

The Proteome of Seed Development in the Model Legume *Lotus japonicus*¹[C][W]

Svend Dam, Brian S. Laursen, Jane H. Ørnfelt, Bjarne Jochimsen, Hans Henrik Stærfeldt, Carsten Friis, Kasper Nielsen, Nicolas Goffard, Søren Besenbacher, Lene Krusell, Shusei Sato, Satoshi Tabata, Ida B. Thøgersen, Jan J. Enghild, and Jens Stougaard*

Centre for Carbohydrate Recognition and Signalling (S.D., B.S.L., J.H.Ø., B.J., L.K., J.S.), Center for Insoluble Protein Structures (I.B.T., J.J.E.), and Department of Molecular Biology (S.D., B.S.L., J.H.Ø., B.J., L.K., I.B.T., J.J.E., J.S.), University of Aarhus, DK-8000 Aarhus, Denmark; Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark (H.H.S., C.F., K.N.); Institut Louis Malardé, 98713 Papeete Tahiti, French Polynesia (N.G.); Australian Research Council Centre of Excellence for Integrative Legume Research, Genomic Interactions Group, Research School of Biological Sciences, Australian National University, Canberra, Australian Capital Territory 2601, Australia (N.G.); Bioinformatics Research Center, University of Aarhus, DK-8000 Aarhus, Denmark (S.B.); and Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan (S.S., S.T.)

We have characterized the development of seeds in the model legume *Lotus japonicus*. Like soybean (*Glycine max*) and pea (*Pisum sativum*), *Lotus* develops straight seed pods and each pod contains approximately 20 seeds that reach maturity within 40 days. Histological sections show the characteristic three developmental phases of legume seeds and the presence of embryo, endosperm, and seed coat in desiccated seeds. Furthermore, protein, oil, starch, phytic acid, and ash contents were determined, and this indicates that the composition of mature *Lotus* seed is more similar to soybean than to pea. In a first attempt to determine the seed proteome, both a two-dimensional polyacrylamide gel electrophoresis approach and a gel-based liquid chromatography-mass spectrometry approach were used. Globulins were analyzed by two-dimensional polyacrylamide gel electrophoresis, and five legumins, LLP1 to LLP5, and two convicilins, LCP1 and LCP2, were identified by matrix-assisted laser desorption ionization quadrupole/time-of-flight mass spectrometry. For two distinct developmental phases, seed filling and desiccation, a gel-based liquid chromatography-mass spectrometry approach was used, and 665 and 181 unique proteins corresponding to gene accession numbers were identified for the two phases, respectively. All of the proteome data, including the experimental data and mass spectrometry spectra peaks, were collected in a database that is available to the scientific community via a Web interface (<http://www.cbs.dtu.dk/cgi-bin/lotus/db.cgi>). This database establishes the basis for relating physiology, biochemistry, and regulation of seed development in *Lotus*. Together with a new Web interface (<http://bioinfoserver.rsbs.anu.edu.au/utis/PathExpress4legumes/>) collecting all protein identifications for *Lotus*, *Medicago*, and soybean seed proteomes, this database is a valuable resource for comparative seed proteomics and pathway analysis within and beyond the legume family.

Legumes constitute the third largest plant family and are an important source of food, feed, and natural compounds of industrial importance. Most legumes possess the ability to form symbiosis with both nitrogen-fixing rhizobia and phosphate-retrieving mycorrhiza, which makes the legumes very useful in pastures and for soil improvement schemes (Graham

and Vance, 2003). The crop legumes account for 27% of the world's primary agricultural crop production, with grain legumes contributing 33% of daily protein nitrogen intake of humans (Graham and Vance, 2003). Legume seed proteins are rich in Lys, complementing the nutritional profiles of cereals and tubers in the diet, but are low in sulfur-containing amino acids (Duranti and Gius, 1997).

Most of the stored nitrogen and sulfur and some of the carbon is found in the protein fraction, mainly in the storage proteins. These proteins accumulate in protein bodies in the cotyledonary parenchyma cells until they undergo hydrolysis upon germination (Tabe et al., 2002). Storage proteins are synthesized on the rough endoplasmic reticulum, and a signal peptide guides transport to the endoplasmic reticulum (zur Nieden et al., 1984). Storage proteins are devoid of catalytic activity and play no structural role in the tissue. Apart from proteins, carbon is stored in lipids

¹ This work was supported by the Danish STF Research Program and the Danish National Research Foundation. S.D. was partly supported by the LOTASSA project.

* Corresponding author; e-mail stougaard@mb.au.dk.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Jens Stougaard (stougaard@mb.au.dk).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.133405

and starch. The levels of protein, lipids, and starch vary between legume species. Soybean (*Glycine max*) and peanut (*Arachis hypogaea*) are oil crops, whereas beans (*Phaseolus vulgaris*), lentils (*Lens culinaris*), peas (*Pisum sativum*), and chickpeas (*Cicer arietinum*) are major staple foods (Domoney et al., 2006).

A typical mature legume seed consist mainly of the embryo. Seed development proceeds through three distinct phases: histodifferentiation, seed filling, and desiccation. Growth of the embryo is characterized by cell division in the histodifferentiation phase and by cell enlargement in the seed-filling phase. Histodifferentiation can be further divided into two phases, where cell division is confined to endosperm and seed coat in the first phase and to the embryo in the second phase (Wang and Hedley, 1993). The endosperm first develops into a syncytium and then cellularizes. The mitotic activity of the seed in the cell division phase has been proposed to be controlled by hormones, environmental factors, and carbon and nitrogen supply (Egli et al., 1989; Munier-Jolain and Ney, 1998; Ozga et al., 2002). During the cell enlargement (seed-filling) phase, seed nitrogen accumulation and protein composition depend on both symbiotic N₂ fixation and nitrogen from the soil. Carbon derives mainly from recent photosynthate rather than from remobilized sources (for review, see Domoney et al., 2006).

Proteins involved in cell division are abundant during early stages of seed development, and the level decreases before the accumulation of the major storage proteins during seed filling (Gallardo et al., 2003).

In legumes, the terms “storage protein” and “globulin” have become synonymous, because the large majority of the storage proteins are soluble in high-salt solution but insoluble in water and as such are globulins according to Osborne’s classical definitions (Chittenden et al., 1908). There are two main groups of globulins. They evolved from a common single-domain ancestor and have sedimentation coefficients of 7S and 11S (Shutov et al., 1995). In pea, 7S and 11S globulins are called vicilins and legumins, respectively, and are often referred to as vicilin- and legumin-like globulins in other legumes. Vicilins are trimeric structures consisting of subunits of approximately 50 kD. Some but not all of the approximately 50-kD polypeptides are proteolytically processed at one or both of two cleavage sites. These cleavage sites are not strictly conserved, but processing occurs at the carboxyl side of an Asn residue. A subset of the processed polypeptides may also be glycosylated. Most vicilins lack Cys, so no disulfide bonds are formed. Convicilins (approximately 70 kD) are homologous to vicilin but contain a polar insert and do not appear to be post-translationally processed besides removal of the signal peptide (Higgins and Spencer, 1981).

The legumins are hexamers composed of subunits of 60 to 80 kD. The legumins consist of disulfide bonded subunits formed by proteolytic cleavage of a single polypeptide precursor at an Asn-Gly or alternatively

an Asn-Phe cleavage site into acidic (approximately 40 kD) and basic (approximately 20 kD) polypeptides (March et al., 1988). In contrast to vicilins, the legumins are rarely glycosylated. For a review of legume seed development, see Weber et al. (2005).

Here, we present data on the seed development and protein composition in the model legume *Lotus japonicus*. Based on the newly released genome sequence, we present a detailed catalogue of proteins present in the seeds at two time points during development and a more detailed analysis of the storage proteins. This work will create the basis for relating physiological development to detailed proteome analysis.

RESULTS

Characterization of the Developing *Lotus* Seed

L. japonicus ecotype Gifu (hereafter referred to as *Lotus*) seeds were collected at 3-d intervals between 7 and 43 d after flowering (DAF; Fig. 1). At each developmental stage, fresh and dry weight of the seeds were measured and water content was calculated in order to define the major phases of seed formation and maturation (Fig. 1). In parallel, seeds were fixed, sectioned,

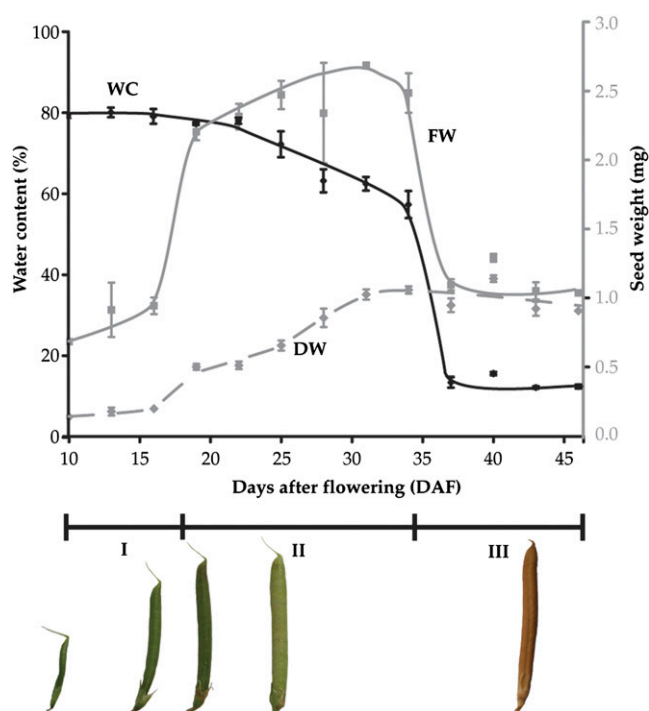


Figure 1. Water content, fresh weight, and dry weight of developing seeds. The graph at top represents the fresh weight (FW; gray line), dry weight (DW; dashed line), and water content (WC; black line) of the developing seeds from 10 to 46 DAF. Error bars represent sd for five replicates. The histodifferentiation (I), seed-filling (II), and desiccation (III) phases are indicated at bottom. Pods representing 7, 16, 19, 25, and 43 DAF are shown.

and stained with toluidine blue to describe the developmental changes at the cellular level (Figs. 2 and 3). The seed weight increased rapidly at 16 to 19 DAF (Fig. 1), and the embryo size also increased dramatically at this stage, indicating the transition from histodifferentiation dominated by cell divisions to seed filling, which is dominated by cell enlargements in the embryo (Fig. 2). The accumulation of storage proteins in the seeds became visible at 25 DAF (Fig. 4). The seed-filling phase ended around 34 DAF, when seeds enter the desiccation phase and dry out (Fig. 1). As expected, the amount of protein in the seeds increased dramatically during seed filling and remained constant in the mature seed (Fig. 4).

At 7 DAF, the embryo in the transition stage occupies only a fraction of the embryo sac, which is filled by the mainly liquid endosperm (Figs. 2 and 3A). At 10 DAF, the embryo is in the heart stage (Figs. 2 and 3B). The embryo is in early torpedo stage at 16 DAF, when the embryo cells are beginning to elongate (Fig. 3C). At 19 DAF, the embryo is in the torpedo stage (Figs. 2 and 3D).

The endosperm has only formed one or two cell layers encircling the embryo sac at 7 DAF (Fig. 3E). Between 7 and 10 DAF, the endosperm cellularizes and fills the embryo sac (Figs. 2 and 3F). The mature seed is high in protein and dominated by the embryo, which takes up most of the embryo sac. In contrast to faba bean (*Vicia faba*) and pea seeds, for example, but like *Medicago*, well-developed endosperm tissue is found in the mature seed (Fig. 3G; Borisjuk et al., 1995, 2002; Abirached-Darmency et al., 2005).

The outer layer of the seed coat consists of one layer of epidermis followed by a layer of hypodermis and several layers of parenchyma cells at 7 DAF (Fig. 3H). At 16 DAF, the epidermis cells are elongated and differentiated almost to macrosclereid cells and hypodermis cells have started to thicken the cell wall (Fig. 3I). At 19 DAF, the seed coat epidermis has differentiated into macrosclereid cells and the hypodermis has differentiated into hourglass-shaped osteosclereid

cells (Fig. 3J). In the seed coat, the macrosclereid cells have become more impenetrable (Ma et al., 2004) and efficiently protect the embryo (Fig. 3K).

Components in Mature *Lotus* Seeds

The main components of legume seeds are starch, protein, and oil. As a prelude to the detailed investigation of proteins involved in seed physiology, biochemistry, and development, the content of these components was determined in mature *Lotus* seed (Table I). In addition, ash and phytic acid contents were determined.

In contrast to peas, in which up to 50% of the dry weight in mature seeds is starch, the mature *Lotus* seeds contained only 0.6% starch. The starch content in *Lotus* is thus more similar to the level found in mature soybean seeds, in which the content varies between 0.19% and 0.91% in different soybean cultivars (Wilson et al., 1978).

The protein content in mature *Lotus* seed is approximately 43% (Table I). For soybean, the protein content is 43.7% (Prakash and Misra, 1988), while the protein level in eight *Medicago* genotypes is between 30% and 40% of the dry weight (Djemel et al., 2005). In contrast, the protein level in pea is between 22% and 26%. Again, *Lotus* is more similar to soybean than to pea.

Lipids constitute approximately 7% of the dry weight of the mature *Lotus* seed. In pea, the lipid content is between 1.4% and 3.3%, while the lipid content in soybean can be higher than 20%. In seven underutilized legumes, the crude lipid content was between 3.77% and 7.04% (Vadivel and Janardhanan, 2000). The fatty acid composition of the lipid fraction was further analyzed (Table II). *Lotus* lipid is composed of approximately 11% C16:0 (palmitate), 4% C18:0 (stearate), and 46% C18:2 (linoleate), which is similar to the values obtained for soybean (Shen et al., 1997). The content of C18:1 (oleate) and C18:3 ω 3 (linolenate) in *Lotus* is approximately 11% and 25%, respectively. For oleate, this is less than half of the level

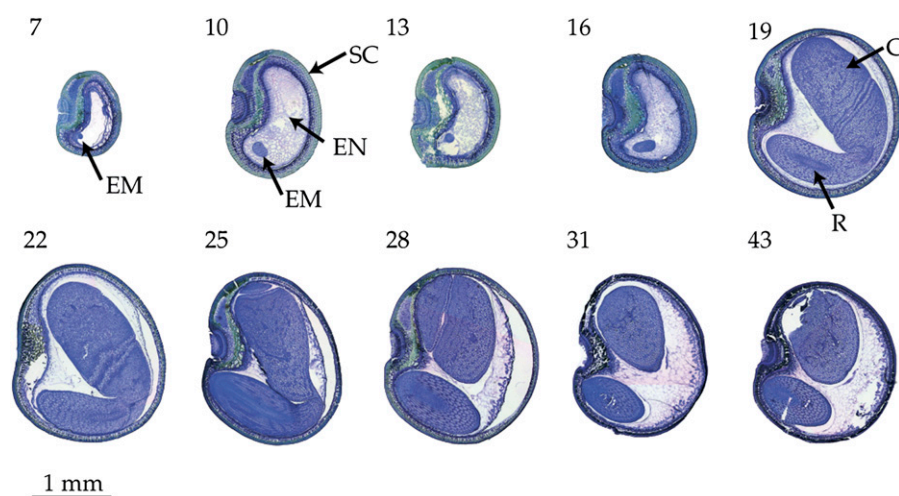


Figure 2. Thin sections of seeds in different developmental stages. The seeds were harvested from 7 to 43 DAF, fixed, sectioned into 5- μ m slices, and stained with toluidine blue. The numbers indicate DAF. The embryo only takes up a small part of the embryo sac at 7 DAF. The endosperm cellularizes between 7 and 10 DAF. Consistent with the increase in fresh weight (Fig. 1), the seed size increases dramatically at 16 to 19 DAF and the embryo expands to occupy most of the seed. This indicates the transition from histodifferentiation to seed filling. At 43 DAF, the seed is mature and desiccated. Some endosperm tissue is still found in the mature seed. C, Cotyledon; EM, embryo; EN, endosperm; R, radicle; SC, seed coat.

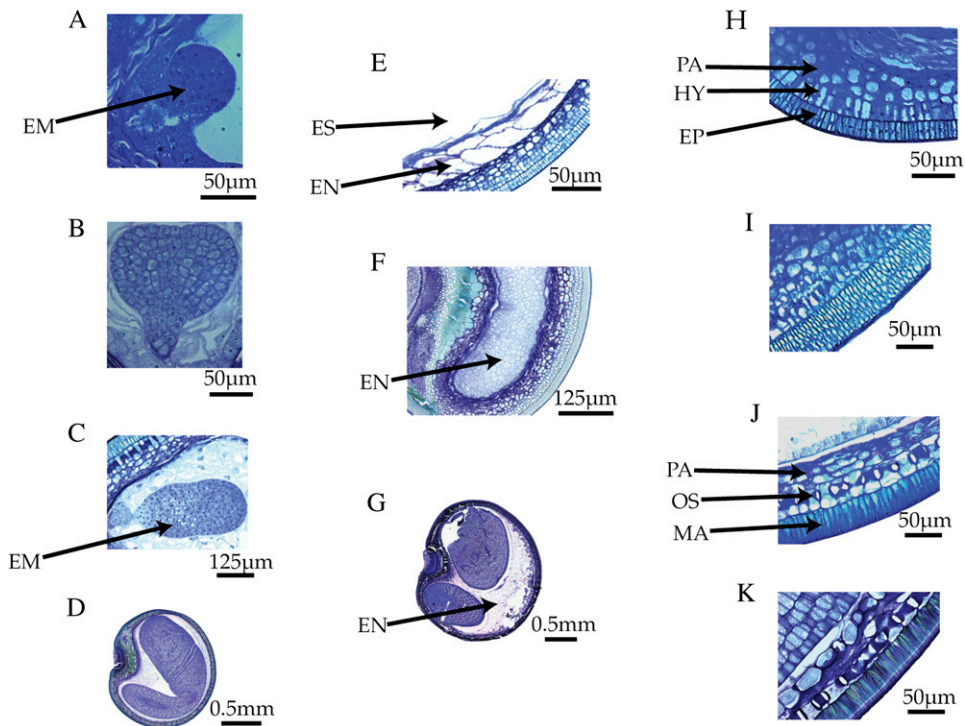


Figure 3. Detailed development of embryo, endosperm, and seed coat. A to D, Thin sections of the embryo at 7, 10, 16, and 22 DAF. At 7 DAF, the embryo is in the transition stage. The embryo (EM) has developed into the heart stage at 10 DAF and close to the torpedo stage at 16 DAF. At 22 DAF, the embryo is in the torpedo stage and the cells are now elongating instead of dividing. E to G, Thin sections showing endosperm tissue at 7, 10, and 43 DAF. The endosperm (EN) has begun to cellularize at 7 DAF and forms one or two cell layers encircling the embryo sac (ES). The endosperm is cellularized at 10 DAF, when it fills the embryo sac. Finally, in the mature seed (43 DAF), the mature embryo occupies most of the embryo sac, but some endosperm tissue is still present. H to K, Thin sections of the seed coat at 7, 16, 19, and 43 DAF. At 7 DAF, the outer layer of the seed coat consists of one layer of epidermis (EP), followed by a layer of hypodermis (HY), and several layers of parenchyma cells (PA). The epidermis cells have elongated and the hypodermis cells have increased their cell wall at 16 DAF. At 19 DAF, the epidermis cells differentiate into macrosclereid (MA) cells and the hypodermis cells differentiate into osteosclereid (OS) cells (hourglass shape). Finally, in the mature seed (43 DAF), the macrosclereid cells have finished differentiation.

in soybean and pea, while the linolenate is more than three times higher in *Lotus*. Compared with *Medicago*, the stearate and linoleate levels are higher in *Lotus*, while the palmitate and oleate levels are lower (Djemel et al., 2005). The relative composition of saturated and unsaturated fatty acid is an important characteristic of

edible oil. Oil with high levels of polyunsaturated fatty acids has nutritional benefits but decreased stability (Liu and White, 1992). Therefore, *Lotus* is a useful model for studying the genetic regulation of pathways responsible for fatty acid synthesis, degradation, and modification, such as desaturation of stearic acid to

Figure 4. Protein profile of seed development. A 5% to 15% gradient SDS-PAGE scan of protein extracted from seeds is shown. The numbers above each lane indicate DAF. Each lane contains protein extracted from 1 mg of seeds (dry weight). Globulins were selectively extracted from mature seeds (43 DAF) and loaded in lane G. The bands that start to show up at 25 DAF and become highly dominant at 43 DAF are in the globulin fraction; indeed, they have been identified as globulins (data not shown).

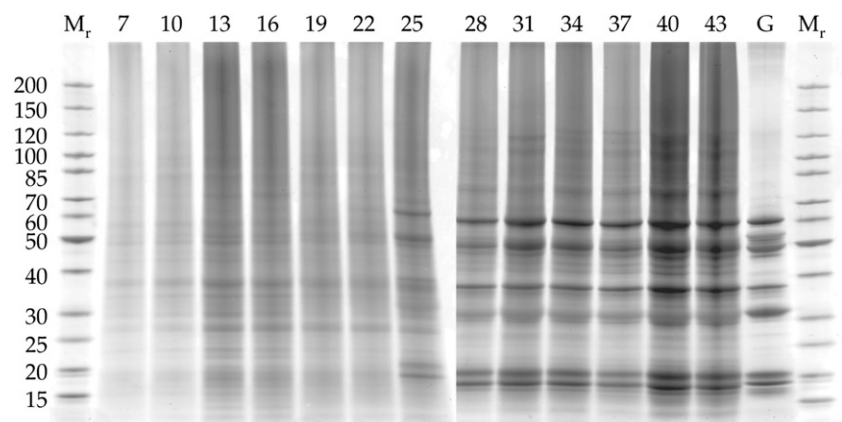


Table I. Components in mature *Lotus* seeds

Compound	Percentage of Dry Matter
Nitrogen	6.89% (approximately 43% protein)
Lipids	6.76%
Starch	0.61%
Phytic acid	0.188%
Ash	4.71%

oleic acid by the plastid stearyl-acyl carrier protein desaturase and establishment of additional double bonds by plastidial and microsomal ω -3 and ω -6 desaturases (Somerville and Browse, 1991). In *Lotus*, the level of ash is approximately 5%. The ash contains both major and minor minerals and an important supply of minerals for food and feed. A chemical analysis of seven different legumes showed ash levels ranging from 3.98% to 6.42% (Vadivel and Janardhanan, 2000).

Characterization of Globulins

Seed globulins were extracted by grinding mature seeds in liquid nitrogen followed by extraction of water-soluble proteins. The supernatant containing soluble proteins was discarded, and the pellet was resuspended in buffer containing 1 M NaCl to dissolve the globulins. Following separation on an isoelectric focusing (IEF) pH 3 to 10 gel in the first dimension and SDS-PAGE in the second dimension, 15 distinct protein spots marked in Figure 5 were analyzed by matrix-assisted laser desorption ionization quadrupole/time-of-flight mass spectrometry (MALDI-Q-TOF MS; Table III). Between five and 22 peptides were identified for each protein identification, and the MS peptide data were utilized to identify the corresponding gene accession numbers in the *Lotus* genome and the derived protein. For the globulins, a signal peptide was predicted from the corresponding gene model using the TargetP 1.1 software (Emanuelsson et al., 2007). In addition, N-terminal sequences were obtained from 10 of the identified protein spots.

Five *Lotus* legumin storage proteins (LLPs) and two *Lotus* convicilin storage proteins (LCPs) were identified. LLP1, LLP2, LLP3, LCP1, and LCP2 were represented by two spots, whereas LLP4 and LLP5 were represented by one spot on the gel (Figure 5; Table III; Supplemental Figure S1; Supplemental Table S1). Scanning of Coomassie Brilliant Blue-stained spots estimated that LLPs and LCPs contribute approximately 90% and 10% of the total globulins, respectively (data not shown).

The LLP1 precursor in Supplemental Figure S1 was assembled from the C-terminal fragment identified in spot 12 and the N-terminal fragment identified in spot 6. Cleavage at the Asn-Gly-Leu-Glu-Glu-X-X-Cys motif (Supplemental Fig. S1) between Asn-289 and Gly-290 is confirmed by the N-terminal sequence of spot

12. The initial 21 amino acid residues in the LLP1 precursor protein is a signal peptide, and in agreement with this prediction, the N-terminal sequence of spot 6 commences at amino acid residue 22. The C-terminal and N-terminal fragments of the LLP2 and LLP3 precursor proteins were identified in spots 13 and 14 and spots 2 and 1, respectively. Cleavage of LLP2 at Asn-394 and Gly-395 and LLP3 at Asn-425 and Gly-426 was confirmed by the N-terminal sequences of proteins 13 and 14, and signal peptide cleavage was confirmed by the N-terminal sequences of spots 2 and 1, respectively (Supplemental Fig. S1). The processed LLP4 was identified from spot 4. The predicted signal peptide is in agreement with the N-terminal sequence of spot 4. All of the identified peptides are positioned N terminal to the Asn-378 and Ala-379 cleavage site (Supplemental Fig. S1), and the C-terminal fragment of the LLP4 precursor was not found in this study. The processed LLP5 was identified from spot 7, and all of the identified peptides were located N terminal to the Asn-290 and Gly-291 cleavage site (Supplemental Fig. S1).

The criteria used to assign LLP1 to LLP5 as legumins was based on similarity toward globulins from other legume species. LLP1 to LLP5 are 68% to 82% similar to 11S globulins in other legumes. Processing of legumin precursors normally results in a basic protein corresponding to the C-terminal fragment, which was observed for the LLP1 to LLP3 C-terminal proteins in spots 12, 13, and 14 (Fig. 5).

The processed LCP1 was identified from peptides in spots 3 and 15. The predicted signal peptide was not in agreement with the N-terminal sequence of protein 15, which commences from amino acid residue 59 (Supplemental Fig. S1). The most N-terminal peptide identified in spot 3 commences at amino acid residue 9 (Supplemental Fig. S1). This indicated precursor processing rather than signal peptide cleavage. The processed LCP2 was identified from spots 5 and 9. Peptides were identified throughout the precursor protein for spot 5, whereas C-terminal peptides were only identified in spot 9. The criteria used to assign LCP as convicilins was based on similarity toward globulins from other legume species. LCP1 and LCP2 are 79% to 81% similar to 7S globulins in other legumes. 7S globulins are divided into two subgroups: vicilins and convicilins. Convicilins differ from vicilins by an additional intrinsically unstructured sequence rich in polar residues located adjacent to the signal peptide. The additional sequence is often between 99

Table II. Fatty acid analysis in mature *Lotus* seeds

Fatty Acid	Percentage of Fatty Acid
C16:0 (palmitate)	11.3%
C18:0 (stearate)	4.4%
C18:1 (oleate)	10.8%
C18:2 (linoleate)	46.3%
C18:3w3 (linolenate)	24.9%
C20 or more	>1%

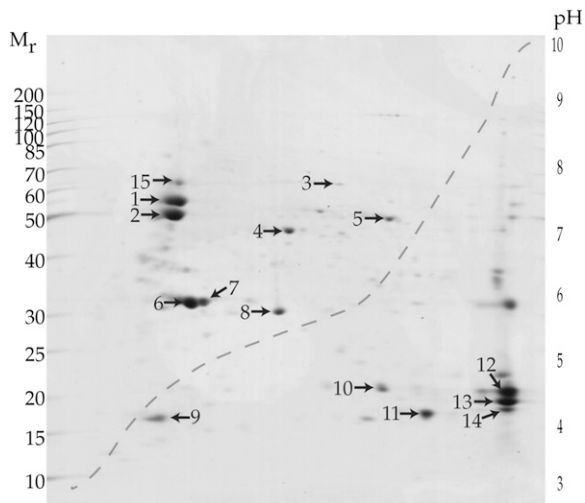


Figure 5. Storage proteins separated by 2-D PAGE. Globulins were extracted from mature seeds (43 DAF) separated on a nonlinear pH 3 to 10 gradient in the first dimension and by 5% to 15% gradient SDS-PAGE in the second dimension. The pH gradient is indicated by the dashed line. The dominant proteins (indicated by arrows) were analyzed (Table III).

and 196 amino acid residues long and increases the average molecular mass of convicilin to approximately 70 kD (Sáenz de Miera et al., 2008). Molecular masses of around 70 kD suggest that LCP1 and LCP2 are convicilins. In addition, a stretch of polar amino acid residues of Gln and Asn is found adjacent to the putative signal peptide.

The remaining spots analyzed were not related to globulins. A lectin was identified in spot 8, with 32% peptide coverage of the precursor protein (Supplemental Fig. S1). A ripening-related protein was identified in spot 11, with 65% peptide coverage. In the gel-based liquid chromatography-mass spectrometry (GeLC-MS/MS) approach (see below), this ripening-related protein was identified as hit 30 in mature seeds. In total for the MALDI and electrospray ionization approaches, 81.5% of spot 11 was covered by peptide identifications. The data obtained for spot 10 did not correspond to any predicted protein.

The Gene Structure of Globulin Genes

The relationship between the proteins and the processing of LLP1 to LLP5 and LCP1 to LCP2 precursor proteins predicted from the protein analysis was further substantiated by determining the structure of the corresponding genes. Sequencing and annotation of the *Lotus* genome was not completed when the gene models used for peptide identification were created. Genome sequences covering approximately 315 Mb of gene space of the *L. japonicus* MG20 ecotype (Sato et al., 2008) were available. To finalize our analysis, unsequenced genome segments and cDNAs were sequenced.

Partial and full-length cDNA sequences corresponding to LLP3 and LLP5 from *Lotus* were obtained by sequencing of EST clones. Alignments toward the genome sequence determined the exon-intron structure shown in Supplemental Figure S2. For LLP5, the exon-intron structure confirmed the gene model predicted by the additional genome sequence obtained by bacterial artificial chromosome sequencing and confirmed the protein sequence shown in Supplemental Figure S1 by adding 178 amino acid residues to the N terminus of the LLP5 protein and one amino acid residue to the C terminus. For LLP3, the partial (346 nucleotides of the 5' end is missing) cDNA sequence from Gifu was not identical with the predicted gene model from the MG20 genome sequence. The *Llp3* gene, therefore, was amplified from Gifu and MG20 and aligned toward the cDNA. This alignment added 32 amino acid residues to the LLP3 protein by removing an intron predicted in the gene model (Supplemental Fig. S1). In addition, five nucleotide mismatches in the exons were identified, corresponding to three amino acid residue differences in LLP3 from Gifu and MG20 (Supplemental Fig. S1).

Alignment of the genome sequence of the *Llp2* genes from the two ecotypes identified 11 nucleotide differences. This resulted in five amino acid residue differences in LLP2 between the two ecotypes (Supplemental Fig. S1).

The *Llp4* gene was completed by additional genome sequencing, and exon-intron predictions were made (Supplemental Fig. S1). The initial Met residue was not predicted in the gene model used for protein identification for LLP4, and in total, 279 amino acid residues were added to the LLP4 protein prediction (Supplemental Fig. S1).

Alignment of a full-length *Lcp1* cDNA from Gifu and MG20 genome sequences confirmed the prediction of exon-intron structure, while alignment of a cDNA lacking the initial 99 nucleotides toward the genome sequence added 27 amino acid residues to the LCP2 protein (Supplemental Fig. S1) by removing a predicted intron. An alignment of *Lcp2* genome sequence of the two ecotypes identified an additional TTC sequence corresponding to Ser-162 in exon 1 of the Gifu sequence (Supplemental Fig. S1).

Combining proteome and genome analysis identification of the LLP and LCP globulins shown in Supplemental Figure S1 was completed, and the corresponding genes were annotated. The LLPs are encoded by genes consisting of three introns except for the *Llp4* gene, which has two introns. The LCP proteins are encoded by genes with five introns.

Large-Scale Protein Identification of Two Developmental Stages

As a first approach toward a full protein profile of the seeds, proteins were extracted from two distinct developmental stages, 16 to 25 DAF (green seeds) and 43 DAF (mature seeds). The proteins were separated

Table III. 2-D-based identifications by MALDI-Q-TOF

Spot ID corresponds to the protein number in Figure 5. The observed M_r and pI values are calculated from the 2-D gel. RRP, Ripening-related protein.

Spot ID	Theoretical M_r /pI	Observed M_r /pI	Protein Identity	Peptides Matched	Sequence Coverage	MOWSE Score	Expected Value	Protein Name
1	66,770/5.43	60,000/4.9	chr5.LjT15N12.20.nc	7	16%	146	1.7e-10	LLP3
2	66,833/5.43	55,000/4.9	chr5.LjT15N12.10.nc	7	16%	85	0.00022	LLP2
3	64,937/6.35	68,000/5.7	chr5.CM0357.580.nc	20	47%	261	5.5e-22	LCP1
4	32,990/6.24	49,000/5.5	LjSGA_091288.1	7	38%	84	0.00028	LLP4
5	65,108/5.84	54,000/6.3	chr5.CM0357.560.nc	22	38%	373	3.5e-33	LCP2
6	54,596/6.32	33,000/5.0	chr1.CM0295.260.nd	8	20%	143	3.5e-10	LLP1
7	33,837/7.96	33,000/5.1	LjSGA_030216.1	6	25%	200	6.9e-16	LLP5
8	30,904/5.73	31,000/5.4	chr5.LjT43D06.110.nd	6	32%	138	1.1e-09	Lectin
9	65,108/5.84	16,000/4.8	chr5.CM0357.560.nc	5	10%	237	1.4e-19	LCP2
11	17,551/5.90	16,000/6.9	chr3.CM0110.130.nc	9	65%	241	5.5e-20	RRP
12	54,596/6.32	20,000/8.1	chr1.CM0295.260.nd	11	33%	232	4.3e-19	LLP1
13	66,833/5.43	18,000/8.1	chr5.LjT15N12.10.nc	5	11%	209	8.7e-17	LLP2
14	66,770/5.43	17,000/8.1	chr5.LjT15N12.20.nc	7	12%	188	1.1e-14	LLP3
15	64,937/6.35	68,000/4.9	chr5.CM0357.580.nc	14	35%	150	6.9e-11	LCP1

by gradient 5% to 15% SDS-PAGE. For each of the two stages, the whole gel lane was cut into approximately 40 slices and subjected to in-gel tryptic digestion, and peptides were identified using a GeLC-MS/MS approach. Two sets of databases, one *Lotus* specific and one containing general protein data, were used for peptide identification. The *Lotus* amino acid (AA) database, the *Lotus* nucleotide (NT) database, and the EST database downloaded from The Institute for Genomic Research (TIGR) belong to one set. The other set consisted of the Swiss-Prot (SPROT) and National Center for Biotechnology Information (NCBI) databases. Individual peptide sequences were verified by inspection of the peptide spectra (see "Materials and Methods" for the criteria used). A total of 665 and 181 protein identifications from green and mature seeds, respectively, corresponded to unique gene accession numbers (Table IV; Supplemental Table S2). Some peptides correspond to more than one gene accession number, either because the accessions represent duplicated genes or gene family members or because all regions of the *Lotus* genome sequence have not been assembled completely. Taking these nonunique protein identifications into account, 920 and 264 proteins were identified from green and mature seeds, respectively, using the *Lotus* AA, NT, and EST contigs (Supplemental Table S2). The overlap between protein identifications from the two developmental stages was assessed by dividing identifications into three subgroups: only green, only mature, and intersection between green and mature seeds. These three subgroups contain 764, 108, and 156 protein identifications, respectively.

For the 920 and 264 protein identifications for green and mature seeds, the molecular process was predicted. Gene Ontology (GO) annotations for the complete set of gene accession numbers from *Lotus* AA, NT, and EST contigs were obtained through a BLAST analysis against UniProt. The subdivisions into classes are

shown in Figure 6. The four dominant molecular processes (primary metabolic process, cellular metabolic process, macromolecule metabolic process, and biosynthetic process) are slightly overrepresented in the mature seeds; however, no big difference is found between green and mature seeds (Fig. 6). For the minor categories, the green seeds generally have a more uniform or higher representation than mature seeds (Fig. 6). The three categories "generation of precursor metabolites and energy," "transport," and "cellular component organization and biogenesis" are relatively lower in the mature seeds (Fig. 6). For the category "carbon utilization," the relative level is about 2-fold higher in mature seeds than in green seeds.

Pathway Analysis

GO annotations based on the 920 and 264 identifications were subsequently used for assigning proteins to biochemical pathways using Cytoscape (Shannon et al., 2003). To establish the background, the complete set of GO annotations for the gene accession numbers from *Lotus* AA, NT, and EST contigs was used. A positive set of protein identifications from, for example, mature seeds (264 gene accession numbers) were

Table IV. Summary of GeLC-MS/MS data

The unique gene accession numbers corresponding to the number of protein identifications from the search in different databases are shown.

Unique Gene Accession Nos. From:	Green Seeds	Mature Seeds
<i>Lotus</i> (AA and NT databases)	556	153
EST database TIGR	76	17
SPROT	17	6
NCBI	16	5
Total	665	181

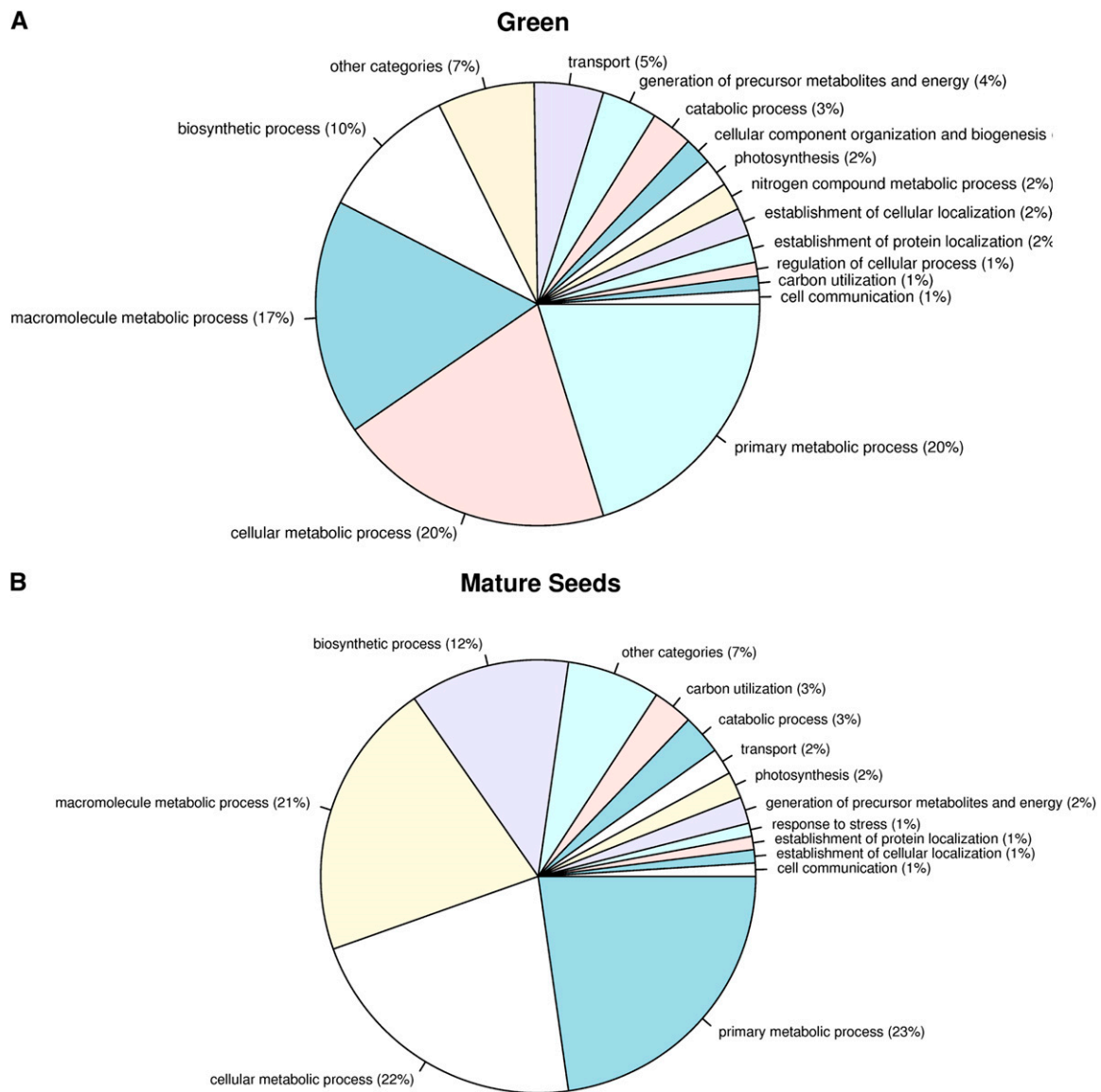


Figure 6. The molecular process of protein identifications. In A and B, the molecular processes of the protein identifications in green and mature seeds, respectively, found using the AA, NT, and EST downloaded from TIGR. Note that it is possible for the same protein to be counted in several categories if it has assignments descending from more than one of the categories depicted. [See online article for color version of this figure.]

compared against the background of the complete set of GO annotations. The biochemical pathways that were thus found to be above the background data are determined as up-regulated pathways in the individual data sets.

As an example of the pathway analysis for green seeds, 25 identified proteins with the corresponding GO annotation “amino acid biosynthesis” were found (Supplemental Table S3). Five of these are predicted to encode Gln synthetase (GS). Of the five GS genes used for identification, three are predicted to be full length and two are predicted to be partial (Supplemental

Table S3). The two partial GS genes correspond to the N and C termini, respectively; however, PCR amplification indicates that the proteins originate from two different genes (data not shown). Four of the five GS are predicted to be located in the cytosol, whereas the fifth GS is predicted to be located in plastids. Three of the protein identifications are predicted to encode a dihydrodipicolinate synthase, which is involved in the biosynthesis of Lys (Supplemental Table S3). In contrast, only one (a Met synthase) of the 25 protein identifications with the predicted GO annotation “amino acid biosynthesis” was found in mature seeds.

Web Interface to the Database

The results presented here may be further investigated using the Web interface found at <http://www.cbs.dtu.dk/cgi-bin/lotus/db.cgi>. This site has a number of options for views and cross-referencing. Guests may log in using the guest user account following the online instructions.

An analysis is associated with the sample on which it is carried out and the experiment that it is part of and may be found by navigating to either one using the tree menu on the left side. The GeLC-MS/MS analyses (divided by tissue type and database in which the peaks were identified) may thus be found by going to *Misc | Tissue/Organ Types | Green Seeds | GeLC-MS/MS green* or *Misc | Tissue/Organ Types | Mature Seeds | GeLC-MS/MS mature*. Alternatively, they can all be accessed from *Misc | Experiments | GeLC-MS/MS Experiment*. The two-dimensional (2-D) gel separation and associated MALDI-Q-TOF analyses (conducted on 14 selected spots on the gel) are found at *Misc | Tissue/Organ Types | Mature Seeds | 2-D_storage_proteins* or *Misc | Experiments | 2-D gel storage proteins*.

When a GeLC-MS/MS or MALDI-Q-TOF analysis is chosen, a node giving access to information on peak peptides will pop out in the tree structure. Furthermore, a node linking to the identified protein (or EST contig or cDNA) will pop out underneath. (Information on peaks and identifications from the GeLC-MS/MS analyses are also listed in Supplemental Table S2.) When a given protein is shown, information on related EST contigs, GO annotations, BLAST (Altschul et al., 1990, 1997) hits against the featured *Lotus* genome, and related experimental analyses in which the protein has been identified is also listed and linked to when relevant. (Data sources are described in "Materials and Methods.") The analyses also link back to the identified proteins, thus making it possible to quickly see in what other analyses that same protein has been identified. EST contig and links to and from related experimental analyses, proteins, and ESTs are shown in a similar way.

DISCUSSION

Model legume studies are important for establishing comparative studies of legume seed development, and we present here the initial microscopical, compositional, and proteome characterization of seed development in *Lotus*. The proteomics analyses established a resource of proteome and genome data that was organized in a database accessible at <http://www.cbs.dtu.dk/cgi-bin/lotus/db.cgi>. This database also holds details of the experimental setup and data from the spectra from the Q-TOF-MS and GeLC-MS/MS. Proteome data obtained from future analyses of seeds and other tissues will also be uploaded to add new information to this resource. Together with seed proteome data from *Medicago* and soybean, our *Lotus* seed proteome database will contribute to establishing the

platform for comparative legume proteomics. So far, 746 unique gene accession numbers identified from green and mature seeds constitute the *Lotus* resource. Similar 2-D gel or Sec-MudPIT-based proteomic studies of seed filling in *Medicago* and soybean identified 226 and 647 unique proteins, respectively (Gallardo et al., 2003, 2007; Hajdich et al., 2005; Agrawal et al., 2008). Comparing these *Medicago* and soybean proteins against the identified *Lotus* seed proteins (BLASTP $e \leq 1 \times 10^{-6}$), 39% of the proteins from *Lotus* were also identified in *Medicago*, and with soybean the overlap was 60%. Comparing our *Lotus* data set with all four proteomic studies from *Medicago* and soybean (873 protein identifications), approximately 27% of the *Lotus* proteins had not been identified previously. Altogether, this shows that further studies are necessary for coverage of the legume seed proteome.

Taking advantage of the 1,619 proteins so far identified in the *Lotus*, *Medicago*, and soybean seed proteomic studies, we attempted to estimate the scope of comparative legume proteomics in predicting biochemical pathways. Of the 1,619 proteins, 1,299 proteins found a homolog in UniprotKB ($e \leq 1 \times 10^{-6}$; Supplemental Table S4). Of these, 612 proteins could be assigned to 213 different enzymatic functions, defined by their Enzyme Commission numbers, and metabolic networks could be derived using the KEGG LIGAND database. In total, enzymes from 123 pathways were identified, and for 42 pathways, five or more enzymes were identified (Supplemental Table S5). Approximately 30% of the pathways in Supplemental Table S5 are involved in amino acid metabolism.

None of the 123 identified pathways are fully covered; however, the merge of proteome data from three legume species increases the coverage for most of the pathways. Although *Lotus*, *Medicago*, and soybean are separate species, we suggest that comparing and compiling proteome data sets can contribute toward a better understanding of seed development and function in legumes. Further details of the pathways assembled from *Lotus*, *Medicago*, and soybean are available at <http://bioinfoserver.rsbs.anu.edu.au/utills/PathExpress4legumes/>.

Globulin Seed Storage Proteins and Globulin Genes in *Lotus*

The overall seed development for *Lotus* is similar to that of other legumes; however, the timing of the shifts between the three growth phases and the presence of endosperm in mature seeds differ between legume species (Borisjuk et al., 1995; Gallardo et al., 2003). The content and composition of lipid, nitrogen, starch, phytic acid, and ash in mature *Lotus* seeds are more similar to those in soybean than in pea. The legumins LLP1 to LLP5 are the major part of the globulins in mature *Lotus* seeds (Fig. 5). This is in contrast to pea, in which the vicilins are the major part of globulins. Tzitzikas et al. (2006) found that two-thirds or more of

the globulins were vicilins for 50 of 59 pea lines; however, in soybeans, the legumins are the major part of the globulins (Yaklich, 2001).

Globulins were identified using a 2-D gel approach, while the proteome was investigated using a GeLC-MS/MS approach. Posttranslational processing of precursor proteins was discovered in the globulins, but in general our methods were not directed toward detection of posttranslational modifications such as phosphorylation and glycosylation. Indications of such modifications were observed, however. For the N-terminal fragments of LLP2, LLP3, and LLP4, the observed molecular masses are approximately 12, 13, and 8 kD, respectively, higher than the theoretical molecular masses (Supplemental Table S1). Glycosylations could cause this molecular mass difference. Glycosylations are normally not found in legumins; however, glycosylated legumins are found in lupin (*Lupinus albus*) and Japanese red cedar (*Cryptomeria japonica*; Duranti et al., 1988; Wind and Häger, 1996). The common N-glycosylation motif, Asn-X-Ser/Thr, is lacking in LLP2 and LLP3. LLP4 has an Asn-X-Ser glycosylation motif at amino acid residues 86 to 88, but a peptide overlapping this potential glycosylation site was identified, indicating that glycosylation at this site may be absent or incomplete (see "Materials and Methods"). The theoretical pI values for the C-terminal fragment of LLP1, LLP2, and LLP3 are higher than the observed pI values (Supplemental Table S1). This could be caused by modification but is rather due to the limited resolution in the high pH range of the IEF strips used for the first dimension.

The N-terminal sequence of LCP1 (spot 15) and identification of LCP2 in spots 5 and 9 indicate that *Lotus* convicilins are processed. This is unusual, since no posttranslational modification other than signal peptide cleavage has been reported for convicilins (Croy et al., 1980; Newbigin et al., 1990). Degradation during extraction cannot be excluded; therefore, processing of LCP1 and LCP2 needs further investigation.

The ancestral angiosperm gene structure for 11S globulin has four introns, while present day 11S globulins have three introns (Häger et al., 1996). Further loss of an intron has given rise to the two-intron genes found in several angiosperms (Häger et al., 1996). In accordance, we found that the *Lotus* *Llp1*, *Llp2*, *Llp3*, and *Llp5* genes have three introns, whereas *Llp4* has two introns.

Legume 7S globulin genes have five introns (Weschke et al., 1987; McHenry and Fritz, 1992), while genes from monocots such as maize (*Zea mays*) have four introns (Belanger and Kriz, 1991). The five-intron structure of the *Lotus* LCPs and the length of the exons correspond to convicilins from pea and bean (Sáenz de Miera et al., 2008).

DNA polymorphism between the *L. japonicus* Gifu and MG20 ecotypes used in this study is generally located outside of the coding regions (Kawaguchi et al., 2001). Therefore, it was interesting to compare the conservation of globulin genes, in which selection

for protein function is less stringent even though protease sites involved in processing and the Cys residue responsible for the assembly of disulfide bonds are maintained (Staswick et al., 1984; Scott et al., 1992; Jung et al., 1998). The genome sequences of the *Llp2*, *Llp3*, and *Lcp2* genes were completed for both ecotypes, and differences in the gene sequences were observed. For *Lcp2*, an insertion/deletion of three nucleotides corresponding to a Ser occurred at nucleotide 483 in the coding sequence. For the *Llp2* and *Llp3* genes located head to head on chromosome 5, the situation is more complex. As shown in Supplemental Figures S3 and S4 these two genes most likely originate from duplication of an ancestral gene that occurred prior to the separation of the Gifu and MG20 ecotypes. According to this hypothesis, changes in the ancestral gene sequence that accumulated in the progenitor of the two ecotypes would be shared between the *Llp2* and *Llp3* genes from the ecotypes. These 156 nucleotide differences are marked with green in Supplemental Figure S4. Subsequent nucleotide changes occurring in the individual ecotypes are marked in gray. Interestingly, these nucleotide changes were found in exons of *Llp2* from Gifu and *Llp3* from MG20, while no mutations were found in the corresponding *Llp2* from MG20 and *Llp3* from Gifu. The second intron of MG20 *Llp3* (Supplemental Fig. S4) is unique in that there is an unusual accumulation of insertions and mutations compared with intron 2 for the three other genes (Supplemental Fig. S4). Only one position in the alignment (marked with yellow in Supplemental Figure S4) does not support the above-described hypothesis. However, a single nucleotide mutation in the *Llp2* or *Llp3* gene of Gifu or MG20 followed by gene conversion is a possible explanation for the observed difference.

Protein Identifications

For the large-scale proteome analysis, the GeLC-MS/MS approach was used and up to 920 and 264 proteins were identified from green and mature seeds, respectively. A possible explanation for this relatively large difference in protein identifications is that peptides from globulins will dominate the MS spectra in mature seeds. Supporting this suggestion, the six protein identifications from mature seeds with the highest MOWSE score are globulins (Supplemental Table S2). Overall, 83% and 41% of the protein identifications were seen only in green and mature seeds, respectively, which indicates a large metabolic difference for these stages; however, the molecular process pattern (Fig. 6) of green and mature seeds is similar.

Protein identifications 7, 8, and 9 from mature seeds do not have any GO annotation (Supplemental Table S2); however, a BLAST search for these three proteins identifies them as homologous of late embryogenesis abundant proteins (LEA proteins). LEA proteins are widespread in higher plants, especially in the seeds,

but generally not found elsewhere. The three LEA proteins identified in *Lotus* seem to be most similar to LEA proteins belonging to subgroup 3. In pea, a subgroup 3 LEA protein is found in mitochondria (Grelet et al., 2005; Tolleter et al., 2007). One of the well-characterized molecular processes of LEA proteins is to protect proteins and other molecules from dehy-

dratation (Gilles et al., 2007). The LEA proteins identified at protein identifications 8 and 9 are found only in mature seeds, whereas protein identification 7 is also found in green seeds, as is protein identification 31 (Supplemental Table S2).

Protein identification 10 in mature seeds is homologous to oleosin. Oleosin is a structural component of

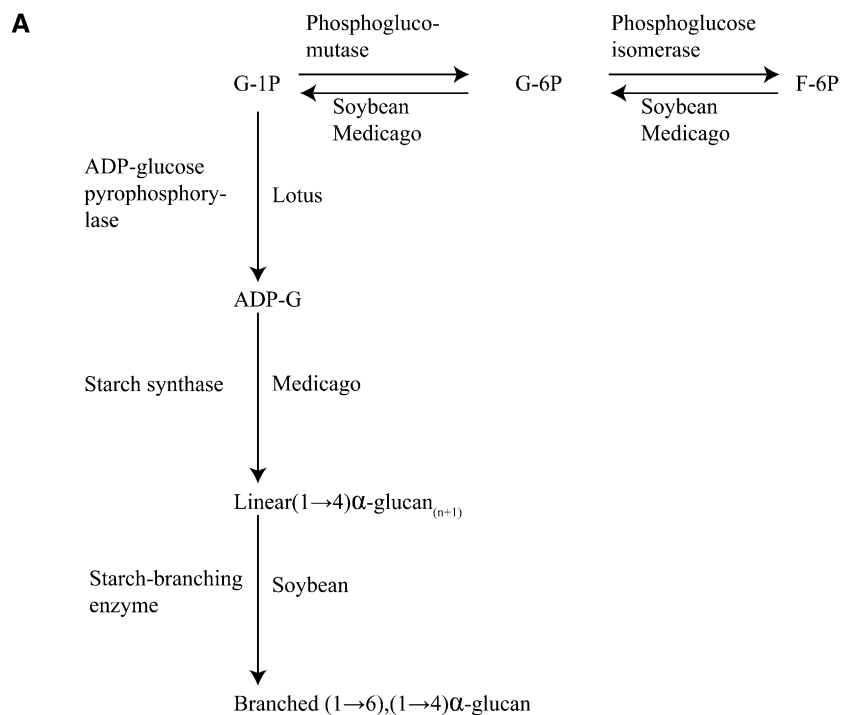
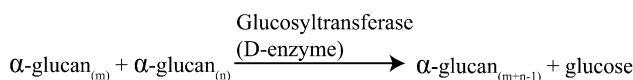
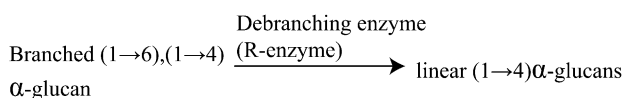
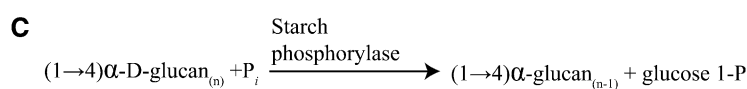
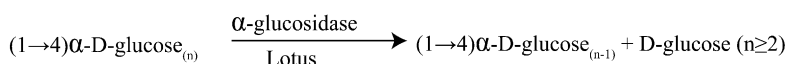
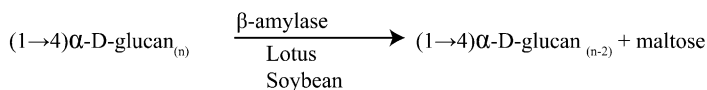
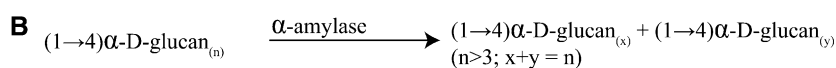


Figure 7. Comparative analysis of starch metabolism in legumes. A, Starch biosynthesis. The key enzyme in starch biosynthesis is ADP-Glc pyrophosphorylase, which converts Glc-1-P (G-1P) to ADP-Glc (ADP-G). B, The irreversible hydrolytic breakdown of starch. C, The phosphorylytic breakdown of starch. *Lotus*, *Medicago*, and soybean indicate the species in which a proteomics-based identification was obtained. F-6P, Fru-6-P; G-6P, Glc-6-P.



the oil body, where it limits coalescence during the cytoplasm compression caused by desiccation.

The 10 protein identifications with the highest MOWSE scores in mature seeds are either globulins or proteins directly or indirectly involved in desiccation, which is in good accordance with the developmental stage of the seed.

Globulins were already found in green seeds as protein identifications 1, 2, 3, 7, and 8. A protein disulfide isomerase PDIL-1 homolog was identified as proteins 6 and 40 in green and mature seeds, respectively, and a PDIL-2 homolog was identified as protein 242 in green seeds. Protein disulfide isomerases play important roles in the formation of disulfide bonds in nascent proteins. Kamauchi et al. (2008) characterized two PDILs in soybean (GmPDIL-1 and GmPDIL-2) that were suggested to be molecular chaperones involved in proper folding or quality of legumins. This may suggest a role for *Lotus* PDILs in disulfide bond formation in the LLPs.

The GO annotations for the identified proteins were used for pathway analysis. For the pathway category "amino acid biosynthesis" in green seeds, five different GS proteins were identified. In plants, GS is encoded by a small multigene family, and cytosolic GS (GS₁) and chloroplastic GS (GS₂) are the two major isoforms. The GS genes are typically expressed in different organs and tissues, and it has been proposed that each GS isoform has a specific function in assimilation and reassimilation of ammonia (Cren and Hirel, 1999). Our data indicate that all of the *Lotus* GS₁ genes were expressed in green seeds. Furthermore, GS₂ was identified in green seeds, and this is to our knowledge the first time that GS₂ was identified in seeds. In pea, three GS₁ genes have been characterized: all are expressed in nodules and two of them are expressed in cotyledons (Tingey et al., 1987). In soybean, only one GS₁ is found in the cotyledons (Morey et al., 2002). Not all 20 amino acids are translocated to the seed coat through the phloem. In bean, mainly Glu and/or Asn are translocated (Mifflin and Lea, 1977), while in pea, mainly Glu, Ala, and Thr are unloaded from the seed coat (25%, 20%, and 15%, respectively). These amino acids are subsequently converted into other amino acids in which GS is an important enzyme synthesizing Gln from Glu. Furthermore, three of the proteins in amino acid biosynthesis are predicted to catalyze a step in the biosynthesis of Lys, which is in good agreement with the high Lys content of legume seeds.

In the mature *Lotus* seed, a Met synthase was identified that is involved in the synthesis of Met from homo-Cys. Met synthase and several other enzymes involved in amino acids metabolism have been identified in *Medicago* seed development (Gallardo et al., 2003, 2007).

Starch Metabolism in *Lotus*, *Medicago*, and Soybean

Histological sections of the *Lotus* seed stained with iodine potassium iodide indicated degradation and/or mobilization of starch from osteosclereid and parenchyma cells at 25 DAF (data not shown), and the

overall starch level decreased prior to desiccation of the *Lotus* seed. In the mature *Lotus* seed, the starch level amounts to 0.6%. A comparably low level was found in the mature soybean and *Medicago* seed, while the pea seed contained approximately 27% starch (Chavan et al., 1998). Interestingly, soybean and *Medicago* also appear to degrade and/or mobilize starch prior to desiccation (Djemel et al., 2005; Stevenson et al., 2006). Refining the bioinformatics-based comparison presented above, we compiled starch enzyme identifications from the five seed proteomic studies (Fig. 7; Gallardo et al., 2003, 2007; Hajduch et al., 2005; Agrawal et al., 2008). Although legume seed starch metabolism may be different from the well-studied processes in leaves and in cereal grains, this approach provided good coverage of known starch biosynthetic and degradation pathways (Fig. 7, A and B). The key regulator enzyme for starch biosynthesis, ADP-Glc pyrophosphorylase, was not identified in any of the legume proteomic studies; however, the enzyme has been identified in a *Lotus* 2-D-PAGE analysis of seeds at 7 DAF (G. Nautrup-Pedersen, S. Dam, B.S. Laursen, A.L. Siegumfelt, H.H. Stærfeldt, C. Friis, K. Nielsen, N. Goffard, S. Sato, S. Tabata, P. Roepstorff, and J. Stougaard, unpublished data). No enzymes belonging to the phosphorylytic breakdown pathway were detected in the five seed proteomics studies from *Lotus*, *Medicago*, or soybean (Fig. 7C), while enzymes of the irreversible hydrolytic pathway were found. β -Amylases were identified in *Lotus* and soybean (Fig. 7B), and an isoamylase-like enzyme together with an enzyme phosphorylating glucan (EC 2.7.9.4), increasing the activity of β -amylase in leaves (Delatte et al., 2006; Edner et al., 2007), was found in soybean. However, isoamylase is also indispensable in starch biosynthesis (Zeeman et al., 1998). The α -glucosidase enzyme was found in a *Lotus* 2-D-PAGE seed analysis (G. Nautrup-Pedersen, S. Dam, B.S. Laursen, A.L. Siegumfelt, H.H. Stærfeldt, C. Friis, K. Nielsen, N. Goffard, S. Sato, S. Tabata, P. Roepstorff, and J. Stougaard, unpublished data). Altogether, this comparative approach identified two pathways probably involved in the starch synthesis and the hydrolytic starch breakdown in the three legumes. Interestingly, degradation of starch in germinating mung bean (*Vigna radiata*), which has a starch content of approximately 54% in mature seeds (Mubarak, 2005), is mainly accomplished by the phosphorylytic breakdown pathway (Matheson and Richardson, 1976; Berkel et al., 1991). Both types of starch degradation have thus been described in legume seeds.

Altogether, these observations highlight the need for improved knowledge of the genes and proteins responsible for the difference in protein, oil, and starch content of different legumes.

MATERIALS AND METHODS

Plant Material

Lotus japonicus ecotype Gifu was grown in a greenhouse with a 16-h-light/8-h-dark cycle, and seeds were harvested at different developmental stages

defined by DAF. Seeds were harvested into 70% ethanol for histochemical studies and directly into a mortar placed on dry ice for protein extractions.

Seed fresh and dry weights were determined as five replicates. The seeds were weighed immediately after harvest (fresh weight) and after drying at 65°C for 72 h (dry weight). Afterward, the water content was calculated.

Fixation, Sectioning, and Staining

A small hole was made in the seeds prior to fixation to allow the fixation liquid to penetrate the seed coat of the mature seeds. The seeds were fixed in 4% (w/v) paraformaldehyde in 50 mM potassium phosphate buffer (pH 7.0) for 2 d at room temperature. The seeds were then vacuum infiltrated to remove trapped air and dehydrated in a graded ethanol series (70%, 80%, 90%, and 96%). Finally, the seeds were embedded in Technovit 7100 (Haereus Klüber) according to the manufacturer's instructions. Thin sections (5 μ m) were stained with toluidine blue, and the slides were observed by light microscopy.

DNA Techniques

Genomic DNA was prepared using plant leaf material from *L. japonicus* Gifu and *L. japonicus* MG20 harvested and stored at -20°C until use. Tissue (0.3–0.5 g) was ground in liquid nitrogen. Then, 5 mL of preheated (60°C) extraction buffer (0.1 M Tris-HCl, pH 8.0, 1.4 M NaCl, 0.02 M EDTA, and 2% cetyltrimethyl ammonium bromide) was added and incubated for 30 min. Five milliliters of chloroform:isoamyl alcohol (24:1) was added, mixed, and centrifuged at 9,300g at room temperature. The aqueous phase was transferred to a new tube, and 10 μ g of RNase per milliliter was added and incubated for 30 min at 37°C. The sample was placed on ice for 5 min, and 0.6 volumes of ice-cold isopropanol was added. The genomic DNA was precipitated at -20°C for 20 min and centrifuged at 14,000g at 4°C for 10 min. The DNA pellet was washed with 70% ethanol, air dried, and dissolved in TE buffer (10 mM Tris-HCl, pH 8.0, and 0.1 mM EDTA, pH 8.0).

Approximately 50 ng of genomic DNA was used for each PCR. PCR products were sequenced using Applied Biosystems BigDye version 3.1 and run on an ABI3130 DNA sequencer.

Partial and full-length cDNA clones originating from a *L. japonicus* seed and seed pod library were sequenced as described above. LjPEST3h1 corresponds to LCP1, LjPEST2e7 corresponds to LCP2, LjPEST2h11 corresponds to LLP3, and LjPEST4d11 corresponds to LLP5.

The exon-intron structures of the globulin genes were predicted using the program FGENESH, last modified on July 24, 2007 (<http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>).

Components in Mature Seeds

The level of nitrogen was determined by the Kjeldahl method. The level of lipids was determined as described (Stoldt, 1952; Rotenberg and Andersen, 1980). The level of starch was determined as described (Åman and Hesselman, 1984), and the level of phytic acid was determined as described (Haug and Lantzsch, 1983).

Preparation of Protein Extracts

Total protein extracts were prepared essentially as described (Wang et al., 2003). Globulins were extracted as follows. Thirty dry seeds were ground in a mortar, and water-soluble proteins were extracted twice in 1 mL of water (10 min, 4°C) followed by centrifugation for 10 min at 13,000g. The insoluble material was resuspended in 1 mL of 0.1 M Tris, pH 8.0, and 1 M NaCl. The solution was incubated for 10 min, followed by centrifugation for 10 min at 13,000g. The supernatant was transferred to a new tube, and 3 volumes of 10% trichloroacetic acid in acetone was added. Proteins were allowed to precipitate for 1 h at -20°C . The solution was centrifuged for 13,000g for 10 min at 4°C, and the resulting pellet was washed twice with 80% acetone. The final pellet was allowed to dry at room temperature and dissolved in 250 μ L of IEF buffer (5 M urea, 2 M thiourea, 2% CHAPS, 2% SB3-10, 10 mM dithiothreitol, 2 mM EDTA, and 0.5% [w/v] immobilized pH gradient [IPG] buffer, pH 3–10 nonlinear [GE Healthcare]) for IEF.

One- and Two-Dimensional Electrophoresis

Reduced proteins were separated in one dimension by SDS-PAGE on 5% to 15% gradient gels using the Gly/2-amino-2-methyl-1,3-propanediol-HCl system described by Bury (1981).

IPG strips (Immobiline Dry Strips, 7 cm, pH 3–10 nonlinear [GE Healthcare]) were rehydrated overnight with 125 μ L of sample in IEF buffer. IEF was carried out on an IPGphor II system (GE Healthcare) using the following program: 200 V (step) for 20 volt hours (Vh), 3,500 V (gradient) for 2,500 Vh, and 3,500 V (step) for 5,000 Vh. After IEF, the proteins were reduced and alkylated by equilibration of the strips for 15 min in Eq buffer (50 mM Tris, pH 8.8, 6 M urea, 30% glycerol, 2% SDS, and 0.002% bromophenol blue) containing 6.5 mM dithiothreitol followed by 15 min in Eq buffer containing 10 mM iodoacetamide. For separation in the second dimension, IPG strips were loaded on 5% to 15% gradient gels as described above. Proteins were visualized by Coomassie Brilliant Blue R-250 staining.

MS-Based Protein Identification

For MS-based protein identification from 2-D gels, spots were excised, washed in water, dehydrated with 50 μ L of acetonitrile for 15 min, and vacuum dried. In-gel digestions were performed by adding 20 μ L of 12.5 μ g mL $^{-1}$ modified porcine trypsin (Promega) in 50 mM NH $_4$ HCO $_3$. The gel pieces absorbed the trypsin solution for 45 min, and the remaining liquid was removed. Finally, 30 μ L of 50 mM NH $_4$ HCO $_3$ was added and the samples were incubated at 37°C for 16 to 18 h. The resulting tryptic peptides were adsorbed to a POROS 20 R2 resin packed in a homemade microscale column (Kusmann et al., 1997), washed with 10 μ L of 1% trifluoroacetic acid, and eluted on a MALDI sample target using 1 μ L of 0.4% recrystallized α -cyano-4-hydroxycinnamic acid (Sigma) in 70% acetonitrile and 0.1% trifluoroacetic acid. Peptides were analyzed by MALDI-MS using a Q-TOF Ultima Global mass spectrometer (Micromass). The MS spectra were combined, background subtracted (polynomial order 4, 60%–80% below curve, tolerance 0.010) and deisotoped (MaxEnt 3) using Masslynx 4.0 (Micromass). The resulting peak lists were used to search for matches in an in-house-compiled database of annotated *Lotus* genome sequences (lotus_060912aa [68,838 sequences, 20,265,406 residues] and lotus_060912nt [416,940 sequences, 122,006,960 residues]) using Mascot 2.1 (Perkins et al., 1999). For each spot, peptide mass fingerprint data and one to five MS/MS spectra were combined into a single search performed with a peptide tolerance of 50 ppm and a MS/MS tolerance value of 0.6 D. Only peptides produced by trypsin with the possibility of one missed cleavage were considered. Searches were performed with carbamidomethylation as a fixed modification of Cys residues and oxidation on Met residues as a variable modification. Hits were considered to be significant if the MOWSE score was greater than 61 for lotus_060912aa and greater than 69 for lotus_060912nt. If no hit was found, the data were searched in an in-house-compiled database of *Lotus* TC sequences obtained from TIGR (*L. japonicus* Gene Index release 3.0) and finally in SPROT (version 49.7 [219,361 sequences, 80,573,946 residues]) and NCBI nr (version 20060522 [3,651,628 sequences, 1,255,333,329 residues]). Hits were considered to be significant if the MOWSE score was greater than 65 for ESTs downloaded from TIGR, greater than 66 for SPROT, and greater than 78 for NCBI.

GeLC-MS identifications were performed in duplicate on samples of seeds at 16 to 25 DAF and mature seeds. Proteins were extracted and separated by one-dimensional SDS-PAGE as described above. The lanes were cut into approximately 40 slices, and proteins were digested with trypsin as described above, with the exception that 20 μ L of 25 μ g mL $^{-1}$ trypsin was used. The samples were then acidified by the addition of 1 μ L of CH $_3$ COOH and filtered through a microspin filter (0.2 μ m) prior to LC-MS/MS.

LC-MS/MS was performed using an LC-Packings Ultimate Nano LC system connected online to a Q-TOF Ultima Global mass spectrometer (Micromass). Peptides were trapped and isocratically desalted on a reverse-phase precolumn (LC-Packings PepMap C18, 300 μ m \times 5 mm) and eluted onto a fused-silica emitter (PicoFrit, New Objective, 10 cm \times 75 μ m packed with Zorbax 300 SB C $_{18}$ 3.5-mm particles; Agilent). The peptides were separated using a linear gradient from buffer A (0.02% hepta-fluorobutyric acid and 0.5% CH $_3$ COOH in water) to buffer B (0.02% hepta-fluorobutyric acid, 0.5% CH $_3$ COOH, and 75% acetonitrile in water) for 90 min using a flow rate of 200 nL min $^{-1}$. Survey scans were acquired over the mass-to-charge ratio range 400 to 2,000, and a maximum of three concurrent MS/MS acquisitions were triggered for precursors detected at an intensity above a predefined threshold and with a charge of +2 or +3. Each MS/MS acquisition

was stopped after 6 s or when the precursor intensity dropped below a given threshold. After data acquisition, the MS/MS data were processed using Masslynx 4.0 (Micromass) as follows. Sequential scans from the same precursor were combined, and the resulting data were further processed if the QA score was above 10.00. The Mass Measure algorithm was used for background subtraction and deisotoping (Background subtract: polynomial order 1, below curve 55.00%; Smooth: smooth window (channels) 4.00, number of smooths 1, mode Savitzky Golay; Centroid: minimum peak width at half height 2; Centroid mode: centroid top 80.00%). A very conservative approach for accepting a hit as a positive identification was applied. The peptide tolerance was set to 1.2 D, and the MS/MS tolerance was set to 0.6 D, and only peptides produced by trypsin with the possibility of one missed cleavage were considered. Oxidation of Met residues and propionamide modification of Cys residues were included as possible modifications. Identifications were only considered if the probability that the match was random was less than 0.01. Peptides with scores less than 20 were not considered, and the MS/MS spectra for peptides scoring between 20 and 40 were manually inspected to verify the assignment of peaks by Mascot. Protein identifications based on only one peptide were all verified by manual inspection of the MS/MS spectra. Manually inspected spectra were only considered significant if they contained a peptide sequence tag of at least three consecutive amino acid residues.

N-Terminal Sequencing

For N-terminal sequencing, proteins were first electroblotted from the 2-D gels to a polyvinylidene difluoride membrane and stained with Coomassie Brilliant Blue R-250. Spots were excised from the membrane and analyzed by automated Edman degradation using an Applied Biosystems Procise 491HT sequencer with online phenylthiohydantoin HPLC analysis. The instrument was operated according to the manufacturer's instructions.

The Database and Interface

The *Lotus* database project integrates data from various experimental analyses (including mass spectrometry and protein-protein interactions), GO (Ashburner et al., 2000) annotations, the sequence databases UniProt (Bairoch et al., 2005) and NR (http://www.ncbi.nlm.nih.gov/blast/blast_databases.shtml), *L. japonicus* genome, cDNA, EST contig, and protein data obtained from the Kazusa DNA Research Institute (Sato et al., 2008), and EST contig, EST, and protein data from The Gene Index Databases (Lee et al., 2005; <http://compbio.dfci.harvard.edu/tgi/>). BLAST (Altschul et al., 1990, 1997) was used to align EST contig (BLASTN), cDNA (BLASTN), and protein (TBLASTN) sequences against the unassembled genome data contained in the database and to find provisional GO annotations taking the annotations from the best matching protein in UniProt. The setup is an implementation of LAMP (Linux, Apache Web server, MySQL, and Perl). The Web interface also uses JavaScript and an extended version of Tree Menu (<http://www.treemenu.net/>).

The MySQL database itself was originally designed using DBDesigner4. From the user's point of view, the database schema centers on the relationship between samples, analyses (each belonging to an experiment), data, and annotations of sequences. An overview of the database schema can be seen in Supplemental Figure S5.

The different analysis types have information in their own specific tables: EstAnalysis, ArrayAnalysis, SeparationAnalysis, InteractionAnalysis, MsAnalysis, SageAnalysis, or ModificationProcess, that all inherit from the common Analysis table structure and in most cases have subordinate tables. Additional descriptive features not available in the existing table structure may be added to an entry by defining feature types (keys) and corresponding values contained in separate tables. This can all be done from the Web interface, and the database is thus user extendable.

The Web interface allows new experiments and analyses to be created. Data may then be filled in online in forms or uploaded in different formats (e.g. XML files that describe the analysis and its data). Furthermore, trusted users may modify and expand the Web interface through the creation of user-defined plots and data views. These scripts may be either global (for all users) or for the individual user only.

Accession Numbers

The genome sequences of *L. japonicus* Gifu *Llp2*, *Llp3*, *Lcp1*, and *Lcp2* correspond to accession numbers FM211901, FM211902, FM211903, and

FM211904, respectively. Additional genome sequences of *L. japonicus* MG20 for *Llp4* and *Llp5* correspond to accession numbers AP010875, AP010876, and AP010877, in which the first two accession numbers correspond to *Llp4*. The cDNA sequences of *L. japonicus* Gifu *Llp3*, *Llp5*, *Lcp1*, and *Lcp2* correspond to accession numbers FM211905, FM211906, FM211907, and FM211908, respectively.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Amino acid sequences and peptide identifications of proteins separated by 2-D PAGE.

Supplemental Figure S2. Exon-intron structures of storage proteins.

Supplemental Figure S3. Hypothesis of evolution of the *Llp2* and *Llp3* genes.

Supplemental Figure S4. ClustalW alignment of *Llp2* and *Llp3* for the two ecotypes MG20 and Gifu.

Supplemental Figure S5. An overview of the database schema.

Supplemental Table S1. Predicted cleavage sites and observed and theoretical molecular weights and pI for 2-D gel.

Supplemental Table S2. Peaks and peptide identifications in the GeLC-MS experiment.

Supplemental Table S3. Pathway analysis of annotation "amino acid biosynthesis" in green seeds.

Supplemental Table S4. Identified proteins in *Lotus*, *Medicago*, and soybean proteomic studies with corresponding identifiers and Enzyme Commission numbers.

Supplemental Table S5. Comparative pathway analysis in legumes.

ACKNOWLEDGMENT

We thank Michael Udvardi for providing the globulin LjPEST clones *Llp3*, *Llp5*, *Lcp1*, and *Lcp2*.

Received November 28, 2008; accepted January 3, 2009; published January 7, 2009.

LITERATURE CITED

- Abirached-Darmency M, Abdel-gawwad MR, Conejero G, Verdeil JL, Thompson R (2005) In situ expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in *Medicago truncatula* and *Pisum sativum* seeds. *J Exp Bot* **56**: 2019–2028
- Agrawal GK, Hajduch M, Graham K, Thelen JJ (2008) In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. *Plant Physiol* **148**: 504–518
- Altschul SE, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410
- Altschul SE, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
- Åman P, Hesselman K (1984) Analysis of starch and other main constituents of cereal grains. *Swed J Agric Res* **14**: 135–139
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29
- Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al (2005) The universal protein resource (UniProt). *Nucleic Acids Res* **33**: D154–D159
- Belanger FC, Kriz AL (1991) Molecular basis for allelic polymorphism of the maize Globulin-1 gene. *Genetics* **129**: 863–872
- Berkel J, Conrads-Strauch J, Steup M (1991) Glucan-phosphorylase forms in cotyledons of *Pisum sativum* L.: localization, developmental change, in-vitro translation, and processing. *Planta* **185**: 432–439

- Borisjuk L, Wang TL, Rolletschek H, Wobus U, Weber H** (2002) A pea seed mutant affected in the differentiation of the embryonic epidermis is impaired in embryo growth and seed maturation. *Development* **129**: 1595–1607
- Borisjuk L, Weber H, Panitz R, Manteuffel R, Wobus U** (1995) Embryogenesis of *Vicia faba* L.: histodifferentiation in relation to starch and storage protein synthesis. *J Plant Physiol* **147**: 203–218
- Bury AF** (1981) Analysis of protein and peptide mixtures: evaluation of three sodium dodecyl sulphate-polyacrylamide gel electrophoresis buffer systems. *J Chromatogr A* **213**: 491–500
- Chavan UD, Shahidi F, Bal AK, McKenzie DB** (1998) Nutrient composition of beach pea (*Lathyrus maritimus* L.) seeds and pod shells at various stages of maturity. *J Food Biochem* **23**: 323–340
- Chittenden RH, Folin O, Gies WJ, Koch W, Osborne TB, Osborne TB, Levene PA, Mandel JA, Mathews AP, Mendel LB** (1908) Joint recommendations of the physiological and biochemical committees on protein nomenclature. *Science* **27**: 554–556
- Cren M, Hirel B** (1999) Glutamine synthetase in higher plants: regulation of gene and protein expression from the organ to the cell. *Plant Cell Physiol* **40**: 1187–1193
- Croy RR, Gatehouse JA, Tyler M, Boulter D** (1980) The purification and characterization of a third storage protein (convicilin) from the seeds of pea (*Pisum sativum* L.). *Biochem J* **191**: 509–516
- Delatte T, Umhang M, Trevisan M, Eicke S, Thorneycroft D, Smith SM, Zeeman SC** (2006) Evidence for distinct mechanisms of starch granule breakdown in plants. *J Biol Chem* **281**: 12050–12059
- Djemel N, Guedon D, Lechevalier A, Salon C, Miquel M, Prosperi JM, Rochat C, Boutin JP** (2005) Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol Biochem* **43**: 557–566
- Domoney C, Duc G, Ellis TH, Ferrandiz C, Firnhaber C, Gallardo K, Hofer J, Kopka J, Kuster H, Madueno F, et al** (2006) Genetic and genomic analysis of legume flowers and seeds. *Curr Opin Plant Biol* **9**: 133–141
- Duranti M, Gius C** (1997) Legume seeds: protein content and nutritional value. *Field Crops Res* **53**: 31–45
- Duranti M, Guerrieri N, Takahashi T, Cerletti P** (1988) The legumin-like storage protein of *Lupinus albus* seeds. *Phytochemistry* **27**: 15–23
- Edner C, Li J, Albrecht T, Mahlow S, Hejazi M, Hussain H, Kaplan F, Guy C, Smith SM, Steup M, et al** (2007) Glucan, water dikinase activity stimulates breakdown of starch granules by plastidial β -amylases. *Plant Physiol* **145**: 17–28
- Egli DB, Ramseur EL, Zhen-wen Y, Sullivan CH** (1989) Source-sink alterations affect the number of cells in soybean cotyledons. *Crop Sci* **29**: 732–735
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H** (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protocols* **2**: 953–971
- Gallardo K, Firnhaber C, Zuber H, Hericher D, Belghazi M, Henry C, Kuster H, Thompson R** (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol Cell Proteomics* **6**: 2165–2179
- Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, 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
- Gilles GJ, Hines KM, Manfre AJ, Marcotte WRJ** (2007) A predicted N-terminal helical domain of a group 1 LEA protein is required for protection of enzyme activity from drying. *Plant Physiol Biochem* **45**: 389–399
- Graham PH, Vance CP** (2003) Legumes: importance and constraints to greater use. *Plant Physiol* **131**: 872–877
- Grelet J, Benamar A, Teyssier E, Avelange-Macherel MH, Grunwald D, Macherel D** (2005) Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiol* **137**: 157–167
- Häger KP, Müller B, Wind C, Erbach S, Fischer H** (1996) Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. *FEBS Lett* **387**: 94–98
- Hajdúch M, Ganapathy A, Stein JW, Thelen JJ** (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
- Haug W, Lantzsch HJ** (1983) Sensitive method for the rapid determination of phytate in cereals and cereal products. *J Sci Food Agric* **34**: 1423–1426
- Higgins TJV, Spencer D** (1981) Precursor forms of pea vicilin subunits: modification by microsomal membranes during cell-free translation. *Plant Physiol* **67**: 205–211
- Jung R, Scott MP, Nam YW, Beaman TW, Bassuner R, Saalbach I, Muntz K, Nielsen NC** (1998) The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell* **10**: 343–357
- Kamauchi S, Wadahama H, Iwasaki K, Nakamoto Y, Nishizawa K, Ishimoto M, Kawada T, Urade R** (2008) Molecular cloning and characterization of two soybean protein disulfide isomerases as molecular chaperones for seed storage proteins. *FEBS J* **275**: 2644–2658
- Kawaguchi M, Motomura T, Imaizumi-Anraku H, Akao S, Kawasaki S** (2001) Providing the basis for genomics in *Lotus japonicus*: the accessions Miyakojima and Gifu are appropriate crossing partners for genetic analyses. *Mol Genet Genomics* **266**: 157–166
- Kussmann M, Nordhoff E, Rahbek-Nielsen H, Haebel S, Rossel-Larsen M, Jakobsen L, Gobom J, Mirgorodskaya E, Kroll-Kristensen A, Palm L, et al** (1997) Matrix-assisted laser desorption/ionization mass spectrometry sample preparation techniques designed for various peptide and protein analytes. *J Mass Spectrom* **32**: 593–601
- Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, Antonescu V, Chan A, Cheung F, Quackenbush J** (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res* **33**: D71–D74
- Liu H, White P** (1992) Oxidative stability of soybean oils with altered fatty acid compositions. *J Am Oil Chem Soc* **69**: 528–532
- Ma F, Cholewa E, Mohamed T, Peterson CA, Gijzen M** (2004) Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. *Ann Bot (Lond)* **94**: 213–228
- March JE, Pappin DJ, Casey R** (1988) Isolation and characterization of a minor legumin and its constituent polypeptides from *Pisum sativum* (pea). *Biochem J* **250**: 911–915
- Matheson NK, Richardson RH** (1976) Starch phosphorylase enzymes in developing and germinating pea seeds. *Phytochemistry* **15**: 887–892
- McHenry L, Fritz PJ** (1992) Comparison of the structure and nucleotide sequences of vicilin genes of cocoa and cotton raise questions about vicilin evolution. *Plant Mol Biol* **18**: 1173–1176
- Mifflin BJ, Lea PJ** (1977) Amino acid metabolism. *Annu Rev Plant Biol* **28**: 299–329
- Morey KJ, Ortega JL, Sengupta-Gopalan C** (2002) Cytosolic glutamine synthetase in soybean is encoded by a multigene family, and the members are regulated in an organ-specific and developmental manner. *Plant Physiol* **128**: 182–193
- Mubarak AE** (2005) Nutritional composition and antinutritional factors of mung bean seeds (*Phaseolus aureus*) as affected by some home traditional processes. *J Foodchem* **89**: 489–495
- Munier-Jolain N, Ney B** (1998) Seed growth rate in grain legumes. II. Seed growth rate depends on cotyledon cell number. *J Exp Bot* **49**: 1971–1976
- Newbigin E, deLumen B, Chandler P, Gould A, Blagrove R, March J, Kortt A, Higgins T** (1990) Pea convicilin: structure and primary sequence of the protein and expression of a gene in the seeds of transgenic tobacco. *Planta* **180**: 461–470
- Ozga JA, van Huizen R, Reinecke DM** (2002) Hormone and seed-specific regulation of pea fruit growth. *Plant Physiol* **128**: 1379–1389
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS** (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551–3567
- Prakash D, Misra PS** (1988) Protein content and amino acid profile of some wild leguminous seeds. *Plant Foods Hum Nutr* **38**: 61–65
- Rotenberg S, Andersen JO** (1980) The effect of dietary citrus pectin on fatty acid balance and on the fatty acid content of the liver and small intestine in rats. *Acta Agric Scand* **30**: 8–12
- Sáenz de Miera LE, Ramos J, Pérez de la Vega M** (2008) A comparative study of convicilin storage protein gene sequences in species of the tribe Viciaeae. *Genome* **51**: 511–523
- Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al** (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Res* **15**: 227–239
- Scott MP, Jung R, Muntz K, Nielsen NC** (1992) A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean. *Proc Natl Acad Sci USA* **89**: 658–662

- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504
- Shen N, Fehr W, Johnson L, White P (1997) Oxidative stabilities of soybean oils with elevated palmitate and reduced linolenate contents. *JAOCS* **74**: 299–302
- Shutov AD, Kakhovskaya IA, Braun H, Baumlein H, Muntz K (1995) Legumin-like and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene. *J Mol Evol* **41**: 1057–1069
- Somerville C, Browse J (1991) Plant lipids: metabolism, mutants, and membranes. *Science* **252**: 80–87
- Staswick PE, Hermodson MA, Nielsen NC (1984) Identification of the cystines which link the acidic and basic components of the glycinin subunits. *J Biol Chem* **259**: 13431–13435
- Stevenson DG, Doorenbos RK, Jane J, Inglett GE (2006) Structures and functional properties of starch from seeds of three soybean (*Glycine max* (L.) Merr.) varieties. *Stärke* **58**: 509–519
- Stoldt W (1952) Vorschlag zur vereinheitlichung der fettbestimmung in lebensmitteln. *Fette und Seifen* **4**: 206–207
- Tabe L, Hagan N, Higgins TJ (2002) Plasticity of seed protein composition in response to nitrogen and sulfur availability. *Curr Opin Plant Biol* **5**: 212–217
- Tingey SV, Walker EL, Coruzzi GM (1987) Glutamine synthetase genes of pea encode distinct polypeptides which are differentially expressed in leaves, roots and nodules. *EMBO J* **6**: 1–9
- Tolte D, Jaquinod M, Mangavel C, Passirani C, Saulnier P, Manon S, Teyssier E, Payet N, Avelange-Macherel MH, Macherel D (2007) Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *Plant Cell* **19**: 1580–1589
- Tzitzikas EN, Vincken JP, de Groot J, Gruppen H, Visser RG (2006) Genetic variation in pea seed globulin composition. *J Agric Food Chem* **54**: 425–433
- Vadivel V, Janardhanan K (2000) Chemical composition of the underutilized legume *Cassia hirsuta* L. *Plant Foods Hum Nutr* **55**: 369–381
- Wang TL, Hedley CL (1993) Genetic and developmental analysis of the seed. In R Casey, DR Davies, eds, *Peas: Genetics, Molecular Biology and Biotechnology*. CAB International, UK/John Innes Institute, Norwich, UK, pp 83–120
- Wang W, Scali M, Vignani R, Spadafora A, Sensi E, Mazzuca S, Cresti M (2003) Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* **24**: 2369–2375
- Weber H, Borisjuk L, Wobus U (2005) Molecular physiology of legume seed development. *Annu Rev Plant Biol* **56**: 253–279
- Weschke W, Baumlein H, Wobus U (1987) Nucleotide sequence of a field bean (*Vicia faba* L. var. minor) vicilin gene. *Nucleic Acids Res* **15**: 10065
- Wilson LA, Birmingham VA, Moon DP, Snyder HE (1978) Isolation and characterization of starch from mature soybeans. *Cereal Chem* **55**: 661–670
- Wind C, Häger K (1996) Legumin encoding sequences from the redwood family (Taxodiaceae) reveal precursors lacking the conserved Asn-Gly processing site. *FEBS Lett* **383**: 46–50
- Yaklich RW (2001) β -Conglycinin and glycinin in high-protein soybean seeds. *J Agric Food Chem* **49**: 729–735
- Zeeman SC, Umemoto T, Lue WL, Au-Yeung P, Martin C, Smith AM, Chen J (1998) A mutant of *Arabidopsis* lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. *Plant Cell* **10**: 1699–1712
- zur Nieden U, Manteuffel R, Weber E, Neumann D (1984) Dictyosomes participate in the intracellular pathway of storage proteins in developing *Vicia faba* cotyledons. *Eur J Cell Biol* **34**: 9–17