

*Seed development and protein accumulation patterns in faba bean (*Vicia faba*, L.)*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Warsame, A. O. ORCID: <https://orcid.org/0000-0002-9281-0443>, Michael, N., O'Sullivan, D. M. ORCID: <https://orcid.org/0000-0003-4889-056X> and Tosi, P. ORCID: <https://orcid.org/0000-0003-4171-6120> (2022) Seed development and protein accumulation patterns in faba bean (*Vicia faba*, L.). *Journal of Agricultural and Food Chemistry*, 70 (30). pp. 9295-9304. ISSN 0021-8561 doi: <https://doi.org/10.1021/acs.jafc.2c02061> Available at <https://centaur.reading.ac.uk/106478/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1021/acs.jafc.2c02061>

Publisher: American Chemical Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Seed Development and Protein Accumulation Patterns in Faba Bean (*Vicia faba*, L.)

Ahmed O. Warsame,* Nicholas Michael, Donal M. O'Sullivan, and Paola Tosi



Cite This: *J. Agric. Food Chem.* 2022, 70, 9295–9304



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

ABSTRACT: A major objective in faba bean breeding is to improve its protein quality by selecting cultivars with enhanced desirable physicochemical properties. However, the protein composition of the mature seed is determined by a series of biological processes occurring during seed growth. Thus, any attempt to explain the final seed composition must consider the dynamics of the seed proteome during seed development. Here, we investigated the proteomic profile of developing faba bean seeds across 12 growth stages from 20 days after pollination (DAP) to full maturity. We analyzed trypsin-digested total protein extracts from the seeds at different growth stages by liquid chromatography-tandem mass spectrometry (LC-MS/MS), identifying 1217 proteins. The functional clusters of these proteins showed that, in early growth stages, proteins related to cell growth, division, and metabolism were most abundant compared to seed storage proteins that began to accumulate from 45 DAP. Moreover, label-free quantification of the relative abundance of seed proteins, including important globulin proteins, revealed several distinct temporal accumulation trends among the protein classes. These results suggest that these proteins are regulated differently and require further understanding of the impact of the different environmental stresses occurring at different grain filling stages on the expression and accumulation of these seed storage proteins.

KEYWORDS: *faba bean, seed development, protein accumulation patterns, legumin, vicilin*

INTRODUCTION

Faba bean (*Vicia faba*, L.) has one of the highest seed protein content among legumes.¹ Therefore, it is considered a good candidate for meeting the increasing global demand for plant proteins. One of the current legume breeding objectives is improving the protein composition to suit quality requirements for human and animal nutrition.^{2,3} In faba bean, the predominant class of seed protein belongs to the globulin protein family, namely, legumin, vicilin, and convicilin.^{3–5} Each of these protein classes is encoded by several structural genes.^{6–8} From a nutritional point of view, legumins are of particular interest due to their higher content of sulfur-containing amino acids (S-AA) compared to other globulins.³ Therefore, cultivars of legumes with higher proportions of sulfur-rich proteins will be of great interest in human and animal nutrition. On the other hand, protein properties such as solubility, heat-induced gelation, foaming, and emulsifying capacity are essential for the food processing industry. However, although there is currently a lack of direct evidence in faba bean, studies in other species show that the accumulation of seed storage proteins is a complex biological process regulated by genes involved in transcription, synthesis, and transport.^{9–11} Thus, a better understanding of seed development and the dynamics of protein deposition is critical for the genetic manipulation of protein composition by conventional breeding, genetic engineering, or gene editing.

Temporal differences in the expression of major seed storage proteins during seed development have been reported in faba bean and other legume species. For instance, in developing faba bean seeds, De Pace et al.¹² reported that the

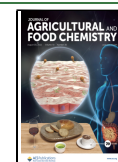
accumulation of certain vicilin protein bands preceded legumin by 4 days, while the legumin A-type bands were observed before those of legumin B-type. In addition, Panitz et al.¹³ showed that the cDNA of legumin B4 and vicilin had a diphasic accumulation pattern. These storage proteins started accumulating at 6 DAP and then disappeared before they started accumulating again at the cotyledon expansion stage. Differences in the timing of protein accumulation or gene expression for specific storage proteins during seed development have also been reported in other legumes, such as Medicago,^{14,15} pea,¹⁶ and soybean.¹⁷ From a nutritional standpoint, temporal differences in protein composition during seed development mean that changes in the environmental conditions at certain seed filling stages would modulate the protein composition and quality in the mature seeds. According to Henriët et al.,¹⁸ under sulfur-deficient growth conditions, the accumulation of legumin proteins was significantly reduced in mature pea seeds. This reduction was accompanied by downregulation of the expression of two legumin genes during seed development. In addition, in soybean, the abundance of several seed storage proteins was reported to vary considerably depending on whether genotypes were grown in the field or a glasshouse,¹⁹ indicating changes in

Received: March 24, 2022

Revised: June 24, 2022

Accepted: June 28, 2022

Published: July 21, 2022



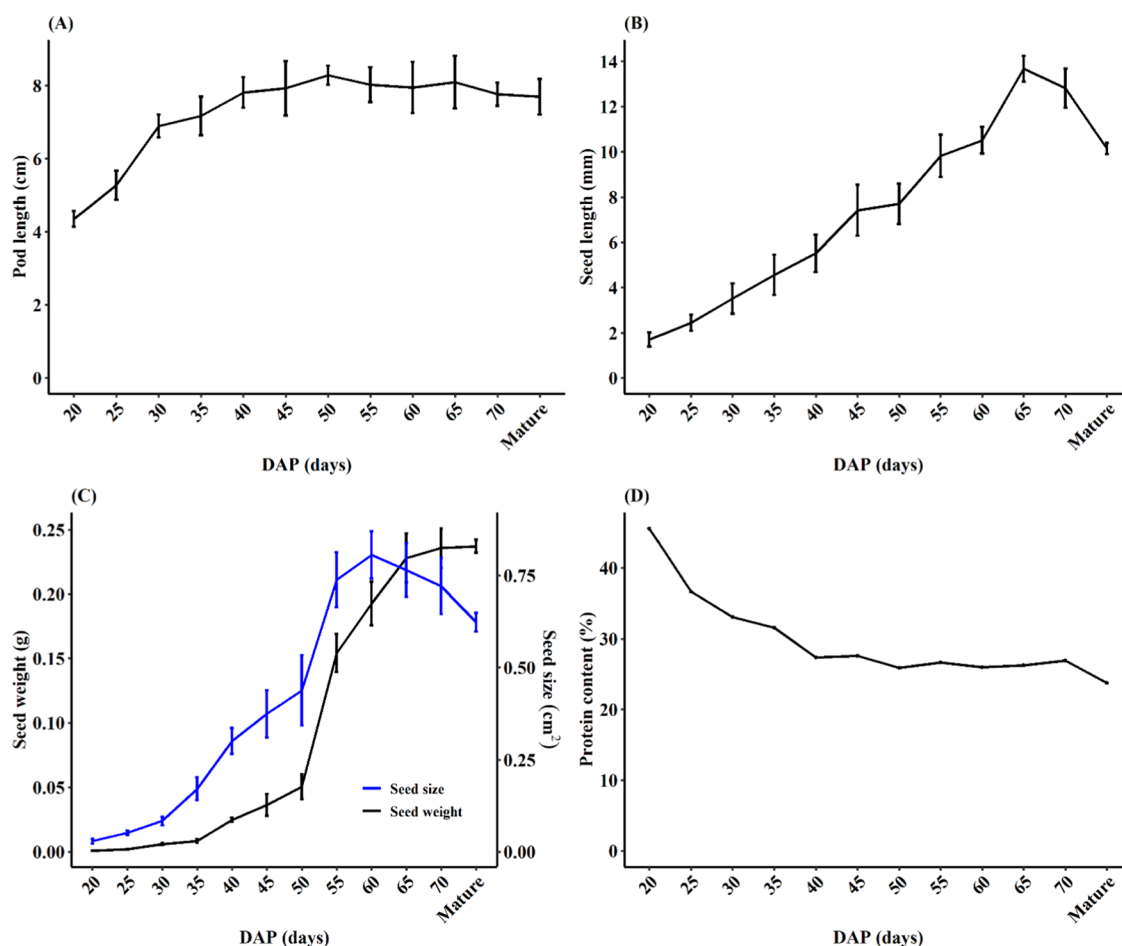


Figure 1. Characteristics of developing faba bean seeds. (A) Pod lengths between 20 days after pollination (DAP) as measured on 5–10 pods per growth stage. (B) Fresh seed length. (C) Weight and size of freeze-dried seeds. (D) Protein content on a dry weight basis for 12 growth stages. The error bars in (A–C) represent mean \pm SD.

protein composition due to prevailing conditions during seed growth. Therefore, better characterization of seed development and accumulation patterns of different storage proteins is the first step toward the rational modification of seed protein composition through genetics or agronomic practices.

Previous studies on the protein accumulation patterns in developing faba bean seeds development and were limited by the number of proteins studied and the sensitivity of the techniques used. For instance, De Pace et al.¹² used one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D SDS-PAGE) to investigate the accumulation of few vicilin and legumin bands. On the other hand, Panitz et al.¹³ studied vicilin and legumin accumulation using a single cDNA for each protein. This study aimed to characterize faba bean seed development and the associated temporal trends in seed protein composition with particular emphasis on storage proteins belonging to legumin, vicilin, and convicilin. We used LC-MS/MS and label-free quantification to identify and quantify seed proteins across 12 seed developmental stages.

MATERIALS AND METHODS

Plant Material and Growth Conditions. In this study, we used the inbred line Hedin/2, a spring faba bean line from Germany with a thousand seed weight of ~500 g.²⁰ This inbred line is of high genetic purity that has been used as a parent in multiple study populations.^{21–23} Also, this line has been used in faba bean genome sequencing project (<https://projects.au.dk/fabagenome/>).

Thirty single plants were grown in the glasshouse using 3 L pots containing homogeneously mixed compost (John Innes No. 2, Clover Peat, U.K.) at the Crop and Environment Laboratory (CEL) of the University of Reading, U.K. Plants were well-watered and received supplemental lighting to achieve 16 h of light per day using high-pressure sodium lamps providing a photosynthetic photon flux density of about $600 \mu\text{mol m}^{-2} \text{s}^{-1}$. The experiment was conducted during the winter season from November to February, with temperatures between 12 and 24 °C, on average.

At the flowering stage, individual flowers were tagged, and the date of pollination date was recorded on the day when the standard petal curved back and spots on the wing petals were visible without undue effort. Flowers were also tripped by hand during tagging to ensure high pod set rates. After successful fertilization, pods were sampled at 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, and 70 DAP. Pods were carefully removed from plants; the number of the sampled node was recorded and then immediately frozen in liquid nitrogen and stored at $-80 \text{ }^{\circ}\text{C}$ until further analysis. The mature seed sample was obtained from pods that naturally dried on the plants.

Pod and Seed Development Measurements. At the end of the sampling process, pods belonging to each developmental stage were combined after visually examining any apparent off-types, such as pods with outlier sizes for the particular growth stage. Before removing seeds, pods were photographed along a 15 cm long ruler with a 1 mm scale. The same was done for the seeds before and after freeze-drying. Images were then processed with ImageJ software,²⁴ where pod and seed lengths and seed size (based on the area of seeds laying on one of the cotyledons) were measured. Sample details are given in Table S1.

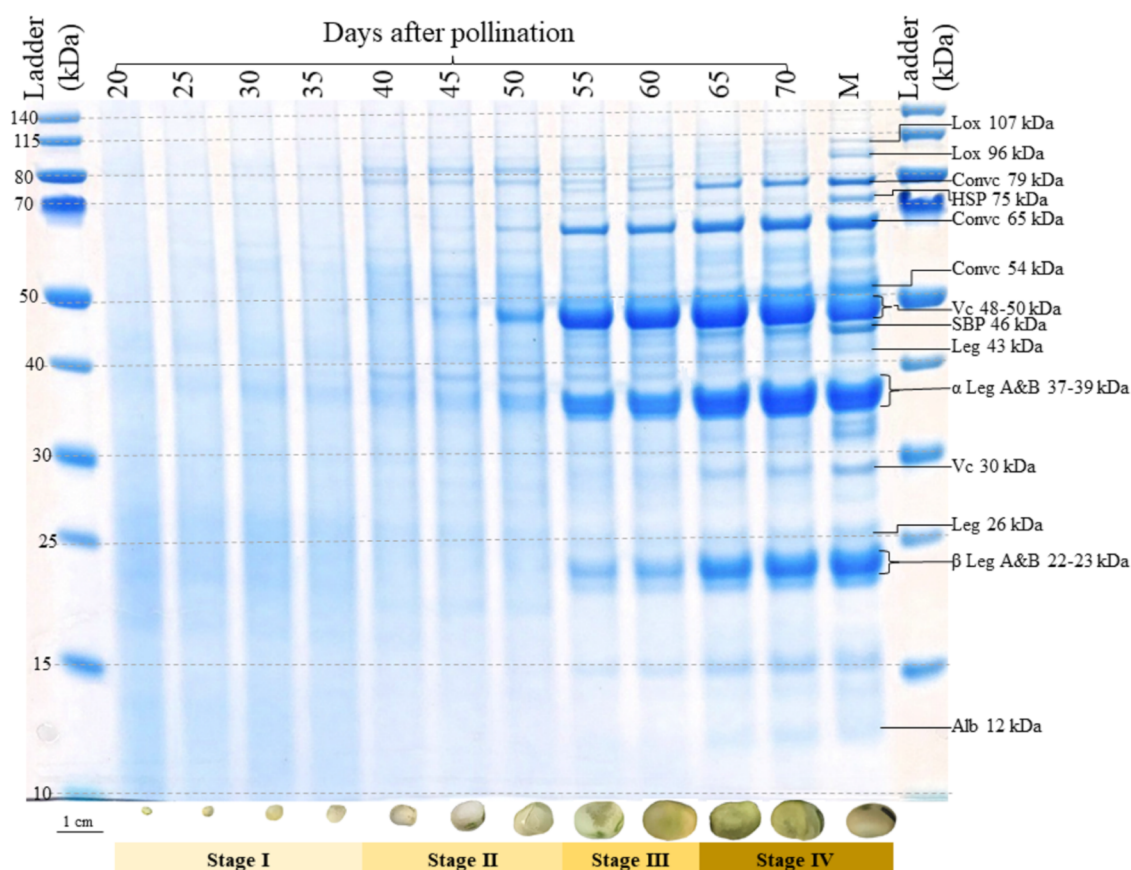


Figure 2. 1D SDS-PAGE gel showing the protein profile of faba bean seeds harvested at 12 developmental stages. Molecular weights of individual bands are estimated with respect to the bands of the MW ladder in the leftmost lane, with sizes of the marker bands given in kDa. Abbreviated names of discrete and most abundant protein bands are based on mass spectroscopic identification of proteins in Warsame et al.⁴ and are listed in Table S5. At the bottom of each lane, sample images of freeze-dried seeds belonging to the pool representing each growth stage are shown; colored bars along the bottom of the gel denote the observed main phases of protein accumulation. A 1 cm scale bar for seed images is given on the bottom left.

Crude Protein Content Analysis. Seeds were freeze-dried (Super Modulyo, Edwards Vacuum, West Sussex, U.K.) until a constant weight was obtained (~78 h) and then ground to homogeneous powder using a mortar and pestle. Then, depending on the sample availability, ~60 to 100 mg of seed flour was analyzed in a LECO carbon/hydrogen/nitrogen determinator (628 Series, LECO). Subsequently, nitrogen content was converted to protein content using the 5.4 conversion factor.²⁵ Due to the limited sample, the protein content at 20 DAP was not measured but imputed from the data points of all five traits described in Figure 1 using the mice R package.²⁶

Total Protein Extraction. Protein extraction was conducted as described by Scollo et al.²⁷ Briefly, to remove phenolic compounds, a cold (~4 °C) aqueous acetone (80% v/v) containing 5 mM sodium ascorbate was added to the samples at a 1:20 sample to buffer ratio. Then, the suspension was vortexed for 1 min and centrifuged at 1500g for 10 min at 4 °C. This step was done twice, and the supernatant was discarded each time. The sample was then washed with cold acetone and air-dried. Next, total proteins were extracted with a solution containing 7 M urea, 2 M thiourea, and 20 mM dithiothreitol (DTT). After shaking the suspension at 300 rpm for 1 h at room temperature, samples were centrifuged at 1500g for 15 min at 4 °C, and the supernatant was collected. The protein concentration of the samples was assessed with the Bradford method²⁸ using a SpectraMax i3x microplate reader (Molecular Devices, U.K.). These samples were separated on NuPAGE 10% Bis–Tris precast SDS-PAGE gels as described by Warsame et al.⁴

Trypsin Digestion. From the total protein extract for each developmental stage, technical triplicates containing ~0.2 mg of

protein were created. Then, 50 mM ammonium bicarbonate was added to each tube to bring the urea concentration to ~1 M. Additional DTT was added to each tube at 10 mM final concentration and then was incubated for 30 min at 37 °C. Next, sulfhydryl groups were alkylated by the addition of iodoacetamide (IAA) at a final concentration of 20 mM. For protein digestion, trypsin (Promega, U.K.) was added at a 1:100 protease/protein ratio (w/w) and the solutions were incubated overnight at 37 °C. The reaction was terminated by freezing samples at –20 °C. To confirm that all proteins in the samples were completely digested, aliquots of the digested samples were loaded on 1D SDS-PAGE gel. Finally, samples were dried in a centrifugal vacuum concentrator. The mass spectrometry analysis was conducted as described by Warsame et al.,⁴ with each of the three technical replicates for each sample being analyzed three times. Thus, nine MS/MS data files were obtained for each developmental stage, except for 60 DAP that was analyzed in duplicate and had six data files.

MS/MS Data Analysis. The raw MS/MS data were searched against the NCBI database using an in-house version of the MASCOT search engine (Matrix Science, U.K.) via Mascot Daemon with the file conversion performed using ProteoWizard. The MASCOT search parameters were set as described by Warsame et al.⁴ Then, to obtain a final list of protein accessions identified in each seed developmental stage, proteins identified in the technical replicates were merged, and the duplicated protein IDs were removed. For functional clustering of the identified proteins, protein sequences were functionally annotated using the MapMan4 framework and its associated online tool Mercator4.²⁹ Considering the relatively large number of the resultant

protein functional clusters obtained from MapMan4, we further grouped proteins into a few clusters according to Bevan et al.³⁰

To assess the relative abundance of proteins across seed developmental stages, we conducted a label-free quantification using the MaxLFQ algorithm³¹ implemented in MaxQuant software.³² In this analysis, to reduce the computational power requirements, a protein database containing the proteins identified by MASCOT analysis was used as a reference for quantification. The details of the quantification parameters are given in Table S2. The downstream analysis, including normalization, imputation, and clustering, was conducted in Perseus software.³³ Considering the differences in the number of proteins identified in each stage, we only used the proteins present in at least 7 of the 12 stages and 5 of the technical replicates for downstream analysis, including the imputation of missing data. The R package “pheatmap” (<https://cran.r-project.org/web/packages/pheatmap>) was used to visualize the normalized average relative abundances obtained from Perseus software.

We used MEGA X software³⁴ to infer the identified globulin proteins' evolutionary relationships using the neighbor-joining method with 10000 replicates and the “ggtree” R package³⁵ to visualize the data.

RESULTS AND DISCUSSION

Features of Developing Faba Bean Seeds. This study used the small-seeded reference inbred line Hedin/2, which carried flowers from nodes 8 to 27 on the main (and sole) stem with seven flowers per node on average. It is important to note that faba bean genotypes can vary significantly in the number of flowers per node, the number of pods, and seeds per pod.^{36,37} Therefore, aspects of the described developmental stages should be regarded as genotype- and environment-dependent. It is worth mentioning that this study does not investigate the very early stages of postfertilization development since collecting sufficient seed samples before 20 DAP was not feasible due to the tiny seed size of Hedin/2 (<2 mm) in the early growth stages.

To provide biological context to the changes in the proteomic profile of the developing seeds, we recorded some of the morphological and biochemical features of the developing seed and pod tissues between 20 DAP and full maturity (Figure 1). The early growth phase is marked by the rapid growth of pods, which nearly doubled in length between 20 and 35 DAP and followed by more gradual growth to reach the maximum length at 50 DAP (Figure 1A). This early rapid pod development before the onset of seed filling has also been observed in soybean³⁸ and serves to give ample space for the seed to develop. In contrast, seeds gradually increased in length almost linearly until 65 DAP (Figure 1B), which usually starts to lose water and shrink at maturity. On a dry seed basis, seed growth was characterized by a gradual increase in weight and area until 50 DAP and then a rapid increase in both parameters between 50 and 60 DAP (Figure 1C). At 60 DAP, the seed area peaked while dry weight increased substantially to reach its maximum at maturity (Figure 1C).

On the other hand, the percent crude protein content on a dry weight basis was the highest during the early stages of seed development but dropped rapidly from >40% to <28% by 40 DAP, remaining in the 26–28% range throughout the later seed developmental stages period (Figure 1D). A similar protein content pattern was reported in developing seeds of mung bean³⁹ and soybean.⁴⁰ The observed higher protein content (i.e., nitrogen content) in the early developmental stages is probably due to the higher concentration of structural proteins, enzymes, free amino acids, and other nitrogen-containing compounds associated with rapid cell division

during these early stages. Also, at the early growth stages, other bulk seed components, chiefly starch, which normally dilutes protein content in the final stages, may not yet have accumulated in the seeds. In the mature seeds, however, the 24% protein content was comparable to the average protein content of Hedin/2 plants grown in the field (data not shown).

1-D SDS-PAGE Protein Profile of the Seed Developmental Stages. The seed protein profile on a 1-D SDS-PAGE gel shows a clear timeline of the accumulation of different seed proteins across the 12 growth stages (Figure 2). The 20–35 DAP period is characterized by a greater apparent abundance of lower molecular weight proteins (<40 kDa) with no visible accumulation of major storage proteins. This period coincides with the stage of rapid seed and pod growth in Figure 1 and is denoted as stage I in Figure 2. From 40 to 50 DAP, seeds transition to the protein accumulation phase (stage II), and some seed storage protein subunits, including convicilin 65 kDa, vicilin 48–50 kDa, and α -legumin A&B 37–39 kDa, are visible by 50 DAP. At 55–60 DAP (stage III), vicilin 48–50 kDa appears to be more abundant than legumins, which is consistent with findings of De Pace et al.,¹² who reported an earlier accumulation of vicilin bands compared to legumins. Finally, seeds enter a phase of accelerated protein accumulation at 65 DAP throughout which the protein profile is quite comparable to that of the mature seed (Figure 2).

Functional Categories of Proteins Identified across Seed Developmental Stages. A total of 1217 proteins were identified by mass spectrometry analysis in the seed samples from 20 DAP to mature seeds (Table S4). This list included 36 of 104 proteins previously identified in protein bands of mature seeds of three different faba bean accessions.⁴ This relatively lower overlap between the two experiments can be attributed to differences in the sample characteristics of the two experiments. The LC-MS analysis of individual protein bands from 1D SDS-PAGE gel slices by Warsame et al.⁴ is more sensitive in detecting low abundant proteins compared to the whole proteome analysis conducted in this study, where the major storage proteins can mask the less abundant protein species. Traditionally, to identify the maximum absolute number of protein species in the total protein extracts from highly complex samples, a fractionation or depletion of the abundant proteins is performed before LC-MS analysis.²⁷ However, since this study aimed to compare the unaltered relative abundances of seed proteins across seed developmental stages with the main interest in the dynamics of the most abundant storage proteins, fractionation or depletion was not considered appropriate. Also, considering the apparent gradient in the protein profile shown in Figure 2, fractionation of the samples would have produced vastly different outcomes in different samples. Compared with other studies on seed proteomics, where the more sensitive nanoflow LC-MS/MS was used, 1168 proteins were identified in mature barley seeds⁴¹ and 704 in total protein extract of cocoa beans.²⁷ This variation underlines the variability in the number of proteins detected among species, which may be affected by sample preparation methods or settings used for the LC-MS analysis.

Proteins identified across the developmental stages could be assigned to 14 functional categories (Figure 3A) using MapMan4, which is a framework specifically designed for plant protein classification.²⁹ After further grouping of the protein clusters, according to Bevan et al.,³⁰ 34% of the protein sequences could not be assigned to any protein category, while the three largest groups, totaling 30% of all proteins identified,

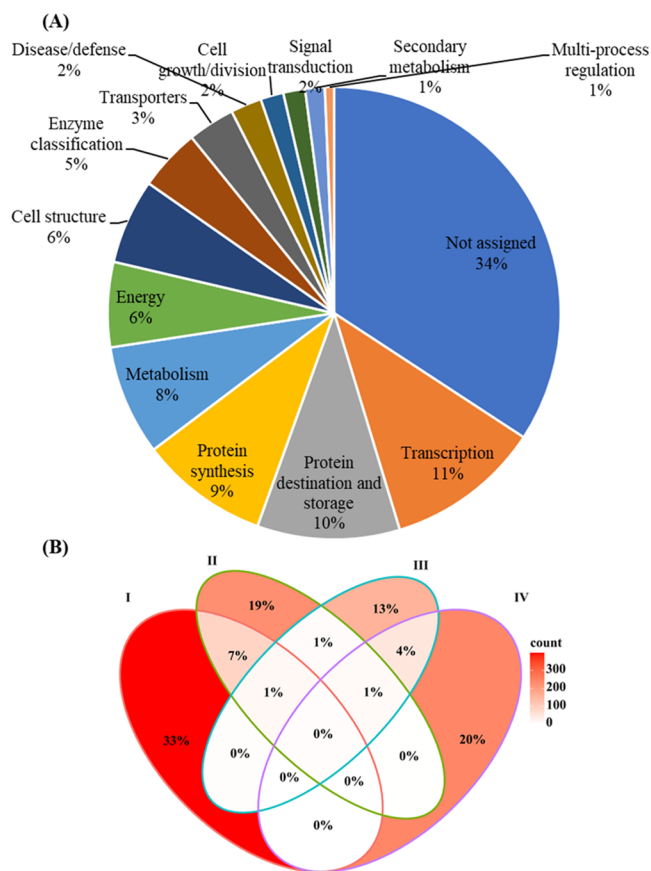


Figure 3. (A) Functional categories of the total list of 1217 proteins identified in 12 faba bean seed growth stages from 20 DAP to the mature stage. (B) Venn diagram showing the percentage of the total number of identified proteins that is specific or common among four developmental stages: I (20–35 DAP), II (40–50 DAP), III (55–60 DAP), and IV (65 to full maturity). The heat scale indicates the strength of overlap (red = high and white = low).

were proteins functioning in transcription, protein destination, and storage, and protein synthesis at 11, 10, and 9%, respectively. These numbers are consistent with the nature of the developing seed, particularly the cotyledon, as a storage organ undertaking active protein synthesis and deposition.

Looking at the developmental stage specificity of the identified proteins, we found that the overall overlap in the protein profile was strikingly low, and the majority of proteins identified at each stage were found only in that stage. For instance, there were 33, 19, 13, and 20% of the proteins specific to stages I (20–35 DAP), II (40–50 DAP), III (55–60 DAP), and IV (65 to full maturity), respectively (Figure 3B). The far greater proportion of the proteins identified in the early stages (ca. 33%) compared to later stages is consistent with the SDS-PAGE patterns in Figure 2. This figure showed a transition from a protein smear caused by many proteins of every possible size at 20 DAP to a clear banding pattern featuring a small number of very abundant protein species. The maximum number of proteins in common between different stages was 7%, shared between stages I and II. This is expected due to shifts in seed proteome across the seed life cycle from actively dividing cells at early growth stages through nutrient (mainly starch and protein) accumulation phase and, eventually, desiccation and dormancy at maturity.⁴²

Growth stages also differed in the dominant protein functional classes. Interestingly, although the major storage proteins did not accumulate significantly before 45 DAP (Figure 2), there was a high relative abundance of proteins involved in protein synthesis and protein destination and storage in the early stages (Figure 4). These earlier stages also had higher percentages of proteins involved in and energy production pathways (Figure 4), which agreed with expression patterns reported in soybean seed proteins³⁸ and their gene cDNAs.¹⁷ It is worth noting that some of the differences in the representation of some protein clusters at certain growth stages could be due to underlying sample properties, particularly in the later developmental stages. At stages III and IV, storage proteins are overwhelmingly abundant, and our method may not have detected the less abundant protein classes.

Temporal Trends in the Relative Abundance of Seed Proteins during Seed Development. As shown in Figure 2, individual proteins could vary progressively in abundance across seed growth and development stages. Thus, label-free quantification was performed to quantify these trends, resulting in calculating the relative abundances of 344 proteins across the developmental time course. Generally, accurate quantification of hundreds or thousands of proteins in a sample is a major challenge in proteomic studies. Each quantification method has its limitations and advantages in terms of cost, sample preparation requirements, and sensitivity.⁴³ Here, considering the relative simplicity of the approach taken, the number of proteins identified and quantified was considered more than satisfactory. Additionally, as mass spectrometry is becoming the gold standard method for proteomic works, standard-flow LC-MS/MS has been regarded as a potential alternative to the more sensitive but expensive nano-LC-MS/MS systems.^{44,45}

Quantified proteins showed characteristic differences in their relative abundance patterns across seed developmental stages and could be clustered accordingly into six groups (Figure 5). For simplicity, the proteins will be referred to by their generic names or accession numbers, as they are not yet adequately annotated in faba bean. Details of the quantified 344 proteins are summarized in Table S5.

At the early growth stages (20–35 DAP), proteins with the highest relative abundance were nonstorage proteins (cluster 1), including those involved in transcription like histones H2A and H2B (gil593699727 and gil470116864, respectively) and energy production, such as glyceraldehyde-3-phosphate dehydrogenase (gil462138). Also, among seven chaperones identified in this analysis, the heat shock protein (gil473217) was consistently upregulated until 45 DAP. This protein is 87% identical to the other HSP71.2 (gil562006), which was previously found abundant in the dry, mature faba bean seeds.⁴ In pea, isoforms of these HSP proteins (PsHSP71.2, PsHSC71.0, and PsHSP70b) were identified in different plant tissues and considered to play distinct biological functions related to repairing and preventing cell damage and acquiring adaptive thermotolerance.⁴⁶

As seeds developed, cluster 2 proteins reached their highest intensity, mainly between 45 and 60 DAP. Though the majority of proteins in this group were not functionally annotated, the group included pectin acetyltransferase 8-like (gil356558882), annexin-like (gil459649445, gil828335547), and putative copper-transporting ATPase 3 (gil734387082). On the other hand, proteins in cluster 3 mainly consisted of seed storage proteins, which started to accumulate around 45 DAP

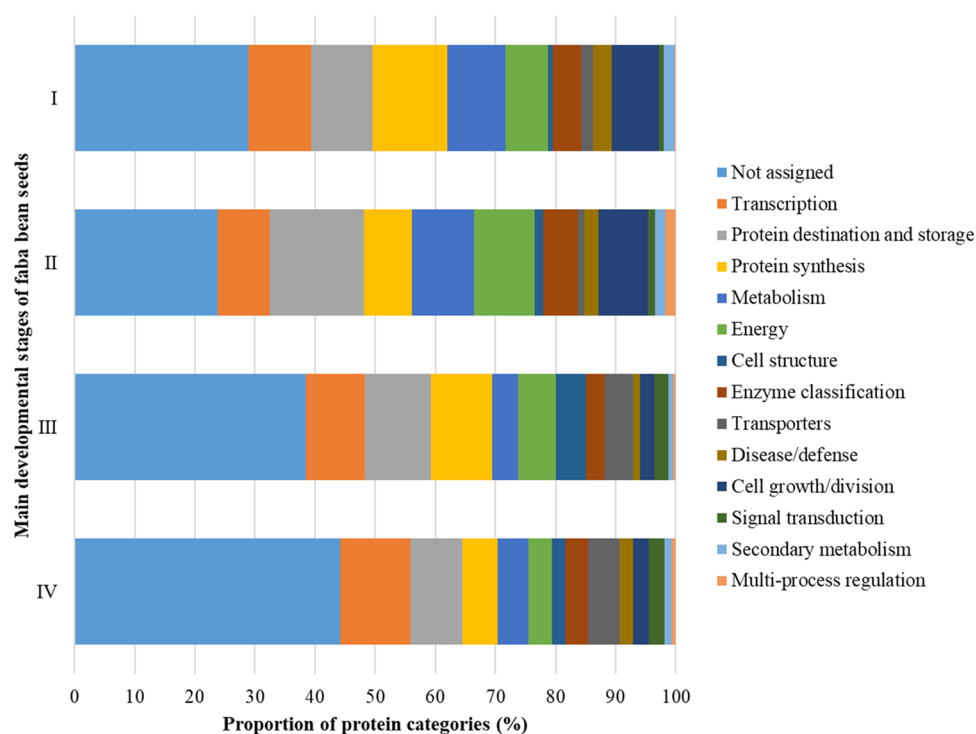


Figure 4. Changes in the functional categories of faba bean seed proteins through four major developmental stages.

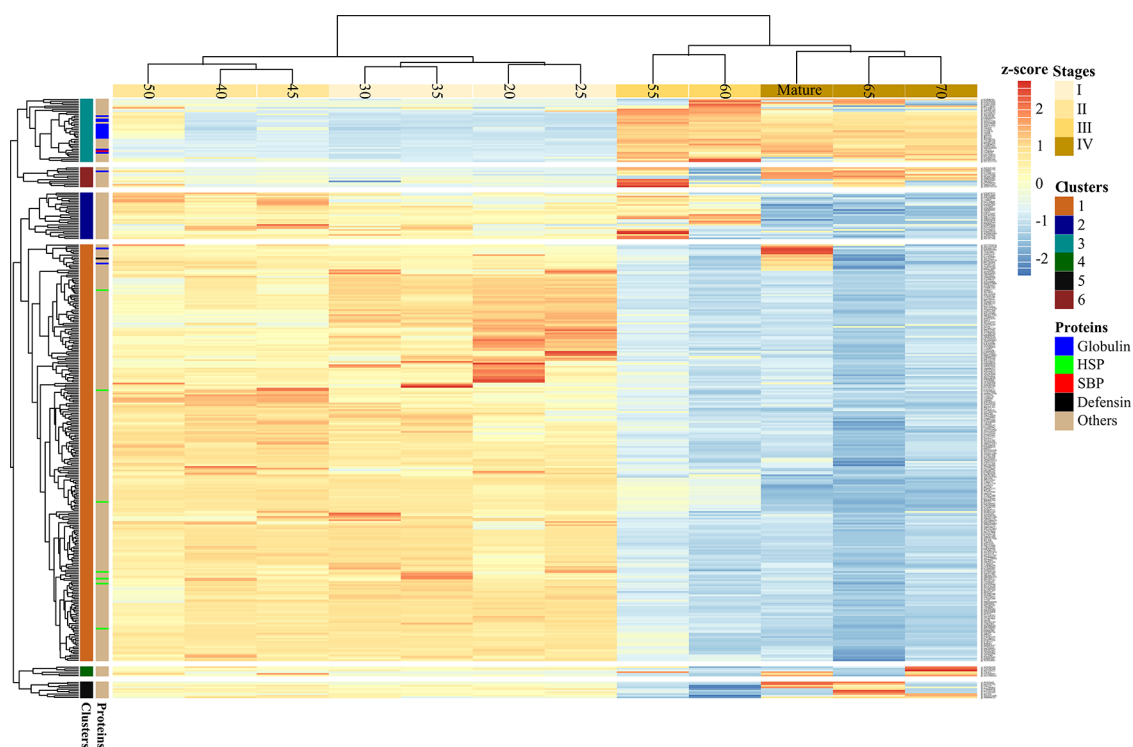


Figure 5. Heatmap showing the relative abundances of 344 proteins between 20 DAP to maturity. The four major developmental phases are indicated by gradient colors ranging from light to dark yellow. Color-coded protein families are globulins (including legumins, vicilins, and convicilins), heat shock proteins (HSPs), sucrose-binding protein (SBP), defensin, and others.

and continued steadily until maturity. Other nonstorage proteins which were differentially accumulated during late grain filling included a sodium/hydrogen exchanger (gil29539109). This protein has been reported to play a critical role in protein trafficking and the biogenesis of protein storage vacuoles (PSV) in *Arabidopsis*.⁴⁷ Furthermore, a putative sugar

phosphate transporter (gil302854600), lectin–glucose complex (gil82408030), and other proteins with protein maintenance functions (gil971508673) were in high relative abundance concomitantly to the phase of high protein accumulation. The remaining clusters (4–6) contained fewer proteins with distinct accumulation patterns. For instance,

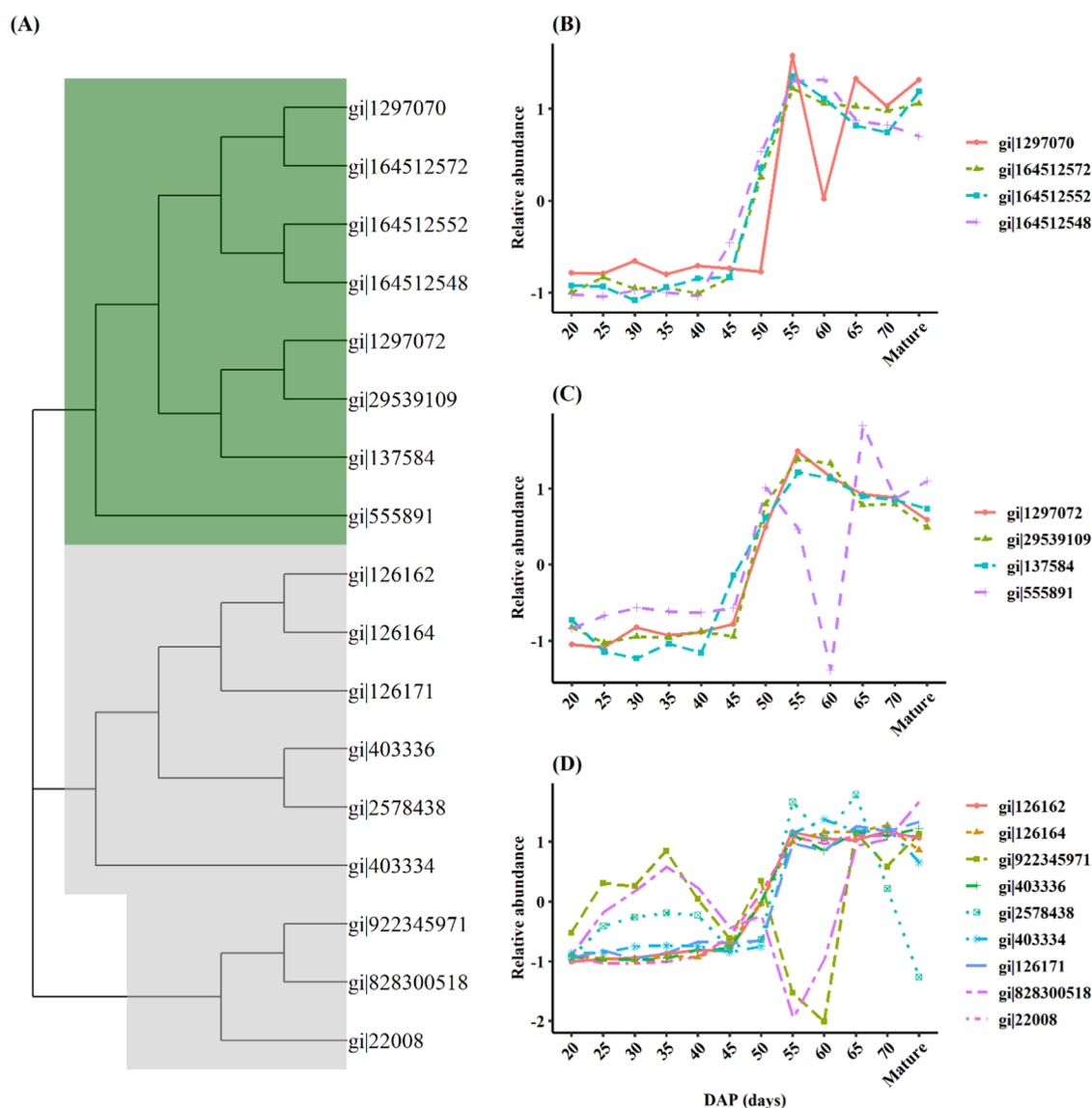


Figure 6. Patterns in the relative abundances of 17 globulin proteins across 12 seed growth stages between 20 DAP to maturity. These proteins are a subset of cluster 3 described in Figure 5. (A) Phylogenetic tree showing relationships between the different globulin genes, where vicilin/convicilins nodes are highlighted with dark green and legumins with gray color (B–D).

proteins in cluster 5 were upregulated mainly at late stages (65 DAP to maturity) and included proteins that may function in defense and stress adaptation, such as ascorbate peroxidase (gil731359393) and mitochondrial chaperonin (gil461736). Another cluster 5 member, a serine protease inhibitor (gil308800626), which is part of a gene family that regulates endogenous proteolysis in seeds and cell death during plant development and senescence,⁴⁸ was also upregulated at 70 DAP and maturity.

Diversity in the Accumulation Patterns among Major Storage Proteins. As previously noted, the globulins were primarily (though not exclusively) found in cluster 3 (Figure 5), characterized by a high accumulation in the late stages of grain filling. However, subtle differences in the onset and peak time of an individual protein could be highly significant given the very high absolute abundance of these storage proteins. Therefore, we compared the relative abundances of 17 globulins, including convicilins, vicilins, and nine legumins,

to reveal the differences in the accumulation patterns within and between these protein classes (Figure 6).

Convicilins and vicilins are collectively called 7S globulins, with convicilin being considered a subunit of vicilin.^{49,50} This hypothesis is supported by their phylogenetic relationship, where they form a separate cluster (Figure 6A). In general, proteins belonging to this globulin group had similar accumulation patterns. They started to accumulate from 45 DAP and reached their maximum abundance between 55 and 60 DAP (Figure 6B,C), after which their relative abundance declined gradually. However, at 60 DAP, one convicilin (gil1297070) and a vicilin-like protein (gil555891) were significantly reduced, with the latter being phylogenetically distant from the other 7S globulins (Figure 6A). The earlier accumulating vicilin (gil137584) was also the most abundant in major protein bands of 48–50 kDa in the mature faba bean seeds.⁴ Based on its apparent molecular weight, this protein is likely to be the vicilin band reported to be synthesized 4 days earlier than legumins by De Pace et al.¹²

On the other hand, legumin proteins were more diverse in their accumulation patterns (Figure 6D). Of the nine legumin-type proteins, 6 had similar accumulation trends, where they started to deposit between 45 and 50 DAP and remained unchanged until maturity (Figure 6D). A biphasic accumulation pattern characterized the other three legumin-type proteins; they had higher relative abundance during early growth stages (25–45 DAP) but downregulated differently in the later growth stages. In particular, the sharp downregulation of two legumins (gil922345971 and gil828300518) between 45 and 60 DAP is intriguing and difficult to explain. A possible reason could be that the transcription of these genes could be sensitive to the prevailing cell conditions at this period. For instance, it has been reported that high sucrose content in the cell reduces the stability of legumin mRNA, leading to significantly reduced legumin protein wrinkled peas.⁵¹ However, this needs further investigation.

Nonetheless, the different temporal trends in the relative abundance of these globulin proteins confirm and expand the current knowledge on the accumulation of globulins in faba bean and other legumes. As was briefly mentioned in the Introduction section, De Pace et al.¹² reported that legumin A subunits appeared to accumulate 2 days before the legumin B-type. In pea, gene expression analysis of developing seeds showed that legumin B-type gene (Psat3g055960) is highly expressed early (16 DAP) compared to legumin A-type (Psat3g058800), which had a maximum expression at 19–23 DAP.¹⁶ Similarly, the expression pattern of legumin K in *Medicago truncatula* indicated that it was synthesized earlier (~16 DAP) compared to legumin A (24 DAP).¹⁵ Overall, the observed diversity in the timing of expression within the classes of storage proteins may be indicative of a complex regulatory system. Thus, we need a better understanding of the genetic basis controlling its fine-tuning and how it impacts the nutritional quality of the seeds.

In conclusion, the different accumulation patterns within and among seed protein classes found in this study suggests that a corresponding diversity may exist in transcriptional, translational, and post-translational regulation of faba bean seed protein expression and accumulation. This opens new avenues for further investigations into the identification of master regulatory factors driving the developmental switch from cell division and growth to protein deposition. For instance, the characterization of promoter sequences that may mediate differential responses to these master regulators can aid in selecting the best targets for genetic improvement of nutritional composition in faba bean. As, by far, the most comprehensive survey of faba bean seed protein accumulation patterns, the results of this study can contribute to the annotation of the currently in production Hedin/2 genome assembly. The availability of a complete reference genome sequence accompanied by detailed expression data will, in turn, permit the identification of important regulatory motifs of seed proteins.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jafc.2c02061>.

Details of developing *V. faba* seed samples used for proteomic analysis (Table S1); summary of the parameter settings used in MaxQuant software for

label-free quantification of relative protein abundances (Table S2); details of seed protein band labels in Figure 2 that were previously identified by LC-MS (Warsame et al.⁴) (Table S3); combined list of all 1217 proteins identified across four major growth stages (Table S4); list of the 344 proteins quantified by MaxQuant across 12 developmental stages of *V. faba* seeds (Table S5) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Ahmed O. Warsame – School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AH, U.K.; orcid.org/0000-0002-9281-0443; Email: warsame@nbi.ac.uk, a.o.warsame105@gmail.com

Authors

Nicholas Michael – School of Chemistry, Food and Pharmacy, University of Reading, Reading RG6 6AH, U.K.

Donal M. O'Sullivan – School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AH, U.K.

Paola Tosi – School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AH, U.K.

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jafc.2c02061>

Funding

The financial support from Islamic Development Bank (IsDB) to A.O.W. was greatly acknowledged. The authors also thank for the additional support from the Rank Prize Nutrition COVID-19 grant received during the preparation of this manuscript.

Notes

The authors declare no competing financial interest.

[§]Department of Biochemistry and Metabolism, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, U.K.

■ ACKNOWLEDGMENTS

The authors thank Deepti Angra and the staff at the Crop and Environment Laboratory (CEL) of University of Reading for their technical support during this work.

■ ABBREVIATIONS

DAP, days after pollination; 1D SD-PAGE, one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis; kDa, kilo Dalton; MW, molecular weight; MS, mass spectrometry; LC, liquid chromatography

■ REFERENCES

- (1) Foyer, C. H.; Lam, H. M.; Nguyen, H. T.; Siddique, K. H.; Varshney, R. K.; Colmer, T. D.; Cowling, W.; Bramley, H.; Mori, T. A.; Hodgson, J. M.; Cooper, J. W.; Miller, A. J.; Kunert, K.; Vorster, J.; Cullis, C.; Ozga, J. A.; Wahlqvist, M. L.; Liang, Y.; Shou, H.; Shi, K.; Yu, J.; Fodor, N.; Kaiser, B. N.; Wong, F. L.; Valliyodan, B.; Considine, M. J. Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* **2016**, *2*, 16112.
- (2) Robinson, G. H. J.; Balk, J.; Domoney, C. Improving pulse crops as a source of protein, starch and micronutrients. *Nutr. Bull.* **2019**, *44*, 202–215.
- (3) Warsame, A. O.; O'Sullivan, D. M.; Tosi, P. Seed storage proteins of faba bean (*Vicia faba*, L.): Current status and prospects for genetic improvement. *J. Agric. Food Chem.* **2018**, *66*, 12617–12626.

- (4) Warsame, A. O.; Michael, N.; O'Sullivan, D. M.; Tosi, P. Identification and quantification of major faba bean seed proteins. *J. Agric. Food Chem.* **2020**, *68*, 8535–8544.
- (5) Müntz, K.; Horstmann, C.; Schlesier, B. Vicia Globulins. In *Seed Proteins* Shewry, P. R.; Casey, R., Eds.; Springer Netherlands: Dordrecht, 1999; pp 259–284.
- (6) Horstmann, C.; Schlesier, B.; Otto, A.; Kostka, S.; Müntz, K. Polymorphism of legumin subunits from field bean (*Vicia faba* L. var. minor) and its relation to the corresponding multigene family. *Theor. Appl. Genet.* **1993**, *86*, 867–874.
- (7) Fuchs, J.; Schubert, I. Localization of seed protein genes on metaphase chromosomes of *Vicia faba* via fluorescence in situ hybridization. *Chromosome Res.* **1995**, *3*, 94–100.
- (8) Sáenz de Miera, L. E.; Ramos, J.; Pérez de la Vega, M. A comparative study of concivillin storage protein gene sequences in species of the tribe Viciaeae. *Genome* **2008**, *51*, 511–523.
- (9) Verdier, J.; Thompson, R. D. Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* **2008**, *49*, 1263–1271.
- (10) Le Signor, C.; Aimé, D.; Bordat, A.; Belghazi, M.; Labas, V.; Gouzy, J.; Young, N. D.; Prospero, J.-M.; Leprince, O.; Thompson, R. D.; Buitink, J.; Burstin, J.; Gallardo, K. Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. *New Phytol.* **2017**, *214*, 1597–1613.
- (11) Gacek, K.; Bartkowiak-Broda, I.; Batley, J. Genetic and molecular regulation of seed storage proteins (SSPs) to improve protein nutritional value of oilseed rape (*Brassica napus* L.) seeds. *Front. Plant Sci.* **2018**, *9*, 890.
- (12) De Pace, C.; Delre, V.; Scarascia Mugnozza, G. T.; Maggini, F.; Cremonini, R.; Frediani, M.; Cionini, P. G. Legumin of *Vicia faba* major: accumulation in developing cotyledons, purification, mRNA characterization and chromosomal location of coding genes. *Theor. Appl. Genet.* **1991**, *83*, 17–23.
- (13) Panitz, R.; Borisjuk, L.; Manteuffel, R.; Wobus, U. Transient expression of storage-protein genes during early embryogenesis of *Vicia faba*: synthesis and metabolization of vicilin and legumin in the embryo, suspensor and endosperm. *Planta* **1995**, *196*, 765–774.
- (14) Gallardo, K.; Le Signor, C.; Vandekerckhove, J.; Thompson, R. D.; Burstin, J. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **2003**, *133*, 664–682.
- (15) Verdier, J.; Kakar, K.; Gallardo, K.; Le Signor, C.; Aubert, G.; Schlereth, A.; Town, C. D.; Udvardi, M. K.; Thompson, R. D. Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol. Biol.* **2008**, *67*, S67.
- (16) Kreplak, J.; Madoui, M.-A.; Cápál, P.; Novák, P.; Labadie, K.; Aubert, G.; Bayer, P. E.; Gali, K. K.; Syme, R. A.; Main, D.; Klein, A.; Bérard, A.; Vrbová, I.; Fournier, C.; D'Agata, L.; Belser, C.; Berrabah, W.; Toegelová, H.; Milec, Z.; Vrána, J.; Lee, H.; Kougbeadjo, A.; Térézol, M.; Huneau, C.; Turo, C. J.; Mohellibi, N.; Neumann, P.; Falque, M.; Gallardo, K.; McGee, R.; Tar'An, B.; Bendahmane, A.; Aury, J.-M.; Batley, J.; Le Paslier, M.-C.; Ellis, N.; Warkentin, T. D.; Coyne, C. J.; Salse, J.; Edwards, D.; Lichtenzweig, J.; Macas, J.; Doležel, J.; Wincker, P.; Burstin, J. A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* **2019**, *51*, 1411–1422.
- (17) Jones, S. I.; Gonzalez, D. O.; Vodkin, L. O. Flux of transcript patterns during soybean seed development. *BMC Genomics* **2010**, *11*, 136.
- (18) Henriët, C.; Aimé, D.; Térézol, M.; Kilandamoko, A.; Rossin, N.; Combes-Soia, L.; Labas, V.; Serre, R.-F.; Prudent, M.; Kreplak, J.; Vernoud, V.; Gallardo, K. Water stress combined with sulfur deficiency in pea affects yield components but mitigates the effect of deficiency on seed globulin composition. *J. Exp. Bot.* **2019**, *70*, 4287–4304.
- (19) John, K. M. M.; Khan, F.; Luthria, D. L.; Matthews, B.; Garrett, W. M.; Natarajan, S. Proteomic and metabolomic analysis of minimax and Williams 82 soybeans grown under two different conditions. *J. Food Biochem.* **2017**, *41*, No. e12404.
- (20) Dieckmann, S.; Link, W. Quantitative genetic analysis of embryo heterosis in faba bean (*Vicia faba* L.). *Theor. Appl. Genet.* **2010**, *120*, 261–270.
- (21) Khazaei, H.; O'Sullivan, D. M.; Stoddard, F. L.; Adhikari, K. N.; Paull, J. G.; Schulman, A. H.; Andersen, S. U.; Vandenberg, A. Recent advances in faba bean genetic and genomic tools for crop improvement. *Legum. Sci.* **2021**, *3*, No. e75.
- (22) Khazaei, H.; Stoddard, F. L.; Purves, R. W.; Vandenberg, A. A multi-parent faba bean (*Vicia faba* L.) population for future genomic studies. *Plant Genet. Resour.* **2018**, *16*, 419–423.
- (23) O'Sullivan, D. M.; Angra, D. Advances in faba bean genetics and genomics. *Front. Genet.* **2016**, *7*, 150.
- (24) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671.
- (25) Mosse, J. Nitrogen-to-protein conversion factor for ten cereals and six legumes or oilseeds. A reappraisal of its definition and determination. Variation according to species and to seed protein content. *J. Agric. Food Chem.* **1990**, *38*, 18–24.
- (26) van Buuren, S.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Software* **2011**, *45*, 1–67.
- (27) Scollo, E.; Neville, D.; Oruna-Concha, M. J.; Trotin, M.; Cramer, R. Characterization of the proteome of *Theobroma cacao* beans by nano-UHPLC-ESI MS/MS. *Proteomics* **2018**, *18*, No. 1700339.
- (28) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.
- (29) Schwacke, R.; Ponce-Soto, G. Y.; Krause, K.; Bolger, A. M.; Arsova, B.; Hallab, A.; Gruden, K.; Stitt, M.; Bolger, M. E.; Usadel, B. MapMan4: A refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* **2019**, *12*, 879–892.
- (30) Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.; Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S. A.; Patel, K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T. M.; Terry, N.; Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.; Lechary, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K. D.; Rieger, M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka, A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.; Piravandi, E.; Obermaier, B.; Hilbert, H.; Düsterhöft, A.; Moores, T.; Jones, J. D.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Delseny, M.; Voet, M.; Volckaert, G.; Mewes, H. W.; Klosterman, S.; Schueller, C.; Chalwatzis, N. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **1998**, *391*, 485–488.
- (31) Cox, J.; Hein, M. Y.; Lubner, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513–2526.
- (32) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372.
- (33) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13*, 731–740.
- (34) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
- (35) Yu, G.; Smith, D. K.; Zhu, H.; Guan, Y.; Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **2017**, *8*, 28–36.

- (36) Duc, G. Faba bean (*Vicia faba* L.). *Field Crops Res.* **1997**, *53*, 99–109.
- (37) Suso, M. J.; Moreno, M. T.; Mondragao-Rodrigues, F.; Cubero, J. I. Reproductive biology of *Vicia faba*: role of pollination conditions. *Field Crops Res.* **1996**, *46*, 81–91.
- (38) Li, L.; Hur, M.; Lee, J.-Y.; Zhou, W.; Song, Z.; Ransom, N.; Demirkale, C. Y.; Nettleton, D.; Westgate, M.; Arendsee, Z.; Iyer, V.; Shanks, J.; Nikolau, B.; Wurtele, E. S. A systems biology approach toward understanding seed composition in soybean. *BMC Genomics* **2015**, *16*, S9.
- (39) Sital, J. S.; Malhotra, J. S.; Sharma, S.; Singh, S. Comparative studies on biochemical components in mung bean [*Vigna radiata* (L.) Wilczek] varieties cultivated in summer and kharif seasons. *Indian J. Agric. Biochem.* **2011**, *24*, 68–72.
- (40) Zhang, H.; Hu, Z.; Yang, Y.; Liu, X.; Lv, H.; Song, B.-H.; An, Y.-Q. C.; Li, Z.; Zhang, D. Transcriptome profiling reveals the spatial-temporal dynamics of gene expression essential for soybean seed development. *BMC Genomics* **2021**, *22*, 453.
- (41) Mahalingam, R. Shotgun proteomics of the barley seed proteome. *BMC Genomics* **2017**, *18*, 44.
- (42) Dante, R. A.; Larkins, B. A.; Sabelli, P. A. Cell cycle control and seed development. *Front. Plant Sci.* **2014**, *5*, 493.
- (43) Brewis, I. A.; Brennan, P. Proteomics technologies for the global identification and quantification of proteins. *Adv. Protein Chem. Struct. Biol.* **2010**, *80*, 1–44.
- (44) Bian, Y.; Zheng, R.; Bayer, F. P.; Wong, C.; Chang, Y.-C.; Meng, C.; Zolg, D. P.; Reinecke, M.; Zecha, J.; Wiechmann, S.; Heinzlmeir, S.; Scherr, J.; Hemmer, B.; Baynham, M.; Gingras, A.-C.; Boychenko, O.; Kuster, B. Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS. *Nat. Commun.* **2020**, *11*, No. 157.
- (45) Distler, U.; Łącki, M. K.; Schumann, S.; Wanninger, M.; Tenzer, S. Enhancing Sensitivity of Microflow-Based Bottom-Up Proteomics through Postcolumn Solvent Addition. *Anal. Chem.* **2019**, *91*, 7510–7515.
- (46) DeRocher, A.; Vierling, E. Cytoplasmic HSP70 homologues of pea: differential expression in vegetative and embryonic organs. *Plant Mol. Biol.* **1995**, *27*, 441–456.
- (47) Wu, X.; Ebine, K.; Ueda, T.; Qiu, Q.-S. AtNHX5 and AtNHX6 Are Required for the Subcellular Localization of the SNARE Complex That Mediates the Trafficking of Seed Storage Proteins in Arabidopsis. *PLoS One* **2016**, *11*, No. e0151658.
- (48) Clemente, M.; Corigliano, M. G.; Pariani, S. A.; Sánchez-López, E. F.; Sander, V. A.; Ramos-Duarte, V. A. Plant serine protease inhibitors: Biotechnology application in agriculture and molecular farming. *Int. J. Mol. Sci.* **2019**, *20*, 1345.
- (49) Müntz, K. Proteases and proteolytic cleavage of storage proteins in developing and germinating dicotyledonous seeds. *J. Exp. Bot.* **1996**, *47*, 605–622.
- (50) O’Kane, F. E.; Happe, R. P.; Vereijken, J. M.; Gruppen, H.; van Boekel, M. A. J. S. Characterization of pea vicilin. 1. Denoting convicilin as the α -subunit of the *Pisum* vicilin family. *J. Agric. Food Chem.* **2004**, *52*, 3141–3148.
- (51) Casey, R.; Domoney, C.; Forster, C.; Hedley, C.; Hitchin, E.; Wang, T. The effect of modifying carbohydrate metabolism on seed protein gene expression in peas. *J. Plant Physiol.* **1998**, *152*, 636–640.

Recommended by ACS

Integrated Proteomics and Metabolomics Analysis Highlights Correlative Metabolite-Protein Networks in Soybean Seeds Subjected to Warm-Water Soaking

Cheol Woo Min, Sun Tae Kim, *et al.*

JULY 01, 2020

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

READ 

Improving Gene Annotation of the Peanut Genome by Integrated Proteogenomics Workflow

Haifen Li, Xuanqiang Liang, *et al.*

MAY 05, 2020

JOURNAL OF PROTEOME RESEARCH

READ 

Differential Proteomic Analysis to Identify Proteins Associated with Apomeiosis in *Boehmeria tricuspis* (Hance) Makino Using an iTRAQ-Based Strategy

Qing Tang, Jianguang Su, *et al.*

OCTOBER 27, 2020

JOURNAL OF PROTEOME RESEARCH

READ 

Proteomic Analysis of Desiccation Tolerance and Its Re-Establishment in Different Embryo Axis Tissues of Germinated Pea Seeds

Wei-Qing Wang, Song-Quan Song, *et al.*

MARCH 19, 2021

JOURNAL OF PROTEOME RESEARCH

READ 

Get More Suggestions >