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

Savithiry S Natarajan*, Farooq H Khan¹, Davanand L Luthria², Mark L Tucker³, QijianSong¹ and Wesley M Garrett⁴

1 USDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA
2 USDA-ARS, Food Composition and Methods Development Laboratory, Beltsville, MD 20705, USA
3 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 β-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 isogene (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 quinine 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: β-conglycinin and glycinin. Other proteins such as β-amylose, 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

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


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 genomics, 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 protocine 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 bpGmpPG11a gene (Glyma12g01480) promoter to a β-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 downstream of the start of translation [24].

The PG11a-GUS-NOS3 construct was cloned into the H1- Eco RI site of pTFT101.1 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 was transformed and regenerated transgenic soybean (Glycinel 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) β-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) β-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 at 20,800 g for 20 min at 4°C. The supernatant containing the solubilized protein was used for each of the 2D-PAGE separations.

**2D-PAGE analysis**

Protein (100 µg) was first separated by isoelectric focusing (IEF) on an ICPHpor II apparatus using13 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 µL rehydration buffer (7M urea, 2M thiourea, 4% CHAPS, 2% pharmalyte, 0.002% bromophenol blue) containing 100µg 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 electrophoresis on a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ). The gels were then stained with colloidal Coomassie250. After de-staining with ddH2O, 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 µL of 10 µg/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 a-cyanohydroxycinnamic acid (CHCA) matrix prepared in 70% acetonitrile containing 0.1% trifluoroacetic acid. Thirty fmol of a commercially prepared trypptic 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:H2O: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 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 C18 Luna column (Phenomenex, Lorance, 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.
soybean seed proteins are shown in (Figure 2) (Table 1). Comparison of another with a pH range of 6 to 11. Representative 2D-PAGE gels for was used for two separate 2D gels, one with a pH range from 4 to 7 and between 4 to 7 and 6 to 11. Therefore, 100 µg of each protein extract total numbers of resolved protein spots were insufficient. Improved this broad pH range, storage proteins were poorly separated and the strips using a pH range of 3 to 10 (data not shown). However, over Analysis was initially carried out with immobilized pH gradient (IPG) three transgenic events and control lines were separated by 2D-PAGE. Analysis of proteins [27]. Three biological replicates of protein extracts from events and control soybean seeds tested in our current investigation. Seed promoter is expressed throughout the soybean seed with the highest [30]. Similarly, we demonstrated using GUS staining that the 35S promoter was expressed strongly in leaves but very weakly in seed 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/aceton 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 digest 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 β-conglycinin and glycinin, respectively. The β-conglycinins are trimeric glycoproteins consisting of three types of non-identical, but homologous polypeptide subunits: α, α′, and β. 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, Gγ1, Gγ2, Gγ3, Gγ4, and Gγ5. 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 (A1A8x), G2 (A2B1a), and G3 (A1aB1b) proteins and group II contains G4 (A5A4B3) and G5 (A3B4) subunits. Beillinson et al. [33] identified additional two genes, glycinin pseudogene (gγ6) and functional gene (Gγ7).

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 a subunits of β-conglycinin, 11 protein spots (spot#7-17) were identified as of β-subunits of β-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 α-subunit and β-subunit of β-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 β-conglycinin spot variation in four conventional soybean genotypes namely BRS 257, 258 and 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 β-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
the variation in the distribution of protein spots could be due to 
that β-conglycinin subunits are products of a multigene family, and 
between PI 567476 and Clark accessions. The authors suggested 
proteins identified by immunoblot and MALDI-TOF/TOF analysis 
expressed protein spots of P34, storage, and seed maturation related 
on protein expression [47]. Koo et al. [48] reported 19 differentially 
soybean seeds, which suggested temperature has a significant effect 
group, location, and environmental variation affect characteristics of 
locations within years. Piper and Boote [46] reported that maturity 
above proteins due to environment and genotype and the magnitude of 
for three years. In addition, Helm et al. [45] also reported variation of 
A3 subunit of glycinin observed in 14 genotypes grown in 8 locations 
significant variation of glycinin and β-conglycinin with the exception of 
in another study, Fehr et al. [44] reported environmental effects caused 
proteins among several wild and cultivated soybean genotypes [42,43]. 
also showed variation of storage proteins, allergen, and anti-nutritional 
at different locations. The previous publications from our laboratory 
soybean grown in a uniform environment to 10% for soybeans grown 
and Murphy [41] reported that glycinin content varies from 7.5% for 
using different proteomics techniques [37,39, 40]. In addition, Hughes 
cultivars of soybeans grown under different environmental conditions 
10-fold variation of glycinin G3 when comparing 2 different varieties. 
 Multiple authors have reported variation in protein profiles in different 
with the exception of 
changes were detected between transgenic and non-transgenic lines and 
2 spots fell in the range of natural variation. The authors concluded that no significant 
was significant, the significance of Control vs. Event 3 and Control vs. 
among the control, and the three events at the 1% probability level. For 
(22,23); one spot of cysteine synthase 
morphological processes and synthesis of 
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 
and other secondary compounds including: 2 spots of 
isovalerate reductase (spot# 22, 23); one spot of cysteine synthase 
(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 
(spott 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.

<table>
<thead>
<tr>
<th>Savi #</th>
<th>Spot #</th>
<th>Control</th>
<th>Event1</th>
<th>Event2</th>
<th>Event3</th>
<th>Bonferroni-corrected p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Abundance</td>
<td>Abundance</td>
<td>Abundance</td>
<td>Abundance</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2275</td>
<td>8.91E+05</td>
<td>2.05E+06</td>
<td>1.16E+06</td>
<td>1.08E+06</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1402</td>
<td>1.54E+06</td>
<td>1.68E+06</td>
<td>1.69E+06</td>
<td>1.97E+06</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>584</td>
<td>7.63E+05</td>
<td>7.14E+05</td>
<td>1.10E+06</td>
<td>7.96E+05</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>412</td>
<td>1.68E+06</td>
<td>2.50E+06</td>
<td>1.34E+06</td>
<td>2.55E+06</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>415</td>
<td>2.19E+06</td>
<td>3.21E+06</td>
<td>2.06E+06</td>
<td>3.30E+06</td>
<td>0.93</td>
</tr>
<tr>
<td>6</td>
<td>912</td>
<td>6.98E+06</td>
<td>6.44E+06</td>
<td>9.40E+06</td>
<td>7.54E+06</td>
<td>0.35</td>
</tr>
<tr>
<td>7</td>
<td>873</td>
<td>3.82E+07</td>
<td>3.40E+07</td>
<td>4.90E+07</td>
<td>3.97E+07</td>
<td>0.06</td>
</tr>
<tr>
<td>8</td>
<td>958</td>
<td>4.98E+05</td>
<td>4.26E+05</td>
<td>6.99E+05</td>
<td>5.46E+05</td>
<td>0.18</td>
</tr>
<tr>
<td>9</td>
<td>862</td>
<td>8.15E+06</td>
<td>7.58E+06</td>
<td>1.28E+07</td>
<td>1.04E+07</td>
<td>0.07</td>
</tr>
<tr>
<td>10</td>
<td>1397</td>
<td>1.26E+06</td>
<td>1.18E+06</td>
<td>1.67E+06</td>
<td>1.59E+06</td>
<td>0.06</td>
</tr>
<tr>
<td>11</td>
<td>2295</td>
<td>1.30E+07</td>
<td>1.28E+07</td>
<td>2.05E+07</td>
<td>1.65E+07</td>
<td>0.02</td>
</tr>
<tr>
<td>12</td>
<td>905</td>
<td>4.73E+06</td>
<td>4.72E+06</td>
<td>7.32E+06</td>
<td>6.23E+06</td>
<td>0.16</td>
</tr>
<tr>
<td>13</td>
<td>856</td>
<td>3.24E+06</td>
<td>3.33E+06</td>
<td>5.33E+06</td>
<td>4.52E+06</td>
<td>0.14</td>
</tr>
<tr>
<td>14</td>
<td>904</td>
<td>2.08E+06</td>
<td>2.29E+06</td>
<td>3.58E+06</td>
<td>2.90E+06</td>
<td>0.27</td>
</tr>
<tr>
<td>15</td>
<td>2276</td>
<td>5.83E+05</td>
<td>5.10E+05</td>
<td>8.46E+05</td>
<td>7.20E+05</td>
<td>0.05</td>
</tr>
<tr>
<td>16</td>
<td>973</td>
<td>2.99E+05</td>
<td>3.35E+05</td>
<td>4.40E+05</td>
<td>4.13E+05</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1159</td>
<td>2.52E+06</td>
<td>2.09E+06</td>
<td>3.19E+06</td>
<td>2.84E+06</td>
<td>0.11</td>
</tr>
<tr>
<td>18</td>
<td>1452</td>
<td>2.87E+06</td>
<td>3.00E+06</td>
<td>2.03E+06</td>
<td>2.56E+06</td>
<td>0.03</td>
</tr>
<tr>
<td>19</td>
<td>419</td>
<td>9.09E+06</td>
<td>7.14E+06</td>
<td>8.46E+06</td>
<td>9.22E+06</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>8.55E+07</td>
<td>8.31E+07</td>
<td>6.66E+07</td>
<td>7.22E+07</td>
<td>0.33</td>
</tr>
<tr>
<td>21</td>
<td>411</td>
<td>4.24E+06</td>
<td>3.25E+06</td>
<td>2.98E+06</td>
<td>3.49E+06</td>
<td>0.08</td>
</tr>
<tr>
<td>22</td>
<td>1124</td>
<td>3.21E+06</td>
<td>2.75E+06</td>
<td>4.05E+06</td>
<td>3.60E+06</td>
<td>0.01**</td>
</tr>
<tr>
<td>23</td>
<td>1118</td>
<td>3.02E+06</td>
<td>2.91E+06</td>
<td>3.92E+06</td>
<td>3.25E+06</td>
<td>0.11</td>
</tr>
<tr>
<td>24</td>
<td>1543</td>
<td>4.76E+05</td>
<td>1.12E+06</td>
<td>9.09E+05</td>
<td>1.02E+05</td>
<td>0.07</td>
</tr>
<tr>
<td>25</td>
<td>952</td>
<td>4.31E+05</td>
<td>5.94E+05</td>
<td>5.78E+05</td>
<td>5.45E+05</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>1060</td>
<td>8.28E+05</td>
<td>6.77E+05</td>
<td>1.15E+06</td>
<td>7.66E+05</td>
<td>0.01**</td>
</tr>
<tr>
<td>27</td>
<td>989</td>
<td>9.09E+06</td>
<td>7.66E+06</td>
<td>6.20E+06</td>
<td>6.39E+06</td>
<td>0.22</td>
</tr>
<tr>
<td>28</td>
<td>577</td>
<td>1.50E+06</td>
<td>2.05E+06</td>
<td>1.52E+06</td>
<td>1.64E+06</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>732</td>
<td>1.44E+06</td>
<td>1.87E+06</td>
<td>1.50E+06</td>
<td>1.29E+06</td>
<td>0.51</td>
</tr>
<tr>
<td>30</td>
<td>538</td>
<td>5.36E+06</td>
<td>7.42E+06</td>
<td>4.60E+06</td>
<td>5.56E+06</td>
<td>0.04</td>
</tr>
</tbody>
</table>

** significant at 1% probability level

Table 1: Average protein levels in the seeds of control and transgenic soybeans and the Bonferroni-corrected p-value
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 tentatively identified as daidzein, genistein, glycitein, 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 the content of isoflavones was observed between 1% and 10% probability level. We concluded that no consistent trend in genetic variation was shown among events 1, 2, 3 and control at the 1% probability level. We concluded that no consistent trend in genetic variation was shown among events 1, 2, 3 and control at the 1% probability level.

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, glycitein, glycitein, 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 the content of isoflavones was observed between 1% and 10% probability level. We concluded that no consistent trend in genetic variation was shown among events 1, 2, 3 and control at the 1% probability level. We concluded that no consistent trend in genetic variation was shown among events 1, 2, 3 and control at the 1% probability level.
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, -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, -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.

### Table 3A: Isoflavone

<table>
<thead>
<tr>
<th>Isoflavone</th>
<th>Control</th>
<th>Event 1</th>
<th>Event 2</th>
<th>Event 3</th>
<th>Bonferroni-corrected p-value</th>
<th>Trait significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycetein (Rt 11.6 min; n=4)</td>
<td>153.250</td>
<td>4.605</td>
<td>3.005</td>
<td>210.450</td>
<td>3.841</td>
<td>1.825</td>
</tr>
<tr>
<td>Glycinin (Rt 12.8 min; n=4)</td>
<td>131.500</td>
<td>6.165</td>
<td>4.688</td>
<td>144.875</td>
<td>5.652</td>
<td>3.901</td>
</tr>
<tr>
<td>Genistin (Rt 18.7 min; n=4)</td>
<td>328.675</td>
<td>12.678</td>
<td>3.857</td>
<td>413.325</td>
<td>3.783</td>
<td>0.915</td>
</tr>
<tr>
<td>Diadzin (Rt 30.2 min; n=4)</td>
<td>50.525</td>
<td>2.644</td>
<td>5.232</td>
<td>50.800</td>
<td>3.477</td>
<td>6.844</td>
</tr>
<tr>
<td>Glycitein (Rt 31.6 min; n=4)</td>
<td>15.275</td>
<td>2.689</td>
<td>17.602</td>
<td>18.925</td>
<td>1.374</td>
<td>7.263</td>
</tr>
<tr>
<td>Genistin (Rt 41.0 min; n=4)</td>
<td>64.425</td>
<td>3.075</td>
<td>4.773</td>
<td>65.750</td>
<td>1.686</td>
<td>2.565</td>
</tr>
<tr>
<td>Conjugate - 1 (Rt 20.4 min; n=4)</td>
<td>211.150</td>
<td>5.910</td>
<td>2.799</td>
<td>218.225</td>
<td>7.049</td>
<td>3.230</td>
</tr>
<tr>
<td>Conjugate - 2 (Rt 20.9 min; n=4)</td>
<td>134.550</td>
<td>6.202</td>
<td>4.610</td>
<td>151.150</td>
<td>12.256</td>
<td>8.108</td>
</tr>
<tr>
<td>Conjugate - 3 (Rt 24.6 min; n=4)</td>
<td>11.150</td>
<td>1.168</td>
<td>10.472</td>
<td>9.800</td>
<td>0.548</td>
<td>5.589</td>
</tr>
<tr>
<td>Conjugate - 4 (Rt 28.9 min; n=4)</td>
<td>608.025</td>
<td>13.237</td>
<td>2.177</td>
<td>760.050</td>
<td>8.421</td>
<td>1.108</td>
</tr>
</tbody>
</table>

### Table 3B: Phenolic acid

<table>
<thead>
<tr>
<th>Phenolic acid</th>
<th>Control</th>
<th>Event 1</th>
<th>Event 2</th>
<th>Event 3</th>
<th>Bonferroni-corrected p-value</th>
<th>Trait significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanillic acid (Rt 27.8 min; n=4)</td>
<td>Average (mg/g)</td>
<td>0.062</td>
<td>0.065</td>
<td>0.063</td>
<td>0.068</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Syringic acid (Rt 31.4 min; n=4)</td>
<td>Average (mg/g)</td>
<td>0.242</td>
<td>0.276</td>
<td>0.266</td>
<td>0.275</td>
<td>0.236</td>
</tr>
<tr>
<td>p-Coumaric acid (Rt 38.5 min; n=4)</td>
<td>Average (mg/g)</td>
<td>0.070</td>
<td>0.082</td>
<td>0.108</td>
<td>0.087</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ferulic acid (Rt 42.2 min; n=4)</td>
<td>Average (mg/g)</td>
<td>0.083</td>
<td>0.084</td>
<td>0.077</td>
<td>0.091</td>
<td>0.012</td>
</tr>
<tr>
<td>Sipinic acid (Rt 43.7 min; n=4)</td>
<td>Average (mg/g)</td>
<td>0.089</td>
<td>0.093</td>
<td>0.078</td>
<td>0.101</td>
<td>0.014</td>
</tr>
</tbody>
</table>

### 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.

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 245-21220-232-006. 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


