Genetic determinants of seed protein plasticity in response to the environment in Medicago truncatula

Kevin Cartelier, Delphine Aimé, Joseph Ly Vu, Lucie Combes-soia, Valérie Labas, Jean-marie Prosperi, Julia Buitink, Karine Gallardo, Christine Le Signor

HAL Id: hal-03177967
https://hal.inrae.fr/hal-03177967v1

Submitted on 23 Mar 2021

Genetic determinants of seed protein plasticity in response to the environment in *Medicago truncatula*

Kevin Cartelier[1], Delphine Aimé[1], Joseph Ly Vu[2], Lucie Combes-Soia[3], Valérie Labas[3], Jean-Marie Prosperi[4], Julia Buitink[2], Karine Gallardo[1]* and Christine Le Signor[1]*

[1]Agroécologie, AgroSup Dijon, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), Université de Bourgogne, Université Bourgogne Franche-Comté, Dijon, France; [2]Univ Angers, Institut Agro, INRAE, IRHS, SFR QUASAV, F-49000 Angers, France; [3]Physiologie de la Reproduction et des Comportements (PRC) UMR85, INRAE, CNRS, Université de Tours, IFCE, Nouzilly, France; [4]Genetic Improvement and Adaptation of Mediterranean and Tropical Plants (AGAP), INRAE, Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Montpellier SupAgro, Montpellier, France.

*Co-corresponding authors, contributed equally to this work

Running title: Seed protein plasticity in *M. truncatula*

Corresponding author details:

Christine Le Signor, christine.le-signor@inrae.fr

Karine Gallardo, karine.gallardo-guerrero@inrae.fr

Same postal address: UMR Agroécologie, pôle GEAPSI, 17 rue de Sully, 21000 Dijon, France

Summary

Since 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 is increasingly important. To elucidate the genetic determinants of seed protein plasticity, major seed proteins from 200 ecotypes of *M. truncatula* grown in four contrasting 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 (GWAS), enabling the identification of candidate genes determining this plasticity. This list was enriched in genes related to transcription, DNA repair, and signal transduction, many of them being stress-responsive. 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 placing 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 protein composition.

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 plant 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 cereal proteins in both food and feed (Perrot, 1995). The major proteins of legume seeds are the globulins, which account for up to 70-80% of total seed proteins. The globulins are of two different types, 7S or 11S, as determined on the basis of 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. The globulins may have different nutritional and functional properties depending on their amino acid composition and structure (Duranti and Gius, 1997; Crévieu-Gabriel *et al.*, 1999; Rubio *et al.*, 2014). For instance, 7S globulins are poor in the sulfur amino acids Met and Cys compared to 11S globulins (Rubio *et al.,* 2014).

There is a large genetic variability for seed protein content and composition in legumes such as pea and soybean (Vollmann *et al.*, 2000; Bourgeois *et al.*, 2011). Environmental changes also have significant effects on these traits, as shown in pea where reduced processing of vicilins was observed in a year associated with drought and heat stress (Bourgeois *et al.*, 2009), or in soybean where high temperatures have been shown to influence seed protein content (Thomas *et al.*, 2003). The balance between the accumulation of the 7S and 11S globulins strongly depends on sulphate uptake by roots, as shown in soybean, pea and the model legume species *Medicago truncatula* (Gayler and Sykes, 1985; Sexton *et al.*, 1998; Zuber *et al.*, 2013; Henriet *et al.*, 2019). 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. Deciphering the genetic basis of phenotypic plasticity, which is the ability of a genotype to modify its phenotype in response to environmental changes (Bradshaw, 2006), is also of prime importance for plant breeding.

Two main models have been proposed in the literature as underlying genetics of plant plasticity: the allelic-sensitivity model and the gene-regulatory model (Pigliucci, 1996; Via, 1993; Via *et al.*, 1995). In the first model, sensitive alleles specifically expressed in certain environments would determine the value of the phenotype in these environments and would be indirectly responsible for phenotypic plasticity (Via, 1993). The second model assumes that expression of the structural genes that determine the phenotype are monitored by plasticity (or regulatory) genes (Pigliucci, 1996). The plasticity between genotypes from the same species, revealed by inter-individual variations in their responses across environments, is expressed as genotype by environment interactions (GxE) (Des Marais *et al.*, 2013; El-Soda *et al.*, 2014). Finlay and Wilkinson (1963) proposed to model this interaction by a joint regression analysis in which the plasticity of a genotype is recorded as the slope of a linear regression of its phenotypes across environments over an environmental index (*i.e.* average performance of all genotypes in each environment). More recently, a Bayesian Finlay-Wilkinson 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). Using these methods, the genetic determinism of plasticity has been studied 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 extensive genetic variability that exists in *Medicago truncatula* for seed globulin composition (Le Signor *et al.*, 2017). GWAS were conducted on proteotyping data of

seeds from 200 ecotypes grown in four environments to identify genomic regions and genes underlying the plasticity of seed protein composition. Genes related to the interconnected sulphur and aspartate family pathways were identified as candidates for governing this plasticity, suggesting approaches for stabilizing the protein composition of legume seeds in a changing climate. One of these possibilities relates to Met recycling that we demonstrate here to be a key determinant of seed globulin composition and plasticity.

Results

Genetic and environmental variability in seed protein composition in *M. truncatula*

The seed protein profile of 200 *M. truncatula* genotypes (Table S1) grown in four environments (E1 to E4) was analyzed by 1-DE (Figure 1a). Nine protein bands (P1 to P9) were matched across all the genotypes, then quantified and annotated using protein maps of *M. truncatula* seeds (Le Signor *et al.*, 2017) and mass spectrometry data obtained for five genotypes displaying allelic variations (*i.e.* protein bands at different positions) (Table S2). Each band was enriched in isoforms for a specific globulin type (either 7S or 11S): P1, P2 and P7 mainly contained 7S globulins, the remaining six 11S globulins/legumins (Figure 1a and Table S2). The 7S globulins were detected under three major bands corresponding to convicilins (P1), unprocessed vicilins of 47 kDa (P2) and processed forms of vicilins at 30 kDa (P7). The alpha chains of legumins (33 to 45 kDa) were also distinguished from the beta chains at around 20 kDa (Figure 1a). By studying the proportion of nutritionally important amino acids (Cys, Met, tryptophan and lysine, Table S2) in the sequence of the globulin isoforms, we found that 7S globulins are particularly low in sulphur amino acids (0.2-0.8%) compared to the 11S globulins (1-3%), but contain the highest proportion of lysine (up to 5%). Vicilins did not contain tryptophan, while convicilins and legumins contain 0.8-1.42% of this amino acid.

A PCA of the relative abundance of each band in each genotype across the four environments separated well the genotypes according to the relative abundancies of the sulfur-poor (7S) and sulfur-rich (11S) globulins along axis 2, which explains 19.3% of the total variation (Figure 1b). A correlation matrix between the relative abundancies of the different bands (Figure 2a, upper triangular matrix) indicated that the strongest positive correlations were between the legumin bands. The

correlation coefficients were rather low (r~0.2-0.3), notably between the alpha and beta chains, which can be explained by differences in their proteolysis during seed development, legumin alpha chains being more susceptible to C-terminal degradation than beta chains (Higashi *et al.,* 2006). The absence of correlation between the relative abundance of Vic-47 (P2) and Vic-30 (P7) can be similarly attributed to a differential processing of vicilin isoforms during seed development (Bourgeois et al., 2009). Not surprisingly, the most negative correlations were found between the relative abundance of the main legumin (P6, P8) and vicilin (Vic-47, P2) bands. An analysis of variance revealed significant genotype, environmental, as well as genotype by environment interaction (GxE) effects for each band. Details about variations in protein band abundance (mean and standard deviation) between the four environments are given in Figure 1c. The environmental effect for the protein bands accounted for on average 14% of total variance, the genotype effect for 50%, and the GxE effect for 12% (Figure 1d). We also calculated the 7S/11S ratio, which reflects globulin composition. For this 7S/11S ratio, the variance explained by GxE interaction (26%) was higher compared to that of each individual band (2 to 21%), showing a strong impact of GxE interaction on globulin composition.

Phenotypic plasticity of major seed proteins across environments

The plasticity of each protein band and of the 7S/11S globulin ratio was assessed using a Bayesian Finlay-Wilkinson regression. The pattern of genotype performance in the four environments can be visualized by plotting the protein abundancies in each environment against the estimated environment effects (Figure 2b). Taking protein band P1 in five genotypes as an example, a plot of regression values of protein abundance *vs* environment effects (Figure 2c) highlighted a large variation in the slope of the lines (corresponding to 1+β): the HM047 genotype performance was stable across the four environments, HM002 performed best in the E3 and E4 environments, whereas the three other genotypes performed best in the E1 and E2 environments. The lower triangular matrix of the correlogram in Figure 2a, comparing the plasticity indices of the different bands, showed the highest positive correlations (r=0.76) for the major alpha-legumin bands (P3 and P5). The most negative correlation coefficients (r=-0.70) were between P7 (30-Vic) and P5 (alpha-legumin). The correlations between relative protein abundancies and plasticity indices, also illustrated in Figure 2a (diagonal values), were significant and positive for six bands (P1, P2, P5, P6, P7, P8). The highest correlations

(0.5-0.6) were observed for P1 (Conv), P6 (α-Leg) and P8 (β-Leg). Low (0.2-0.3) or non-significant correlation for the other bands suggest the existence of plasticity-specific genetic determinants. A PCA of the plasticity indices of the nine protein bands separated well the individuals according to the genetic structure of the population (Figure 2d) along axis 1, which explains 38.7% of the variation. The two genetic groups were characterized by accessions originating from different geographical origins: west Mediterranean for Group 1 *vs* centre/east Mediterranean 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.

Identifying loci and candidate genes for seed globulin plasticity

Protein band abundancies in the four environments and their plasticity indices were used to decipher the underlying genetic determinism by GWAS. The plasticity index of the 7S/11S globulin ratio did not yield results due to its bimodal distribution within the collection, which did not fit the linear GWAS model. Distribution of the significant SNPs along the eight chromosomes of *M. truncatula* highlighted hotspots associated with variations in protein band abundance in one or more environments as well as hotspots associated with plasticity indices (Figure 3). The distribution 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 out of 61 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). This study highlighted hotspots for both plasticity and protein band abundance in one or several environments as exemplified by a region on chromosome 1 (at ∼ 50Mb) or a region at the top of chromosome 2. In addition, 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).

Next, we compared the lists of genes under GWAS intervals for each environment (E1: 507 genes, E2: 645 genes, E3: 552 genes and E4: 545 genes; ontology enrichment by environment presented 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. A total of 781 SNP were retrieved from plasticity indices 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 which we therefore refer to as plasticity-specific candidate genes (PSG) and 164 genes were also identified from protein abundance (Table S4). 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).

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 1). Among the most represented were transmembrane receptor protein serine/threonine kinase signaling pathway (GO I, 13 genes) and DNA repair (GO VI, 11 genes). The three other most represented GO terms (II, IV, and V) were aspartate 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 placing these genes in their corresponding pathways, specific pathways potentially involved in globulin plasticity were revealed (Figure 4). In particular, 11 PG were related to the recycling of Met. Five encoded *S*-adenosylmethionine-dependent methyltransferases that act in the methyl cycle regenerating homocysteine to yield Met. Four are related to the Met salvage pathway, which synthesizes Met from 5'-methylthioadenosine 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 that converts adenine into AMP, a methylthioribose-1-phosphate isomerase that synthesizes 1,2-dihydro-3-keto-5-methylthiopentene, and two acireductone dioxygenases that synthesizes 2-keto-4-(thiomethyl) butyrate, the keto-acid precursor of Met, from 1,2-dihydro-3-keto-5-methylthiopentene. 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 3 (*HMT3,* Arabidopsis homolog: At3g22740), which converts SMM into Met (Bourgis *et al.*, 1999). The chromosomal position of *HMT3* and *MMT*

perfectly colocalized with GWAS hotspots for both plasticity indices and globulin abundance in several environments (Figure 3). 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 determining both globulin composition and plasticity.

Response of candidate genes for globulin plasticity to abiotic stresses

We studied the regulation of the PG or PSG in seeds in response to abiotic stresses using transcriptomics data previously obtained from *M. truncatula s*eeds collected during early (stage I), mid (stage II), and late (stage III, abscission and maturity) stages of 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 aspartate metabolisms depicted in Figure 4, 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 aspartate metabolisms, 15 were differentially expressed in response to water stress or nutrient deficiencies, especially sulfur deficiency (Figure 4 and Table S6).

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

Using the *M. truncatula* gene expression atlas (Benedito *et al.*, 2008), we showed that 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 for playing a role in globulin synthesis or processing since they were annotated as transcription factors (six genes) or peptidases. Moreover, 12 PG 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 monitor globulin plasticity. Among the subset of genes expressed during storage protein synthesis, occurring from 16 to 24 days after pollination (Benedito *et al.*, 2008), is *HMT3,* a good candidate for regenerating Met within the seed for seed globulin accumulation and plasticity. To investigate the potential link 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 5a). 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 5b). More interestingly, *hmt3* seeds accumulated less 11S globulins than wild-type seeds under both sulfur-sufficient and sulfur-deficient conditions. The higher accumulation of sulfur-poor 7S globulins in *hmt3* seeds compared to wild-type was sufficient to maintain nitrogen content in the sulfur-sufficient condition but not under sulfur deficiency (Figure 5b). 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 5c).

Candidate regulators of seed protein plasticity

Transcription factors 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 transcription factor or chromatin remodelers, revealed that both PG and PSG lists were enriched in regulators (15% and 16% respectively, Table S7). The most highly represented transcription factor families in the PG dataset were *mTERF* (Mitochondrial Transcription Termination Factor) and *AP2-EREBP* (Apetala2-Ethylene-Responsive Element Binding Proteins) (Figure S1). 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 trihelix GT-2 like transcription factor whose expression increased during seed filling up to late maturation, suggesting it could participate in storage activities. Since most of the regulators in Table S9 were not preferentially expressed in seeds (76/82), regulatory events at the level of the mother plant might contribute to seed globulin plasticity. Transcriptomics data under abiotic stresses (Zhang *et al.*, 2014; de Bang *et al.*, 2017) 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, homeobox 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).

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. By studying the 1-DE seed protein profiles of 200 *M. truncatula* ecotypes grown under four environmental conditions and using Bayesian Finlay-Wilkinson regression to quantify the plasticity of globulins, we showed that the *M. truncatula* collection displayed a large genetic variability for globulin plasticity, making it possible to investigate its genetic determinism by GWAS. The results revealed that some regions underlying seed globulin composition in one or several environments were also identified from globulin plasticity indices (Figure 3) and would be relevant to the allelic-sensitivity model proposed by Via (1993). Other regions or genes were only identified from plasticity indices (PSG), supporting the gene-regulatory model in favor of the existence of plasticity-specific determinants (Pigliucci, 1996). 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 globulin plasticity are discussed hereafter.

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 aspartate amino acid family pathways (Table 1), 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 4). 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 were regulated by abiotic stresses in developing seeds, shoots or roots (15/17, Figure 4). There have been significant advances towards understanding the role of the sulfur and aspartate family pathways in plant stress responses (Khan *et al.*, 2008; Galili, 2011; Batista-Silva *et al.*, 2019). Our data suggest that these pathways may also play a prominent role in the plastic response of seed proteins to environmental

changes. One PSG encodes a key enzyme of the aspartate family pathway, aspartate-semialdehyde dehydrogenase, which produces aspartate semialdehyde and acts at the branch point between lysine and Met metabolisms (Ravanel *et al.*, 1998). The identification of another PSG as corresponding to LL-diaminopimelate aminotransferase, which catalyzes a subsequent step of lysine biosynthesis, suggests that 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 Met recycling rather than on *de novo* Met 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 produced during the synthesis of ethylene and polyamines, two hormones well known for their role in the plant response to environmental stimuli (Gill and Tuteja, 2010; Booker and DeLong, 2015). Genes of the Met salvage pathway were up-regulated in response to water stress in roots (methylthioribose-1-phosphate isomerase and acireductone dioxygenase1), shoots (methylthioribose-1-phosphate isomerase) and late developing seeds (methylthioribose-1-phosphate isomerase) (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 *S*-adenosylmethionine (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 *S*-adenosylmethionine 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 4 and Table 7) and thus is 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. Characterization of seed protein composition of a *Tnt1* insertion mutant of *M. truncatula* for *HMT3* (Figure 5) showed that HMT3 is a key contributor towards 11S globulin

accumulation in legume seeds, the 7S/11S ratio being higher in *hmt3* seeds. These observations complement well the data from Cohen *et al.* (2017) who showed an increase of protein-bound Met in *Arabidopsis* mutant seeds overexpressing *HMT3*. The results also indicate that *HMT3* acts as a plasticity gene by contributing to stabilizing the 7S to 11S globulin balance under contrasting sulfur supply (Figure 5c). 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. Targeting the SMM cycle to decrease the 7S/11S globulin ratio might help to improve and stabilize the amino acid balance of legume seeds, 11S globulins being rich in sulfur amino acids compared to vicilin-type 7S globulins, which did not contain tryptophan, the second most limiting amino acid in legume seeds (Table S2).

Transcription factors and DNA repair proteins are likely to play a role in globulin plasticity

The enrichment of the PG dataset in genes for transmembrane receptor protein serine/threonine kinases (Table 1) and transcription factors (Table S7) suggested the importance of signaling and regulatory processes in the plastic response of globulins to environmental variations. Of the six transcription factors from the PG list that were preferentially expressed in seeds, one (Chr1g0190851) merits attention since it is expressed at seed filling stages. Its closest Arabidopsis homolog, AtGT-2 Like (AT5G28300), as other GT factors (trihelix DNA-binding factor) 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 influence the plasticity of seed proteins. DNA repair might also contribute to seed globulin plasticity owing to the overrepresentation of this biological process in the PG dataset (Table 1). 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). 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 good candidates for contributing to 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 research especially in relation to methionine recycling pathways and their contribution to the amino acid balance of legume seeds.

Experimental Procedures

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

A population of 200 ecotypes from the *M. truncatula* HapMap project (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 (IRHS) 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 m³ box pallets filled with a mixture of soil and sand, 5 plants per 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 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 seed lots were analyzed).

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, 2μL of protein extracts mixed with 8 μl of loading buffer (5X = TrisHCl 0.5M, SDS 4%, glycerol 20% and 1 mg bromophenol blue) were separated by one-dimensional electrophoresis (1-DE) using the XCell4 *SureLock*™ Midi-Cell system (Thermo Fisher Scientific, Illkirch-Graffenstaden, France) 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). Gels were stained with Coomassie Brilliant Blue R-250 (Mathesius *et al.*, 2001) and scanned using the Odyssey infrared imaging system (LI-COR Biosciences, Bad Homburg, Germany). Protein band detection and quantification were performed using the CLIQS software (http://www.totallab.com). When there was a shift in protein bands between genotypes, the bands were matched to obtain quantitative values in the entire core-collection. The molecular weight of each protein band was calculated using a low-range protein ladder (Bio-Rad, Marnes-La-Coquette, France). 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, and car).

Identification of mature seed proteins

Protein bands were annotated in comparison with previously established protein maps of mature *M. truncatula* seeds (Le Signor *et al.*, 2017). Eight major bands in the seed protein profile of the reference line Jemalong A17 were analyzed by mass spectrometry, which confirmed the correspondence with previously established protein maps (Table S2). To confirm the correspondence between the major protein bands displaying a position shift across the collection, four additional genotypes (HM079, HM083, HM160, HM187, Table S1) were selected for band excision and protein identification (Table S2). Protein band digestion and peptide extraction were performed according to Labas *et al.* (2015). The peptides were then 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 (Table S2). The proportion of Cys, Met, tryptophan and lysine in the protein sequence of each globulin isoform enriched in the 1-DE protein bands was calculated after removal of the peptide signal determined using PSORT Prediction (http://psort1.hgc.jp/form.html).

Calculation of plasticity indices

For each protein band, a plasticity index was calculated using the bayesian Finlay-Wilkinson regression method implemented in the R package (R Core Team, 2014; Lian and de los Campos, 2016). Bayesian Finlay-Wilkinson 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 kth replicate of the ith variety observed in the jth environment, g_i is the main effect of ith variety, h_j is the main effect of the

j^{th} environment, and ε_{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 (β). For each genotype, β plasticity indices were calculated from the relative abundancies of each of the nine protein bands in the four environments.

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 by 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.

Procedure for GWAS and candidate 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 rMVP package (Yin *et al.*, 2020), 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 quantile-quantile (Q-Q) plots using the rMVP package.

Manhattan, and Q-Q plots obtained for globulin plasticity are provided in Figure S2. 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. To select candidate genes, a first list with the most significant SNPs was made (*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 (r² 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 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/.

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* genotype 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 Tnt1 insertion in *HMT3* was detected in plants from NF16472. Plants from NF7660 carried an insertion in exon 6 of *HMT3* (NF7660_high5). The Tnt1 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 watered with a nitrate and sulfur-containing solution, as described in Zuber at al. (2013). At flowering, half of the plants (one block) were subjected to sulfur deficiency

until maturity as described in Zuber *et al.* (2013). Upon 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 Tukey's test.

Data availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol and Moreno, 2020) partner repository (http://www.ebi.ac.uk/pride) with the dataset identifier PXD022244 and 10.6019/PXD022244. All the other data supporting the findings of this study are available within the paper and its supplementary files.

Acknowledgements

Author contributions

KG and CLS conceived the research plan and supervised the study, with the contribution of JB who coordinated the ANR-REGULEG project. KC and DA performed the 1-DE gels. KC performed

GWAS with the contribution of CLS. JB and JLV produced the CC200 seeds in the environments E1 and E2. J-MP produced the CC200 seeds in the environments E3 and E4. VL and LC performed the mass spectrometry analyses of seed protein bands. KC, KG and CLS analyzed all the data and wrote the manuscript.

Conflicts of interest

The authors declare that they have no conflicts of interest for this work.

Supporting Material legends

Figure S1. The most highly represented transcription factor families in the PG dataset.

Figure S2. Manhattan plots, Q-Q plots, number of SNP and genes identified by GWAS from plasticity indices.

Table S1. List of the 200 *M. truncatula* ecotypes used in this study.

Table S2. Proteins identified by LC-MS/MS in the 1-DE bands from *M. truncatula* seeds.

Table S3. Enriched GO terms in the list of genes identified by GWAS for seed protein composition in each of the four environments.

Table S4. List of candidate genes for globulin plasticity and their abiotic stress regulation.

Table S5. Plasticity-related genes (PG) bearing SNPs in their coding sequence that lead to an amino acid change or a premature stop codon.

Table S6. PG associated with sulfur, Met and aspartate family pathways.

Table S7. Enrichment analysis for genes preferentially expressed in seeds or regulated in response to abiotic stresses in the different post-GWAS datasets.

Table S8. PG preferentially expressed in seeds and their regulation under various stress conditions.

Table S9. PG annotated as regulators and their response to stress conditions.

References

Alexa A and Rahnenfuhrer J (2016) topGO: Enrichment Analysis for Gene Ontology. R package version 2.24.0. *Bioconductor*. Available at: http://bioconductor.org/packages/topGO/ [Accessed October 1, 2016].

Bang, T.C. de, Lay, K.S., Scheible, W.-R. and Takahashi, H. (2017) Small peptide signaling pathways modulating macronutrient utilization in plants. *Current Opinion in Plant Biology*, **39**, 31–39.

Batista-Silva, W., Heinemann, B., Rugen, N., Nunes-Nesi, A., Araújo, W.L., Braun, H.-P. and Hildebrandt, T.M. (2019) The role of amino acid metabolism during abiotic stress release. *Plant, Cell & Environment*, **42**, 1630–1644.

Benedito, V.A., Torres-Jerez, I., Murray, J.D., et al. (2008) A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal*, **55**, 504–513.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 289–300.

Booker, M.A. and DeLong, A. (2015) Producing the Ethylene Signal: Regulation and Diversification of Ethylene Biosynthetic Enzymes. *Plant Physiology*, **169**, 42–50.

Bourgeois, M., Jacquin, F., Cassecuelle, F., et al. (2011) A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics*, **11**, 1581–1594.

Bourgeois, M., Jacquin, F., Savois, V., Sommerer, N., Labas, V., Henry, C. and Burstin, J. (2009) Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics*, **9**, 254–271.

Bourgis, F., Roje, S., Nuccio, M.L., et al. (1999) S-Methylmethionine Plays a Major Role in Phloem Sulfur Transport and Is Synthesized by a Novel Type of Methyltransferase. *The Plant Cell*, **11**, 1485–1497.

Bradshaw, A.D. (2006) Unravelling phenotypic plasticity ? why should we bother? *New Phytol*, **170**, 644–648.

Browning, S.R. and Browning, B.L. (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, **81**, 1084–1097.

Cheng, X., Krom, N., Zhang, S., Mysore, K.S., Udvardi, M. and Wen, J. (2017) Enabling Reverse Genetics in *Medicago truncatula* Using High-Throughput Sequencing for Tnt1 Flanking Sequence Recovery. *Methods in Molecular Biology (Clifton, N.J.)*, **1610**, 25–37.

Cohen, H., Salmon, A., Tietel, Z., Hacham, Y. and 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.

Crévieu-Gabriel, I., Gomez, J., Caffin, J.-P. and Carré, B. (1999) Comparison of pig and chicken pepsins for protein hydrolysis. *Reprod. Nutr. Dev.*, **39**, 443–454.

Des Marais, D.L., Hernandez, K.M. and 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.

Duranti, M. and Gius, C. (1997) Legume seeds: protein content and nutritional value. *Field Crops Research*, **53**, 31–45.

El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M. and Aarts, M.G.M. (2014) Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends in Plant Science*, **19**, 390–398.

Finlay, K. and Wilkinson, G. (1963) The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.*, **14**, 742–754.

Galili, G. (2011) The aspartate-family pathway of plants. *Plant Signaling & Behavior*, **6**, 192–195.

Gallardo, K., Firnhaber, C., Zuber, H., Héricher, D., Belghazi, M., Henry, C., Küster, H. and 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**, 2165–2179.

Gao, M.-J., Lydiate, D.J., Li, X., Lui, H., Gjetvaj, B., Hegedus, D.D. and Rozwadowski, K. (2009) Repression of Seed Maturation Genes by a Trihelix Transcriptional Repressor in Arabidopsis Seedlings. *The Plant Cell*, **21**, 54–71.

Gayler, K.R. and Sykes, G.E. (1985) Effects of Nutritional Stress on the Storage Proteins of Soybeans. *Plant Physiol.*, **78**, 582–585.

Gill, S.S. and Tuteja, N. (2010) Polyamines and abiotic stress tolerance in plants. *Plant Signaling & Behavior*, **5**, 26–33.

Graham, P.H. and Vance, C.P. (2003) Legumes: Importance and Constraints to Greater Use. *Plant Physiol.*, **131**, 872–877.

Henriet, C., Aimé, D., Térézol, M., et al. (2019) Water stress combined with sulfur deficiency in pea affects yield components but mitigates the effect of deficiency on seed globulin composition S. Kopriva, ed. *Journal of Experimental Botany*, **70**, 4287–4304.

Khan, N.A., Singh, S. and Umar, S. eds. (2008) *Sulfur Assimilation and Abiotic Stress in Plants*, Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: http://link.springer.com/10.1007/978-3-540-76326-0 [Accessed September 25, 2020].

Kusmec, A., Srinivasan, S., Nettleton, D. and Schnable, P.S. (2017) Distinct genetic architectures for phenotype means and plasticities in Zea mays. *Nature Plants*, **3**, 715–723.

Labas, V., Grasseau, I., Cahier, K., et al. (2015) Qualitative and quantitative peptidomic and proteomic approaches to phenotyping chicken semen. *Journal of Proteomics*, **112**, 313–335.

Lacaze, X., Hayes, P.M. and Korol, A. (2009) Genetics of phenotypic plasticity: QTL analysis in barley, Hordeum vulgare. *Heredity*, **102**, 163–173.

Le Signor, C., Aimé, D., Bordat, A., et al. (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.

Li, C., Wu, X., Li, Yongxiang, Shi, Y., Song, Y., Zhang, D., Li, Yu and Wang, T. (2019) Genetic architecture of phenotypic means and plasticities of kernel size and weight in maize. *Theor. Appl. Genet.*, **132**, 3309–3320.

Lian, L. and Campos, G. de los (2016) FW: An R Package for Finlay–Wilkinson Regression that Incorporates Genomic/Pedigree Information and Covariance Structures Between Environments. *G3*, **6**, 589–597.

Mathesius, U., Keijzers, G., Natera, S.H., Weinman, J.J., Djordjevic, M.A. and 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.

Osborne, T.B. (1909) *The vegetable proteins, by Thomas B. Osborne.*, London,: Longmans, Green,. Available at: http://www.biodiversitylibrary.org/bibliography/18912.

Parreira, J.R., Balestrazzi, A., Fevereiro, P. and Araújo, S.D.S. (2018) Maintaining Genome Integrity during Seed Development in *Phaseolus vulgaris* L.: Evidence from a Transcriptomic Profiling Study. *Genes*, **9**, 463.

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

Perez-Riverol, Y. and Moreno, P. (2020) Scalable Data Analysis in Proteomics and Metabolomics Using BioContainers and Workflows Engines. *PROTEOMICS*, **20**, 1900147.

Perrot, C. (1995) Les protéines de pois : de leur fonction dans la graine à leur utilisation en alimentation animale. *Productions animales*, **8**, 151–164.

Pigliucci, M. (1996) How organisms respond to environmental changes: from phenotypes to molecules (and vice versa). *Trends in Ecology & Evolution*, **11**, 168–173.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Ravanel, S., Gakière, B., Job, D. and Douce, R. (1998) The specific features of methionine biosynthesis and metabolism in plants. *PNAS*, **95**, 7805–7812.

Righetti, K., Vu, J.L., Pelletier, S., et al. (2015) Inference of Longevity-Related Genes from a Robust Coexpression Network of Seed Maturation Identifies Regulators Linking Seed Storability to Biotic Defense-Related Pathways. *Plant Cell*, tpc.15.00632.

Ronfort, J., Bataillon, T., Santoni, S., Delalande, M., David, J.L. and Prosperi, J.-M. (2006) Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biology*, **6**, 28.

Rubio, L.A., Pérez, A., Ruiz, R., Guzmán, M.Á., Aranda-Olmedo, I. and 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.

Rzewuski, G., Cornell, K.A., Rooney, L., Bürstenbinder, K., Wirtz, M., Hell, R. and Sauter, M. (2007) OsMTN encodes a 5′-methylthioadenosine nucleosidase that is up-regulated during submergence-induced ethylene synthesis in rice (Oryza sativa L.). *Journal of Experimental Botany*, **58**, 1505–1514.

Sauter, M., Moffatt, B., Saechao, M.C., Hell, R. and Wirtz, M. (2013) Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochemical Journal*, **451**, 145–154.

Schroeder, H.E. (1982) Quantitative studies on the cotyledonary proteins in the genusPisum. *Journal of the Science of Food and Agriculture*, **33**, 623–633.

Sexton, P.J., Paek, N.C. and 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 Research*, **59**, 1–8.

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

Tan, Q., Zhang, L., Grant, J., Cooper, P. and Tegeder, M. (2010) Increased Phloem Transport of S-Methylmethionine Positively Affects Sulfur and Nitrogen Metabolism and Seed Development in Pea Plants. *Plant Physiology*, **154**, 1886–1896.

Thomas, J.M.G., Boote, K.J., Allen, L.H., Gallo-Meagher, M. and Davis, J.M. (2003) Elevated Temperature and Carbon Dioxide Effects on Soybean Seed Composition and Transcript Abundance. *Crop Science*, **43**, 1548–1557.

Via, S. (1993) Adaptive phenotypic plasticity: target or by-product of selection in a variable environment? *American Naturalist*, **142**, 352–365.

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S.M., Schlichting, C.D. and Van Tienderen, P.H. (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, **10**, 212–217.

Vollmann, J., Fritz, C.N., Wagentristl, H. and Ruckenbauer, P. (2000) Environmental and genetic variation of soybean seed protein content under Central European growing conditions. *Journal of the Science of Food and Agriculture*, **80**, 1300–1306.

Wang, H., Lin, J., Li, X.G. and 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.

Waterworth, W.M., Bray, C.M. and West, C.E. (2019) Seeds and the Art of Genome Maintenance. *Front. Plant Sci.*, **10**, 706.

Xi, J., Qiu, Y., Du, L. and Poovaiah, B.W. (2012) Plant-specific trihelix transcription factor AtGT2L interacts with calcium/calmodulin and responds to cold and salt stresses. *Plant Science*, **185–186**, 274–280.

Yang, S.F. and Hoffman, N.E. (1984) Ethylene Biosynthesis and its Regulation in Higher Plants. *Annu. Rev. Plant. Physiol.*, **35**, 155–189.

Yin, L., Zhang, H., Tang, Z., et al. (2020) rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study. *bioRxiv*, 2020.08.20.258491.

Zhang, J.-Y., Cruz De Carvalho, M.H., Torres-Jerez, I., et al. (2014) Global reprogramming of transcription and metabolism in *M edicago truncatula* during progressive drought and after rewatering: Drought-stressed *Medicago truncatula*. *Plant Cell Environ*, **37**, 2553–2576.

Zhou, X. and Stephens, M. (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods*, **11**, 407–409.

Zuber, H., Poignavent, G., Le Signor, C., et al. (2013) Legume adaptation to sulfur deficiency revealed by comparing nutrient allocation and seed traits in *Medicago truncatula*. *Plant J*, **76**, 982–996.

Figure and table legends

Figure 1. 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 three genotypes (G1: HM107, G2: HM194, G3: HM106) in two environments (E1 and E2). The P1, P2, and P3 bands mainly contained 7S globulins (in green): convicilins (Conv), 47 kDa vicilins (47-Vic) and 30 kDa vicilins (30-Vic, as described in Le Signor *et al.,* 2017). The other bands, in red, mainly contained 11S globulins: alpha chains (P3 to P6, α-Leg) and basic chains (P8 and P9, β-Leg). **(b)** Principal component analysis based on the relative protein abundance of the nine protein bands of the 200 genotypes (7S globulins, green; 11S globulins, red). 75% confidence ellipses per environment are shown in blue (E1), grey (E2), purple (E3) and orange (E4). **(c)** Mean values and standard deviation of relative protein abundance per environment (Tukey's multiple comparison test is indicated using letters). **(d)** Relative percentage of total variance explained by genotype (G), environment (E) and genotype by environment interaction (GxE). The variances were calculated using a two-way random effects ANOVA (lmer function in R) from the abundance of each of the nine protein bands and for the 7S/11S globulin ratio.

Figure 2. Plasticity of relative globulin abundance across environments. (a) Correlation matrix between plasticity indices (ß) of protein bands (lower triangular), between protein band abundance (upper triangular) and between protein band abundance and plasticity indices (diagonal) (ns, non-significant correlations, $P>0.05$). **(b)** Schematic representation of the ß meaning (modified from Finlay & Wilkinson, 1963): the dashed line corresponds to the behavior of an average genotype (ß+1 = 1), the blue line in the left plot reflects the behavior of a genotype adapted to favorable environments, and the blue line in the right plot reflects the behavior of a genotype specifically adapted to unfavorable environments. The slope of the regression line corresponds to 1+ß**. (c)** Finlay-Wilkinson regression for the relative content of the P1 protein band displayed for five genotypes. The horizontal axis shows the estimated environmental effect and the vertical axis shows the relative content of P1 for each genotype. **(d)** Principal component analysis showing distributions of the 200 genotypes according to the ß indices of the nine protein bands (7S globulins, green; 11S globulins, red). The genetic group (G) to which the genotypes belong is indicated (G1 in red, G2 in blue).

Figure 3. Position of GWAS hotspots on the v5 physical map (in Mbp) of *Medicago truncatula*. The hotspots identified from the relative abundance of globulins in each environment, E1 to E4, are shown with red (11S globulins) or green (7S globulins) dots. For each dot, there were at least 20 SNPs in a 500 kbp window. The hotspots identified from globulin plasticity indices are shown using triangles within the chromosomes (light colored triangle: ≥5 SNPs in a 500 kbp window; dark colored triangle: ≥10 SNPs in a 500 kbp window). The plain lines encircle genomic regions only identified from plasticity indices and dashed lines encircle the genomic regions identified from both plasticity indices and protein abundance (minimum two environments). Genes from GO enrichment analysis related to aspartate, sulfur and methionine metabolisms (Table 1) that colocalized with hotspots for plasticity indices are indicated. ADI1 and ADI2: acireductone dioxygenases 1 and 2, ASD: aspartate-semialdehyde dehydrogenase, DAP-AT: LL-diaminopimelate aminotransferase, HMT3: homocysteine *S*-methyltransferase 3, MMT: methionine *S*-methyltransferase, SHMT: glycine hydroxymethyltransferase, thiE: thiamine phosphate synthase.

Figure 4. Plasticity-linked genes related to the sulfur and aspartate family pathways. The plasticity-linked genes identified by GWAS are shown in red. Stars indicate that the gene was identified from plasticity indices only (PSG list). Eleven are in pathways regenerating Met *via* three pathways: (1) from 5'-methylthioadenosine (MTA) produced in the course of ethylene synthesis (Met salvage pathway), (2) from *S*-Adenosyl-L-homocysteine (AdoHcy) produced during *S*-adenosylmethionine-dependent methylation reactions (Methyl cycle), (3) from S-methylmethionine (SMM) produced from methionine and *S*-adenosylmethionine (SMM cycle). Two genes connected to sulfur metabolism are involved in the synthesis of sulfoquinovosyldiacylglycerol (SQDG) or thiamine. Dashed arrows represent multiple reactions. Six genes are related to the metabolism of non-sulfur amino acids, particularly aspartate. The table indicates whether the genes were preferentially expressed in seeds (PE) and/or regulated by stresses in seeds (left part), shoots and roots (right part). Red ticks in the left part indicate a preferential expression in seeds (seed PE) or a regulation in response to abiotic stresses occurring during seed development at low temperature (14°C), high temperature (26°C) or under water stress (WS). Red ticks in the right part indicate a regulation in shoots and/or roots in response to WS or nutrient deficiencies (K, potassium; N, nitrogen; P, phosphorus; S, sulphur). Detailed information on the regulation of gene expression by stresses is

available in Table S4. A dash indicates that the data were not available for that gene. **Metabolites.** ACC: 1-aminocyclopropane-1-carboxylate; AdoHcy: *S*-Adenosyl-L-homocysteine, 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, Gly: glycine, HET-P: 4-methyl-5-β-hydroxyethylthiazole phosphate, HMP-PP: 4-amino-2-methyl-5-hydroxymethylpyrimidine diphosphate, KMTB: 2-oxo-4-methylthiobutyrate, Lys: lysine, Met: Methionine, MTA: 5'-methylthioadenosine, MTR: 5'-methylthioribose, MTR-P: 5'-methylthioribose-1-phosphate, NAD+: Nicotinamide adénine dinucléotide, NMN: Nicotinamide mononucleotide, OAS: *O*-acetylserine, Ser: Serine, SMM: *S*-methylmethionine, SQDG: Sulfoquinovosyldiacylglycerol, TMP: Thiamine monophosphate, UDP-glucose: Uridine diphosphate-glucose. **Enzymes.** ADI1 and ADI2: acireductone dioxygenases 1 and 2, APRT: adenine phosphoribosyltransferase, ASD: aspartate-semialdehyde dehydrogenase, CPS: carbamoyl-phosphate synthase (glutamine-hydrolysing), DAP-AT: LL-diaminopimelate aminotransferase, Guanine-7-MTase: guanine-7-methyltransferase, HMT3: homocysteine *S*-methyltransferase 3, hTrm4 MTase: tRNA (cytosine34-C5)-methyltransferase, MMT: methionine *S*-methyltransferase, MTase: methyltransferase, MRI: methylthioribose-1-phosphate isomerase, NMNAT: NMN adenylyltransferase, PARP: poly(ADP-ribose) polymerase, PRMT: protein arginine methyltransferase, SHMT: glycine hydroxymethyltransferase, SQD2: sulfoquinovosyltransferase 2, thiE: thiamine phosphate synthase.

Figure 5. 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 sulfur-sufficient (+S) and sulfur-deficient (-S) conditions (7S globulins, green; 11S globulins, red). **(b)** This table shows the mean relative abundance (± standard deviation) of each protein band, of the 7S to 11S globulin ratio and seed nitrogen content (N%) in each condition. Different letters indicate significant variations (Tukey test, N = 8). **(c)** GxE interaction significance (*, P-value < 0.05; ***, P-value < 0.001) is shown along with differences in the plasticity of the 7S to 11S globulin ratio between the wild-type and *hmt3* mutant. ns, non-significant differences.

Table 1. GO terms enriched in the list of genes (556) identified by GWAS for globulin plasticity. 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.

Table 1

GO group	GO.ID	Term	Significant/ annotated	Expected	Classic
I	GO:0007178	Transmembrane receptor protein serine/threonine kinase signaling pathway	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
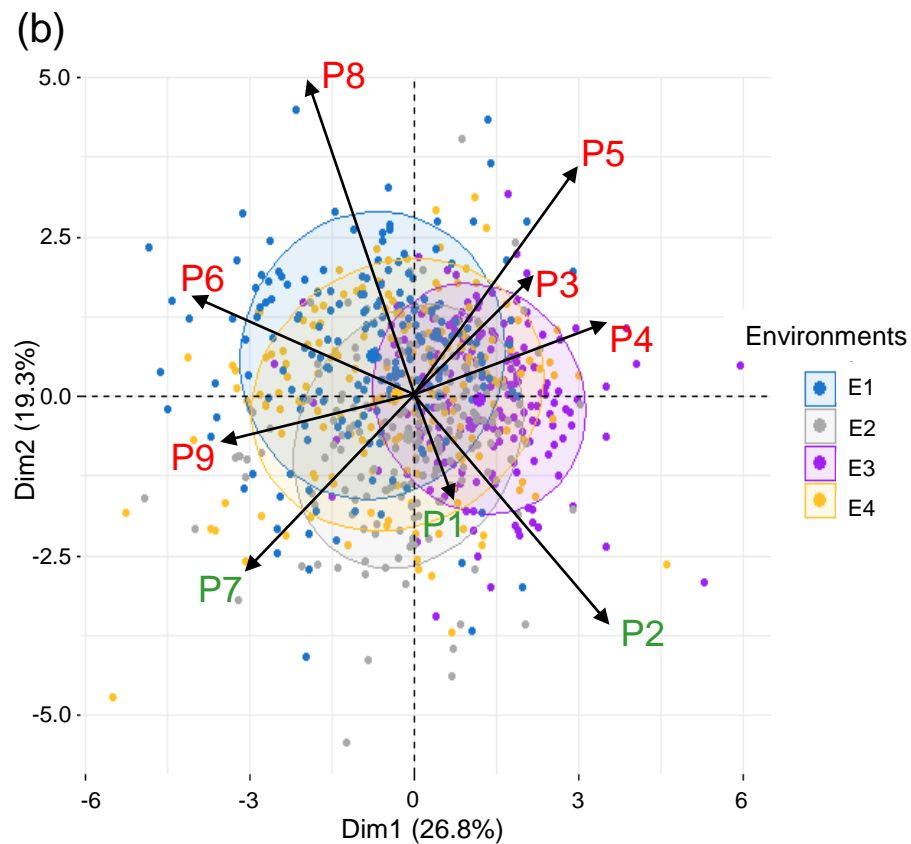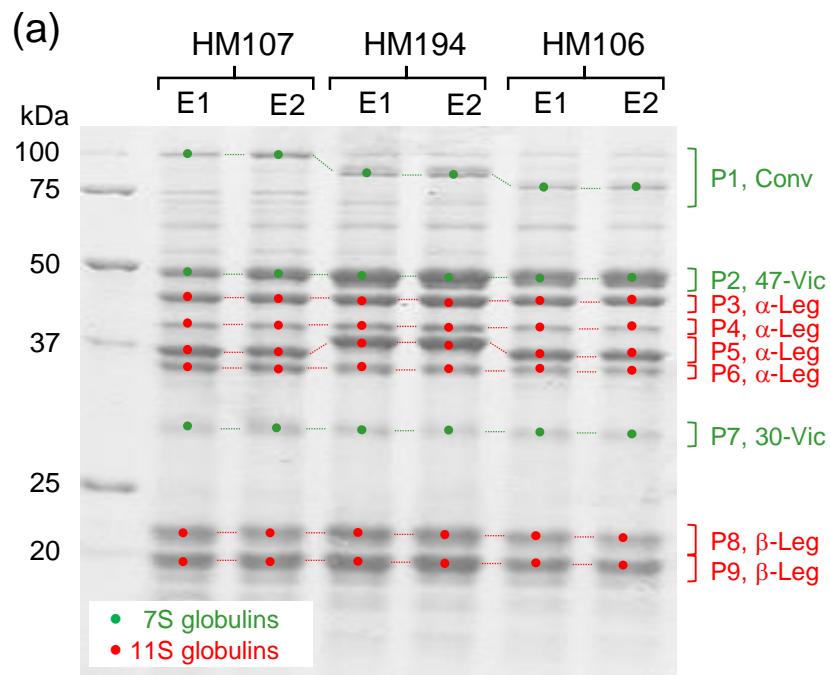
(a)

HM107 HM194 HM106

E1 E2 E1 E2 E1 E2

kDa

100
75
50
37
25
20

P1, Conv
P2, 47-Vic
P3, α-Leg
P4, α-Leg
P5, α-Leg
P6, α-Leg
P7, 30-Vic
P8, β-Leg
P9, β-Leg

● 7S globulins
● 11S globulins

(b)

(c)

Protein band	E1		E2		E3		E4	
P1	7.24 ± 2.6	b	7.9 ± 2.87	c	7.03 ± 2.27	b	6.63 ± 1.97	a
P2	16.7 ± 2.49	a	19.6 ± 2.57	c	20.1 ± 2.43	d	17.9 ± 2.6	b
P3	10.7 ± 1.4	b	10.6 ± 1.18	b	10.8 ± 1.29	b	10.1 ± 1.53	a
P4	4.13 ± 1.16	b	3.79 ± 0.88	a	5.78 ± 0.96	d	5.21 ± 1.12	c
P5	13.9 ± 1.42	b	13.4 ± 1.43	a	13.9 ± 1.14	b	13.2 ± 1.41	a
P6	7.16 ± 1.34	c	6.35 ± 1.13	b	6.06 ± 1.05	a	7.18 ± 1.3	c
P7	4.29 ± 1.07	bc	4.42 ± 1.05	c	3.51 ± 0.86	a	4.15 ± 1.09	b
P8	12.6 ± 1.79	c	11.2 ± 1.5	a	11.2 ± 1.39	a	11.5 ± 1.48	b
P9	16.6 ± 1.65	c	16.4 ± 1.62	bc	15.7 ± 1.51	a	16.3 ± 1.87	b
7S/11S	0.43 ± 0.07	a	0.52 ± 0.08	c	0.48 ± 0.07	b	0.45 ± 0.07	a

(d)

Effect	P1	P2	P3	P4	P5	P6	P7	P8	P9	7S/11S
G	71%	47%	54%	27%	64%	49%	52%	49%	38%	47%
E	4%	27%	4%	41%	4%	17%	13%	14%	4%	17%
GxE	13%	20%	2%	7%	18%	8%	19%	6%	21%	26%
Residual	13%	7%	39%	26%	15%	26%	16%	31%	38%	11%

(a) Relative protein band abundance

	P1	P2	P3	P4	P5	P6	P7	P8	P9	7S/11S
P1	0.59	-0.11	-0.17	-0.17	-0.16	-0.25	-0.14	-0.16	-0.29	0.51
P2	-0.20	0.34	ns	0.19	ns	-0.50	-0.12	-0.59	-0.20	0.56
P3	-0.60	ns	ns	ns	0.23	ns	-0.31	ns	-0.18	-0.13
P4	0.29	ns	-0.30	ns	0.30	-0.24	-0.35	-0.17	-0.35	-0.11
P5	-0.60	ns	0.76	-0.40	0.24	-0.31	-0.31	0.24	-0.33	-0.30
P6	ns	ns	0.20	ns	ns	0.59	0.19	0.23	0.25	-0.44
P7	0.31	0.17	-0.60	0.45	-0.70	ns	0.31	-0.18	0.20	0.07
P8	-0.50	0.31	0.28	-0.10	0.36	0.19	-0.30	0.46	Ns	-0.56
P9	ns	0.29	-0.50	0.34	-0.40	-0.20	0.27	ns	ns	-0.29
7S/11S	-0.40	-0.20	0.27	-0.40	-0.20	0.27	-0.40	-0.20	0.27	0.21

β

(b)

Genotype adapted to high-ranking environments

Genotype performance vs Environment effect — slope 1+β

Genotype adapted to low-ranking environments

Genotype performance vs Environment effect

(c) P1

HM047, HM002, HM312, HM176, HM040, Slope = 1

Genotype performance vs Environment effect

E4, E3, E1, E2

(d)

Dim2 (17.3%) vs Dim1 (38.7%)

P2, P8, P9, P4, P7, P5, P6, P3, P1

Groups: G1, G2

Globulins: 11S 7S

PLASTICITY INDEX ▲ ▲

ABUNDANCE (E1 to E4) ● ●

Seed preferential expression (PE) and gene regulation in response to stresses

Gene	PE	Seeds			Shoots and/or roots				
		14C	26C	WS	WS	K	N	P	S
ADI1*	X	X	✓	X	✓	X	X	X	✓
ADI2*	X	X	✓	X	X	X	X	✓	✓
APRT*	-	X	X	X	-	X	X	✓	✓
ASD*	X	✓	X	X	X	X	X	✓	✓
CPS*	X	X	X	✓	X	X	X	X	X
DAP-AT*	-	X	✓	X	-	X	✓	✓	✓
Guanine-7-MTase	X	X	X	X	X	X	X	X	X
HMT3	✓	X	X	✓	✓	X	X	X	✓
hTrm4 MTase	X	-	-	-	X	X	✓	✓	✓
MMT	X	X	X	X	X	X	X	X	✓
MRI	X	✓	✓	✓	✓	X	✓	✓	X
MTase#1*	X	X	✓	X	X	X	X	X	X
MTase#2*	-	✓	X	X	-	X	✓	X	X
NMNAT*	-	-	-	-	-	X	X	X	X
PARP*	-	X	✓	✓	-	X	X	X	X
PRMT	✓	X	X	X	✓	X	X	✓	✓
SHMT*	X	✓	✓	✓	X	X	✓	X	✓
SQD2*	X	X	✓	X	X	✓	✓	✓	X
thiE*	X	X	X	X	✓	X	✓	X	X

Pathway diagram labels:

S → Sulfate → Sulfite → Sulfide

UDP-glucose → UDP-sulfoquinovose → SQD2* → SQDG
thiE*
HMP-PP, Gly, NAD+, HET-P → Thiamine ← TMP

OAS ← Ser ← N → Asp → NMN
SHMT*
Gly, Glu ← Gln, CPS*
Arg ← Carbamoyl-P
Cys, GSH, NAD+
NMNAT* → NAD+ → PARP* → Nicotinamide
Asp-P → ASD* → AspSA → Cystathionine → Homocysteine

Adenosine
SMM Cycle: HMT3, MMT, SMM, Met
KMTB ← ADI1* / ADI2* → DHKMP
MRI → MTR-P
Methionine salvage pathway

Methyl cycle
AdoHcy
S-adenosylmethionine
Polyamines → MTA → MTR → adenine → APRT* → AMP
Nicotianamine, ACC → Ethylene
PRMT → Methylation reactions
MTase#1* - MTase#2*
Guanine 7 MTase
hTrm4 MTase

2,3-dihydropicolinate → 2,3,4,5-tetrahydrodipicolinate → DAP-AT* → LL-2,6-diaminopimelate → Meso-2,6-diaminopimelate → Lys

(a)

WT   hmt3

kda   +S   -S   +S   -S

250
150
100   ← P1
75
50   ← P2
         ← P3
37   ← P4
         ← P5
         ← P6
         ← P7
25
20   ← P8
         ← P9
15

(b)

Protein band	Sulfur sufficiency (+S)		Sulfur deficiency (-S)		
	Wt	hmt3	Wt	hmt3	
P1	9.42 ± 1.48 a	11.2 ± 0.92 b ↗	14.8 ± 0.91 c	14.75 ± 1.36 c	ns
P2	20.2 ± 0.92 a	26.9 ± 1.33 b ↗	35.5 ± 1.65 c	44.17 ± 1.34 d	↗
P3	10.1 ± 0.33 d	9.54 ± 0.41 c ↘	8.39 ± 0.29 b	7.18 ± 0.42 a	↘
P4	5.89 ± 0.41 d	3.89 ± 0.65 c ↘	1.55 ± 0.28 b	0.89 ± 0.14 a	↘
P5	12.4 ± 0.58 d	10.7 ± 0.60 c ↘	6.6 ± 0.16 b	4.22 ± 0.37 a	↘
P6	5.22 ± 0.49 c	3.81 ± 0.44 a ↘	4.58 ± 0.31 b	3.39 ± 0.49 a	↘
P7	1.49 ± 0.30 a	1.73 ± 0.22 a ns	2.18 ± 0.23 b	2.47 ± 0.34 b	ns
P8	11 ± 0.42 d	8.89 ± 0.79 c ↘	5.21 ± 0.62 b	2.65 ± 0.43 a	↘
P9	24.3 ± 1.07 b	23.5 ± 0.63 b ns	21.3 ± 1.02 a	20.29 ± 1.13 a	ns
7S/11S	0.45 ± 0.04 a	0.66 ± 0.04 b ↗	1.1 ± 0.08 c	1.6 ± 0.11 d	↗
N%	8.54 ± 0.15 c	8.42 ± 0.13 bc ns	8.28 ± 0.15 b	8.06 ± 0.16 a	↘

(c)

Protein band	G x E
P1	*
P2	*
P3	*
P4	***
P5	*
P6	ns
P7	ns
P8	ns
P9	ns
7S/11S	***
N %	ns

7S/11S globulins

WT   hmt3

2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2

+S   -S