frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 July 2018
doi: 10.3389/fpls.2018.01025

Variation in Seed Allergen Content From Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain[1]*, Severin E. Stevenson[2], Cavell Brownie[3], Corinne Herouet-Guicheney[4], Rod A. Herman[5], Gregory S. Ladics[6], Laura Privalle[7], Jason M. Ward[8], Nancy Doerrer[9] and Jay J. Thelen[2]

[1] Syngenta Crop Protection, LLC, Research Triangle Park, NC, United States, [2] Department of Biochemistry, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, [3] Department of Statistics, North Carolina State University, Raleigh, NC, United States, [4] Regulatory Science, Bayer SAS, Sophia Antipolis, France, [5] Dow AgroSciences, Zionsville, Indianapolis, IN, United States, [6] Pioneer Hi-Bred, DuPont Agricultural Biotechnology, Wilmington, DE, United States, [7] BASF Plant Science, Research Triangle Park, NC, United States, [8] Regulatory Affairs, Royal Canin, U.S.A., St. Charles, MO, United States, [9] Protein Allergenicity Technical Committee, Health and Environmental Sciences Institute, Washington, DC, United States

OPEN ACCESS

Edited by:
Setsuko Komatsu,
Fukui University of Technology, Japan

Reviewed by:
Carla Pinheiro,
Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal
Abu Hena Mostafa Kamal,
University of Texas at Arlington, United States

***Correspondence:**
Scott McClain
scottmcclain24@gmail.com

Specialty section:
This article was submitted to Plant Proteomics, a section of the journal Frontiers in Plant Science

Received: 28 November 2017
Accepted: 25 June 2018
Published: 23 July 2018

Citation:
McClain S, Stevenson SE, Brownie C, Herouet-Guicheney C, Herman RA, Ladics GS, Privalle L, Ward JM, Doerrer N and Thelen JJ (2018) Variation in Seed Allergen Content From Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana. Front. Plant Sci. 9:1025. doi: 10.3389/fpls.2018.01025

Soybean (*Glycine max*) is an important food stock, and also considered an allergenic food with at least eight well characterized allergens. However, it is a less prevalent allergen source than many other foods and is rarely life-threatening. Soybean is incorporated into commonly consumed foods, and therefore, the allergens pose a potential concern for individuals already sensitized. The protein profile of soybean can be affected by several factors including genetic and environmental. To investigate how soybean allergen content may be affected by genetics and/or environment, nine soy allergens were quantified from three commercial soybean varieties grown at nine locations in three states within a single climate zone in North America; Iowa, Illinois, and Indiana, United States. Quantitation was achieved using liquid chromatography-selected reaction monitoring (LC-SRM) tandem mass spectrometry with AQUA peptide standards specific to the nine target allergens. Quantitation of allergen concentration indicated that both genetics and location affected specific allergen content. Seven of the nine allergens were significantly influenced by genetics, with the exceptions of glycinin G4 and KTI 3. The allergens P34, Gly m Bd 28k, glycinin G3, and KTI 1 showed statistically significant impact from location as well, but at a lower threshold of significance compared with genetics (cultivar/variety). This dataset contributes to our understanding of the natural variation of endogenous allergens, as it represents a sampling of soybeans grown in a controlled, distributed plot design under agronomic conditions common for commercial soybean food and feed production. The aim was to build upon our recent understanding of how allergens are expressed as part of the overall soybean proteome.

Keywords: allergen, soybean, quantitative proteomics, mass spectrometry, multiple reaction monitoring, AQUA

INTRODUCTION

Soybean is used in the manufacturing of a vast variety of foods, but is considered a major allergenic food in the United States, and much of the developed world. Over 30 potential soybean allergen sequences have been identified, sixteen of which have been confirmed with some data to support sensitization and elicitation (Stevenson et al., 2012). A comprehensive listing of peer-reviewed allergens is available[1] with a recent review finding that eight soybean allergens have a clear status as both clinically relevant allergens with known sequences (Ladics et al., 2014). Other databases or listings of allergens are also available such as www.allergenonline.org and www.allergome.org. The primary soybean allergens are seed storage proteins which are of particular interest in certain foods because of their physical and nutritional properties. These properties include gelation, pH, solubility as well as emulsification; they are known to affect the way foods are processed (Utsumi and Kinsella, 1985; Cai and Chang, 1999; Mujoo et al., 2003). Although soybean is a widely recognized allergenic food crop, a specific and multiplexed method to quantitatively measure protein allergens has only recently been developed and utilized (Houston et al., 2011; Stevenson et al., 2012).

One expectation of any newly developed analytical capability that is tied to human health is the evaluation of important variables affecting the interpretation of quantitative expression data, particularly in the case of allergens where determining exposure thresholds can be critical. As has been noted by initial work with soybeans using tandem mass spectrometry approaches, absolute quantitative analysis of allergen expression in food is novel (Houston et al., 2011; Stevenson et al., 2012). Analytically robust studies are essential to provide useful information on the natural variability of allergens. In the case of soybean, understanding the genetic variation in soybean and the impact from environmental conditions on allergen content is important because: (1) there are many individual allergens and the relationship of expression between one or more of the allergens has not been fully characterized; (2) soybeans are planted across broad geographical locations and environmental influence is known to be a primary variable in allergome profiles (Stevenson et al., 2012); and (3) soybeans have a wide array of uses that rely on extensive breeding which creates a variety of genome profiles capable of potentially impacting the expressed proteome. However, commodity soybeans are blended from numerous field sources and individual varieties are not generally considered a point-source of food exposure. Therefore, slight differences in allergen content on a variety basis might be inconsequential in the context of allergen exposure from commodity soybeans and food products. Nevertheless, an understanding of the natural variability of seed allergen content is an expectation on a variety of fronts, including regulatory questions around genetically modified soybean varieties and in commodity soybean processing, as soybean is likely to remain a prominent food allergen source on a population and allergy prevalence basis. From a technical standpoint, describing the variance of target analytes (the soybean allergens in this case) is a key to using any assay for describing significant variance under controlled experimentation. This is particularly true in understanding a complex plant response to factors that may affect how much allergen load an allergen source carries.

In this study, nine allergens (or sub-units of multimers) that have each been substantially characterized with respect to their allergenicity were chosen for analysis (Utsumi, 1992; Stevenson et al., 2012) with two of the Kunitz trypsin inhibitors (KTI 1 and KTI 3) included. In some cases, these allergens exist *in vivo* as heteromers, necessitating the measurement of individual subunits during quantitation. For example, glycinin accounts for at least 30% of the soybean seed protein and consists of six subunits; each subunit has a specific basic chain linked to an acidic chain via disulfide bonding. The assembly of the basic and acidic subunits is achieved through a series of co- and post-translational modifications of the original polypeptides known as glycinin precursors (Ereken-Tumer et al., 1981; Utsumi, 1992; Petruccelli and Anon, 1995; Riblett et al., 2001). We were able to determine the absolute quantities (i.e., concentrations) of one chain from each of the four glycinin (11S globulin; Gly m 6) subunits (glycinin G1, Bx chain; glycinin G2, A2 chain; glycinin G3, A chain; glycinin G4, A5 chain). The second most abundant storage protein in soybean, β-conglycinin (vicilin; Gly m 5), a trimer, consists of α′, α, and β subunits (Mori et al., 1981) and the available protein sequence for the α-subunit of β-conglycinin was used in this study. Both Gly m Bd 30K (P34) and Gly m Bd 28K allergens are Asn-linked glycoproteins belonging to the thiol protease family (Ogawa et al., 1993; Tsuji et al., 1997). The P34 protein is known to elicit severe allergic reactions even at low abundance and is one of the most immune-dominant soybean allergens (Herman, 2003).

The goal of this work was to investigate and compare two variables that could quantitatively influence soybean seed allergen content; geographical growth location (environment; external influence) and genetics of the soybean varieties (genotype; internal influence). To accomplish this, samples were taken of soybeans from three varieties grown together at nine different field locations, but within a single climate zone representing the major soybean growing region in the United States. Although within this one growing zone, it was expected that environmental conditions vary from site to site since locations were spread apart through three states of United States. The goal of utilizing samples of soybean from this field trial across many sites was to examine the hypothesis that the combination of broadly assumed environmental variance and genetics of the soybean varieties would offer insight into how protein expression varies for important soybean allergens. It is recognized that although some characterization of environmental conditions are recorded for the underlying field trial from which samples were obtained, the work herein is a parallel examination of specific protein allergen expression that is apart from the field trial design itself since quantitative measures of environmental parameters were not included in the allergen statistical analysis and were considered beyond the scope of this allergen expression survey.

[1] http://comparedatabase.org

Nine allergens were quantitatively measured to provide a basis for interpreting the most influential variable on the expression of each allergen. These sequences were previously available as part of the FARRP allergenonline.org database for which there was pertinent sequence information and high quality clinical data supporting these as relevant allergens. A recent study that compared allergen levels of four non-GM soybean varieties grown in three different climate zones in North America showed that environment had a greater impact on allergen expression variability than genetics (Stevenson et al., 2012). In this paper, we address the same interacting variables by comparing three soy varieties grown at nine locations within a single North American climate zone. It is understood that each soybean variety has inherent, but undefined differences in background genetics that likely influence the final compositional makeup of soybean seed, including protein allergens. The environmental "load" or putative impact from the environment is characterized by a selection of varieties adapted to single climate zone having been grown at nine distinct locations that have differences in seasonal agronomic conditions.

MATERIALS AND METHODS

Plant Material

Three soybean varieties, A, B1, and B2, grown in 2009 in nine replicated field trials were used for analysis. The varieties are branded soybean seed produced in bulk as commodity soybean fit for growing in the United States region denoted in this study. They are listed and then coded with the letter designation for discussion herein as the following: A = Stine 2500-2 Brand; B1 = Stine 3300-2 Brand; B2 = Stine 3308-2 Brand. All three are non-transgenic and commercially available varieties and represent two different maturity groups; A is maturity group 2; B1 and B2 are group 3. Each field trial consisted of single plots of each variety and all plots at a site were subjected to the same agronomic practices during the 2009 United States growing season. Samples of 20 g each were taken from the harvest of one plot for each variety at each location. The field trials were located in three contiguous mid-western states in the United States: Iowa (IA), Illinois (IL), and Indiana (IN) (Supplementary Figure 1). IA was the location of seven of the nine plots, near the towns of Adel, Glidden, IA Falls, Marcus, Mediapolis, Perry and Winterset, and one plot each was located close to Fithian, IL, and Sharpsville, IN. Site-specific soil description and weather conditions are described in Supplementary Table 5.

Protein Extraction and Quantification

Total protein was extracted from three sub-samples (five seeds/sub-sample) of soybean seed taken from each variety at each site. Each sub-sample was ground to a fine powder using a Waring blender. A phenol extraction method compatible with soybean was employed similarly to Hajduch et al., 2007. Briefly, ground soybean material was suspended in 2.5 mL of Tris-buffered phenol (pH 8.0; Fisher Scientific, Inc., Pittsburgh, PA, United States) and 2.5 mL extraction buffer [0.9 M sucrose, 0.1 M Tris–HCl pH 8.0, 10 mM EDTA, 0.4% (v/v) 2-mercaptoethanol]. The mixture was homogenized with a T25 Basic S1 Disperser (IKA Works, Inc., Wilmington, NC, United States) on speed 4 with the S25N-10 tool for 2 min. The homogenized mixture was inverted at 4°C for 2 h and then centrifuged at 5000 × *g* and 4°C for 20 min. The phenol phase was removed, and a single back extraction was performed using an additional 2.5 mL of phenol following the same procedure. Phenol fractions were combined, and protein precipitated using five volumes of 95% (v/v) methanol with 0.1 M ammonium acetate then incubating at −20°C overnight. The precipitated protein was pelleted by centrifugation at 5000 × *g* and 4°C for 5 min and washed by re-suspending in 10 mL fresh methanol-ammonium acetate solution three times, 80% (v/v) acetone two times, and 70% (v/v) ethanol once, with 20 min incubations at −20°C before pelleting for each. Protein was stored in 80% (v/v) acetone at −20°C. The amount of total protein extracted was quantified as in Stevenson et al. (2012) by using the Bradford (1976) method.

In-Solution Digestion and Mass Spectrometry Analysis

Protein was aliquoted and digested prior to mass spectrometry as described in Lee et al. (2010) and Stevenson et al. (2012). Briefly, frozen previously quantitated protein was thawed at 4°C, vortexed until homogenous and 10 μg portioned into 1.5 mL polypropylene tubes. Sample volumes equalized by bringing to 10.0 μL using 8 M urea buffer. Disulfide bonds were reduced with dithiothreitol (DTT), to 10 mM for 1 h at room temperature. Reduced cysteines were carboxyamidomethylated with iodoacetamide to 50 mM, at room temperature in the dark for 1 h. Urea was diluted to 0.67 M and iodoacetamide was neutralized with equimolar DTT and incubating at room temperature for 15 min. Samples were chilled to 4°C and trypsin (sequencing grade – modified, Promega, Madison, WI, United States) added to 1.7 ng/mL. Digestion was allowed to proceed for 16 h at 37°C. After digestion, reactions were cooled to 4°C and each preparation was spiked with AQUA peptide standards (Sigma-Aldrich, The Woodlands, TX, United States) in 50% acetonitrile (Sigma-Aldrich, Saint Louis, MO, United States) and 1% (v/v) formic acid (ACS grade, 88% purity; Acros Organics, NJ, United States) within their linear quantitative ranges, determined as described by Houston et al. (2011). Typically, this is 100 fmol of peptide standard per 1.0 μg of soy protein, targeting each standard within a 10-fold concentration range of the native peptide signal. Target peptide and protein surrogate information are displayed in Supplementary Table 1. Digests containing peptide standards (for each allergen) were frozen at −80°C and lyophilized to dryness using a Centrivap. Lyophilized digests were dissolved to 0.2 μg/μL in 5% (v/v) acetonitrile, 0.1% (v/v) formic acid by vortexing and were centrifuged at 21,000 × *g* for 2 min at room temperature. The supernatant was transferred to a 96-well plate for LC-SRM analysis.

Mass spectrometric analysis was performed similarly to the previously described method in Stevenson et al. (2012), with 1 μg on column (5 μL loadings) using a TSQ Vantage EMR (Thermo Fisher Scientific Inc., San Jose, CA, United States), which was within the linear range for quantification for

all peptides. Automated sampling and chromatography were performed using an Eksigent nanoLC-Ultra 1D plus (AB SCIEX, Dublin, CA, United States). Sample peptides were desalted and chromatographically separated using reverse-phase chromatography with acidified acetonitrile and water mobile phases. Separations were achieved using a 25-min run with a 12.5-min non-linear gradient from 2 to 60% acetonitrile at 500 nL/min. Peptides were measured as they eluted in pre-determined retention time windows, with dwell times optimized to obtain at least ten measurements across the chromatographic peak with at least five transitions monitored per peptide (SRM details provided in Supplementary Table 2). All chromatographic elution profiles for peptides of interest were scrutinized to ensure matching retention times and relative fragment ion intensity ratios (for AQUA and native peptide precursor and product masses, respectively). Peak integration and calculations were performed by LCQuan software (Thermo Fisher Scientific Inc., San Jose, CA, United States). Absolute quantities were calculated by multiplying the response ratio (sum of the fragment mass peak areas for endogenous peptide signal/sum of the fragment mass peak areas for AQUA peptide signal) by the amount of AQUA peptide analyzed.

Statistical Analysis

These analyses were based on the following linear model for the concentrations (μg/mg) for a single allergen:

$$Y_{ijk} = \mu + G_i + L_j + GL_{ij} + \varepsilon_{ijk} \quad (1)$$

where Y_{ijk} is the concentration of protein for the kth sample from Variety i in Location j, μ is the overall mean, G_i is an effect for the ith Variety, L_j is an effect for the jth Location, GL_{ij} is a Variety by Location interaction effect, and ε_{ijk} is a random error that includes sampling error and measurement error.

The SAS[2] (SAS Institute, Inc., Cary, NC, United States) procedure for general linear models (Proc GLM) was used to carry out the mixed effects ANOVA based on the linear model above with variety defined as a fixed factor. The analysis was carried out separately for each allergen. Both transformed and untransformed raw data were used to initially survey the data analysis. Resulting F-tests for analysis of untransformed data are summarized in **Table 1**.

The ratio of β-conglycinin α-subunit to total glycinin concentration was calculated for each sample, and ANOVA was carried out on the ratios to test whether the ratio differed by variety or location. Results for F-tests are at the bottom of **Table 1**.

Multivariate analysis (MANOVA) was used to demonstrate the presence of variety-by-allergen interaction for the summed allergen groups, β-conglycinin α-subunit, total glycinin, and total KTI. Pairs of each of these allergen groups were analyzed to evaluate variety (genotype) and location effects and the possible interaction of these variables with variation of the allergen pairs (see Supplementary Table 3 for analysis and p-values). MANOVA was performed on the summed concentration values for the four glycinin sub-units (total glycinin group) and the summed values for the two KTI sub-units (total KTI); log concentration values were used for analysis. The allergens P34 and Gly m Bd28k were the least abundant allergens and were excluded from MANOVA analysis.

[2]www.sas.com

RESULTS

Influence of Environment and Genetic Background on Allergen Expression

Three varieties of soybean were grown and harvested from nine locations in a single climate zone in North America (Supplementary Figure 1). There were significant differences for each of the allergen concentrations [except for glycinin G4 (Gly G4) and Kunitz trypsin inhibitor (KTI 3)] across each sample type, due to either location or soybean variety (**Figure 1**). Variety and location effects (including statistical interaction) explained a significant fraction of the statistical variation for all allergens (**Table 1**). Soybean variety (genotype) showed the strongest effect statistically as all but glycinin G4 (Gly G4) and KTI 3 showed differences at a significance level of $p < 0.01$. And, of these seven showing strong impact from location, five of them had p-values past the third decimal (reported as < 0.000). In comparison, only four of the allergens, P34, Gly m Bd28k, glycinin G3, and KTI 1, showed significant influence from location and this was at a significance level of $p < 0.05$, but not nearly as consistently low in p-value as shown by that of variety. Variety effects were particularly strong (> 60% of total) for β-conglycinin α-subunit, glycinins G1 and G2, KTI 1, and also for P34 (about 40% of total).

As an example of the greater expression range (i.e., from smallest to largest concentrations) for KTI 1 and the easily distinguishable variety difference, the B2 variety had ∼3-fold lower concentration of KTI 1 compared to the other two varieties. Overall, KTI 1 and KTI 3 were detected at similar levels across site locations, 1.2–3.8 and 2.9–4.2 μg allergen/mg of protein, respectively, and were consistent with the published abundance data for these proteins (Stevenson et al., 2012). However, abundance of the KTI 1 allergen was affected differently compared to KTI 3; the KTI 1 concentration was significantly affected by the variety and by location, whereas KTI 3 was not significantly affected by either.

A statistical evaluation of the location effects showed that the strongest impact was on glycinin G3 and P34; about 50 and 30% of the total statistical variation, respectively (**Table 2**). Results are similar whether or not data were log-transformed (log-transformed data analyses not shown). Residual (i.e., within location and variety) variance was greatest for glycinin G4 and KTI 3, about 47 and 32%, respectively. Importantly, the residual variance displayed in **Table 2** helps to explain why neither glycinin G4 nor KTI 3 showed an impact from either variety or location as separate variables (**Table 1**); these are the two highest residual values indicating that some other variable must be explaining how these allergens are expressed. Note that these are the only two allergens with both location and variety demonstrating a non-significant impact on expression levels (p-values > 0.05; **Table 1**). The residual statistic is also highly significant (**Table 1**) for both of these allergens.

TABLE 1 | *F*-tests for significance: genotype (variety), location, and variety by location interaction effects.

Allergenic protein	Numerator source	Denominator mean square source	Num DF*	Den DF**	*F*-value	*p*-value
P34 (Gly m Bd 30k)	Genotype	Genotype × Location	2	16	19.60	0.000
P34 (Gly m Bd 30k)	Location	Genotype × Location	8	16	5.60	0.002
P34 (Gly m Bd 30k)	Genotype × Location	Residual	16	54	9.42	0.000
β-conglycinin α-subunit	Genotype	Genotype × Location	2	16	250.17	0.000
β-conglycinin α-subunit	Location	Genotype × Location	8	16	1.60	0.202
β-conglycinin α-subunit	Genotype × Location	Residual	16	54	1.94	0.036
Gly m Bd 28k	Genotype	Genotype × Location	2	16	8.71	0.003
Gly m Bd 28k	Location	Genotype × Location	8	16	3.56	0.015
Gly m Bd 28k	Genotype × Location	Residual	16	54	3.03	0.001
Glycinin G1	Genotype	Genotype × Location	2	16	74.19	0.000
Glycinin G1	Location	Genotype × Location	8	16	1.58	0.208
Glycinin G1	Genotype × Location	Residual	16	54	2.43	0.008
Glycinin G2	Genotype	Genotype × Location	2	16	26.56	0.000
Glycinin G2	Location	Genotype × Location	8	16	1.47	0.242
Glycinin G2	Genotype × Location	Residual	16	54	2.91	0.002
Glycinin G3	Genotype	Genotype × Location	2	16	6.73	0.008
Glycinin G3	Location	Genotype × Location	8	16	6.57	0.001
Glycinin G3	Genotype × Location	Residual	16	54	33.78	0.000
Glycinin G4	Genotype	Genotype × Location	2	16	1.76	0.204
Glycinin G4	Location	Genotype × Location	8	16	1.62	0.197
Glycinin G4	Genotype × Location	Residual	16	54	3.35	0.000
KTI 1	Genotype	Genotype × Location	2	16	276.81	0.000
KTI 1	Location	Genotype × Location	8	16	3.17	0.024
KTI 1	Genotype × Location	Residual	16	54	10.10	0.000
KTI 3	Genotype	Genotype × Location	2	16	1.42	0.270
KTI 3	Location	Genotype × Location	8	16	1.24	0.339
KTI 3	Genotype × Location	Residual	16	54	6.55	0.000
β-conglycinin α-subunit/total Glycinin	Genotype	Genotype × Location	2	16	10.12	0.001
β-conglycinin α-subunit/total Glycinin	Location	Genotype × Location	8	16	3.20	0.074
β-conglycinin α-subunit/total Glycinin	Genotype × Location	Residual	16	54	32.67	0.000

*The source of the numerator in this calculation was varied between genotype, location, and genotype*location. The source of the Mean Square used in the denominator was either genotype*location or residual. *Num DF = numerator (as indicated in the Source column) degrees of freedom. **Den DF = denominator (as indicated in Denominator Mean Square column) degrees of freedom.*

Therefore, it is assumed that any variability in expression for glycinin G4 and KTI 3 allergens is largely a representation of sampling/experimental error or other variable not included in our analysis; it appears that neither environment nor variety has an impact on the levels of these two allergens.

Total Allergen Content of Three Commercial Varieties Grown in Nine Different Locations

The total allergen content for soybean sample was computed by summing the concentration values for all nine allergens (**Figure 2**); a breakdown of each allergen concentration as summed for the nine site locations and displayed for each variety is shown in Supplementary Table 4. The summed values for the allergens (**Figure 2**) are shown in order to represent the "total allergen load." However, each allergen has independent distribution around the mean and summing the values across the allergens does not allow a convenient way to represent statistical spread around the mean. Therefore, the data bars are free of standard error or deviation bars. The lowest concentration of allergens was observed in variety A (149 μg/mg at IA Falls) and the highest level in variety B2 (193 μg/mg at Glidden). Variety B2 displayed the highest concentration at each location, except at Sharpsville (B1 was highest at 192 μg/mg for Sharpsville), and varied the least across all of the sites displaying a concentration differential from lowest to highest of approximately 9 μg/mg. In contrast, more variation from location to location was observed for the other two soybean varieties. The concentration differential (difference from lowest to highest value across locations) for variety A was approximately 29 μg/mg and the differential for the variety B1 was 30 μg/mg, both of which were more than three times greater in their differential (i.e., variation across locations) than that of the variety B2.

Correlation Among Soybean Allergens

One question that was asked *a priori* with regard to the relationship of allergen content was whether the allergens correlated with other allergens according to their protein family. This analysis was approached by summing the concentration

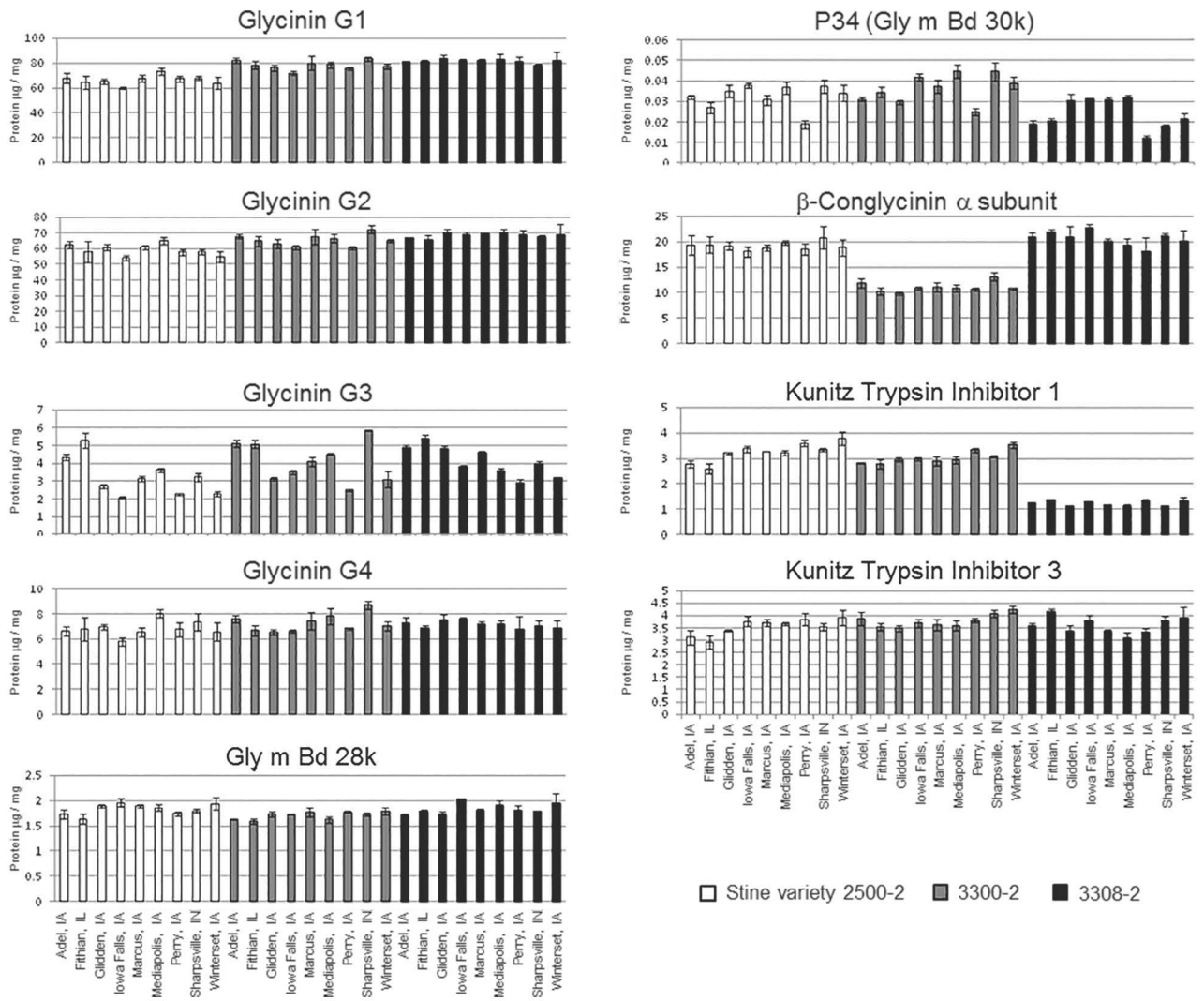

FIGURE 1 | Concentration of nine allergens by location and soybean variety. The concentration of each allergen (µg/mg total protein) is plotted separately in each panel for each allergen. The grayscale shading differentiates the three varieties and the *X*-axis provides the growth location. Error bars denote standard deviation from three replicates (sub-samples of grain, each prepared separately). Statistical analysis of this data is presented in **Table 1**.

values for the allergen isoforms that group into a single protein or protein group; for example, glycinin and Kunitz trypsin inhibitor, respectively. The focus was limited to the most abundant of the allergen groupings in this study; the glycinin and Kunitz trypsin inhibitor allergens. The summed allergen types were then compared to the single representative of the conglycinin protein family, β-conglycinin α-subunit, also one the abundant allergens.

To investigate possible correlations between changes in one allergen concentration versus that of another, log concentrations for one allergen were plotted against another for various pairs of allergens. β-conglycinin α-subunit was then plotted against each of these totals; an example plot is shown in **Figure 3**. Clustering of points associated with the three varieties is evident. This clustering suggests that variety impacts the relationships between allergens and thus, computing simple correlations between pairs of allergens, using all observations and ignoring variety, was not useful. MANOVA was therefore used to demonstrate the presence of variety-by-allergen interaction for the allergen pairs involving β-conglycinin α-subunit, total glycinin and total KTI.

In applying additional MANOVA analyses (Supplementary Table 3), results showed that, as with single allergens or related allergen types in the prior analysis, the soybean variety (noted as genotype) had a significant impact on the relationship between allergens. As an example, when β-conglycinin α-subunit, total glycinin, and total KTI were each compared to the other allergens the variety was a significant variable; in most cases, growing location did not have a statistical impact unless the raw data was log-transformed (Supplementary Table 3; e.g., Total glycinin and total KTI; location F-value 2.73; p-value = 0.0133). With regard to the initial premise, the relationship between allergens was followed across the three varieties and demonstrated that an increase in one allergen type was shown to not necessarily

TABLE 2 | Relative contribution to total statistical variance of variety and location for each of the nine allergen concentrations: main and interaction effects.

		% of Total Variance Contributed by each Variable			
Allergenic protein	**Total variance**	**Variety**	**Location**	**Variety by Location**	**Residual**
P34 (Gly m Bd 30k)	8.89E-05	42.95	31.86	18.57	6.62
β-conglycinin α-subunit	28.23	92.55	0.67	1.62	5.16
Gly m Bd 28k	0.016	25.41	25.31	19.86	29.41
Glycinin G1	77.69	80.13	1.90	5.79	12.18
Glycinin G2	37.00	60.62	3.36	14.01	22.01
Glycinin G3	1.27	17.91	52.28	27.31	2.50
Glycinin G4	0.49	4.48	10.90	37.19	47.43
KTI 1	1.27	94.10	2.22	2.77	0.91
KTI 3	0.12	3.28	5.55	59.18	31.98

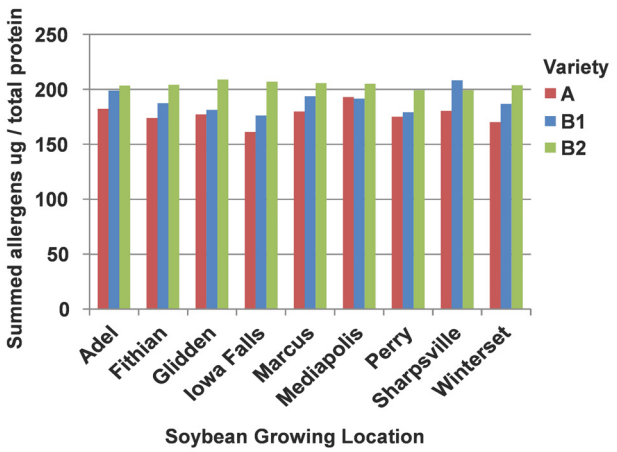

FIGURE 2 | Summed allergen content by location for each soybean variety. The concentrations of each of the allergens were summed by location and plotted for each of the three soybean varieties.

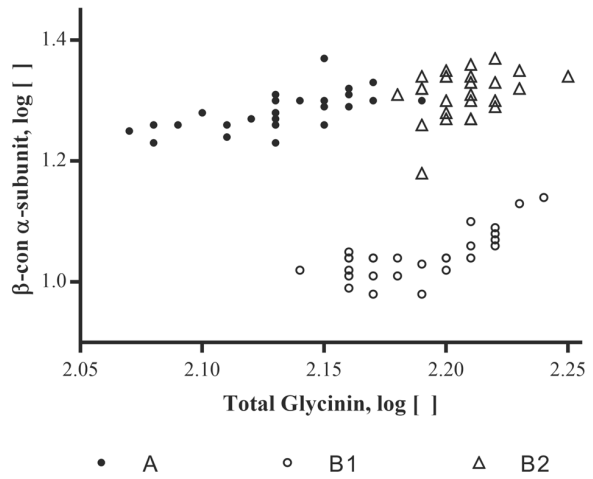

FIGURE 3 | Soybean varieties plotted for their log concentrations of β-conglycinin α-subunit versus the log concentrations of the summed total of the glycinin subunits (Total glycinin). Log concentrations are of the mean μg allergen(s)/mg total protein; each site location per variety and allergen (n = 3).

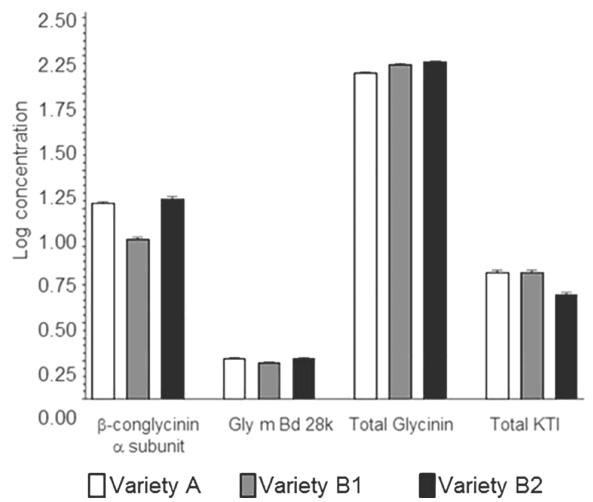

FIGURE 4 | Mean log concentrations of protein allergens by group and variety. Paralogous protein allergen concentrations were summed (individual allergens summed together into allergen groups for glycinin and Kunitz trypsin inhibitor proteins, respectively, for each site location). Data displayed is the mean and standard deviation for each of the varieties at all site locations.

be accompanied by a decrease in another (**Figure 4**; P34 data not shown due to the extremely low concentration values). This indicates that allergen expression does not vary in a predictable manner based on the expression of other allergens, at least based on the varieties examined. In other words, the allergens appear to vary in their expression in response to one or more variables independently of one another. For example, when KTI 1 and KTI 3 were examined as a class of related proteins (i.e., trypsin inhibitors) they were shown to not vary as a related pair of proteins across the three soybean varieties. MANOVA analysis (Supplementary Table 3) showed highly significant variety-by-allergen interaction indicating that an increase in one of these trypsin inhibitors is not accompanied by a similar increase in the other; location by allergen interaction was also significant, though not as strong as the variety by allergen interaction. A more graphical explanation of the effect shows that when KTI 1 concentrations were plotted against KTI 3 (**Figure 5**) varieties A and B1 are similar, but concentrations of KTI 1 for these two varieties are substantially higher in comparison to B2. Further analysis by ANOVA on the summed concentrations of KTI 1 and KTI 3 demonstrated that the sum differs significantly by variety (F = 51.71, df = 2,16, p < .0001) though not by location (F = 1.88, df = 8,16, p = 0.136); again, another reflection of the impact of soybean variety on allergen concentration. This degree of independence for KTI 1 and KTI 3 was also observed by Stevenson et al. (2012) in the way in which these two allergens and their expression were impacted by genotype and location (2012).

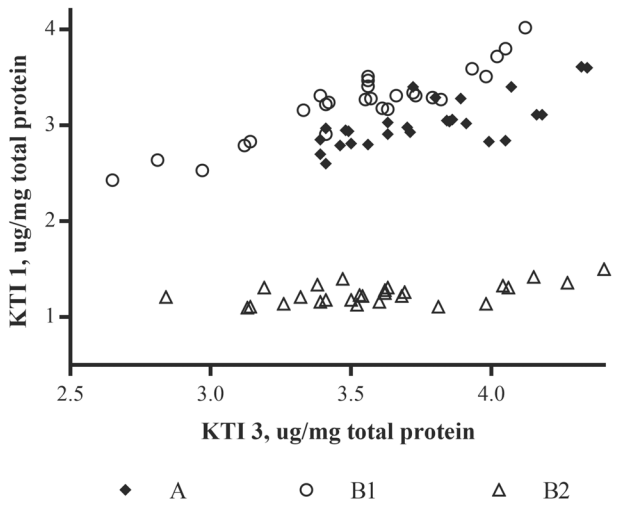

FIGURE 5 | Plot of Kunitz trypsin inhibitor concentrations in soybean seed as measured by LC-SRM mass spectrometry; relationship between KTI 1 and KTI 3 concentrations relative to each of the three soybean varieties. Data represented is for each site location per variety and allergen (*n* = 3).

An interesting exception to the general lack of correlation among the allergens by variety or location was a comparison of β-conglycinin α-subunit to the total glycinin content. The ratio of β-conglycinin α-subunit to total glycinin concentration was calculated for each sample, and ANOVA was carried out on the ratios to test whether the ratio differed by variety or location. The main effect for variety was highly significant and indicated that the concentration of β-conglycinin α-subunit relative to glycinin was dependent on the variety of soybean (bottom of **Table 1**). This may not be surprising since β-conglycinin α-subunit and glycinin are storage proteins and the abundance of one may influence the content of the other in terms of overall seed storage protein levels and their potential biological limitations for expression in the seed. This has been observed before in early attempts to quantify the amounts of these two important proteins in soybeans which were based on studies of different varieties grown at the same location (Boyer, 1982).

DISCUSSION

Environmental factors and genetic traits have long been known to influence quality and productivity of agricultural crops (Harada et al., 1983). Environmental factors encompass both biotic and abiotic elements that influence a plant's growth and development. Biotic factors affecting agricultural production include, but are not limited to, pests, parasites, weeds, and predators. Abiotic factors include temperature, rainfall, irrigation, and pesticide and herbicide application. While many studies have addressed the correlation between genetic and environmental factors affecting crop productivity and composition there are fewer studies that have addressed the impact of these factors on specific protein expression. A recent investigation on allergen variability within four conventional soybean varieties harvested from Georgia, Iowa, Kansas, Nebraska, Ontario, and Pennsylvania during seasonal growth conditions in 2007 revealed diverse environmental growth conditions (spanning three agricultural climate zones) had a greater influence on allergen variability than genetic background (Stevenson et al., 2012). However, due to the single sampling from each location, it was unclear whether such environmental variation was also possible within a limited geographical region.

In the current study, we investigated the effect of environment and soybean variety on the levels of the same soybean seed allergens from field productions of three commercial varieties in nine Midwestern locations (seven in IA, and one each in IL and IN) within a single climate zone. Soybean plots were planted within the typical United States soybean growing region during the growing season of 2009. All plots received conventional herbicide treatment with commercially available herbicides in accordance with typical agronomic practice. The results presented here clearly demonstrate that both genetic and environmental factors influence protein allergen content in soybean seed. It was clear that in the case of the KTI proteins the KTI 3 protein had a much greater range in abundance for two of the varieties (B1 and B2) compared to the variety A (**Figure 5**). Glycinin and β-conglycinin α-subunit abundances were affected by both soybean variety and by growing location. This is important since they are not only major soybean allergens, but they are major seed storage proteins accounting for over 30% of total protein. The four subunits of glycinin displayed differences in seed concentrations with glycinin G1 and glycinin G2 being most abundant followed by glycinin G4 and glycinin G3, and this is consistent with previous studies (Houston et al., 2011; Stevenson et al., 2012) in terms of ranking these isoforms by their abundance. Glycinin G1 and glycinin G2 also display similar patterns of accumulation as influenced by the variety, but were less influenced by growing location (i.e., environmental influence).

To analyze how the concentration of one allergen coincided with the concentration of another, the ratio of β-conglycinin α-subunit was compared to the sum of the four glycinin isoforms and the conclusion reached was that variety does impact the relationship between β-conglycinin α-subunit and total glycinin. Earlier studies have demonstrated that glycinin content can indeed vary by soybean variety and there is evidence that at least 20% of Japanese soybean varieties lack the group IIa $A_5A_4B_3$ glycinin subunit (Harada et al., 1983; Hughes and Murphy, 1983), which would be expected to impact a quantitative measure of the glycinin subunit(s). Environmental growth conditions are also known to significantly impact the levels of glycinin and β-conglycinin levels in mature soybean seeds (Helm et al., 1998; Fehr et al., 2003; Stevenson et al., 2012). Murphy and Resurreccion (1984) reported that environmental differences were more influential than varietal differences in the abundances of glycinin and β-conglycinin of four cultivars (i.e., varieties) grown in different years at different locations. Although the different locations were within the state of IA, there appeared to be a greater magnitude of differences among sites compared to the small differences observed when two varieties were grown side-by-side under the same conditions within the same

year. The conclusions that Murphy and Resurreccion (1984) draw are different than those observed herein and are likely a result of a combination of factors; (a) much different soybean genetic backgrounds and (b) relatively coarse detection methods used which they acknowledge can have a differential effect on measured glycinin content as a proportion of total protein. In particular, there was relatively little variation (3.3–10.5%) which may have been outside the range of precision given the unknown level of variation in the total protein extraction method that was common at the time. In other work, Barros et al. (2010), showed with non-specific protein profiling that site location can be a dominant influence when protein profiling across maize varieties grown during a single season even when the distance between growing locations is relatively short at ∼140 miles (∼225 km); this was apart from the fact that variety did not appear to be an important variable correlated with expression. In contrast, our study indicates that although environmental factors could be detected statistically, the variety of soybean was more dominant in explaining the observed variation in the concentrations of the glycinins and β-conglycinin α-subunit. As an example, the β-conglycinin α-subunit expression was strongly influenced by the variety with almost a twofold lower concentration in the variety B1 compared to the other two varieties (Supplementary Table 4). The overall dominance of the variety effect on allergen content in this study may be due to the fact that seven of the nine growing locations were relatively close to one another and not necessarily representative of the overall variation in United States soybean growing regions. Indeed, studies carried out by Stevenson et al. (2012) demonstrated that growth locations that are geographically distinct exert greater influence than varietal differences on allergen expression.

CONCLUSION

These data employ a new quantitative assay for the determination of a select set of biologically important proteins in a complex matrix. A quantitative approach lessens the impact of assay-specific, sample-to-sample variation while inherently increasing the ability to discriminate study design-specific variables (i.e., better precision). We examined two of the most important variables that affect protein abundance in a complex soybean seed matrix, genetic (variety) and environmental influences. These hypothesized influential variables on allergen expression are characterized as general variables as we do not know the tangible variances of the multiple genes that may vary within a cultivar. With regard to environmental variables, the site locations for the field trial from which the samples in this study were taken are within one growing region for soybean, though there is an expectation that site-to-site variances in actual environmental conditions such as weather, soil, etc., are important agronomically. However, specific measures of both genetics (detailed gene descriptions) and abiotic/biotic environmental conditions were beyond the scope of the work herein in terms of a statistical correlation approach.

We identified that as few as three different soybean varieties can display statistically detectable differential accumulation of protein allergens. Allergens were the focus because analytical detection of these proteins could lead to a better understanding of how best to interpret potential changes in their accumulation. The expectation is that a quantitative assessment of these allergens as well as any other proteins by the methods employed here can be used for the purpose of investigating plant protein expression from a targeted proteomic perspective. Although commodity soybeans are blended to a high degree and the natural variation in expression would be tempered, the ability to measure allergens will likely provide an important analytical technique to support clinical studies of soybean allergy. The concept of natural variability will also play a role in understanding regulatory questions regarding soybean allergens, as a recent review has highlighted (Ladics et al., 2014). A multiplex approach, especially one that is reliably quantitative, can support a greater degree of precision in the design of future studies that seek to understand allergen threshold exposure and related aspects of clinical food allergy. Such a quantitative approach represents a major improvement over serum-based assays by reducing the inconsistency and variability associated with such studies. In future surveys of the potential impact of genetics and environment on allergen expression, it can be imagined that extensive specific field trial designs could catalog and measure many of the discrete variables. This would allow much more extensive means comparisons and principle component type analyses to better understand more than the very top-level variables of "genetics" and "environment" we have assessed in this work. In the near term, this work supports a better understanding of the broad sources of variation with regard to natural protein allergen expression that we consider a stepping stone to providing quantitative allergen detection methodology.

AUTHOR CONTRIBUTIONS

JT and SS sponsored and performed all the laboratory analyses and data compilation. CB provided the statistical analyses. CH-G supported for the sample collection and coordination. CH-G, RH, GL, LP, JW, ND, and SM provided the study design guidance and manuscript editing. SM provided the primary manuscript text, tables, and figures and the editing coordination.

FUNDING

The work presented in this manuscript is the result of a collaborative effort organized under the umbrella of the Health and Environmental Sciences Institute (HESI), a non-profit scientific organization that facilitates public-private partnerships in health and environmental health. As part of this institute, the Protein Allergenicity Technical Committee (PATC) funded part of the work presented in this manuscript, namely: the isolation and quantification of proteins from seed samples and mass spectrometry of those samples employing both spectral counting absolute quantification using synthetic isotopically labeled peptides. Harvested soybean samples analyzed for allergen content were provided by Bayer CropScience.

ACKNOWLEDGMENTS

We thank Lauren Kitchen (PepPro Analytics, LLC), as well as the Health and Environmental Sciences Institute editorial review board, for proofreading and editing this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01025/full#supplementary-material

REFERENCES

Barros, E., Lezar, S., Anttonen, M. J., van Dijk, J. P., Röhlig, R. M., Kok, E. J., 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. doi: 10.1111/j.1467-7652.2009.00487.x

Boyer, J. S. (1982). Plant productivity and environment. *Science* 218, 443–448. doi: 10.1126/science.218.4571.443

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Cai, T., and Chang, K. C. (1999). Processing effect on soybean storage proteins and their relationship with tofu quality. *J. Agric. Food Chem.* 47, 720–727. doi: 10.1021/jf980571z

Ereken-Tumer, N., Richter, J. D., and Nielsen, N. C. (1981). Structural characterization of the glycinin precursors. *J. Biol. Chem.* 257, 4016–4018.

Fehr, W. R., Hoeck, J. A., Johnson, S. L., Murphy, P. A., Nott, J. D., Padilla, G. I., et al. (2003). Genotype and environment influence on protein components of soybean. *Crop Sci.* 43, 511–514. doi: 10.2135/cropsci2003.0511

Hajduch, M., Casteel, J. E., Tang, S., Hearne, L. B., Knapp, S., and Thelen, J. J. (2007). Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. *J. Proteome Res.* 6, 3232–3241. doi: 10.1021/pr070149a

Harada, K., Toyokawa, Y., and Kitamura, K. (1983). Genetic analysis of the most acidic 11S globulin subunit and related characters in soybean seeds. *Jpn. J. Breed.* 33, 23-33. doi: 10.1270/jsbbs1951.33.23

Helm, R. M., Cockrell, G., Herman, E., Burks, A. W., Sampson, H. A., and Bannon, G. A. (1998). Cellular and molecular characterization of a major soybean allergen. *Int. Arch. Allergy Immunol.* 117, 29–37. doi: 10.1159/000023987

Herman, E. M. (2003). Genetically modified soybeans and food allergies. *J. Exp. Bot.* 54, 1317–1319. doi: 10.1093/jxb/erg164

Houston, N. L., Lee, D. G., Stevenson, S. E., Ladics, G. S., Bannon, G. A., McClain, S., et al. (2011). Quantitation of soybean allergens using tandem mass spectrometry. *J. Proteome Res.* 10, 763–773. doi: 10.1021/pr100913w

Hughes, S. A., and Murphy, P. A. (1983). Varietal influence on the quality of glycinin in soybeans. *J. Agric. Food Chem.* 31, 376–379. doi: 10.1021/jf00116a048

Ladics, G. S., Budziszewski, G. J., Herman, R. A., Herouet-Guicheney, C., Joshi, S., Lipscomb, E. A., et al. (2014). Measurement of endogenous allergens in genetically modified soybeans – short communication. *Regul. Toxicol. Pharm.* 70, 75–79. doi: 10.1016/j.yrtph.2014.06.006

Lee, D. G., Houston, N. L., Stevenson, S. E., Ladics, G. S., McClain, S., Privalle, S., et al. (2010). Mass spectrometry analysis of soybean seed proteins: optimization of gel-free quantitative workflow. *Anal. Methods* 2, 1577–1583. doi: 10.1039/c0ay00319k

Mori, T., Utsumi, S., Inaba, H., Kitamura, K., and Harada, K. (1981). Differences in subunit composition of glycinin among soybean cultivars. *J. Agric. Food Chem.* 29, 20–23. doi: 10.1021/jf00103a006

Mujoo, R., Trinh, D. T., and Ng, P. K. W. (2003). Characterization of storage proteins in different soybean varieties and their relationship to tofu yield and texture. *Food Chem.* 82, 265–273. doi: 10.1016/S0308-8146(02)00547-2

Murphy, P. A., and Resurreccion, A. P. (1984). Varietal and environmental differences in soybean glycinin and β-conglycinin content. *J. Agric. Food Chem.* 32, 911–915. doi: 10.1021/jf00124a052

Ogawa, T., Tsuji, H., Bando, N., Kitamura, K., Zhu, Y. L., Hirano, H., et al. (1993). Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci. Biotech. Biochem.* 57, 1030–1033. doi: 10.1271/bbb.57.1030

Petruccelli, S., and Anon, M. C. (1995). Partial reduction of soy protein isolate disulfide bonds. *J. Agric. Food Chem.* 43, 2001–2006. doi: 10.1021/jf00056a008

Riblett, A. L., Herald, T. J., Schmidt, K. A., and Tilley, K. A. (2001). Characterization of β-conglycinin and glycinin soy protein fractions from four selected soybean genotypes. *J. Agric. Food Chem.* 49, 4983–4989. doi: 10.1021/jf0105081

Stevenson, S. E., Woods, C. A., Hong, B., Kong, X., Thelen, J. J., and Ladics, G. S. (2012). Environmental effects on allergen levels in commercially grown non-genetically modified soybeans: assessing variation across North America. *Front. Plant Sci.* 3:196. doi: 10.3389/fpls.2012.00196

Tsuji, H., Bando, N., Hiemori, M., Yamanishi, R., Kimoto, M., Nishikawa, K., et al. (1997). Purification of characterization of soybean allergen Gly m Bd 28K. *Biosci. Biotech. Biochem.* 61, 942–947. doi: 10.1271/bbb.61.942

Utsumi, S. (1992). Plant food protein engineering. *Adv. Food Nutr. Res.* 36, 89–208. doi: 10.1016/S1043-4526(08)60105-9

Utsumi, S., and Kinsella, J. E. (1985). Structure-function relationships in food proteins: subunit interactions in heat-induced gelation of 7S, 11S, and soy isolate proteins. *J. Agric. Food Chem.* 33, 297–303. doi: 10.1021/jf00062a035

Disclaimer: The views and conclusions expressed in this article are those of the authors and do not necessarily represent the policies or positions of their organizations.

Conflict of Interest Statement: Some authors are employed by organizations that develop and market transgenic seed.

Copyright © 2018 McClain, Stevenson, Brownie, Herouet-Guicheney, Herman, Ladics, Privalle, Ward, Doerrer and Thelen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplementary Material

Variation in Seed Allergen Content from Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain1*, Severin E. Stevenson2, Cavell Brownie3 Corinne Herouet-Guicheney4, Rod A. Herman5, Gregory S. Ladics6, Laura Privalle7, Jason M. Ward8, Nancy Doerrer9, Jay J. Thelen10

*** Correspondence:** Scott McClain: scottmcclain24@gmail.com

Supplementary Figure 1:

Supplementary Figure 1. Approximate field trial site locations for soybean seed samples used in this investigation. Note that the field trials for this study comprised a single climate zone for North America, but are spread across many miles in three U.S. states; Iowa (IA), Illinois (IL) and Indiana (IN).

Supplementary Material

Variation in Seed Allergen Content from Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain1*, Severin E. Stevenson2, Cavell Brownie3 Corinne Herouet-Guicheney4, Rod A. Herman5, Gregory S. Ladics6, Laura Privalle7, Jason M. Ward8, Nancy Doerrer9, Jay J. Thelen10

*** Correspondence:** Scott McClain: scottmcclain24@gmail.com

Supplementary Table 1. Target protein and peptide information for quantifying each respective protein allergen using mass spectrometry. (fmol = femtomole; g/mol is molar mass of peptide; inj = injection).

Target Name	Glyma # (for protein target)	Allergen Symbol	Peptide Sequence	fmol/inj (1ug)	g/mol
Glycinin G1	Glyma03g32030.1	GlyG1	VLIVPQNFVVAAR	200	1425.9
Glycinin G2	Glyma03g32020.1 Glyma03g32020.2	GlyG2	NLQGENEEEDSGAIVTVK	100	1931.9
Glycinin G3	Glyma19g34780.1	GlyG3	FYLAGNQEQEFLQYQPQK	100	2231.1
Glycinin G4	Glyma10g04280.1	GlyG4	VESEGGLIQTWNSQHPELK	100	2152.1
Beta-conglycinin, Alpha subunit	Glyma20g28650.1 Glyma20g28650.2 Glyma20g28660.1	BconA1	LITLAIPVNKPGR	100	1391.9
Kunitz trypsin inhibitor 1	Glyma01g10900.1	KTI1	DTVDGWFNIER	100	1351.6
Kunitz trypsin inhibitor 3	Glyma08g45530.1	KTI3	FIAEGHPLSLK	100	1211.7
Gly m Bd 28k	Glyma11g15870.1	AllGly28	DGPLEFFGFSTSAR	100	1530.7
34 kDa maturing seed protein	Glyma08g12270.1	P34	EESETLVSAR	1	1120.5

Supplementary Material

Variation in Seed Allergen Content from Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain1*, Severin E. Stevenson2, Cavell Brownie3 Corinne Herouet-Guicheney4, Rod A. Herman5, Gregory S. Ladics6, Laura Privalle7, Jason M. Ward8, Nancy Doerrer9, Jay J. Thelen10

*** Correspondence:** Scott McClain: scottmcclain24@gmail.com

Supplementary Table 2. Mass spectrometry assay parameters.

Peptide	Peptide Name	Precursor (Th)	Precursor Charge	Product (Th)	Product Charge	Ion Type
EESETLVSAR	P34	560.778	2	432.256	1	y4
EESETLVSAR	P34	560.778	2	545.34	1	y5
EESETLVSAR	P34	560.778	2	646.388	1	y6
EESETLVSAR	P34	560.778	2	862.462	1	y8
EESETLVSAR	P34	560.778	2	991.505	1	y9
EESETLVSAR	P34	560.778	2	333.188	1	y3
EESETLVSAR[HeavyR]	P34	565.782	2	442.264	1	y4
EESETLVSAR[HeavyR]	P34	565.782	2	555.348	1	y5
EESETLVSAR[HeavyR]	P34	565.782	2	656.396	1	y6
EESETLVSAR[HeavyR]	P34	565.782	2	872.471	1	y8
EESETLVSAR[HeavyR]	P34	565.782	2	1001.513	1	y9
EESETLVSAR[HeavyR]	P34	565.782	2	343.196	1	y3
NLQGENEEEDSGAIVTVK	GlyG2	966.463	2	1390.669	1	y13
NLQGENEEEDSGAIVTVK	GlyG2	966.463	2	1276.626	1	y12
NLQGENEEEDSGAIVTVK	GlyG2	966.463	2	1147.584	1	y11
NLQGENEEEDSGAIVTVK	GlyG2	966.463	2	1018.541	1	y10
NLQGENEEEDSGAIVTVK	GlyG2	966.463	2	889.498	1	y9
NLQGENEEEDSGAIVTVK[HeavyK]	GlyG2	970.47	2	1398.683	1	y13
NLQGENEEEDSGAIVTVK[HeavyK]	GlyG2	970.47	2	1284.64	1	y12
NLQGENEEEDSGAIVTVK[HeavyK]	GlyG2	970.47	2	1155.598	1	y11
NLQGENEEEDSGAIVTVK[HeavyK]	GlyG2	970.47	2	1026.555	1	y10
NLQGENEEEDSGAIVTVK[HeavyK]	GlyG2	970.47	2	897.513	1	y9
FIAEGHPLSLK	KTI3	404.564	3	460.312	1	y4
FIAEGHPLSLK	KTI3	404.564	3	557.365	1	y5
FIAEGHPLSLK	KTI3	404.564	3	694.424	1	y6
FIAEGHPLSLK	KTI3	404.564	3	751.446	1	y7
FIAEGHPLSLK	KTI3	404.564	3	880.488	1	y8
FIAEGHPLSLK	KTI3	404.564	3	951.525	1	y9
FIAEGHPLSLK[HeavyK]	KTI3	407.236	3	468.327	1	y4
FIAEGHPLSLK[HeavyK]	KTI3	407.236	3	565.379	1	y5
FIAEGHPLSLK[HeavyK]	KTI3	407.236	3	702.438	1	y6
FIAEGHPLSLK[HeavyK]	KTI3	407.236	3	759.46	1	y7
FIAEGHPLSLK[HeavyK]	KTI3	407.236	3	888.502	1	y8

FIAEGHPLSLK[HeavyK]	KTI3	407.236	3	959.539	1	y9
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	854.444	2	y15
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	962.481	2	y17
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	486.292	1	y4
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	1367.67	1	y11
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	1239.611	1	y10
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	1138.563	1	y9
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	740.881	2	y12
VESEGGLIQTWNSQHPELK	GlyG4	718.027	3	445.192	1	b3
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	858.451	2	y15
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	966.488	2	y17
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	494.306	1	y4
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	1375.684	1	y11
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	1247.625	1	y10
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	1146.578	1	y9
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	744.888	2	y12
VESEGGLIQTWNSQHPELK[HeavyK]	GlyG4	720.698	3	445.192	1	b3
LITLAIPVNKPGR	BconA	464.629	3	457.288	1	y4
LITLAIPVNKPGR	BconA	464.629	3	571.331	1	y5
LITLAIPVNKPGR	BconA	464.629	3	670.399	1	y6
LITLAIPVNKPGR	BconA	464.629	3	767.452	1	y7
LITLAIPVNKPGR	BconA	464.629	3	880.536	1	y8
LITLAIPVNKPGR	BconA	464.629	3	951.573	1	y9
LITLAIPVNKPGR	BconA	464.629	3	1064.657	1	y10
LITLAIPVNKPGR	BconA	464.629	3	1165.705	1	y11
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	467.296	1	y4
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	581.339	1	y5
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	680.407	1	y6
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	777.46	1	y7
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	890.544	1	y8
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	961.581	1	y9
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	1074.665	1	y10
LITLAIPVNKPGR[HeavyR]	BconA	467.965	3	1175.713	1	y11
FYLAGNQEQEFLQYQPQK	GlyG3	1116.042	2	663.346	1	y5
FYLAGNQEQEFLQYQPQK	GlyG3	1116.042	2	791.404	1	y6
FYLAGNQEQEFLQYQPQK	GlyG3	1116.042	2	904.488	1	y7
FYLAGNQEQEFLQYQPQK	GlyG3	1116.042	2	1051.557	1	y8
FYLAGNQEQEFLQYQPQK	GlyG3	1116.042	2	1180.599	1	y9
FYLAGNQEQEFLQYQPQK	GlyG3	1116.042	2	372.224	1	y3
FYLAGNQEQEFLQYQPQK[HeavyK]	GlyG3	1120.049	2	671.36	1	y5
FYLAGNQEQEFLQYQPQK[HeavyK]	GlyG3	1120.049	2	799.418	1	y6
FYLAGNQEQEFLQYQPQK[HeavyK]	GlyG3	1120.049	2	912.502	1	y7
FYLAGNQEQEFLQYQPQK[HeavyK]	GlyG3	1120.049	2	1059.571	1	y8
FYLAGNQEQEFLQYQPQK[HeavyK]	GlyG3	1120.049	2	1188.613	1	y9
FYLAGNQEQEFLQYQPQK[HeavyK]	GlyG3	1120.049	2	380.238	1	y3
VLIVPQNFVVAAR	GlyG1	713.432	2	425.312	1	b3
VLIVPQNFVVAAR	GlyG1	713.432	2	501.28	2	y9
VLIVPQNFVVAAR	GlyG1	713.432	2	1001.552	1	y9
VLIVPQNFVVAAR	GlyG1	713.432	2	326.243	1	b2
VLIVPQNFVVAAR	GlyG1	713.432	2	904.499	1	y8
VLIVPQNFVVAAR	GlyG1	713.432	2	1100.621	1	y10
VLIVPQNFVVAAR	GlyG1	713.432	2	776.441	1	y7
VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	425.312	1	b3

VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	506.284	2	y9
VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	1011.56	1	y9
VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	326.243	1	b2
VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	914.508	1	y8
VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	1110.629	1	y10
VLIVPQNFVVAAR[HeavyR]	GlyG1	718.437	2	786.449	1	y7
DTVDGWFNIER	KTI1	676.317	2	531.288	1	y4
DTVDGWFNIER	KTI1	676.317	2	921.457	1	y7
DTVDGWFNIER	KTI1	676.317	2	678.356	1	y5
DTVDGWFNIER	KTI1	676.317	2	1036.484	1	y8
DTVDGWFNIER	KTI1	676.317	2	864.436	1	y6
DTVDGWFNIER[HeavyR]	KTI1	681.322	2	541.296	1	y4
DTVDGWFNIER[HeavyR]	KTI1	681.322	2	931.466	1	y7
DTVDGWFNIER[HeavyR]	KTI1	681.322	2	688.365	1	y5
DTVDGWFNIER[HeavyR]	KTI1	681.322	2	1046.492	1	y8
DTVDGWFNIER[HeavyR]	KTI1	681.322	2	874.444	1	y6
DGPLEFFGFSTSAR	AllGly28	765.865	2	1148.536	1	y10
DGPLEFFGFSTSAR	AllGly28	765.865	2	1019.494	1	y9
DGPLEFFGFSTSAR	AllGly28	765.865	2	872.426	1	y8
DGPLEFFGFSTSAR	AllGly28	765.865	2	725.357	1	y7
DGPLEFFGFSTSAR	AllGly28	765.865	2	1415.695	1	y13
DGPLEFFGFSTSAR	AllGly28	765.865	2	1358.673	1	y12
DGPLEFFGFSTSAR	AllGly28	765.865	2	1261.621	1	y11
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	1158.545	1	y10
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	1029.502	1	y9
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	882.434	1	y8
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	735.365	1	y7
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	1425.703	1	y13
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	1368.682	1	y12
DGPLEFFGFSTSAR[HeavyR]	AllGly28	770.869	2	1271.629	1	y11

Supplementary Material

Variation in Seed Allergen Content from Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain1*, Severin E. Stevenson2, Cavell Brownie3 Corinne Herouet-Guicheney4, Rod A. Herman5, Gregory S. Ladics6, Laura Privalle7, Jason M. Ward8, Nancy Doerrer9, Jay J. Thelen10

*** Correspondence:** Scott McClain: scottmcclain24@gmail.com

Supplementary Table 3. MANOVA results for pairs of the most abundant allergens. Glycinin sub-units were summed to create "total glycinin" values for analysis. KTI1 and KTI3 were similarly summed to create "Total KTI". Beta-conglycinin is the same as "β-conglycinin α-subunit" listed elsewhere. Variety is equivalent to the descriptor "genotype".

Beta-conglycinin and total glycinin				
Model Effect	**DF**	**Den, DF**	**F value**	***p*-value**
Genotype	2	54	291.07	<0.0001
Location	8	54	3.65	0.0018
Location*Genotype	16	54	2.53	0.0057
Allergengrp	1	54	189838	<0.0001
Genotype*allergengrp	2	54	1739.29	<0.0001
Location*allergengrp	8	54	8.59	<0.0001
Location*Genotype*allergen	16	54	2.99	0.0013
Beta-conglycinin and total KTI				
Model Effect	**DF**	**Den, DF**	**F value**	***p*-value**
Genotype	2	54	176.21	<0.0001
Location	8	54	4.04	0.0008
Location*Genotype	16	54	3.74	0.0001
Allergengrp	1	54	33313.7	<0.0001
Genotype*allergengrp	2	54	2621.23	<0.0001
Location*allergengrp	8	54	20.73	<0.0001
Location*Genotype*allergen	16	54	7.47	<0.0001
Total glycinin and total KTI				

Model Effect	**DF**	**Den, DF**	**F value**	***p*-value**
Genotype	2	54	84.31	<0.0001
Location	8	54	2.73	0.0133
Location*Genotype	16	54	3.66	0.0002
Allergengrp	1	54	1112747	<0.0001
Genotype*allergengrp	2	54	2544.47	<0.0001
Location*allergengrp	8	54	78.36	<0.0001
Location*Genotype*allergen	16	54	29.61	<0.0001
KTI1 and KTI3				
Model Effect	**DF**	**Den, DF**	**F value**	***p*-value**
Genotype	2	54	946.27	<0.0001
Location	8	54	14.34	<0.0001
Location*Genotype	16	54	8.28	<0.0001
Protein_name	1	54	10159.9	<0.0001
Genotype*Protein_name	2	54	4555.33	<0.0001
Location*Protein_name	8	54	10.15	<0.0001
Location*Genotype*Protein	16	54	4.93	<0.0001

Supplementary Material

Variation in Seed Allergen Content from Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain1*, Severin E. Stevenson2, Cavell Brownie3 Corinne Herouet-Guicheney4, Rod A. Herman5, Gregory S. Ladics6, Laura Privalle7, Jason M. Ward8, Nancy Doerrer9, Jay J. Thelen10

*** Correspondence:** Scott McClain: scottmcclain24@gmail.com

Supplementary Table 4. Mean concentrations (μg/mg total protein) by variety for each allergen.

Allergen protein	**Variety A**	**Variety B1**	**Variety B2**
P34 (Gly m Bd 30k)	0.032	0.036	0.024
β-conglycinin α-subunit	19.18	11.08	20.57
Gly m Bd 28k	1.83	1.71	1.83
Glycinin G1	66.4	78.15	81.54
Glycinin G2	59.06	65.38	68.54
Glycinin G3	3.21	4.09	4.13
Glycinin G4	6.81	7.25	7.14
KTI1	3.24	3.03	1.24
KTI3	3.54	3.77	3.6

Supplementary Material

Variation in Seed Allergen Content from Three Varieties of Soybean Cultivated in Nine Different Locations in Iowa, Illinois, and Indiana

Scott McClain1*, Severin E. Stevenson2, Cavell Brownie3 Corinne Herouet-Guicheney4, Rod A. Herman5, Gregory S. Ladics6, Laura Privalle7, Jason M. Ward8, Nancy Doerrer9, Jay J. Thelen10

*** Correspondence:** Scott McClain: scottmcclain24@gmail.com

Supplementary Table 5. Site-Specific Soil Description and Weather Conditions Associated with the 2009 Grain Production Trials.

Closest Town	GPS Coordinates	Soil Type	Soil pH	Organic Matter (%)	Month	Avg Max Temperature (°F)	Av. Min Temperature (°F)	Total Rainfall (Inch)
Adel, Iowa	41.70291 -94.16245	Caniesteo Silty Clay	7.6 – 8.4	0.1 – 7.0	May June July August September October	77 81 80 79 77 53	56 61 60 61 54 38	2.95 5.20 3.10 4.79 0.88 7.15
Glidden, Iowa	42048600 -94.601260	Nicollet Loam, Webster Clay Loam	6.7 – 7.0	5.3 – 6.3	May June July August September October	74 81 80 79 78 54	54 61 60 60 51 39	1.89 4.81 3.63 4.66 0.45 4.22
Iowa Falls, Iowa	42.559240 -93.470050	Webster Silty Clay Loam, Clarion Loam, Caniesteo Silty Clay Loam	6.9 – 7.9	1.69 – 4.9	May June July August September October	65 75 76 74 73 50	46 57 56 56 48 36	2.34 3.65 2.07 0.94 2.86 4.64
Marcus, Iowa	42.877536 -95.819992	Primghar Silty Clay Loam	6.5	5.5	May June July August September October	73 78 79 79 76 51	49 59 59 59 53 37	0.82 0.83 0.65 1.16 1.11 4.53
Mediapolis, Iowa	41.020920 -91.141540	Taintor Silty Clay Loam	6.4	2.35	May June July August September October	77 83 80 82 77 57	55 62 58 60 51 39	4.76 7.32 4.60 9.47 1.01 7.14
Perry, Iowa	41.792190 -94.067280	Clarion Loam,	5.6 –	1.43 – 3.37	May June	74 80	53 61	2.95 5.20

		Caniesteo Silty Clay Loam	7.8		July	80	59	3.10
					August	79	60	4.79
					September	77	51	0.88
					October	53	37	7.15
Winterset, Iowa	41.032420 -94.095100	Sharpsburg Silty Clay	6.0	1.76	May	76	51	3.88
					June	84	59	6.84
					July	84	57	2.97
					August	83	58	5.38
					September	82	49	1.42
					October	56	35	6.25
Fithian, Illinois	40.177133 -87.838050	Drummer Silty Clay Loam	6.1	4.5 - 7.0	May	80	59	5.90
					June	82	63	3.95
					July	78	61	4.01
					August	80	63	5.20
					September	76	57	1.93
					October	58	43	7.75
Sharpsville, Indiana	40.374250 -86.215917	Patton Silty Clay Loam	6.7	4.5	May	86	55	4.87
					June	86	61	3.70
					July	85	58	3.36
					August	85	58	2.94
					September	79	54	1.53
					October	60	41	5.84
					November	59	30	1.52