Les progrès récents des technologies de séquençage ouvrent de nouvelles perspectives de recherche de gènes d'intérêt en vue d'améliorer les caractéristiques de qualité des graines. Elles permettent dorénavant d'explorer le génome entier de centaines de génotypes au sein de collections d'écotypes. Des millions de polymorphismes de séquences sont alors identifiés et utilisés pour des études de génétique d'association à l'échelle du génome. Cette approche, nommée GWAS pour « Genome Wide Association Studies », consiste à identifier les formes de gènes (ou variants alléliques) responsables des variations de caractères d'intérêt au sein d'une collection d'écotypes (**Figure 3A**). L'identification des gènes en jeu est facilitée par l'annotation des séquences d'un génome de référence, dont l'obtention est en cours pour plusieurs légumineuses, y compris le pois, dans le cadre de consortiums internationaux.

Cette approche GWAS a récemment été utilisée dans l'objectif d'établir un inventaire des gènes contrôlant l'accumulation des différentes protéines de réserve au sein de la graine de *M. truncatula* (**Figure 3A**), puis à transférer les acquis vers le pois (**Figure 3B**). Cette action est partie du constat que la majorité des connaissances sur la régulation de la synthèse des protéines de réserve provient d'approches ciblées sur la plante modèle des crucifères, *Arabidopsis* (Santos-Mendoza *et al.*, 2008), et, bien que des voies de transport/maturation des protéines de réserve vers les vésicules de stockage ont été proposées chez les légumineuses (Vitale and Hinz, 2005; Maruyama *et al.*, 2006), la liste des gènes impliqués reste à établir. Les rares gènes décrits sont impliqués dans le clivage et le compactage des protéines de réserve (Hara-Nishimura *et al.*, 1993; Jung *et al.*, 1998; Gruis *et al.*, 2002; Paris and Neuhaus, 2002; Shimada *et al.*, 2003).

La stratégie utilisée est présentée sur la **Figure 3A**. Les protéines des graines de 175 génotypes de *M. truncatula* (core-collection CC192, Ronfort *et al.*, 2006, Branca *et al.*, 2011) ont été quantifiées après séparation électrophorétique. Une recherche d'associations entre ces données dites de protéotypage et les variations de séquences des différents génotypes de *M. truncatula* (medicagohapmap.org, Stanton-Geddes *et al.*, 2013) a permis d'identifier un nombre important de gènes intervenant au niveau transcriptionnel ou post-transcriptionnel dans le contrôle de la composition protéique des graines (Le Signor *et al.*, 2017). Comme exemple de résultats issus de cette approche, il est à citer :

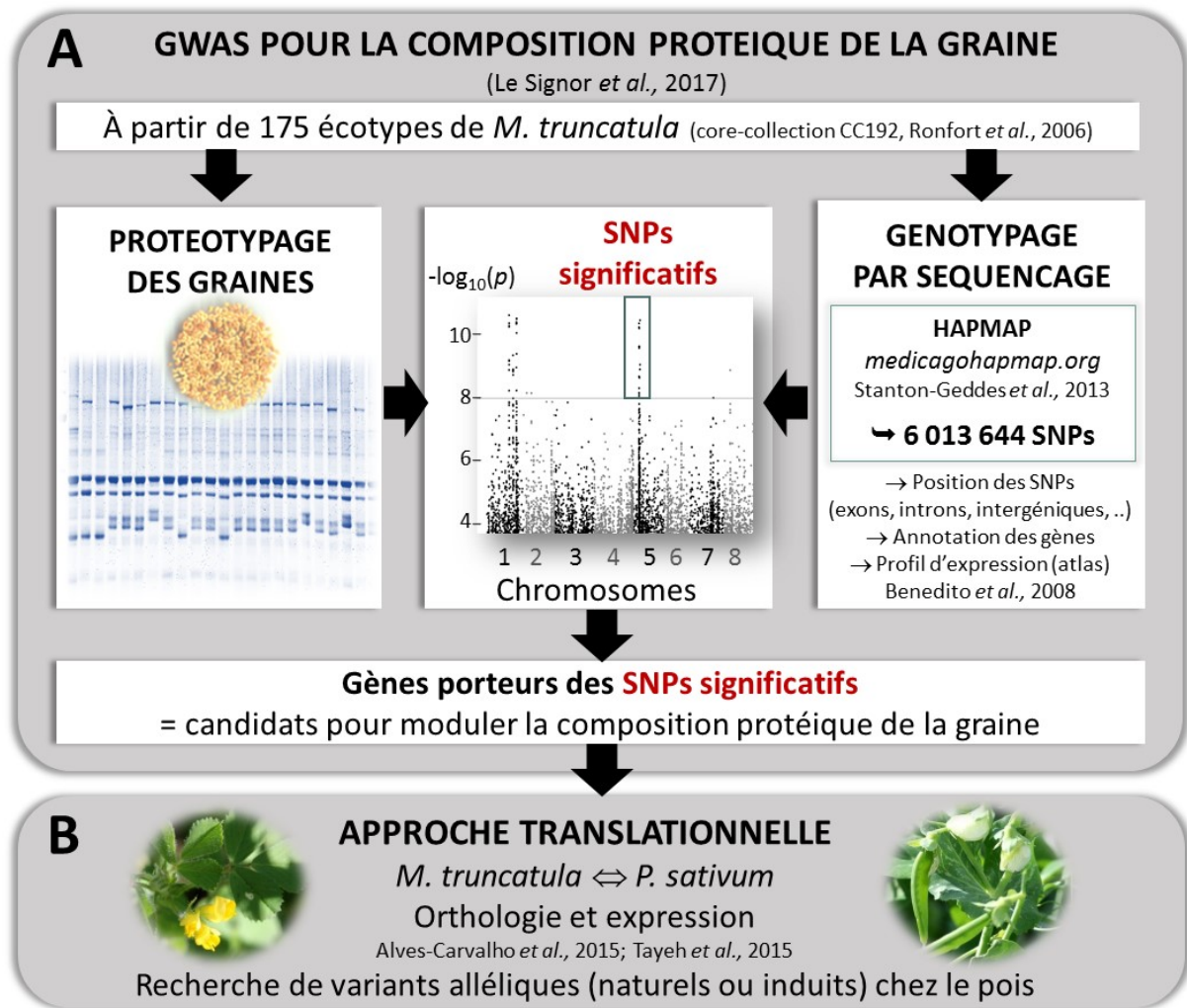


Figure 3. Génétique d'association génome entier (GWAS, A) et approche translationnelle (B) utilisées pour identifier des gènes candidats pour la qualité des graines de légumineuses.

A. Le protéotypage a été réalisé en gels d'électrophorèse mono-dimensionnelle à partir des extraits protéiques des graines de 175 écotypes de *M. truncatula* issus de la core-collection CC192. Ces écotypes ont été séquencés dans le cadre du projet HAPMAP, ce qui a permis d'identifier plus de 6 millions de polymorphismes pour un seul nucléotide (SNP). Des études statistiques d'association entre ces polymorphismes de séquences et les données de protéotypage ont permis d'identifier des gènes porteurs de SNPs candidats pour moduler la composition protéique de la graine.

B. Une approche translationnelle a ensuite permis de transférer les acquis vers le pois protéagineux en recherchant les gènes orthologues dans les bases de données publiques. L'exploitation de la variabilité génétique naturelle et induite chez le pois permettra d'identifier différentes formes de ces gènes pour des études fonctionnelles et technologiques.

↪ **La construction d'un modèle de régulation de la synthèse des globulines** en étudiant les profils d'expression, lors du développement de la graine, des gènes codant les facteurs de transcription identifiés par l'approche présentée **Figure 3A**. Ces gènes peuvent s'exprimer de manière concomitante à l'expression des gènes codant les globulines 7S ou 11S, suggérant un rôle dans l'activation de la synthèse des globulines. D'autres peuvent présenter un profil inverse, suggérant un rôle dans la répression de la synthèse des globulines. Un modèle de régulation impliquant 14 facteurs de transcription a ainsi été développé. Ce modèle comprend ABI5 et FUSCA3 (**Figure 2**), validant l'approche GWAS utilisée pour identifier des régulateurs clés de la composition protéique des graines.

↪ **Un inventaire des gènes associés aux différentes étapes de la mise en réserve des globulines** a été réalisé en ciblant les gènes dont la classification fonctionnelle suggérait un rôle dans la maturation ou le transport des protéines et en analysant systématiquement leurs séquences (domaines et homologies) et leurs profils d'expression au cours du développement de la graine. Ce travail a permis de mieux situer le rôle probable de 38 gènes dans les différentes

étapes de l'accumulation des globulines, depuis leur entrée dans le réticulum endoplasmique jusqu'à leur mise en réserve dans les vésicules de stockage ou corpuscules protéiques. Parmi ces gènes, 13 codent des protéases susceptibles de participer à la maturation des protéines de réserve, dont une enzyme VPE (Vacuolar Processing Enzyme) impliquée dans le clivage des formes précurseurs des légumineuses conduisant à la séparation des chaînes alpha et bêta et à la formation d'hexamères nécessaires au compactage de ces protéines (Gruis *et al.*, 2004).

→ **L'identification de 111 gènes susceptibles de moduler des modifications post-traductionnelles autres que protéolytiques, comme des glycosylations et phosphorylations.** Des modifications pourraient se produire au niveau des globulines et influencer leurs propriétés fonctionnelles (Casanova *et al.*, 2008), et d'autres pourraient réguler les processus conduisant à l'accumulation des globulines. Comme exemple, une protéine kinase fortement exprimée lors du remplissage des graines a été identifiée dont la séquence est très proche de la protéine BIN2 (BRASSINOSTEROID INSENSITIVE2) décrite chez *Arabidopsis* comme phosphorylant et stabilisant ABI5 lors de la germination des graines (Hu and Yu, 2014). En supposant une homologie fonctionnelle entre la kinase identifiée chez *M. truncatula* et BIN2, il serait intéressant d'étudier son rôle possible dans la régulation d'ABI5 lors de la synthèse des globulines 7S.

Optimisation de la composition des graines pour la meilleure valeur nutritionnelle possible

Si les facteurs antinutritionnels présents dans les graines (e.g. facteurs anti-trypsiques) ont été réduits dans les différentes variétés de légumineuses à graines cultivées, la composition protéique reste à améliorer. Ces améliorations peuvent être réalisées en manipulant les gènes codant les protéines de réserve (régulations cis au niveau du gène de structure ou du promoteur) ou d'autres gènes (trans) qui régulent l'expression ou l'abondance de ces protéines. Comme exemples d'applications, chez le soja et le haricot des mutations ont été identifiées dans les régions cis-ou trans-régulatrices des protéines de réserve (Hayashi *et al.*, 2009 ; Tsubokura *et al.*, 2012 ; Wang *et al.*, 2014 ; Pandurangan *et al.*, 2016) et des marqueurs spécifiques ont été développés pour sélectionner des variétés améliorées pour la qualité de la graine (Jegadeesan *et al.*, 2012 ; Song *et al.*, 2014).

Chez le pois, des variations alléliques au locus *Vc-2*, où se trouvent plusieurs gènes de vicilines, influencent la teneur en azote (Chinoy *et al.*, 2011), montrant l'influence de ces régulations cis. Concernant les régulations trans, nous disposons maintenant de l'identité de nombreux gènes candidats pour moduler la valeur nutritionnelle des graines. Pour 80% des gènes clés trans-régulateurs identifiés par GWAS chez *M. truncatula*, il existe une séquence orthologue dans les bases de données du pois, rendant possible la recherche de variants alléliques pour ces gènes (**Figure 3B**). Les données en cours d'obtention chez la féverole (projet Investissement d'Avenir PeaMust) permettront d'élargir la recherche de gènes orthologues à cette espèce. De larges collections d'écotypes existent chez le pois et la féverole (CRB protéagineux Dijon, Burstin *et al.*, 2015) qui peuvent être exploitées pour rechercher par une approche ciblée des variants alléliques de gènes candidats. Une nouvelle variabilité génétique a été créée chez le pois par la méthode TILLING (Targeting Induced Local Lesions in Genomes, McCallum *et al.*, 2000) qui permet de moduler la composition protéique des graines dans un même fond génétique (Dalmais *et al.*, 2008). L'approche consiste à rechercher par génétique inverse des plantes porteuses de mutations ponctuelles dans des gènes d'intérêt, puis à réaliser des rétrocroisements successifs avec le génotype sauvage afin d'éliminer les mutations résiduelles pour ensuite récolter les graines et tester leur valeur nutritionnelle. Ces ressources ouvrent des perspectives intéressantes d'identification d'allèles favorables à combiner dans un même génotype pour créer du matériel génétique innovant.

REMERCIEMENTS

Les données de génétique quantitative sur l'espèce modèle *M. truncatula* ont été obtenues dans le cadre de projets financés par l'Agence Nationale de la Recherche : QualityLegSeed (ANR-06-GPLA-0008) et GENOPEA (ANR-09-GENM-026). Nous remercions Florence Naudé (équipe ECP, UMR Agroécologie) pour la production du chromatogramme présenté Figure 1 et Delphine Aimé (équipe FILEAS, UMR Agroécologie) pour les analyses protéomiques (Figure 3).

REFERENCES BIBLIOGRAPHIQUES

- Altenbach S.B., Pearson K.W., Meeker G., Staraci L.C., Sun 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: 513-22.
- 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.
- 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: 504-13.
- Bordat A., Savoies V., Nicolas M., Salse J., Chauveau A., Bourgeois M., Potier J., Houtin H., Rond C., Murat F., Marget P., Aubert G., Burstin J. 2011. Translational Genomics in Legumes Allowed Placing *In Silico* 5460 Unigenes on the Pea Functional Map and Identified Candidate Genes in *Pisum sativum* L. G3 (Bethesda). 1: 93-103.
- 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.
- 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.
- 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., et al. 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences of the United States of America* 108: E864-870.
- Burstin J., Gallardo K., Mir R.R., Varshney R.K., Duc G. 2011. Improving Protein Content and Nutrition Quality (Eds Pratap A and Kumar J) CAB International. In *Biology and Breeding of Food Legumes*, p. 314-328.
- Burstin J., Salloignon P., Chabert-Martinello M., Magnin-Robert J.B., Siol M., Jacquin F., Chauveau A., Pont C., Aubert G., Delaitre C., Truntzer C., Duc G. 2015. Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics* 16: 105.
- Casanova F., Nakamura A., Masuda H., Lima L., Fialho E. 2008. Functionality of phosphorylated vicilin exposed to chemical and physical agents. *Food Chem.* 107: 1138-1143.
- Casieri L., Gallardo K., Wipf D. 2012. Transcriptional response of *Medicago truncatula* sulphate transporters to arbuscular mycorrhizal symbiosis with and without sulphur stress. *Planta* 235: 1431-4.
- Chandler P.M., Higgins T.J., 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.
- Chinoy C., Welham T., Turner L., Moreau C., Domoney C. 2011. The genetic control of seed quality traits: effects of allelic variation at the *Tri* and *Vc-2* genetic loci in *Pisum sativum* L. *Euphytica* 180: 107-122.

- Dalmais M., Schmidt J., Le Signor C., Moussy F., Burstin J., Savoie V., Aubert G., Brunaud V., de Oliveira Y., Guichard C., et al. 2008. UTILLdb, a *Pisum sativum* *in silico* forward and reverse genetics tool. *Genome Biol.* 9: R43.
- Damerval C., Maurice A., Josse J. M., De Vienne D. 1994. Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* 137: 289-301.
- D'Erfurth I., Le Signor C., Aubert G., Sanchez M., Vernoud V., Darchy B., Lherminier J., Bourion V., Bouteiller N., Bendahmane A., Buitink J., Prospero J-M., Thompson R., Burstin J., Gallardo K (2012) A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytol.* 196: 738-51.
- Gallardo K., Courty P-E., Le Signor C., Wipf D., Vernoud V. 2014. Sulfate transporters in the plant's response to drought and salinity: regulation and possible functions. *Front. Plant Sci.* 5: 580.
- 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.* 6: 2165-79.
- Giovannetti M., Tolosano M., Volpe V., Kopriva S., Bonfante P. 2014. Identification and functional characterization of a sulfate transporter induced by both sulfur starvation and mycorrhiza formation in *Lotus japonicus*. *New Phytol.* 204: 609-19.
- Gruis D., Schulze J., Jung R. 2004. Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. *The Plant Cell* 16: 270-290.
- Gruis D.F., Selinger D.A., Curran J.M., Jung R. 2002. Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. *The Plant Cell* 14: 2863-2882.
- Guéguen J., Lemarié J. (1996). "Composition, structure et propriétés physico-chimiques des protéines de légumineuses et d'oléagineux", *Protéines végétales*, B. Godon (ed.), Partie 3 : Propriétés biochimiques et physico chimiques des protéines végétales, Lavoisier, Paris, pp 80-119.
- Hayashi M., Kitamura K., and Harada K. 2009. Genetic mapping of *Cgdef* gene controlling accumulation of 7S globulin (β -conglycinin) subunits in soybean seeds. *J. Hered.* 100: 802-806.
- Hara-Nishimura I., Takeuchi Y., Nishimura M. 1993. Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *The Plant Cell* 5: 1651-1659.
- Hu Y., Yu D. 2014. BRASSINOSTEROID INSENSITIVE2 interacts with ABSCISIC ACID INSENSITIVE5 to mediate the antagonism of brassinosteroids to abscisic acid during seed germination in *Arabidopsis*. *The Plant Cell* 26: 4394-4408.
- Jegadeesan S., Yu K., Woodrow L., Wang Y., Shi C., and Poysa V. 2012. Molecular analysis of glycinin genes in soybean mutants for development of gene-specific markers. *Theor. Appl. Genet.* 124: 365-372.
- Jung R., Scott MP., Nam Y.W., Beaman T.W., Bassüner R., Saalbach I., Müntz K., Nielsen N.C. 1998. The role of proteolysis in the processing and assembly of 11S seed globulins. *The Plant Cell* 10: 343-357.
- Kroj T., Savino G., Valon C., Giraudat J., Parcy F. 2003. Regulation of storage protein gene expression in *Arabidopsis*. *Development.* 130: 6065-73.
- Le Signor C., Aimé D., Bordat A., Belghazi M., Labas V., Gouzy J., Young N.D., Prospero 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 Phytologist*, sous presse (DOI: 10.1111/nph.14500).

- Maruyama N., Mun L.C., Tatsuhara M., Sawada M., Ishimoto M., Utsumi S. 2006. Multiple vacuolar sorting determinants exist in soybean 11S globulin. *The Plant Cell* 18: 1253-1273.
- McCallum C.M., Comai L., Greene EA., Henikoff S. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123: 439-42.
- Molvig L., Tabe L.M., Eggum B.O., Moore A.E., Craig S., Spencer D., Higgins T.J. 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. USA.* 94: 8393-8.
- Ng W-K.D., Hall T.C. 2008. PvALF and FUS3 activate expression from the phaseolin promoter by different mechanisms. *Plant Molecular Biology* 66: 233-44.
- Nordlee J.A., Taylor S.L., Townsend J.A., Thomas L.A., Bush R.K. 1996. Identification of a Brazil-nut allergen in transgenic soybeans. *N Engl J Med.* 334: 688-92.
- Osborne T.B. 1924. *The Vegetable Proteins.* (London: Longmans, Green).
- Pandurangan S., Diapari M., Yin F., Munholland S., Perry G.E., Chapman B.P., Huang S., Sparvoli F., Bollini R., Crosby WL., 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: 389.
- Paris N., Neuhaus J-M. 2002. BP-80 as a vacuolar sorting receptor. *Plant Mol. Biol.* 50: 903-914.
- Pastorello E.A., Pompei C., Pravettoni V., Brenna O., Farioli L., Trambaioli C., Conti A. 2001. Lipid transfer proteins and 2S albumins as allergens. *Allergy* 56 Suppl 67:45-7.
- Prudent M., Vernoud V., Girodet S., Salon C. 2016. How nitrogen fixation is modulated in response to different water availability levels and during recovery: A structural and functional study at the whole plant level. *Plant and Soil* 399: 1-12
- 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.* 27: 2692-708.
- 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.
- Salon C., Avice J-C., Alain O., Prudent M., and Voisin A-S. (2011). "Plant N fluxes and modulation by nitrogen, heat, and water stresses: a review based on comparison of legumes and non-legume plants," in *Abiotic Stress in Plants - Mechanisms and Adaptations*, eds A. Shanker and B. Venkateswarlu (Rijeka: InTech), 78-118.
- Santi P. (2017) Moins de viande, de sel, de sucre... les recommandations de l'agence sanitaire. *Le Monde*, 24.01.2017. En savoir plus sur http://www.lemonde.fr/sante/article/2017/01/24/moins-de-viande-et-de-charcuteries-moins-de-sucre-les-recommandations-de-l-agence-sanitaire_5068115_1651302.html#aBJedHzvD9lpTV2u.99
- Santos-Mendoza M., Dubreucq B., Baud S., Parcy F., Caboche M., Lepiniec L. 2008. Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *The Plant Journal* 54: 608-620.
- Shimada T., Fuji K., Tamura K., Kondo M., Nishimura M., Hara-Nishimura I. 2003. Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 100: 16095-16100.
- Sieh D., Watanabe M., Devers E.A., Brueckner F., Hoefgen R., Krajinski F. 2013. The arbuscular mycorrhizal symbiosis influences sulfur starvation responses of *Medicago truncatula*. *New Phytol.* 197: 606-16.
- Song B., Shen L., Wei X., Guo B., Tuo Y., Tian F., Han Z., Wang X., Li W and Liu S. 2014. Marker-assisted backcrossing of a null allele of the α -subunit of soybean (*Glycine max*) β -conglycinin

- with a Chinese soybean cultivar (a). The development of improved lines. *Plant Breed.* 133: 638-648.
- Stanton-Geddes J., Paape T., Epstein B., Briskine R., Yoder J., Mudge J., Bharti A.K., Farmer A.D., Zhou P., Denny R., et al. 2013. Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS One* 8: e65688.
- Tabé L.M., Droux M. 2002. Limits to sulfur accumulation in transgenic lupin seeds expressing a foreign sulfur-rich protein. *Plant Physiol.* 128: 1137-48.
- Tan-Wilson A.L., Wilson K. A. 2012. Mobilization of seed protein reserves. *Physiologia Plantarum* 145: 140-153.
- 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.L., Warkentin T.D., Brunel D., Marget P., Le Paslier M-C., 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: 1257-73.
- Thompson R., Burstin J., Gallardo K. 2009. Post-genomics studies of developmental processes in legume seeds. *Plant Physiol.* 151: 1023-9.
- Tsubokura Y., Hajika M., Kanamori H., Xia Z., Watanabe S., Kaga A., Katayose Y., Ishimoto M., Harada K. 2012. The β -conglycinin deficiency in wild soybean is associated with the tail-to-tail inverted repeat of the α -subunit genes. *Plant Mol. Biol.* 78: 301-9.
- Vandecasteele C., Teulat-Merah B., Morère-Le Paven M-C., Leprince O., Ly Vu B., Viau L., Ledroit L., Pelletier S., Payet N., Satour P., Lebras C., Gallardo K., Hugué T., Limami A.M., Prospero J-M, Buitink J. 2011. Quantitative trait loci analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in *Medicago truncatula*. *Plant Cell Environ.* 34: 1473-87.
- Vitale A., Hinz G. 2005. Sorting of proteins to storage vacuoles: how many mechanisms? *Trends in Plant Science* 10: 316-323.
- Wang J., Liu L., Guo Y., Wang Y-H., Zhang L., Jin L-G., Guan R-X., Liu Z-X., Wang L-L., Chang R-Z., Qiu L-J. 2014. A dominant locus, *qBSC-1*, controls β subunit content of seed storage protein in soybean (*Glycine max* (L.) Merri.). *J Integr Agric.* 13: 1854-1864.
- 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.
- Zuber H., Poignavent G., Le Signor C., Aimé D., Vieren E., Tadla C., Belghazi M., Labas V., Santoni A-L., Wipf D., Buitink J., Avice J-C., Salon C., Gallardo K. 2013. Legume adaptation to sulfur deficiency revealed by comparing nutrient allocation and seed traits in *Medicago truncatula*. *Plant Journal* 76: 982-96.