



HAL
open science

A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues

Karine Gallardo, Christian Firnhaber, H el ene Zuber, Delphine H ericher, Maya Belghazi, Celine C. Henry, Helge K uster, Richard Thompson

► To cite this version:

Karine Gallardo, Christian Firnhaber, H el ene Zuber, Delphine H ericher, Maya Belghazi, et al.. A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Molecular and Cellular Proteomics*, 2007, 6 (12), pp.2165-2179. 10.1074/mcp.M700171-MCP200 . hal-02660247

HAL Id: hal-02660247

<https://hal.inrae.fr/hal-02660247>

Submitted on 30 May 2020

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

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

A Combined Proteome and Transcriptome Analysis of Developing *Medicago truncatula* Seeds

EVIDENCE FOR METABOLIC SPECIALIZATION OF MATERNAL AND FILIAL TISSUES*[§]

Karine Gallardo^{‡§}, Christian Firnhaber[¶], H el ene Zuber[‡], Delphine H ericher[‡], Maya Belghazi^{||**}, C eline Henry^{‡‡}, Helge K uster[¶], and Richard Thompson[‡]

A comparative study of proteome and transcriptome changes during *Medicago truncatula* (cultivar Jemalong) seed development has been carried out. Transcript and protein profiles were parallel across the time course for 50% of the comparisons made, but divergent patterns were also observed, indicative of post-transcriptional events. These data, combined with the analysis of transcript and protein distribution in the isolated seed coat, endosperm, and embryo, demonstrated the major contribution made to the embryo by the surrounding tissues. First, a remarkable compartmentalization of enzymes involved in methionine biosynthesis between the seed tissues was revealed that may regulate the availability of sulfur-containing amino acids for embryo protein synthesis during seed filling. This intertissue compartmentalization, which was also apparent for enzymes of sulfur assimilation, is relevant to strategies for modifying the nutritional value of legume seeds. Second, decreasing levels during seed filling of seed coat and endosperm metabolic enzymes, including essential steps in Met metabolism, are indicative of a metabolic shift from a highly active to a quiescent state as the embryo assimilates nutrients. Third, a concomitant persistence of several proteases in seed coat and endosperm highlighted the importance of proteolysis in these tissues as a supplementary source of amino acids for protein synthesis in the embryo. Finally, the data revealed the sites of expression within the seed of a large number of transporters implied in nutrient import and intraseed translocations. Several of these, including a sulfate transporter, were preferentially expressed in seeds compared with other plant organs. These findings provide new directions for genetic improvement of grain legumes. *Molecular & Cellular Proteomics* 6:2165–2179, 2007.

Seed development starts with the formation of a protective seed coat, the testa, derived from the maternal ovule integuments, which enclose the filial compartments, the endosperm, and embryo. These tissues are in dynamic interaction, as shown by the recent report that the maternal control of seed coat cell elongation and the zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis thaliana* (1). Once these tissues are differentiated, storage compounds accumulate in the endosperm and/or embryo, depending on the species, desiccation tolerance is acquired, and finally metabolic activity declines to a quiescent state (2, 3). At maturity, seed coats were found to be the primary determinant of seed dormancy (4).

During seed filling, the young seed coat supports storage compound synthesis in the filial tissues by transmitting organic nutrients from the phloem, mainly sugars, glutamine, and asparagine (5). Phloem sulfur supply in the form of sulfate, S-methylmethionine (SMM)¹ and glutathione, also influences significantly seed composition (6, 7). The morphology and composition of the filial organs vary greatly among species. In mature seeds of the *Gramineae*, the endosperm serves as the major filial storage tissue, rich in starch but with a protein content of less than 16%. In contrast, the principal storage organ of grain legumes is the cotyledon with protein contents ranging from 20% to as much as 40%, and the mature seed has little endosperm tissue. Like the major storage proteins of barley and maize grains (prolamin, glutelin), the predominant proteins of legume seeds (legumin, vicilin) are low in tryptophan (<1%) and in the sulfur-containing amino acids cysteine and methionine (<1.5%). The nature of the storage proteins that determine the amino acid composition of the total protein

From the [‡]UMR102 INRA/ENESAD, Genetics and Ecophysiology of Grain Legumes, F-21000 Dijon, France, [¶]Genomics of Legume Plants, Institute for Genome Research and Systems Biology, Center for Biotechnology, Bielefeld University, D-33594 Bielefeld, Germany, ^{||}UMR6175 INRA, Mass Spectrometry Platform for Proteomics, F-37380 Nouzilly, France, and the ^{‡‡}Unit e de Biochimie et Structure des Prot eines INRA, Mass Spectrometry Platform for Proteomics, F-78352 Jouy-en-Josas, France

Received, April 16, 2007, and in revised form, August 14, 2007

Published, MCP Papers in Press, September 11, 2007, DOI 10.1074/mcp.M700171-MCP200

¹ The abbreviations used are: SMM, S-methylmethionine; dap, days after pollination; TC, tentative consensus sequence; TIGR MtGI, *M. truncatula* Gene Index of the Institute for Genomic Research; AdoMet, S-adenosylmethionine; AdoHcy, S-adenosylhomocysteine; Hcy, homocysteine; APS, 5'-adenylyl sulfate; mRNA, messenger RNA; α -TIP, α -tonoplast intrinsic protein; RuBisCO, ribulose-bisphosphate carboxylase/oxygenase; Emb, embryo; Eo, endosperm; Sc, seed coat; L, leaves; F, flowers; R, roots; S, stems; qRT, quantitative reverse transcription; BLAST, Basis Local Alignment Search Tool; EST, Expressed Sequence Tag; BAC, bacterial artificial chromosome.

fraction present in seeds is genetically programmed. However, their rate of accumulation also depends on nutrient availability during seed filling (8). Because legume and cereal seeds are major human and livestock food sources, much research and breeding effort are concentrated on optimizing their nutritional value.

Although separate transcriptome and proteome analyses of developing seeds of legumes (9–11), *Arabidopsis* (12), and cereals (13–17) have proven invaluable in identifying changes in expression during seed filling, new strategies are needed which, by comparing transcript and protein patterns, can distinguish protein accumulation driven directly by transcript abundance from that post-transcriptionally regulated. Such proteome-transcriptome comparisons have recently been reported for human, animals, bacteria, archaea, and yeast (18–23) and are also becoming feasible in plants, especially where extensive genomic or expressed sequence information is available, such as *Arabidopsis* (24, 25).

Our objective was to perform a comparative study of the proteome and transcriptome at both spatial and temporal levels during seed development. We have chosen to use the annual barrel medic *Medicago truncatula*, a model legume characterized by a process of seed development very similar to that of other legumes, the only notable difference being a layer of endosperm remaining at maturity (26, 27). Six stages spanning important developmental phases (early seed filling to desiccation) and the three major seed tissues (seed coat, endosperm, embryo), isolated at the switch to storage product accumulation, were analyzed by proteomics and transcriptomics. Such seed tissue analyses have been performed in tomato seeds (28) or in cereal grains (14, 15, 29–31) but have not been reported for a legume species. Comparisons of proteome and transcriptome data provided novel indications as to which processes related to seed development are regulated at the level of the transcriptome and which are controlled at the proteome level. Furthermore, the seed tissue analysis revealed the partitioning of metabolic pathways between filial and maternal tissues. In particular, there is a remarkable compartmentalization of Met metabolism, which is of agronomic interest with respect to legume seeds. Complementary information derived from the transcriptomics dataset was used 1) to further investigate whether differential expression between tissues is specific to Met biosynthetic enzymes or holds more generally for other amino acids, and 2) to identify candidate genes with possible roles in the transfer of amino acids and other nutrients between the seed compartments.

EXPERIMENTAL PROCEDURES

Plant Materials and Growth Conditions—Two independent series of 15 *M. truncatula* plants (cultivar Jemalong, line A17) representing two biological replicates were used for the proteome and transcriptome analyses. The two batches were independently grown in a growth chamber (22/19 °C day/night temperatures, 16-h photoperiod at 220 $\mu\text{E m}^{-2} \text{s}^{-1}$ light intensity, 60–70% relative humidity) at two distinct periods of time (i.e. two subsequent years). In each experiment, plants

were not nodulated and were fertilized three times a week (N:P:K 20:20:20). Individual flowers were tagged on the day of flower opening (i.e. 24 h after pollination) over a one-month period. Pods were harvested at a similar time of the day during the light cycle to avoid circadian effects, from 8 to 44 days after pollination (dap), and developing seeds were collected on Petri dishes placed on ice to prevent any dehydration, weighed (Sartorius ISO 9001 Scale, Quality Control Services, Portland, OR), and rapidly frozen in liquid nitrogen. Because of the small size of *M. truncatula* seeds, ~4000 seeds (5 g of seeds at the 12 dap reference stage and 1 g of seeds for each of the five developmental stages ranging from 14 to 36 dap) were collected per biological replicate on a set of 10 plants for the time course analysis. For seed tissue analyses, a total of 250 seeds were collected per biological replicate on a set of 10 plants. Seed coat, endosperm, and embryo of the freshly harvested seeds were manually separated under a magnifying glass (magnification, $\times 3.5$) on Petri dishes placed on ice. Once isolated, seed tissues were weighed and immediately frozen in liquid nitrogen. From the five remaining plants of the two biological replicates, flowers, leaves, stems, and roots were collected at flowering. All seed and plant tissue samples were separately ground in liquid nitrogen using mortar and pestle. The powder was stored at -80 °C until mRNA and/or protein extraction.

A third batch of plants (third biological replicate) was grown in another year under the conditions described above. A more restricted quantity of seed and tissue samples was collected from this third experiment (0.3 g per sample) to further validate the data obtained via a set of genes (see “qRT-PCR and SYBR Green Detection”).

Isolation of Total RNA, Labeling, Microarray Hybridization, and Data Acquisition/Analysis—The microarray experiments were performed using seeds of six stages ranging from 12 to 36 dap and three seed tissues (seed coat, endosperm, and embryo), each collected on two batches of plants grown independently in two subsequent years to provide two biological replicates. Four technical replicates were performed per biological replicate, including a dye swap. The Mt16kOLI1 microarrays used in this study consist of a set of 16,086 70-mer oligonucleotides (*Medicago* Genome Oligo Set, version 1.0, Operon) representing all tentative consensus sequences (TCs) from the Institute for Genomic Research *M. truncatula* Gene Index 5 (TIGR MtGI). These 70-mer oligonucleotide probes were printed in two duplicates according to Hohnjec *et al.* (32). Total RNA was extracted from 14 dap seed tissues (seed coat, endosperm, and embryo) and from whole seeds collected at 12, 14, 16, 20, 24, and 36 dap, with the phenol/SDS method described by Verwoerd *et al.* (33). As references for co-hybridizations on the microarray slides, one batch of embryo RNA and one batch of 12 dap seed RNA were prepared per biological replicate. The 12 dap stage was chosen as a common reference for the time course because it precedes the synthesis of the storage products (Fig. 1). For each hybridization, 20 μg of total RNA was reverse-transcribed, purified by ultrafiltration (Microcon YM-30, Millipore), and used to synthesize first strand cDNA targets via incorporation of aminoallyl-dUTP as described by Küster *et al.* (34). Reverse transcription efficiency was checked using 2 μl of the reaction mixture on 1% (w/v) agarose gels after ethidium bromide staining. Since a low quantity of cDNA was observed for 14–16 dap seeds and seed coat, presumably because of the presence of polyphenolic compounds in seed coat impeding reverse transcription, the RNA extraction procedure described by Heim *et al.* (35) was successfully applied to these samples. The cDNAs were subsequently labeled with Cy3- or Cy5-*N*-hydroxysuccinimide esters and purified. Labeling efficiency was checked by scanning separated Cy-coupled targets on 1% (w/v) agarose gels with a Typhoon scanner (GE Healthcare/Amersham Biosciences). The embryo reference sample was co-hybridized with the samples derived from seed coat or endosperm. For the time course analysis, the 12 dap reference sample was co-hybridized with

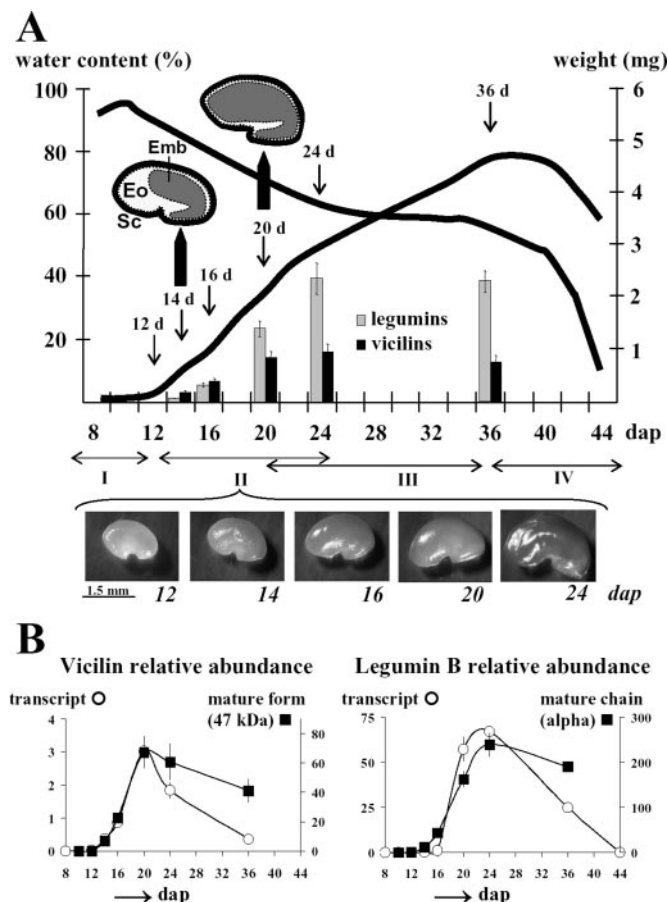


FIG. 1. Developmental stages subjected to mRNA and protein profiling. *A*, changes in seed dry weight (rising curve) and water content (downward curve) during the different phases of seed development: (I) embryogenesis, (II) storage protein synthesis and deposition (i.e. seed filling), (III) seed maturation, and (IV) desiccation. The bars indicate the cumulative volume of the protein spots identified either as vicilins or as legumins. Arrows indicate the various seed samples subjected to transcriptomics and proteomics. Schemes of the developing seeds show the different tissue types (Emb, Eo, Sc) are based on the structure at 14 and 20 dap. Photographs show the phenotypes of the developing seeds at various stages of seed filling. *B*, developmental changes in transcript and protein abundance for vicilin (TC76528) and legumin B (TC85214). Transcript abundance was measured by qRT-PCR with respect to the constitutively expressed *msc27* gene. The mean values (\pm S.D.) of two repeated experiments are presented. Protein abundance (closed symbols) was determined by quantitative proteomics according to “Experimental Procedures.”

the samples derived from the different stages of seed development. Hybridization in the Automatic Slide Processor station (GE Healthcare/Amersham Biosciences) and manual washing steps of the Mt16kOLI1 microarray slides were performed as described by Küster *et al.* (34). Dried slides were scanned with a pixel size of $10 \mu\text{m}$ at optimal settings using a Scanarray 4000 (PerkinElmer Life Sciences).

Spot detection, image segmentation, and quantification were performed using the ImaGene 5.5 software (BioDiscovery, Los Angeles, CA), including manual grid adjustments and spot flagging if necessary. Spots with intensities less than ~ 2 -fold background ($R \leq 1$ for both channels) were automatically flagged “empty.” ImaGene output

files were imported into the ArrayLIMS and EMMA (2.0) microarray analysis software tools (36). During import, spots flagged empty or “poor” (flag value 2 and 1) were removed. After local background subtraction and after applying a floor value of 20, the resulting signal intensities were used for data normalization using a local regression (Lowess) procedure applied globally. Subsequently, M values (\log_2 intensity ratios) and A values (average signal intensities) were calculated according to Dudoit *et al.* (37). Genes significantly up- or down-regulated were identified based on t -lists obtained from EMMA (Supplemental Table 1). Genes were regarded as being differentially expressed if signals were detectable on more than 66% of replicated spots along with $p < 0.05$ and $-1 \geq M \geq 1$ (i.e. 2-fold regulation) for statistical significance. Genes were manually annotated based on automatic BLAST hits of oligo corresponding TIGR MtGI TCs in the curated databases TrEMBL, TrEMBLnew, Protein Information Research, and Swiss-Prot as well as in the Medicago EST Navigation System database. A manual assignment of the seed-expressed genes to functional classes was performed according to the MapMan scheme (38) as described in Supplemental Table 1. Gene expression profiles selected for example by functional class were hierarchically clustered using the Cluster 3.0 software (39) via the average linkage method using an uncentered correlation and visualized using the Java TreeView (1.0.12) software. Data (from ImaGene output files) were submitted to ArrayExpress under accession numbers E-MEXP-904 and E-MEXP-907.

Quantitative RT-PCR and SYBR Green Detection—Verification of differential gene expression was performed by SYBR qRT-PCR from seed and tissue samples collected on three series of plants (three biological replicates) at the stages analyzed by transcriptomics and from additional stages (8, 10, and 44 dap) and plant tissues (flowers, roots, leaves, and stems). Total RNAs were extracted by using the method described by Chang *et al.* (40). RNAs ($20 \mu\text{g}$) were incubated in presence of 20 units of RNase-free RQ1 DNase (Promega, Madison, WI). Non-reverse-transcribed RNA samples were checked for absence of contaminating genomic DNA by PCR using primers for the constitutively expressed *msc27* gene (TC85211) (41). Samples were reverse-transcribed using the iScript cDNA synthesis Kit (Bio-Rad, Hercules, CA), and diluted in a final volume of 1 ml. Primers were designed preferably in 3' regions of the genes to amplify fragments of 50–150 base pairs (Supplemental Table 2). The qRT-PCR reactions were carried out in duplicate in a Bio-Rad iCycler in a final volume of $25 \mu\text{l}$ containing $5 \mu\text{l}$ of diluted cDNA, 200 nM of each primer, and $12.5 \mu\text{l}$ of iQ SYBR Green Supermix for 2 min at 95°C , 40 cycles of 20 s at 95°C , 20 s at 60°C , and 30 s at 72°C . To establish the presence of a single PCR product and the absence of primer dimers, melting analysis (i.e. heat dissociation of oligonucleotides) was applied immediately after PCR by heating PCR products from 59 to 96°C . Relative gene expression was calculated according to the relative standard curve method (ΔCT) using *M. truncatula msc27* as a constitutively expressed gene. The expression stability of the *msc27* gene in the different test samples was verified by comparison with two other constitutively expressed genes encoding γ -tubulin (TC81141) and the eukaryotic translation initiation factor 5A-2 (TC76568) (data not shown). As a control for the data obtained, we additionally performed digital expression profiling via the expression summary tool from the TIGR MtGI 8, allowing prediction of global expression profiles based on the percentage of ESTs corresponding to a given gene in different cDNA libraries.

Total Protein Extraction and Two-dimensional Electrophoresis—Total proteins were extracted as described by Gallardo *et al.* (9) from each of the seed stages, and tissues were analyzed in parallel with the Mt16kOLI1 microarrays, i.e. from seeds ranging from 12 to 36 dap and seed tissues at 14 dap, all collected on two series of plants grown independently in two subsequent years. From each of these two

biological replicates, two independent protein extractions were performed, and two replicated two-dimensional gels were prepared from each protein extract (four technical replicates per biological replicate). Total proteins were extracted from developing seeds in 20 μ l/mg of seed dry matter of thiourea/urea lysis buffer (see the corresponding seed fresh weight in Fig. 1A and Ref. 42). Total proteins of seed tissues were extracted in 2 μ l of the same buffer per mg of seed fresh weight. Protein concentration was measured according to Bradford (43). Proteins were first separated by IEF using a constant volume (20 μ l) of the protein extracts from developing seeds and 150 μ g of proteins from seed coat, embryo, and endosperm. Proteins were separated using gel strips forming an immobilized nonlinear pH 3 to 10 gradient (Immobiline DryStrip, 24 cm; GE Healthcare/Amersham Biosciences), allowing an accurate visualization of the *M. truncatula* seed proteome by minimizing overlaps. Strips were rehydrated in the IPGphor system (GE Healthcare/Amersham Biosciences) for 7 h at 20 °C with the thiourea/urea lysis buffer containing 2% (v/v) Triton X-100, 20 mM DTT, and the protein extracts. IEF was performed at 20 °C in the IPGphor system for 7 h at 50 V, 1 h at 300 V, 2 h at 3.5 kV, and 7 h at 8 kV. Prior to the second dimension, each gel strip was incubated at room temperature for 2 \times 15 min in 2 \times 15 ml equilibration buffer as described in Gallardo *et al.* (44). Proteins were separated in vertical polyacrylamide gels according to Gallardo *et al.* (44).

Protein Staining and Image Analysis—Gels were stained with Coomassie Brilliant Blue G-250 (Bio-Rad) according to Mathesius *et al.* (45). Image acquisition was done using the Odyssey Infrared Imaging System (LI-COR Biosciences, Lincoln, NE) at 700 nm with a resolution of 169 μ m. Image analyses were carried out with the ImageMaster 2D Platinum version 5.0 software (GE Healthcare/Amersham Biosciences) according to the instruction manual. After matching the spots detected during the time course, a synthetic gel was created, allowing the visualization of all the polypeptides. This composite reference map was then used for protein pattern comparison during the time course and for matching with two-dimensional gels from the seed tissues. An attempt was made not to include spots where overlap with other spots was readily apparent.

Spot Volume Normalization—We have normalized the volume of each spot (*i.e.* spot abundance) to total spot volume in each gel for the three seed tissues. In the context of the time course of seed development, this method of normalization is problematic because the storage proteins accumulate to 70% of total spot volume, and the proportion they represent of the total proteins changes drastically during seed filling. Therefore, this method of normalization approximates spot abundance relative to storage protein concentration. To circumvent this problem, we have used the scaling procedure described previously (9, 42), which involves the normalization, in each two-dimensional gel, of the volume of each spot to the volume of a set of housekeeping proteins, which showed little variation in intensity during seed development. This method of normalization allowed a reliable comparison with the microarray data, whose normalization is based on the expression ratios of the total number of probes using a Lowess procedure, with probes corresponding to the storage protein genes only representing a very small fraction of the total number of probes (less than one per thousand) and the corresponding signal representing a vanishingly small proportion of the total.

Statistical Analyses of Protein Variations—For each spot, the quantitative data obtained during the time course were submitted to a one-way analysis of variance using the SAS software (46). Then, a Dunnett's *t* test was performed to compare each stage of the time course (14, 16, 20, 24, and 36 dap) to the 12 dap stage used as reference in the transcriptomics experiments. Statistically significant differences (with 95% confidence intervals) in the quantities of individual protein spots as compared with the reference stage were identified (Supplemental Table 3). Similarly, for each spot, the quan-

titative seed tissue data were submitted to a one-way analysis of variance followed by a Dunnett's *t* test to compare the seed tissues (seed coat, endosperm) to the embryo used as reference in the transcriptomics experiments. Statistically significant differences ($p < 0.05$) in the quantities of individual protein spots as compared with the embryo were identified (Supplemental Table 3).

Protein Identification and Comparison with Transcriptomics Data-sets—Data are reported for 224 protein spots identified by MS. Of these 224 proteins, 56 were previously identified by MALDI-TOF MS (9). In the current study, the identity of 31 proteins excised from two-dimensional gels derived from whole seeds was obtained by MALDI-TOF MS (Voyager DE super STR, Applied Biosystems, Foster City, CA) equipped with a nitrogen laser emitting at 337 nm. The excised gel plugs were destained in 50% acetonitrile and 50 mM ammonium bicarbonate (v/v). After gel-drying for 30 min, the digestion was performed in 25 μ l of 50 mM ammonium bicarbonate (pH 8.0) with 0.5 μ g of modified trypsin (sequencing grade, Promega, Charbonnières-les-Bains, France) for 16 h in a thermomixer (Eppendorf, Le Pecq, France) at 37 °C with vortexing at 500 rpm. One microliter of supernatant was mixed on the stainless steel MALDI plate with 1 μ l of α -cyano-4-hydroxycinnamic acid (Sigma-Aldrich, Saint Quentin Fallavier, France) at 4 mg/ml in acetonitrile/TFA (50:50; v/v) 0.3% and dried at room temperature. Spectra were recorded in positive reflector mode with 20 kV as accelerating voltage, a delayed extraction time of 130 ns, and a 62% grid voltage. Mass spectra were treated by Data Explorer 4.2 (Applied Biosystems) with the following parameters: noise filter/smooth (noise removal of 2), 0.5% base peak intensity, 0.5% maximum peak area, resolution-dependent settings option, and peak resolution of 10,000. Spectra were deisotoped and internally calibrated using the autolytic trypsin fragments characterized by $[M+H]^+$ 842.509 and 2211.104 Da. Spectral profiles were collected in the mass range 750–3500 Da. The known cluster matrix masses (47), trypsin auto-cleavable peptides, and human keratin masses (48) were removed for database searches. A search in the MtGI database (226,923 ESTs assembled in 18,612 TCs and 18,238 singletons, from 55 cDNA libraries, last release: 8.0 January 19, 2005) and in the *M. truncatula* translated sequences downloaded from NCBI was done using the MS-FIT program in local version (Protein Prospector v3.2.1). All peptide masses were assumed to be monoisotopic and protonated molecular ions $[M+H]^+$. The MS-FIT search parameters were trypsin specificity, two missed cleavages, 30 ppm mass accuracy, carbamidomethylation, and other modifications (N-terminal pyroglu, oxidation of Met, protein N terminus acetylated). A match was considered significant when the sequence coverage was at least 20% with more than four nonoverlapping peptides. In most cases, the molecular weight search score was above 1e+003 (the lowest score was 297), and the numbers of peptides identified was high, with an average of 10 peptides per protein hit. The spectrums, scores, percent matches, percent sequence coverage, peptide masses, and mass errors are provided in Supplemental Table 3.

Subsequently, a total of 137 additional spots, including those that could not be identified unambiguously by MALDI-TOF and spots derived from the separated seed tissues and/or from the whole seeds, were subjected to nano-LC-MS/MS sequencing. Protein spots were excised from two-dimensional gels and reduced with 10 mM DTT for 45 min at 56 °C, alkylated with 55 mM iodoacetamide for 30 min in the dark at room temperature, and incubated overnight at 37 °C with 12.5 ng/ μ l trypsin (sequencing grade; Roche Applied Science) in 25 mM ammonium bicarbonate. The tryptic fragments were extracted, dried, reconstituted with 2% (v/v) acetonitrile and 0.1% formic acid, and sonicated for 10 min. Tryptic peptides were sequenced by nano-LC-MS/MS (Q-TOF-Ultima Global equipped with a nano-ESI source coupled with a Cap LC nanoHPLC, Waters Micromass, Waters, Saint Quentin en Yvelines, France) in the Data Dependent Acquisition mode

allowing the selection of three precursor ions per survey scan. Only doubly and triply charged ions were selected for fragmentation over a mass range of m/z 400–1300. A spray voltage of 3.2 kV was applied. The peptides were loaded on a C18 column (Atlantis dC18, 3 μm , 75 μm \times 150 mm Nano Ease, Waters) and eluted with a 5–60% linear gradient with water/acetonitrile 98/2 (v/v) containing 0.1% formic acid (buffer A) and water/acetonitrile 20/80 (v/v) containing 0.1% formic acid (buffer B) over 30 min at a flow rate of 200 nL/min. MS/MS raw data were processed (smooth 3/2 Savitzky Golay and no deisotoping) using the ProteinLynx Global Server 2.05 software (Waters), and peak lists were exported in the micromass pkl format. Peak lists of precursor and fragment ions were matched automatically to proteins in the NCBI nonredundant database March 16, 2007: 4,761,919 sequences, 1,643,098,755 residues) and the GenBank Viridiplantae other than Arabidopsis and *Oryza sativa* EST database (April 29, 2006: 140,695,050 sequences, 26,889,187,900 residues) using the MASCOT version 2.2 program (Matrix Science, London, UK). We first performed a search in the NCBI nr database and then in the GenBank EST database, where we applied a restriction to Viridiplantae other than Arabidopsis and *Oryza sativa* because of sequence redundancy between the two databases. The MASCOT search parameters were trypsin specificity, one missed cleavage, carbamidomethyl Cys and oxidation of Met, and 0.2 Da mass tolerance on both precursor and fragment ions. All proteins identified have a MASCOT score above the significance level corresponding to $p < 0.05$. To validate protein identification based on multiple peptides, only matches with individual ion scores above 20 were considered. In most cases, at least two different nonoverlapping peptide sequences of more than 6 amino acids with a mass tolerance < 0.05 Da were obtained. Moreover, among the positive matches based on one unique peptide, only spectra containing a series of at least 5 consecutive y or b ions with individual ions scores above a threshold value calculated by the MASCOT algorithm with our search parameters were accepted (identity threshold of 41 for the NCBI nr database and of 60 for the GenBank EST database for Viridiplantae other than Arabidopsis and *Oryza sativa*). These validation criteria are a good compromise to limit the number of false positive matches without missing real proteins of interest. When the same set of peptides matched different EST accessions, we retained the *M. truncatula* ESTs and checked that they belong to the same contig from the TIGR MtGI. Of the 137 spots analyzed, 123 were unambiguously identified. The protein sequences were subjected manually to BLAST searches against the MtGI database to identify the corresponding TC sequence. The peptide sequences from nano-LC-MS/MS with accession number, description, protein and peptide MASCOT scores, MS/MS fragmentation for unique peptide-based identifications, sequence coverage, and BLAST probability scores are provided in Supplemental Table 3.

The protein spots that did not fit these validation criteria (14 of 137) were reanalyzed by HPLC coupled with a LCQ Deca XP+ ion trap (Thermo Electron, San Jose, CA) using a nano electrospray interface according to Méchin *et al.* (17). Ionization (1.3–1.5 kV ionization potential) was performed with liquid junction and a noncoated capillary probe (10-mm inner diameter; New Objective). Peptide ions were analyzed using Xcalibur 1.4 (Thermo Electron) with the following data-dependent acquisition steps: 1) full MS scan (m/z ratio 400–1900, centroid mode), 2) ZoomScan on a selected precursor (scan at high resolution in profile mode on a m/z window of 4), and 3) MS/MS (stability parameter = 0.22, 50 ms activation time, 40% collision energy, centroid mode). Steps 2 and 3 were repeated for the two major ions detected in step 1. Dynamic exclusion was set to 30 s. A search in the MtGI database (226,923 ESTs clustered, on 36,878 TCs, from 55 cDNA libraries, last release: 8.0 January 19, 2005) was performed with Bioworks 3.1 (Thermo Electron). Trypsin digestion, Cys carboxyamidomethylation, and Met oxidation were set to enzy-

matic cleavage, static, and possible modifications, respectively. Precursor mass and fragment mass tolerance were 1.4 Da and 1 Da, respectively. Identified tryptic peptides were filtered according to their cross-correlation score (Xcorr) higher than 1.7, 2.2, and 3.3 for mono-, di-, and tri-charged peptides, respectively. A minimum of two different peptides was required. In the case of identification with two or three MS/MS spectra, similarity between the experimental and the theoretical MS/MS spectra was visually confirmed. The peptide sequences with accession number, cross-correlation scores, ΔCn , and percent sequence coverage are provided in Supplemental Table 3.

The identification of *M. truncatula* protein sequences allowed us to search for the corresponding genes on the Mt16kOL1. To compare protein and transcript patterns, protein accumulation values were expressed as \log_2 ratios of spot abundance in the seed samples relative to that in the reference sample (12 dap for the time course and embryo for the tissues). Protein patterns and the corresponding transcript profiles were hierarchically clustered using the Cluster 3.0 software (39) with the commonly used average linkage method and were visualized using Java TreeView (1.0.5), except for the genes displaying no significant regulation during the time course ($-1 < M < 1$), which were independently subjected to a hierarchical clustering. Pearson correlation coefficients (r) were used to evaluate the levels of correlation between transcript and protein profiles during the time course of seed development and in seed tissues (Statistica software, version 7.0, StatSoft). An r value of 1.0 indicates perfect correlation, whereas a value of 0 indicates no correlation, and $r = -1$ indicates perfect anti-correlation. A p value < 0.05 was considered to be statistically significant. The transcriptome/proteome comparisons made, including the Pearson correlation coefficients, are available in Supplemental Table 4.

RESULTS

Integrative Analysis of the Proteome and Transcriptome during Seed Development—Protein and RNA were prepared from seeds collected at six key stages characterized previously at the physiological level (9): 12 dap (at the end of the embryonic cell division phase and preceding the synthesis of storage products), 14 and 16 dap (early stages of seed filling), 20 dap (peak of accumulation of storage compounds), 24 dap (end of seed filling and late maturation), and 36 dap (stage of physiological maturity and desiccation) (Fig. 1). Proteome analysis of these seeds was performed by two-dimensional electrophoresis (Supplemental Fig. 1), and transcriptome changes were monitored using the Mt16kOL1 oligonucleotide microarrays. The cDNAs from 12 dap seeds (reference stage) were cohybridized on the microarrays with cDNAs from 14, 16, 20, 24, and 36 dap. To allow a comparison between proteome and transcriptome data, protein abundance during the time course was compared with that at the 12 dap reference stage. A high resolution composite two-dimensional map pinpointing the location of 790 polypeptides profiled during seed development was established (Supplemental Fig. 1). The two-dimensional map includes 56 proteins identified previously (9) and 168 additional spots analyzed by MS in the present study (Supplemental Table 3).

Of the 790 polypeptides profiled during seed development, 615 varied significantly at characteristic stages as compared with the 12 dap stage. These variations are accompanied by significant changes in the transcriptome because about 50%

of the transcripts (~4800 mRNAs) detected in developing seeds varied in abundance. A complete list of all genes significantly regulated (≥ 2 -fold, $p < 0.05$) at either the transcript or protein level, having been reannotated and classified into functional classes as defined by the MapMan ontology (38), is provided in Supplemental Tables 1 and 3. As expected, the number of differentially accumulated proteins and transcripts increased from 14 to 24 dap, corresponding to the seed filling phase. Strikingly, a 2-fold increase in the total number of up-regulated transcripts was observed between 24 dap and physiological seed maturity (36 dap), whereas the number of up-regulated proteins detected did not increase in the same period. Interestingly, the most marked changes are in genes related to transcription and RNA processing (class #27, Supplemental Fig. 2 and Supplemental Table 1). These transcripts presumably contribute to the stored mRNA pool used for protein synthesis during germination (49). They represent potential indicators of germination performance.

Comparison of Protein Profiles with Transcript Patterns during Seed Development—The proteome and transcriptome profiles were compared to investigate at which level protein accumulation is regulated, *i.e.* if regulation takes place at the level of the transcript or rather at the level of the protein itself (protein degradation and other post-translational modifications). A total of 208 spots matched to TCs corresponding to oligonucleotide probes on the Mt16kOLI1 microarray. For 156 of these, RNA and protein expression profiles throughout seed development were obtained. The protein and transcript profiles, with values expressed as mean of \log_2 ratios using the 12 dap sample as reference, were hierarchically clustered. Nine distinct groups of profiles were identified that reflect different modes of regulation (Fig. 2).

Groups 1–4, Proteins and Transcripts Coordinately Expressed—First, we found that 50% of the identified proteins display a profile similar to that of the corresponding transcripts, suggesting that protein accumulation is primarily regulated by transcript abundance. These proteins belong to group 1 (increased levels in the course of seed development), group 2 (transiently increased levels), group 3 (decreased levels), and group 4 (constant levels) (Fig. 2). A positive linear correlation between transcript and protein levels was observed for 80% of the genes clustered in groups 1 and 2. The r and p values obtained are provided in Supplemental Table 4. Members of the major seed storage protein families, the legumins (or 11S globulins) and vicilins (or 7S globulins), belong to the group 1 of proteins that are transcriptionally up-regulated. Their synthesis begins in a specific temporal order as described previously (9): vicilins at 14 dap and legumins at 16 dap, reaching a maximum at 20–24 dap (Fig. 1B). The late embryogenesis abundant proteins, markers of seed maturation, also belong to this group. The other proteins belonging to groups 1 and 2 are transcriptionally up-regulated, either during reserve accumulation or maturation, and may play associated roles, such as in protein degradation, (*e.g.* the subtilisin-type

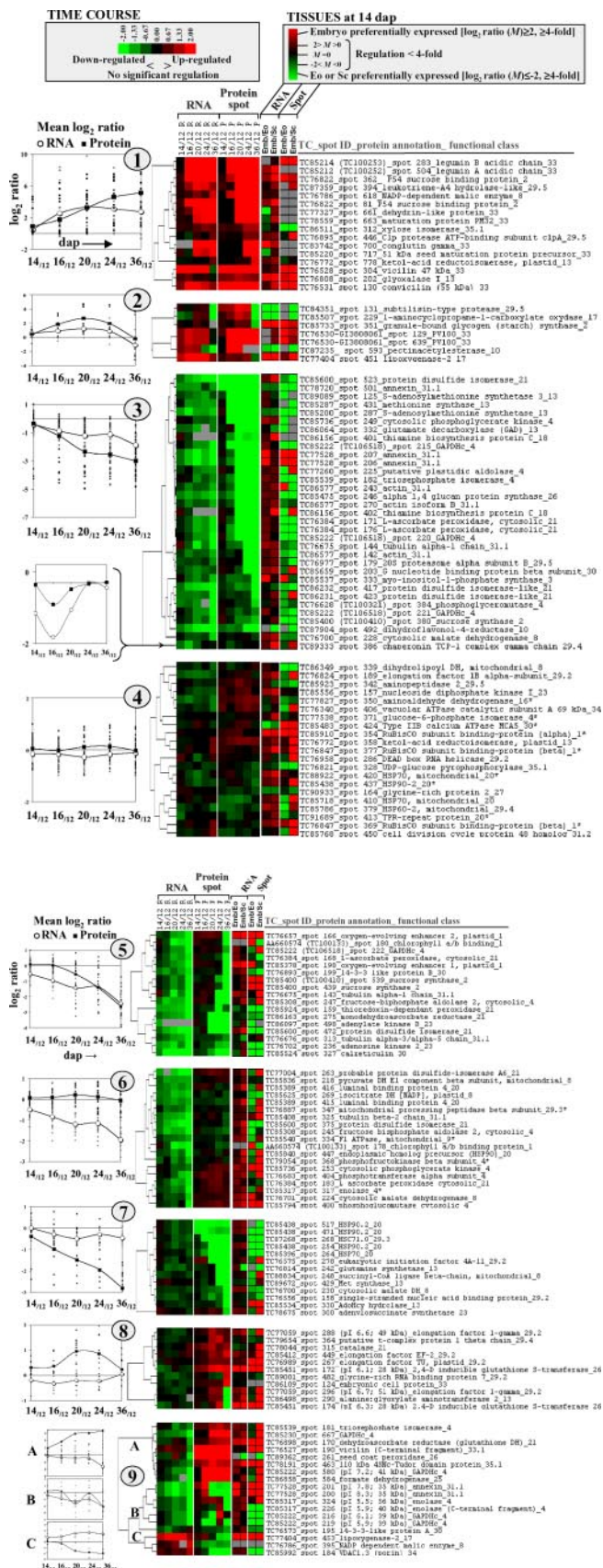
protease or the Clp protease, Ref. 50), in sugar transport (P54 sucrose-binding proteins), starch synthesis (starch synthase), or in cell wall modification (pectin acetyltransferase).

Group 3 is composed of 33 protein spots whose abundance decreases from the early stages of seed filling onwards. A highly significant correlation ($r > 0.9$, $p < 0.05$) between transcript and protein levels was found for 13 genes in this group, and for 14 further genes, the r value exceeded 0.5 (Supplemental Table 4). Many of them are related to cytoskeletal organization in the cell (tubulin and actin), redox regulation, abiotic stresses, glycolysis, and Met metabolism. Among enzymes of Met metabolism identified were a Met synthase and two isoforms of *S*-adenosylmethionine (AdoMet) synthetase, previously shown to be associated with the status of metabolic activity in seeds (9, 51, 52). Their decreased levels would be consistent with the switch from active metabolism to a quiescent state. Finally, group 4 includes proteins whose accumulation did not vary significantly during seed development ($-1 < M < 1$, *i.e.* less than 2-fold regulation), suggesting a continuing function throughout this process. Many are molecular chaperones involved in protein folding either in the plastid (ribulose biphosphate carboxylase-oxygenase (RuBisCO) subunit-binding proteins) or in the mitochondrion (heat shock protein family members), and several are required for protein synthesis (RNA helicase, elongation factor 1B α -subunit).

Groups 5 and 6, Sequences Displaying Apparent Preferential Transcript Turnover—Fifteen sequences of 36 have a correlation coefficient between transcript and protein patterns below 0.5 reflecting different degrees of stability (Supplemental Table 4). The levels of these proteins remain constant while transcript abundance decreases. They include those essential for mitochondrial and chloroplast function during seed development. Some are present throughout seed filling (group 5), such as proteins involved in photosynthesis (two oxygen evolving enhancer proteins and a chlorophyll a/b binding protein), a process shown to produce oxygen for respiration and reserve synthesis in developing seeds (5). Others are present up to desiccation (group 6), such as the mitochondrial processing peptidase ensuring the continued functioning of the mitochondrion.

Group 7, Sequences Displaying Apparent Preferential Protein Turnover—The proteins whose abundance decreases while the transcript level remains constant are involved in amino acid metabolism (glutamine synthetase, Met synthase, and *S*-adenosylhomocysteine (AdoHcy) hydrolase) or related to stress, suggesting that the transcripts encoding these proteins may form part of a pool of stored mRNAs that could be reutilized during seed development (*e.g.* in response to a stress) or later during germination. Further studies are needed to characterize the occurrence of protein and/or transcript turnover by using radiolabeled precursors for their synthesis (49, 53).

Groups 8 and 9, Protein and Transcript Poorly Correlated—Our study revealed several proteins displaying a profile differing from that of the corresponding transcripts during seed



development ($r < 0.5$ for 21 sequences of 29). The majority of these proteins are involved in the regulation of protein synthesis (translation elongation factors) either in the plastid or in the cytosol, in oxidative stress or in detoxification (e.g. catalase and glutathione S-transferase). The underlying causes for the association of genes observed in groups 8 and 9 will be the subject of further investigation.

Protein and Transcript Distribution in Seed Coat, Endosperm, and Embryo—We have examined the distribution of the proteins and transcripts in the three major seed tissues, seed coat, endosperm, and embryo, at the onset of seed filling (14 dap, Fig. 1A). Protein distribution in the three tissues was investigated by using as reference the high resolution composite two-dimensional map established from the time course data (Supplemental Fig. 1), and transcript distribution was studied by cohybridizing, on the Mt16kOLI1 microarrays, cDNAs derived from either the seed coat or the endosperm with embryo cDNAs. To allow a comparison between proteome and transcriptome data, protein spot abundance in the seed coat or endosperm was compared with that in the embryo reference tissue. A classification of the genes differentially expressed (at least 2-fold with a p value < 0.05) at the transcript or protein level in embryo versus seed coat or endosperm is presented in Supplemental Fig. 3, and a complete list is provided in Supplemental Tables 1 and 3.

Despite the difficulty of collecting separated tissue components from seeds as small as those of *M. truncatula* at an early stage of development, we identified a large number of proteins and transcripts enriched in the separated tissues. The

Fig. 2. Clustering of expression profiles of proteins identified by MS (Supplemental Fig. 1) with their corresponding transcript profiles. Log₂ expression ratios (M) were calculated between the seed stages analyzed (14, 16, 20, 24, and 36 dap) and the 12 dap reference stage as well as between the reference tissue Emb versus Eo or Sc. The color code for seed tissues is indicated on the top right. The green color indicates a preferential expression in seed coat or endosperm, and the red color indicates a preferential expression in embryo. Genes up-regulated in the whole seed during the time course compared with the 12 dap reference or in the embryo are marked in red (color code on the top left). Gray indicates no detectable expression. For each group, the log₂ ratios at each stage of the clustered genes are plotted with the group mean profile outlined for transcripts (open symbols) and proteins (closed symbols) to point out either transcriptional regulation or differences in transcript or protein accumulation during seed development. For each polypeptide is indicated the corresponding TIGR MtGI TC-ID, the spot-ID, protein annotation, and the functional class assigned according to the MapMan scheme. Constitutively expressed proteins used in the normalization procedure are marked by asterisks. Several polypeptides with variations in their pI values, such as annexin and cytosolic glyceraldehyde-3-phosphate dehydrogenase (group 9 (B)), reflect seed filling-associated post-translational modifications. A search of the PROSITE database (Swiss Institute of Bioinformatics, Geneva, Switzerland) for motifs revealed, for example, 14 potential phosphorylation sites in sequences for both annexin and cytosolic glyceraldehyde-3-phosphate dehydrogenase.

proteomes of the three tissues are highly contrasted, with 95% of the spots detected being differentially accumulated (≥ 2 -fold, $p < 0.05$) in embryo *versus* seed coat or endosperm (Fig. 3). The microarray data revealed 718, 1382, and 2029 genes predominantly expressed in endosperm, seed coat, and embryo, respectively (about one-third of all genes expressed in these tissues). The effectiveness of seed component separation was confirmed by the detection of marker gene products known to be associated with a particular seed tissue. For example, the seed coat peroxidase was not detected, at either transcript or protein level, in the endosperm and embryo extracts (spot 261 in Fig. 3A, 10-fold higher expressed in seed coat *versus* embryo), and storage proteins and their corresponding transcripts were preferentially accumulated in embryo cells (4-fold higher expressed in embryo, Fig. 3A). This study carried out at the seed tissue level yielded a number of novel findings, including the identification of endosperm-specific gene products, with few being reported to date in legumes (54).

The correlation between mRNA and protein expression ratios in seed tissues is shown in Supplemental Table 4. Positive correlations between mRNA and protein expression ratios in embryo *versus* seed coat ($r = 0.5$, $p < 0.05$) and embryo *versus* endosperm ($r = 0.4$, $p < 0.05$) were observed for the entire dataset (156 genes). When considering each group depicted in Fig. 2, different degrees of correlation were observed. First, those genes transiently up-regulated during seed filling (group 2 in Fig. 2) show coordinate distribution of RNA and protein ($r = 0.99$, $p < 0.05$) whether located in embryo cells (e.g. starch synthase) or in the seed coat (e.g. pectin acetyltransferase, 1-aminocyclopropane-1-carboxylate oxidase). Among the gene products up-regulated during seed filling and preferentially detected in seed coat is a subtilisin-type protease with a possible role in endogenous nitrogen remobilization (spot 131 in Figs. 3B and 4A). Second, a correlation value of 0.55 ($p < 0.05$) between mRNA and protein expression ratios in seed tissues was found for genes of groups 3 that are down-regulated during seed development. An extensive protein turnover in the seed coat and/or endosperm may occur as the embryo enlarges and assimilates nutrients since most of the proteins whose abundance decreases were preferentially detected in these surrounding tissues. The lower abundance of the corresponding transcripts in the seed coat and endosperm at the same stage may reflect a more rapid turnover of the transcripts (group 3, Fig. 2). Finally, low to moderate correlations ($r = -0.07$ to 0.5) between mRNA and protein expression ratios in seed tissues were observed for groups 1 and 4 to 9, reflecting differences in the steady-state levels of the proteins and transcripts in the seed tissues at the onset of seed filling.

qRT-PCR for Selected Genes and Distribution in the Different Plant Organs—We have conducted qRT-PCR experiments to profile the expression of 11 selected genes in the corresponding seed tissues and developmental stages collected on three series of plants grown independently (Figs. 4B

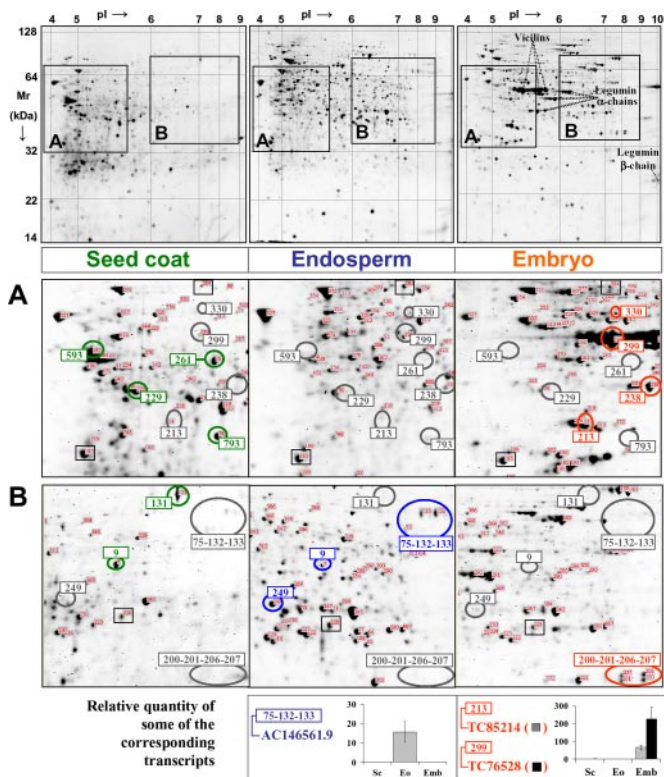


Fig. 3. Typical two-dimensional protein profiles obtained from seed coat (A), endosperm (B), and embryo extracts at the onset of storage protein synthesis (14 dap). The location of the various vicilin and legumin spots is indicated in a typical two-dimensional gel from embryo (see Supplemental Table 3 for a complete list of protein identity). Two representative parts of the two-dimensional gels are enlarged in A and B. Some proteins preferentially accumulated in the seed coat, endosperm, or embryo are *encircled*, whereas some constitutively accumulated proteins are marked by *squares*. Seed coat-located spots (*green*) are Met synthase (9, EST singleton NF075D02ST1F1016), seed coat peroxidase (261, TC89362), pectin acetyltransferase (593, TC87235), 1-aminocyclopropane-1-carboxylate oxidase (229, TC85507), cytosolic fructose-bisphosphate aldolase (793, TC85308), and subtilisin-type protease (131, TC84351). Endosperm-located spots (*blue*) are Met synthase (9), subtilisin-like proteins (75-132-133, gene accessible in the BAC clone mth2-12j18 #AC146561.9), and cytosolic phosphoglycerate kinase (249, TC85736). The embryo-located spots (*orange*) are AdoHcy hydrolase (330, TC85534), vicilin (299, TC76528), legumin α -chains (213, TC85214; 238, TC85218), and annexin (200-201-206-207, TC77528). For some of the endosperm- and embryo-located proteins, the distribution of the corresponding transcripts is shown at the *bottom* of the figure.

and 5A–D). The choice of these genes was based on: 1) their proteomics- and/or microarray-derived patterns, 2) their possible roles during seed filling based on the MapMan ontology (38), and 3) preferential expression in seeds, based on EST frequencies accessible via the TIGR MtGI e-northern. The selected seed coat or endosperm tissue-associated genes are potentially involved in Met metabolism (Hcy S-methyltransferase, 5 in Fig. 6), proteolysis (subtilisin, Fig. 4B), or in the transport of amino acids (Fig. 5A) or sulfate (Fig. 5B), whereas the selected embryo-associated genes correspond

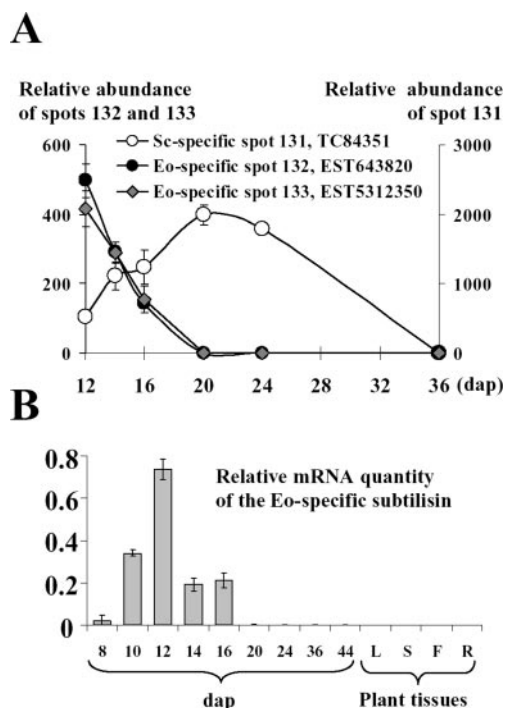


FIG. 4. Kinetics of synthesis of serine proteases in developing seeds. *A*, protein level of the endosperm-specific subtilisin spots 132 and 133 (EST643820 and EST5312350, respectively) and of the seed coat-specific subtilisin spot 131 (TC84351) from 12 to 36 dap. Spot volume was obtained by using the ImageMaster 2D Platinum software as noted in “Experimental Procedures.” *B*, relative quantity of the transcript corresponding to the endosperm-specific subtilisin spots 132 and 133 (matched the same gene sequence on the BAC clone *mth2-12j18* #AC146561.9, Fig. 3) during seed development and in plant tissues (L, leaves; S, stems; F, flowers; R, roots), as well as in Sc, Eo, and Emb. The mRNA quantity was measured by qRT-PCR with respect to the constitutively expressed *msc27* gene.

to storage proteins (Fig. 1B) and to proteins potentially involved in nitrogen sensing (PII protein, Fig. 7), protein transport (Fig. 5C), or solute/water exchange of storage vacuoles (α -tonoplast intrinsic protein (α -TIP) aquaporins, Fig. 5D). The qRT-PCR data from developing seed and tissues collected on three series of plants (Supplemental Table 2) were consistent with those obtained using microarrays, thus confirming the robustness of our results, but the dynamic range of induction measured for a sequence was ~ 3 -fold higher with qRT-PCR, a common observation as mentioned previously (34). The distribution of these transcripts within the plant was examined in further qRT-PCR experiments (Figs. 4B and 5A–D). Eight genes of 11 were preferentially expressed in immature seeds compared with the other plant organs (e.g. protein transporter, sulfate transporter, and subtilisin), implying a specialized function in developing seeds. Five of these seed-specific genes were expressed in embryo cells during storage protein deposition. One of them has high sequence similarity (73%) to the Tim17:22 gene family of protein transporters found in the mitochondrial inner membrane (Fig. 5C). Knowing the central role of mitochondria in storage cells of developing seed (5), it

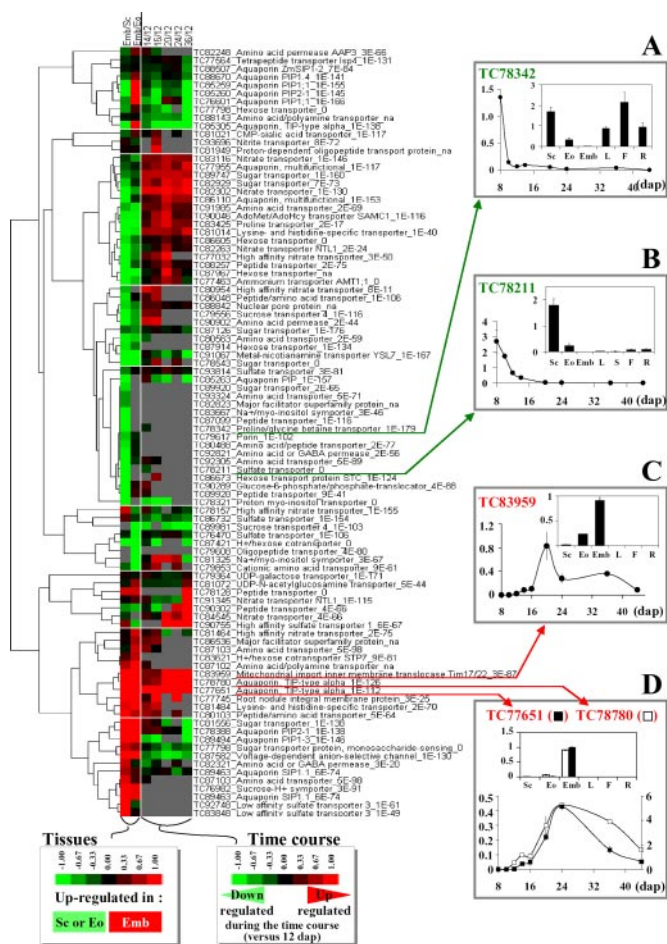


FIG. 5. Heat map of hierarchically clustered expression profiles of selected genes encoding various transporters. Mt16kOLI1 microarray data were subjected to Cluster 3.0 software and visualized using Java TreeView. Genes up-regulated during the time course (14–36 dap) compared with the reference seed stage (12 dap) or preferentially expressed in embryo are marked in red. Note that genes significantly down-regulated in embryo (≥ 2 -fold, $p \leq 0.05$, \log_2 ratio ≤ -1 , marked in bright green) correspond to genes preferentially expressed in seed coat or endosperm. Gray indicates no detectable expression. Four of these genes found by microarrays to be up-regulated in seed tissues and at characteristic stages of seed development were also analyzed by qRT-PCR (boxes A–D). Expression is shown during seed development (from 8 dap to the mature stage) and in the seed and plant tissues (bars). The mRNA quantity was calculated with respect to the constitutively expressed *msc27* gene, and the mean values (\pm S.D.) of two repeated experiments are presented. *A*, TC8342, putative amino acid transporter; *B*, TC78211, sulfate transporter; *C*, TC83959, mitochondrial import inner membrane translocase; *D*, TC77651 and TC78780, two aquaporins (α -TIPs).

would be of interest to investigate the specialized function of this putative translocase. The four other genes encode storage proteins and α -TIPs, aquaporins regulating the flux of water and solutes across the storage vacuole membranes (55). The two α -TIP genes were expressed concomitantly with legumins, suggesting a role in the deposition of storage proteins (Figs. 1B and 5D).

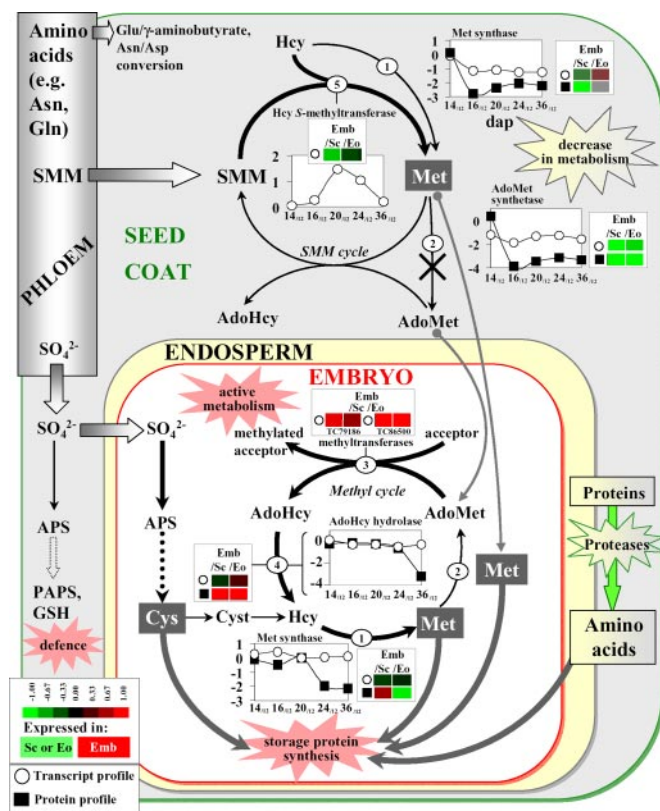


FIG. 6. A working model of accumulation of sulfur-containing amino acids Met and Cys for storage protein synthesis during seed filling. Intertissue compartmentalization of the pathways was deduced from expression profiling data of the corresponding genes (see also Fig. 7). Genes preferentially expressed (≥ 2 -fold, $p \leq 0.05$, \log_2 ratios ≥ 1) in the embryo are marked in *red*, whereas a preferential expression in the seed coat or endosperm is marked in *green*. In addition, the time course (\log_2 expression ratios of seeds 14–36 dap versus seeds 12 dap) is depicted (transcript pattern, *open symbols*; protein pattern, *closed symbols*). Pathways proposed to be preferentially active during seed filling are marked in *bold*. Proteases in the seed coat and endosperm important to provide amino acids for protein synthesis within the embryo are also depicted. Substrates: *CdRP*, 1-(2-carboxyphenylamino)-1-deoxy-D-ribose-5-phosphate; *Cyst*, cystathionine; *GSH*, glutathione. Enzymes: 1, seed coat-located Met synthase (TC85287, spot 431), embryo-located Met synthase, (TC89672, spot 429); 2, AdoMet synthetase (TC85200, spot 287); 3, S-AdoMet-dependent methyltransferase (tissue expression data are shown for the genes encoding a protein carboxyl methylase (TC79186) and a sterol 24-C-methyltransferase (TC86500)); 4, AdoHcy hydrolase (TC85534, spot 330); 5, Hcy S-methyltransferase (TC89059, no protein spot identified). The *dashed arrows* indicate multiple enzyme steps. *PAPS*, 3'-phosphoadenylyl sulfate; *GSH*, glutathione. Detailed information about the transcripts and protein spots is available in Supplemental Tables 1 and 3.

DISCUSSION

Our study has provided a comprehensive dataset that documents the dynamics of the proteome and transcriptome during seed development and in discrete seed tissues, and the relationship between mRNA and protein patterns from early stages of seed filling to desiccation. The results showed

that the *M. truncatula* seed is subject to many proteome and transcriptome changes during its development. The abundance of about 77% of the protein spots and 50% of the transcripts detected in seeds varied more than 2-fold during the time course. In Arabidopsis seeds, a lower proportion of transcripts (35%) was at least 2-fold regulated in the same period (12). This suggests a more complex network of gene expression during the development of legume seeds. Interestingly, the most highly regulated Arabidopsis genes tended to be preferentially expressed in seeds compared with other plant organs (12), suggesting a more specialized function in seeds for those genes. Genes involved in carbohydrate metabolism and glycolysis are among the most regulated in both Arabidopsis and *M. truncatula* seeds and follow the same expression profiles. For example, in both species, genes involved in glycolysis are differentially expressed (groups 3, 4, 5, 6, and 9, Fig. 2) and those involved in starch synthesis (e.g. starch synthase in group 2, Fig. 2) are transiently expressed during seed filling. In young seeds, starch transiently accumulates and later disappears, probably to provide carbon skeletons for the synthesis of other reserve compounds (12).

An original finding raised by our study is that amino acid metabolism genes are among the most highly regulated in the *M. truncatula* seed. By specifying the distribution of the proteins and transcripts in the seed coat, endosperm, and embryo at the switch toward storage functions, we observed a compartmentalization of Met metabolism between seed tissues that is of agronomic interest. These results are discussed hereafter in light of the proteome-transcriptome comparison. Here, we relied on complementary information derived from our transcriptome dataset, for example, 1) to specify whether differential expression between tissues is specific to Met biosynthetic enzymes or holds more generally for other amino acids, and 2) to identify candidate genes with possible roles in the transfer of nutrients/amino acids between the seed compartments.

The Significance of the Complex Compartmentalization of Amino Acid Metabolic Pathways between the Seed Tissues: Example of Met Metabolism—The most novel findings of the study relate to the compartmentalization of metabolic pathways in component tissues of the seed, in particular those concerned with the biosynthesis of certain amino acids. A differential expression of enzymes involved in Met metabolism between seed tissues has been revealed, reflecting a partitioning of biosynthetic metabolic pathways. A Met synthase (spot 431), producing Met from homocysteine (Hcy), and its transcript (TC85287) were preferentially accumulated within the seed coat (Fig. 6). Protein and RNA abundance both decreased in linear correlation ($r = 0.95$, $p = 0.01$) as seed filling progressed (group 3 in Figs. 2 and 6). Another Met synthase detected at the protein level in the seed coat and endosperm (spot 9, Fig. 3) also decreased during this period. This implies a decrease in Met synthesis from Hcy in the tissues surrounding the embryo during seed filling. Other

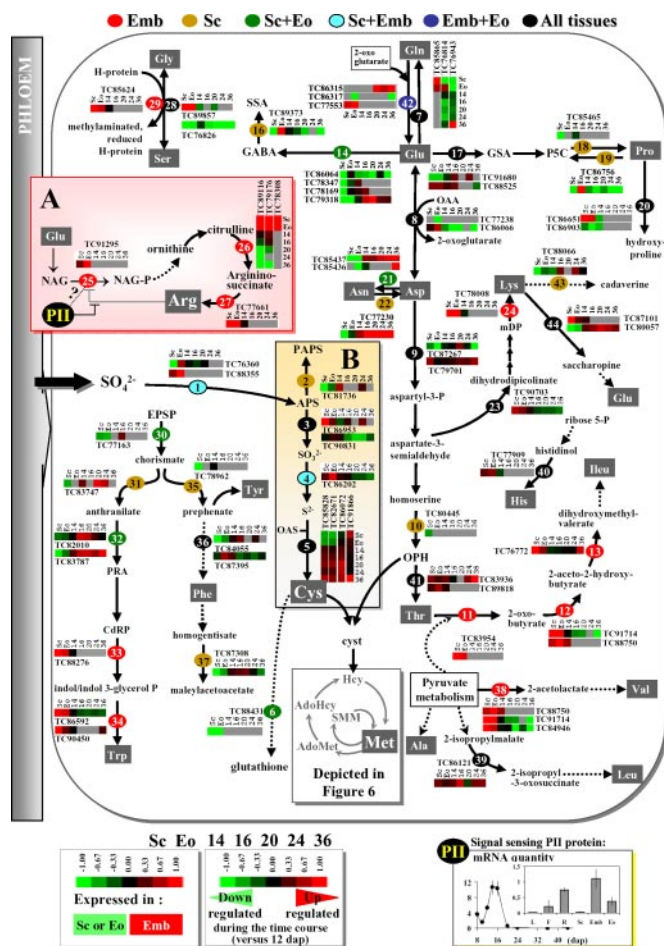


FIG. 7. Global view of the compartmentalization of amino acid biosynthetic pathways between seed tissues. The Arg and sulfur metabolisms are indicated in *A* and *B*, respectively. For each gene, defined by the TIGR MtGI TC-ID, the corresponding *bar* of a heat map obtained using Cluster 3.0/Java TreeView is shown based on Mt16kOLI1 microarray data. Log₂ ratios ≥1 (≥2-fold, *p* ≤ 0.05, *red*) indicate genes significantly up-regulated during the time course (seeds from 14 to 36 dap compared with the 12 dap stage) or preferentially expressed in embryo compared with the surrounding tissues (Sc or Eo). Values ≤−1 (*green*) indicate genes significantly down-regulated during the time course or preferentially expressed in the seed coat or endosperm. From *left to right*, the *bar* represents gene expression in embryo *versus* Sc, embryo *versus* Eo, and seeds at 14 dap *versus* 12 dap (14), 16 dap *versus* 12 dap (16), 20 dap *versus* 12 dap (20), 24 dap *versus* 12 dap (24), and 36 dap *versus* 12 dap (36). For ease of visualization, *colors* on the *arrows* indicate the seed tissue(s) where the genes were preferentially expressed (the *color code* is given within the figure). The *dashed arrows* indicate multiple enzyme steps. Some enzymes, including Cys synthase (72), were detected under several isoforms, which are located in different organelles. The organelles were not depicted for ease of visualization. The gene expression profile for PII (TC88736) was obtained by qRT-PCR during seed development (from 8 dap to the mature stage) and in the seed and plant tissues (*bars*). The mRNA quantity was calculated with respect to the constitutively expressed *msc27* gene, and the mean values (± S.D.) of two repeated experiments are presented. Substrates: AASA, α-amino adipate semialdehyde; *CdRP*, 1-(2-carboxyphenylamino)-1-deoxy-D-ribose-5-phosphate; *Cyst*, cystathionine; *EPSP*, 5-enolpyruvylshikimate-3-phosphate; *GABA*, γ-ami-

genes related to Met metabolism were also preferentially expressed within the seed coat, in particular a gene encoding Hcy S-methyltransferase that catalyzes the irreversible conversion of SMM to Met (Fig. 6). The transcript increases strikingly in relative abundance at 20 dap, parallel to the massive accumulation of the major storage proteins (Fig. 1B), and decreases subsequently as the embryo prepares for quiescence. The predominance of the Hcy S-methyltransferase transcript over those encoding Met synthase suggests that the production of free Met at this stage is from SMM rather than from the pathway involving Met synthase.

SMM, a transport and storage form of Met unique to plants (56), may be imported directly from the phloem, or it could be synthesized within the seed by a methyl transfer from AdoMet to Met via the action of AdoMet:Met S-methyltransferase (SMM cycle). The AdoMet precursor, needed for the SMM cycle to be active, is synthesized from Met by the action of AdoMet synthetase. The dramatic decrease during seed filling of the AdoMet synthetase transcripts TC89089 (endosperm-located) and TC85200 (endosperm- and seed coat-located) and of their corresponding proteins (group 3 in Figs. 2 and 6), together with a report of high concentrations of SMM in the phloem (6), suggest that seed SMM is mainly derived from the phloem rather than from the SMM cycle intraseed and that Met synthesis occurs within the seed coat from this pool of SMM, thus providing free Met for transfer to the embryo and incorporation into storage proteins. This may avoid depletion of the embryo Cys pool otherwise sourced for Met synthesis.

Met synthase and AdoMet synthetase (group 3, Fig. 2) are fundamental in controlling the transition from a quiescent to a highly active metabolic state during germination (51). The disappearance of these enzymes during seed development

nobutyrate; *GSA*, glutamic-γ-semialdehyde; *MDP*, meso-2,6-diaminoheptanedioate; *NAG(-P)*, N-acetyl-L-glutamate (phosphate); *OAS*, O-acetylserine (activated serine); *OAA*, oxaloacetate; *OPH*, O-phosphohomoserine; *PAPS*, 3'-phosphoadenylyl sulfate; *P5C*, pyrroline-5-carboxylate; *PRA*, 5-phosphoribosyl-anthranilate; *SSA*, succinic semialdehyde. Enzymes: 1, ATP sulfurylase; 2, APS kinase; 3, adenosine 5'-phosphosulfate reductase; 4, sulfite reductase; 5, cysteine synthase; 6, glutathione synthetase; 7, glutamine synthetase; 8, aspartate transaminase; 9, aspartate kinase; 10, homoserine kinase; 11, threonine dehydratase; 12, acetolactate synthase; 13, ketol-acid reductase; 14, glutamate decarboxylase; 16, GABA transaminase; 17, pyrroline-5-carboxylate synthase; 18, pyrroline-5-carboxylate reductase; 19, proline dehydrogenase; 20, prollyl 4-hydroxylase; 21, asparagine synthase; 22, L-asparaginase; 23, dihydrodipicolinate synthase; 24, diaminopimelate decarboxylase; 25, acetylglutamate kinase; 26, argininosuccinate synthase; 27, argininosuccinate lyase; 28, serine hydroxymethyltransferase/glycine hydroxymethyltransferase; 29, glycine dehydrogenase; 30, chorismate synthase; 31, anthranilate synthase; 32, phosphoribosylanthranilate transferase; 33, indole 3-glycerol phosphate synthase; 34, Trp synthase; 35, chorismate mutase; 36, prephenate dehydratase; 37, homogentisicase; 38, acetolactate synthase; 39, 3-isopropylmalate dehydratase; 40, histidinol dehydrogenase; 41, threonine synthase; 42, glutamate synthase; 43, lysine decarboxylase; 44, lysine 2-oxoglutarate reductase.

coincides with previous data obtained from the entire seed or isolated filial tissue (9–11, 17, 57). Their restriction in the seed coat and endosperm to the onset of seed filling reflects a metabolic shift in these tissues from a highly active to a quiescent state as the embryo accumulates the storage compounds. Such a metabolic shutdown has been inferred for other types of metabolism (3, 11). Further evidence for a reduction in metabolic activities is the sharp decrease of endosperm-located glycolytic enzymes (e.g. phosphoglycerate kinase, phosphoglyceromutase, and aldolase), and of their transcripts during seed filling (group 3, Fig. 2).

We further analyzed the proteins related to Met metabolism that were specifically accumulated in embryo cells. Among them was an AdoHcy hydrolase (Fig. 3A). This enzyme converts AdoHcy, resulting from AdoMet-dependent methylation of nucleic acids, proteins, lipids, and other metabolites to Hcy (58, 59). Since AdoHcy produced during the activated methyl cycle is a potent competitive inhibitor of methyltransferases that are essential for cell growth and development, AdoHcy hydrolase (i.e. depletion of the AdoHcy pool) may be crucial to ensure transmethylation in embryo cells during seed filling. In support of this hypothesis, the AdoHcy hydrolase level remained high up to 24 dap and then decreased as the seed entered the desiccation phase (Fig. 6). Furthermore, transcripts of several genes encoding AdoMet-dependent methyltransferases were detected in embryo cells (e.g. TC79186 and TC86500 in Fig. 6). Interestingly, in addition to the seed coat and endosperm-located Met synthase, another isoform of Met synthase (TC89672) was expressed with a similar profile to AdoHcy hydrolase (group 7, Fig. 2). Although the corresponding transcript was present in all three seed tissues, the protein was only detected in the endosperm and to a higher extent in the embryo where it may regenerate Met from Hcy produced in the course of the activated methyl cycle (Fig. 6). Because this Met synthase isoform is highly expressed up to 20 dap, it may contribute to Met synthesis up to this stage. Moreover, the observation that AdoHcy hydrolase and Met synthase transcripts persist at the end of seed development is consistent with the finding that both the *de novo* Met biosynthetic pathway and the activated methyl cycle operate very early during seed germination (59).

Based on these data, a model for Met synthesis in developing seeds is proposed (Fig. 6), in which Met is synthesized within the embryo from Hcy produced in the course of the activated methyl cycle up to mid-stages of seed filling (20 dap). To meet the high demand for protein synthesis during seed filling, Met is also synthesized in the seed coat from the pool of SMM and subsequently transported to the embryo. In support of this model, Hcy S-methyltransferase, which synthesizes Met from SMM, showed a higher activity than the SMM cycle enzyme AdoMet:Met S-methyltransferase in developing *Arabidopsis* seeds, and the incorporation of ^{35}S into protein was the same whether $[^{35}\text{S}]\text{Met}$ or $[^{35}\text{S}]\text{SMM}$ was supplied (60).

Compartmentalization of Other Amino Acid Biosynthetic Pathways—To investigate whether differential expression between tissues is specific to Met biosynthetic enzymes or holds more generally for other amino acids, we have further used information from the transcriptomics dataset (Fig. 7). Genes encoding asparagine synthase and L-asparaginase, involved in the interconversion of asparagine/aspartate, are mainly expressed in the seed coat and the endosperm. This suggests that these tissues are metabolizing asparagine, which is one of the principal amino acids delivered by the phloem, before translocation to the embryo (Fig. 6). These genes are also preferentially expressed in the pea seed coat (2). The seed coat is also the site of expression of the gene encoding glutamate decarboxylase (14 in Fig. 7; see also Fig. 6) that converts glutamate to γ -aminobutyrate, which has been associated with various physiological responses, including regulation of carbon flux into the tricarboxylic acid cycle (61). In contrast, a number of genes are apparently expressed preferentially in embryo cells, such as those encoding enzymes of tryptophan, valine, isoleucine, and arginine biosynthesis. Significantly, all of the detected genes encoding enzymes of the Arg biosynthetic pathway were preferentially expressed in embryo cells. Recently, it has been shown that the key regulatory enzyme of Arg synthesis, *N*-acetyl glutamate kinase (25 in Fig. 7A), interacts with the PII protein to relieve inhibition of the kinase activity by Arg (62). Here, we show that the *pii* gene (TC88736) is highly expressed in embryo cells at 14 and 16 dap, concomitantly with storage protein synthesis (Fig. 7), and may therefore be required for a rate of Arg synthesis sufficient for incorporation into storage proteins, which contain ~10% Arg residues.

Another finding was that the seed compartments accommodate two distinct pathways of sulfur assimilation. Transcripts for the first enzyme of sulfate reduction (ATP sulfurylase, 1 in Fig. 7) were detected in both the seed coat (TC76360) and the embryo (TC88355) at the onset of seed filling. This enzyme catalyzes the formation of 5'-adenylyl sulfate (APS) from SO_4^{2-} , which can be utilized in two pathways. In the first pathway, APS is converted by the action of APS kinase (2) to 3'-phosphoadenylyl sulfate, a sulfur donor for the synthesis of defense-related secondary metabolites, including glucosinolates (63). In the second pathway, APS is reduced to provide sulfide ions for Cys synthesis, through two reactions catalyzed by adenosine 5'-phosphosulfate reductase (3) and sulfite reductase (4), respectively (Fig. 7B). The genes encoding enzymes involved in APS reduction and sulfide incorporation into Cys (Cys synthase, 5) are expressed in both the seed coat and the embryo, whereas the genes encoding APS kinase (2) and glutathione synthetase (6) are expressed specifically within the seed coat and the endosperm. These findings suggest that sulfate in the tissues surrounding the embryo is mainly incorporated into glutathione and defense-related secondary metabolites, whereas most of the sulfate entering the embryo is utilized for the

synthesis of Cys. This compartmentalization, along with that observed for the Met biosynthetic pathways, may avoid a depletion of the embryo Cys pool in favor of its incorporation into proteins during seed filling (Fig. 6). The partitioning of these metabolic pathways provides new directions for genetic improvement of grain legumes, low in the essential sulfur-rich amino acids, to enhance value for human consumption and animal feed.

An Intraseed Nitrogen Remobilization May Contribute to Storage Protein Accumulation—Among the metabolic events occurring at the switch to seed filling is protein turnover in the seed coat and endosperm, which releases amino acids for the developing embryo. It has been shown that whole seeds cultured *in vitro* in the absence of exogenous nitrogen are able to accumulate embryo storage proteins by recycling nitrogenous compounds from the embryo surrounding tissues (64), whereas isolated embryos are not. Several proteases, expressed preferentially either in the endosperm (aspartic protease 1, oligopeptidase A, and serine proteases) or in the seed coat (20S proteasome subunit, leucine aminopeptidase and serine protease) (Supplemental Table 3), are good candidates for involvement in this process of remobilization (Fig. 6). Three endosperm-associated subtilisin-like spots (75, 132, and 133, Figs. 3B and 4) matched one gene sequence accessible in the BAC clone mth2-12j18 (GenBank AC146561.9). The predicted molecular masses and pI values of the polypeptides correspond with co- and post-translational modifications of a prepro-enzyme (65). The transcript and protein spots peaked in the prestorage phase and were not detected in other tissues (Fig. 4). In contrast, a seed coat-associated subtilisin increased from the onset of seed filling (14 dap) and reached a maximum level at 20 dap when endosperm reserves were consumed (spot 131 in Fig. 3B, Fig. 4A, group 2 in Fig. 2). The subtilisins detected in the endosperm and in seed coat are encoded by different genes and display contrasted protein patterns during seed development, suggesting that they may play distinct roles in each tissue. Moreover, the linear correlation between transcript and protein levels ($r = 0.9$, $p < 0.05$, Supplemental Table 3) suggests that the amount of subtilisin-like serine proteases in developing seeds is controlled by transcript abundance. The profile of expression of the seed coat-associated subtilisin is identical to that of a cell wall-related pectin acetyltransferase expressed in the same tissue (spot 593, group 2 in Fig. 2). As there are reports of matrix-associated plant subtilisins (66–68), a role in cell wall modification is possible.

Transporters Implied in Nutrient Import and Intraseed Translocations—The exchange of metabolites between seed tissue compartments implies a need for the corresponding transporters. The transcriptome study revealed more than 90 genes encoding transporters potentially involved in the uptake of nutrients or in their translocation within the seed (Fig. 5). Many of them were preferentially expressed in the embryo surrounding tissues and are thus probably involved in the

uptake of nutrients from the phloem or in the efflux of endogenous nutrients en route to the embryo. For example, among the 17 identified genes encoding amino acid transporters, 10 were preferentially expressed in seed coat and/or endosperm, such as the seed coat-derived amino acid transporter TC-ID #78342. One transcript (TC90046) encodes a mitochondrial carrier protein with high sequence similarity (72%) to the Arabidopsis transporter SAMC1, proposed to catalyze the uptake of AdoMet in exchange for AdoHcy to regulate methyltransferase activities in mitochondria and plastids (69). This gene was preferentially expressed in seed coat and endosperm and reaches a maximum at 20 dap, presumably to transport the remaining pool of AdoMet needed for methyltransferase activities (Fig. 6). In addition, we identified five sulfate transporters preferentially expressed in the embryo (TC86732, TC83848, and TC92748) or in the seed coat (TC93814 and TC78211). TC-ID #78211 is preferentially expressed in developing seeds compared with other plant organs and at much higher levels during the prestorage phase (Fig. 5B). The protein encoded shows 78% sequence similarity to the Arabidopsis transporter SULTR3;5, which facilitates sulfate transport in the vascular system (70). The relative contribution of sulfate and SMM import to seed sulfur metabolism merits further investigation.

Conclusions—In addition to documenting the relationship between mRNA and protein patterns from early stages of seed filling to desiccation, this study has revealed a specialization of the filial and maternal tissues, in particular regarding Met and sulfur metabolism. This specialization may be not only controlled by the tissue-specific genetic programs of either maternal (seed coat) or zygotic (embryo), but also by the energy and oxygen status of the different tissues. Because of their location, the embryos grow in an environment of low light and oxygen availability (71), which may affect ATP production and biosynthetic activities. As a way to control biosynthetic fluxes, embryos become photosynthetically active during maturation (5, 12). In developing seeds of soybean, photosynthesis is followed by an increase in ATP levels more prominently within the inner regions of the embryo where biosynthetic activities occur. Several proteins from both photosynthetic light reactions (oxygen evolving enhancer and chlorophyll *a/b*-binding proteins) and dark reactions (RuBisCO and RuBisCO subunit binding proteins) were detected in the *M. truncatula* developing embryo (Fig. 2 and Supplemental Table 3), suggesting that photosynthesis also provides oxygen and increases the energy state at the switch toward storage functions to regulate biosynthetic activities.

Acknowledgments—We thank G. Aubert and J. Verdier (Unité Mixte de Recherche Génétiques et Ecophysiologie of Grain Legumes (UMRLEG)-INRA/ENESAD) for their very valuable technical support with the qRT-PCR experiments and B. Valot (Platform for Proteomics, UMR de Génétique végétale INRA/Université Paris Sud/CNRS/INA PG, Ferme du Moulon, Gif sur Yvette) for help with the data obtained using the ion trap mass spectrometer, C. Le Signor (UMRLEG) for

advice and help with statistical analysis of the proteome data, F. Moussy (UMRLEG) for useful assistance with plant growth, and J. Burstin (UMRLEG) for critical reading of the manuscript and helpful discussions.

* This work was supported by the FP6 EU project "Grain Legumes" (FOOD-CT-2004-506223), by a travel grant from the COST action number 843 (COST-STSM-843-00091), and by the International NRW Graduate School in Bioinformatics and Genome Research (to H. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

** Present address: Centre d'Analyses Protéomiques de Marseille, IFR Jean Roche, Faculté de médecine, 51 bd Pierre Dramard, 13916 Marseille, France.

§ To whom correspondence should be addressed: UMR102 Genetics and Ecophysiology of Grain Legumes, INRA/ENESAD, Domaine d'Epoisses, 21110 Bretenières, France. Tel.: 33-3-80693391; Fax: 33-3-80693263; E-mail: gallardo@epoisses.inra.fr.

REFERENCES

1. Garcia, D., Fitz Gerald, J. N., and Berger, F. (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* **17**, 52–60
2. Bewley, J. D., and Black, M. (1994) *Seeds: Physiology of Development and Germination*, pp. 51–100, Plenum Press, New York
3. Buitink, J., Leger, J. J., Guisle, I., Vu, B. L., Wuillème, S., Lamirault, G., Bars, A. L., Meur, N. L., Becker, A., Küster, H., and Leprince, O. (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. *Plant J.* **47**, 735–750
4. Bethke, P. C., Libourel, I. G., Aoyama, N., Chung, Y. Y., Still, D. W., and Jones, R. L. (2007) The Arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. *Plant Physiol.* **143**, 1173–1188
5. Weber, H., Borisjuk, L., and Wobus, U. (2005) Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* **56**, 253–279
6. Bourgis, F., Roje, S., Nuccio, M. L., Fisher, D. B., Tarczynski, M. C., Li, C., Herschbach, C., Rennenberg, H., Pimenta, M. J., Shen, T. L., Gage, D. A., and Hanson, A. D. (1999) S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell* **11**, 1485–1498
7. Tabe, L. M., and Droux, M. (2001) Sulfur assimilation in developing lupin cotyledons could contribute significantly to the accumulation of organic sulfur reserves in the seed. *Plant Physiol.* **126**, 176–187
8. Higgins, T. J., Chandler, P. M., Randall, P. J., Spencer, D., Beach, L. R., Blagrove, R. J., Kortt, A. A., and Inglis, A. S. (1986) Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. *J. Biol. Chem.* **261**, 11124–11130
9. Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R. D., and Burstin, J. (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **133**, 664–682
10. Firnhaber, C., Pühler, A., and Küster, H. (2005) EST sequencing and time course microarray hybridizations identify more than 700 *Medicago truncatula* genes with developmental expression regulation in flowers and pods. *Planta* **222**, 269–283
11. Hajdich, M., Ganapathy, A., Stein, J. W., and Thelen, J. J. (2005) A systematic proteomic study of seed filling in soybean: establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiol.* **137**, 1397–1419
12. Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* **14**, 1191–1206
13. Finnie, C., Maeda, K., Ostergaard, O., Bak-Jensen, K. S., Larsen, J., and Svensson, B. (2004) Aspects of the barley seed proteome during development and germination. *Biochem. Soc. Trans.* **32**, 517–519
14. Zhang, H., Sreenivasulu, N., Weschke, W., Stein, N., Rudd, S., Radchuk, V., Potokina, E., Scholz, U., Schweizer, P., Zierold, U., Langridge, P., Varshney, R. K., Wobus, U., and Graner, A. (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J.* **40**, 276–290
15. Drea, S., Leader, D. J., Arnold, B.C., Shaw, P., Dolan, L., and Doonan, J. H. (2005) Systematic spatial analysis of gene expression during wheat caryopsis development. *Plant Cell* **17**, 2172–2185
16. Grimaneli, D., Perotti, E., Ramirez, J., and Leblanc, O. (2005) Timing of the maternal-to-zygotic transition during early seed development in maize. *Plant Cell* **17**, 1061–1072
17. Méchin, V., Thévenot, C., Le Guilloux, M., Prioul, J. L., and Damerval, C. (2007) Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant Physiol.* **143**, 1203–1219
18. De Nobel, H., Lawrie, L., Brul, S., Klis, F., Davis, M., Alloush, H., and Coote, P. (2001) Parallel and comparative analysis of the proteome and transcriptome of sorbic acid-stressed *Saccharomyces cerevisiae*. *Yeast* **18**, 1413–1428
19. Kolkman, A., Olsthoorn, M. M., Heeremans, C. E., Heck, A. J., and Slijper, M. (2005) Comparative proteome analysis of *Saccharomyces cerevisiae* grown in chemostat cultures limited for glucose or ethanol. *Mol. Cell. Proteomics* **4**, 1–11
20. Li, L., Li, Q., Rohlin, L., Kim, U., Salmon, K., Rejtar, T., Gunsalus, R. P., Karger, B. L., and Ferry, J. G. (2007) Quantitative proteomic and microarray analysis of the archaeon *Methanosarcina acetivorans* grown with acetate versus methanol. *J. Proteome Res.* **6**, 759–771
21. Resch, A., Leicht, S., Saric, M., Pasztor, L., Jakob, A., Götz, F., and Nordheim, A. (2006) Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling. *Proteomics* **6**, 1867–1877
22. Unwin, R. D., and Whetton, A. D. (2006) Systematic proteome and transcriptome analysis of stem cell populations. *Cell Cycle* **5**, 15
23. Wagner, W., Feldmann, R. E. Jr., Seckinger, A., Maurer, M. H., Wein, F., Blake, J., Krause, U., Kalenka, A., Burgers, H. F., Saffrich, R., Wuchter, P., Kuschinsky, W., and Ho, A. D. (2006) The heterogeneity of human mesenchymal stem cell preparations: evidence from simultaneous analysis of proteomes and transcriptomes. *Exp. Hematol.* **34**, 536–548
24. Holmes-Davis, R., Tanaka, C. K., Vensel, W. H., Hurkman, W. J., and McCormick, S. (2005) Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics* **5**, 4864–4884
25. Mooney, B. P., Miernyk, J. A., Greenlief, C. M., and Thelen, J. J. (2006) Using quantitative proteomics of Arabidopsis roots and leaves to predict metabolic activity. *Physiol. Plant.* **128**, 237–250
26. Choi, H. K., Mun, J. H., Kim, D. J., Zhu, H., Baek, J. M., Mudge, J., Roe, B., Ellis, N., Doyle, J., Kiss, G. B., Young, N. D., and Cook, D. R. (2004) Estimating genome conservation between crop and model legume species. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15289–15294
27. Djemel, N., Guedon, D., Lechevalier, A., Salon, C., Miquel, M., Prosperi, J. M., Rochat, C., and Boutin, J. P. (2005) Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol. Biochem.* **43**, 557–566
28. Sheoran, I. S., Olson, D. J., Ross, A. R., and Sawhney, V. K. (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. *Proteomics* **5**, 3752–3764
29. Sreenivasulu, N., Altschmied, L., Panitz, R., Hähnel, U., Michalek, W., Weschke, W., and Wobus, U. (2002) Identification of genes specifically expressed in maternal and filial tissues of barley caryopses: a cDNA array analysis. *Mol. Genet. Genomics* **266**, 758–767
30. Finnie, C., and Svensson, B. (2003) Feasibility study of a tissue-specific approach to barley proteome analysis: aleurone layer, endosperm, embryo and single seeds. *J. Cereal Sci.* **38**, 217–227
31. Sreenivasulu, N., Altschmied, L., Radchuk, V., Gubatz, S., Wobus, U., and Weschke, W. (2004) Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *Plant J.* **37**, 539–553
32. Hohnjec, N., Vieweg, M. F., Pühler, A., Becker, A., and Küster, H. (2005) Overlaps in the transcriptional profiles of *Medicago truncatula* roots inoculated with two different *Glomus* fungi provide insights into the genetic program activated during arbuscular mycorrhiza. *Plant Physiol.* **137**, 1283–1301

33. Verwoerd, T., Dekker, B., and Hoekema, A. (1989) A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Res.* **17**, 2362
34. Küster, H., Hohnjec, N., Krajinski, F., El-Yahyaoui, F., Manthey, K., Gouzy, J., Dondrup, M., Meyer, F., Kalinowski, J., Brechenmacher, L., van Tuinen, D., Gianinazzi-Pearson, V., Pühler, A., Gamas, P., and Becker, A. (2004) Construction and validation of cDNA-based Mt6k-RIT macro- and microarrays to explore root endosymbioses in the model legume *Medicago truncatula*. *J. Biotechnol.* **108**, 95–113
35. Heim, U., Weber, H., Bäumllein, H., and Wobus, U. (1993) A sucrose-synthase gene of *Vicia faba* L.: expression pattern in developing seeds in relation to starch synthesis and metabolic regulation. *Planta* **191**, 394–401
36. Dondrup, M., Goesmann, A., Bartels, D., Kalinowski, J., Krause, L., Linke, B., Rupp, O., Sczyrba, A., Pühler, A., and Meyer, F. (2003) EMMA: a platform for consistent storage and efficient analysis of microarray data. *J. Biotechnol.* **106**, 135–146
37. Dudoit, S., Yang, Y., Callow, M., and Speed, T. (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **12**, 111–139
38. Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O. E., Palacios-Rojas, N., Selbig, J., Hannemann, J., Piques, M. C., Steinhäuser, D., Scheible, W. R., Gibon, Y., Morcuende, R., Weicht, D., Meyer, S., and Stitt, M. (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol.* **138**, 1195–1204
39. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14863–14868
40. Chang, S., Puryear, J., and Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant. Mol. Biol. Rep.* **11**, 113–116
41. Bouton, S., Viau, L., Lelievre, E., and Limami, A. (2005) A gene encoding a protein with a proline-rich domain (MtPPRD1), revealed by suppressive subtractive hybridization (SSH), is specifically expressed in the *Medicago truncatula* embryo axis during germination. *J. Exp. Bot.* **56**, 825–832
42. Nagaraj, S., Lei, Z., Watson, B., Sumner, L., Gallardo K., Dumas-Gaudot, E., Recorbet, G., Robert, F., Thiery, O., Valot, B., Mathesius, U., and Triplett, E. (2006) Proteomics, In: *The Medicago truncatula handbook* (Mathesius, U., Journet, E. P., and Sumner, L. W., eds) Ardmore, OK
43. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254
44. Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J., and Job, D. (2002) Proteomics of Arabidopsis seed germination: a comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol.* **129**, 823–837
45. 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
46. SAS Institute (1999) *SAS/STAT User's Guide*. SAS Institute Inc., Cary, NC
47. Harris, W. A., Janecki, D. J., and Reilly, J. P. (2002) Use of matrix clusters and trypsin autolysis fragments as mass calibrants in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **16**, 1714–1722
48. Ding, Q., Xiao, L., Xiong, S., Jia, Y., Que, H., Guo, Y., and Liu, S. (2003) Unmatched masses in peptide mass fingerprints caused by cross-contamination: an updated statistical result. *Proteomics* **3**, 1313–1317
49. Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004) The effect of alpha-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* **134**, 1598–1613
50. Cahoon, A. B., Cunningham, K. A., and Stern, D. B. (2003) The plastid *clpP* gene may not be essential for plant cell viability. *Plant Cell Physiol.* **44**, 93–95
51. Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J., and Job, D. (2002) Importance of methionine biosynthesis for Arabidopsis seed germination and seedling growth. *Physiol. Plant.* **116**, 238–247
52. Rajjou, L., Gallardo, K., Job, C., and Job, D. (2006) *Proteome Analysis for the Study of Developmental Processes in Plants* (Finnie, C., ed), pp. 151–184, Blackwell Publishing, Oxford, UK
53. Fan, J., Yang, X., Wang, W., Wood, W. H., Becker, K. G., and Gorospe, M. (2002) Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10611–10616
54. Edwards, M. E., Choo, T. S., Dickson, C. A., Scott, C., Gidley, M. J., and Reid, J. S. G. (2004) The seeds of *Lotus japonicus* lines transformed with sense, antisense, and sense/antisense galactomannan galactosyltransferase constructs have structurally altered galactomannans in their endosperm cell walls. *Plant Physiol.* **134**, 1153–1162
55. Hakman, I., and Oliviussen, P. (2002) High expression of putative aquaporin genes in cells with transporting and nutritive functions during seed development in Norway spruce (*Picea abies*). *J. Exp. Bot.* **53**, 639–649
56. Mudd, S. H., and Datko, A. H. (1990) The S-methylmethionine cycle in *Lemna paucicostata*. *Plant Physiol.* **93**, 623–630
57. Vensel, W. H., Tanaka, C. K., Cai, N., Wong, J. H., Buchanan, B. B., and Hurkman, W. J. (2005) Developmental changes in the metabolic protein profiles of wheat endosperm. *Proteomics* **5**, 1594–1611
58. Giovanelli, J., Mudd, S. H., and Datko, A. H. (1985) In vivo regulation of *de novo* methionine biosynthesis in a higher plant (*Lemna*). *Plant Physiol.* **77**, 450–455
59. Ravanel, S., Block, M. A., Rippert, P., Jabrin, S., Curien, G., Rebeille, F., and Douce, R. (2004) Methionine metabolism in plants: chloroplasts are autonomous for *de novo* methionine synthesis and can import S-adenosylmethionine from the cytosol. *J. Biol. Chem.* **279**, 22548–22557
60. Ranocha, P., McNeil, S. D., Ziemak, M. J., Li, C., Tarczynski, M. C., and Hanson, A. D. (2001) The S-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity. *Plant J.* **25**, 575–584
61. Bouche, N., Fait, A., Bouchez, D., Moller, S. G., and Fromm, H. (2003) Mitochondrial succinic-semialdehyde dehydrogenase of the gamma-aminobutyrate shunt is required to restrict levels of reactive oxygen intermediates in plants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6843–6848
62. Chen, Y. M., Ferrar, T. S., Lohmeier-Vogel, E., Morrice, N., Mizuno, Y., Berenger, B., Ng, K. K., Muench, D. G., and Moorhead, G. B. (2006) The PII signal transduction protein of *Arabidopsis thaliana* forms an arginine-regulated complex with plastid N-acetyl glutamate kinase. *J. Biol. Chem.* **281**, 5726–5733
63. Leustek, T., and Saito, K. (1999) Sulfate transport and assimilation in plants. *Plant Physiol.* **120**, 637–644
64. Gallardo, K., Kurt, C., Thompson, R., and Ochatt, S. (2006) *In vitro* culture of immature *M. truncatula* grains under conditions permitting embryo development comparable to that observed *in vivo*. *Plant Sci.* **170**, 1052–1058
65. Beilinson, V., Moskalenko, O. V., Livingstone, D. S., Reverdatto, S. V., Jung, R., and Nielsen, N. C. (2002) Two subtilisin-like proteases from soybean. *Physiol. Plant.* **115**, 585–597
66. Batchelor, A. K., Boutillier, K., Miller, S. S., Labbe, H., Bowman, L., Hu, M., Johnson, D. A., Gijzen, M., and Miki, B. L. (2000) The seed coat-specific expression of a subtilisin-like gene, SCS1, from soybean. *Planta* **211**, 484–492
67. Tanaka, H., Onouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., Machida, C., and Machida, Y. (2001) A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* **128**, 4681–4689
68. Hamilton, J. M., Simpson, D. J., Hyman, S. C., Ndimba, B. K., and Slabas, A. R. (2003) Ara12 subtilisin-like protease from *Arabidopsis thaliana*: purification, substrate specificity and tissue localization. *Biochem. J.* **370**, 57–67
69. Palmieri, L., Arrigoni, R., Blanco, E., Carrari, F., Zanor, M. I., Studart-Guimaraes, C., Fernie, A. R., and Palmieri, F. (2006) Molecular identification of an *Arabidopsis thaliana* S-adenosylmethionine transporter: analysis of organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. *Plant Physiol.* **142**, 855–865
70. Buchner, P., Takahashi, H., and Hawkesford, M. J. (2004) Plant sulphate transporters: co-ordination of uptake, intracellular and long-distance transport. *J. Exp. Bot.* **55**, 1765–1773
71. Rolletschek, H., Radchuk, R., Klukas, C., Schreiber, F., Wobus, U., and Borisjuk, L. (2005) Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. *New Phytol.* **167**, 777–786
72. Wirtz, M., and Droux, M. (2005) Synthesis of the sulfur amino acids: cysteine and methionine. *Photosynth. Res.* **86**, 345–362