Received: 10 June 2021 | Revised: 28 September 2021 | Accepted: 19 October 2021 | Accepted article online: 26 November 2021

DOI: 10.1002/pmic.202100143

Proteomics
Proteomics and Systems Biology

DATASET BRIEF

Quantitative proteomic analyses reveal the dynamics of protein and amino acid accumulation during soybean seed development

Nazrul Islam[1] | **Hari B. Krishnan**[2] | **Savithiry Natarajan**[1]

[1] Soybean Genomics and Improvement Laboratory, USDA Agricultural Research Service, Beltsville, Maryland, USA

[2] Plant Genetics Research Unit, USDA Agricultural Research Service, University of Missouri, Columbia, Missouri, USA

Correspondence
Savithiry Natarajan, Soybean Genomics and Improvement Laboratory, USDA Agricultural Research Service, 10300, Baltimore Avenue, Beltsville, MD 20705, USA.
Email: savi.natarajan@usda.gov

Funding information
Agricultural Research Service, USDA

Abstract

Using high throughput tandem mass tag (TMT) based tagging technique, we identified 4172 proteins in three developmental stages: early, mid, and late seed filling. We mapped the identified proteins to metabolic pathways associated with seed filling. The elevated abundance of several kinases was observed from the early to mid-stages of seed filling, indicating that protein phosphorylation was a significant event during this period. The early to late seed filling stages were characterized by an increased abundance of proteins associated with the cell wall, oil, and vacuolar-related processes. Among the seed storage proteins, 7S (β-subunit) and 11S (Gy3, Gy4, Gy5) steadily increased in abundance during early to late stages of seed filling, whereas 2S albumin exhibited a decrease in abundance during the same period. An increased abundance of proteases, senescence-associated proteins, and oil synthesis proteins was observed from the mid to late seed filling stages. The mid to late stages of seed filling was also characterized by a lower abundance of transferases, transporters, Kunitz family trypsin, and protease inhibitors. Two enzymes associated with methionine synthesis exhibited lower abundance from early to late stages. This study unveiled several essential enzymes/proteins related to amino acid and protein synthesis and their accumulation during seed development. All data can be accessed through this link: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=38784ecbd0854bb3801afc0d89056f84. (Accession MSV000087577)

KEYWORDS
development, mass spectrometry, pathways, proteome, seed, soybean

The metabolic network that controls the protein and amino acid synthesis and their accumulation remain poorly defined. A better understanding of metabolic control of synthesis and accumulation of proteins/amino acids in soybean seed will help breeders, geneticists, and plant physiologists to alter their metabolic pathways to produce seed with a more beneficial balance of protein and amino acids. Using TMT based tagging technique, we identified 4172 proteins and mapped them to metabolic pathways of amino acids and protein accumulation in developing soybean seeds.

Seeds of soybean line "M92-220" were obtained from Dr Robert Stupar at the University of Minnesota (St. Paul, MN). The seeds were planted in 2019 at the experimental field of University of Missouri/USDA (Bradford Research Center, Columbia). A complete randomized design with three biological replicates was adopted for the experiment. Cultural practices were typical of those utilized for soybean production in the Midwest USA. Developing seeds representing three distinct reproductive (R) developmental stages [1, 2] were

Abbreviations: TMT, Tandem Mass Tag; MS, Mass spectrometry; FN, Fast neutron; FDR, False Discovery Rate; ANOVA, Analysis of Variance; HSD, Honestly Significant Difference

collected and prepared for protein extraction as previously reported [3]. Proteins were prepared for mass spectrometry analysis as described earlier [4]. Mass spectrometric data were collected using a mass spectrometer (Orbitrap Fusion mass spectrometer, Thermo Fisher), coupled with a Proxeon EASY-nLC 1200 LC pump (Thermo Fisher) as previously described [5]. Peptide spectral matches were filtered to a <1% peptide, and <1% false discovery rate (FDR) of protein abundance was also employed using linear discriminant analysis and probabilistic scoring [6]. Proteins were quantified by summing reporter ion signal to noise for all matching PSMs as described previously [4, 5, 7]. Peptide matching multiple proteins were assigned to the protein with the most unique assigned matches. Peptides used for protein quantification were filtered to >100 signal to noise and >0.5 MS^2 isolation specificity.

We identified 4172 proteins and performed ANOVA and post-hoc test using the Tukey HSD procedure, followed by FDR at $p < 0.05$. A 1.2-fold change was then used as a minimum threshold to compare significant changes among the proteins found in the three stages of seed filling.

Three genes, oleosin family protein (Glyma.19G004800), aspartyl protease family protein (Glyma.07G262600), and seed maturation protein (Glyma.20G147600) were selected for gene expression analyses by probe RT-PCR technique, and the amount of the PCR product was measured by a standard curve equation ($R^2 = 0.982$ and slope = -3.7).

We identified proteins in all three stages, such as early, mid and late seed filling. The relative expression pattern of all proteins across the biological replicates is presented in Figure 1S, and their relative abundance is listed in Table 1S.

Based on the ANOVA analyses, the differentially expressed proteins were classified into seven groups (Figure 2S and Table 2S). As indicated, 1997 total proteins appeared in the group a (early to mid), followed by 1603 proteins in group b (early to late) and 1556 proteins in group c (mid to late). Interestingly, within each individual group, the highest number of differentially expressed proteins (641) was found in group b followed by c (144) and group a (15). Among all the proteins selected for analysis, a total of 362 protein were observed in group abc. This group has unique attributes because, unlike like other proteins, these proteins appeared in all three stages, and their abundance varied significantly at each stage.

The dynamics of protein abundance among the seven groups (1 to 7) across all three stages of seed filling is shown in Figure 1. These proteins are listed in supporting Tables a (Table 3S), b (Table 4S), c (Table 5S), ab (Table 6S), bc (Table 7S), ac (Table 8S), and abc (Table 9S).

In group a (1-early to mid-stage), 12 proteins (91%) showed higher and only three protein showed lower abundance (Figure 1, Table 3S). Among the 12 proteins that showed higher abundance between early to mid-stage of seed filling include carbohydrate kinase (1.59-fold), protein kinase (1.52-fold), glutathione transferases (1.39 and1.31-fold), amino acid transferase (1.30-fold), seipin (1.51-fold), leucine-rich repeats (LRR) (1.41-fold), dihydrosphingosine phosphate lyase (1.37-fold), and delta 1-pyrroline-5-carboxylate synthase 2 (1.34-fold). The lower abundant proteins include glucose-methanol-choline (GMC) oxidoreductase, peroxin 14, and PLC-like phosphodiesterases.

Unlike group a, 641 proteins exhibited differential protein abundance in group b (2- early to late stages) (Figure 1, Table 4S), of which 68 proteins (10%) showed higher abundance and 573 proteins (90%) showed lower abundance. The high abundant proteins include callose synthase five increased by 7.44-fold followed by dehydrin-late embryogenesis abundant (LEA) family (5.34-fold), oleosin family protein (4.48-fold), two beta vacuolar processing enzymes by 2.40-fold and 1.98-fold, in addition to other proteins associated with carbohydrate metabolism, protein phosphorylation, energy production, nucleotide metabolism, and proteases. On the other hand, the lower abundant proteins include an HSP20-like chaperones superfamily protein (0.15-fold), spermidine hydroxycinnamoyl transferase (0.18-fold), ascorbate peroxidase 3 (0.19-fold), pyrophosphorylase 4 (0.19-fold), and allene oxide cyclase 4 (0.19-fold), an enzyme that catalyses an essential step in jasmonic acid biosynthesis, this group also contain several photosystem proteins and protein processing proteins.

In group c (3- mid to late-stage) (Figure 1, Table 5S), 144 proteins were differentially expressed, of which only 11 proteins exhibited increased abundance, and 135 proteins showed lower abundance (Table 5S). Among the higher abundant proteins, aspartyl and cysteine proteases increased by 2-fold, followed by senescence protein (1.8-fold). Among the lower abundant proteins, two glutathione S-transferase tau 9 and 18 were showed lower abundance by 0.33-fold and 0.49-fold, respectively. Similarly, three HXXXD-type acyl-transferase family proteins exhibited lower protein abundance by 0.35, 0.42, and 0.51-fold, respectively. In addition, several other key proteins exhibited lower protein abundance, such as two Kunitz family trypsin and protease inhibitor protein by 0.51-fold and 0.68-fold, respectively. The SecY transport protein showed lower abundance to 0.52-fold, in addition to and several ribosomal proteins.

Group ab (4) (Figure 1, Table 6S) represents proteins that exhibited either higher or lower protein abundance between early to mid and early to late stages but insignificant changes between mid to late stage. In this group, 561 proteins were differentially expressed, of which only 95 proteins (17%) showed higher abundance, and 466 proteins (83%) exhibited lower abundance in both mid and late stages. Some of the most abundant proteins are 2-methylene-furan-3-one reductase (4 to 6-fold), urease (4 to 5-fold), Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily (3 to 5-fold), and methanethiol oxidase (2-fold). The lower abundant proteins are primarily associated with cell development processes and secondary metabolites production.

Group bc (5) (Figure 1, Table 7S) represents a class of proteins that showed significant change between early to late and mid to late stages. We identified 1009 proteins in this group, of which 858 proteins exhibited lower abundance at the late stage when compared to early and mid-stages. Some of the most lower abundant proteins include NAD(P)-linked oxidoreductase superfamily protein (~0.13-fold), chalcone-flavanone isomerase family protein (~0.09-fold), P-loop containing nucleoside triphosphate hydrolases superfamily protein (~0.12-fold), and TCP-1/cpn60 chaperonin family protein (~0.13-fold). Among the higher abundant proteins, seed maturation protein increased by 156-fold between early and late stages, followed by

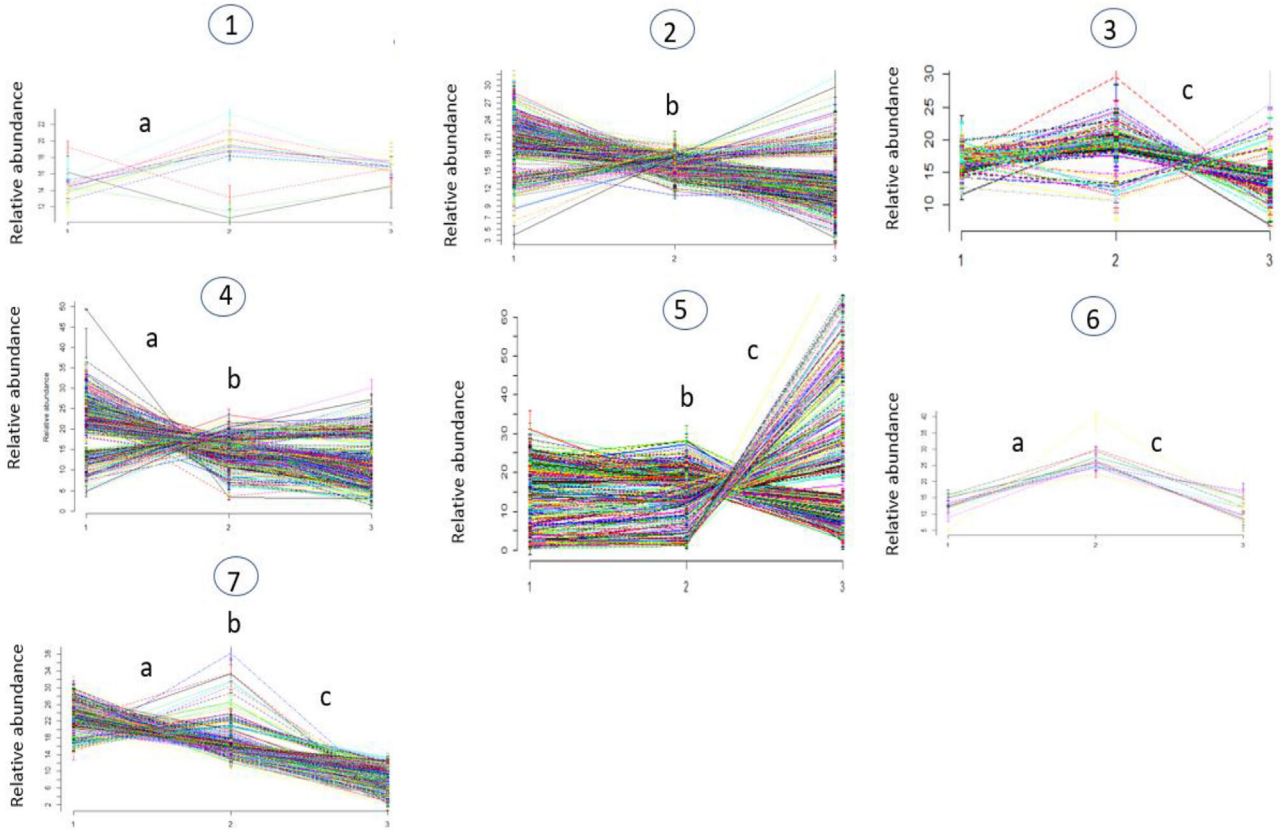

FIGURE 1 Protein abundant patterns across three developmental stages. 1, 2, 3 in X axis indicate early (E), mid (M), and late (L) stages respectively, a- significant number of proteins between early and mid-stages, b- significant number of proteins between early and late stages, c- significant number of proteins between mid-and late stages. The number 1 to 7 within the circle indicate the seven patterns of protein abundance. The list of proteins within each group is shown in the supplementary Tables (Table 3S to Table .9S). Error bars denote the SD obtained from three replicates of seed samples.

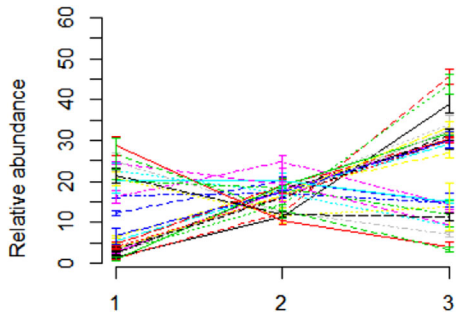

FIGURE 2 Differentially expressed storage proteins across three stages of seed filling were collected from three biological replicates. Error bars indicate SD obtained from three replicates of seed sample. The list of proteins within each group is shown in Table 12S. Developmental stages indicate 1-early, 2-mid, and 3-late stages.

oleosin family protein (74-fold), cystathionine beta-synthase (CBS) protein (73-fold), and another seed maturation protein by 71-fold.

In group ac (6) (Figure 1, Table 8S), only 15 proteins were differentially expressed, of which nine proteins exhibited lower abundance and six higher abundance between mid to late stages. However, all the 15 proteins showed higher abundance between early to mid-stages. The high abundant proteins included two storage proteins, albumin (2.4 to 8.1-fold) and RmlC-like cupins superfamily protein (1.2 to 2.4-fold), formate dehydrogenase (1.83 to 2.92-fold), and heat shock family protein (1.38 to 2.65-fold).

Group abc (7) (Figure 1 Table 9S) is unique compared to the other groups in that the protein abundance is specific to each stage. It includes 362 proteins related to a wide variety of functions. These proteins were divided into two groups: (1) exhibited lower abundance at an early stage, high at mid-and low at the late stages, and (2) exhibited higher abundance at an early stage and then lower at mid and late stages. Group 1 contained 38 proteins of which NAD(P)-binding Rossmann-fold increased by (2.36-fold), followed by stearoyl-acyl-carrier-protein desaturase (2.07-fold), pectin acetyl esterase family protein (2.03-fold), lipoxygenase (1.99-fold), fatty acid desaturase (1.75-fold), phosphatidylethanolamine-binding protein (1.71-fold), and vacuolar sorting receptor 3 stearoyl-acyl-carrier-protein desaturases (1.70-fold).

As a first step towards understanding protein synthesis and processes, we mapped the differentially expressed proteins on mRNA metabolic process, translation initiation, translation (Figure 3S, Table 10S), Golgi apparatus, and endoplasmic reticulum (Figure 3S, Table 11S). Most of the proteins mapped to the translation initiation pro-

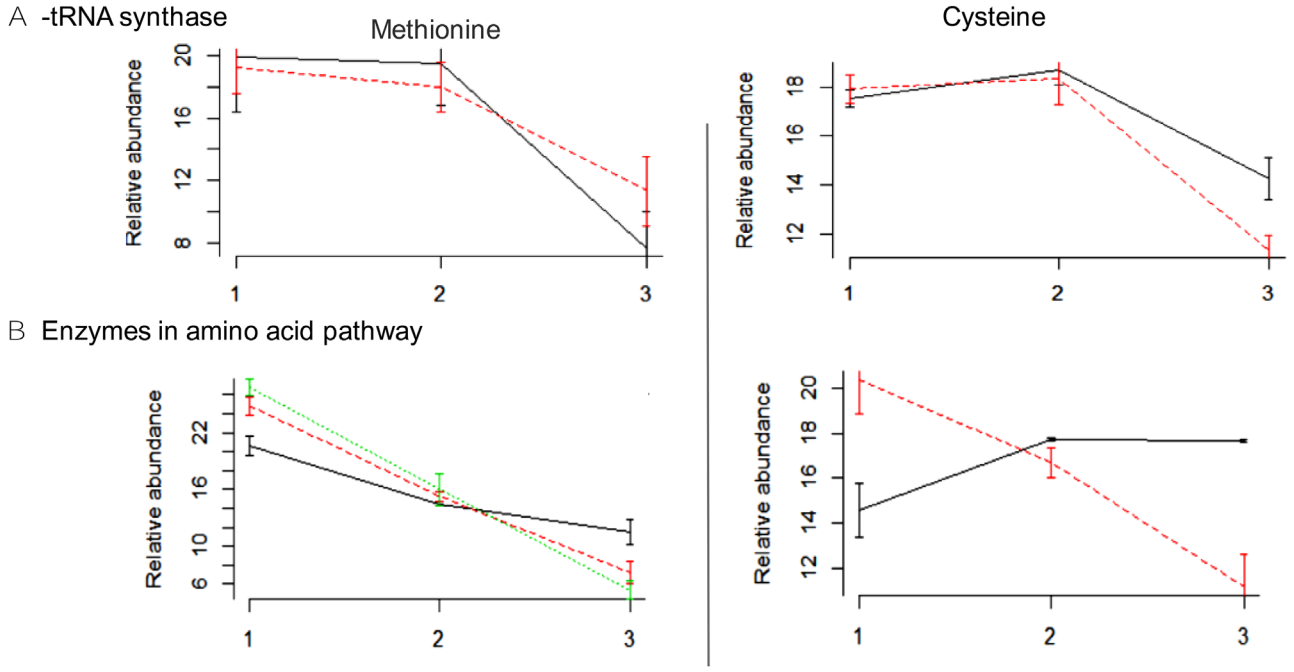

FIGURE 3 Relative abundance of -tRNA synthase (A) and enzymes of methionine and cysteine amino acids (B) across three stages of seed development. Error bars indicate SD obtained from three replicates of seed sample. The list of proteins within each group is shown in Table 14S. Developmental stages indicate 1-early, 2-mid, and 3-late stages.

cess showed lower abundance from early to mid and mid to late stages except two proteins, namely eIF3L; translation initiation factor 3 subunit L, which exhibited lower abundance between mid to late stages, and eIF1; translation initiation factor 1 which did not show significant changes between mid and late stages (Figure 3S and Tables 10 S).

In our study, most of the proteins that were mapped to GA and ER exhibited a similar trend of abundance from mid to late stages except B-cell receptor-associated protein (BAP31), which showed higher abundance between mid to late stages (Figure 3S and Table 11S).

We identified 25 storage proteins annotated as cupin family Figure 2 (Table 12S). Song et al. (2016) listed 18 accessions V1 (Glyma1.1 annotation) of cupin family proteins with conventional annotation [8], of which 16 accessions were matched in the V2 accessions (Williams 82 assembly version 2 annotation) of our study. Twelve of the 16 conventional annotated proteins were found among the significant proteins in our study. We could not find the conventional annotations for the 13 proteins and kept the annotations listed in the database. As evident from the supporting Table 12S, 25 genes located at 12 chromosomes encode the seed storage proteins. Interestingly, the relative abundance of 14 genes belongs to group abc, meaning that each protein's abundance was independent at each stage (Figure 2). All these proteins exhibited a steady increase in their abundance from early-stage to late-stage. These include 7S (β-subunit) (Glyma.20g146200), another 7S (β-subunit) (Glyma.20g148200), and 11S (Gy3, Gy4, Gy5). Unlike the 7S (β-subunit) and 11S (Gy3, Gy4, Gy5), the 2S albumins exhibited higher abundance at the early stage of seed development and lower abundance towards maturity except for one protein that was not annotated (Glyma.15G119700).

One of the major limitations of soybean seed is its low content of sulphur-containing amino acids such as methionine and cysteine [9]. We, therefore, mapped the differentially expressed protein on the sulphur metabolic pathways. As evident from Figure 4S and Table 13S, the relative abundance of the most sulfur metabolic proteins/enzymes exhibited higher abundance at early stages of seed development when compared to mid and late stages except metE; 5 methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (Glyma.20G055900), which showed higher abundance at mid and late stages. Like other enzymes, cysE; serine O-acetyltransferase that converts serine to cysteine exhibited higher abundance (21.7-fold) at an early stage. The abundance became less than half (9.8-fold) at the late stage of seed development.

Thirteen transcription factors/regulators (TFs) were differentially expressed across the three stages of seed development (Figure 5S and Table 15S). These are three basic transcription factors 3; two GATA type zinc finger transcription factor family protein; five winged-helix DNA-binding transcription factor family proteins. Among the 13 TFs, only two, WRKY family transcription factor family protein and transcription regulators; zinc ion binding exhibited higher abundance from early to the late stages of seed development (Figure 5S).

To verify the protein abundance, the gene expression analyses was performed using two genes based on a more than three-fold increase in abundance (Glyma.19G004800, from 0.80 to 2.51; Glyma.20G147600 from 1.443 to 2.790) and one that did not exhibit significant abundance in protein as control (Glyma.07G262600 from 12.647 to 10.651) (Table 1S). The gene expression analyses (Figure 6S) were consistent with the protein abundant patterns for the three genes.

Metabolic event during seed development determines the quality of seed composition. In the early to mid-stages of seed development, we observed the elevated abundance of several kinases, indicating that protein phosphorylation was a major event that occurred during this period. Consistent with the higher abundance of phosphorylation, most of the kinases in the glycolytic pathway exhibited higher abundance from the early stage to mid-stages of seed development. Two of the kinases are pyruvate kinases (Glyma.19G000700 and Glyma.10G201100) along with the fructokinase, enhance the final step of glycolysis where the conversion of phosphoenolpyruvate to pyruvate occurs with phosphorylation [10]. Similar to our study, Louis et al. (2012) characterized the phosphorylation pattern of developing Arabidopsis, rapeseed, and soybean seeds and identified fewer phosphoproteins at later stages of seed maturation in soybean and Arabidopsis [11]. As noted by the authors lower number of phosphoproteins at the later stages of seed filling might be related to the increased accumulation of storage protein [11]. It might also be related to the transcription factors that terminate phosphorylation at the later stages of seed development [12].

In the mid to late stages, we observed a higher abundance of the aspartyl, cysteine proteases, and senescence proteins. However, the two key proteases, Kunitz trypsin and another protease inhibitor decreased from mid to late stage. The higher abundance of aspartyl, cysteine proteases, and senescence proteins in the mid to late stages of seed development might be associated with protein remobilization [13].

In early to late stages, we observed a significant increase in several proteins such as callose synthase 5, a cell wall-associated protein, dehydrin, a desiccation protein, oleosin family protein, a structural protein located within vascular plant oil bodies, and beta vacuolar processing enzymes, probably because of seed desiccation. Consistent with our study, Jones et al. (2013) identified 40 gene models using RNA-seq that peak at the late stages of seed development, including hydrophilic proteins such as LEA, dehydrin, and oleosin. Similarly, Asakura et al. (2012) investigated gene expression in developing soybean seeds using DNA-microarray. They concluded that the expression of gene encoding oleosin gradually increased from early to the late stages of seed development [14]. At the final stage of seed development, seeds undergo a period of desiccation to lead to the quiescent period before imbibition and germination. The higher abundance of these proteins helps plants stabilizing cell membranes towards maturity [15].

Most of the proteins that were mapped to GA and ER exhibited a similar trend of abundance from mid to late stages except B-cell receptor-associated protein (BAP31), which showed higher abundance between mid to late stages. The functional annotation of BAP31 in plant systems is limited. However, the functions of this protein were well studied in mammalian systems and microbial systems [16, 17]. In the mammalian system, it involves protein sorting, transport, signalling, degradation, phosphorylation, and methylation [18, 19]. Interaction of this protein with transcription factor 2 (elf2) is also reported [19].

The two major seed proteins in soybean are 7S globulins and 11S globulins. The 7S β-conglycinin is made up of three subunits, namely α, β, and γ while the11S glycinin is encoded by at least five different gene families [20, 21]. In our investigation, 7S (β-subunit) (Glyma.20g146200) and several subunits of 11S (Gy3, Gy4, Gy5) exhibited higher abundance from early to maturity, while most of the 2S albumin showed decreased abundance during the same period. Similar to our study, Jones et al. (2013) conducted comprehensive transcriptome analyses of developing soybean seeds at seven stages after fertilization [22]. The authors reported ten gene models associated with storage proteins, glycinin, and beta-conglycinin, which accumulated in the seeds at the 4th stage and peaked at the 5/6th stage [22]. Similar to this study, decreased accumulation of some storage proteins at the later stages of seed development was also reported from other gene expression analyses in soybean [14, 23, 24] and in Arabidopsis [25].

While soybean is considered a valuable commodity of vegetable protein and oil supply for human food and animal feed, it has a limited amount of sulfur-containing amino acids such as methionine and cysteine. Therefore, several strategies, including metabolic engineering, have been undertaken to enhance the sulfur-containing amino acids in soybean seed [26, 27].

Cysteine biosynthesis includes an essential step of incorporating serine to cysteine controlled by a key enzyme, serine O-acetyltransferase. In this study, this enzyme showed higher abundance at the early stage. However, the amount became less than half at the later stage of seed development. We anticipate that a genetic engineering approach to secure the steady abundance of this enzyme across the developmental stages might play a crucial role in enhancing the cysteine content of soybean seeds. Related to methionine production, two enzymes responsible for production exhibited lower abundance from early to late stages (Figure 3). Although one (S-adenosylmethionine synthetase) of the two enzymes of cysteine production showed a similar trend as the methionine enzymes, homocysteine methyltransferase exhibited increase abundance from early to mid-stages. It showed minimum changes from mid to late stages (Figure 3). One of the plausible explanations of higher abundance of homocysteine methyltransferase between early to mid-stage is that seed secured the required supply of methionine until mid-stages seed development by converting homocysteine to methionine.

In summary, we identified the differential expression of several key proteins across the three stages of seed filling. This study also unveiled the differential expression of storage proteins during seed development. We also identified an increased abundance of a WRKY family transcription factor from early to late stages. To our knowledge, the involvement of this transcription factor in developing seeds has not been reported before. The results from our study will assist scientists and breeders in developing new value-added soybeans with improved protein quality traits.

ACKNOWLEDGMENTS

The authors would like to thank Thermo Center at Harvard Medical School for performing the quantitative protein analyses. Funding for this research was provided by Agricultural Research Service, USDA. Mention of trade name, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Accession MSV000087577 https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=38784ecbd0854bb3801afc0d89056f84

ORCID

Nazrul Islam https://orcid.org/0000-0002-6969-9791
Hari B. Krishnan https://orcid.org/0000-0001-6437-3672

REFERENCES

1. Jones, S. I., Gonzalez, D. O., & Vodkin, L. O. (2010). Flux of transcript patterns during soybean seed development. *BMC Genomics*, *11*, 136. https://doi.org/10.1186/1471-2164-11-136
2. Ritchie, S. W., H, J., Thompson, H. E. & Benson, G. O. How a soybean plant develops. Special Report No. 53 Ames IA: Iowa State University of Science and Technology Cooperative Extension Service 199 1985.
3. N. Islam, R M. Stupar, S. Qijian, D L. Luthria, W. Garrett, A O. Stec, J. Roessler, S. Natarajan, Genomic changes and biochemical alterations of seed protein and oil content in a subset of fast neutron induced soybean mutants. *BMC Plant Biol.* 2019, *19*, 420. https://doi.org/10.1186/s12870-019-1981-
4. J. Navarrete-Perea, Q. Yu, S P. Gygi, J A. Paulo, Streamlined tandem mass tag (SL-TMT) protocol: An efficient strategy for quantitative (phospho)proteome profiling using tandem mass tag-synchronous precursor selection-MS3. *Journal of Proteome Research* 2018, *17*, 2226–2236.
5. G C. Mcalister, D P. Nusinow, M P. Jedrychowski, M. Wühr, E L. Huttlin, B K. Erickson, R. Rad, W. Haas, S P. Gygi, MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem.* 2014, *86*, 7150–7158. https://doi.org/10.1021/ac502040v
6. E L. Huttlin, M P. Jedrychowski, J E. Elias, T. Goswami, R. Rad, S A. Beausoleil, J. Villén, W. Haas, M E. Sowa, S P. Gygi, A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* 2010, *143*, 1174–1189.
7. G C. Mcalister, E L. Huttlin, W. Haas, L. Ting, M P. Jedrychowski, J C. Rogers, K. Kuhn, I. Pike, R A. Grothe, J D. Blethrow, S P. Gygi, Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Analytical Chemistry* 2012, *84*, 7469–7478.
8. Bo Song, L. An, Y. Han, H. Gao, H. Ren, X. Zhao, X. Wei, H B. Krishnan, S. Liu, Transcriptome profile of near-isogenic soybean lines for β-conglycinin α-subunit deficiency during seed maturation. *Plos One* 2016, *11*, e0159723.
9. H B. Krishnan, Engineering soybean for enhanced sulfur amino acid content. *Crop Science* 2005, *45*, 454–461. https://doi.org/10.2135/cropsci2005.0454
10. H. Muirhead, Isoenzymes of pyruvate kinase. *Biochemical Society Transactions* 1990, *18*, 193–196.
11. L J. Meyer, J. Gao, D. Xu, J. Thelen, Phosphoproteomic analysis of seed maturation in Arabidopsis, rapeseed, and soybean. *Plant Physiology* 2012, *159*, 517–528.
12. C. Sonmez, I. Baurle, A. Magusin, R. Dreos, S. Laubinger, D. Weigel, C. Dean, RNA 3' processing functions of Arabidopsis FCA and FPA limit intergenic transcription. *Proceedings of the National Academy of Sciences of the United States of America* 2011, *108*, 8508–8513.
13. M. Poret, B. Chandrasekar, R. A. L. Van Der Hoorn, J. -C. Avice, Characterization of senescence-associated protease activities involved in the efficient protein remobilization during leaf senescence of winter oilseed rape. *Plant Science* 2016, *246*, 139–153.
14. T. Asakura, T. Tamura, K. Terauchi, T. Narikawa, K. Yagasaki, Y. Ishimaru, K. Abe, Global gene expression profiles in developing soybean seeds. *Plant Physiology and Biochemistry* 2012, *52*, 147–153.
15. E. A. Bray, Molecular responses to water deficit. *Plant Physiology* 1993, *103*, 1035–1040.
16. Atabekova, A. K., Pankratenko, A. V., Makarova, S. S., Lazareva, E. A., RA Owens, AG Solovyev, SY Morozov, Phylogenetic and functional analyses of a plant protein related to human B-cell receptor-associated proteins. *Biochimie* 2017, *132*, 28–37.
17. Pankratenko, A. V., Atabekova, A. K., Lazareva, E. A., Baksheeva, V. E., OA Zhironkina, EY Zernii, RA Owens, AG Solovyev, SY Morozov, Plant-specific 4/1 polypeptide interacts with an endoplasmic reticulum protein related to human BAP31. *Planta* 2017, *245*, 193–205.
18. C. -. C. Jia, J. Du, X. Liu, R. Jiang, Y. Huang, T. Wang, Y. Hou, B. Wang, B-cell receptor-associated protein 31 regulates the expression of valosin-containing protein through Elf2. *Cellular Physiology and Biochemistry* 2018, *51*, 1799–1814.
19. Fu, W., Sun, H., Zhao, Y., Chen, M., X Yang, Y Liu, W Jin, BCAP31 drives TNBC development by modulating ligand-independent EGFR trafficking and spontaneous EGFR phosphorylation. *Theranostics* 2019, *9*, 6468–6484.
20. V H Thanh, K Shibasaki, Beta-conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 1977, *490*, 370–384.
21. Vu Huu, T., Kazuyoshi, O., Kazuo, S., Isolation and characterization of the multiple 7s globulins of soybean proteins. *Plant Physiol.* 1975, *56*, 19–22. https://doi.org/10.1104/pp.56.1.19
22. S I. Jones, L O. Vodkin, Using RNA-Seq to profile soybean seed development from fertilization to maturity. *Plos One* 2013, *8*, e59270.
23. D. W. Meinke, J. Chen, R. N. Beachy, Expression of storage-protein genes during soybean seed development. *Planta* 1981, *153*, 130–139.
24. Nielsen, N. C., Dickinson, C. D., Cho, T. J., Thanh, V. H., B J Scallon, R L Fischer, T L Sims, G N Drews, R B Goldberg, Characterization of the glycinin gene family in soybean. *Plant Cell* 1989, *1*, 313–328.
25. F. Y. Peng, R. J. Weselake, Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *Bmc Genomics [Electronic Resource]* 2011, *12*, 286.
26. W.-S. Kim, J. Sun-Hyung, N W. Oehrle, J M. Jez, H B. Krishnan, Overexpression of ATP sulfurylase improves the sulfur amino acid content, enhances the accumulation of Bowman–Birk protease inhibitor and suppresses the accumulation of the β-subunit of β-conglycinin in soybean seeds. *Scientific Reports* 2020, *10*, 14989.
27. K. Saito, Sulfur assimilatory metabolism. The long and smelling road. *Plant Physiology* 2004, *136*, 2443–2450.

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202100143 in the Supporting Information section at the end of the article.

How to cite this article: Islam, N., Krishnan, H. B., & Natarajan, S. (2022). Quantitative proteomic analyses reveal the dynamics of protein and amino acid accumulation during soybean seed development. *Proteomics*, *22*, e2100143. https://doi.org/10.1002/pmic.202100143