JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Impact of Environment on the Biomass Composition of Soybean (*Glycine max*) seeds

Tamara McClure,[†,‡] Jean-Christophe Cocuron,[†,‡] Veronika Osmark,[§] Leah K. McHale,[‡,∥,⊥] and Ana Paula Alonso*[,†,‡,⊥]

[†]The Ohio State University, Department of Molecular Genetics, Columbus, Ohio 43210, United States

[‡]The Ohio State University, Center for Applied Plant Sciences, Columbus, Ohio 43210, United States

[§]Høgskolen i Sør Trøndelag, Trondheim, 7004, Norway

[∥]The Ohio State University, Department of Horticulture and Crop Science, Columbus, Ohio 43210, United States

[⊥]The Ohio State University, Center for Soybean Research, Columbus, Ohio 43210, United States

Supporting Information

ABSTRACT: Factors including genetics, fertilization, and climatic conditions, can alter the biomass composition of soybean seeds, consequently impacting their market value and usage. This study specifically determined the content of protein and oil, as well as the composition of proteinogenic amino acids and fatty acids in seeds from 10 diverse soybean cultivars grown in four different sites. The results highlighted that different environments produce a different composition for the 10 cultivars under investigation. Specifically, the levels of oleic and linoleic acids, important contributors to oil stability, were negatively correlated. Although the protein and oil contents were higher in some locations, their "quality" was lower in terms of composition of essential amino acids and oleic acid, respectively. Finally, proteinogenic histidine and glutamate were the main contributors to the separation between Central and Northern growing sites. Taken together, these results can guide future breeding and engineering efforts aiming to develop specialized soybean lines.

KEYWORDS: *Soybean, crop improvement, biomass composition, oil, protein, essential fatty acids, essential amino acids*

1. INTRODUCTION

The soybean plant (*Glycine max* (L.) Merrill) is a species of legume that produces pods containing beans. Current soybean cultivars are phenotypically distinct from their wild relative *Glycine soja* due to the selection pressures occurring over thousands of years of agricultural production[1] and ongoing plant breeding efforts.[2] The fresh beans from green pods can be eaten as a vegetable, and dry mature seeds are used not only in animal feeds and human foods that contain soybean oil or meal, but also in industry (solvents, lubricants, inks, plastics, waxes, etc.).[3] The market value of soybean stems largely from the significant oil and protein content of its seeds which accounts for approximately 40% and 20% of total dry weight, respectively, giving the seeds appreciable versatility.[4,5] Because of these attributes, soybean accounts for a significant amount of the world's vegetable oil, animal fodder and food for human consumption.

Soybean is an integral part of the diet of many Asian cultures with applications in items such as soy milk, soy sauce, soy paste, edamame, tempeh, miso, tofu, and natto.[3] Among western cultures soybean is used mostly for its meal and oil. In 2014, 47.7% of soybeans produced in the United States were crushed domestically for oil and meal, 46.9% was exported, and the remaining was used for seed and other purposes. Soybean exports have increased by about 85% from 2000 to 2014, from 996 to 1843 million bushels, for a total price of $18.6 billion in 2014.[6] The United States is a top soybean producer along with Brazil and followed by Argentina, China, and India. According to the United States Department of Agriculture, soybeans comprise about 90% of U.S. oilseed production and make up the world's largest protein source in animal feed and second largest source of vegetable oil, emphasizing not only the United States' role as a major producer, but also the global demand for soybean. U.S. soybean production is the highest in Midwestern states including Illinois, Iowa, Minnesota, Indiana, Nebraska and Ohio.[7] Research regarding soybean is relevant in the context of these major regions of production.

Both breeding programs and genetic modification efforts are used for the purpose of creating specialized soybean lines. Successful efforts include the creation of soybean lines with increased target nutrients, and others with resistances to herbicides, pesticides, and pathogens.[8−10] Breeding has been a widely employed strategy for improving soybean for over two thousand years,[2] and genetically modified soybeans have been commercially grown in the U.S. since 1996.[11] Current soybean breeding programs and genetic modification efforts aim to improve nutritional value, among other traits. Specifically, soybean seed biomass components such as protein and oil content and composition have been targeted in research due to their importance to the soybean market.[12]

Traditional soybean oil is mainly comprised of palmitic acid (C16:0, approximately 10% of crude oil), stearic acid (C18:0, approximately 4%), oleic acid (C18:1, approximately 22%), and

Received: March 29, 2017
Revised: June 13, 2017
Accepted: July 19, 2017
Published: July 19, 2017

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

essential linoleic (C18:2, omega-6, approximately 54%) and linolenic (C18:3, omega-3, 10%) acids, which are necessary for health and must be obtained by consumption.[13] Though essential and desirable for human health, linolenic and linoleic acids are responsible for oxidative instability of soybean oil, which has been historically addressed through partial hydrogenation.[4,5] Once hydrogenated and refined, soybean oil is a stable source of vegetable oil; however this is at the cost of health, as the trans-fats that result from the hydrogenation process have been shown to increase the risk of coronary heart disease.[5,14] Naturally high oleic acid content in soybean (target value >70% of total oil) is valuable due to this monounsaturated omega-9 fatty acid's stability at high temperatures, which reduces the formation of trans-fats during vegetable oil production and use.[5] For the same reason, low linolenic soybean (target value <3% of total oil) has also been pursued to decrease the amount of oxidative instability. Mid to high oleic acid soybeans (30−70% of total oil) have been obtained.[13,15,16] Additionally, low linolenic acid soybean oil genotypes have also been developed through breeding.[17,18] In parallel, genetic engineering efforts produced soybean varieties with high oleic acid content, such as Vistive Gold and Plenish.

Regarding protein composition, current priorities for improving soybean meal aim at increasing the total protein and the relative levels of the sulfur containing amino acids cysteine and methionine that are only present in a small amount and are critical for animal health.[9,19] However, current understanding of the variation in amino acid profiles among cultivars is inadequate and there have only been a few studies on the effect of growing environment on seed amino acid content or on the combined effects of environment and cultivar on proteinogenic amino acid composition.[20,21]

Additional efforts to simultaneously increase the contents of both oil and protein in soybean seeds have had limited success due to an apparent inverse correlation of these two components.[4,22,23] Therefore, maximizing both may require alternative strategies. Both breeding programs and genetic modification efforts are used for the purpose of creating specialized soybean lines.

Genotype and environmental factors have been shown to affect some biomass components in soybean.[24,25] In order to develop product-specific soybeans, these major factors must be further investigated and understood. Determining correlations between biomass components and both genotype and environmental factors in soybean will guide breeding and genetic engineering efforts, and allow for the adjustment of agricultural practices to optimize the creation of product-specific soybean lines. The goal of this study was 2-fold: (i) provide a comprehensive biochemical characterization of select Ohio-adapted soybean lines, and (ii) determine the relative impact of location on traits that influence the market value of soybean. For this purpose, 10 different soybean cultivars (Table 1) were grown at four locations in Ohio that covered three different soil types (Supporting Information (SI) Figure S1). We anticipate that this study will guide future breeding and engineering efforts aiming to develop specialized soybean lines.

2. MATERIALS AND METHODS

2.1. Chemicals. Triheptadecanoin standard, 3 N methanolic-HCl and toluene were purchased from Sigma-Aldrich (St. Louis, MO). Hexane, isopropanol, cation exchange resin (DOWEX 50WX8, 100−200 mesh), ammonium hydroxide, hydrochloric acid (HCl) and LC-MS grade acetonitrile were ordered through ThermoFisher (Pittsburgh, PA).

Table 1. Soybean Cultivars Investigated in This Study

cultivar	classification	status
dennison	commodity	public cultivar
H09−4	high protein	breeding line
HM09-W043	high protein	breeding line
HM09-W053	commodity	public cultivar (branded)
HM09-W150	commodity	public cultivar (branded)
HM09-W153	food grade	public cultivar (branded)
Ohio FG1	food grade	public cultivar
Ohio FG5	food grade	public cultivar
Summit	commodity	public cultivar
Wyandot	commodity/food grade	public cultivar

2.2. Plant Materials. For this study, 10 soybean cultivars: Dennison,[26] H09−4 (sister line to Highpro1[27]), HM09-W043, HM09-W053, HM09-W150, HM09-W153, Ohio FG1,[28] Ohio FG5,[29] Summit,[30] and Wyandot (Table 1) were grown as part of the Ohio Advanced Line Tests. Each cultivar was grown in the Ohio locations Plain City (PC), South Charleston (WE), Hoytville (NW), and Wooster (WO) during the summer of 2013 (SI Figure S1). PC and WE plots consisted of six 4 m long rows with 38 cm spacing. Approximately 500 seeds were planted per plot for a seeding density of ∼50 seeds/m^2 in the harvested area. The inner four rows were harvested at maturity in fall of 2013 with a Massey Ferguson eight plot combine. NW and WO plots consisted of eight 4.9 m rows with 19 cm spacing for the inner six rows and 72 cm spacing for the two outermost rows. Approximately 600 seeds were planted per plot for a seeding density of ∼55 seeds/m^2 in the harvested area, and the inner six rows were harvested in the fall of 2013 with a Wintersteiger Classic plot combine. Three of the growth locations, WE, NW, WO, are Ohio Agricultural Research and Development Center (OARDC) farms and employed similar agricultural practices whereas PC is non-OARDC and was managed according to farmer practices.

2.3. Biomass Extraction. Mature soybean seeds were dried in a lyophilizer and weighed to obtain the initial dry weight. Seeds were individually ground using a mortar and pestle, and 50 mg of the soybean powder was transferred in a 2 mL screw cap tube for further analysis. Fatty acids and proteins were extracted as previously described[31] with the following modifications: (i) the fatty acid extractions were carried out with 50 μL of triheptadecanoin (10 mg/mL) as an internal standard; and (ii) proteins were extracted using 1 mL of extraction buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 1% SDS). The process was repeated twice for a final volume of 3 mL of protein-containing supernatant to be used for quantification.

2.4. Biomass Quantification. Total oil content and oil composition of each seed sample was determined by Gas Chromatography−Mass Spectrometry (GC-MS) analysis of fatty acid methyl esters as previously described[31] with some modifications. Fatty acids were resuspended in 300 μL of toluene, and methylated with 1 mL of 3N methanolic/HCl. Transmethylation was carried out for 120 min at 80 °C, and the reaction was quenched using 500 μL sodium bisulfate 5% (w/v). Oil content was quantified using GC-MS analysis as previously described[32] using a run time of 6 min. Protein concentration of each sample was measured using the Biorad DC Protein Assay kit, as previously described.[33]

2.5. Proteinogenic Amino Acids. Two hundred 50 μL of protein extract was hydrolyzed using 6 N HCl for 24 h at 120 °C. HCl was evaporated at 60 °C under a stream of nitrogen, and samples were resuspended in 0.01 N HCl. Then, proteinogenic amino acids were loaded on a cation exchange resin (Dowex 50 × 8). After washing with water (5 × 1 mL), amino acids were eluted with 1 N NH_4OH (5 × 1 mL). After 60 min under nitrogen flow, the samples were lyophilized overnight. Proteinogenic amino acids were resuspended in 500 μL of 0.01 N HCl, and analyzed by Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) using the method previously described.[31] The quantification of each amino acid was possible by running known

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

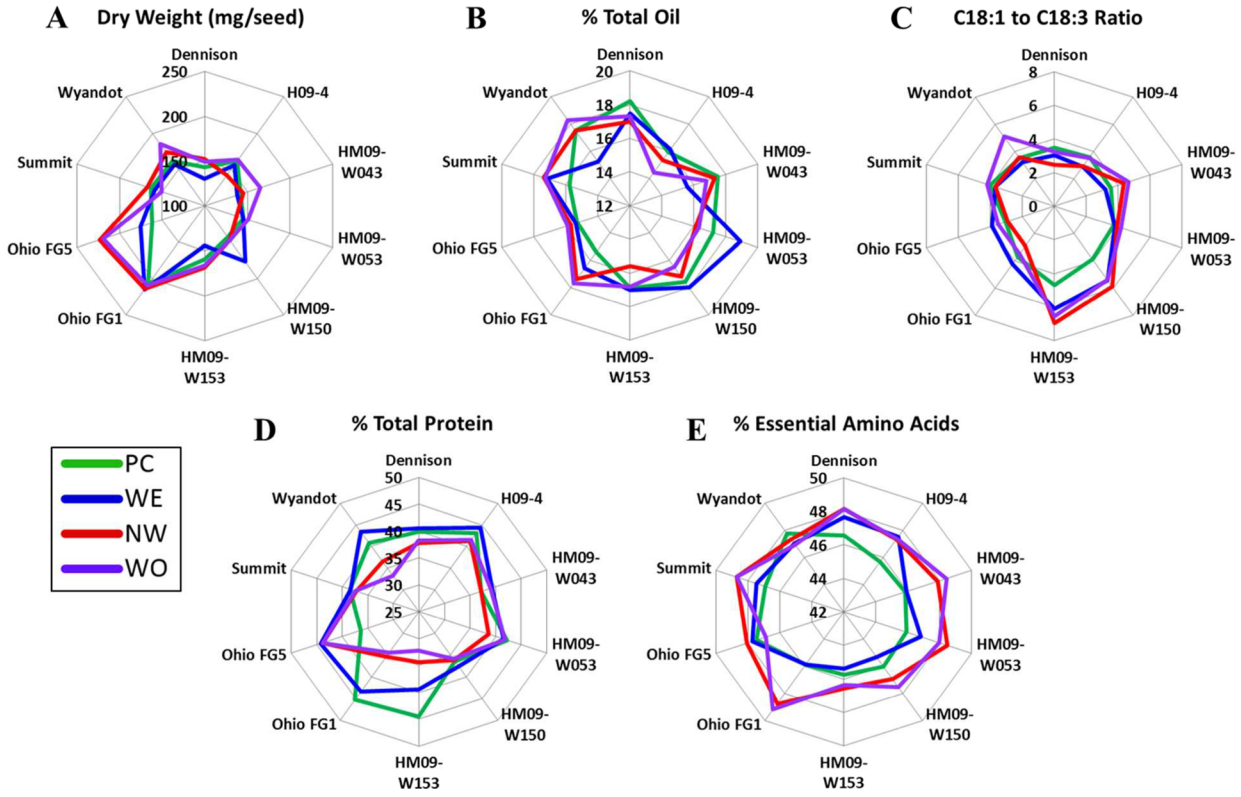

Figure 1. Biomass characterization of 10 soybean cultivars grown in four different locations. (A) Radar showing the average seed dry weight in mg/seed (n = 20), (B) Radar depicting the average total oil in % dry weight (w/w; n = 4), (C) Radar showing the average ratio of oleic (C18:1) acid to linolenic (C18:3) acid (n = 4), (D) Radar depicting the average total protein in % dry weight (w/w; n = 4), and (E) Radar showing the average total essential amino acids (% total amino acids; n = 4). The green, blue, red and purple lines refer to the four growing sites: Plain City (PC), South Charleston (WE), Hoytville (NW), and Wooster (WO), respectively.

concentrations of external standards. The quantitation method included 16 amino acids: alanine, arginine, aspartate, glutamate, glycine, histidine, methionine, leucine, isoleucine, lysine, phenylalanine, proline, serine, threonine, tyrosine, and valine. It is important to note that in this procedure cysteine was converted into cystine,[34] tryptophan was degraded, and asparagine and glutamine were deamidated into their respective acids. Therefore, the quantities of glutamate and aspatarte reported in this study are the sum of glutamate + glutamine, and aspartate + asparagine, respectively.

2.6. Statistical Analyses. For each biomass component under study, the mean and standard deviation of the four biological replicates were calculated for each location and each cultivar. To determine statistical significance, a student's t test was used to compare each parameter under study with respect to location. P-values of <0.05 were considered significant. Pearson-correlation analysis, unsupervised principal component analysis (PCA), and supervised partial least-squares discriminate analysis (PLS-DA) were performed using Metaboanalyst v3.0, a free web-based statistical software (www.metaboanalyst.ca).[35] Biomass components were first normalized using log2 function, then mean-centered and divided by the standard deviation of each variable.

3. RESULTS AND DISCUSSION

3.1. Impact of Growing Location on Soybean Biomass Composition. *3.1.1. Seed Dry Weight.* This entire work incorporated a total of 10 diverse soybean cultivars (Table 1) grown at four different sites (SI Figure S1). In order to perform the calculations necessary to quantify and characterize biomass composition, the dry weight for each biological replicate seed was measured (Figure 1A, SI Table S1). Eight of the 10 cultivars under study, Dennison, H09−4, HM09-W043, HM09-W150, HM09-W153, Ohio FG5, Summit and Wyandot showed significant differences (p < 0.05) in dry weight depending on growth location. For instance, Ohio FG5 seeds were the heaviest at the NW and WO sites (223.8 ± 28.9 and 219.8 ± 26.0 mg/seed, respectively), but were up to 28% lighter at PC and WE where row spacing was wider. These variations in seed weight for Ohio FG5 cultivar can be due to differences in soil type and/or the row spacing. On the other hand, the dry weights of Ohio FG1 and HM09-W053 seeds were not affected by the locations. Ohio FG1 had the heaviest seeds overall with dry weights over 205 mg/seed independently of the growing site.

Seed size is particularly important for food-grade soybean.[36] Large seed (>200 mg/seed) is a very desirable trait for the production of tofu, edamame, miso, and soy milk, whereas smaller seeds (<120 mg/seed) are used for natto, soy sauce, and tempeh. Among the four food-grade soybean cultivars investigated in this study (Table 1), large seeds were harvested for Ohio FG1 in the four locations, and Ohio FG5 when it was grown in NW and WO. These results underline the impact of the growing site on seed weight, and consequently on market application and value.

3.1.2. Seed Oil Content and Composition. Total oil content was determined as a percentage of total dry weight for seeds of each cultivar grown at each location along with fatty acid composition (Figure 1B, SI Table S2). Total fatty acid levels varied between 15.0 and 18.9% (w/w) with H09−4 seeds harvested in WO and HM09-W053 seeds from WE site contributing the low and high extremes, respectively. Four of the 10 cultivars, HM09-W043, HM09-W053, Ohio FG1, and Wyandot, showed significant differences (p < 0.05) in total oil

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

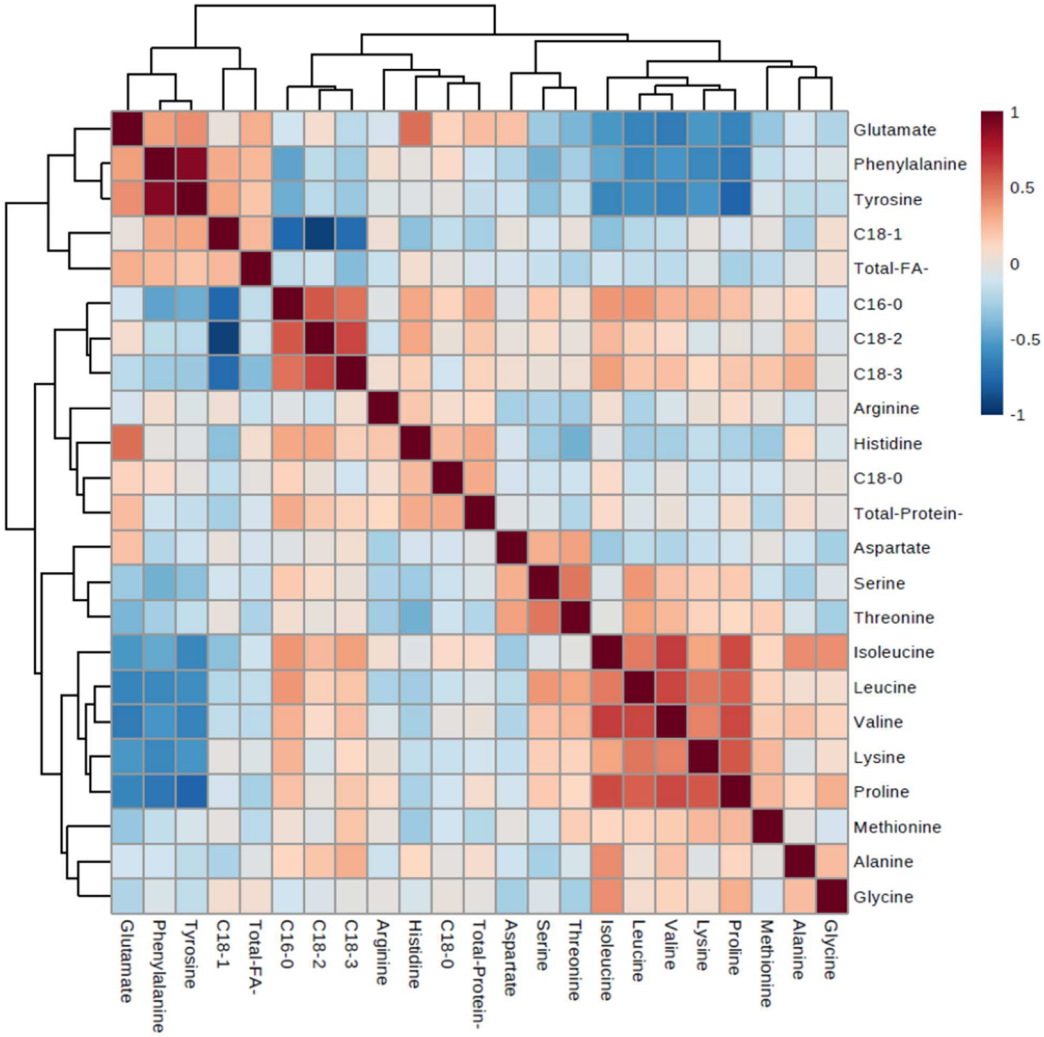

Figure 2. Pearson correlation between the biomass components of 10 soybean cultivars grown in four different locations. Heat map of the Pearson correlation coefficients depicting the strength of association between all biomass components under study where 1 is the strongest positive correlation and −1 is the strongest negative correlation.

content depending on growth location. These variations can be due to differences in soil type and/or row spacing. The percentage of fatty acids in HM09-W053 decreased from 18.9 ± 0.5 to 16.1 ± 0.4% in seeds harvested in WO and PC, respectively. Overall, growing site had the strongest impact on the total oil content of the Wyandot cultivar with 18.3 ± 0.3% at WO with 19 cm row spacing and only 15.2 ± 0.4% at WE with 38 cm row spacing.

Fatty acid composition was determined in seeds from the 10 soybean cultivars harvested from four different sites. Palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) were found to be the major fatty acid species (SI Table S2). Interestingly, for each cultivar under investigation, the growing location significantly affected the content of at least one fatty acid species. Since naturally high oleic acid and low linolenic acid contents in soybean are valuable to improve the stability of oil, we were especially interested in the levels of these two fatty acid species, though the variation in this study does not approach what is required by the soybean oil market.

For oleic acid (C18:1), six out of the 10 cultivars, Dennison, H09-4, HM09-W043, HM09-W150, HM09-W153, and Ohio FG5, produced seeds with significant differences with respect to growth location (SI Table S2). For example, the highest overall oleic acid levels were measured in HM09-W153 seeds grown at NW and WO with narrow row spacing (43.9 ± 2.9 and 41.0 ± 4.6% of total oil content, respectively). In addition, there were significant differences in linolenic acid (C18:3) across locations for the Dennison, H09-4, HM09-W043, HM09-W150, Ohio FG1 and Wyandot cultivars (SI Table S2). The largest variation in C18:3 levels was measured in Dennison seeds from PC and NW with 9.0 ± 0.6 and 7.3 ± 0.3% of total fatty acids, respectively. Overall the lowest level was measured in HM09-W153 seeds consistently so for all locations (between 6.2 and 6.4% of total oil content). The ratio of C18:1 and C18:3 fatty acids was also calculated and presented in Figure 1C: a larger C18:1 to C18:3 ratio reflects a more desirable fatty acid composition because it improves soybean oil stability. Ratios varied across both cultivars and locations; the lowest was 2.5 for Dennison seeds harvested in NW with 19 cm row spacing whereas the top two highest ratios were found to be in HM09-W150 and HM09-W153 at the same site (6.0 and 6.9, respectively). Interestingly, when these latter two cultivars were grown in PC with wide row spacing, the ratio between C18:1 and C18:3 fell to 3.9 and 4.7 for HM09-W150 and HM09-W153, respectively.

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

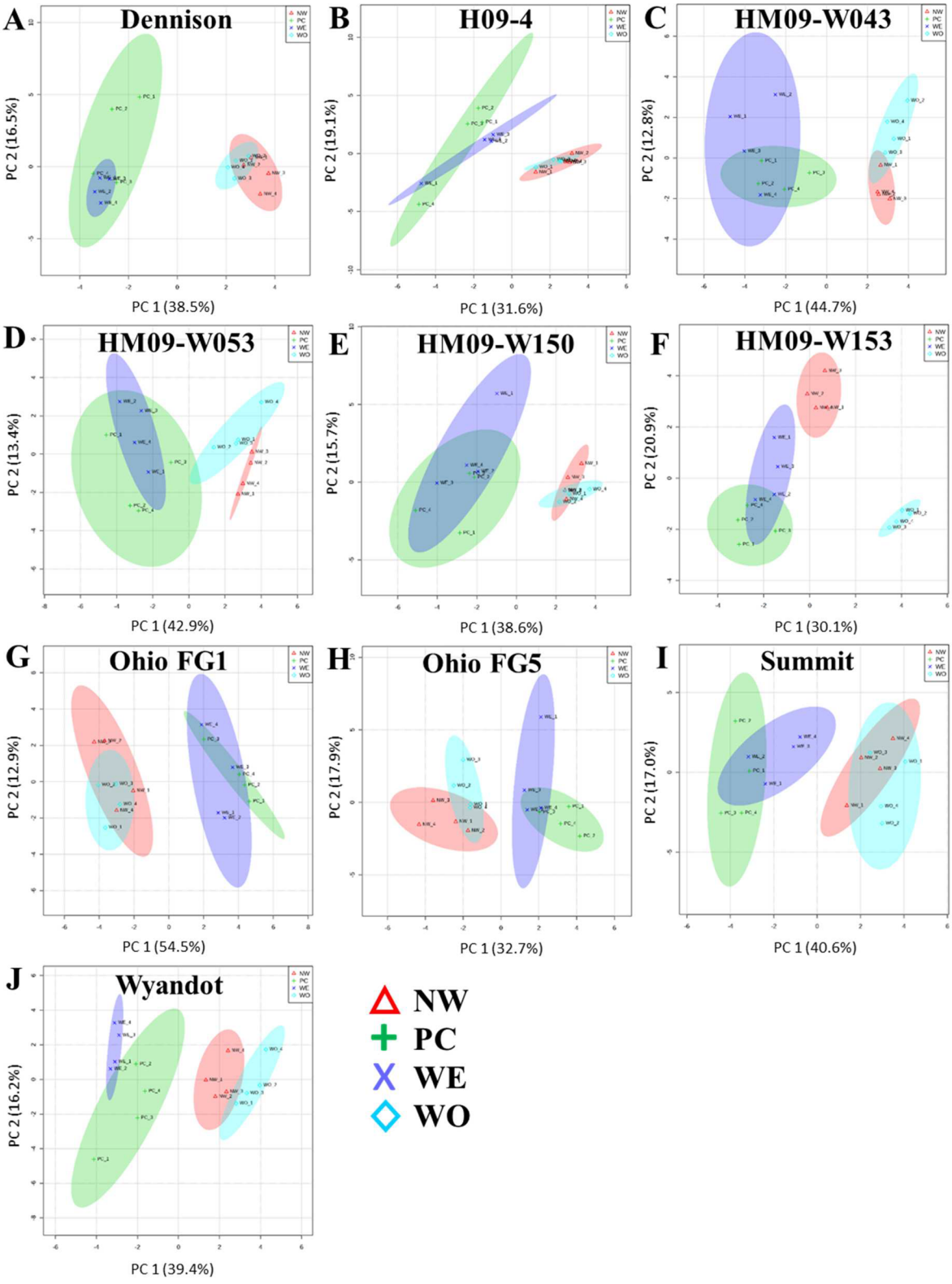

Figure 3. Principal component analysis of the biomass components for each soybean cultivar grown in four different locations. (A) Dennison, (B) H09-4, (C) HM09-W043, (D) HM09-W053, (E) HM09-W150, (F) HM09-W153, (G) Ohio FG1, (H) Ohio FG5, (I) Summit, (J) Wyandot cultivars. The shaded red, green, purple, and blue regions in the PCA plots represent 95% confidence intervals for the four locations: Hoytville (NW), Plain City (PC), South Charleston (WE), and Wooster (WO), respectively.

3.1.3. Seed Protein Content and Amino Acid Composition. In addition, total protein content and amino acid composition were also determined for each of the groups (Figures 1D and E, SI Table S2). The total protein content varied across both cultivar and location; the lowest and highest levels were found to be 32.3 and 45.3% proteins (w/w) for Wyandot harvested in WO with 19 cm row spacing and Ohio FG1 grown in PC with 38 cm row spacing, respectively. Four cultivars, HM09-W153,

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

Ohio FG1, Ohio FG5 and Wyandot, showed significant differences in total protein levels depending on growth location. The PC and WE sites with wide row spacing produced significantly higher protein contents than the other two locations for Wyandot, Ohio FG1 and HM09-W153 seeds. The largest location-dependent variation was measured in HM09-W153 seeds harvested from PC and WO with 44.5 ± 2.4 and 32.2 ± 3.0% proteins (w/w), respectively. The differences measured in total protein can be due to the location and/or the row spacing. High-protein varieties of soybean produce a meal with a greater protein content and nutritional value, and are hence desirable for feeding livestock.[37] Among the 10 cultivars under investigation, H09-4 and HM09-W043 are classified "high-protein" (Table 1). Although their protein content was found to be stable across the four locations (between 41.3 and 44.4, and between 37.2 and 39.3% proteins for H09-4 and HM09-W043, respectively), Ohio FG1 reached higher levels when it was grown in PC.

The composition in proteinogenic amino acids was determined by liquid chromatography tandem mass spectrometry after acid hydrolysis of soybean proteins as described in section 2.5. For all samples, the relative amounts of 16 amino acids were calculated (SI Table S2). Eight of the nine essential amino acids, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, and valine were measurable using this method. There were significant differences between locations regarding essential amino acid content for six of the 10 cultivars: Dennison, HM09−043, HM09-W053, HM09-W150, Ohio FG1 and Summit (SI Table S2). Interestingly, all these cultivars, aside from Dennison, reached higher levels of essential amino acids when they were harvested from two specific sites: NW and WO with narrow row spacing. For instance essential amino acid content for Ohio FG1 at the NW and WO locations was 48.8 ± 0.5 and 49.2 ± 0.9%, respectively, whereas content at the PC and WE locations with wide row spacing was of 45.9 ± 0.7 and 45.9 ± 0.9%, respectively. For all the cultivars under investigation, methionine was found to be the less abundant amino acid (SI Table S2). This essential sulfur-containing amino acid has been shown to be critical for animal health, and current efforts for improving soybean meal aim at increasing its levels.[9,19]

All of the above results indicated that the growing site had an impact on the seed total dry weight and/or composition for the 10 soybean cultivars under investigation. These results are in accordance with previous studies which demonstrated that growing location can considerably modify oil and fatty acid composition as well as protein content.[24,25] The most interesting findings are summarized as follows: (i) although some cultivars, such as Ohio FG1, had a very stable seed size and therefore yield, the biomass composition in terms of percentage and composition of oil and proteins was significantly different across locations, which can have an influence on the market application and value; (ii) knowing that the typical C18:1 content in soybean is approximately 22% of total fatty acids,[13] HM09-W153 reached high levels of oleic acid at the NW site (up to an average of 43.9%), which emphasizes that breeding and/or metabolic engineering must take location into consideration in order to improve soybean oil composition; (iii) locations producing seeds with high protein levels had lower percentage in essential amino acids, and vice versa. This negative correlation between protein content and the percentage in essential amino acids corroborates previous studies,[38,39] and indicate that higher protein levels can be achieved by breeding and/or metabolic engineering in soybean but at the detriment of nutritional quality.

3.2. Correlation and PCA Analyses. *3.2.1. Correlations between Biomass Components.* Pearson-correlation analyses of the biomass components from 10 soybean cultivars grown in four different locations showed that there were strong correlations between several compounds (Figure 2; SI Table S3). First, phenylalanine and tyrosine were found to be strongly positively correlated ($r = 0.90$, $p < 0.001$). This finding can be explained by the fact that these two amino acids are both produced by the shikimate pathway (aka chorismate biosynthesis pathway).[40] Therefore, breeding efforts toward the increase of the essential amino acid phenylalanine would concomitantly enhance the amount of the nonessential amino acid tyrosine. Second, the fatty acid C18:2 strongly correlated ($p < 0.001$) with C16:0 and C18:3 with an r value of 0.58 and 0.63, respectively. Additionally, we measured a strong negative correlation ($p < 0.001$) between these three fatty acids and C18:1 ($-0.74 < r < -0.93$). These results support breeding and metabolic engineering efforts aiming at improving the stability of oil, that is to say increasing oleic acid (C18:1) and lowering linolenic acid (C18:3) contents.[41] Third, among the 10 cultivars selected for this study that were grown in four locations, there was no strong negative correlation between the total fatty acid and total protein, which contrasts with previous studies.[21,38] Based on this single year study, our results indicate that enhancing both protein and oil content in soybean seeds is achievable through breeding and metabolic engineering. Indeed loci have been previously identified which contribute to both protein and oil.[42] In addition, the expression of a fungal diacylglycerol acyltransferase from *Umbelopsis ramanniana* in soybean seeds increased oil from 20.0% to 21.5% with no significant impact on protein content nor yield.[43] These studies provide evidence that seed protein and oil content do not have a requisite pleiotropic relationship and corroborate our findings. Finally, proline was found to positively correlate ($p < 0.001$) with four essential amino acids: isoleucine, leucine, valine, and lysine ($0.55 < r < 0.62$). Interestingly, this group of amino acids had a significant negative correlation ($p < 0.001$) with glutamate, phenylalanine, and tyrosine. These results may guide breeding and metabolic engineering efforts aiming to improve the composition in essential amino acids in soybean proteins.[44] For instance, enhancing proline levels in soybean seems an attractive venue to increase the content in essential amino acids.

3.2.2. Independently of the Soybean Cultivar, The Two Central Locations Group Away from the Two Northern Locations. PCAs for each of the 10 cultivars were examined in order to visualize differences attributed to growing site (Figure 3). For all 10 cultivars, the principal component 1 separated the PC and WE locations, which grouped together, from both NW and WO. The growing sites NW and WO grouped together for all cultivars with the exception of HM09-W053 and HM09-W153 (Figures 3D and F). The trends observed across the 10 cultivars suggest that similarities in biomass composition delineate two larger groups, one consisting of PC and WE and the other of NW and WO. To verify this tendency, a PLS-DA was performed after gathering PC and WE into a "Central Region with 38 cm row spacing" group and NW and WO into a "Northern Region with 19 cm row spacing" (Figure 4A). The component 1 separated the two Northern from the Central locations, which supported the idea that soybeans grown in PC are most similar in biomass composition to those grown in WE,

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

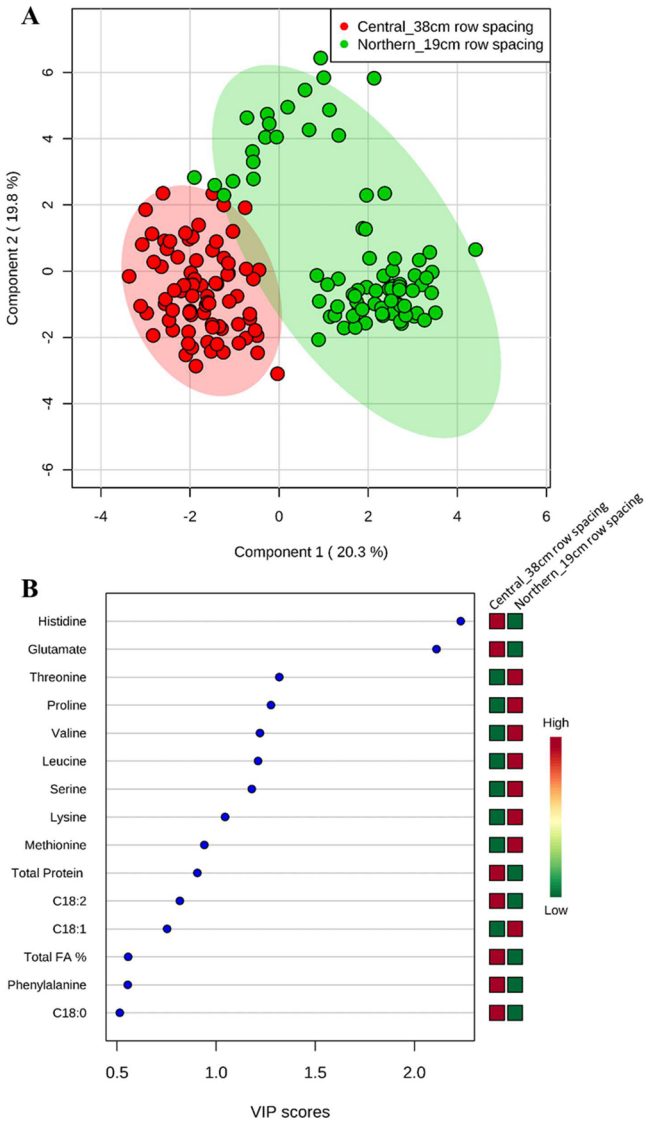

Figure 4. Partial least-squares discriminant analysis of the biomass components for 10 cultivars grown in the Central region with 38 cm row spacing and the Northern region with 19 cm row spacing. (A) PLS-DA plot of biomass components for all cultivars showing separation with respect to the Central locations with 38 cm row spacing and Northern locations with 19 cm row spacing. Shaded red and green regions in the PLS-DA plot represent 95% confidence intervals for the Central and Northern groups, respectively. (B) Variable important in projection (VIP) scores showing the top 15 variables that contribute the most to the PLS-DA model.

and similarly for NW and WO. Interestingly, three of the sites, WE, NW, WO, are Ohio Agricultural Research and Development Center (OARDC) farms and employed the similar agricultural practices whereas PC is non-OARDC. PC and WE grouped together away from NW and WO, which indicates similarities in environmental conditions such as soil type, nutrient availability and weather patterns. The variable importance in projection (VIP) corresponding to the component 1 of the PLS-DA pinpointed the biomass components that contributed the most to the separation between Northern and Central locations (Figure 4B). Interestingly, 10 of the top 15 compounds were found to be amino acids of which seven were essential. Histidine, glutamate, and phenylalanine were more abundant in Central locations with wide row spacing, whereas the level of the other amino acids was higher in the North regions with narrow row spacing. There is some precedence for this finding as the percentage of total protein and oil commonly varies by 1−3% and 1−2%, respectively, over locations, specifically, the nitrogen source and availability have been previously implicated in the level of methionine and cysteine in seeds.[45,46] Although the total protein and fatty acid contents were larger in the Central growing sites with wide row spacing, their "quality" was lower in terms of composition in essential amino acids and C18:1, respectively (Figure 4B). Further field trials would be required to determine if the variations observed between Central and Northern regions are due to the location or the row spacing. Several studies previously analyzed the effect of row spacing on soybean biomass composition and their results were conflicting. For instance, wide rows have been found to be associated with marginally increased[47] or marginally decreased[48] protein levels, to be affiliated with increased oil, or to have[48] inconsistent[49] or insignificant[50] effects. The relatively larger variations observed in our study may be essentially due to the growing environment.

The market value and use of soybean seeds depend upon their biomass composition. Therefore, high-throughput methods to determine the percentage of oil and proteins, and to analyze the fatty acid and proteinogenic amino acid profiles, such as the ones described here, are required to estimate the best use and price. In addition, factors, such as genetics, fertilization, and climatic conditions can modify biomass composition and hence strongly impact soybean value. Such biochemical quantifications will be critical to understand the impact of planting environment as well as their interaction with genetics to better anticipate these effects in the future, which will ultimately contribute to soybean improvement.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.7b01457.

Figure S1. Soil regions of Ohio and growth locations in this study. Table S1. Dry weight of soybean seeds harvested from four different locations. Table S2. Biomass composition of soybean seeds grown in four different locations. Table S3. Pearson correlation between biomass components of 10 soybean cultivars grown in four locations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +001-614-688-7404; fax: +001-614-247-8937; e-mail: alonso.19@osu.edu.

ORCID

Ana Paula Alonso: 0000-0003-4696-1811

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank The Ohio Soybean Council (award OSC 14-2-12) for funding this research. We are grateful to the Targeted Metabolomics Laboratory (metabolomics.osu.edu) for access to the GC-MS and LC-MS/MS equipment funded by the

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

Translational Plant Sciences Targeted Investment in Excellence (TIE).

REFERENCES

(1) Flint-Garcia, S. A. Genetics and consequences of crop domestication. *J. Agric. Food Chem.* **2013**, *61*, 8267−76.

(2) Rincker, K.; Nelson, R.; Specht, J. E.; Sleper, D. A.; Cary, T.; Cianzio, S. R.; Casteel, S.; Conley, S.; Chen, P.; Davis, V.; Fox, C.; Graef, G.; Godsey, C.; Holshouser, D.; Jiang, G.-L.; Kantartzi, S. K.; Kenworthy, W.; Lee, C.; Mian, R.; McHale, L. K.; Naeve, S.; Orf, J. H.; Poysa, V.; Schapaugh, W.; Shannon, G.; Uniatowski, R.; Wang, D.; Diers, B. Genetic improvement of U.S. soyrban in maturity group II, III, and IV. *Crop Sci.* **2014**, *54*, 1419−1432.

(3) Dorff, E., The soybean, agriculture's jack-of-trades, is gaining ground across Canada. *Statistics Canada* **2007**, *96−325-XIE*, 1−13.

(4) Clemente, T. E.; Cahoon, E. B. Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiol.* **2009**, *151*, 1030−40.

(5) Medic, J.; Atkinson, C.; Hurburgh, C. R., Jr Current knowledge in soybean composition. *J. Am. Oil Chem. Soc.* **2014**, *91*, 363−384.

(6) Ash, M.; Knisley, S., Oil Crops Yearbook. *USDA - Economic Research Service* 2016. http://www.ers.usda.gov/data-products/oil-crops-yearbook.aspx (accessed August 11, 2016).

(7) Denicoff, M. R.; Prater, M. E.; Bahizi, P., Soybean Transportation Profile. *United States Department of Agriculture* **2014**. https://www.ams.usda.gov/sites/default/files/media/Soybean%20Transportation%20Profile.pdf.10.9752/TS203.10-2014 (accessed August 11, 2016).

(8) Hartman, G. L.; Miles, M. R.; Frederick, R. D. Breeding for resistance to soybean rust. *Plant Dis.* **2005**, *89*, 664−666.

(9) Krishnan, H. B. Engineering soybean for enhanced sulfur amino acid content. *Crop Sci.* **2005**, *45*, 454−461.

(10) Liu, S.; Kandoth, P. K.; Warren, S. D.; Yeckel, G.; Heinz, R.; Alden, J.; Yang, C.; Jamai, A.; El-Mellouki, T.; Juvale, P. S.; Hill, J.; Baum, T. J.; Cianzio, S.; Whitham, S. A.; Korkin, D.; Mitchum, M. G.; Meksem, K. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* **2012**, *492*, 256−60.

(11) Fernandez-Cornejo, J.; Wechsler, S.; Livingston, M.; Mitchell, L. Genetically engineered crops in the United States. *USDA - Economic Research Service* **2014**, *162*, 1−54.

(12) Van, K.; McHale, L. K. Meta-Analyses of QTLs Associated with Protein and Oil Contents and Compositions in Soybean [Glycine max (L.) Merr.] Seed. *Int. J. Mol. Sci.* **2017**, *18* (6), 1180−1200.

(13) La, T. C.; Pathan, S. M.; Vuong, T.; Lee, J. D.; Scaboo, A. M.; Smith, J. R.; Gillen, A. M.; Gillman, J.; Ellersieck, M. R.; Nguyen, H. T.; Shannon, J. G. Effect of High-Oleic Acid Soybean on Seed Oil, Protein Concentration, and Yield. *Crop Sci.* **2014**, *54*, 2054−2062.

(14) Danaei, G.; Ding, E. L.; Mozaffarian, D.; Taylor, B.; Rehm, J.; Murray, C. J.; Ezzati, M. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med.* **2009**, *6*, e1000058.

(15) Bachlava, E.; Burton, J. W.; Brownie, C.; Wang, S.; Auclair, J.; Cardinal, A. J. Heritability of oleic acid content in soybean seed oil and its genetic correlation with fatty acid and agronomic traits. *Crop Sci.* **2008**, *48*, 1764−1772.

(16) Kim, H. J.; Ha, B. K.; Ha, K. S.; Chae, J. H.; Park, J. H.; Kim, M. S.; Asekova, S.; Shannon, J. G.; Son, C. K.; Lee, J. D. Comparison of a high oleic acid soybean line to cultivated cultivars for seed yield, protein and oil concentrations. *Euphytica* **2015**, *201*, 285−292.

(17) Fehr, W. R.; Welke, G. A.; Hammond, E. G.; Duvick, D. N.; Cianzio, S. R. Inheritance of reduced linolenic acid content in soybean genotypes A16 and A17. *Crop Sci.* **1992**, *32*, 903−906.

(18) Rahman, S. M.; Anai, T.; Kinoshita, T.; Takagi, Y. Combining ability in loci for high oleic and low linolenic acids in soybean. *Crop Sci.* **2001**, *43*, 527−531.

(19) Wang, T. L.; Domoney, C.; Hedley, C. L.; Grusak, M. A. Can. we improve the nutritional quality of legume seeds? *Plant Physiol.* **2003**, *131*, 886−891.

(20) Grieshop, C. M.; Fahey, G. C. Comparison of quality characteristics of soybeans from Brazil, China, and the United States. *J. Agric. Food Chem.* **2001**, *49*, 2669−2673.

(21) Karr-Lilienthal, L. K.; Grieshop, C. M.; Merchen, N. R.; Mahan, D. C.; Fahey, G. C. Chemical composition and protein quality comparisons of soybeans and soybean meals from five leading soybean-producing countries. *J. Agric. Food Chem.* **2004**, *52*, 6193−6199.

(22) Cober, E. R.; Voldeng, H. D. Developing high-protein, high-yield soybean populations and lines. *Crop Sci.* **2000**, *40*, 39−42.

(23) Filho, M. M.; Destro, D.; Miranda, L. A.; Spinosa, W. A.; Carrão-Panizzi, M. C.; Montalván, R. Relationships among Oil Content, Protein Content and Seed Size in Soybeans. *Braz. Arch. Biol. Technol.* **2001**, *44*, 23−32.

(24) Bhardwaj, H. L.; Hamama, A. A.; Rangappa, M.; Joshi, J. M.; Sapra, V. T. Effects of soybean genotype and growing location on oil and fatty acids in tofu. *Plant Foods Hum. Nutr.* **2003**, *58*, 197−205.

(25) Vollman, J.; Fritz, C. N.; Wagentristl, H.; Ruckenbauer, P. Environmental and genetic variation of soybean seed protein content under Central European growing conditions. *J. Sci. Food Agric.* **2000**, *80*, 1300−1306.

(26) St Martin, S. K.; Feller, M. K.; McIntyre, S. A.; Fioritto, R. J.; Dorrance, A. E.; Berry, S. A.; Sneller, C. H. Registration of 'Dennison' Soybean. *J. Plant Regist* **2008**, *2*, 21−21.

(27) Mian, M. A. R.; McHale, L. K.; Li, Z.; Dorrance, A. E. Registration of 'Highpro1' Soybean with Protein and High Yield Developed from a North x South Cross. *J. Plant Regist* **2016**, *11*, 51−54.

(28) StMartin, S. K.; CalipDuBois, A. J.; Fioritto, R. J.; Schmitthenner, A. F.; Min, D. B.; Yang, T. S.; Yu, Y. M.; Cooper, R. L.; Martin, R. J. Registration of 'Ohio FG1' soybean. *Crop Sci.* **1996**, *36*, 813−813.

(29) St Martin, S. K.; Mills, G. R.; Fioritto, R. J.; McIntyre, S. A.; Dorrance, A. E.; Berry, S. A. Registration of 'Ohio FG5' Soybean. *Crop Sci.* **2006**, *46*, 2709−2709.

(30) McHale, L. K.; Feller, M. K.; McIntyre, S. A.; Berry, S. A., St; Martin, S. K.; Dorrance, A. E. Registration of 'Summit,' a High-Yielding Soybean with Race-Specific Resistance to Phytophthora sojae. *J. Plant Regist* **2013**, *7*, 36−41.

(31) Cocuron, J. C.; Anderson, B.; Boyd, A.; Alonso, A. P. Targeted metabolomics of *Physaria fendleri*, an industrial crop producing hydroxy fatty acids. *Plant Cell Physiol.* **2014**, *55*, 620−633.

(32) Tsogtbaatar, E.; Cocuron, J. C.; Sonera, M. C.; Alonso, A. P. Metabolite fingerprinting of pennycress (Thlaspi arvense L.) embryos to assess active pathways during oil synthesis. *J. Exp. Bot.* **2015**, *66*, 4267−4277.

(33) Lowry, O. H.; Rosebrough, N. J.; Farr, A. L.; Randall, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **1951**, *193*, 265−75.

(34) Cocuron, J. C.; Tsogtbaatar, E.; Alonso, A. P. High-throughput quantification of the levels and labeling abundance of free amino acids by liquid chromatography tandem mass spectrometry. *Journal of Chromatography A* **2017**, *1490*, 148−155.

(35) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0–making metabolomics more meaningful. *Nucleic Acids Res.* **2015**, *43*, W251−7.

(36) Zhang, B.; Chen, P. Y.; Florez-Palacios, S. L.; Shi, A. N.; Hou, A. F.; Ishibashi, T. Seed quality attributes of food-grade soybeans from the US and Asia. *Euphytica* **2010**, *173*, 387−396.

(37) Baker, K. M.; Utterback, P. L.; Parsons, C. M.; Stein, H. H. Nutritional value of soybean meal produced from conventional, high-protein, or low-oligosaccharide varieties of soybeans and fed to broiler chicks. *Poult. Sci.* **2011**, *90*, 390−395.

(38) Pathan, S. M.; Vuong, T.; Clark, K.; Lee, J. D.; Shannon, J. G.; Roberts, C. A.; Ellersieck, M. R.; Burton, J. W.; Cregan, P. B.; Hyten, D. L.; Nguyen, H. T.; Sleper, D. A. Genetic Mapping and Confirmation of Quantitative Trait Loci for Seed Protein and Oil Contents and Seed Weight in Soybean. *Crop Sci.* **2013**, *53*, 765−774.

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761

(39) Warrington, C. V.; Abdel-Haleem, H.; Hyten, D. L.; Cregan, P. B.; Orf, J. H.; Killam, A. S.; Bajjalieh, N.; Li, Z.; Boerma, H. R. QTL for seed protein and amino acids in the Benning x Danbaekkong soybean population. *Theor. Appl. Genet.* **2015**, *128*, 839−850.
(40) Tzin, V.; Galili, G. The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in Arabidopsis thaliana. *Arabidopsis Book* **2010**, *8* (e0132), 1−18.
(41) Pham, A. T.; Shannon, J. G.; Bilyeu, K. D. Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *Theor. Appl. Genet.* **2012**, *125*, 503−15.
(42) Hwang, E. Y.; Song, Q. J.; Jia, G. F.; Specht, J. E.; Hyten, D. L.; Costa, J.; Cregan, P. B. A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* **2014**, *15* (1), 1−12.
(43) Lardizabal, K.; Effertz, R.; Levering, C.; Mai, J.; Pedroso, M. C.; Jury, T.; Aasen, E.; Gruys, K.; Bennett, K. Expression of Umbelopsis ramanniana DGAT2A in seed increases oil in soybean. *Plant Physiol.* **2008**, *148*, 89−96.
(44) Ufaz, S.; Galili, G. Improving the content of essential amino acids in crop plants: goals and opportunities. *Plant Physiol.* **2008**, *147*, 954−961.
(45) Paek, N. C.; Imsande, J.; Shoemaker, R. C.; Shibles, R. Nutritional control of soybean seed storage protein. *Crop Sci.* **1997**, *37*, 498−503.
(46) Grabau, L. J.; Blevins, D. G.; Minor, H. C. Stem Infusions Enhanced Methionine Content of Soybean Storage Protein. *Plant Physiol.* **1986**, *82*, 1013−1018.
(47) Dougherty, C. T. Influence of Irrigation, Row Spacing, Plant Population, and Inoculation on Yield of Soybeans in Canterbury. *N. Z. J. Agric. Res.* **1969**, *12*, 367.
(48) Donovan, L. S.; Dimmock, F.; Carson, R. B. Some effects of planting pattern on yield, per cent oil and per cent protein in Mandarin (Ottawa) soybeans. *Can. J. Plant Sci.* **1963**, *43*, 131−140.
(49) Boydak, E.; Alpaslan, M.; Hayta, M.; Gercek, S.; Simsek, M. Seed composition of Soybeans grown in the Harran region of Turkey as affected by row spacing and irrigation. *J. Agric. Food Chem.* **2002**, *50*, 4718−4720.
(50) Al-Tawaha, A. M.; Seguin, P. Seeding date, row spacing, and weed effects on soybean isoflavone concentrations and other seed characteristics. *Can. J. Plant Sci.* **2006**, *86*, 1079−1087.

DOI: 10.1021/acs.jafc.7b01457
J. Agric. Food Chem. 2017, 65, 6753−6761