

# A Comparison of Protein and Phenolic Compounds in Seed from GMO and Non-GMO Soybean

Savithiry S Natarajan<sup>1\*</sup>, Farooq H Khan<sup>1</sup>, Davanand L Luthria<sup>2</sup>, Mark L Tucker<sup>1</sup>, Qijian Song<sup>1</sup> and Wesley M Garrett<sup>3</sup>

<sup>1</sup>USDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA

<sup>2</sup>USDA-ARS, Food Composition and Methods Development Laboratory, Beltsville, MD 20705, USA

<sup>3</sup>USDA-ARS, Animal Biosciences and Biotechnology Laboratory, Beltsville, MD 20705, USA

## Abstract

Soybean protein is a valuable and important component in human and animal diets. Approximately 94% of the soybean planted in the US is genetically modified (GM) to enhance quality and productivity. Since value-added traits are continuously being developed by genetic modification, it is important to determine if any unintended changes occur in GM soybean seeds. In this investigation, we have selected three different transgenic lines, denoted event 1, 2 and 3 with a single *Agrobacterium tumefaciens* T-DNA insert that included genes for a herbicide-resistance selectable gene (*bar*) and a  $\beta$ -glucuronidase (*GUS*) reporter gene expressed using a double 35S Cauliflower Mosaic Virus (CaMV) promoter and a soybean polygalacturonase (*Glyma12g01480*) promoter, respectively. The transgenic lines and non-transgenic progenitor isolate (control) were used for both proteomic and phenolic compound analysis. Seed proteins were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Out of approximately 1300 protein spots detected per protein extract, 30 spots were selected for further analysis based on software-determined differences (ANOVA) in their relative abundance in the protein gels for the control and three events. Subsequent statistical analysis after Bonferroni correction indicated that the abundance of only two of the thirty protein spots were significantly different at the 1% probability level. Two protein spots, an isoflavone reductase and a quinone oxidoreductase-like protein, in event 2 were significantly different from the control and the other two transgenic events. All thirty protein spots were analyzed and identified by mass spectrometry (MS) followed by a search of the NCBI databases using the Mascot search engine. In addition to protein, two classes of phenolic compounds, isoflavonoids and phenolic acids, were analyzed by LC-MS. The results indicated no systematic differences in the amount or profile for either class of phenolic compounds in the control or three transgenic events.

**Keywords:** Soybean; Transgenic; Proteomics; Two-dimensional gel electrophoresis; MALDI-TOF/TOF; Phenolic acids; Isoflavones

## Introduction

Soybean and soy derived products are consumed worldwide for benefits derived from their high protein and bioactive phenolic components, namely isoflavones and phenolic acids. Soybean proteins are used in human foods in a variety of forms. Consumption of soybeans reduced the risk of cancer, decreased risk factors for cardiovascular disease, and reduced chances of other chronic illnesses [1]. Soybean seed contains 2 major storage proteins:  $\beta$ -conglycinin and glycinin. Other proteins such as  $\beta$ -amylase, cytochrome c, lectin, lipoxygenase, urease, Bowman-Birk Inhibitor (BBI) and trypsin are also present [2]. Recently, Clemente et al. [3] reported that soybean trypsin and chymotrypsin inhibitor Bowman-Birk (BBI) were linked to the prevention and treatment of colorectal cancer. Soybean isoflavones also were reported to reduce the risk of breast, prostate, intestine, and stomach cancer [4,5].

Over the past few decades, genetically modified (GM) crops have played a significant role in increasing the productivity and nutritional value of crops, e.g., increasing tolerance to herbicides, improving resistance to pathogens and producing recombinant pharmaceutical molecules including human growth hormone and coagulation factor IX [6-9].

Since 1996, GM crops have been commercially available in United States and many other countries have also approved the commercial use of GM crops. Clive [10] recently reported that approved GM crops are currently grown on approximately 180 million hectares in more than 25 countries. Foods derived from GM crops are subjected to rigorous safety evaluation such as characterization of intended modification with specific tests for allergenicity and toxicity. In

addition, potential unintended effects are evaluated on the basis of agronomic characteristics, compositional analysis, and evaluation of key nutrients [11].

A safety assessment of GMO crops is mandatory in US and other countries. Numerous international organizations have played vital roles in the formulation of universal safety/risk assessments of GMO crops (Codex Alimentarius Commission (CAC), 2003 and 2009; Food and Agriculture Organization of the United Nations (FAO), 2011; World Health Organization (WHO), 1993 and 2000; Organization of Economic Cooperation and Development (OECD), 1993). According to Millstone et al. [12], the safety assessment of new crops is based on the concept of "substantial equivalence". If the chemical composition of a new crop is substantially similar to that of existing crop, it is not considered to pose a health risk [13]. Unintended effects may occur due to unforeseen interactions with other proteins or biochemical, or effects caused by random insertion into the genome, which can alter normal plant processes. It is therefore important that each genetic modification be examined on a case-by-case basis for unintended

**\*Corresponding author:** Savithiry S Natarajan, Soybean Genomics and Improvement Laboratory, ARS-USDA, 10300 Baltimore Avenue, Beltsville, USA, Tel: 301-504-5258; Fax: 301-504-5728; E-mail: [savi.natarajan@ars.usda.gov](mailto:savi.natarajan@ars.usda.gov)

**Received** August 28, 2014; **Accepted** September 11, 2014; **Published** September 18, 2014

**Citation:** Natarajan SS, Khan FH, Luthria DL, Tucker ML, Song Q, et al. (2014) A Comparison of Protein and Phenolic Compounds in Seed from GMO and Non-GMO Soybean. J Data Mining Genomics Proteomics 5: 161. doi:10.4172/2153-0602.1000161

**Copyright:** © 2014 Natarajan SS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

effects [14]. In 2010, the sequence of the soybean genome became available and greatly improved our ability to access unintended effects [15]. However, a clear understanding of unintended effects in regard to protein and metabolites due to transgenic modifications is still lacking [16-19].

Holistic profiling approaches such as genomic, proteomics, transcriptomics, and metabolomics have broadened the spectrum of compounds that can be identified and analyzed in contrast to earlier targeted analytical approaches [20-23]. In the present study, we utilize a combination of a proteomic and a more classical targeted approach. Proteins were separated using two-dimensional gel electrophoresis (2D-PAGE) and individual proteins identified by mass spectrometry. In addition, two classes of phenolic compounds, phenolic acids and isoflavones, were analyzed using high performance liquid chromatography (HPLC) with diode array (DAD) and mass spectral detections (MSD).

## Materials and Methods

### Plant material and generation of transgenic plants

A soybean phage genomic library was screened for clones with similarity to a PG11 cDNA. A 17 kb genomic insert was sequenced and when the soybean genomic sequence became available the two sequences were compared, with only a few differences found between the two sequences. Polymerase chain reaction (PCR) was used to fuse a 1951 bp *GmPG11a* gene (Glyma12g01480) promoter to a  $\beta$ -glucuronidase (GUS) reporter gene immediately downstream from the ATGGUS open reading frame. The GUS open reading frame included an intron from the castor bean catalase gene 18 bp down from the start of translation [24]. The PG11a-GUS-NOS3 construct was cloned into the *Bam* HI-*Eco* RI site of pTF101.1 in the opposite orientation to a herbicide-resistance selectable marker gene (bialaphos resistance, *bar*) [25]. The herbicide-resistance gene was constitutively expressed using a double CaMV 35S promoter [26]. The Iowa State University Transformation Facility transferred this construct into *A. tumefaciens* (EHA101) and then transformed and regenerated transgenic soybean (*Glycine max* Williams 82). Seven independent events displayed herbicide resistance and all seven tested positive for the GUS gene in a PCR genomic DNA assay. Five events displayed strong GUS staining in an abscission assay. Segregation analysis of second and third generation seed indicated that three events with strong GUS staining had a single copy of the transgene. Homozygous seed from third generation plants for each of the following three events were used for this study: ST-16-3-10, ST83-28-1-15 and ST83-37-7-10. Seeds were collected from multiple plants for each event, mixed and several seeds were used for analysis.

### Protein extraction from transgenic soybean seeds

Protein was extracted from transgenic and isogenic control soybean seeds using a modified Trichloroacetic acid (TCA) / acetone method [27]. Soybean seeds were first pulverized to fine powder in liquid nitrogen using a mortar and pestle. Fifty mg of seed powder was homogenized in 1 mL of precipitation solution containing 10% (w/v) TCA in acetone with 0.07% (v/v)  $\beta$ -mercaptoethanol. The protein was allowed to precipitate for 24 hrs at -20°C. The precipitated protein then twice underwent 1 hr rinses with cold acetone containing 0.07% (v/v)  $\beta$ -mercaptoethanol, with each rinse followed by centrifugation at 20,800 g for 20 min at 4°C. The supernatant from each rinse was discarded. The protein precipitate was dried using a vacuum centrifuge for 30 minutes, and then resolubilized in 1.0 mL of lysis buffer [7M urea, 2M thiourea, 4% (w/v) CHAPS, 0.5% (w/v) DTT]. The mixture

was homogenized with repetitive pipetting, followed by the use of a sonicator for 10 minutes, and centrifugation at 20,800 g for 20 minutes at 4°C. The supernatant containing the solubilized protein was used for each of the 2D-PAGE separations.

### 2D-PAGE analysis

Protein (100  $\mu$ g) was first separated by isoelectric focusing (IEF) on an IPGphor II apparatus using 13 cm immobilized pH gradient (IPG) strips of pH ranges 4-7 and 6-11 (GE Healthcare, Piscataway, NJ). The dry IPG strips were hydrated 12 hrs in 250  $\mu$ L rehydration buffer (7M urea, 2M thiourea, 4% CHAPS, 2% pharmalyte, 0.002% bromophenol blue) containing 100  $\mu$ g protein. The voltage settings for IEF were 500 V for 1 hr, 1000 V for 1 hr, 5000 V for 1 hr, and 8000 V to a total 24 kVhr. Following electrophoresis, the protein in the strips was reduced through incubation with equilibration buffer (50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, 1% DTT) and subsequently alkylated with the same buffer by substituting 2.5% iodoacetamide for DTT. Incubations for both reduction and alkylation were timed for 30 minute and took place on a shaker at room temperature. The second dimensional protein separation was achieved by polyacrylamide gel electrophoresis on a 12.5% polyacrylamide gel using a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ). The gels were then stained with colloidal CoomassieG-250. After de-staining with ddH<sub>2</sub>O, gels were scanned using a GE ImageScanner III (GE Healthcare, Piscataway, NJ).

### Protein visualization and image analysis

Protein expression analysis was conducted through the use of Progenesis SameSpots (TotalLab, Newcastle, England). Scanned images in Maya Embedded Language (MEL) file format were first uploaded and underwent a quality check for color saturation and ensured consistency in image resolution across all samples. The images were then aligned and spots were automatically detected. In addition to automated spot detection, a thorough visual inspection was used to eliminate non-spot fragments falsely reported by the software. Subsequent to the spot detection, the experimental design allowed the software to report differentially expressed spots across control and transgenic soybean gels. All of the differentially expressed spots with a p-value below 0.05 were chosen for subsequent analysis.

### In-gel digestion of protein spots

Soybean protein spots differentially expressed across control and transgenic samples were excised with a 1.5 mm Spot Picker (The Gel Company, San Francisco, CA, USA). Protein digestion was performed using trypsin as described previously [27]. For further removal of the gel stain, the gel plugs were hydrated with 25mM ammonium bicarbonate on a shaker for 10 minutes, and then dried with acetonitrile for 10 minutes. The hydration and dehydration cycle was performed twice. The gel plugs were then thoroughly dried under vacuum and incubated overnight at 37°C with 20  $\mu$ L of 10  $\mu$ g/mL porcine trypsin (Promega, sequencing grade, Madison, WI) in 20 mM ammonium bicarbonate for protein digestion. The resulting peptides were eluted from the gel in 50% acetonitrile and 5% trifluoroacetic acid. The extract was vacuum-dried and the dried peptides dissolved in 50% acetonitrile and 0.1% trifluoroacetic acid.

### Mass spectrometry

Samples were spotted on a MALDI plate, co-crystallized with a 5 mg/ml concentration of  $\alpha$ -cyanohydroxycinnamic acid (CHCA) matrix prepared in 70% acetonitrile containing 0.1% trifluoroacetic acid. Thirty fmol of a commercially prepared tryptic digest of bovine

serum albumin (Michrom Bioresources, Inc. Auburn, CA, USA) was spotted onto the 13 calibration wells of the sample plate and 5 peptides with masses in the range of 927.493 m/z to 1881.905 m/z were used for the calibration. The mass spectrometer used was an AB SCIEX TOF/TOF™ 5800 System (AB SCIEX, Framingham, MA, USA) operated in positive ion reflector mode to analyze tryptic peptides. Prior to analysis of unknowns, a plate model calibration was run to optimize mass accuracy and to update the instrument's default calibration parameters. The instrument was operated in batch mode during peptide analysis, which entails first performing an MS survey scan on all spots of interest, followed by sequential MS/MS analysis of peaks detected in the MS scan. Acquisition of MS/MS data was controlled by an interpretation method that acquired MS/MS spectra on the strongest precursors first on up to 100 precursors detected in the MS scan. An exclusion mass list was prepared to prevent MS/MS analysis of common human keratin contaminant and minor porcine trypsin autolysis peaks. MS spectra for both standards and unknowns were acquired in positive ion reflector mode with 400 shots of a 349 nm Nd:YAG laser operating at 404 Hz. MS/MS spectra were also acquired in positive ion reflector mode with 250-1000 laser shots firing at a rate of 1010 Hz. Collision energy was set to 1kV and collision induced dissociation (CID) was enabled with air as the collision gas in the CID cell. When possible, known trypsin autolysis peaks at m/z 842.51 and 2,211.10 were used to internally calibrate the MS spectra.

Protein identification was performed using the Mascot search engine (<http://www.matrixscience.com>) against the NCBI non-redundant database with the taxonomy filter *Viridiplantae* (green plants). The parameters for database searches included: monoisotopic mass, trypsin as the digestive enzyme with allowance for 1 missed cleavage,

peptide tolerance of 50 parts per million (ppm), MS/MS tolerance of 0.6 Da, allowance of 1+ peptide charge, fixed modification for carbamidomethylation of cysteine residues, and variable modifications for oxidation of methionine residues as well as N-terminal pyroglutamic acid resulting from glutamic acid or glutamine. Positive identification of proteins by MS/MS analysis required a single peptide having a significant ion score. Samples identified as uncharacterized/unknown identity were subjected to sequence alignment match via BLAST against UniProt knowledgebase, where sequence similarity of 85% and above was used as a minimum criteria for inclusion.

### Extraction and analysis of isoflavones from soybeans

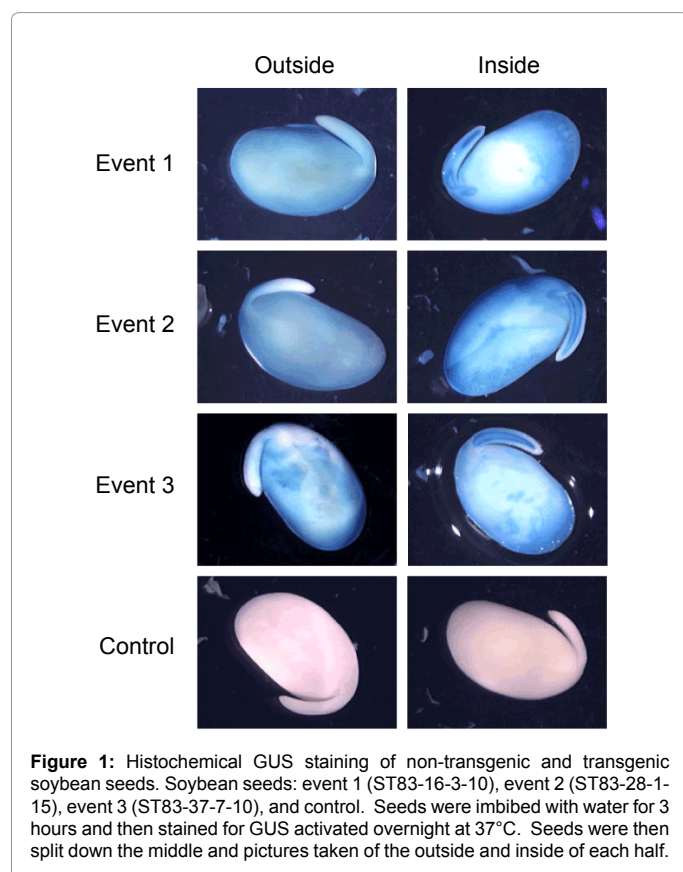
Ground soybean samples (250 mg) were placed in 15 ml polypropylene conical tube with 5 ml of hexane. The mixture was placed in an ultrasonic bath for 15 min. The mixture was centrifuged and the hexane layer, which contained oil, was removed and discarded. The residue was extracted twice with 5 ml of with the previously optimized solvent mixture (EtOH:H<sub>2</sub>O:DMSO, 75:20:5, v/v/v). The supernatant from the two extraction cycles were pooled together in a volumetric flask and the volume of the combined extract was adjusted to 10 ml with extraction solvent. Appropriate aliquots of extracts were filtered through a 0.45 µm PVDF syringe filter for isoflavone analysis by LC-MS. Analysis of isoflavones was carried out using an Agilent 1290 Infinity LC system coupled to a diode array and a mass spectrometry detector from Agilent Technologies (Palo Alto, CA, USA). Separation of isoflavones was achieved using a reversed phase C<sub>18</sub> Luna column (Phenomenex, Lorange, CA, USA, 150×4.6 mm; particle size 5 µm), preceded by a guard column (Phenomenex, 4×3.0 mm). The structures for all isoflavones were confirmed by comparison of retention time, UV spectra and mass spectral analysis. Peak areas were integrated for quantitation. Comparison of extraction efficiencies was achieved by comparing HPLC peak areas.

### Extraction and analysis of phenolic acids from soybeans

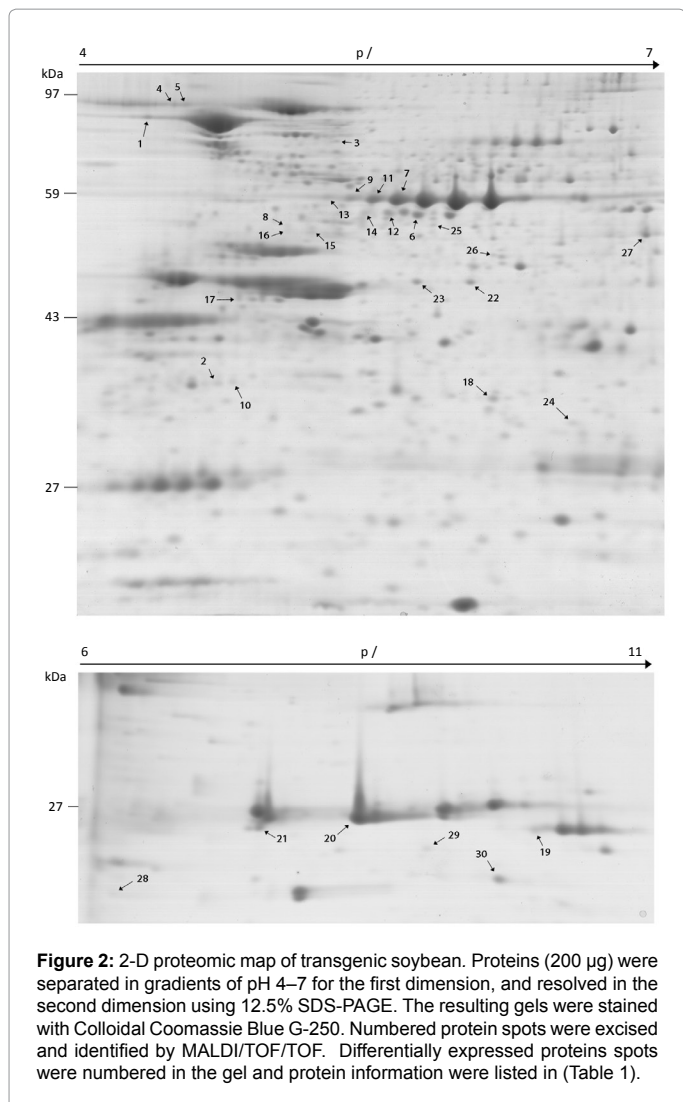
Ground soybean samples (250 mg) were hydrolyzed in 2N NaOH containing 10 mM EDTA and 1% ascorbic acid for 30 min in an ultrasonic bath at 56°C as described previously [28]. After hydrolysis, the samples were cooled to an ambient temperature and the pH of the extract was adjusted to 2.5 with 6N HCl. Phenolic acids were isolated from the acidified extract with ethyl acetate (5 ml×2). The mixture was vortexed for 30 sec and centrifuged on a bench top centrifuge (Damon IEC HN-SII, Ramsey, Minnesota, USA) at 5000 rpm for 10 min. The upper organic ethyl acetate layer containing hydrolyzed phenolic acids was carefully transferred into a separate tube and evaporated under nitrogen gas. The dried residue was re-dissolved in 2 ml of 80% aqueous methanol filtered through a 0.45 µm PVDF syringe filters into HPLC vials for analysis. Four replicates of hydrolysis, extraction, and analysis were carried out with each sample. The structures for identified phenolic acids were confirmed by comparison of retention time, UV and mass spectral analysis as reported earlier [29]. Peak areas were integrated for quantitation. Comparison of phenolic acids was achieved by total peak area under the peak as detected by the UV-diode array detector.

### Results and Discussion

Transgenic soybean seeds consist of single gene insertions containing both the *bar* herbicide gene and the β-glucuronidase (*GUS*) reporter gene controlled by the double CaMV 35S promoter and *PG11a* gene promoter, respectively. Although the 35S promoter is generally used as a constitutive promoter in plants, its expression level varies in



**Figure 1:** Histochemical GUS staining of non-transgenic and transgenic soybean seeds. Soybean seeds: event 1 (ST83-16-3-10), event 2 (ST83-28-1-15), event 3 (ST83-37-7-10), and control. Seeds were imbibed with water for 3 hours and then stained for GUS activated overnight at 37°C. Seeds were then split down the middle and pictures taken of the outside and inside of each half.



**Figure 2:** 2-D proteomic map of transgenic soybean. Proteins (200 µg) were separated in gradients of pH 4–7 for the first dimension, and resolved in the second dimension using 12.5% SDS-PAGE. The resulting gels were stained with Colloidal Coomassie Blue G-250. Numbered protein spots were excised and identified by MALDI/TOF/TOF. Differentially expressed proteins spots were numbered in the gel and protein information were listed in (Table 1).

different tissues. One study of soybean demonstrated that the CaMV 35S promoter was expressed strongly in leaves but very weakly in seed [30]. Similarly, we demonstrated using GUS staining that the *PG11a* promoter is expressed throughout the soybean seed with the highest expression in the vascular cylinder of the root radical (Figure 1). Seed morphology and coat color were similar between the 3 transgenic events and control soybean seeds tested in our current investigation.

### Analysis of proteins

A modified TCA/acetone method was used to extract soybean proteins [27]. Three biological replicates of protein extracts from three transgenic events and control lines were separated by 2D-PAGE. Analysis was initially carried out with immobilized pH gradient (IPG) strips using a pH range of 3 to 10 (data not shown). However, over this broad pH range, storage proteins were poorly separated and the total numbers of resolved protein spots were insufficient. Improved separation and spot resolution were achieved using pH gradients between 4 to 7 and 6 to 11. Therefore, 100 µg of each protein extract was used for two separate 2D gels, one with a pH range from 4 to 7 and another with a pH range of 6 to 11. Representative 2D-PAGE gels for soybean seed protein are shown in (Figure 2) (Table 1). Comparison of

2D gel images for each of the transgenic events to that of control images identified 30 proteins spots that differed among treatments. Each of the 30 differentially expressed spots were excised manually from the gels and digested with trypsin. The tryptic digests were purified and analyzed by MALDI/TOF/TOF mass spectrometry. (Table 2) includes MASCOT information for each excised protein referenced by its spot number including: predicted protein identity, theoretical isoelectric point (pI), molecular weight (Mr), the original species that the protein was identified in, number of peptides matched, MOWSE score, and gene ID or accession number of the best match.

### Variation of storage proteins

Soybean seeds contain two major classes of storage proteins, the 7S and 11S proteins, which are normally referred to as  $\beta$ -conglycinin and glycinin, respectively. The  $\beta$ -conglycinins are trimeric glycoproteins consisting of three types of non-identical, but homologous polypeptide subunits:  $\alpha$ ,  $\alpha'$ , and  $\beta$ . They form seven different combinations with the molecular weight of 180 kDa [31]. The second class of storage protein, glycinin, is hexameric with molecular weights of 360 kDa and consists of acidic (A) and basic (B) polypeptides. Glycinin is encoded by five non-allelic genes, Gy1, Gy2, Gy3, Gy4, and Gy5. It codes for five precursor protein molecules, G1, G2, G3, G4, and G5, respectively [32]. Based on physical properties, these five subunits are classified into two distinct major groups; group I consists of G1 (A1aBx), G2 (A2B1a), and G3 (A1aB1b) proteins and group II contains G4 (A5A4B3) and G5 (A3B4) subunits. Beilinson et al. [33] identified additional two genes, glycinin pseudogene (*gy6*) and functional gene (*Gy7*).

In this investigation a total of 30 differentially expressed protein spots were detected over pH ranges 4-7 and 6-11. Among the 30 spots, a total of 21 spots were identified as storage proteins. More specifically, 6 protein spots (spot#1-6) were identified as  $\alpha$  subunits of  $\beta$ -conglycinin, 11 protein spots (spot#7-17) were identified as  $\beta$ -subunits of  $\beta$ -conglycinin, and 4 spots (spot#18-21) were identified as glycinin G1 subunits. Based on the results, these protein spots are distributed over a wide range of pI and molecular weights and varied among the 3 transgenic events. Transgenic events 1, 2, and 3 had between 1.0 to 1.5 fold increase/decrease in protein abundance for the  $\alpha$ -subunit and  $\beta$ -subunit of  $\beta$ -conglycinin. The significance of the protein level of each storage protein spot among the control and three events were tested using ANOVA procedure of SAS [34], as multiple comparisons were involved, a Bonferroni correction at 1% threshold was used to determine the significance of each protein spot. If a significant difference was observed for any spot, a comparison of the control vs. each of the three events was conducted using Dunnett's test.

Table 1- includes results and statistics for the relative abundance of each protein spot for each of the 3 events compared with the non-transgenic control. There was a broad range of variability of both acidic and basic storage proteins components among the transgenic events. Gomes et al. [35] investigated alpha subunit of  $\beta$ -conglycinin spot variation in four conventional soybean genotypes namely BRS 257, 258, 267 and Embrapa 48, using 2D-PAGE analysis and also by ID-PAGE. They showed variation of total number of protein spots in BRS 257, BRS258, Embrapa 48 and BRS 267 is 102, 124, 113, and 99 respectively. They are also reported 46 differentially expressed proteins (storage, allergenic, maturation, agglutinin, and trypsin inhibitors) in 2D gels among 4 non-GM soybeans. Similarly, significant differences of both  $\beta$ -conglycinin and glycinin storage proteins using proteomics was reported in 14 Canadian commercial soybean varieties by Zarkada et al. [36] and in 90 Brazilian soybean cultivars by [37]. Houston et al. [38] quantified soybean allergens in 20 non-GM soybeans and observed

Savi #	Spot #	Control	Event1	Event2	Event3	Bonferroni-corrected p-value
		Abundance	Abundance	Abundance	Abundance	
1	2275	8.91E+05	2.05E+06	1.16E+06	1.08E+06	1
2	1402	1.54E+06	1.68E+06	1.69E+06	1.97E+06	1
3	584	7.63E+05	7.14E+05	1.10E+06	7.96E+05	0.3
4	412	1.68E+06	2.50E+06	1.34E+06	2.55E+06	0.1
5	415	2.19E+06	3.21E+06	2.06E+06	3.30E+06	0.93
6	912	6.98E+06	6.44E+06	9.40E+06	7.54E+06	0.35
7	873	3.82E+07	3.40E+07	4.90E+07	3.97E+07	0.06
8	958	4.98E+05	4.26E+05	6.99E+05	5.46E+05	0.18
9	862	8.15E+06	7.58E+06	1.28E+07	1.04E+07	0.07
10	1397	1.26E+06	1.18E+06	1.67E+06	1.59E+06	0.06
11	2295	1.30E+07	1.28E+07	2.05E+07	1.65E+07	0.02
12	905	4.73E+06	4.72E+06	7.32E+06	6.23E+06	0.16
13	856	3.24E+06	3.33E+06	5.33E+06	4.52E+06	0.14
14	904	2.08E+06	2.29E+06	3.58E+06	2.90E+06	0.27
15	2276	5.83E+05	5.10E+05	8.46E+05	7.20E+05	0.05
16	973	2.99E+05	3.35E+05	4.40E+05	4.13E+05	1
17	1159	2.52E+06	2.09E+06	3.19E+06	2.84E+06	0.11
18	1462	2.87E+06	3.00E+06	2.03E+06	2.56E+06	0.03
19	419	9.09E+06	7.14E+06	8.46E+06	9.22E+06	1
20	400	8.55E+07	8.31E+07	6.66E+07	7.22E+07	0.33
21	411	4.24E+06	3.25E+06	2.98E+06	3.49E+06	0.08
22	1124	3.21E+06	2.75E+06	4.05E+06	3.60E+06	0.01**
23	1118	3.02E+06	2.91E+06	3.92E+06	3.25E+06	0.11
24	1543	4.76E+05	1.12E+06	9.09E+05	1.02E+06	0.07
25	952	4.31E+05	5.94E+05	5.78E+05	5.45E+05	1
26	1060	8.28E+05	6.77E+05	1.15E+06	7.66E+05	0.01**
27	989	9.09E+06	7.66E+06	6.20E+06	6.39E+06	0.22
28	577	1.50E+06	2.05E+06	1.52E+06	1.64E+06	1
29	732	1.44E+06	1.87E+06	1.50E+06	1.29E+06	0.51
30	538	5.36E+06	7.42E+06	4.60E+06	5.56E+06	0.04

\*\* significant at 1% probability level

**Table 1:** Average protein levels in the seeds of control and transgenic soybeans and the Bonferroni-corrected p-value

10-fold variation of glycinin G3 when comparing 2 different varieties. Multiple authors have reported variation in protein profiles in different cultivars of soybeans grown under different environmental conditions using different proteomics techniques [37,39, 40]. In addition, Hughes and Murphy [41] reported that glycinin content varies from 7.5% for soybean grown in a uniform environment to 10% for soybeans grown at different locations. The previous publications from our laboratory also showed variation of storage proteins, allergen, and anti-nutritional proteins among several wild and cultivated soybean genotypes [42,43]. In another study, Fehr et al. [44] reported environmental effects caused significant variation of glycinin and  $\beta$ -conglycinin with the exception of A3 subunit of glycinin observed in 14 genotypes grown in 8 locations for three years. In addition, Helm et al. [45] also reported variation of above proteins due to environment and genotype and the magnitude of the differences between the 2 cultivars for 11S/7S ratio varied among locations within years. Piper and Boote [46] reported that maturity group, location, and environmental variation affect characteristics of soybean seeds, which suggested temperature has a significant effect on protein expression [47]. Koo et al. [48] reported 19 differentially expressed protein spots of P34, storage, and seed maturation related proteins identified by immunoblot and MALDI-TOF/TOF analysis between PI 567476 and Clark accessions. The authors suggested that  $\beta$ -conglycinin subunits are products of a multigene family, and the variation in the distribution of protein spots could be due to

post-translational modifications such as a sequence of glycosylation, deglycosylation, and proteolysis [49]. Several other transgenic studies in soybeans and other crops have been also reported in the literature, which concluded that the occurred variation is within the natural variation of conventional cultivars [13,22,50,51].

### Variation of other proteins

Other differentially expressed proteins found in this investigation were involved in primary metabolic processes and synthesis of nucleotides and other secondary compounds including: 2 spots of isoflavone reductase (spot# 22, 23); one spot of cysteine synthase (spot#24), one spot of Pv42P (*Phaseolus vulgaris*) (spot# 25), one spot of quinine oxidoreductase-like protein (spot#26), one spot of alcohol dehydrogenase (spot# 27); one spot of nucleoside diphosphate kinase (spot# 28); one spot of peptidyl-prolylcis-trans Isomerase 1 (spot# 29); one spot of seed maturation proteins (spot# 30). Except for the spots 22 and 26, no significant difference of the protein level was observed among the control, and the three events at the 1% probability level. For the spots 22 and 26, only the difference between Control and event 2 was significant, the significance of Control vs. Event 3 and Control vs. Event 4 were not significant. The authors concluded that no significant changes were detected between transgenic and non-transgenic lines and the differences occurred in 2 spots fell in the range of natural variation.

Spot #	Protein	PI #	MOWSE Score	# of PM	Mr (Da)
1	Beta-conglycinin, alpha chain	gi 121281	669	7	121281
2	Beta-conglycinin, alpha chain-like [Glycine max]	gi 356535993	67	1	68392
3	Alpha' subunit of beta-conglycinin [Glycine max]	gi 9967361	277	2	65160
4	Alpha' subunit of beta-conglycinin [Glycine max]	gi 9967361	566	6	65160
5	Alpha' subunit of beta-conglycinin [Glycine max]	gi 9967361	533	6	65160
6	Beta-conglycinin alpha subunit [Glycine max]	gi 335353923	943	11	70451
7	*Beta-conglycinin beta subunit [Glycine max]	F7J077	1364	18	50442
8	*Beta-conglycinin beta subunit [Glycine max]	F7J077	529	5	50442
9	*Beta-conglycinin beta subunit [Glycine max]	F7J077	1462	20	50442
10	Beta-conglycinin beta subunit [Glycine max]	gi 341603995	299	3	50010
11	Beta-conglycinin beta subunit [Glycine max]	gi 63852207	1523	19	48358
12	Beta-conglycinin beta subunit [Glycine max]	gi 341603995	906	11	50010
13	Beta-conglycinin beta subunit [Glycine max]	gi 63852207	881	13	48358
14	Beta-conglycinin beta subunit [Glycine max]	gi 341603995	432	3	50010
15	Beta-conglycinin beta subunit [Glycine max]	gi 341603995	911	12	50010
16	Beta-conglycinin beta subunit [Glycine max]	gi 341603995	172	1	50010
17	Beta-conglycinin beta subunit [Glycine max]	gi 341603995	908	11	50010
18	Glycinin G1 precursor [Glycine max]	gi 121276	100	1	56299
19	Glycinin A3B4 [Glycine max]	gi 225440	521	6	27447
20	Glycinin G1 precursor [Glycine max]	gi 121276	464	5	56299
21	Glycinin G1 precursor [Glycine max]	gi 121276	405	5	56299
22	Isoflavone reductase homolog 2 [Glycine max]	gi 351726399	811	9	33919
23	Isoflavone reductase homolog 2 [Glycine max]	gi 351726399	826	11	33919
24	Cysteine synthase [Glycine max]	gi 351727525	858	6	34362
25	*Pv42p [Phaseolus vulgaris]	Q41108	119	1	41312
26	Quinone oxidoreductase-like protein At1g23740	gi 356571378	180	1	34660
27	*Alcohol dehydrogenase 1 [Glycine max]	Q8LJR2	501	4	40007
28	Nucleoside diphosphate kinase [Glycine max]	gi 26245395	252	3	16402
29	Peptidyl-prolyl cis-trans isomerase 1 / Cyclophilin 1 [Glycine max]	gi 75304514	359	4	18395
30	Seed maturation protein PM30 [Glycine max]	gi 351727184	228	2	15145

\* Refers to identity based on BLASTp on resulting sequence from MASCOT

Table 2: Identification of differentially expressed protein between control and transgenic soybean seed varietes

## Variation in isoflavones and phenolic acids content

Isoflavonoids are the principal group of phenolic compounds present in soybeans. Lee et al. [52] extracted soybean protein and investigated the effects of year, location, and genotype on soybean isoflavones. The authors concluded that environmental and genotype effects were the most important sources of isoflavones content variation in soybeans. In another study, Wei et al. [53] investigated the differences in isoflavones content in locally grown soybeans and genetically modified imported soybeans. They reported that the isoflavone content of imported genetically modified soybeans were similar to the regular Taiwanese grown soybeans. Taie et al. [54] reported that varying levels of organic fertilization produced significant influence on the phenolic content and antioxidant activity of soybeans. Similarly, influence of postharvest and storage conditions on phenolic content was reviewed by Amarowicz et al. [55]. The authors revealed that the variation in polyphenol content was often negligible as compared to the differences in content between various plant varieties.

In the present study, we compared the isoflavone content from four soybean samples and the results are presented in (Table 3A). Identification of isoflavones was achieved by comparison of retention times, ultraviolet, and mass spectral data with authentic commercial standards or results published in the literature. The six isoflavones were identified as daidzein, glysetin, glysetein, daidzin, genistin, and genistein. Four additional peaks were tentatively identified as malonyl

and acetyl conjugates of the three isoflavones at retention times 20.4, 20.9, 24.6, and 26.9 mins respectively. These compounds were present in all four soybean samples. Similar compounds have been identified and reported previously [56]. Quantification in the present study was achieved by comparing peak areas under the curve for the identified isoflavones. The relative standard deviation for four replicate analyses for most samples was less than 5%. There were minor differences in individual isoflavones between control and transgenic samples. The total quantity of isoflavones extracted from four different samples varied between 10-20%. Based on the statistical data no systematic variation in different isoflavones content was observed between different events and control sample. For example, two isoflavonoids, genistein and daidzein, showed no significant differences between event 1 and control but showed difference among events 1, 2, 3 and control at the 1% probability level. We concluded that no consistent trend in genetic variation was observed in our study with isoflavones, but the results indicated that the variation is due to environmental effects since the seeds used in this investigation were harvested in different times. Berman et al. [57] reported that the second generations of transgenic soybeans were compositionally equivalent to the conventional soybeans. However comparative large variations in isoflavone content that ranged from 551-7584 µg/g with an average of 2973 µg/g were reported in soybeans grown in Northern and Southern China from maturity groups 0-6 [58]. Similar high variations in isoflavones content in different cultivars and growing conditions in Kansas grown soybeans were also reported [59].

Table 3A Isoflavone	Control			Event 1			Event 2			Event 3			Bonferroni-corrected p-value	Trait significance
	Average	St Dev	%RSD	Average	St Dev	%RSD	Average	St Dev	%RSD	Average	St Dev	%RSD		
Diadzin (Rt 11.6 min; n=4)	153.250	4.605	3.005	210.450	3.841	1.825	186.500	3.080	1.651	204.625	2.786	1.362	<0.001	(1,3),2,C*
Glycetin (Rt 12.8 min; n=4)	131.500	6.165	4.688	144.875	5.652	3.901	124.175	4.388	3.534	122.025	3.242	2.657	0.002	
Genistin (Rt 18.7 min; n=4)	328.675	12.678	3.857	413.325	3.783	0.915	383.850	10.094	2.630	365.450	8.721	2.386	<0.001	(1,2),3,C*
Diadzein (Rt 30.2 min; n=4)	50.525	2.644	5.232	50.800	3.477	6.844	83.200	2.810	3.377	82.925	4.172	5.032	<0.001	(2,3),(1,C)*
Glycetein (Rt 31.6 min; n=4)	15.275	2.689	17.602	18.925	1.374	7.263	23.100	1.608	6.962	20.700	1.610	7.780	0.009	
Genistein (Rt 41.0 min; n=4)	64.425	3.075	4.773	65.750	1.686	2.565	117.650	2.983	2.535	78.500	4.511	5.747	<0.001	2,3,(1,C)*
Conjugate - 1 (Rt 20.4 min; n=4)	211.150	5.910	2.799	218.225	7.049	3.230	255.575	6.282	2.458	260.550	4.064	1.560	<0.001	1,(3,2),C*
Conjugate - 2 (Rt 20.9 min; n=4)	134.550	6.202	4.610	151.150	12.256	8.108	123.300	17.664	14.326	134.425	2.333	1.735	0.462	
Conjugate - 3 (Rt 24.6 min; n=4)	11.150	1.168	10.472	9.800	0.548	5.589	9.500	0.383	4.031	13.725	0.377	2.750	<0.001	3,(1,2,C)*
Conjugate - 4 (Rt 26.9 min; n=4)	608.025	13.237	2.177	760.050	8.421	1.108	701.800	11.453	1.632	707.650	17.885	2.527	<0.001	1,(3,2),C*

Table 3B Phenolic acid	Control		Event 1		Event 2		Event 3		Bonferoni corrected p-value	Trait significance
	Average (mg/g)	%RSD	Average (mg/g)	%RSD	Average (mg/g)	%RSD	Average (mg/g)	%RSD		
Vanilic acid (Rt 27.8 min; n=4)	0.062	4.602	0.065	3.663	0.063	4.731	0.068	2.546	0.002	
Syringic acid (Rt 31.4 min; n=4)	0.242	3.254	0.276	3.707	0.266	2.053	0.275	1.349	0.236	
<i>p</i> - Coumaric acid (Rt 38.5 min; n=4)	0.070	1.964	0.082	2.815	0.108	1.854	0.087	1.629	< 0.001	2,(3,1),C*
Ferulic acid (Rt 42.2 min; n=4)	0.083	2.544	0.084	2.035	0.077	2.165	0.091	2.055	0.012	
Sinapic acid (Rt 43.7 min; n=4)	0.089	2.255	0.093	2.537	0.098	3.431	0.101	2.286	0.014	

**Table 3:** LC-MS analysis of phenolic compounds from control and genetically modified soybean. (A) HPLC area of four isoflavones was used for comparison and four separate extractions and analysis were carried out for each sample. (B) Quantification of five phenolic acids was achieved using authentic commercial standards and four separate extractions and analysis were carried out for each sample.

It is well documented in the literature that the phenolic acid content in seeds of several crops is influenced by cultivars, growing conditions, and the methodologies used for analysis [28, 60-62]. Table 3B includes information for the identity and quantification of phenolic acids from the four soybean samples. Identification of phenolic acids was achieved by comparison of retention times, ultraviolet and mass spectral data with authentic commercial standards. The five prominent phenolic acids were identified as vanillic, syringic, *p*-coumaric, ferulic, and sinapic acids. In addition, there were four minor peaks detected at retention times 14.4, 47.2, 50.6, and 56.4 mins respectively. These compounds were present in minor amounts in all four soybean samples. The total quantity of phenolic acids extracted from four different samples ranged between 0.55 mg/g to 0.62 mg/g. The relative standard deviation of four replicate analyses was below 5%. Similar phenolic acids profiles have been reported in soybean, flaxseed, and olives [63]. However, our results are different from the data published by Taie et al. [54], where authors observed chlorogenic acid in addition to the other phenolic acids in soybeans grown under organic growing conditions. Based on the statistical data, only one out of five acids, *p*-coumaric acid, showed significant difference among transgenic events 2, 3, 1 and control at the 1% probability level with no consistent trend. The variations in the phenolic phytochemicals content observed in the present study are within the natural range of variations observed in various conventional cultivars grown and processed under different environmental conditions or analyzed by different methods.

## Conclusions

We separated soybean proteins using 2D-PAGE and identified 30 proteins that appeared to be differentially expressed in at least one of 3 transgenic events in comparison to non-transgenic control seeds. Each of these proteins (spots) was excised from the gels and their identity determined with MALDI-TOF/TOF tandem mass spectrometry. Phenolic acids and isoflavones were extracted and analyzed by LC-MS analysis. The results indicated that minor variations in proteins, isoflavones, and phenolic acids profiles exist between control and transgenic soybean seeds. The variations observed in the present study are generally within the natural range of variations observed in conventional cultivars grown and processed under different environmental conditions. However, additional long term studies with different cultivars grown over multiple years with more transgenic insertions with their isogenic lines are needed to precisely evaluate the impact of genetic transformation. Detailed metabolomic profile analysis looking at multiple classes of phytochemicals in both control and transgenic lines are needed to confirm the current results.

## Acknowledgement

We thank Dr. Richard Sicher for critical review of this manuscript. Funding for this research was provided by ARS project 1245-21220-232-00D. 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.

## References

1. Badger TM, Ronis MJ, Simmen RC, Simmen FA (2005) Soy protein isolate and protection against cancer. *J Am Coll Nutr* 24: 146S-149S.
2. Friedman M, Brandon DL (2001) Nutritional and health benefits of soy proteins. *J Agric Food Chem* 49: 1069-1086.
3. Clemente A, Moreno FJ, Marin-Manzano Mdel C, Jimenez E, Domoney C (2010) The cytotoxic effect of Bowman-Birk isoinhibitors, IBB1 and IBB2, from soybean (*Glycine max*) on HT29 human colorectal cancer cells is related to their intrinsic ability to inhibit serine proteases. *Mol Nutr Food Res* 54: 396-405.
4. Messina M, Barnes S (1991) The role of soy products in reducing risk of cancer. *J Natl Cancer Inst* 83: 541-546.
5. Rowland I, Faughnan M, Hoey L, Wahala K, Williamson G, et al. (2003) Bioavailability of phyto-oestrogens. *Br J Nutr* 89 Suppl 1: S45-58.
6. Aluru MR, Rodermerl SR, Reddy MB (2011) Genetic modification of low phytic acid 1-1 maize to enhance iron content and bioavailability. *J Agric Food Chem* 59: 12954-12962.
7. Cunha NB, Murad AM, Cipriano TM, Araujo AC, Aragao FJ, et al. (2011) Expression of functional recombinant human growth hormone in transgenic soybean seeds. *Transgenic Res* 20: 811-826.
8. Cunha NB, Murad AM, Ramos GL, Maranhao AQ, Brigido MM, et al. (2011) Accumulation of functional recombinant human coagulation factor IX in transgenic soybean seeds. *Transgenic Res* 20: 841-855.
9. Herman EM, Helm RM, Jung R, Kinney AJ (2003) Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol* 132: 36-43.
10. Clive J (2013) Global Status of Commercialized Biotech/GM Crops: 2013. ISAAA Brief 46.
11. Goodman RE, Panda R, Ariyaratna H (2013) Evaluation of endogenous allergens for the safety evaluation of genetically engineered food crops: review of potential risks, test methods, examples and relevance. *J Agric Food Chem* 61: 8317-8332.
12. Millstone E, Brunner E, Mayer S (1999) Beyond 'substantial equivalence'. *Nature* 401: 525-526.
13. Cheng KC, Beaulieu J, Iqura E, Belzile FJ, Fortin MG, et al. (2008) Effect of transgenes on global gene expression in soybean is within the natural range of variation of conventional cultivars. *J Agric Food Chem* 56: 3057-3067.
14. Kleter GA, Kuiper HA (2002) Considerations for the assessment of the safety of genetically modified animals used for human food or animal feed. *Livestock production science* 74: 275-285.
15. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178-183.
16. Barbosa HS, Arruda SCC, Azevedo RA, Arruda MAZ (2012) New insights on proteomics of transgenic soybean seeds: evaluation of differential expressions of enzymes and proteins. *Anal Bioanal Chem* 402: 299-314.
17. Brandao AR, Barbosa HS, Arruda MA (2010) Image analysis of two-dimensional gel electrophoresis for comparative proteomics of transgenic and non-transgenic soybean seeds. *J Proteomics* 73: 1433-1440.
18. Nordlee JA, Taylor SL, Townsend JA, Thomas LA, Bush RK (1996) Identification of a Brazil-nut allergen in transgenic soybeans. *New Engl J Med* 334: 688-692.
19. Sorin C, Negroni L, Balliau T, Corti H, Jacquemot MP, et al. (2006) Proteomic analysis of different mutant genotypes of *Arabidopsis* led to the identification of 11 proteins correlating with adventitious root development. *Plant Physiol* 140: 349-364.
20. Barros E, Lezar S, Anttonen MJ, van Dijk JP, Rohlig RM, et al. (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnol J* 8: 436-451.
21. Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, et al. (2004) Unintended effects and their detection in genetically modified crops. *Food Chem Toxicol* 42: 1089-1125.
22. Lehesranta SJ, Davies HV, Shepherd LV, Nunan N, McNicol JW, et al. (2005) Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. *Plant Physiol* 138: 1690-1699.
23. Zolla L, Rinalducci S, Antonioli P, Righetti PG (2008) Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. *Journal of proteome research* 7: 1850-1861.
24. Tanaka A, Mita S, Ohta S, Kyojuka J, Shimamoto K, et al. (1990) Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an increased level of mRNA and an efficient splicing of the intron. *Nucleic Acids Res* 18: 6767-6770.
25. Thompson CJ, Movva NR, Tizard R, Cramer R, Davies JE, et al. (1987) Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. *EMBO J* 6: 2519-2523.
26. Kay R, Chan A, Daly M, McPherson J (1987) Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. *Science* 236: 1299-1302.
27. Natarajan S, Xu CP, Caperna TJ, Garrett WA (2005) Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal Biochem* 342: 214-220.
28. Luthria DL, Pastor-Corrales MA (2006) Phenolic acids content of fifteen dry edible bean (*Phaseolus vulgaris* L.) varieties. *J Food Compos Anal* 19: 205-211.
29. Memon AA, Memon N, Bhangar MI, Luthria DL (2013) Assay of phenolic compounds from four species of ber (*Ziziphus mauritiana* L.) fruits: comparison of three base hydrolysis procedure for quantification of total phenolic acids. *Food Chem* 139: 496-502.
30. Kim MJ, Kim JK, Kim HJ, Pak JH, Lee JH, et al. (2012) Genetic modification of the soybean to enhance the beta-carotene content through seed-specific expression. *PLoS One* 7: e48287.
31. Thanh VH, Shibasaki K (1976) Heterogeneity of beta-conglycinin. *Biochim Biophys Acta* 439: 326-338.
32. Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallan BJ, et al. (1989) Characterization of the glycinin gene family in soybean. *Plant Cell* 1: 313-328.
33. Beilinson V, Moskalenko OV, Livingstone DS, Reverdatto SV, Jung R, et al. (2002) Two subtilisin-like proteases from soybean. *Physiol Plant* 115: 585-597.
34. SAS-Institut-Inc. (2012) SAS software 9.4.
35. Gomes LS, Senna R, Sandim V, Silva-Neto MA, Perales JE, et al. (2014) Four conventional soybean [*Glycine max* (L.) Merrill] seeds exhibit different protein profiles as revealed by proteomic analysis. *J Agric Food Chem* 62: 1283-1293.
36. Zarkadas CG, Gagnon C, Gleddie S, Khanizadeh S, Cober ER, et al. (2007) Assessment of the protein quality of fourteen soybean [*Glycine max* (L.) Merr.] cultivars using amino acid analysis and two-dimensional electrophoresis. *Food Res Int* 40: 129-146.
37. Carrao-Pannizi MC, Kwanyuen P, Erhan SZ, Lopes ION (2008) Genetic variation and environmental effects on beta-conglycinin and glycinin content in Brazilian soybean cultivars. *Pesq agropec bras Brasilia* 43: 1105-1114.
38. Houston NL, Lee DG, Stevenson SE, Ladics GS, Bannon GA, et al. (2011) Quantitation of soybean allergens using tandem mass spectrometry. *Journal of proteome research* 10: 763-773.
39. Murphy PA, Resurreccion AP (1984) Varietal and environmental differences in soybean glycinin and beta-conglycinin content. *J Agric Food Chem* 32: 911-915.
40. Zilic SM, Barac MB, Pesic MB, Mladenovic Drinic SD, Ignjatovic-Micic DD, et al. (2011) Characterization of proteins from kernel of different soybean varieties. *J Sci Food Agric* 91: 60-67.
41. Hughes SA, Murphy PA (1983) Varietal influence on the quantity of glycinin in soybeans. *J Agric Food Chem* 31: 376-379.
42. Natarajan SS, Xu C, Bae H, Caperna TJ, Garrett WM (2006) Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J Agric Food Chem* 54: 3114-3120.
43. Xu CP, Caperna TJ, Garrett WM, Cregan P, Bae HH, et al. (2007) Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. *J Sci Food Agric* 87: 2511-2518.
44. Fehr WR, Hoek JA, Johnson SL, Murphy PA, Nott JD, et al. (2003) Genotype and environment influence on protein components of soybean. *Crop Sci* 43: 511-514.



45. Helm R, Cochrell G, Herman E, Burks A, Simpson HA et al. (1998) Cellular and molecular characterization of a major soybean allergen. *Int Arch Allergy Immunol* 117: 29-37.
46. Piper EL, Boote KJ (1999) Temperature and cultivar effects on soybean seed oil and protein concentrations. *J Am Oil Chem Soc* 76: 1233-1241.
47. Gibson LR, Mullen RE (1996) Soybean seed composition under high day and night growth temperatures. *J Am Oil Chem Soc* 73: 733-737.
48. Koo SC, Bae DW, Seo JS, Park KM, Choi MS, et al. (2011) Proteomic analysis of seed storage proteins in low allergenic soybean accession. *J Korean Soc Appl. Biol Chem* 54: 332-339.
49. Sengupta C, Deluca V, Bailey DS, Verma DP (1981) Post-translational processing of 7S and 11S components of soybean storage proteins. *Plant Mol Biol* 1: 19-34.
50. Natarajan SS, Luthria D, Bae H, Lakshman D, Mitra A (2013) Transgenic soybean and soybean protein analysis: An overview. *J Agric Food Chem* 61: 11736-11743.
51. Ruebelt MC, Lipp M, Reynolds TL, Astwood JD, Engel KH, et al. (2006) Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 2. Assessing natural variability. *J Agric Food Chem* 54: 2162-2168.
52. Lee SJ, Yan W, Ahn JK, Chung IM (2003) Effects of year, site, genotype and their interactions on various soybean isoflavones. *Field Crop Res* 81: 181-192.
53. Wei Q, Jone W, Fang T (2004) Study on isoflavones isomers contents in Taiwan's soybean and GM soybean. *Journal of Food and Drug Analysis* 12: 324-331.
54. Taie HAA, El-Mergawi R, Radwan S (2008) Isoflavonoids, flavonoids, phenolic acids profiles and antioxidant activity of soybean seeds as affected by organic and bioorganic fertilization. *Am-Eurasian J Ag Env Sci* 4: 207-213.
55. Amarowicz R, Carle R, Dongowski G, Durazzo A, Galensa R, et al. (2009) Influence of postharvest processing and storage on the content of phenolic acids and flavonoids in foods. *Mol Nutr Food Res* 53: S151-183.
56. Rostagno MA, Palma M, Barroso CG (2003) Ultrasound-assisted extraction of soy isoflavones. *J Chromatogr A* 1012: 119-128.
57. Berman KH, Harrigan GG, Riordan SG, Nemeth MA, Hanson C, et al. (2010) Compositions of forage and seed from second-generation glyphosate-tolerant soybean MON 89788 and insect-protected soybean MON 87701 from Brazil are equivalent to those of conventional soybean (*Glycine max*). *J Agric Food Chem* 58: 6270-6276.
58. Zhang J, Ge Y, Han F, Li B, Yan S, et al. (2014) Isoflavone Content of Soybean Cultivars from Maturity Group 0 to VI Grown in Northern and Southern China. *J Am Oil Chem Soc* 91: 1019-1028.
59. Swanson M, Stoll M, Schapaugh W, Takemoto L (2004) Isoflavone content of Kansas soybeans. *Am J Undergrad Res* 2: 27-32.
60. Kim JK, Kim EH, Lee OK, Park SY, Lee B et al. (2013) Variation and correlation analysis of phenolic compounds in mungbean (*Vigna radiata* L.) varieties. *Food Chem* 141: 2988-2997.
61. Luthria DL, Mukhopadhyay S, Krizek DT (2006) Content of total phenolics and phenolic acids in tomato (*Lycopersicon esculentum* Mill.) fruits as influenced by cultivar and solar UV radiation. *J Food Compos Anal* 19: 771-777.
62. Mpofo A, Sapirstein HD, Beta T (2006) Genotype and environmental variation in phenolic content, phenolic acid composition, and antioxidant activity of hard spring wheat. *J Agric Food Chem* 54: 1265-1270.
63. Alu'datt MH, Rababah T, Ereifej K, Alli I (2013) Distribution, antioxidant and characterisation of phenolic compounds in soybeans, flaxseed and olives. *Food Chem* 139: 93-99.