Metabolite changes in nine different soybean varieties grown under field and greenhouse conditions

K.M. Maria John^a, Savithiry Natarajan^{b#}, and Devanand L. Luthria^{a*}

^aBeltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland 20705, United States

^bSoybean Genomics and Improvement Laboratory, Plant Sciences Institute, ARS, USDA, Beltsville, MD 20705

*Corresponding author

Devanand Luthria, Ph.D., EMIB

Research Chemist

Food Composition Methods Development Lab

Bldg. 161, Lab 202, BARC (E)

10300 Baltimore Avenue

Agricultural Research Service, US Department of Agriculture

Beltsville, MD 20705

Tel: +1 301 504 7247; Fax: +1 301 504 8314

Email: Dave.Luthria@ars.usda.gov

[#]Co-Corresponding Author

ABSTRACT

Global food security remains a worldwide concern due to changing climate, increasing population, and reduced agriculture acreages. Greenhouse cultivation increases productivity by extending growing seasons, reducing pest infestations and providing protection against short term drastic weather fluctuations like frost, heat, rain, and wind. In the present study, we examined and compared the metabolic responses of nine soybean varieties grown under field and greenhouse conditions. Extracts were assayed by GC-FID, GC-MS, and LC-MS analyses for the identification of 10 primary (amino acids, organic acids, and sugars) and 10 secondary (isoflavones, fatty acid methyl esters) metabolites. Sugar molecules (glucose, sucrose, and pinitol) and isoflavone aglycons were increased but the isoflavones glucoside content decreased in the greenhouse cultivated soybeans. The amino acids and organic acids varied between the varieties. The results show that clustering (PCA and PLS-DA) patterns of soybean metabolites were significantly influenced by the genetic variation and growing conditions.

Keywords: *Cultivars; greenhouse; metabolites; multivariate analyses; soybean*

1. Introduction

Soybeans are one of the major foods consumed in many Asian countries. Soybeans are an excellent source of protein, essential fatty acids, carbohydrates, numerous vitamins, minerals, isoflavones, and fiber. They provide the world's largest source of animal protein feed and the second largest source of vegetable oil. The United States is the leading soybean producer and exporter. Soybean seeds contain high levels of isoflavones exhibiting a wide array of bioactivities such as minimization of postmenopausal symptoms (Wang & Murphy 1994; Messina 1999; Nakatsu et al., 2014; Pusparini, Yenny & Hidayat 2015).

Soybean based foods such as soy-milk, tofu, soy sauce, soy sprouts, and oil are commonly consumed worldwide. Awareness of the health benefits of soybeans has resulted in an increased demand for soy based food products. Besides their use as food and feed, soybeans have also been recently used in the development of commercial products such as plastics, lubricants, and adhesives (United Soybean Board, 2015). Due to the increased demand for soybeans and their products, there is a distinct need to increase production and improve soybean quality.

Bioactive isoflavones exist in free and conjugated (glucoside, acetyl and malonyl esters) forms in soybeans. Isoflavone glucosides have poor absorbability in the human intestine compared to their corresponding aglycons (Izumi et al., 2000). In fermented soy foods, significant conversion of glucosides to aglycons occurs, hence their absorbability is improved (Kim et al., 2011; Park et al., 2010; Baek et al., 2010; Fan et al., 2009; Kang et al., 2011; Lee et al., 2012; Wu et al., 2004). This has resulted in the development of several soybean fermented products like miso, natto, pickled tofu, soy sauce, and doenjang (Jeon et al., 2012; Shin et al., 2013; Lee et al., 2013).

Due to the high economic importance of soybean, a wide array of new varieties are continuously being introduced into the global market (Ha et al., 2013). In addition, the influence of environmental conditions on nutrient profiles are being investigated by researchers across the globe. Greenhouses and other controlled environments have been investigated to improve crop yield, reduce pest damages, extended growing seasons, and provide protections against short term drastic weather fluctuations like frost, heat, rain, and wind. In addition, greenhouses have also been used to study the effect of climate change. Crops like tomatoes (Choi et al., 2014), strawberries (Gündüz & Ozdemir, 2014), and peppers (Keyhaninejad et al., 2012) were reported to have improved yield and nutrient content under greenhouse conditions. Whereas in soybeans, the field and greenhouse conditions were studied only for their drought tolerance, disease, and toxicology studies (de Paiva Rolla et al., 2014; Dann et al., 1998; Pflieger et al., 2011). However, there is limited research available on understanding changes in the phytochemical content in soybeans grown under greenhouse as compared to the ambient field conditions. Furthermore, only limited varieties have been used in previous studies.

Plant metabolic profiles are directly influenced by genetics, physiological conditions, and the analytical methodologies used. The objective of the current research was to carry out a systematic comparison of the primary and secondary metabolic profiles of nine soybean varieties grown in the field and greenhouse conditions. Furthermore, we investigated application of multivariate analysis to differentiate soybean varieties grown under two conditions (field vs greenhouse).

2. Materials and methods

2.1. Chemicals and reagents

HPLC-grade chemicals (methanol, and acetonitrile) were purchased from Burdick and Jackson (Muskegon, MI, USA). Derivatizing agents methoxyamine hydrochloride, pyridine, and N,O-bis (trimethylsilyl) trifluoroacetamide (BSTFA) containing 1% trimethylchlorosilane (TMCS) were purchased from Sigma Aldrich (St. Louis, MO, USA). Individual standards for the primary and secondary metabolites were purchased from Sigma Aldrich (St. Louis, MO, USA) and Nu-Chek Prep, inc. (Elysian, MN, USA).

2.2. Soybean Samples Grown in Field and Greenhouse

Nine different soybean varieties (PI#549032, #549046, #157434, #548488, #88788, #555397, #636694, #632431, #639740) from United States Department of Agriculture (USDA) germplasm collection were selected for this study. Detailed information for the nine varieties (all soybean genotypes including wild type) is presented in Table 1. The varieties were cultivated under two distinct conditions, field (FI) and greenhouse (GH), during the period of June and July, 2014 at USDA facility in Beltsville, Maryland. For field experiments, the soybean seeds were planted using a randomized complete block design with three replications (hillplots were spaced at 91 cm between hills in rows spaced from 61 to 76 cm with five seeds per hill). Seeds were harvested at maturity from each hillplot. The average temperature during the growing season in the field was 71 °F with natural photoperiod. For greenhouse experiments, the photoperiod (grown in 16h light/8 h dark for the first 2 months, then 12 h light/12 h dark to initiate flowering) and temperature (72 °F) conditions were maintained as similar as possible to those in the field by computer control. Five seeds of each of the genotypes were planted in individual 6-quart pots that were filled with moistened Potting Mix (Canadian sphagnum moss, perlite, vermiculite, dolomitic limestone, starter fertilizer, trace elements and a wetting agent). Soil from the field was mixed with the potting mix to provide *Bradyrhizobium japonicum*

inoculation similar to that present in the field to permit normal nodulation and nitrogen fixation. After a month, Osmocote 14-14-14 solid fertilizer was applied. During second month, Bloom Booster fertilizer was applied to enhance the reproductive stages (flowering and pod formation). The pods were collected at full maturity (R9) and stored at 4°C until analyzed.

2.3. Sample Preparation for Primary Metabolite Analysis

Ground soybean samples (100 mg) collected from the field and greenhouse were individually extracted with 1 ml of methanol:water:chloroform (2.5:1:1, v:v:v) using ultrasonic assisted extraction for 10 min. The extracts were centrifuged at 10,000 rpm for a period of 10 min and the supernatants were collected. This extraction procedure was repeated two additional times and the collected supernatants were pooled and evaporated to dryness using a speedvac. The resulting pellet was derivatized as follows: samples were treated with 200 µl of methoxyamine hydrochloride in pyridine (20 mg/ml) followed by a 90-min incubation at 35°C with methyloximation. N,O-bis (trimethylsilyl) trifluoroacetamide (BSTFA, 100 µl) containing 1% trimethylchlorosilane (TMCS), and then incubated at 40 °C for 30 min to accelerate the rate of the silylation reaction (Maria John et al., 2013). Two replicate samples were collected for each cultivar under both growing conditions (greenhouse and field). Three extractions and analyses were carried out for each soybean sample, resulting in six analyses for each variety under a single growing condition.

2.4. Primary Metabolite Analysis by Gas Chromatography-Mass Spectrometry (GC-MS)

Analysis of samples was performed using an Agilent GC-MS system (7890A) with an autosampler (Agilent 7693) equipped with a HP-5MS capillary column (30 m length × 0.25 mm i.d. × 0.25 µm film thickness-Agilent J&W GC column). The injector temperature was 250 °C, and the injection volume was 1 µl. The oven temperature was programmed as follows: 80 °C for

2 min, then ramped to 300 °C at a rate of 10 °C/min, and held at 300 °C for 3 min. The transfer line temperature was set at 250 °C. The ionization potential was set at -70 V (electron energy) with a source temperature of 200°C. The detector voltage was 1450 V and the mass range was set at 50–600 m/z with an acquisition rate of 10 spectra per second. Two replicate samples were collected for each variety under both growing conditions (greenhouse and field). Three extractions and analyses were carried out for each soybean sample, resulting in six analyses for each variety under a single growing condition.

2.5. Fatty Acid Methyl Ester Extraction and Analysis

Ground soybean (100 mg) samples were extracted twice with hexane (5 ml). Each time the mixture was placed in a ultrasonic bath (power 600 watts) for a period of 15 min. The pooled extracts were centrifuged at 5000 rpm for 10 min and the supernatant was collected and evaporated to dryness under slow stream of nitrogen gas. The residue that primarily composed of soybean oil was re-suspended in 2 ml hexane and 1 ml was separated and evaporated to dryness for the preparation of fatty acid methyl ester derivatives by transesterification of extracted soybean oil. The pellet was derivatized by adding 5 ml of acidified methanol; a slow dropwise addition of 10 ml of acetyl chloride to 90 ml of cold methanol in an ice bath. The mixture was stirred at ambient temperature overnight. The next day, 3 ml water was added and the fatty acid methyl esters were extracted with 2 ml of hexane. The hexane layer was separated and analyzed with GC-FID (Luthria, Chen & Sprecher 1997).

Fatty acid methyl esters (FAMES) were analyzed using an Agilent GC-FID system (6890N) with an autosampler (Agilent 7683) equipped with a DB-WAX capillary column (30 m length × 0.25 mm i.d. × 0.25 µm film thickness-Agilent J&W GC column) and flame ionization detection (FID). The injector temperature was 250 °C, and the injection volume was 1 µl. The

oven temperature program commenced at 170 °C with an equilibration time of 3 min and finally reached 215 °C with the ramp of 1°C. The detector was maintained at 270 °C with a hydrogen flow at 45 ml/min, air flow at 450 ml/min, and make up nitrogen flow at 45 ml/min. Identification of FAMES were made by comparison of retention times with those of authentic standards. Two replicate samples were collected for each variety under both growing conditions (greenhouse and field). Three extractions and analyses were carried out for each soybean sample, resulting in six analyses for each variety under a single growing condition.

2.6. Sample Preparation and Isoflavones Analyses

Ground soybean samples (100 mg) were extracted with 1 ml of methanol:water (80:20, v/v) using ultrasonic assisted extraction for 10 min. The mixture was centrifuged at 10,000 rpm for 10 min. The supernatant was collected and the pellet was re-extracted with 1 ml of methanol:water (80:20, v/v) twice using the same procedure as above. The collected supernatants were pooled and evaporated to dryness using a speed vac. The dried extract was redissolved in 500 µl of methanol and filtered (PTFE syringe filter with a 0.45-µm pore size, (Agela Technologies Inc, Newark, DE, USA) prior to analysis. Two replicate samples were collected for each variety under both growing conditions (greenhouse and field). Three extractions and analyses were carried out for each soybean sample, resulting in six analyses for each variety under a single growing condition.

2.7. Liquid Chromatography-Mass Spectrometry (LC-MS) analysis

Secondary metabolites of the soybean samples were analyzed using LC-ESI-MS (Agilent) with Luna 5 µ C18 (2) 100A column (150 × 4.6 mm). The mobile phase consisted of water (A) and acetonitrile (B) with 0.1% formic acid (v/v). The initial gradient of the mobile phase was maintained at 5% acetonitrile for 1 min, gradually increased to 100% acetonitrile at 60

min, maintained at 100% acetonitrile for 1 min, and then decreased to 5% acetonitrile and maintained for 1 min. Five microliters of the sample was injected and the flow rate was maintained at 0.3 ml/min. ESI was performed in the negative (-) and positive (+) ion mode over a range of 100–1,000 m/z. The operating parameters were as follows: ion source temperature, 300 °C; cone gas flow, 50 l/h; desolvation gas flow, 600 l/h; capillary voltage, 2.8 kV; and cone voltage, up to 35 V. Two replicate samples were collected for each variety under both growing conditions (greenhouse and field). Three extractions and analyses were carried out for each soybean sample, resulting in six analyses for each variety under a single growing condition.

2.8. Data Processing

GC-MS and LC-MS data files were converted to net CDF (*.cdf) formats by using either in build common data format (*.cdf), converter (LC-MS) or MassLynx software (GC-MS). After CDF conversion, metAlign was used to convert the files into readable CSV format with peak retention time, peak area, and corresponding mass (m/z) data matrix. The matrix was later exported to Microsoft Excel (Microsoft, Redmond, WA, USA) for further analysis. For the GC-FID peak areas of the five identified methyl esters were used. The GC-FID data sets were stored in CSV format and Microsoft Excel (Microsoft, Redmond, WA, USA) was used for analysis.

2.9. Multivariate Analysis (MVA)

Multivariate statistical analysis of primary metabolites (amino acids and sugars), fatty acid methyl esters (FAMES), and isoflavone metabolites was carried out using SIMCA software 13.0 (Umetrics, Umeå, Sweden). Principal component analysis (PCA), and Partial least-squares-discriminate analysis (PLS-DA) was carried out using auto-scaled and log-transformed data from the GC-MS and LC-MS to identify the metabolites that were different for the field and greenhouse growing conditions. Based on the variable importance in projection (VIP) values,

corresponding to loadings scores, and a threshold of 0.05 for Student's t-test of individual samples, key metabolites were selected and identified. The $\log_{10}$ value of the peak area was used to compare the levels of metabolites between samples. Annotation of peaks were based on standard retention times, m/z, and literature references.

3. Results and discussion

3.1. Phenotypic variations of soybean based on metabolic content

Two replicate samples were collected for each variety under both growing conditions (greenhouse and field). Three extractions and analyses were carried out for each soybean sample, resulting in six analyses for each variety under a single growing condition. The primary and secondary metabolites were analyzed by GC-FID, GC-MS and LC-MS methods. Typical GC-MS (sugars and amino acids), GC-FID (fatty acid methyl esters after transesterification of extracted soybean oil) and LC-MS (isoflavones) chromatograms are presented in Fig. 1. Details on the identification of different metabolites is presented in Table 2.

Fig. 2 shows the PLS-DA analysis to investigate the similarities and differences between nine soybean varieties grown under field (A-C) or greenhouse (D-F) grown conditions. In general, all the analyses showed four distinct clustering patterns. However, the clustering pattern varied between greenhouse and field grown samples. For aminos acids and sugars, the field grown samples showed four clusters, I (FI1 and FI2), II (FI5), III (FI3 and FI4) and IV (FI6, FI7, FI8 and FI9) whereas greenhouse condition showed four clusters, I (GH1, GH2), II (GH8 and GH9), III (GH3, GH5, GH6 and GH7) and IV (GH4). Similarly, variety difference in the two growing conditions were observed in the patterns of the fatty acid methyl esters for field I (FI1), II (FI8 and FI9), III (FI2 and FI3) and IV (FI4, FI5, FI6 and FI7) and greenhouse I (GH1), II

(GH2 and GH6), III (GH3, GH4, GH5, GH8 and GH9) and IV (GH7) and isoflavones for field I (FI2), II (FI1 and FI5), III (FI6, FI7, FI8 and FI9) and IV (FI3 and FI4) and greenhouse I (GH1, GH2), II (GH3 and GH4), III (GH6, GH7, GH8 and GH9) and IV (GH5).

The two wild soybean varieties were distinctly separated based on amino acid and sugar analyses by GC-MS for both growing conditions (greenhouse and field). A similar separation was also obtained for the two wild soybeans based on isoflavone analysis by LC-MS. Similarly, all four US soybean varieties clustered together for all metabolites analyzed in the present study. Thus both primary and secondary metabolites are significantly influenced by variety types and growing conditions.

3.2. Influence of cultivation condition on metabolic content of soybean phenotypes

Non-supervised analysis of the entire GC-MS (Fig. 3A) and LC-MS (Fig. 3C) data sets using PCA models revealed that greenhouse and field grown samples showed variations. The results presented in Fig. 3A show that for most varieties there were apparent differences between the greenhouse and the field grown samples when the entire data set (identified and unidentified peaks) was included. About 35% of the variation was captured with two components (PC1 21.2% and PC2 13.0%). Similar variances between the field and the greenhouse samples was obtained with fatty acid methyl esters analyzed by GC-FID (Fig. 3B) and LC-MS analysis (Fig. 3C) of isoflavones. In addition, a distinct clustering pattern was observed for wild soybeans (PI#549032 and PI#549046) (Table 1) with GC-MS (amino acids, sugars, and organic acids) and LC-MS (isoflavones) analysis. However, with FAMES analysis by GC-FID significant overlap was observed between greenhouse and field grown soybean samples.

3.3. Individual metabolic variation of soybean phenotypes grown under two different conditions

Metabolites contributing to the distribution of the samples were identified by comparing the VIP values greater than 0.8 and *p*-value statistics. Several variables which discriminates the varieties based on their growing conditions were observed. Only annotated variables based on the availability of standards, their retention time and spectral data were further analyzed. A total of 10 primary metabolites (L-valine, L-threonine, butanedioic acid, L-aspartic acid, L-proline, L-phenylalanine, D-pinitol, D-glucose, myo-inositol, D-sucrose), 5 FAMES (methyl hexadecenoate, methyl octadecenoate, methyl octadecenoate (*cis*-9), methyl octadecadienoate (all *cis*-9,12), methyl octadecatrienoate (all *cis*-6,9,12)), and 5 isoflavones (genistin, daidzin, daidzein, glycitein, and genistein) are compared in the field and greenhouse grown samples (Fig. 4). Sucrose, pinitol, and glucose content were increased in all the varieties grown under greenhouse conditions compared to those grown infield. Similarly, methyl hexadecenoate, methyl octadecenoate, methyl octadecenoate (*cis*-9) and methyl octadecadienoate (all *cis*-9,12) contents were increased under greenhouse cultivated soybean as compared to the corresponding field grown soybean. The five individual amino acid metabolites in different varieties did not show consistent trend between field and greenhouse grown samples. Similar inconsistencies were observed for the isoflavones.

3.4. Field vs greenhouse metabolic content

Total amino acids, sugars, fatty acid methyl esters, and isoflavones were calculated by summing the chromatographic peak areas (log value) of individually identified primary and secondary metabolites. Fig. 5 shows the total amino acids (Fig. 5A), total sugars (Fig. 5B), total fatty acid methyl esters (Fig. 5C), total isoflavones (Fig. 5D), total isoflavone glucosides (Fig. 5E) and total isoflavone aglycons (Fig. 5F) in field and greenhouse grown nine soybean varieties. Total amino acids detected by GC-MS analysis showed no consistent trend for all varieties

grown under field and green house conditions (Fig. 5A). However, significant difference among various varieties were observed between field and greenhouse grown samples. Varieties FI1, FI5, FI8, and FI9 showed higher yields (~4 to 14%) of total amino acids in field grown samples as compared to greenhouse grown samples. Only very low changes (~0.5 to 1%) in the total amino acids were observed between field and greenhouse grown samples for two varieties FI2 and GH3. On the other hand a marginal decrease (~5 to 8%) in total amino acid levels were observed in field samples of two varieties (FI6 and FI7). Sugar metabolites identified and analyzed in the present study showed consistent trend for all varieties. The sugar content for all field grown samples were lower than the green house grown samples. The percent decreased ranged from (~3 to 9%).

The total peak areas of the five identified fatty acid methyl esters grown under greenhouse conditions were higher as compared to the field grown samples, except for varieties GH8 and GH9. The seven varieties namely GH1, GH2, GH3, GH4, GH5, GH6, and GH7 showed ~3 to 5% increase in fatty acid methyl esters content under green house conditions. In case of varieties, GH8 and GH9, ~1 to 3% decrease in their total fatty acid methyl esters content was observed as compared to field grown samples (FI8 and FI9).

The peak areas of the five identified isoflavones also showed similar trend as total amino acids. Significant variations among different varieties were observed. Seven varieties GH2, GH3, GH4, GH5, GH6, GH8, and GH9 showed higher yields (~0.9 to 6%) of total isoflavones in greenhouse grown samples as compared to field samples. This increase was mostly contributed by their total aglycons content (~2 to 8%) than their glucoside content. Due to low glucoside content, only marginal increase (0.6 to 0.8%) in the total isoflavones content were observed for field grown samples for two varieties FI1 and FI7.

The results presented here provide an assessment of soybean response to controlled and field grown conditions at the level of the metabolome. While considerable work has been carried out on other crop species, like tomato (Ferne, Tadmor & Zamir 2006; Fernie & Klee 2011), maize (Harrigan et al., 2007; Lisec, et al., 2011) strawberry (Gündüz & Ozdemir 2014) and pepper (Keyhaninejad, Richins & O'Connell, 2012), the metabolic changes of soybean under controlled conditions have not been reported before. For controlled environments (greenhouses), studies have focused on understanding the genetic factors underlying abscisic acid levels in case of drought (Setter et al., 2011), osmolyte accumulation (Yancey 2005), and the relationships between yield, ash content, and isotope distributions (Cabrera-Bosquet et al., 2011). General increases in metabolite levels under drought stress, including changes in amino acids, sugars, sugar alcohols, and intermediates of the TCA cycle, were reported in maize cultivated under greenhouse conditions (Witt et al., 2012).

The influence of treatments, namely, drought, diseases, and herbicides on soybean grown in greenhouse have been reported (de Paiva Rolla et al., 2014; Jaidee, Polthanee, Saenjan, 2012; Lemes et al., 2011; Miles et al., 2008; Twizeyimana et al., 2007). Previous results show that greenhouse conditions have been used for evaluating the response of a treatment on yield of soybean production. Most of the earlier studies used 1-3 varieties to investigate the response to a particular treatment. But the metabolic responses from the present study with nine soybean varieties showed that the amino acids had varied responses within the greenhouse. It is clear that amino acid and organic acid levels were dependent on their variety. However, for sugars, most of the fatty acid methyl esters, and isoflavones, significant increases were observed for greenhouse grown samples irrespective of the varieties.

Prior literature on comparison of plants grown under greenhouse and field conditions primarily focused on biotic or abiotic stresses. In the present study, we compared primary (amino acids, organic acids, and sugars) and secondary (isoflavones, fatty acid methyl esters) metabolites of nine soybean varieties grown in the field and greenhouse. The results showed significant differences in metabolic profiles between field and greenhouse grown samples. Significant difference in sugars and isoflavones profiles were observed in field and greenhouse grown samples. In addition, genetic variation in soybeans significantly influenced the clustering patterns as two wild type soybeans, and four soybean varieties produced in US showed two clusters based on metabolite profiles. Thus understanding of genetic, environmental, and processing conditions are essential for classification of food substances based on analytical data and multivariate analysis.

Author Information

Corresponding Author

Email:Dave.Luthria@ars.usda.gov; phone: +1 301 504 7247; fax: +1 301 504 8314

References

- Baek, J.G., Shim, S.M., Kwon, D.Y., Choi, H.K., Lee, C.H., & Kim, Y.S. (2010). Metabolite profiling of Cheonggukjang, a fermented soybean paste, inoculated with various *Bacillus* strains during fermentation. *Bioscience, Biotechnology, and Biochemistry*, 74, 1860–1868.
- Cabrera-Bosquet, L., Sanchez, C., Rosales, A., Palacios-Rojas, N., & Araus, J.L. (2011). Near-infrared reflective spectroscopy (NIRS) assessment of $\delta^{18}\text{O}$ and nitrogen and ash contents for improved yield potential and drought adaptation in maize. *Journal of Agricultural and Food Chemistry*, 59, 467–474.
- Choi, S.H., Kim, D.S., Kozukue, N., Kim, H.J., Nishitani, Y., Mizuno, M., Levin, C.E., & Friedman, M. (2014). Protein, free amino acid, phenolic, β -carotene, and lycopene content, and antioxidative and cancer cell inhibitory effects of 12 greenhouse-grown commercial cherry tomato varieties. *Journal of Food Composition and Analysis*, 34, 115–127.
- Dann, E., Diers, B., Byrum, J., & Hammerschmidt, R. (1998). Effect of treating soybean with 2,6-dichloroisonicotinic acid (INA) and benzothiadiazole (BTH) on seed yields and the level of disease caused by *Sclerotinia sclerotiorum* in field and greenhouse studies. *European Journal of Plant Pathology*, 104, 271–278.
- de Paiva Rolla, A.A., de Fátima Corrêa Carvalho, J., Fuganti-Pagliarini, R., Engels, C., do Rio, A., Marin, S.R., de Oliveira, M.C., Beneventi, M.A., Marcelino-Guimarães, F.C., Farias, J.R., Neumaier, N., Nakashima, K., Yamaguchi-Shinozaki, K., & Nepomuceno, A.L. (2014). Phenotyping soybean plants transformed with rd29A:AtDREB1A for drought tolerance in the greenhouse and field. *Transgenic Research*, 23, 75–87.

- Fan, J., Zhang, Y., Chang, X., Saito, M., & Li, Z. (2009). Changes in the radical scavenging activity of bacterial-type Douchi, a traditional fermented soybean product, during the primary fermentation process. *Bioscience, Biotechnology, and Biochemistry*, 73, 2749–2753.
- Fernie, A.R., & Klee, H.J. (2011). The use of natural genetic diversity in the understanding of metabolic organization and regulation. *Frontiers in Plant Science*, 2, 59, 1-10.
- Fernie, A.R., Tadmor, Y., & Zamir, D. (2006). Natural genetic variation for improving crop quality. *Current Opinion in Plant Biology*, 9, 196–202.
- Gündüz, K., & Ozdemir, E. (2014). The effects of genotype and growing conditions on antioxidant capacity, phenolic compounds, organic acid and individual sugars of strawberry. *Food Chemistry*, 155, 298-303.
- Ha, C.V., Le, D.T., Nishiyama, R., Watanabe, Y., Tran, U.T., Dong, N.V., & Tran, L.-S.P. (2013). Characterization of the newly developed soybean cultivar DT2008 in relation to the model variety W82 reveals a new genetic resource for comparative and functional genomics for improved drought tolerance. *BioMed Research International*, 759657.
- Harrigan, G.G., Stork, L.G., Riordan, S.G., Ridley, W.P., Macisaac, S., Halls, S.C., Orth, R., Rau, D., Smith, R.G., Wen, L., Brown, W.E., Riley, R., Sun, D., Modiano, S., Pester, T., Lund, A., & Nelson, D. (2007). Metabolite analysis of grain from maize hybrids grown under drought and watered conditions during the 2002 field season. *Journal of Agricultural and Food Chemistry*, 55, 6169–6176.
- Izumi, T., Piskula, M.K., Osawa, S., Obata, A., Tobe, K., Saito, M., Kataoka, S., Kubota, Y., & Kikuchi, M. (2000). Soy isoflavone aglycons are absorbed faster and in higher amounts than their glucosides in humans. *Journal of Nutrition*, 130, 1695-1699.

378 Jaidee, R., Polthanee, A., & Saenjan, P. (2012). Growth and yield of soybean cultivars as
 379 affected by ground water levels and phosphorus rates grown under greenhouse and field
 380 conditions. *Australian Journal of Crop Science*, 6, 81-92.

381 Jeon, H.Y., Seo, D.B., Shin, H.J., & Lee, S.J. (2012). Effect of *Aspergillus oryzae* challenged
 382 germination on soybean isoflavone content and antioxidant activity. *Journal of*
 383 *Agricultural and Food Chemistry*, 60, 2807–2814.

384 Kang, H.J., Yang, H.J., Kim, M.J., Han, E.S., Kim, H.J., & Kwon, D.Y. (2011). Metabolomic
 385 analysis of Meju during fermentation by ultra performance liquid chromatography
 386 quadrupole time of flight mass spectrometry (UPLC-Q-TOF MS). *Food Chemistry*, 127,
 387 1056-1064.

388 Keyhaninejad, N., Richins, R.D., & O’Connell, M.A. (2012). Carotenoid content in field-grown
 389 versus greenhouse-grown peppers: different responses in leaf and fruit. *HortScience*, 47,
 390 852–855.

391 Kim, J., Choi, J.N., Kang, D., Son, G.H., Kim, Y.S., Choi, H.K., Kwon, D.Y., & Lee, C.H.
 392 (2011). Correlation between antioxidative activities and metabolite changes during
 393 Cheonggukjang fermentation. *Bioscience, Biotechnology, and Biochemistry*, 75, 732-739.

394 Lee, S.H., Lee, J., Jung, M.H., & Lee, Y.M. (2013). Glyceollins, a novel class of soy
 395 phytoalexins, inhibit angiogenesis by blocking the VEGF and bFGF signaling pathways.
 396 *Molecular Nutrition & Food Research*, 57, 225–234.

397 Lee, S.Y., Kim, H.Y., Lee, S., Lee, J.M., Muthaiya, M.J., Kim, B.S., Oh, J.Y., Song, C.K., Jeon,
 398 E.J., Ryu, H.S., & Lee, C.H. (2012). Mass spectrometry-based metabolite profiling and
 399 bacterial diversity characterization of Korean traditional Meju during fermentation.
 400 *Journal of Microbiology and Biotechnology*, 22, 1523-1531.

401 Lemes, E.M., MacKowiak, C.L., Blount, A., Marois, J.J., Wright, D.L., Coelho, L., & Datnoff,
 402 L.E. (2011). Effects of silicon applications on soybean rust development under
 403 greenhouse and field conditions. *Plant Disease*, 95, 317-32.

404 Lisec, J., Römisch-Margl, L., Nikoloski, Z., Piepho, H.P., Giavalisco, P., Selbig, J., Gierl, A., &
 405 Willmitzer, L. (2011). Corn hybrids display lower metabolite variability and complex
 406 metabolite inheritance patterns. *The Plant Journal*, 68, 326-36.

407 Luthria, D.L., Chen, Q., & Sprecher, H. (1997). Metabolites produced during the peroxisomal β -
 408 oxidation of linoleate and arachidonate move to microsomes for conversion back to
 409 linoleate. *Biochemical and Biophysical Research Communications*, 233, 438-441.

410 Maria John, K.M., Jung, E.S., Lee, S., Kim, J.S., & Lee, C.H. (2013). Primary and secondary
 411 metabolites variation of soybean contaminated with *Aspergillus sojae*. *Food Research*
 412 *International*, 54, 487-494.

413 Messina, M.J. (1999). Legumes and soybeans: Overview of their nutritional profiles and health
 414 effects. *The American Journal of Clinical Nutrition*, 70, 439S-450S.

415 Miles, M.R., Morel, W., Ray, J.D., Smith, J.R., Frederick, R.D., & Hartman, G.L. (2008). Adult
 416 plant evaluation of soybean accessions for resistance to *Phakopsora pachyrhizi* in the
 417 field and greenhouse in Paraguay. *Plant Disease*, 92, 96-105.

418 Nakatsu, C.H., Armstrong, A., Clavijo, A.P., Martin, B.R., Barnes, S., & Weaver, C.M. (2014).
 419 Fecal bacterial community changes associated with isoflavone metabolites in
 420 postmenopausal women after soy bar consumption. *PLoS ONE*, 9, 0108924.

421 Park, M.K., Cho, I.H., Lee, S., Choi, H.K., Kwon, D.Y., & Kim, Y.S. (2010). Metabolite
 422 profiling of Cheonggukjang, a fermented soybean paste, during fermentation by gas

chromatography-mass spectrometry and principal component analysis. *Food Chemistry*, 122, 1313-1319.

Pfleeger, T., Olszyk, D., Lee, E.H., & Plocher, M. (2011). Comparing effects of low levels of herbicides on greenhouse- and field-grown potatoes (*Solanum tuberosum* L.), soybeans (*Glycine max* L.), and peas (*Pisum sativum* L.). *Environmental Toxicology and Chemistry*, 30, 455-68.

Pusparini, Yenny, & Hidayat, A. (2015). Effect of soy isoflavone supplementation on endothelial dysfunction and oxidative stress in equol-producing postmenopausal women. *Endocrine, Metabolic and Immune Disorders-Drug Targets*, 15, 71-79.

Setter, T.M., Yan, J., Warburton, M., Ribaut, J.M., Xu, Y., Sawkins, M., Buckler, E.S., Zhang, Z., & Gore, M.A. (2011). Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *Journal of Experimental Botany*, 62, 701–716.

Shin, S.H., & Lee, Y.M. (2013). Glyceollins, a novel class of soybean phytoalexins, inhibit SCF-induced melanogenesis through attenuation of SCF/c-kit downstream signaling pathways. *Experimental & Molecular Medicine*, 45, doi:10.1038/emm.2013.20.

Twizeyimana, M., Ojiambo, P.S., Ikotun, T., Paul, C., Hartman, G.L., & Bandyopadhyay, R. (2007). Comparison of field, greenhouse, and detached-leaf evaluations of soybean germplasm for resistance to *Phakopsora pachyrhizi*. *Plant Disease*, 91, 1161-1169.

United soybean board. Soy-based materials and green building constructions technologies. http://soynewuses.org/wp-content/uploads/44420_Final_GreenBuildingWhitePaper.pdf.

As cited on August/27/2015.

445 Wang, H.J., & Murphy, P.A. (1994). Isoflavone content in commercial soybean foods. *Journal of*
446 *Agricultural and Food Chemistry*, 42, 1666–1673.

447 Witt, S., Galicia, L., Lisec, J., Cairns, J., Tiessen, A., Araus, J.L., Palacios-Rojas, N., & Fernie,
448 A.R. (2012). Metabolic and phenotypic responses of greenhouse-grown maize hybrids to
449 experimentally controlled drought stress. *Molecular Plant*, 5, 401-17.

450 Wu, Q., Wang, M., Sciarappa, W.J., & Simon, J.E. (2004). LC/UV/ESI-MS analysis of
451 isoflavones in edamame and tofu soybeans. *Journal of Agricultural and Food Chemistry*,
452 52, 2763-2769.

453 Yancey, P.H. (2005). Organic osmolytes as compatible, metabolic and counteracting
454 cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology*,
455 208, 2819–2830.

456

Figure captions

Fig. 1. GC-MS (amino acids and sugars), GC-FID (fatty acid methyl esters) and LC-MS (isoflavones) metabolic profiles of soybeans varieties grown under field and greenhouse conditions.

Fig. 2. PLS-DA analysis of soybean grown under field (A-C) and greenhouse (D-F) conditions analyzed by GC-MS (amino acids and sugars), GC-FID (fatty acid methyl esters) and LC-MS (isoflavones).

Fig. 3. PCA analysis of soybean grown under field and green house conditions analyzed by (A) GC-MS (amino acids and sugars), (B) GC-FID (targeted analysis of fatty acid methyl esters) and (C) LC-MS (isoflavones).

Fig. 4. Individual metabolite variations in nine soybean varieties grown under field (green) and greenhouse (red) conditions (see table 1 for sample descriptions).

Fig. 5. Total amino acids (A), total sugars (B), total fatty acids (C), total isoflavones (D), total isoflavone glucosides (E) and total isoflavone aglycons (F) in nine soybean varieties grown under field (white) and greenhouse (black) conditions.

478 Table 1. Soybean cultivars analysed for Amino acids, Fatty acids, and isoflavones

S.No	USDA Accession #	Genotypes	Origin / Information	Maturity Group	Field grown	Greenhouse grown
1	PI# 549032	Wild soybean (<i>G.soja</i>)	China	III	FI 1	GH 1
2	PI# 549046	Wild soybean (<i>G.soja</i>)	China	IV	FI 2	GH 2
3	PI# 157434	Soybean Landraces	S. Korea	IV	FI 3	GH 3
4	PI# 548488	Soybean Landraces	Heilongjiang, China	V	FI 4	GH 4
5	PI# 88788	Soybean Landraces	China	III	FI 5	GH 5
6	PI# 555397	Soybean Bred for Seed Traits	Beltsville, MD	IVS	FI 6	GH 6
7	PI# 636694	Soybean Bred for Seed Traits	Beltsville, MD	IVS	FI 7	GH 7
8	PI# 632431	Soybean Bred for Seed Traits	Ohio	IV	FI 8	GH 8
9	PI# 639740	Soybean Bred for Seed Traits	Illinois	IV	FI 9	GH 9

479 Table 2. Metabolites analyzed by three chromatographic methods, GC-MS, GC-FID, and LC-MS.

S. No	RT (min)	Compound Name	Targetted ion (m/z)	LOQ (µg/ml)	Calibration	R ²	Reference
1	6.83	L-Valine	144	0.19	Y=0.3178x + 21.27	0.9666	Std
2	8.09	L-Threonine	119	0.39	Y=0.3166x + 20.041	0.9684	Std
3	8.13	Butanedioic acid	165	-	-	-	MS, NIST
4	9.62	L-Aspartic acid	116	-	-	-	MS, NIST
5	10.93	L-Proline	142	3.12	Y=5.3682x + 51.305	0.9835	Std
6	11.22	L-Phenylalanine	120	0.39	Y=0.2916x + 19.589	0.9684	Std
7	14.65	D-Pinitol	273	-	-	-	MS, NIST
8	15.30	D-Glucose	205	1.56	Y=0.7679x + 20.673	0.9615	Std
9	17.20	Myo-Inositol	217	1.56	Y=0.1241x + 8.6784	0.9671	Std
10	21.94	D-Sucrose	361	1.56	Y=0.2587x + 5.1676	0.9611	Std
11	12.17	Palmitic (16.0)	271	1.13	Y=1162.9x + 0.7086	0.9994	Std
12	19.85	Stearic (18.0)	299	1.50	Y=1188.2x + 0.4023	0.9995	Std
13	20.74	Oleic (18.1)	297	0.75	Y=3555.6x + 0.6379	0.9996	Std
14	22.91	Linolic (18.2)	293	1.50	Y=1001.1x + 0.2331	0.9993	Std
15	26.17	Linolenic (18.3)	263	3.00	Y=1095.6x + 0.2025	0.9992	Std
S. No	RT (min)	Compound Name	[M-H] ⁻ (m/z)	LOQ (µg/ml)	Calibration	R ²	Reference
16	9.61	Daidzin	461	0.33	Y=1152.2x + 0.4428	0.9993	Std
17	10.32	Glycitin	491	0.25	Y=811.7x + 0.1607	0.9997	Std
18	15.62	Genistin	431	1.07	Y=715.5x + 0.8687	0.9996	Std
19	27.20	Daidzein	283	0.42	Y=2283.7x + 1.2891	0.9995	Std
20	38.45	Genestein	269	1.07	Y=489.55x + 0.3627	0.9996	Std

Fig. 1. GC-MS (amino acids and sugars), GC-FID (fatty acid methyl esters) and LC-MS (isoflavones) metabolic profiles of soybeans varieties grown under field and greenhouse conditions.

Fig. 2. PLS-DA analysis of soybean grown under field (A-C) and greenhouse (D-F) conditions analyzed by GC-MS (amino acids and sugars), GC-FID (fatty acid methyl esters) and LC-MS (isoflavones).

Fig. 3. PCA analysis of soybean grown under field and green house conditions analyzed by (A) GC-MS (amino acids and sugars), (B) GC-FID (targeted analysis of fatty acid methyl esters) and (C) LC-MS (isoflavones).

Fig. 4. Individual metabolite variations in nine soybean varieties grown under field (green) and greenhouse (red) conditions (see table 1 for sample descriptions).

Fig. 5. Total amino acids (A), total sugars (B), total fatty acids (C), total isoflavones (D), total isoflavone glucosides (E) and total isoflavone aglycons (F) in nine soybean varieties grown under field (white) and greenhouse (black) conditions.