A TMT-based quantitative proteomic analysis provides insights into the protein changes in the seeds of high- and low- protein content soybean cultivars

Cheol Woo Min • Ravi Gupta • Nguyen Van Truong • Jin Woo Bae • Jong Min Ko • Byong Won Lee • Sun Tae Kim

Received: 6 September 2020 / Revised: 14 September 2020 / Accepted: 14 September 2020

© Korean Society for Plant Biotechnology

Abstract The presence of high amounts of seed storage proteins (SSPs) improves the overall quality of soybean seeds. However, these SSPs pose a major limitation due to their high abundance in soybean seeds. Although various technical advancements including mass-spectrometry and bioinformatics resources were reported, only limited information has been derived to date on soybean seeds at proteome level. Here, we applied a tandem mass tags (TMT)-based quantitative proteomic analysis to identify the significantly modulated proteins in the seeds of two soybean cultivars showing varying protein contents. This approach led to the identification of 5,678 proteins of which 13 and 1,133 proteins showed significant changes in Daewon (low-protein content cultivar) and Saedanbaek (high-protein content cultivar) respectively. Functional annotation revealed that proteins with increased abundance in Saedanbaek were mainly associated with the amino acid and protein metabolism

involved in protein synthesis, folding, targeting, and degradation. Taken together, the results presented here provide a pipeline for soybean seed proteome analysis and contribute a better understanding of proteomic changes that may lead to alteration in the protein contents in soybean seeds.

Keywords *Glycine max*, LC-MS/MS, Low-abundance proteins, Protamine sulfate precipitation, Seed storage proteins, Tandem mass tags

Introduction

Soybean seeds (*Glycine max* Merr. L.) are one of the most important global sources of vegetative proteins and oils for humans and livestock global. Moreover, other nutrients present in soybean seeds including isoflavones, phytate, soy-saponin, and others, exhibit health-promoting effects in treating metabolic disorders, cardiovascular diseases, and cancer (Badole et al. 2015). Multiple efforts have been put to elucidate the differential profile of transcriptome (Lambirth et al. 2015; Schmidt et al. 2011), metabolome (Schmidt et al. 2011), and proteome (Min et al. 2015; Pandurangan et al. 2012; Xu et al. 2015) using soybean seeds showing different protein contents.

The classical workflow of soybean seed proteomics including two-dimensional proteomic analysis (2-DGE) allowed the identification of a few hundreds of significantly modulated proteins due to the presence of high abundant proteins (HAPs) which account for 75% of total proteins in soybean seeds (Min et al. 2019). Recently, this limitation have been overcome by of the development of a variety of methods for the pre-fractionation of total soybean proteins (Gupta et al. 2016) using protamine sulfate (PS) (Kim et al. 2013, 2015), calcium (Krishnan et al. 2009), and polyethylene glycol (PEG) (Kim et al. 2001). Moreover, advancements in the

C. W. Min • R. Gupta • N. V. Truong • S. T. Kim (✉)
Department of Plant Bioscience, Life and Industry Convergence
Research Institute, Pusan National University, Miryang, 50463,
Republic of Korea
e-mail: stkim71@pusan.ac.kr

R. Gupta
Department of Botany, School of Chemical and Life Sciences,
Jamia Hamdard, New Delhi, 110062, India

J. W. Bae
Department National Institute of Crop Science, Rural Development
Administration, Wanju, 55365, Republic of Korea

J. M. Ko
Department of Functional Crops, National Institute of Crop Science,
Rural Development Administration, Miryang, 50424, Republic of
Korea

B. W. Lee
Department of Central Area Crop Science, National Institute of
Crop Science, Rural Development Administration, Suwon, 16429,
Republic of Korea

liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based methodologies and analytical software for downstream data processing such as MaxQuant (Tyanova et al. 2016a), Perseus (Tyanova et al. 2016b), Proteome discoverer (Thermo Fisher Scientific, Waltham, MA, USA), and Skyline (Maclean et al. 2010), have led to the improvement of sensitivity, reliability, and coverage in proteome analysis.

Although the methodological developments have contributed to the identification of thousands of proteins per MS run with complete proteome coverage (Boersema et al. 2015; Niu et al. 2018), soybean seed proteomics is still poorly conducted because of the presence of SSPs and utilization of majorly 2-DGE based proteomics approaches (Min et al. 2019). Previously, Kim's group carried out high-throughput proteome analysis using soybean seeds by shot-gun proteomic approaches including label-free (Min et al. 2017) and tandem mass tag (TMT) labeling quantitative analysis (Min et al. 2020b). In particular, a TMT-based quantitative analysis of filling stages of soybean seeds identified 5,918 proteins, the highest number of proteins reported to date in soybean seeds (Min et al. 2020b). Moreover, the utilization of the PS precipitation method with shot-gun proteome pipeline, especially the TMT labeling approach, were carried out and comparison of the total, PS-supernatant (PS-S), and PS-pellet (PS-P) proteins, revealed enrichment of various low-abundance proteins (LAPs) related to diverse seed metabolism (Min et al. 2019).

Here, we are reporting a comparative seed proteome profiling of two soybean cultivars differing into protein content. Altogether, this study resulted in the identification of 1,146 differentially modulated proteins (13 and 1,133 protein showed different abundance profiles), providing a list of potential protein candidates using two soybean seed cultivars differing in protein and oil contents.

Materials and Methods

Plant materials

Soybean seeds (Daewon, and Saedanbaek) were sown in the experimental fields of the National Institute of Crop Science (NICS), Rural Development Administration (RDA), in Miryang, South Korea, in June. The soil was supplemented with a standard RDA N-P-K fertilizer (N-P-K=3-3-3.3 kg/10 acre). Seeds were harvested in October 2018 (average temperature, 23.5±3.5°C; average day length, 12 hours 17 min) (Min et al. 2016).

Protein extraction, protein digestion, and TMT labeling

Total proteins from two different cultivars of soybean seeds were isolated using the PS precipitation method with trichloroacetic acid (TCA)/acetone precipitation method (Gupta et al. 2015; Kim et al. 2015). Briefly, for PS precipitation method, one gram of each seed powder was homogenized with 10 mL of ice-cold Tris-Mg/NP-40 extraction buffer (0.5 M Tris-HCl, pH 8.3, 2% (v/v) NP-40, 20 mM MgCl₂) and centrifuged at 15,922 g for 10 min at 4°C. The clear homogenate was incubated on ice for 30 min with 0.15% (w/v) PS stock solution. The extract was centrifuged at 15,922 g for 10 min at 4°C to divide the PS-S and PS-P fractions, respectively, as described previously (Kim et al. 2015). Finally, the PS-S fraction was dissolved in 80% acetone containing 0.07% β -mercaptoethanol and stored -20°C until further analysis. Trypsin digestion by filter-aided sample preparation (FASP), TMT labeling and peptide pre-fractionation by basic pH reverse phase chromatography were carried out as described previously (Gupta et al. 2020; Kim et al. 2018; Min et al. 2020a; Wiśniewski et al. 2009). A total of 12 peptide fractions were collected, lyophilized in a vacuum centrifuge and stored at -80°C until further LC-MS/MS analysis.

LC-MS/MS analysis

Obtained peptides were dissolved in solvent-A (water/Acetonitrile (ACN), 98:2 v/v; 0.1% formic acid) and separated by reversed-phase chromatography using a UHPLC Dionex UltiMate® 3000 (Thermo Fisher Scientific, USA) instrument (Pajarillo et al. 2015). For trapping the sample, the UHPLC was equipped with Acclaim PepMap 100 trap column (100 μ m $\times$ 2 cm, nanoViper C18, 5 μ m, 100 Å) and subsequently washed with 98% solvent A for 6 min at a flow rate of 6 μ L/min. The sample was continuously separated on an Acclaim PepMap 100 capillary column (75 μ m $\times$ 15 cm, nanoViper C18, 3 μ m, 100 Å) at a flow rate of 400 nL/min. The LC analytical gradient was run at 2% to 35% solvent B (100% ACN and 0.1% formic acid) over 90 min, then 35% to 95% over 10 minutes, followed by 90% solvent B for 5 minutes, and finally 5% solvent B for 15 minutes. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was coupled with an electrospray ionization source to the quadrupole-based mass spectrometer QExactive™ Orbitrap High-Resolution Mass Spectrometer (Thermo Fisher Scientific, MA, Waltham, USA). The resulting peptides were electro-sprayed through a coated silica emitted tip (Scientific Instrument Service, NJ, Amwell Township, USA) at an ion

spray voltage of 2000 eV. The MS spectra were acquired at a resolution of 70,000 (200 m/z) in a mass range of 350–1650 m/z. The automatic gain control (AGC) target value was 3×10^6 and the isolation window for MS/MS was 1.2 m/z. Eluted samples were used for MS/MS events (resolution of 35,000), measured in a data-dependent mode for the 15 most abundant peaks (Top15 method), in the high mass accuracy Orbitrap after ion activation/dissociation with Higher Energy C-trap Dissociation (HCD) at 32 collision energy in a 100–1650 m/z mass range (Pajarillo et al. 2015). The AGC target value for MS/MS was 2×10^5 . The maximum ion injection time for the survey scan and MS/MS scan was 30 ms and 120 ms, respectively.

Data analysis by MaxQuant, Perseus, and R software

The acquired raw data were analyzed with the MaxQuant software (version 1.5.3.30) as described previously (Tyanova et al. 2016a; Gupta et al. 2018; Min et al. 2020b). All three technical replicates were cross-referenced against the Uniprot *Glycine max* database (75,674 entries, UP000008827, <http://www.uniprot.org>). TMT data processing was performed using default precursor mass tolerances set by the Andromeda search engine, which is set to 20 ppm for the first search and 4.5 ppm for the main search. Reporter mass tolerance has to set the minimum as 0.003 Da. The product mass tolerance was set to 0.5 Da and a maximum of two missed tryptic cleavage were allowed. Carbamidomethylation of cysteine residues and acetylation of lysine residues and oxidation of methionine residues were specified as fixed and variable modifications respectively. A reverse nonsense version of the original database was generated and used to determine the FDR which was set to 1% for peptide identifications. Statistical analysis was carried out using Perseus software (ver. 1.5.8.5) and R software as described previous report (Min et al. 2020a; Tyanova et al. 2016b). For removing the batch effect within TMT-6plex, data normalization was carried out using an internal reference scaling method as described previously (Plubell et al. 2017; Gupta et al. 2019). Missing values imputation was carried out from a normal distribution (width: 0.3, downshift: 1.8) using Perseus software (Tyanova et al. 2016b). Multiple Sample test controlled by the Benjamini-Hochberg FDR threshold of 0.05, was applied to identify the significant differences in the protein abundance (> 1.5 -fold change). The functional classification and pathway analysis were carried out using AgriGO v2.0 (Tian et al. 2017) web-based software for GO enrichment analysis, KEGG pathway analysis by DAVID proteome annotation web-based software

(Jiao et al. 2012), and MapMan software (version 3.6.0 RC1), respectively.

Results

Quantitative proteomic analysis using soybean seeds

To investigate the differential modulation of soybean seed proteome in high- and low-protein containing cultivars, seed proteins were isolated from Daewon and Saedanbaek and subjected to protamine sulfate precipitation method for depletion of major seed storage proteins (SSPs) (Kim et al. 2015). SSPs depleted fraction, referred as PS-S fraction, from two different cultivars (marked by DS; Daewon PS-S fraction and SS; Saedanbaek PS-S fraction, respectively) were sequentially subjected to trypsin digestion by filter-aided sample preparation (FASP) method and TMT-6plex labeling in the same manner as reported previously (Min et al. 2020a, 2020b) (Fig. 1A). Sequentially, pre-fractionation by basic-pH reversed-phase (BPRP) using in-house developed stage-tip was carried out to decrease the complexity of multiplex labeling sample mixtures (Han et al. 2014). This approach led to the identification of 51,278 peptides and 22,483 unique peptides matching to 5,678 protein groups from three technical replicates of TMT labeling sample sets (Fig. 1A). Particularly, TMT labeling combined with pre-fractionation approach showed improvements of the resolution and identification of protein as observed by 4,892 (84.3%) while a previous label-free study (Min et al. 2017)

Fig. 1 SDS-PAGE analysis reveals a clear separation of soybean seed proteins fractionated by protamine sulfate (PS) precipitation method. Abbreviation used: T; Total fraction, PS-S; PS-supernatant fraction, PS-P; PS-pellet fraction

A

TMT set & Replicates		Sample labeling		Number of identified Proteins	Number of Identified Peptides	Number of Identified Unique Peptides	Sequence Coverage
		TMT reagent	Sample Info				
TMT 6-plex batch set_1	Rep_1	126	DS	5,127	48,398	20,905	24.7
		127	DS	5,124			
		128	DS	5,131			
		129	SS	5,159			
		130	SS	5,163			
		131	SS	5,165			
TMT 6-plex batch set_2	Rep_2	126	DS	5,131	48,521	20,956	24.8
		127	DS	5,133			
		128	DS	5,139			
		129	SS	5,166			
		130	SS	5,175			
		131	SS	5,180			
TMT 6-plex batch set_3	Rep_3	126	DS	5,102	48,492	20,875	24.7
		127	DS	5,106			
		128	DS	5,106			
		129	SS	5,145			
		130	SS	5,140			
		131	SS	5,142			

B

Fig. 2 TMT-based quantitative proteomic analysis of Daewon and Saedanbaek seed proteins. (A) Different replicates of each sample were labeled with TMT-6plex isobaric labeling reagent as listed in the table. (B) Venn diagram showing the comparison of recently published proteome data analyzed by label-free approach versus TMT-based proteome analysis result of the current study

using PS-S fraction of soybean seed protein identified a comparatively lower number of protein (247 unique proteins, 0.4%) than present study (Fig. 1B).

Data normalization and statistical analysis

For normalization and removal of batch effects within TMT data sets, we applied an internal reference scaling (IRS) method to 4,610 proteins showing more than 70% valid intensity values (Fig. 2A). As per the normalization steps, TMT data sets were normalized at the peptide spectrum match (PSM) level into the MaxQuant software (Yu et al. 2020). Sequentially, PSM-level normalized reporter ion intensities of each TMT data set were applied to the further IRS method for normalization (Plubell et al. 2017). These multiple-step normalization procedures showed the correction of batch effects that occurred by TMT-6plex reagents (Fig.

3A). IRS normalization of the data showed an improvement of the median coefficient of variation (CV) values of each sample from 19.63% to 6.06% (Fig. 3B). Besides, Pearson correlation coefficients showed a high degree of correlation among different replicates of each sample with an average R^2 value of 0.996 (Fig. 2B). Of these 4,610 proteins, the sequential application of fold change (FC) calculation and Student's *t*-test controlled by a *Benjamini-Hochberg* FDR were applied to identify the statistically significant proteins between Daewon and Saedanbaek seeds ($FDR < 0.05$, $FC > 1.5$) (Fig. 2C). This resulting in the identification of 1,146 differential proteins, of these 1,133 and 13 proteins showed increased and decreased abundances in cluster_1 and 2 respectively (Table S1 and Fig. 4A). The PCA plot analysis revealed that PS-S proteins in Daewon vs Saedanbaek cultivars were separated at the PC1 accounting for a maximum 95.9% variation (Fig. 2D).

Fig. 3 The cluster plots (A) and boxplots (B) revealing the normalization efficiency through the IRS method. The normalization carried out for the correction of the batch effect occurred by TMT reagents. The CV values showing the improvement of the quantitative reproducibility of the normalized proteins from 19.63% to 6.06%

Functional annotation of differential proteins

MapMan analysis of 1,146 differential proteins showed up- and down-regulation of various proteins in the metabolism and cell function overview categories. Proteins with increased abundance in Saedanbaek, involved in cluster_1, were mainly related to the CHO metabolism (9.3%), photosynthesis (9.3%), secondary metabolism (8.5%), lipid metabolism (15.1%), and amino acid metabolism (13.2%) (Table S2). In the cell function overview category, majority of these proteins were found to be associated with protein degradation (11.9%), stress-related protein (10.8%), signaling (9.1%), transport (8.2%), RNA regulation (7.6%), protein targeting (7.2%), protein synthesis (5.9%) (Table S2). Particularly, in the case of the protein degradation category, various types of protease including subtilases, serine, cysteine, and aspartate protease, among others showed increased abundance Saedanbaek (Fig. 4B). Furthermore, 32 proteins related to protein

synthesis including various isoform of ribosomal proteins, initiation, and elongation factors also showed increased abundance in Saedanbaek (Fig. 4B). In addition to protein synthesis, the increased abundance of 4, 9, 39, and 20 proteins related to amino acid activation, protein folding, protein targeting, and post-translational modifications respectively were observed in Saedanbaek cultivar (Fig. 4B).

GO enrichment analysis of identified proteins showed an increased abundance of the proteins associated with the major metabolic pathway. In particular, proteins involved in cluster_2 showed increased abundance of proteins associated with protein metabolic process (GO:0019538), protein localization (GO:0008104), protein transport (GO:0015031), protein folding (GO: 0006457), and protein catabolic process (GO:0030163), among others in biological process categories (Table S3). In order to get further functional insights of proteins involved in cluster_2, KEGG pathway analysis was carried out using DAVID functional annotation web-based

Fig. 4 Proteome analysis of Daewon and Saedanbaek seeds by TMT-based quantitative proteomics approach. (A) Venn diagram showing the distribution of total identified and significant proteins followed by a narrow-down approach between samples. (B) Multi-scatter plots reveal the reproducibility across the three replicates with Pearson correlation values. (C) Volcano plot showing the fold change differences between Daewon and Saedanbaek seed samples. (D) Principle component analysis showing a clear separation of significant proteins

software (Jiao et al. 2012). KEGG pathway analysis showed that proteins with increased abundance in Saedanbaek were majorly associated with various metabolic pathways including biosynthesis of secondary metabolites, biosynthesis of amino acids, carbon metabolism, and protein processing in the endoplasmic reticulum (Table S4).

Discussion

Recently, the next-generation proteomics approaches including label-free and isotope labeling-based quantitative analysis have been showing significant improvements in protein quantification and thus identification of differential proteins (Boersema et al. 2015; Min et al. 2019). However, soybean seed proteomics is still elusive due to several limitations including a narrow range of detection, low reproducibility, and difficulty to detect LAPs due to the presence of high abundant proteins (HAPs) (Gygi et al. 2000; Thompson et

al. 2003). Therefore, a number of HAPs depletion methods have been developed specifically for the enrichment of LAPs from soybean seeds using protamine sulfate (Kim et al. 2015), calcium (Krishnan et al. 2009), and PEG (Kim et al. 2001). Our previous study showed a broad application of PS for the enrichment of LAPs from different plant samples including seeds and leaves of rice, soybean, pea, and peanut (Kim et al. 2015). Moreover, a previous report revealed that LAPs related to various major metabolism in filling and matured stages of soybean seeds were successfully enriched and identified in PS-S fraction using TMT-based quantitative analysis (Min et al. 2020b). Therefore, here we utilized the PS precipitation method in combination with TMT-based quantification to identify the differential proteins from the seeds of Daewon and Saedanbaek differing in total protein contents (Gupta et al. 2020; Min et al., 2020a, 2020b). This approach led to the identification of 1,146 significantly modulated proteins (FDR < 0.05, FC > 1.5) by the comparison between Daewon and Saedanbaek cultivars. Moreover, further

Fig. 5 Hierarchical clustering and functional annotation of the significantly modulated proteins. (A) Heatmap showing clustering of 1,146 significantly modulated proteins into two major clusters based on their abundance patterns. (B) The MapMan functional classification reveals the proteins involved in cluster_2 were mainly associated with protein and amino acid metabolism

functional classification of the increased abundance proteins, particularly in Saedanbaek cultivar showed accumulation of various LAPs associated with major seed metabolic pathways including photosynthesis, major/minor CHO metabolism, amino acid metabolism, lipid metabolism, and secondary metabolism, among others.

For the accumulation of storage compounds such as proteins and lipids, an enormous amount of energy is required for which the diffusion of oxygen in plant tissues is prerequisite for the energy production in mitochondria (Krishnan and Coe, 2001; Galili et al. 2014). Therefore, energy production through photosynthetic activity is required and critical for the accumulation of reserved metabolites during seed desiccation (Fait et al. 2006). Here, we identified 22 proteins including psbP, psb28 subunits, ATP synthase, plastocyanin, ferredoxin, NADH-ubiquinone oxidoreductase chain 1, Ribulose biphosphate carboxylase, Glyceraldehyde-3-phosphate dehydrogenase, and among others associated with photosynthesis showing increased abundance in Saedanbaek which is similar to that reported previously (Table S1) (Min et al. 2020b).

Besides, enrichment of LAPs led to the identification of 24 and 39 proteins related to major/minor CHO and lipid metabolism which showed increased abundance in the Saedanbaek cultivar (Table S2). Out of these, six and ten proteins related to starch synthesis/degradation and lipid degradation, respectively, showed increased abundance along with an increased abundance of two raffinose synthase proteins in the Saedanbaek cultivar (Table S1 and S2).

Moreover, during seed maturation stages, 10 to 15% of lipids are converted to raffinose family oligosaccharides (RFOs) when the supply of exogenous resources from maternal plants are limited (Kambhampati et al. 2020). These RFOs are produced by carbon remobilization from lipid along with sucrose during the development of seeds (Kambhampati et al. 2020).

In addition, we observed the accumulation of 34 proteins mainly associated with amino acid metabolism including GABA, glutamate, aspartate, branched-chain amino acids, tryptophan, serine, glycine, cysteine, and histidine synthesis (Table S2). Furthermore, MapMan functional classification of metabolism overview revealed the increased abundance of 3 proteins (more than 1.5 and 2.0-FC increase) involved in nitrogen metabolism which have an important role in determining the total amount of storage proteins. The ammonia derived from nitrogen uptake by maternal vegetative tissues is the primary source for supply the nitrogen predominantly as amino acid such as glutamine and asparagine to the seeds (Ohya et al. 2017). In addition, amino acids participate in the synthesis of storage proteins and thereby contributing the carbon remobilization through proteolysis activity during the late seed developmental stages (Galili et al. 2014; Kambhampati et al. 2020). Here, 64 proteases showed increased abundance in Saedanbaek that might be having a crucial role in the remobilization of endogenous nitrogenous products such as amino acid or proteins to storage proteins (Gallardo et al. 2006, 2007). Taken together, our results suggest a positive correlation of various metabolism-related proteins involved

in major/minor CHO metabolism, photosynthesis, nitrogen, amino acid metabolism, and among others with a higher protein content of soybean seeds.

Acknowledgment

This work was supported by a 2-Year Research Grant of Pusan National University.

References

- Badole SL, Patil KY, Rangari VD (2015) Antihyperglycemic activity of bioactive compounds from soybeans. In: Glucose intake and utilization in pre-diabetes and diabetes: Implications for cardiovascular disease. Academic Press, Boston, pp. 225–227
- Boersema PJ, Kahraman A, Picotti P (2015) Proteomics beyond large-scale protein expression analysis. *Curr. Opin. Biotechnol.* 34:162–170
- Fait A, Angelovici R, Less H, et al (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* 142:839–854
- Galili G, Avin-Wittenberg T, Angelovici R, Fernie AR (2014) The role of photosynthesis and amino acid metabolism in the energy status during seed development. *Front. Plant Sci.* 5:1–6
- Gallardo K, Firnhaber C, Zuber H, et al (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, Kurt C, Thompson R, Ochatt S (2006) *In vitro* culture of immature *M. truncatula* grains under conditions permitting embryo development comparable to that observed *in vivo*. *Plant Sci.* 170:1052–1058
- Gupta R, Min CW, Kim SW, et al (2020) A TMT-based quantitative proteome analysis to elucidate the TSWV induced signaling cascade in susceptible and resistant cultivars of *Solanum lycopersicum*. *Plants* 9:290
- Gupta R, Min CW, Kim SW, et al (2015) Comparative investigation of seed coats of brown- versus yellow-colored soybean seeds using an integrated proteomics and metabolomics approach. *Proteomics* 15:1706–1716
- Gupta R, Min CW, Kim YJ, Kim ST (2019) Identification of Msp1-induced signaling components in rice leaves by integrated proteomic and phosphoproteomic analysis. *Int. J. Mol. Sci.* 20:1–17
- Gupta R, Min CW, Kramer K, et al (2018) A multi-omics analysis of *Glycine max* leaves reveals alteration in flavonoid and isoflavonoid metabolism upon ethylene and abscisic acid treatment. *Proteomics* 18:1–10
- Gupta R, Min CW, Wang Y, et al (2016) Expect the unexpected enrichment of “hidden proteome” of seeds and tubers by depletion of storage proteins. *Front. Plant Sci.* 7:1–7
- Gygi SP, Corthals GL, Zhang Y, et al (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* 97:9390–9395
- Han D, Jin J, Woo J, et al (2014) Proteomic analysis of mouse astrocytes and their secretome by a combination of FASP and stage tip-based, high pH, reversed-phase fractionation. *Proteomics* 14:1604–1609
- Jiao X, Sherman BT, Huang DW, et al (2012) DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28:1805–1806
- Kambhampati S, Aznar-Moreno JA, Hostetler C, et al. (2020) On the inverse correlation of protein and oil: Examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants. *Metabolites* 10:1–15
- Kim DK, Park J, Han D, et al. (2018) Molecular and functional signatures in a novel Alzheimer’s disease mouse model assessed by quantitative proteomics. *Mol. Neurodegener.* 13:1–19
- Kim ST, Cho KS, Jang YS, Kang KY (2001) Two-dimensional electrophoretic analysis of rice proteins by polyethylene glycol fractionation for protein arrays. *Electrophoresis* 22:2103–2109
- Kim YJ, Lee HM, Wang Y, et al. (2013) Depletion of abundant plant RuBisCO protein using the protamine sulfate precipitation method. *Proteomics* 13: 2176–2179
- Kim YJ, Wang Y, Gupta R, et al. (2015) Protamine sulfate precipitation method depletes abundant plant seed-storage proteins: A case study on legume plants. *Proteomics* 15: 1760–1764
- Krishnan HB, Coe EH (2001) Seed Storage Proteins. In: Encyclopedia of Genetics. Academic Press, New York, pp.1782–1787
- Krishnan HB, Oehrle NW, Natarajan SS (2009) A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: A case study using *Glycine max*. *Proteomics* 9:3174–3188
- Lambirth KC, Whaley AM, Blakley IC, et al. (2015) A comparison of transgenic and wild type soybean seeds: Analysis of transcriptome profiles using RNA-Seq. *BMC Biotechnol.* 15:89
- Maclean B, Tomazela DM, Shulman N, et al. (2010) Skyline : an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26:966–968
- Min CW, Gupta R, Agrawal GK, et al (2019) Concepts and strategies of soybean seed proteomics using the shotgun proteomics approach. *Expert Rev. Proteomics* 16:795–804
- Min CW, Hyeon H, Gupta R, et al. (2020a) Integrated proteomics and metabolomics analysis highlights correlative metabolite-protein networks in soybean seeds subjected to warm-water soaking. *J. Agric. Food Chem.* 68:8057–8067
- Min CW, Kim YJ, Gupta R, et al. (2016) High-throughput proteome analysis reveals changes of primary metabolism and energy production under artificial aging treatment in *Glycine max*

- seeds. *Appl. Biol. Chem.* 59:841–853
- Min CW, Lee SH, Cheon YE, et al. (2017) In-depth proteomic analysis of *Glycine max* seeds during controlled deterioration treatment reveals a shift in seed metabolism. *J. Proteomics* 169:125–135
- Min CW, Park J, Bae JW, et al. (2020b) In-depth investigation of low-abundance proteins in matured and filling stages seeds of *Glycine max* employing a combination of protamine sulfate precipitation and TMT-based quantitative proteomic analysis. *Cells* 9:1517
- Min CW, Gupta R, Kim SW, et al. (2015) Comparative biochemical and proteomic analyses of soybean seed cultivars differing in protein and oil content. *J. Agric. Food Chem.* 63:7134–7142
- Niu L, Yuan H, Gong F, et al (2018) Protein extraction methods shape much of the extracted proteomes. *Front. Plant Sci.* 9:802
- Ohyama T, Ohtake N, Sueyoshi K, et al. (2017) Amino acid metabolism and transport in soybean plants. In: Amino acid - New insights and roles in plant and animal. IntechOpen, London
- Pajarillo EAB, Kim SH, Lee JY, et al (2015) Quantitative proteogenomics and the reconstruction of the metabolic pathway in *Lactobacillus mucosae* LM1. *Korean J. Food Sci. Anim. Resour.* 35:692–702
- Pandurangan S, Pajak A, Molnar SJ, et al. (2012) Relationship between asparagine metabolism and protein concentration in soybean seed. *J. Exp. Bot.* 63:3173–3184
- Plubell DL, Wilmarth PA, Zhao Y, et al. (2017) Extended multiplexing of tandem mass tags (TMT) labeling reveals age and high fat diet specific proteome changes in mouse epididymal adipose tissue. *Mol. Cell. Proteomics* 16:873–890
- Schmidt MA, Barbazuk WB, Sandford M, et al. (2011) Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant Physiol.* 156:330–345
- Thompson A, Kuhn K, Kienle S, et al. (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75:1895–1904
- Tian T, Liu Y, Yan H, et al. (2017) AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45:W122–W129
- Tyanova S, Temu T, Cox J (2016a) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11:2301–2319
- Tyanova S, Temu T, Sinitcyn P, et al. (2016b) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13:731–740
- Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* 6:359–362
- Xu XP, Liu H, Tian L, et al (2015) Integrated and comparative proteomics of high-oil and high-protein soybean seeds. *Food Chem.* 172:105–116
- Yu SH, Kiriakidou P, Cox J (2020) Isobaric matching between runs and novel PSM-level normalization in MaxQuant strongly improve reporter ion-based quantification. *bioRxiv*.